

Human-centered Hybrid Intelligence

**Provision of human input and its utilization by AI in creative
work and problem-solving**

Cumulative dissertation with the aim of achieving a doctoral degree at the
Faculty of Mathematics, Informatics, and Natural Sciences
Department of Informatics
Universität Hamburg

Submitted by

Izabel Cvetkovic

2025

Hamburg, Germany

Date of disputation:

17.10.2025

Evaluators:

First Evaluator: Eva Bittner

Second Evaluator: Alexander Richter

Abstract

Motivation

Artificial Intelligence (AI) applications are rapidly moving from futuristic concepts to becoming indispensable instruments for organizations navigating volatile markets, geopolitical shocks, and accelerating technological change. AI tools are now on the agenda for the majority of organizations, yet relatively few have embedded them deeply enough to influence processes that drive innovation and competitive advantage, such as creative work and complex problem-solving. While 70% of organizations experiment with generative AI or predictive analytics, only a few have scaled these systems beyond isolated pilots. Most teams still struggle with cognitive overload, unstructured workflows, and collaboration frictions - symptoms that current point solutions cannot address. Closing this gap requires deliberately designed human-AI systems that weave algorithmic assistance into the way diverse groups think, decide, and act together. Hybrid Intelligence Systems (HIS) hold great promise for addressing these challenges by combining the complementary strengths of humans and AI. Humans bring empathy, common sense, and contextual awareness, while AI excels at processing large amounts of data, identifying patterns, and handling repetitive tasks. This collaboration can create opportunities for AI not only to support human efforts but also to act as a creative partner in both individual and team-based problem-solving. Realizing the potential of HIS requires careful consideration of how humans and AI can best work together. A key limitation in current AI development is the lack of a socio-technical perspective. Much of the focus has been on technological capabilities, often neglecting the human and organizational factors critical for effective integration. A human-centered approach to HIS is needed to address questions such as what types of input (e.g., tasks and information) humans are willing to provide to AI, how AI can effectively utilize this input and transform it into outputs that practitioners perceive as relevant, credible, and actionable, and how systems can be designed to align with human needs and goals. By addressing these questions, it is possible to create systems that not only augment human capabilities but also foster trust, agency, and collaboration. This dissertation takes a human-centered approach to designing HIS for creative work and complex problem-solving. By investigating the interplay between human input and AI output, it develops validated design principles and prototypes to bridge the gap between technical feasibility and socio-technical integration. The research contributes to the advancement of human-centered AI by ensuring that HIS empower individuals and teams in ethically sound and practical ways, ultimately supporting innovation and decision-making in modern creative work and complex problem-solving.

Research Design

This cumulative dissertation is anchored in the Design-Science Research (DSR) paradigm and unfolds along its three intertwined cycles—relevance, rigor, and design. Each cycle is

linked to the core research questions, ensuring that theoretical inquiry and practical problem-solving continually inform each other. Motivated by the epistemological interests in understanding the socio-technical interplay and human-centered design of HIS, this dissertation adopts an interdisciplinary, mixed-method approach to define the problem space and generate solutions. Literature reviews were conducted to capture existing scientific knowledge and identify research gaps. These reviews informed both descriptive and prescriptive contributions to the cumulative generation of design knowledge. Qualitative methods, including semi-structured interviews, were used to identify real-world challenges and elicit user requirements. These methods allowed the exploration of the types of human input and the corresponding AI output necessary for an effective HIS design. A combination of qualitative and quantitative methods was used to assess design knowledge, ensuring qualitative depth and quantitative breadth. Qualitative approaches, including interviews and content analysis, provided in-depth insights into user perceptions and the feasibility of the proposed solutions. Quantitative methods, such as surveys and structural equation modelling, were employed to validate theoretical constructs and measure the broader impact of HIS on creative work and complex problem-solving. In addition, primary and secondary data stemming from artefact evaluations served as the basis for further complementary statistical analyses. By integrating these diverse research methods, this dissertation ensures a rigorous and comprehensive approach to addressing the challenges of HIS. Iterating between literature reviews, empirical discovery, artefact construction, and multi-method evaluation keeps this research firmly rooted in rigorous scholarship while driving toward actionable, human-centered solutions. In doing so, the research not only advances theoretical understanding of HIS but also delivers validated design knowledge that practitioners can apply in practice.

Results

This cumulative dissertation provides validated design knowledge for human-centered HIS and their integration in creative and complex problem-solving contexts. The findings are divided into descriptive and prescriptive contributions, addressing theoretical and practical challenges of designing HIS for creative work and problem-solving. The former focus on understanding the prerequisites and opportunities for human input provision and its utilization by AI. These contributions include identifying the conditions under which individuals are inclined to delegate tasks or share information with AI, such as trust, motivation, and task complexity. For instance, a task delegability framework was refined and validated within a specific knowledge work context, providing empirical evidence of factors influencing the decision to delegate tasks to AI. Additionally, a taxonomy of user interactions with large language models (LLMs) was developed, capturing archetypes of user behaviors and highlighting the diverse ways in which AI can support human creativity and problem-solving. The prescriptive knowledge contributions emphasize actionable design knowledge for the development of HIS. Design principles have been identified to optimize task allocation in HIS, structure creative work in teams, and enable effective

human-AI co-creation, *inter alia*. These principles were instantiated in various artifacts, including conversational agents and a hybrid facilitation system. Evaluations demonstrated how HIS can enhance individual and group creative processes, offering scalable solutions for diverse working settings. Furthermore, developed prompt templates for LLMs demonstrated helpfulness, quality, and relevance of LLM-generated outputs, highlighting the effectiveness of prompt engineering in supporting creative work and problem-solving.

Contribution

This dissertation advances the fields of hybrid intelligence, creativity support systems, and human-centered AI by distilling ten publications into a coherent body of validated design knowledge. By synthesizing the findings, it addresses the theoretical and practical challenges of human-centered HIS in creative work and problem-solving contexts. In the context of hybrid intelligence research, the dissertation identifies the prerequisites for human input provision by investigating which tasks and types of information people prefer to delegate or share with AI. The findings show that humans choose to offload routine, time-consuming, and less motivating tasks to AI, while retaining responsibilities involving common sense, human contact, or moral considerations. These insights about what people delegate and why they do so form the basis of a refined task delegability framework, illustrating how task complexity, motivation, and trust influence collaboration preferences. Ethical considerations are also addressed through research on privacy preferences and the conditions under which users are comfortable sharing sensitive personal data with AI systems. These findings highlight the importance of designing HIS that prioritize user trust and transparency while respecting privacy concerns. Building on these descriptive insights, the dissertation highlights opportunities for HIS. It does so by examining different human inputs and corresponding AI outputs that can maximize collaboration efficiency. Additionally, a taxonomy of user interaction patterns with LLMs sheds light on the diverse ways users engage with AI systems, offering an understanding of user behaviors and preferences that can guide the development of adaptive HIS. The dissertation also advances research on creativity support and problem-solving by offering prescriptive design principles and frameworks that guide the development of HIS in these contexts. These include actionable guidelines for designing systems that allocate tasks in HIS according to complementary strengths and support co-creation processes. The presented design knowledge enables the development of HIS that are adaptable to diverse user needs and organizational contexts, enhancing both individual and team-based work processes. Finally, the dissertation introduces overarching design principles that synthesize its central findings.

The results of this dissertation also have practical implications for organizations seeking to integrate HIS into creative work and problem-solving contexts. By adopting the design principles and frameworks presented, organizations can develop systems that empower users to contribute meaningful input to AI while leveraging AI's strengths in data

processing and performing routine tasks. Furthermore, the dissertation demonstrates how HIS can improve facilitation by combining complementary human and AI capabilities to address complex tasks and challenges. The contributions provide a robust foundation for practical implementation, helping organizations understand what types of human-AI configurations they should strive for and thereby advancing the design and integration of HIS to support creative work and problem-solving in diverse organizational settings.

Limitations

This cumulative dissertation is subject to several limitations relating to the definition of the problem space, the production of the solution space, and the evaluation of the generated design knowledge. While efforts were made to ensure methodological rigor and practical relevance, these limitations reflect the inherent constraints of the research context. When defining the problem space, the dissertation relied on qualitative methods, such as semi-structured interviews, to capture real-world challenges and requirements for HIS. While these methods provided valuable insights, the findings are subject to the interviewees' subjective interpretations. Although established coding and analysis techniques were employed to minimize bias and ensure consistency, complete objectivity cannot be guaranteed. Moreover, the selection of participants and scenarios may limit the generalizability of the identified challenges and user needs. Although diverse perspectives were sought to mitigate this risk, the findings remain context-specific and may not apply to other domains or use cases. A notable limitation is the lack of organizational context in some aspects of the research. While the studies focused on understanding user perceptions and HIS outputs, the integration of HIS within specific organizational structures and workflows was not fully addressed. This restricts the projectability of produced design knowledge, i.e., the practical applicability of the findings for organizations looking to deploy HIS in real-world environments, where factors such as team dynamics, organizational culture, and resource constraints play a critical role. Design knowledge was generated by synthesizing findings from the literature to produce the solution space. The selection of relevant literature and research contexts was guided by established methodologies, but it is possible that certain perspectives or emerging insights were overlooked. This selection process could influence the comprehensiveness of the generated design principles and frameworks. The evaluation of the design knowledge also presents limitations. While qualitative methods provided depth, quantitative methods ensured breadth, with approaches ranging from small-scale interviews to large-scale surveys. However, some studies had a limited number of participants, and some evaluations were conducted in controlled or semi-naturalistic settings rather than real-world environments. These constraints may affect the ecological validity and scalability of the findings. Finally, the cumulative nature of the dissertation poses inherent challenges in maintaining coherence across diverse studies. While each paper contributes to the overarching research goal, integrating the findings may result in certain nuances being overlooked. This is

particularly relevant for interdisciplinary topics, where balancing technical, ethical, and socio-technical perspectives is challenging.

Future Research

This dissertation sets out several areas for future research into the design and integration of HIS in creative work and problem-solving contexts. While it advances understanding of HIS and generates actionable design knowledge, further research is required to address the identified limitations and expand the applicability of the findings. In the context of socio-technical integration, future research should focus on incorporating organizational contexts more deeply. This would involve studying how HIS can be adapted to suit different organizational structures, workflows, and team dynamics. For example, research could examine the influence of cultural and resource-related factors on the deployment and effectiveness of HIS. Furthermore, longitudinal studies in real-world organizational settings could shed light on the long-term impact of HIS on creative work and problem-solving processes. Further exploration is also required to validate and refine the design principles and frameworks presented in this dissertation. These contributions should be evaluated with larger and more diverse user groups in naturalistic environments to enhance their ecological validity. Furthermore, interdisciplinary approaches combining perspectives from psychology, sociology, and organizational science could offer a more comprehensive understanding of HIS. Research into ethical considerations and user trust is a crucial area for future work. While this dissertation addresses trust in AI and privacy preferences, further studies could examine how HIS can be developed to adapt dynamically to users' evolving ethical concerns and trust levels. Investigating how HIS can incorporate user feedback in real time to align with their expectations would also be valuable. From a methodological perspective, expanding the range of evaluation methods — for example, by using real-time interaction data or observational studies — could provide richer insights into the dynamics of HIS. Additionally, the role of LLMs in supporting creative work requires further investigation, particularly with regard to understanding how their outputs can be fine-tuned to meet specific user needs. Lastly, the taxonomy of user interactions with LLMs developed in this dissertation provides a foundation for adaptive HIS design. Future studies could expand this taxonomy to encompass a broader range of use cases and interaction styles. They could also examine how adaptive systems can dynamically respond to different user archetypes to ensure inclusivity and personalisation in HIS. Addressing these areas will enable future research to build on the contributions of this dissertation and further advance the development of human-centered HIS that empower users and align with the complexities of modern creative work and complex problem-solving.

Keywords:

Creative work, Complex-problem solving, Conversational Agents, Large Language Models, Human-centered AI, Hybrid Intelligence Systems

Kurzfassung

Motivation

Anwendungen der künstlichen Intelligenz (KI) entwickeln sich rasch von futuristischen Konzepten zu unverzichtbaren Instrumenten für Unternehmen, die mit unbeständigen Märkten, geopolitischen Schocks und einem beschleunigten technologischen Wandel zurechtkommen müssen. KI-Tools stehen heute auf der Agenda der meisten Unternehmen, doch nur relativ wenige haben sie tief genug eingebettet, um Prozesse zu beeinflussen, die Innovation und Wettbewerbsvorteile gewährleisten, wie z. B. kreative Arbeit und das Lösen von komplexen Problemen. Während 70 % der Unternehmen mit generativer KI oder prädiktiver Analytik experimentieren, haben nur wenige diese Systeme über isolierte Pilotprojekte hinaus skaliert. Die meisten Teams sind immer noch mit kognitiver Überlastung, unstrukturierten Arbeitsabläufen und Reibungsverlusten bei der Zusammenarbeit konfrontiert - Symptome, die sich mit den derzeitigen Insellösungen nicht beheben lassen. Um diese Lücke zu schließen, sind bewusst konzipierte Mensch-KI-Systeme erforderlich, die algorithmische Unterstützung in die Art und Weise einbinden, wie verschiedene Gruppen gemeinsam denken, entscheiden und handeln. Hybride Intelligenzsysteme (HIS) sind vielversprechend bei der Bewältigung dieser Herausforderungen, da sie die sich ergänzenden Stärken von Menschen und KI kombinieren. Der Mensch bringt Einfühlungsvermögen, gesunden Menschenverstand und kontextbezogenes Bewusstsein mit, während die KI sich durch die Verarbeitung großer Datenmengen, die Erkennung von Mustern und die Bewältigung sich wiederholender Aufgaben auszeichnet. Durch diese Zusammenarbeit kann die KI nicht nur die menschliche Arbeit unterstützen, sondern auch als kreativer Partner bei der individuellen und teambasierten Problemlösung fungieren. Um das Potenzial von HIS auszuschöpfen, muss sorgfältig geprüft werden, wie Menschen und KI am besten zusammenarbeiten können. Eine wesentliche Einschränkung in der derzeitigen KI-Entwicklung ist das Fehlen einer soziotechnischen Perspektive. Der Schwerpunkt liegt auf den technologischen Fähigkeiten und vernachlässigt oft die menschlichen und organisatorischen Faktoren, die für eine wirksame Integration entscheidend sind. Ein menschenzentrierter Ansatz für KI ist erforderlich, um Fragen zu beantworten wie die, welche Arten von Eingaben (z. B. Aufgaben und Informationen) Menschen bereit sind, der KI zur Verfügung zu stellen, wie KI diese Eingaben effektiv nutzen und in Ergebnisse umwandeln kann, die von Praktikern als relevant, glaubwürdig und umsetzbar wahrgenommen werden, und wie Systeme so gestaltet werden können, dass sie mit menschlichen Bedürfnissen und Zielen übereinstimmen. Durch die Beantwortung dieser Fragen ist es möglich, Systeme zu schaffen, die nicht nur die menschlichen Fähigkeiten erweitern, sondern auch Vertrauen, Handlungsfähigkeit und Zusammenarbeit fördern. Diese Dissertation verfolgt einen menschenzentrierten Ansatz bei der Entwicklung von HIS für kreative Arbeit und das Lösen von komplexen Problemen. Durch die Untersuchung des Zusammenspiels zwischen

menschlichen Eingaben und KI-Ausgaben werden validierte Designprinzipien und Prototypen entwickelt, um die Kluft zwischen technischer Machbarkeit und soziotechnischer Integration zu überbrücken. Die Forschung trägt dazu bei, die menschenzentrierte KI voranzubringen, indem sie sicherstellt, dass HIS Einzelpersonen und Teams auf ethisch vertretbare und praktische Weise befähigen und letztlich Innovation sowie Entscheidungsfindung in der modernen Kreativ- und Problemlösungsarbeit unterstützen.

Forschungsdesign

Diese kumulative Dissertation folgt dem Paradigma der Design Science Research (DSR) und entfaltet sich entlang ihrer drei miteinander verknüpften Zyklen: Relevanz, Strenge und Gestaltung. Jeder Zyklus ist mit den zentralen Forschungsfragen verknüpft, um sicherzustellen, dass sich theoretische Untersuchungen und praktische Problemlösungen kontinuierlich gegenseitig beeinflussen. Motiviert durch das erkenntnistheoretische Interesse am Verständnis des sozio-technischen Zusammenspiels und der menschenzentrierten Gestaltung von HIS, wird in dieser Dissertation ein interdisziplinärer, multimethodischer Ansatz gewählt, um den Problemraum zu definieren und Lösungen zu entwickeln. Um das vorhandene wissenschaftliche Wissen zu erfassen und Forschungslücken zu identifizieren, wurden Literaturrecherchen durchgeführt. Diese Übersichten lieferten sowohl deskriptive als auch präskriptive Beiträge zur kumulativen Generierung von Designwissen. Um reale Herausforderungen zu identifizieren und Nutzeranforderungen zu sammeln, wurden qualitative Methoden, einschließlich semistrukturierter Interviews, eingesetzt. Diese Methoden ermöglichten die Erkundung der Arten von menschlichen Eingaben und der entsprechenden KI-Ausgaben, die für ein effektives HIS-Design erforderlich sind. Bei der Bewertung des Designwissens kam eine Kombination aus qualitativen und quantitativen Methoden zum Einsatz, die qualitative Tiefe und quantitative Breite miteinander verknüpften. Qualitative Ansätze, einschließlich Interviews und Inhaltsanalysen, lieferten tiefe Einblicke in die Wahrnehmungen der Nutzer und die Machbarkeit der vorgeschlagenen Lösungen. Quantitative Methoden, wie Umfragen und Strukturgleichungsmodelle, wurden eingesetzt, um theoretische Konstrukte zu validieren und die breiteren Auswirkungen von HIS auf kreative Arbeit und das Lösen von komplexen Problemen zu messen. Darüber hinaus dienten Primär- und Sekundärdaten aus Artefaktevaluierungen als Grundlage für weitere ergänzende statistische Analysen. Durch die Integration dieser verschiedenen Forschungsmethoden gewährleistet diese Dissertation einen rigorosen und umfassenden Ansatz zur Bewältigung der Herausforderungen von HIS. Der Wechsel zwischen Literaturrecherche, empirischer Entdeckung, Artefaktkonstruktion und Multimethoden-Evaluierung sorgt dafür, dass diese Forschung fest in der Wissenschaft verwurzelt ist und gleichzeitig zu umsetzbaren, menschenzentrierten Lösungen führt. Auf diese Weise erweitert die Forschung nicht nur das theoretische Verständnis von HIS, sondern liefert auch validiertes Designwissen, das Praktiker in der Praxis anwenden können.

Ergebnisse

Diese kumulative Dissertation liefert validiertes Designwissen für menschenzentrierte HIS und deren Integration in kreative und komplexe Problemlösungskontexte. Die Ergebnisse sind in deskriptive und präskriptive Beiträge unterteilt, die sich mit den theoretischen und praktischen Herausforderungen der Gestaltung von HIS für kreative Arbeit und Problemlösung befassen. Erstere konzentrieren sich auf das Verständnis der Voraussetzungen und Möglichkeiten für die Bereitstellung menschlichen Eingaben und dessen Nutzung durch KI. Zu diesen Beiträgen gehört die Ermittlung der Bedingungen, unter denen Menschen geneigt sind, Aufgaben zu delegieren oder Informationen mit KI zu teilen, wie etwa Vertrauen, Motivation und Aufgabenkomplexität. So wurde beispielsweise ein Rahmen für die Delegierbarkeit von Aufgaben verfeinert und in einem spezifischen Wissensarbeitskontext validiert, der empirische Belege für Faktoren liefert, die die Entscheidung, Aufgaben an KI zu delegieren, beeinflussen. Darüber hinaus wurde eine Taxonomie der Benutzerinteraktionen mit großen Sprachmodellen (LLMs) entwickelt, die Archetypen des Benutzerverhaltens erfasst und die vielfältigen Möglichkeiten aufzeigt, wie KI die menschliche Kreativität und Problemlösung unterstützen kann. Die präskriptiven Wissensbeiträge betonen umsetzbares Designwissen für die Entwicklung von HIS. Es wurden Gestaltungsprinzipien aufgestellt, um die Aufgabenverteilung im HIS zu optimieren, die kreative Arbeit in Teams zu strukturieren und eine effektive KI-Ko-Kreation zu ermöglichen, unter anderem. Diese Prinzipien wurden in verschiedenen Artefakten instanziiert, darunter Konversationsagenten und ein hybrides Moderationssystem. Die Auswertungen haben gezeigt, wie HIS die kreativen Prozesse von Einzelpersonen und Gruppen verbessern kann und skalierbare Lösungen für verschiedene Arbeitsumgebungen bietet. Darüber hinaus zeigten die entwickelten Prompt-Vorlagen für LLMs, dass die von LLMs generierten Ergebnisse hilfreich, qualitativ hochwertig und relevant sind, was die Effektivität von Prompt-Engineering bei der Unterstützung kreativer Arbeit und Problemlösung unterstreicht.

Beitrag

Diese Dissertation bringt die Bereiche hybride Intelligenz, kreativitätsunterstützende Systeme und menschenzentrierte KI voran, indem sie zehn Publikationen zu einem kohärenten Korpus validierten Designwissens verdichtet. Durch die Synthese der Ergebnisse werden theoretische und praktische Herausforderungen der menschenzentrierten KI in kreativen Arbeits- und Problemlösungskontexten angesprochen. Im Kontext der Forschung zu hybrider Intelligenz identifiziert die Dissertation die Voraussetzungen für die Bereitstellung menschlicher Eingaben, indem sie untersucht, welche Aufgaben und Arten von Informationen Menschen lieber an KI delegieren oder mit ihr teilen. Die Ergebnisse zeigen, dass Menschen sich dafür entscheiden, routinemäßige, zeitaufwändige und weniger motivierende Aufgaben an KI zu delegieren, während sie Aufgaben, die den gesunden Menschenverstand, den menschlichen Kontakt oder

moralische Erwägungen betreffen, beibehalten. Diese Erkenntnisse darüber, was Menschen delegieren und warum sie dies tun, bilden die Grundlage eines verfeinerten Rahmens für die Delegierbarkeit von Aufgaben, der zeigt, wie Aufgabenkomplexität, Motivation und Vertrauen die Präferenzen bei der Zusammenarbeit beeinflussen. Ethische Erwägungen werden auch durch die Erforschung der Datenschutzpräferenzen und der Bedingungen, unter denen die Benutzer sensible persönliche Daten mit KI-Systemen teilen, angesprochen. Die Ergebnisse zeigen, wie wichtig es ist, ein HIS zu entwerfen, das das Vertrauen und die Transparenz der Nutzer in den Vordergrund stellt und gleichzeitig die Belange des Datenschutzes berücksichtigt. Aufbauend auf diesen deskriptiven Erkenntnissen zeigt die Dissertation Möglichkeiten für HIS auf. Dazu werden verschiedene menschliche Eingaben und entsprechende KI-Ausgaben untersucht, die die Effizienz der Mensch-KI Zusammenarbeit maximieren können. Darüber hinaus beleuchtet eine Taxonomie der Interaktionsmuster von Nutzern mit LLMs die verschiedenen Arten, wie Nutzer mit KI-Systemen interagieren, und bietet so ein Verständnis von Nutzerverhalten und -präferenzen, das die Entwicklung adaptiver HIS anleiten kann. Die Dissertation bringt auch die Forschung im Bereich der Kreativitätsunterstützung und Problemlösung voran, indem sie präskriptive Designprinzipien und Rahmenwerke anbietet, die die Entwicklung von HIS in diesen Kontexten leiten. Dazu gehören umsetzbare Richtlinien für die Gestaltung von Systemen, die Aufgaben in HIS entsprechend komplementärer Stärken zuweisen und Mitgestaltung unterstützen. Das vorgestellte Designwissen ermöglicht die Entwicklung von HIS, die an unterschiedliche Nutzerbedürfnisse und organisatorische Kontexte anpassbar sind und sowohl individuelle als auch teambasierte Arbeitsprozesse verbessern. Abschließend werden in der Dissertation übergreifende Gestaltungsprinzipien vorgestellt, die die zentralen Erkenntnisse zusammenfassen.

Die Ergebnisse dieser Dissertation haben auch praktische Auswirkungen für Organisationen, die HIS in kreative Arbeits- und Problemlösungskontexte integrieren wollen. Durch die Anwendung der vorgestellten Gestaltungsprinzipien und Rahmenbedingungen können Unternehmen Systeme entwickeln, die es den Benutzern ermöglichen, sinnvolle Beiträge zur KI zu leisten und gleichzeitig die Stärken der KI bei der Datenverarbeitung und der Ausführung von Routineaufgaben zu nutzen. Darüber hinaus zeigt die Dissertation, wie HIS durch die Kombination sich ergänzender menschlicher und KI Fähigkeiten die Bewältigung komplexer Aufgaben und Herausforderungen erleichtern kann. Die Beiträge bieten eine solide Grundlage für die praktische Umsetzung und helfen Organisationen zu verstehen, welche Art von Mensch-KI-Konfigurationen sie anstreben sollten. Dadurch wird die Entwicklung und Integration von HIS zur Unterstützung von kreativer Arbeit und Problemlösung in verschiedenen organisatorischen Umgebungen vorangetrieben.

Limitationen

Diese kumulative Dissertation unterliegt mehreren Beschränkungen in Bezug auf die Definition des Problemraums, die Erstellung des Lösungsraums und die Bewertung des generierten Designwissens. Obwohl Anstrengungen unternommen wurden, um methodologische Strenge und praktische Relevanz zu gewährleisten, spiegeln diese Einschränkungen die inhärenten Beschränkungen des Forschungskontexts wider. Bei der Definition des Problemraums stützte sich die Dissertation auf qualitative Methoden, wie z. B. semistrukturierte Interviews, um reale Herausforderungen und Anforderungen an HIS zu erfassen. Diese Methoden lieferten zwar wertvolle Einblicke, die Ergebnisse unterliegen jedoch den subjektiven Interpretationen der Befragten. Obwohl etablierte Kodierungs- und Analysetechniken eingesetzt wurden, um Verzerrungen zu minimieren und Konsistenz zu gewährleisten, kann eine vollständige Objektivität nicht garantiert werden. Außerdem kann die Auswahl der Teilnehmer und Szenarien die Verallgemeinerbarkeit der ermittelten Herausforderungen und Nutzerbedürfnisse einschränken. Obwohl verschiedene Perspektiven gesucht wurden, um dieses Risiko zu mindern, bleiben die Ergebnisse kontextspezifisch und sind möglicherweise nicht auf andere Bereiche oder Anwendungsfälle übertragbar. Eine bemerkenswerte Einschränkung ist das Fehlen eines organisatorischen Kontexts in einigen Aspekten der Forschung. Während sich die Studien auf das Verständnis der Benutzerwahrnehmung und der HIS-Ergebnisse konzentrierten, wurde die Integration von HIS in spezifische Organisationsstrukturen und Arbeitsabläufe nicht vollständig berücksichtigt. Dies schränkt die praktische Anwendbarkeit des gewonnenen Gestaltungswissens ein, d.h. die praktische Anwendbarkeit der Ergebnisse für Organisationen, die HIS in realen Umgebungen einsetzen wollen, wo Faktoren wie Teamdynamik, Organisationskultur und Ressourcenbeschränkungen eine entscheidende Rolle spielen. Das Designwissen wurde durch die Synthese der Erkenntnisse aus der Literatur generiert, um den Lösungsraum zu schaffen. Die Auswahl der relevanten Literatur und des Forschungskontextes wurde von etablierten Methoden geleitet, aber es ist möglich, dass bestimmte Perspektiven oder neue Erkenntnisse übersehen wurden. Dieser Auswahlprozess könnte die Vollständigkeit der generierten Designprinzipien und -rahmen beeinflussen. Die Bewertung des Designwissens weist ebenfalls Einschränkungen auf. Während die qualitativen Methoden für Tiefe sorgten, gewährleisteten die quantitativen Methoden Breite, wobei die Ansätze von kleinen Interviews bis zu groß angelegten Umfragen reichten. Einige Studien hatten jedoch nur eine begrenzte Anzahl von Teilnehmern, und einige Bewertungen wurden in kontrollierten oder semi-naturalistischen Umgebungen und nicht in realen Umgebungen durchgeführt. Diese Einschränkungen können die ökologische Validität und Skalierbarkeit der Ergebnisse beeinträchtigen. Schließlich stellt der kumulative Charakter der Dissertation eine inhärente Herausforderung dar, wenn es darum geht, die Kohärenz zwischen den verschiedenen Studien zu wahren. Zwar trägt jede Arbeit zu dem übergeordneten Forschungsziel bei, doch kann die Integration der Ergebnisse dazu führen, dass bestimmte Nuancen übersehen werden. Dies gilt insbesondere für interdisziplinäre Themen, bei denen die Abwägung

zwischen technischen, ethischen und soziotechnischen Perspektiven eine Herausforderung darstellt.

Ausblick

In dieser Dissertation werden mehrere Bereiche für künftige Forschung zur Gestaltung und Integration von HIS in kreativen Arbeits- und Problemlösungskontexten aufgezeigt. Während sie das Verständnis von HIS vorantreibt und umsetzbares Designwissen generiert, sind weitere Forschungen erforderlich, um die identifizierten Einschränkungen anzugehen und die Anwendbarkeit der Ergebnisse zu erweitern. Im Kontext der soziotechnischen Integration sollte sich die künftige Forschung auf eine stärkere Einbeziehung des organisatorischen Kontexts konzentrieren. Dies würde bedeuten, zu untersuchen, wie HIS an unterschiedliche Organisationsstrukturen, Arbeitsabläufe und Teamdynamiken angepasst werden können. So könnte beispielsweise der Einfluss kultureller und ressourcenbezogener Faktoren auf den Einsatz und die Effektivität von HIS untersucht werden. Darüber hinaus könnten Langzeitstudien in realen Unternehmenskontexten Aufschluss über die langfristigen Auswirkungen von HIS auf kreative Arbeit und Problemlösungsprozesse geben. Weitere Untersuchungen sind auch erforderlich, um die in dieser Dissertation vorgestellten Gestaltungsprinzipien und Rahmenbedingungen zu validieren und zu verfeinern. Diese Beiträge sollten mit größeren und vielfältigeren Benutzergruppen in naturalistischen Umgebungen evaluiert werden, um ihre ökologische Validität zu erhöhen. Darüber hinaus könnten interdisziplinäre Ansätze, die Perspektiven aus der Psychologie, Soziologie und Organisationswissenschaft kombinieren, ein umfassenderes Verständnis von HIS ermöglichen. Die Erforschung ethischer Überlegungen und des Vertrauens der Nutzer ist ein wichtiger Bereich für künftige Arbeiten. Während sich diese Dissertation mit dem Vertrauen in KI und den Präferenzen für den Schutz der Privatsphäre befasst, könnten weitere Studien untersuchen, wie HIS so entwickelt werden können, dass sie sich dynamisch an die sich entwickelnden ethischen Bedenken und Vertrauensniveaus der Nutzer anpassen. Es wäre auch wertvoll zu untersuchen, wie HIS das Feedback der Benutzer in Echtzeit einbeziehen können, um sich an ihre Erwartungen anzupassen. Aus methodischer Sicht könnte eine Erweiterung des Spektrums an Evaluierungsmethoden - z. B. durch die Verwendung von Echtzeit-Interaktionsdaten oder Beobachtungsstudien - umfassendere Einblicke in die Dynamik von HIS liefern. Darüber hinaus muss die Rolle der LLM bei der Unterstützung der kreativen Arbeit weiter untersucht werden, insbesondere im Hinblick auf die Frage, wie ihre Ergebnisse auf die spezifischen Bedürfnisse der Nutzer abgestimmt werden können. Schließlich bietet die in dieser Dissertation entwickelte Taxonomie der Benutzerinteraktionen mit LLMs eine Grundlage für das adaptive HIS-Design. Zukünftige Studien könnten diese Taxonomie erweitern, um ein breiteres Spektrum von Anwendungsfällen und Interaktionsstilen zu erfassen. Sie könnten auch untersuchen, wie adaptive Systeme dynamisch auf verschiedene Benutzerarchetypen reagieren können, um Inklusivität und Personalisierung in HIS zu gewährleisten. Die Beschäftigung mit diesen

Bereichen wird es der zukünftigen Forschung ermöglichen, auf den Beiträgen dieser Dissertation aufzubauen und die Entwicklung von menschenzentrierten HIS weiter voranzutreiben, die die Nutzer befähigen und sich an die Komplexität der modernen Kreativ- und Problemlösungsarbeit anpassen.

Keywords: Kreative Arbeit, Lösen komplexer Probleme, konversationelle Agenten, große Sprachmodelle, menschenzentrierte KI, hybride Intelligenzsysteme

Table of Contents

I. List of Figures	XXIII
II. List of Tables	XXV
III. List of Abbreviations	XXVII
1. Introduction	1
1.1 Motivation and Problem Statement	1
1.2 Research Goal and Research Questions	4
1.3 Thesis Outline.....	7
2 Theoretical Foundations	10
2.1 Human-centered AI	10
2.1.1 User.....	12
2.1.2 Technology	15
2.1.3 Ethics	22
2.2 Provision of human input and its utilization by AI	25
2.2.1 Input	26
2.2.2 Process	27
2.2.3 Output	28
3 Research Design	31
3.1 Research Paradigm.....	31
3.2 Research Strategy	32
3.2.1 Relevance.....	35
3.2.2 Rigor	36
3.2.3 Design	37
3.3 Knowledge Utilization and Contribution	40
3.4 Research Methods.....	41
3.4.1 Artifact Design.....	42
3.4.2 Artifact Evaluation.....	44
4 Publications	49
5 Theoretical Contributions.....	69
5.1 Provision of human input and its utilization by AI	69
5.1.1 Prerequisites for human-centered input provision	69

5.1.2 Opportunities for effective human input utilization by AI	73
5.2 Human-centered AI in creative work and complex problem-solving.....	76
5.2.1 Design and development of human-centered AI for creative work and complex problem-solving	77
5.2.2 Impact of human-centered AI integration on creative work and complex problem-solving	83
5.3 Overall Theoretical Contribution.....	87
6 Practical Contributions	93
6.1 Integration of AI in creative work and complex problem-solving	93
6.2 Design knowledge for human-centered HIS.....	94
6.2.1 Facilitating creative work	94
6.2.2 Customizing LLMs for problem-solving.....	94
6.3 Responsible privacy practices.....	95
7 Limitations	98
8 Implications for Future Research	101
8.1 Extending evaluation contexts and methodologies	101
8.2 Contextualizing AI design knowledge in broader domains	101
8.3 Evolving role of LLMs.....	103
9 Hybride Wissensarbeit	106
9.1 Von IT-gestützten Wissensarbeitssystemen hin zu Hybriden Wissensarbeitssystemen	106
9.2 Merkmale und Gestaltungsdimensionen Hybrider Wissensarbeitssysteme 109	
9.3 Anwendungsbeispiel: Kollaboratives Schreiben im Journalismus	111
9.4 Forschungsziele und Ausblick	111
9.5 Literatur	112
10 Task Delegability to AI: Evaluation of a Framework in a Knowledge Work Context.....	114
10.1 Introduction.....	114
10.2 Theoretical Background.....	116
10.2.1 Human-AI Collaboration	116
10.2.2 Task allocation in Human-AI Collaboration	116

10.2.3	Task Delegability	117
10.3	Method	119
10.3.1	Questionnaire	119
10.3.2	Subjects	120
10.3.3	Data Analysis	121
10.4	Results	121
10.4.1	Confirmatory Factor Analysis	121
10.4.2	Linear Regression	122
10.4.3	Analysis of covariance	123
10.5	Discussion	124
10.5.1	Theoretical implications	124
10.5.2	Design implications	125
10.5.3	Limitations and future research	126
10.6	References	127
11	What should AI know? Information Disclosure in Human-AI Collaboration.	132
11.1	Introduction	132
11.2	Theoretical Background	134
11.2.1	Design Thinking and Artificial Intelligence	134
11.2.2	Information Disclosure	135
11.3	Preliminary Results and Proposed Research Approach	137
11.4	Expected Contribution and Conclusion	141
11.5	References	142
12	A User-centric Taxonomy for Conversational Generative Language Models	150
12.1	Introduction	150
12.2	Related Research	152
12.3	Research Methodology	154
12.3.1	Data Collection with a Prompt-a-thon	154
12.3.2	Development of Taxonomy	155
12.3.3	Clustering and Development of Archetypes	158
12.4	Results	158

12.4.1	User-centric Taxonomy of GLM Usage	158
12.4.2	Clustering and Derived Operator Archetypes.....	163
12.5	Discussion	166
12.5.1	Theoretical Contribution.....	168
12.5.2	Practical Contribution	169
12.6	Limitations and Further Research	170
12.7	Conclusion	171
12.8	References.....	171
13	Conversational Agent as a Black Hat: Can Criticizing Improve Idea Generation?	176
	Abstract.....	176
13.1	Introduction.....	176
13.2	Related Work	178
13.2.1	Conversational Agents.....	178
13.2.2	Design thinking.....	178
13.2.3	Conversational Agents in Design Thinking – With AI to better ideas	179
13.3	Method	180
13.3.1	Design Requirements & corresponding Design Principles.....	180
13.3.2	Study design.....	182
13.3.3	Implementation of the Criticism	184
13.4	Results	185
13.4.1	User Evaluation of the prototype	185
13.4.2	Impact on criticism on Idea Generation.....	185
13.4.3	Impact of criticism on the perception of the CA	186
13.5	Discussion	187
13.5.1	User evaluation of the prototype.....	188
13.5.2	Impact of criticism on Idea Generation	188
13.5.3	Impact of criticism on the perception of the CA	189
13.6	Contributions and limitations.....	190
13.7	Recommendations for future research	191
13.8	References.....	191

14 Augmented Facilitation: Designing a multi-modal Conversational Agent for Group Ideation.....	194
14.1 Introduction.....	194
14.2 Theoretical Background.....	196
14.2.1 Human-AI Collaboration	196
14.2.2 Facilitation & Ideation	196
14.2.3 Conversational Agents as (Co-)Facilitators.....	198
14.3 Design Science Research Approach	199
14.3.1 Problem Understanding	199
14.3.2 Objectives of a Solution.....	200
14.3.3 Design & Development	203
14.3.4 Demonstration and Evaluation.....	204
14.4 Results.....	205
14.5 Discussion	208
14.6 Contribution and Conclusion	210
14.7 References.....	212
15 Designing a Hybrid Intelligence System for the Facilitation of Group Collaboration	218
15.1 Introduction.....	218
15.2 Theoretical Background & Related Work	220
15.2.1 Facilitation of Collaboration	220
15.2.2 Hybrid Intelligence	220
15.3 Method	221
15.3.1 Objectives of a Solution.....	222
15.3.2 Development and Demonstration	226
15.3.3 Evaluation	227
15.4 Results.....	228
15.4.1 Design Requirements and Principles	228
15.4.2 User evaluation results.....	230
15.4.3 Evaluation of the chat logs and working outputs.....	232
15.5 Discussion	232
15.6 References.....	234

15.7	Appendix 1: Task Allocation	239
15.8	Appendix 2: Search terms and databases.....	239
15.9	Appendix 3: Design Requirements.....	240
15.10	Appendix 4: Questionnaire and Interview Guideline.....	241
16	Human-AI Collaboration in conceptualizing Design Science Research Studies: Perceived Helpfulness of Generative Language Model's Suggestions	244
16.1	Introduction.....	244
16.2	Background	246
16.2.1	Design science research tool support.....	246
16.2.2	Generative language models.....	247
16.3	Methodology	249
16.3.1	Study approach	249
16.3.2	Survey	252
16.4	Results.....	254
16.5	Discussion	257
16.5.1	Perception of AI suggestions	257
16.5.2	Usage of AI suggestion in a DSR support tool.....	258
16.5.3	Towards a DSR support tool with embedded, generative AI capabilities 260	
16.6	Limitations.....	261
16.7	Conclusion and outlook.....	262
16.8	References.....	263
17	The More Is Not the Merrier: Effects of Prompt Engineering on the Quality of Ideas Generated By GPT-3.....	268
17.1	Introduction.....	268
17.2	Background	270
17.2.1	Idea generation support.....	270
17.2.2	Generative language models (GLM)	270
17.3	Method	271
17.3.1	Prompt template development	271
17.3.2	Idea generation.....	275

17.3.3	Idea evaluation.....	276
17.3.4	Data analysis.....	276
17.4	Results.....	277
17.5	Discussion	278
17.5.1	Answer to the research question	278
17.5.2	Contribution to theory.....	279
17.5.3	Implications for practice	280
17.6	Limitations and outlook	280
17.7	Conclusion	282
17.8	References.....	283
18	Quid pro Quo: Information disclosure for AI feedback in Human-AI collaboration.....	287
18.1	Introduction.....	287
18.2	Theoretical background	290
18.2.1	Human-AI collaboration	290
18.2.2	Information disclosure	293
18.3	Hypothesis development.....	296
18.4	Method	301
18.4.1	Vignette development and validation	303
18.4.2	Instruments.....	305
18.4.3	Main analysis	306
18.5	Results.....	308
18.5.1	Robustness analyses.....	316
18.6	Discussion	317
18.6.1	Contributions	317
18.6.2	Limitations & Outlook.....	321
18.7	Conclusion	323
18.8	Appendix A. Supplementary data	324
18.9	Appendix.....	325
18.9.1	Vignettes	325
18.9.2	SEM Outputs – Correlation matrices.....	326
18.10	References.....	328

19 References.....338

I. List of Figures

Figure 1. Assignment of RQs to DSR cycles. Adapted from Hevner (2007).....	6
Figure 2. Hybrid Intelligence with the core components addressed in this thesis highlighted in the grey box. Adapted from Oeste-Reiß et al. (2021)	16
Figure 3. DSR strategy.....	40
Figure 4. Utilization of and contribution to the design knowledge. Adapted from Drechsler & Hevner (2018).	41
Figure 5. Original framework (left) vs new framework suitable for postgraduate research (right). Adapted from Cvetkovic & Bittner (2022).....	70
Figure 6. Proposed AI outputs for different human inputs: speech shares, cognitive load, and breaking rules of creative teamwork (Cvetkovic et al., 2022, 2025)	72
Figure 7. Relationship between trust in AI and Task Delegability to AI. Adapted from Cvetkovic & Bittner (2022)	73
Figure 8. A User-centric Taxonomy of GLM Usage (von Brackel-Schmidt et al., 2023).....	76
Figure 9. Design knowledge for a CA for individual ideation	77
Figure 10. Design knowledge for a multi-modal CA for the facilitation of creative teamwork	78
Figure 11. Instantiation of a multi-modal CA for the facilitation of creative work in teams	79
Figure 12. A framework for task allocation in an HFS (Cvetkovic et al., 2024).....	79
Figure 13. Instantiation of the HFS (Cvetkovic et al., 2024).....	81
Figure 14. Snippet of a GLM-based Brainstorming app prototype (Memmert et al., 2024)	82
Figure 15. Prompting techniques performing better than baseline (left) and worse than baseline (right) (Memmert et al., 2024).....	86
Figure 16. Contributions along the Socio-Technical Continuum.	88

The figures of the articles included in this dissertation are not listed.

II. List of Tables

Table 1. Thesis outline	7
Table 2. Similarities and differences across the use cases and technologies addressed in the thesis	22
Table 3. IPO model, adapted from Curry et al. (2006)	25
Table 4. Contributions of articles to RQs.....	33
Table 5. Included journal publications.....	49
Table 6. Included conference publications.....	49
Table 7. Related journal submissions, not included in the dissertation	50
Table 8. First article of the cumulative dissertation	51
Table 9. Second article of the cumulative dissertation.....	52
Table 10. Third article of the cumulative dissertation	54
Table 11. Fourth article of the cumulative dissertation.....	55
Table 12. Fifth article of the cumulative dissertation	57
Table 13. Sixth article of the cumulative dissertation.....	58
Table 14. Seventh article of the cumulative dissertation	60
Table 15. Eighth article of the cumulative dissertation.....	62
Table 16. Ninth article of the cumulative dissertation	64
Table 17. Tenth article of the cumulative dissertation.....	66
Table 18. Tasks that postgraduate researchers (do not) want to delegate to AI. Adapted from Cvetkovic & Bittner (2022).	71
Table 19. Examples of human input identified in the literature and proposed AI output for each respective input.....	75
Table 20. DPs for Hybrid Facilitation Systems (Cvetkovic et al., 2024).....	80
Table 21. Participant ratings for AI suggestions per design component. Adapted from Memmert et al. (2023)	85
Table 22. Overarching DPs for human-centered HIS	88

The tables of the articles included in this dissertation are not listed.

III. List of Abbreviations

AI	Artificial Intelligence
CA	Conversational Agent
DR	Design Requirement
DP	Design Principle
DSR	Design Science Research
HFS	Hybrid Facilitation System
HI	Hybrid Intelligence
HIS	Hybrid Intelligence Systems
HKWS	Hybrid Knowledge Work System
ID	Information Disclosure
IS	Information Systems
ML	Machine Learning
RG	Research Goal
RQ	Research Question
SLR	Systematic Literature Review

The abbreviations in the articles included in this dissertation are not listed.

1. Introduction

1.1 Motivation and Problem Statement

"Ask not what AI could do, but what AI should do." This thought-provoking statement, inspired by Lubars and Tan (2017), encapsulates the essence of this dissertation. It challenges us to shift the narrative around artificial intelligence (AI) from looking solely at the potential capabilities to considering ethical, practical, and human-centered implementation. Central to this human-centered perspective is the notion of human input - the critical contributions humans provide to guide, enhance, and ethically ground AI systems. This focus is especially critical as we navigate an era where AI's influence permeates every aspect of work and life. For clarity, this work defines AI as computational systems designed to grasp and, to some extent, imitate human intelligence (McCarthy, 2007). Hereafter, the term AI refers to AI-based systems unless otherwise specified.

While challenges such as pandemics, economic disruptions, and societal transformations have always been part of human history, the interconnectedness and speed of today's globalized world amplify their complexity. Organizations now face problems that are not only multifaceted but also evolve rapidly, requiring innovative and adaptive strategies. In this volatile landscape, creative work, defined as the generation of novel and valuable ideas (Amabile, 1996) and complex problem-solving, characterized by interdependent variables, uncertainty, and dynamic constraints (Tim Brown & Katz, 2019), have become essential tools for navigating challenges and staying competitive. Activities such as product ideation, collaborative writing, and innovation management increasingly require diverse expertise and iterative exploration. The innovation processes often suffer from a lack of structure and overwhelming cognitive demands on knowledge workers, who are tasked with navigating vast quantities of information under intense time pressure (Schallmo & Lang, 2020). These challenges are compounded by the need to foster effective collaboration within teams, where interpersonal dynamics and diverse perspectives further complicate the path to success (Bittner & Leimeister, 2014).

Recent advancements in AI are transforming how we approach these problems, offering new pathways for innovation and collaboration. AI excels at automating repetitive tasks, processing large datasets, and identifying patterns that would otherwise elude human perception (Eziefule et al., 2022; Mahadevkar et al., 2024). For example, breakthroughs in deep learning have enabled AI to achieve unprecedented levels of performance in tasks such as image recognition and natural language processing (NLP), solidifying its role as a transformative technology (LeCun et al., 2015). The recently awarded Nobel prize¹ "*for foundational discoveries and inventions that enable machine learning with artificial neural networks*" speaks for itself – AI is no longer a futuristic concept but a pivotal force driving

¹ <https://www.nobelprize.org/prizes/physics/2024/press-release/>

change across industries, revolutionizing how we solve problems, innovate, and collaborate in the contemporary world. With its capabilities, AI is increasingly used to augment human efforts rather than replace them. This shift from automation to augmentation allows for the development of systems that empower humans to focus on creative and interpretative tasks while delegating routine or data-intensive activities to machines (Shneiderman, 2020). With improvements in generative AI, the augmentation now goes beyond such routine and data-intensive tasks to also encompass creative and problem-solving tasks. This sets the stage for Hybrid Intelligence Systems (HIS) – an integration of human and artificial intelligence designed to leverage their complementary strengths. Dellermann, Ebel, et al. (2019) outline three essential characteristics for HIS: 1. a common goal that aligns human and AI efforts, 2. complementary strengths that optimize their respective capabilities, and 3. lifelong learning that ensures continuous improvement and adaptability. Crucially, achieving these characteristics depends significantly on human input. This encompasses, e.g., tasks, feedback, domain-specific insights, contextual knowledge, and ethical guidelines that direct how AI systems learn, adapt, and align with human objectives (Lukyanenko et al., 2022). HIS thus emphasizes synergy, where AI complements humans rather than acting independently (Dellermann, Ebel, et al., 2019). Richter and Schwabe (2025) reinforce this perspective by arguing that teamwork with AI should not be measured solely by mutual understanding but rather by the complementary contributions each party can offer, highlighting the dynamic nature of roles within human-AI ensembles.

In practice, these complementary contributions are already visible. For instance, in the legal domain, HIS can assist in legal research by sifting through vast amounts of case law and legal documents to identify relevant precedents. Human input guides AI's selection criteria, ensuring relevance and ethical appropriateness. AI systems can quickly retrieve pertinent information, allowing lawyers to focus on crafting legal strategies and arguments (Seufert & Meier, 2023). In collaborative writing, AI tools like ChatGPT assist teams in generating draft content or identifying knowledge gaps, while human input refines narratives and ensures coherence (Hutson, 2023). Similarly, in complex problem-solving, AI can provide actionable insights by analyzing interrelated variables and generating solution scenarios, enabling humans to focus on contextual and ethical decision-making (Topol, 2019). Yet, realizing this potential requires addressing critical questions about the design and orchestration of these systems.

One central question revolves around the allocation of roles and responsibilities within human-AI ensembles (Berente et al., 2021). Determining which tasks should be performed by humans versus AI – and how these roles should dynamically evolve is a key challenge (Richter & Schwabe, 2025). Research suggests that repetitive, routine tasks are better suited for AI, while humans excel in areas requiring creativity, context sensitivity, and ethical judgment (Dellermann, Calma, et al., 2019). Human input is integral here, guiding the protocols for defining task boundaries, specifying ethical priorities, and calibrating AI outputs through structured prompts to ensure AI operates within human-defined constraints

(Shneiderman, 2020; W. Xu et al., 2023). Beyond role allocation, effective HIS must also ensure seamless interaction between humans and AI. This involves designing interfaces and workflows that enable clear communication, facilitate task orchestration, and support mutual learning. W. Xu et al. (2023) emphasize that these interactions must be carefully structured to align AI outputs with human goals, ensuring clarity in the allocation of tasks and the interpretation of results.

Another critical aspect of HI is fostering trust in AI and its transparency. For users to engage meaningfully with AI systems, they must understand how these systems operate, process data, and generate outputs (Liao & Wortman Vaughan, 2023; Lukyanenko et al., 2022). Transparency is not only a foundation of user trust but also essential for maintaining human agency in collaborative decision-making (Gunning & Aha, 2019). Equally important is the willingness of users to share information with AI systems, a process inherently reliant on human input, which is often shaped by concerns around privacy, data security, and the perceived value of AI support (Al-Natour et al., 2021). Addressing these concerns requires embedding human-centered design principles (DPs) into the development of HIS (W. Xu, 2019), that ensure technology aligns with human values, supports user autonomy, and prioritizes ethical considerations.

In this dissertation, we specifically emphasize the contexts of creative work and complex problem-solving (hereafter also referred to as problem-solving, unless otherwise specified). While countless other domains (e.g., decision-making, healthcare diagnostics, or financial forecasting) offer valuable use cases for HIS, these two focal areas emphasize open-ended exploration, the melding of diverse perspectives, and the balancing of objective facts with subjective insights. Creative work and complex problem-solving scenarios thus present unique design challenges for HIS, including the need to encourage ideation, support real-time collaboration, and manage the interplay between AI-driven suggestions and human judgment. For instance, AI's real-time processing capabilities can facilitate innovation management by dynamically structuring brainstorming sessions or highlighting underexplored areas, significantly improving the speed and quality of idea generation and evaluation (Bittner et al., 2021). Tavanapour and Bittner (2018) illustrate how AI systems can facilitate collaborative creativity by structuring unorganized data and offering alternative perspectives. However, their findings highlight the importance of tailoring AI capabilities to the context and objectives of human users. By centering on these contexts, this dissertation aims to illustrate how HIS can generate impactful and practically relevant outcomes in tasks that benefit from structured support for knowledge integration, idea development, and iterative feedback rather than purely rule-based or routine tasks.

Furthermore, this dissertation focuses specifically on one-on-one and team settings as primary application contexts due to their prevalence and importance in real-world creative work and problem-solving scenarios. Individual (one-on-one) settings, such as solitary brainstorming or writing tasks, represent fundamental units of creative work where AI can augment personal creativity. Team settings, on the other hand, introduce complex social

dynamics, requiring AI systems to understand interpersonal interactions and support effective group collaboration. By targeting both contexts, this dissertation ensures that developed HIS artifacts are comprehensive, versatile, and broadly applicable.

Summing up, while HIS hold significant promise, their design and deployment present notable challenges. The interaction between humans and AI has been described as one of the most significant managerial issues of our time (Berente et al., 2021). Achieving effective human-AI collaboration requires a delicate balance between human and machine strengths, guided by clear communication, trust, and shared goals. A key enabler of that balance is human input, e.g., the knowledge, expertise, preferences, and contextual understanding that people contribute to AI systems. While many discussions around AI focus on algorithmic performance or data availability, it is equally vital to examine under which conditions humans provide input and how AI leverages that input to produce meaningful support. This perspective on human input sets the stage for investigating how and why users disclose information to AI, how trust can be fostered, and how human insight shapes the outcomes of creative work and complex problem-solving.

This dissertation addresses these challenges by designing and empirically evaluating HIS artifacts that operationalize the principles of augmentation over automation, effective role allocation, and the central importance of human input in AI-driven processes. It seeks to answer questions about the roles humans and AI should assume, how their interactions can be orchestrated to achieve shared objectives, and what ethical and contextual considerations are necessary for successful implementation. In doing this, it addresses both theoretical and design knowledge gaps (Avdiji & Winter, 2019). By advancing both theoretical and practical understanding of HIS, this work contributes to the growing body of research on human-centered AI and offers actionable insights for the design of AI that empowers rather than replaces human capabilities.

1.2 Research Goal and Research Questions

Current research lacks systematic and integrated design knowledge on human-centered design of HIS in the context of creative work and complex problem-solving. Specifically, the provision of human input and its utilization by AI to achieve the goals of HIS need to be addressed from different perspectives. While it is relevant to design and evaluate HIS in different contexts, the early work of this dissertation has recognized an opportunity in creative work and complex problem-solving, where AI offers immense possibilities for improvement. Hence, this dissertation postulates the following overall research goal (RG) in an attempt to pave the way toward closing the above-described research gap:

RG: Enable human-centered HIS for creative work and complex problem-solving by developing validated design knowledge for the provision of human

input and its effective utilization by AI while leveraging the complementary strengths of humans and AI.

As the RG emphasizes design focus resulting in producing design knowledge, this dissertation adopts a design science research (DSR) paradigm as its methodological framework of choice. To guide the systematic examination of the overarching RG in the scope of a cumulative dissertation, we postulate five overarching research questions (RQ) and organize them along the three recognized DSR cycles: relevance, rigor, and design, following Hevner (2007) and Thuan et al. (2019) (Figure 1).

Addressing the relevance cycle, we analyze the status quo of HIS in the environment - creative work and complex problem-solving - to identify opportunities and challenges for potential solutions to be generated. As this work adopts a human-centered design of HIS, this RQ focuses on human factors, especially the ones potentially driving or hindering the development and adoption of AI in the context of HIS. As human-AI ensembles cannot be uncoupled and should thus not be observed separately, a socio-technical lens is adopted to ensure that the interplay between the system actors is accurately and holistically represented. Accordingly, the first RQ is split into two sub-questions and follows:

RQ 1a: What are the prerequisites for the provision of human input and its utilization by AI in human-centered HIS?

RQ 1b: What are the opportunities tied to the provision of human input and its utilization by AI in human-centered HIS?

Having established prerequisites and opportunities, we transition into the design cycle, where artifacts, such as DPs or software prototypes, are iteratively designed, developed, and evaluated. To address this, the second research question is divided into two parts: the first concerning the design of artifacts and the second, their evaluation.

Vis-à-vis the design aspect, after identifying user scenarios suitable for HIS, determining the types of tasks that AI could and should undertake, and understanding the factors affecting the willingness to share data with AI to benefit from its support, we developed prototypes where AI assumed structured, routine tasks that align with its complementary strengths, allowing humans to focus on creative and interpretive outputs.

As this dissertation progressed, the rapid advancements in generative AI transformed its capabilities, equipping it with abilities such as nuanced contextual understanding, natural language generation (NLG), and multi-turn conversational engagement. These advances not only expanded the opportunities for implementing AI in diverse scenarios but also evolved user needs and expectations. This prompted us to extend the scope of AI

implementation to complex problem-solving scenarios, enabling AI to undertake more demanding tasks such as generating ideas and providing suggestions based on human input. In this context, we utilized existing state-of-the-art AI solutions and focused on designing prompt templates to enhance AI output in given scenarios.

To guide this exploration, we pose the question:

RQ 2a: How can AI be designed and developed, or utilized in the context of human-centered HIS for creative work and complex problem-solving?

After designing several artifacts to address the prerequisites and opportunities identified in RQ 1, we moved on to the evaluation phase to validate the resulting design knowledge as outlined in RQ 2a. For design knowledge to be applicable in practice, it must undergo validation to confirm its utility and relevance in real-world contexts (Tremblay et al., 2010). This involves establishing the utility of the artifacts and the knowledge underpinning their development. To assess the validity of these designs and ensure their practical applicability, the following RQ is posed:

RQ 2b: How does the human-centered integration of AI affect creative work and complex problem-solving?

This “affect” is examined along both objective indicators (e.g., working output value, novelty, efficiency) and subjective user perceptions (e.g., of usefulness, cognitive load).

Addressing the rigor cycle, we draw from the outcomes of prior RQs to make new contributions to the knowledge base. Thus, the final RQ is the following:

RQ 3: What contributions can be made to the knowledge base regarding the design and integration of AI in human-centered HIS to enhance creative work and complex problem-solving?

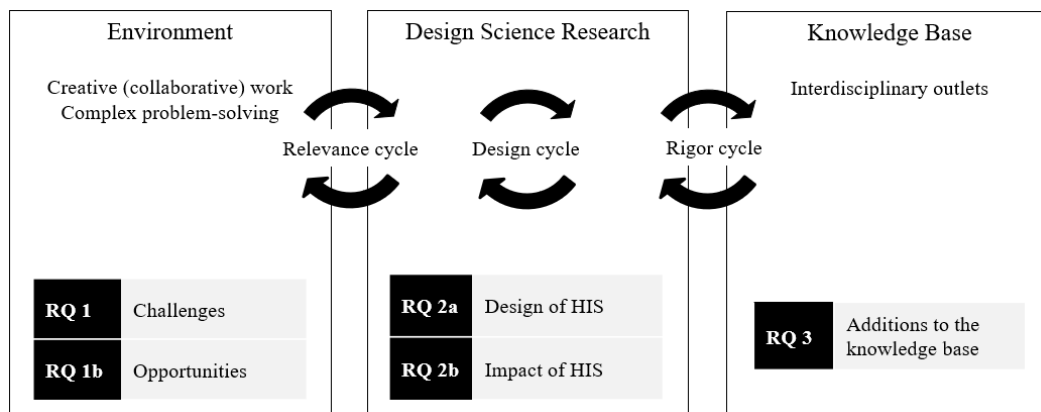


Figure 1. Assignment of RQs to DSR cycles. Adapted from Hevner (2007)

These RQs are addressed in ten interconnected articles, each with individual research goals, questions, and contributions. Each article contributes to answering multiple RQs and drives the accumulation and evolution of design knowledge toward the overall RG of our DSR project. The relationship between the articles and concrete research activities and how they are linked to respective RQs is outlined in Section 3.2.

1.3 Thesis Outline

This dissertation consists of a wrapper and ten articles, as shown in Table 1.

Table 1. Thesis outline

Wrapper	1. Introduction	2. Theoretical Foundations	3. Research Design	4. Publications
	5. Theoretical Contributions	6. Practical Contributions	7. Limitations	8. Implications for Future Research
Publications	9. Article 1	Hybride Wissensarbeit		
	10. Article 2	Task Delegability to AI: Evaluation of a Framework in a Knowledge Work Context		
	11. Article 3	What should AI know? Information Disclosure in Human-AI Collaboration		
	12. Article 4	A User-centric Taxonomy for Conversational Generative Language Models		
	13. Article 5	Conversational Agent as a Black Hat: Can Criticizing improve Idea Generation?		
	14. Article 6	Augmented Facilitation: Designing a multi-modal Conversational Agent for Group Ideation		
	15. Article 7	Designing a Hybrid Intelligence System for the Facilitation of Group Collaboration		
	16. Article 8	Human-AI Collaboration in Conceptualizing Design Science Research Studies: Can Generative Language Models Provide Helpful Suggestions?		
	17. Article 9	The more is not the merrier: Effects of Prompt Engineering on the Quality of Ideas produced by GPT-3		
	18. Article 10	Quid pro Quo: Information Disclosure for AI Benefits in Human-AI Collaboration		

The wrapper provides a cohesive summary, linking each publication's contributions to the overall RG and RQs. It begins with an introduction that presents the research motivation, outlines the problem, formulates the RQs, and briefly discusses the approach to address them in Chapter 1. Chapter 2 establishes the theoretical and conceptual foundations that support the dissertation's research. Chapter 3 details the research design, encompassing the research paradigm, strategy, and methods that integrate the publications within the overarching DSR framework of the dissertation. Chapter 4 offers a summary of each included publication. Theoretical contributions are discussed in Chapter 5, followed by practical contributions in Chapter 6. Chapter 7 addresses the thesis' limitations, and Chapter 8 provides directions for future research. Sections 9 through 18 contain the ten individual publications that form the core of this cumulative dissertation.

2 Theoretical Foundations

This chapter provides the theoretical and conceptual groundwork for the dissertation. Its primary purpose is twofold: (1) Establish a unifying framework for the dissertation’s publications, ensuring that shared concepts and principles can be consistently referenced in later chapters (particularly when discussing individual studies and their contributions in Chapter 5); (2) Define the key terms and concepts (e.g., human-centered AI, Hybrid Intelligence, Conversational Agents (CAs), Large Language Models) that recur across the publications, so that readers have a coherent understanding of how these terms are used throughout the dissertation.

To achieve these goals, Chapter 2 is structured around a human-centered AI perspective, a lens that highlights the interplay between human needs, technological capabilities, and ethical considerations. In particular, it focuses on: (1) User Dimension, which explores the concept of user-centric design and the specific use cases of creative work and complex problem-solving; (2) Technology Dimension, which delves into the topic of HI and discusses AI-driven tools such as CAs and LLMs; (3) Ethics Dimension, which addresses ethical concerns like transparency, privacy, and trust.

Building on these three foundational pillars, Section 2.2 then introduces the “Input-Process-Output” (IPO) model for human input provision and its utilization by AI, offering a structured view of how humans and AI cooperate and jointly produce outcomes in HIS. This model underpins the dissertation’s discussions of the types of input humans share with AI, the processes that transform that input into useful outputs, and the types of resulting AI outputs.

2.1 Human-centered AI

As a response to emerging concerns about AI development, such as its controllability, inherent biases, spreading of fake news, and many more, leading research institutes such as Stanford, MIT, and UC Berkeley have established human-centered AI research institutes (W. Xu, 2019). Researchers from these institutes believe that AI research and development should follow three objectives: 1. It should mimic human intelligence; 2. It should augment human intelligence rather than replace it; and 3. It should focus on the AI’s impact on humans.

Human-centered AI has since risen as an approach to AI that emphasizes designing systems around human needs, preferences, and values to ensure that technology remains safe, fair, and beneficial. Unlike traditional AI, which often prioritizes pure efficiency or accuracy, human-centered AI focuses on enhancing human well-being, autonomy, and decision-making. This approach integrates interdisciplinary insights from fields like psychology, ethics, and human-computer interaction to address the complexities of human behavior and social impact (Shneiderman, 2020).

Research shows that human-centered AI can improve user satisfaction, trust, and overall system reliability by prioritizing transparency, interpretability, and user participation in AI development. For example, Mosqueira-Rey et al. (2023) and Deng et al. (2020) highlight that user-centric DPs, such as co-design and participatory methods, can help AI systems better align with real-world contexts and reduce biases that might otherwise emerge from isolated, technical approaches to model building. Furthermore, Shneiderman (2020) argues that human-centered AI can mitigate risks associated with autonomous systems by ensuring that humans maintain control and oversight over AI functions, thus enhancing accountability and safety in complex applications like healthcare and autonomous driving.

This paradigm aligns closely with the principles of *socio-technical systems design*, which focuses on the interaction between people and technology within organizational contexts. Both frameworks advocate for designing systems that prioritize human needs and values, fostering collaboration between humans and machines rather than viewing technology as a mere replacement for human effort (Baxter & Sommerville, 2011).

Human-centered AI exemplifies its potential to enhance human experiences across various domains, as seen in applications like Woebot², Microsoft's Seeing AI³, and personalized learning platforms like Duolingo⁴. Woebot provides mental health support through empathetic CAs, offering users personalized coping strategies while ensuring their privacy is respected. Similarly, Microsoft's Seeing AI empowers visually impaired individuals by using computer vision to describe their environments, enhancing independence without compromising personal data security. In the educational sector, Duolingo adapts its lessons based on individual performance and preferences, creating a tailored learning experience that respects user anonymity.

At its core, human-centered AI can be understood through three interrelated dimensions: User, Technology, and Ethics (W. Xu, 2019). It places humans at the center – all three aspects of it converge on the central role of humans. Specifically, the user factor ensures that AI systems are controllable by giving humans the ability to make final decisions, usable and explainable by providing effective interaction design and understandable output. The technology aspect ensures that human abilities are augmented through a complementary integration of AI, and the ethics aspect addresses ethical issues such as information privacy by providing meaningful human control (van Diggelen et al., 2021; W. Xu et al., 2023).

In the following, the three core human-centered AI factors will be introduced, with specific reference to the subtopics relevant to the dissertation.

² <https://woebothealth.com/>

³ <https://www.microsoft.com/en-us/garage/wall-of-fame/seeing-ai/>

⁴ <https://www.duolingo.com/>

2.1.1 User

The user aspect of human-centered AI emphasizes the importance of user-centric design approaches in AI development (Amershi et al., 2019). This involves identifying user needs, defining users and user scenarios, focusing on human-AI collaboration, and augmenting human capabilities (W. Xu & Gao, 2024). Recent research has shown that incorporating user-centric DPs can significantly improve the effectiveness and acceptance of AI systems (Ehsan & Riedl, 2020; Rai, 2020; Yang et al., 2020). Human-AI collaboration is a key focus of human-centered AI, promoting the development of systems that facilitate effective interaction between users and AI. Research by Seeber et al. (2020) argues that successful human-AI collaboration requires careful consideration of task characteristics, AI capabilities, and human cognitive processes.

This approach aims to create a synergistic relationship where both human and AI capabilities are leveraged to achieve optimal outcomes. The concept of augmenting human capabilities ties into the idea of HI, where AI systems are designed to enhance human cognitive abilities rather than replace them. A paper by Dellermann, Ebel, et al. (2019) theorizes that HIS can lead to superior problem-solving and decision-making outcomes compared to either humans or AI working alone.

Human-AI collaboration offers significant opportunities in tasks requiring analytical thinking, problem-solving, and creativity. This dissertation explores two key scenarios: 1) facilitating creative work in both one-on-one and team settings, and 2) supporting complex problem-solving processes. These scenarios highlight the potential of AI to enhance tasks that benefit from the combination of human insight and computational power, with a focus on creativity and problem-solving as distinct areas of application.

Creative work

Collaboration is central to modern organizational environments, where teams often work together across dispersed locations, departments, or even within the same physical space. In the context of creative work, facilitation serves as a “neutral supporter” (Leimeister, 2014) that enhances group dynamics, promotes effective communication, and manages conflicts, all while helping the team achieve agenda-driven outcomes (Niederman et al., 1996). Effective facilitation contributes to structuring meetings, encouraging participation, and guiding the creation of high-quality group outputs. Human facilitators take on diverse responsibilities, from setting meeting objectives and promoting trust among team members to ensuring task completion and documenting outcomes (Beranek et al., 1993; Clawson et al., 1993).

Creativity and innovation are vital components of many organizational processes, with companies continuously seeking new ways to foster these qualities among both individuals and teams (Meinel & Leifer, 2012). In particular, creative innovation approaches such as Design Thinking (DT) have emerged as a valuable approach. Such approaches are used

across various industries for tasks like empathizing with customers, ideation, prototyping, and product development (Carlgren et al., 2014). Creative innovation process is structured into distinct stages, where facilitators play a key role in managing team dynamics, moderating interactions, and organizing resources (Tim Brown & Katz, 2011; Hjalmarsson et al., 2015). To complement facilitation, organizations deploy creativity support systems that scaffold exploration, externalize evolving concepts, and act as a cognitive “memory” that designers can search, reflect on, and build upon in later cycles so that teams can iterate rapidly and collaboratively (Shneiderman, 2007).

Building on these foundations, co-creative AI now joins teams as a “virtual collaborator” (Siemon et al., 2018), offering fresh stimuli and steering creative work toward novel directions. Given the intensity and multifaceted nature of facilitation, AI has significant potential to assist with its logistical and managerial aspects. For instance, AI can produce and analyze transcripts, recommend productive next steps, and suggest potential solutions based on previous projects (Bittner et al., 2021; Chen et al., 2021). This would allow human facilitators to concentrate on more nuanced tasks that require emotional intelligence, such as building trust and encouraging creative risk-taking among team members. Moreover, group-creativity research shows that evaluation apprehension, the fear of negative social judgment, reduces idea fluency in brainstorming (Paulus & Dzindolet, 1993). By shifting the criticizing role to a neutral digital agent, facilitation technologies can lower this social cost and thereby unlock a larger pool of raw ideas (Siemon, 2022). Further, AI can enhance collaborative creativity by managing repetitive tasks, analyzing vast datasets, and assisting with ideation processes. For example, Tavanapour and Bittner (2018) found that AI can play a role in brainstorming sessions by encouraging participants to elaborate on their ideas, thereby ensuring that contributions are detailed and comprehensive. Additionally, AI can assess creative outputs based on criteria such as novelty and feasibility, allowing practitioners to focus on developing the most promising solutions while maintaining a structured workflow (Maher & Fisher, 2012). From an external-cognition perspective, interactive artefacts can act as *cognitive off-loading* devices that extend the limited working memory of idea generators (Kirsh, 2010; Scaife & Rogers, 1996). Such visual scaffolds not only organize information but also stimulate *cognitive conflict*, a well-established trigger for higher originality in teams (Nemeth et al., 2004). As companies continue to face global challenges and rapid technological changes, the capacity for AI to support creativity and innovation in collaborative environments becomes increasingly essential.

While previous studies indicate the potential for AI to enhance facilitation of creative work, there is a need to explore how AI can provide scalable support across varied facilitation tasks in DT and other collaborative approaches. Current AI implementations are limited, mostly addressing individual-level tasks. Further research is required to develop comprehensive AI solutions that can also support complex, team-based facilitation activities, assisting human facilitators in areas such as team interaction, conflict resolution, and task management within dynamic collaborative settings (Bittner et al., 2021; Schallmo

& Lang, 2020). Accordingly, this dissertation addresses these challenges by designing and evaluating AI artifacts that go beyond single-user scenarios (Cvetkovic, Gierlich-Joas, et al., 2023; Cvetkovic et al., 2024), aiming to enhance group-level processes, manage multifaceted team tasks, and provide context-specific support for human facilitators in collaborative innovation settings.

Complex problem-solving

Complex problem-solving represents another critical domain where HIS can yield significant benefits. In contemporary rapidly evolving business landscape, organizations frequently encounter complex challenges that require multifaceted solutions. Complex problems differ from simpler problems primarily due to their inherent uncertainty, dynamic interdependencies among numerous variables, lack of clearly defined procedures for solving them, and evolving constraints over time (Fischer et al., 2012). Unlike simple problems, which typically have well-defined structures, known procedures, and predictable outcomes, complex problems are ambiguous and require iterative exploration and adaptive decision-making. As such, these problems often exceed the cognitive capacity of individual problem-solvers (Joksimovic et al., 2023; Newell & Simon, 1972; Raisch & Krakowski, 2021).

Over the years, digital support for tackling complex problem-solving has progressed from model-driven decision support systems that let individual managers run what-if simulations, through group decision-support platforms that helped dispersed teams visualise assumptions and debate alternatives (Shim et al., 2002), to the latest set-ups in which AI joins as a collaborator, co-exploring dynamic challenges with human experts (Memmert & Bittner, 2022). AI systems can augment human problem-solving capabilities in several ways. This can be viewed as a form of *cognitive scaffolding* (Wood et al., 1976) that structures the interaction so as to minimize intrinsic cognitive load (Sweller, 1988) while exploiting the analogy-based reasoning inherent in advanced AI models. Further, AI can process and analyze vast amounts of data to identify patterns and relationships that might elude human perception (Shneiderman, 2020; Silver et al., 2021). This can help in framing the problem more accurately and comprehensively. Likewise, AI can generate and evaluate multiple solution scenarios rapidly, allowing human decision-makers to explore a broader range of options (Edwards et al., 2024).

Despite the potential benefits of human-AI collaboration in complex problem-solving, several critical research gaps remain. Generally, while studies have explored AI's role in specific tasks or domains (Brynjolfsson & McAfee, 2017; Davenport & Ronanki, 2018; Huang & Rust, 2018), these need to be continuously updated as AI's capabilities increase, and user needs correspondingly change. Furthermore, a lot of existing research focuses on the technical capabilities of AI systems, with less attention paid to the human side of the collaboration. There's a need for more in-depth studies on how workers experience, interact

with, and adapt to AI collaborators in real-world settings (Davenport & Kirby, 2016; Richter & Schwabe, 2025; Seeber et al., 2020).

To address these gaps, this dissertation empirically explores two specific use cases of complex problem-solving: the conceptualization phase of DSR and brainstorming solutions for critical societal and environmental issues. In the DSR context, the thesis examines how generative AI can support researchers in systematically deriving kernel theories, issues, design principles, and requirements (Memmert et al., 2023). Regarding the second use case, it investigates how different prompt-engineering techniques influence AI's capacity to produce high-quality ideas, emphasizing novelty and practicality in problem-solving scenarios (Memmert et al., 2024). By focusing on these use cases, this research provides concrete insights into how AI can effectively complement human cognitive capabilities when addressing ambiguous and multifaceted challenges.

2.1.2 Technology

Advances in AI have made it possible for AI to transition from simple automation tools to more sophisticated roles that augment human capabilities. This transition is largely due to developments in machine learning (ML), NLP, and data analytics. These advancements enable AI systems to interpret complex information, recognize patterns, and offer structured insights that support human tasks in dynamic work environments (Russell & Norvig, 2021). In particular, AI's ability to process and analyze large datasets in real time has enabled its integration into areas where high-level decision-making and innovation are required (Davenport & Ronanki, 2018). Most recently, the development of transformer models has revolutionized NLP. The GPT (Generative Pre-trained Transformer) series, culminating in GPT-4x by OpenAI⁵ has demonstrated unprecedented capabilities in language understanding and generation.

Hybrid Intelligence

The concept of HI originates from the fathers of AI in the 1960es, when AI was discussed as a possible “amplifier” of human’s “intellectual power”, or “augmenting human intellect” through a “man-computer symbiosis” (Ryjov, 2021). HI leverages the technological capabilities of AI to create systems where AI supports rather than replaces human effort. It is characterized by three core principles: a *common goal* shared by both humans and AI, *complementary strengths* that allocate tasks to each based on their unique capabilities, and *lifelong learning* that supports ongoing knowledge development for both human and AI (Dellermann, Ebel, et al., 2019). These principles echo foundational collaboration theories: joint-activity theory emphasizes the establishment of common ground for effective coordination (Klein et al., 2005), shared-mental-model research shows that partners, human or artificial, perform better when they hold aligned representations of goals, tasks, and capabilities (Mathieu et al., 2000), distributed-cognition perspectives remind us that

⁵ <https://openai.com/chatgpt/overview/>

cognitive work is spread across humans, AI algorithms, and the artefacts that link them (Hutchins, 2000).

This dissertation tackles the first two HI principles; the learning aspect is out of the scope of the overall research goal. The first two principles presume a symbiotic relationship between human cognitive strengths, such as intuition and creativity, and AI's computational power, which excels at data-intensive, structured tasks. This symbiotic relationship reflects the middle levels of automation, which determine the balance of human control versus AI-driven task management across various applications (Parasuraman et al., 2000).

This approach is foundational in knowledge-intensive work, where HI allows for an efficient division of labor: AI handles routine or analytical processes, while humans engage in more complex, interpretive work (Dellermann, Ebel, et al., 2019). By emphasizing collaboration rather than replacement, HI encourages the development of adaptive, flexible systems that prioritize human oversight and decision-making (Rahwan et al., 2019).

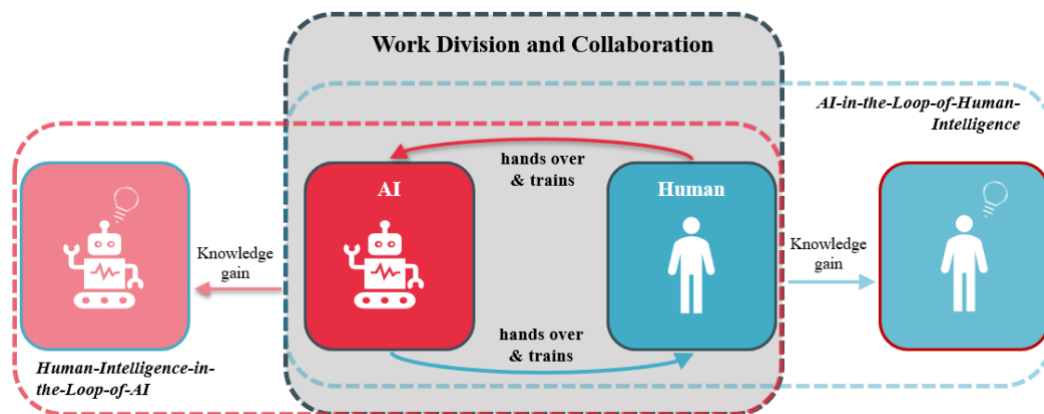


Figure 2. Hybrid Intelligence with the core components addressed in this thesis highlighted in the grey box. Adapted from Oeste-Reiß et al. (2021)

Applications of HI span multiple industries. In healthcare, for example, AI tools assist doctors by analyzing patient data, identifying patterns, and generating diagnostic suggestions, but it is ultimately the physician who integrates these insights into a treatment plan (Topol, 2019). Similarly, in finance, AI systems analyze complex datasets to inform investment decisions, while human analysts provide the contextual and ethical judgment necessary to guide high-stakes choices (Brynjolfsson & McAfee, 2014). In education, HI frameworks are used to support personalized learning: AI systems analyze student performance and provide recommendations, while teachers adapt these insights to individual learning needs and classroom dynamics (Luckin et al., 2016).

HI can be particularly beneficial in creative work environments where teams engage in collaborative creativity, brainstorming, and problem-solving. In such settings, AI can support human teams by, e.g., clustering content, providing relevant background data, or suggesting alternative perspectives (Bittner et al., 2021). This role as a “co-creator” enables

AI to help structure complex tasks, allowing human collaborators to direct their cognitive efforts toward the creative and interpretative dimensions of work. It is hypothesized that HI can enhance productivity and increase user satisfaction by empowering participants with supportive AI tools that streamline aspects of the collaborative process (Shneiderman, 2020).

Effective HI implementation requires seamless integration, clear task boundaries, user-friendly interfaces, and a commitment to transparency, *inter alia* (Amershi et al., 2019). The latter allows users to understand AI outputs, fostering trust and enhancing collaboration (Gunning & Aha, 2019). As AI increasingly supports complex tasks, HI frameworks ensure that users retain the agency to make informed decisions, thus maintaining a balance between computational efficiency and human oversight. A well-designed HIS should reflect human-centered principles, providing transparency in decision-making, supporting user autonomy, and fostering trust. These demands require AI systems that can engage users in natural, intuitive ways, facilitating productive collaboration without imposing a steep learning curve or disrupting established workflows (Shneiderman, 2020; W. Xu et al., 2023).

This dissertation tackles these topics by empirically refining a task delegation framework that identifies the roles of motivation, risk, task complexity, and trust in shaping each delegation decision (Cvetkovic & Bittner, 2022). Further, it proposes a model for information disclosure to AI that encompasses trust and transparency as core components (Cvetkovic et al., 2025). It then derives DPs to build and evaluate HIS prototypes that demonstrate, in practice, how complementary strengths can be orchestrated within creative work (Cvetkovic, Gierlich-Joas, et al., 2023; Cvetkovic et al., 2024) and complex problem-solving settings (Memmert et al., 2023, 2024), thereby closing the conceptual loop set out above.

Conversational Agents

In the context of the previously discussed, conversational AI has emerged as a promising solution, especially suited to the collaborative, creative scenarios described previously. CAs and LLMs represent some of the most effective approaches for achieving natural, accessible human-AI interaction (Adamopoulou & Moussiades, 2020; Diederich et al., 2022; Radziwill & Benton, 2017). These systems align well with the human-centered AI framework's target values (W. Xu et al., 2024), as they embody AI solutions that reflect human cognitive capabilities and values by allowing users to interact naturally, intuitively, and responsively with AI. By enabling a conversational interface, CAs and LLMs create an inclusive interaction model that resonates with how people naturally communicate, making these technologies particularly effective in settings requiring user-centric facilitation, creativity, and complex problem-solving.

CAs are AI-driven systems designed to simulate human-like interactions through natural language, via text (chatbots) or audio (e.g., Siri, Alexa) (Brandtzæg & Følstad, 2018).

Initially, CAs were simple, rule-based systems with limited conversational capabilities, such as ELIZA in 1966 (Weizenbaum, 1966), but advances in NLP and ML have enabled them to handle more dynamic, complex conversations (Devlin et al., 2019; Vaswani et al., 2017). The use of NLP technology is crucial to the functionality of modern CAs. It involves transforming unstructured speech or text data into a structured format that machines can interpret and act upon. NLP comprises two core components: Natural Language Understanding (NLU) and Natural Language Generation (NLG). NLU enables machines to comprehend grammar, semantics, and context to extract meaning, while NLG focuses on generating coherent and contextually appropriate responses based on structured data (Nadkarni et al., 2011).

The development and adoption of CAs have been accelerated by ML and the use of extensive training data, allowing these systems to improve response quality over time. By learning from past conversations, modern CAs acquire new knowledge, refine their performance, and provide increasingly accurate and relevant information (Lewandowski et al., 2023). CAs are also becoming more accessible through the proliferation of bot platforms provided by companies like Telegram, Slack, and Facebook. These platforms allow external developers to create tailored CAs that integrate with various applications, extending their functionalities for polling, entertainment, and workflow automation (Khan & Das, 2018). As the volume of data generated by internet users continues to grow, CAs gain access to richer datasets, further enhancing their capacity to analyze and extract meaningful insights (Khan & Das, 2018). These capabilities empower CAs to recognize user intentions, process contextual information, and generate meaningful responses in real-time, making them valuable tools in e.g., customer service, healthcare, and education (Poser, Wiethof, & Bittner, 2022; Radziwill & Benton, 2017).

CAs have been increasingly integrated into creative work, serving both individual users and teams. In one-on-one settings, CAs function as intelligent partners, assisting users in tasks such as brainstorming, writing, and problem-solving. For instance, Wiethof et al. (2020) developed an intelligent collaborative agent to enhance collaborative writing, aiming to create a synergy between human creativity and AI capabilities. Similarly, Poser, Küstermann, et al. (2022) designed and evaluated a CA to facilitate idea generation in organizational innovation processes, demonstrating the agent's potential to stimulate creative thinking. In team environments, CAs act as facilitators, promoting effective communication and coordination among group members. Winkler et al. (2021) explored how interactions with smart personal assistants can improve group performance, highlighting the role of CAs in enhancing collaborative problem-solving. Tepper et al. (2018) demonstrated that a CA can assist users in coping with chat information overload by helping them understand the main topics discussed as well as by drawing connections between collaborators, links, and resources. Furthermore, Kim et al. (2021) investigated how CAs can guide and structure team discussions, and Elshan et al. (2022) and Siemon et al. (2022) showed how CAs can relieve teams from innovation blockages and influence

their motivation to contribute. In digital collective brainstorming, CAs have been shown to guide participants in individual idea generation and then through a collective idea evaluation (Strohmann et al., 2017). In addition, this CA was able to facilitate team interactions, by e.g., recognizing “killer phrases” that hinder creative processes. In this context, Bittner et al. (2021) identified key areas of expertise that a CA should cover to support both human facilitators and creative teams. The authors categorize the potential of CAs acting as facilitators along the areas of task, process, and interaction facilitation as well as method-specific facilitation. For example, to assist teams with tasks and processes, CAs could act as a “Process Keeper” by helping the team stick to the DT process and execute it in the intended way (Bittner et al., 2021).

Despite their versatility, CAs have inherent limitations that become particularly pronounced in creative work. While they excel at executing structured tasks, they struggle with interpreting unstructured conversations, addressing complex interpersonal dynamics, and understanding nuanced contexts. For instance, Toxtli et al. (2018) identified challenges in CA communication with multiple users, noting difficulties in tracking conversational threads and maintaining context. Similarly, research by Lasecki et al. (2013) highlighted that CAs often struggle with real-time decision-making in dynamic environments, necessitating human intervention to ensure accuracy and relevance. Furthermore, CAs frequently encounter obstacles in processing ambiguous language and sarcasm, leading to misunderstandings in team interactions (Ghosh et al., 2018). Their limited capacity to recognize and adapt to the emotional states of users can impede effective collaboration, as emotional intelligence plays a crucial role in team dynamics (Brave et al., 2005). For example, during brainstorming or ideation sessions, human facilitators often adapt dynamically to participants’ emotions and social cues - capabilities that current CAs lack. Additionally, CAs may lack the cultural competence required to navigate diverse team settings, potentially resulting in miscommunications and reduced cohesion (Evers et al., 2008).

In light of these drawbacks, a HIS combining a CA and human facilitators emerges as a potential solution. In previous research, CAs have acted as a replacement for human facilitators in an automated fashion. However, the augmentation of human facilitators in the context of HIS has advantages: while the CA overtakes task and process facilitation by analyzing and structuring data quickly and reliably, human facilitators can act on the side of interaction interventions, by, e.g., monitoring team dynamics and flexibly intervening.

In this dissertation, the identified limitations of current CAs are addressed in three successive studies. Cvetkovic, Rosenberg, and Bittner (2023) prototype a CA that delivers structured, real-time criticism during individual ideation and empirically shows how this can raise idea novelty without introducing the social friction typical of human criticism. Cvetkovic, Gierlich-Joas, et al. (2023) extend this work to team settings with a multi-modal CA that clusters, visualizes and summarizes live input, thereby mitigating information-overload and context-tracking problems noted by Toxtli et al. (2018) and Lasecki et al.

(2013). Finally, Cvetkovic et al. (2024) evaluate a hybrid facilitation system (HFS) in which a CA assumes task- and process-oriented interventions while a human facilitator remains responsible for nuanced interaction cues, directly tackling the emotional-intelligence and cultural-competence gaps outlined above and demonstrating how complementary strengths can be orchestrated in practice.

Large Language Models

LLMs, such as GPT-3 and GPT-4, represent a significant breakthrough in NLP, with the ability to generate coherent, human-like text based on user prompts. These models are trained on extensive datasets, enabling them to understand context, respond to open-ended questions, and generate creative content across various domains (Tom Brown et al., 2020). As such, they represent advanced AI systems designed to understand and generate human-like text. They utilize deep learning techniques, particularly neural networks with numerous parameters, to predict and produce coherent language sequences. LLMs can perform tasks including summarization, translation, content creation, and complex question-answering, making them versatile tools in collaborative work and complex problem-solving (Zhang et al., 2021).

The development of LLMs has evolved significantly, starting with early generative pre-trained transformer models like GPT-1, which laid the foundation for task-agnostic NLP systems. GPT-2 and GPT-3 marked milestones by significantly scaling model size and improving accuracy without the need for task-specific training (Tom Brown et al., 2020). For example, GPT-3, trained on 175 billion parameters, has demonstrated impressive capabilities in generating coherent and contextually rich text for tasks such as news article writing, conceptual design, and even creativity testing (Di Fede et al., 2022; Zhu & Luo, 2022). The advent of GPT-3.5, which powered the first version of ChatGPT, introduced conversational capabilities, allowing for multi-step, chat-based interactions. Unlike earlier one-off interactions, chat-based LLMs enable iterative problem-solving, facilitating deeper understanding and dynamic adjustment during the conversation. These capabilities, further enhanced by the release of GPT-4, underscore the ongoing innovation in the LLM landscape (Dwivedi et al., 2023). ChatGPT represents a convergence of CAs and LLMs, combining the conversational architecture of traditional chatbots with the generative and adaptive capabilities of advanced LLMs.

LLMs have found applications in diverse fields. In the creative domain, they have been employed for tasks such as brainstorming, where they generate "sparks" of inspiration by offering alternative perspectives or suggestions (Gero et al., 2022). For instance, LLMs have been used in divergent thinking tasks like the "alternative use test," showcasing their ability to produce novel and varied ideas (Stevenson et al., 2022). Additionally, in design contexts, LLMs have supported conceptualization by generating preliminary solutions to design challenges (Zhu & Luo, 2022). In problem-solving scenarios, LLMs demonstrate flexibility by generating task-specific content without requiring pre-defined training

datasets. This adaptability makes them well-suited for addressing open-ended, knowledge-rich problems typical of e.g., DSR (Hevner et al., 2004). The ability to provide content-level support in real-time, based on human input, distinguishes LLMs from traditional AI systems reliant on static repositories of information.

LLMs face several challenges, particularly in the realm of user interaction. Effective interaction with LLMs hinges on crafting precise and contextually appropriate prompts. Small variations in input prompts can lead to significantly different outputs, necessitating expertise in *prompt engineering* - the art of optimizing prompts to elicit desired responses (Mishra et al., 2022). This requirement poses difficulties for users, who may struggle to frame their inputs effectively (Jiang et al., 2022). Moreover, the transition from one-off, static prompts to dynamic, conversational interactions introduces both opportunities and complexities. While multi-step conversations allow for iterative refinement of ideas, they require a framework to guide users in structuring their inputs and leveraging the system's full potential (Wu et al., 2022). Techniques such as prompt chaining, i.e., breaking tasks into smaller steps, have been proposed to enhance the efficacy of LLM interactions in such settings.

Despite the rapid advancements and widespread adoption of LLMs, their meaningful and effective use case scenarios are still in the process of being identified and refined. While LLMs demonstrate impressive capabilities in generating creative ideas and solving complex problems, there remains a significant gap in understanding how these systems can be best leveraged in diverse contexts. Furthermore, research on the user experience aspects of LLMs, particularly in terms of their perceived helpfulness and usefulness, is limited. Exploring how users interact with LLMs, what factors influence their satisfaction, and how these systems can be designed to align with user needs is essential for realizing their potential in real-world applications. Addressing these gaps will provide valuable insights into optimizing LLM functionality and ensuring their seamless integration into human-centered HIS.

In summary, the possibilities offered by LLMs and CAs in facilitating creative work and problem-solving within HIS are still partly untapped. Thus, research must delve into designing, developing, and testing artifacts that leverage the adaptive capabilities of LLMs and the conversational structure of CAs, aiming to create human-centered systems that align with human cognitive processes. Advancing our knowledge in this area will be crucial to realizing the full potential of HIS, fostering a new generation of AI tools that act as genuine partners in creative work and problem-solving.

These gaps form the basis for three of this dissertation's empirical studies. von Brackel-Schmidt et al. (2023) build a taxonomy of prompting behaviours, showing how practitioners guide ChatGPT when they tackle open-ended tasks; this directly tackles the "how do users interact?" gap. Memmert et al. (2023) embed GPT-3 into a DSR canvas and analyse the perceived helpfulness, relevance, and novelty of its suggestions, thus providing

systematic evidence on LLM usefulness in complex problem-solving workflows. Memmert et al. (2024) experiment with six prompt-engineering techniques and quantify their impact on idea novelty and value, translating the rough art of prompt design into actionable guidance. Together, these studies move the discussion from “LLMs could help” to concrete, evaluated artefacts and DPs that show when, why, and with which prompts LLMs become effective collaborators in creative work and complex problem-solving.

Bridging Technological Capabilities and Use Cases

The technologies just reviewed do not stand alone; their value unfolds only when they are matched to the right collaboration setting and role within a HIS. The overview table below therefore bridges the technical distinctions (rule-based chatbots versus transformer-based LLMs) with the use cases and facilitator–collaborator roles examined in this dissertation. It makes explicit the particular AI components used in the thesis, what tasks that component addresses, and how it augments human creativity or problem-solving. This technology-to-use-case mapping provides the reference frame for the design and evaluation work set out in the following chapters.

Table 2. Similarities and differences across the use cases and technologies addressed in the thesis

Level	Chatbots	LLMs	Shared Aspects
Technology	• Mix of rule-based and NLP-driven systems • Predefined dialogue • Limited flexibility	• Transformer models • Prompt engineering • High adaptability	• Human-centered HI • NLP
Use Case	• Individual brainstorming • Group brainstorming • Group empathy mapping	• Individual brainstorming • Individual problem-solving • Group problem-solving	• Brainstorming • Human-AI team constellation
Role of AI	(Co-)facilitator: • provides constructive criticism on ideas • manages tasks and processes, e.g., saves and organizes group input, facilitates the convergent phase to arrive at a solution	Collaborator: • Builds up on human input to provide creative sparks • Analyzes context • Helps solve open-ended problems	Augments human creativity and decision-making

2.1.3 Ethics

The ethics pillar of human-centered AI emphasizes the importance of developing AI systems that are not only technologically advanced but also ethically sound and socially responsible. Key ethical considerations include protection of human values and privacy as

well as user participation in data collection, algorithm optimization and testing (W. Xu et al., 2024). Correspondingly, this part addresses information privacy in the context of human-centered HIS.

As AI systems become integral to work environments, ethical considerations, especially vis-à-vis transparency and data privacy, are crucial in all organizational settings (Sudeeptha et al., 2024). In team settings, the effectiveness of interaction and outcomes is shaped by the quality of communication, cooperation, and shared atmosphere among participants. These dynamics can be influenced by AI's capacity to collect and analyze various types of data, such as social signals, or speech patterns, which would allow it to moderate and optimize team interactions (Bittner et al., 2021; Elshan et al., 2022). However, introducing AI into teamwork settings poses several challenges, as it touches on significant technological, psychological, and societal issues, with privacy and trust as foundational concerns (Berente et al., 2021). For AI to be adopted in such scenarios, users need to be comfortable disclosing varied information.

Information disclosure (ID) refers to the willingness of individuals to share personal or sensitive data with external entities, including AI systems. The decision to disclose personal information is not merely transactional but guided by a mental framework known as *privacy calculus* (Culnan & Armstrong, 1999). This model suggests that individuals weigh perceived privacy risks against potential benefits before deciding to share information (Knijnenburg & Kobsa, 2013; Laufer & Wolfe, 1977). Privacy calculus has been extensively studied in digital interactions across domains such as social media, e-commerce, and virtual communication (Karl et al., 2022; Kroll & Stieglitz, 2021), and it is increasingly relevant in the context of AI-mediated collaboration.

Virtual meetings are one of the primary application areas for AI-facilitated collaboration, especially as remote work has surged in recent years (Cook, 2020). Integrating AI into these environments allows for enhanced meeting facilitation, improved communication, and task structuring based on real-time analysis of social signals and interaction data. However, this integration raises ethical concerns regarding surveillance and privacy. Research suggests that users often resist sharing data with AI systems in such settings, fearing privacy intrusion and loss of control (Mantello et al., 2023). Instances like Zoom's temporary implementation of attendee attention-tracking highlight these concerns, as the feature was quickly removed in response to public backlash over privacy issues (Li et al., 2024).

The specific context of ID, such as a professional or social setting, can greatly influence privacy calculus. For instance, while individuals might readily share data in a low-risk social environment, they may be more cautious in professional settings where disclosure may affect their reputation, security, or privacy (Belen-Saglam et al., 2022). Additional variables, including the nature of the information requested, the potential benefits of sharing data, and the context of AI-mediated collaboration, contribute to the complexity of disclosure decisions (H. Xu et al., 2008). When users are aware that their data will be used

to improve the effectiveness of collaboration, they may view the benefits as justifying potential risks, especially if the AI system transparently communicates its intentions and protections for personal data (Molina & Sundar, 2022). Recent field work with Microsoft's productivity assistant confirms this dynamic: this tool can provide benefit only when it ingests detailed calendar and e-mail traces, yet many employees feel unease about being monitored (Cranefield et al., 2023).

Trust is fundamental to ethical human-AI interactions and a critical determinant of ID. Studies have shown that users are more willing to share information when they trust the AI system, which is influenced by factors such as the system's transparency, fairness, and perceived usefulness (Hoff & Bashir, 2015). In work contexts, trust in AI aligns with users' expectations that the system will handle their data securely and act in their best interests (Masaviru, 2016). For instance, transparency in AI decision-making processes can significantly enhance trust, as users gain a clearer understanding of how their data is being utilized and to what ends (Patil & Kobsa, 2005; Patil & Lai, 2005). Without trust, users may resist sharing data or engaging meaningfully with AI systems, thus limiting the technology's potential benefits.

In addition to trust, the type of data being collected, especially sensitive information like emotions or cognitive states, plays a significant role in disclosure intent. AI-mediated emotion detection or sentiment analysis, for example, has the potential to enhance meeting productivity and interpersonal dynamics but often encounters resistance due to perceived invasions of personal boundaries (Park et al., 2021). This tension underscores the importance of designing AI systems that balance functionality with respect for individual autonomy and privacy preferences, allowing users to feel secure and in control of their data.

To ethically navigate the delicate balance between AI's need for data and users' privacy concerns, human-centered DPs must be embedded in AI development. Ethical frameworks such as federated learning provide one approach by enabling decentralized data sharing without requiring direct access to personal information, thereby maintaining privacy while still allowing AI to learn and improve (Wei et al., 2020). Federated learning aligns well with ethical AI design by fostering trust and addressing privacy concerns in collaborative environments. Additionally, implementing privacy nudges, design elements that encourage users to make informed decisions about data sharing, has proven effective in promoting transparency and enhancing user autonomy without restricting their freedom of choice (Ioannou et al., 2021). Furthermore, regulatory initiatives such as the European Union's AI Act impose standards for transparency, accountability, and data minimization, which AI developers must adhere to when implementing systems in sensitive settings like virtual meetings. Such regulations underscore the responsibility of AI designers to implement user-centric, privacy-conscious AI that respects both individual rights and collective group dynamics (Dell'Acqua et al., 2023). As AI applications in work continue to evolve, adherence to ethical guidelines and transparent data practices will be crucial to establishing AI as a trusted partner rather than an intrusive overseer. Moving forward, a human-centered

approach to AI ethics will be essential, ensuring that as technology advances, it does so in alignment with users' needs, values, and rights in both virtual and physical organizational spaces.

Building on the ethical challenges sketched above, this thesis tackles the open questions around ID in two successive steps. First, the research-in-progress paper (Cvetkovic et al., 2022) identifies the kinds of team-level data an AI facilitator would need, maps these data to concrete service bundles an AI could return, and formulates two initial privacy-calculus hypotheses. The subsequent, fully-fledged study (Cvetkovic et al., 2025) builds on this by empirically examining seven information types (e.g., talk time, interruptions, team sentiment) in a vignette experiment with 600 participants. It shows that trust in AI is the decisive driver of ID across all information types, while perceived benefits and risks mediate that effect. Together, these two publications contribute to the ethical aspect of the thesis: they specify which personal signals are ethically sensitive, when workers are willing to share them, and that trust-building could make AI-supported facilitation more acceptable in practice.

2.2 Provision of human input and its utilization by AI

To better understand how it is addressed in this dissertation, HIS can be broken down into three components: input (from human or AI), process (during which the input is utilized and transformed), and output (also from human or AI, depending on the sequence in which parts of the tasks are executed by human-AI parts). Effective functioning of HIS hinges on the seamless provision of human input and its optimal utilization by AI. This thesis thus addresses the human as the input provider and investigates how AI can utilize this input and what output it can give back accordingly. Applying the IPO model (Curry et al., 2006) to this context provides a structured framework for understanding how task allocation, information disclosure, and prompt engineering flow through HIS to enhance creative work and problem-solving. Thus, this section defines what constitutes human input in HIS, explores the process by which the input is handed over to AI, and the AI outputs that demonstrate how it utilizes human inputs.

Table 3. IPO model, adapted from Curry et al. (2006)

Human Input	Process	AI Output
Tasks	Task allocation	Takeover of routine tasks
Information	Information disclosure	Real-time interventions
Prompts	Prompt engineering	Contextual suggestions

2.2.1 Input

Following Morana, Pfeiffer, and Adam (2020), *Input* comprises three basic classes of human contributions: behavioural traces (e.g., actions, clicks, gestures), linguistic expressions (spoken or typed natural-language text), and physiological or affective biosignals (voice tone, heart-rate, etc.), which together convey the user's cognition, affect and behaviour to the assistance system. In the context of this dissertation, the provision of human input encompasses different types of content, all essential for enabling AI to assist effectively in creative work and problem-solving processes. This input is provided in three main forms:

Tasks refer to the specific activities or responsibilities (Gozli, 2019) that need to be addressed within the HIS context. These tasks can range from data analysis to logistical management and group facilitation. By clearly defining tasks, humans establish the operational boundaries within which AI operates, clarifying “who does what” in the team (Latto et al., 2025). The nature of the task often determines how AI can assist, whether by providing organizational support, generating content, or summarizing insights. This foundational step sets the stage for the subsequent task allocation and orchestration processes, ensuring that AI's involvement aligns with the objectives and requirements of the human collaborators (Seeber et al., 2020; Traumer et al., 2017). The study by Cvetkovic and Bittner (2022) addresses this aspect explicitly by empirically refining a task delegability framework, clarifying what tasks humans prefer delegating to AI.

Information refers to data that humans share with AI, which serves as a foundation for AI's ability to assist effectively. This encompasses a variety of potential information types that humans could provide. In the context of, e.g., creative work in teams, this includes project or task-specific data, process-related data, and personal as well as team-related interaction data (Cvetkovic et al., 2022). For instance, sharing information about team roles, project milestones, or deadlines enables AI systems to align their recommendations with the team's objectives and timeline (Bittner et al., 2021).

Prompts represent instructions or queries provided to AI to elicit specific outputs. This input type has gained prominence with the rise of LLMs where the quality and specificity of a prompt significantly influence the relevance and utility of AI responses (Mishra et al., 2022). For instance, well-crafted prompts can direct LLMs to generate relevant content, analyze data, or provide procedural guidance tailored to the context of a project (Reynolds & McDonell, 2021). The design of prompts requires careful consideration, as even subtle variations can lead to vastly different AI outputs. The thesis addresses this topic in Memmert et al. (2024), by deriving and empirically comparing different prompts for creative applications of LLMs.

2.2.2 Process

With respect to different input types in HIS, the process delineates the orchestration of the provision of the respective input to AI.

Task allocation lies at the heart of HIS, where responsibilities are distributed based on the complementary strengths of humans and AI (Dellermann, Ebel, et al., 2019). In creative work and problem-solving settings, AI should thus take on repetitive, structured, or data-intensive tasks, while humans focus on creative and strategic dimensions. The orchestration process ensures that task allocation is clear, contextually appropriate, and aligned with the overarching goals of the collaboration. For example, AI can manage the grouping and categorization of content during brainstorming sessions, allowing humans to focus on ideation and refinement (Przybilla et al., 2019). Dynamic task allocation mechanisms must also account for evolving circumstances, such as improvements in AI capabilities. Addressing this aspect, Cvetkovic et al. (2024) develop an HFS that operationalizes complementary strengths and specifies allocation of tasks between human and AI facilitators.

Information disclosure involves sharing personal or project data with AI systems to enhance task execution. Examples include sharing insights into team dynamics, emotional cues, or contextual details about ongoing projects. Disclosure is guided by trust and privacy calculus, with individuals weighing the potential benefits of improved AI support against privacy risks (Knijnenburg & Kobsa, 2013; H. Xu et al., 2008). For effective HIS, participants must feel comfortable disclosing information that allows AI to understand and adapt to group dynamics, optimizing its contributions to group goals. This disclosure process is empirically addressed in detail by Cvetkovic et al. (2025), clarifying how different contextual variables and information types affect user willingness to share sensitive data.

Prompt engineering involves designing inputs that elicit desired responses from AI. For instance, in the context of creative work, prompts might guide AI to generate creative suggestions, organize brainstorming outputs, or synthesize complex datasets. Techniques such as zero-shot, one-shot, or few-shot prompting, where examples are included to instruct the AI, have been shown to improve output relevance and accuracy in various tasks (Tom Brown et al., 2020). By iteratively refining prompts based on AI responses, users can develop more nuanced and targeted interactions with the system. The process of effective prompt engineering is not just about linguistic precision but also about understanding the context in which AI operates. For example, in creative problem-solving, prompts must balance open-endedness with specificity to encourage diverse outputs while maintaining task relevance. In this sense, prompt engineering becomes an iterative process, where users test and adapt prompts to optimize AI performance in a given scenario (Jiang et al., 2022). Recent advancements in conversational interfaces have shifted the paradigm from one-off prompts to multi-step interactions. Conversational AI systems allow for dynamic, evolving

exchanges where users can clarify, adjust, or refine their inputs over time. This interactive process reduces the cognitive load on users, as they no longer need to formulate perfect prompts upfront but can instead co-develop responses with the AI in a collaborative manner (Zamfirescu-Pereira et al., 2023). Despite these advancements, prompt engineering remains a skill that requires training and practice. Especially novice users may struggle to articulate their needs effectively, leading to suboptimal outputs. To address this, HIS must integrate support mechanisms, such as prompt templates or guided interactions, to help users craft effective prompts without requiring deep technical expertise. These mechanisms can include examples, predefined templates, or even AI-generated suggestions for improving prompt clarity and specificity (Ouyang et al., 2022). The effects of specific prompt techniques are explicitly evaluated by Memmert et al. (2024), filling a gap in understanding how humans can effectively interact with LLMs to elicit creative and relevant outputs.

2.2.3 Output

AI Output reflects the tangible contributions that AI provides in response to human input and the processes of orchestration. The Output can take form of multimodal feedback, ranging from visual or graphical displays and spoken or other auditory cues to informational messages, recommendations, and socially expressive signals that guide or motivate the user (Morana, Pfeiffer, & Adam, 2020). Valuable output should not only align with human expectations but also enhance task outcomes.

One of AI's core strengths lies in its ability to perform *structured, repetitive tasks* that would otherwise burden humans. In HIS, this includes delegating logistical and administrative activities to AI, such as organizing data, summarizing discussions, and tracking progress toward group goals (Bittner et al., 2021; Dellermann, Ebel, et al., 2019). For example, in creative teamwork sessions, AI can document ideas in real-time, categorize them into thematic clusters, and maintain a clear record of contributions for future reference (Przybilla et al., 2019). By doing so, AI not only alleviates the cognitive load on human participants but also ensures that no critical insights are lost during high-paced interactions. Such structured outputs empower human collaborators to focus on high-level tasks, such as ideation or decision-making, while delegating routine processes to AI. This structured support is addressed in this thesis through prototypes developed in Cvetkovic, Gierlich-Joas, et al. (2023) and Cvetkovic et al. (2024), demonstrating AI's capacity to facilitate creative work by clustering and summarizing content and performing task management.

Given relevant information is disclosed, AI can facilitate group collaboration by providing *real-time feedback* that enhances team dynamics and productivity. This feedback often takes the form of summaries, sentiment analysis, or reminders, ensuring that participants remain aligned and engaged throughout the process (Elshan et al., 2022). For instance, during brainstorming sessions, AI can identify unaddressed discussion points or suggest topics that require further exploration, promoting balanced participation among team

members (Siemon & Strohmman, 2021). Feedback outputs are particularly valuable in addressing interpersonal challenges within teams. By analyzing emotional cues or conversational dynamics, AI can suggest strategies to foster trust, mitigate conflicts, or encourage quieter team members to contribute. However, the acceptance of such feedback hinges on the importance of trust and transparency in the system's design (Masaviru, 2016). This aspect is addressed conceptually in Cvetkovic et al. (2022) by proposing concrete AI interventions aligned with disclosed information and subsequently tested empirically by Cvetkovic et al. (2025), to assess whether the perceived benefits of the proposed feedback outweigh the perceived risks of disclosing corresponding information and how this is affected by trust in AI and transparency in providing the feedback.

Given well-crafted prompts, the output could be *contextual suggestions* that may include actionable recommendations, creative ideas, or prioritized task lists tailored to specific needs and goals (Tom Brown et al., 2020; Reynolds & McDonell, 2021). For example, in complex problem-solving scenarios, AI might analyze previous project data, identify potential risks, and propose strategies to mitigate them. Similarly, in creative work, AI can generate alternative perspectives or build on human input to offer new directions for exploration (Gero et al., 2022). Such outputs exemplify the complementary strengths of humans and AI, where the latter acts as a "co-creator" that augments rather than replaces human decision-making (Zhu & Luo, 2022). This type of output is specifically addressed in studies by Memmert et al. (2023) and Memmert et al. (2024), showing how prompt engineering can yield high-quality, contextually relevant suggestions.

In summary, while advancements in human-centered AI and HIS have demonstrated their potential to enhance creativity, collaboration, and problem-solving, critical gaps remain. The dynamic nature of identifying suitable input-output pairs, coupled with the complexities of orchestrating processes and interactions within specific use case boundaries, underscores the need for ongoing research. Central to these challenges is the imperative to maintain a human-centered focus, ensuring that AI systems are not only effective but also transparent, ethical, and aligned with user needs and values. By addressing these gaps and reinforcing human agency in the design and deployment of HIS, this research aims to contribute to the development of AI as a collaborative partner, fostering trust, innovation, and meaningful human-AI synergy across diverse contexts.

3 Research Design

3.1 Research Paradigm

The practical problems addressed in this dissertation and the associated research goal are characterized by a necessity for an integrative, problem-solving research paradigm resulting in concrete design knowledge. An applied Information Systems (IS) research paradigm, DSR, focuses on solving complex problems through the creation and evaluation of innovative artefacts (Gregor & Hevner, 2013; Hevner et al., 2004) and is thus the methodology of choice. The artefacts can signify different contributions based on their maturity, completeness and abstraction level. So do design theories represent more mature, abstract and complete design artefacts, while software instantiations convey more specific, limited and less mature knowledge. In between the two fall, e.g., constructs, methods, models and DPs and constitute a middle-ranged contribution described as nascent design theory (Gregor & Hevner, 2013). DSR puts emphasis on the development and evaluation of prescriptive knowledge and is characterized as constructive research (Dresch et al., 2015), in contrast to explanatory research, representative of social/behavioral sciences. Hence, the research objectives of DSR are typically more pragmatic, concentrating on what can be achieved and what is beneficial for shaping potential futures, rather than on what currently exists. However, this does not mean that DSR and behavioral sciences are strictly separable. As Hevner et al. (2004) note “Technology and behavior are [...] inseparable in IS research.”

DSR aims to produce actionable design knowledge that can guide the development of systems and solutions in real-world settings, as emphasized by Hevner (2007) and Thuan et al. (2019). Hevner (2007) presents DSR as a process driven by three distinct but interconnected cycles: the relevance cycle, the design cycle, and the rigor cycle. The relevance cycle ensures that the research addresses real-world problems by understanding the context and identifying opportunities for innovation. The design cycle focuses on the creation of artefacts, theories, models, or systems, that are developed and iteratively refined to meet these challenges. Finally, the rigor cycle ensures that the solutions are grounded in existing theoretical knowledge, contributing to both the research community and practical application.

Given that the goal of this dissertation is to enhance human-centered HIS for creative work and problem-solving, DSR's focus on both rigor and relevance aligns with the research aims. The relevance cycle ensures that the challenges and opportunities in the selected use cases are accurately understood, while the rigor cycle builds on existing knowledge to refine the design of HIS. The iterative process of artefact creation and evaluation fits well with the design and validation of human-centered HIS, allowing for continuous improvement in response to real-world challenges. Through DSR, this dissertation not only

contributes to the theoretical body of knowledge but also provides practical, validated solutions that improve the process and outcomes of creative work and problem-solving (Baskerville et al., 2018; Hevner, 2007).

3.2 Research Strategy

The goal of this dissertation is to enable human-centered HIS for creative work and problem-solving by developing validated design knowledge that supports the effective provision of human input and its utilization by AI, leveraging the complementary strengths of both humans and AI. Following DSR's objectives, this dissertation advances both theoretical insights and practical guidelines, focusing on creating artifacts that align with the strengths and collaborative dynamics of human-AI teams (Baskerville et al., 2018).

Throughout this cumulative dissertation, DSR activities extend beyond theoretical exploration to include empirical investigation within the application environment. These artifacts, implemented iteratively in alignment with DSR's three cycles - relevance, rigor, and design - are validated through building and evaluation phases to ensure their utility, measured by their ability to meet predefined goals (Hevner, 2007; Hevner et al., 2004; March & Smith, 1995). Each artifact independently contributes to the overarching design knowledge while also integrating with others to form a coherent knowledge base that evolves and accumulates across several publications (Rothe et al., 2020; vom Brocke et al., 2020).

By adhering to rigorous DSR guidelines and standards, including established reference processes and normative steps (Kuechler & Vaishnavi, 2012; Peffers et al., 2007) this dissertation advances the intersection of DSR and (Human-Computer Interaction) HCI (Adam et al., 2021; Hevner & Zhang, 2011), producing cumulative insights that bridge scholarly knowledge and practical application in the design of HIS.

Serving the overarching RG, the interconnected articles in this thesis address multiple RQs (see Table 4). First, the thematic journal publication, **Art. 1** (Oeste-Reiß et al., 2021) sets the stage for the use case of HIS in knowledge work by outlining new design opportunities for the reorganization of work at the interface of humans and AI (RQ 1b). Furthermore, **Art. 2** (Cvetkovic & Bittner, 2022) identifies knowledge work tasks in postgraduate research that humans are willing to delegate to AI; **Art. 3** (Cvetkovic et al., 2022) identifies tasks in the space of creative work in teams where AI could augment humans based on their inputs, and **Art. 4** (von Brackel-Schmidt et al., 2023) shows how AI is used for different complex problem-solving scenarios to build and sharpen understanding or overtake tasks. With this, these articles contribute to the identification of suitable HI use cases as well as the corresponding opportunities tied with the provision of human input and its utilization by AI. **Art. 2** (Cvetkovic & Bittner, 2022) also delves into the problem space, i.e., the prerequisites for a willful task delegation to AI (RQ 1a). Additionally, the produced knowledge in the problem space from **Art. 2** is complemented by **Art. 3** (Cvetkovic et al.,

2022), by drawing on the information privacy literature to hypothesize potential challenges in human input utilization by AI with regards to privacy concerns, and **Art. 10** (Cvetkovic et al., 2025) by confirming that these privacy concerns are indeed a major factor in determining whether certain information is shared with AI. Human-centered design knowledge for HIS is developed to enhance creative work and problem-solving in previously identified use cases (RQ 2a). **Art. 5** (Cvetkovic, Rosenberg, & Bittner, 2023) presents a CA that assumes a specific feedback-providing role in creative individual idea generation. In **Art. 6**, Cvetkovic, Gierlich-Joas, et al. (2023) design and develop a CA that facilitates group work by enriching and distilling unstructured silos of information. **Art. 7** (Cvetkovic et al., 2024) goes one step further and instantiates a prototypic HIS in which a human facilitator and a CA facilitator work together to moderate group work, each performing tasks pertaining to their complementary strengths. **Art. 8** (Memmert et al., 2023) and **Art. 9** (Memmert et al., 2024) utilize (at the time of each manuscript preparation) state-of-the-art LLMs and develop prompt templates to enhance complex problem-solving and idea generation, respectively. By means of the respective evaluations in articles (**Art. 5-9**), the effects of the created solutions on processes and outcomes of creative work and complex problem-solving are assessed (RQ 2b). With respect to RQ 3, all published articles (**Art. 1-10**) contribute to the knowledge base by presenting validated prescriptive design knowledge and insights into boundary conditions affecting HIS.

Table 4. Contributions of articles to RQs

#	Article	RQs	Contribution to RQs
Art. 1	<i>Oeste-Reiß et. al (2021)</i>	1b	Identification of opportunities for HIS in the knowledge work context; Specification of a HIS use case on the example of journalism
		3	Presentation of three archetypes of human-machine tasks within hybrid knowledge work systems and an outline of three design dimensions of such systems
Art. 2	<i>Cvetkovic & Bittner (2022)</i>	1a	Identification of prerequisites for task delegation to AI in knowledge work by recording the factors aiding or hindering the handover of tasks to AI; Recording of tasks that should be delegated to AI
		3	Validation of a task delegability framework in a knowledge work context

Art. 3	<i>Cvetkovic et al. (2022)</i>	1a	Identification of prerequisites for information provision to AI in collaborative creative work
		1b	Identification of opportunities that AI could offer given human input provision
		3	Proposal of hypotheses to test ID to AI in creative work in teams
Art. 4	<i>von Brackel-Schmidt et al. (2023)</i>	1b	Identification of use cases for LLM usage for complex-problem solving; Identification of prompt objectives, styles and settings of users.
		3	A user-centric taxonomy of LLM usage
Art. 5	<i>Cvetkovic, Rosenberg, & Bittner (2023)</i>	2a	Proposal and instantiation of design knowledge for CAs in creative work
		2b	Validation of the proposed design knowledge
Art. 6	<i>Cvetkovic, Gierlich-Joas et al. (2023)</i>	3	Proposal of CA-supported facilitation of tasks and processes
Art. 7	<i>Cvetkovic et al. (2024)</i>	2a	Proposal and instantiation of design knowledge for HIS in creative work in teams
		2b	Validation of the proposed design knowledge
		3	Task allocation framework for HIS in creative work in teams
Art. 8	<i>Memmert et al. (2023)</i>	2a	Proposal and instantiation of design knowledge for prompting LLMs in creative work and complex problem-solving

Art. 9	<i>Memmert et al. (2024)</i>	2b	Validation of the proposed design knowledge
		3	Proposal of roles in HIS for value co-creation with LLMs
Art. 10	<i>Cvetkovic et al. (2025)</i>	1a	Proposal of a comprehensive research model of ID to AI in virtual meetings
		3	Validation of the research model

To support the accumulation and evolution of design knowledge within a unified DSR project, the research follows the three established DSR cycles: relevance, rigor, and design (Hevner, 2007; Hevner et al., 2004; Thuan et al., 2019). Each cycle is guided by distinct epistemological perspectives, aligning with diverse knowledge interests and methodological approaches (Niehaves, 2005). The following sections elaborate on the research activities conducted within these cycles to address the RQs and illustrate the interconnectedness of the individual publications across the DSR framework. Figure 3 depicts how the papers contribute to multiple RQs and span different DSR cycles, highlighting their integration and alignment within the overarching research project. Additionally, the ways in which knowledge is utilized and generated are contextualized, explained, and visualized to provide a comprehensive overview of the contributions of this cumulative dissertation.

3.2.1 Relevance

Guided by pragmatism as a foundational philosophy, this research emphasizes generating practically useful insights, evaluated through their real-world implications (Hevner & Chatterjee; Venkatesh et al., 2013). Within the relevance cycle, a mixed epistemological approach is adopted to analyze the requirements and opportunities of HIS for creative work and problem-solving. This includes both positivist elements, such as quantitative analyses (e.g., confirmatory factor analysis, linear regression) to empirically test stipulated frameworks, and interpretivist perspectives, particularly when collecting and interpreting subjective perceptions of artefacts (Niehaves, 2007; Venkatesh et al., 2013).

The relevance cycle serves the identification of challenges and opportunities in real-world application domains (Hevner et al., 2004; Kuechler & Vaishnavi, 2012; Peffers et al., 2007; vom Brocke et al., 2020). Accordingly, RQ 1a and RQ 1b address the requirements and opportunities for the provision of human input and its utilization by AI in human-centered HIS in creative work and problem-solving domains. The former informs the research by framing the problem space (Figure 3, activity A) through the systematic identification of

the practical challenges in the domain environment, stakeholder needs, and theoretical underpinnings (Hevner et al., 2004). This framing ensures that the research is grounded in real-world scenarios, focusing on issues relevant to practitioners and end-users. By systematically capturing requirements from key stakeholders in the environment, the alignment of emerging design knowledge and tangible artifacts with their specific challenges can be effectively ensured (Gregor & Hevner, 2013; vom Brocke et al., 2020). To this end, Cvetkovic and Bittner (2022), Cvetkovic et al. (2022), and Cvetkovic et al. (2025) accessed contextual knowledge via the relevance cycle to inform design activities (Figure 3, Activities A and B). Accordingly, RQ 1a addresses the requirements for effective human input provision and its utilization by AI in human-centered HIS. These requirements capture the challenges related to task delegation to AI (Cvetkovic & Bittner, 2022), and information disclosure to AI in collaborative settings (Cvetkovic et al., 2025; Cvetkovic et al., 2022), providing the foundation for defining the problem space.

Conversely, RQ 1b emphasizes the identification of opportunities, aligning the relevance cycle with the solution space (Figure 3, activity B). Opportunities are framed around AI's potential to augment creative work and problem-solving, fostering innovative practices while respecting human agency and ethical constraints. For example, Cvetkovic and Bittner (2022) identify research tasks that researchers want to delegate to AI, and Cvetkovic et al. (2022) identify how and which tasks, processes and interactions can be (co)-facilitated by AI in the space of creative work in teams. These insights inform the design and evaluation cycles, ensuring that proposed solutions are both relevant and actionable. The iterative nature of the relevance cycle allows for continuous refinement, with findings from the solution space feeding back into problem definition and stakeholder understanding (Figure 3, activity G) (Peppers et al., 2007; vom Brocke et al., 2020).

3.2.2 Rigor

The rigor cycle relies primarily on positivist assumptions (Kuhn, 1962), aiming to build on established knowledge and validate findings through empirical testing, which supports the development of design knowledge with potential generalizability (Gregor & Hevner, 2013; van Aken, 2004). Through this structured, evidence-based approach, the research aims to create outcomes that contribute broadly to the field while addressing real-world application needs (Hevner et al., 2004).

To ensure the rigor of this DSR project, both the usage of existing knowledge to inform the environment and the design (Figure 3, Activity H) and the contribution of new knowledge to the knowledge base (Activities E and F) were systematically implemented. By drawing on Ω -knowledge (descriptive and explanatory) and λ -knowledge (prescriptive), activity H played a critical role in grounding the design activities in established theories, frameworks, and empirical findings (Drechsler & Hevner, 2018; Hevner, 2007). This ensured that the design process was not only innovative but also aligned with existing scientific and practical knowledge (Hevner et al., 2004; Jones & Gregor, 2007).

For instance, kernel theories such as Privacy Calculus (Culnan & Armstrong, 1999) used in Cvetkovic et al. (2022) and Cvetkovic et al. (2025) and the Task Delegability Framework (Lubars & Tan, 2017) which inspired Cvetkovic and Bittner (2022), and Cvetkovic et al. (2024), informed the conceptualization of challenges and design opportunities in HIS. This activity directly supported RQ 1a and RQ 1b by refining the understanding of opportunities and prerequisites for HIS in creative and problem-solving domains. Additionally, theoretical advancements in prompt engineering (e.g., Mishra et al. (2022)) and creativity support systems (Shneiderman, 2007) provided actionable guidelines for designing HIS artifacts (Kuechler & Vaishnavi, 2012; Walls et al., 1992). Hereby, Activity H refers to RQ 2a and RQ 2b by drawing on knowledge to inform design activities. Furthermore, for Cvetkovic, Rosenberg, and Bittner (2023), Cvetkovic, Gierlich-Joas, et al. (2023), and Cvetkovic et al. (2024) literature reviews were conducted to derive design requirements from the existing body of knowledge. These reviews not only synthesized insights from prior research but also served to supplement the practice-based requirements identified during stakeholder engagement and environmental analysis. This dual approach ensured a holistic understanding of the requirements for HIS and provided a robust foundation for the subsequent design and development activities (vom Brocke et al., 2020). The alignment between practice-driven and literature-driven requirements further reinforced the rigor and relevance of the design process.

Via Activity E, the insights from RQ 1a flow into the knowledge base. Theoretical models tested in Cvetkovic and Bittner (2022) and Cvetkovic et al. (2025) contribute to the descriptive and explanatory Ω -knowledge by empirically validating the prerequisites and challenges of human-centered input provision and utilization in HIS. These studies extend existing theories, e.g., the Privacy Calculus (Culnan & Armstrong, 1999) and the Framework of Delegation to Agentic IS artefacts (Baird & Maruping, 2021), by contextualizing them within the domain of HIS. This flow from the problem space to the knowledge base ensures that the identified challenges and empirical findings are systematically cataloged, offering a foundation for future theoretical and practical advancements.

Through activity F, the novel contributions derived from design efforts (Cvetkovic, Gierlich-Joas, et al., 2023; Cvetkovic, Rosenberg, & Bittner, 2023; Cvetkovic et al., 2024; Memmert et al., 2023, 2024) were fed back into the knowledge base. These contributions include a taxonomy, DPs, validated artifacts, and conceptual models that advance λ -knowledge and provide prescriptive guidance for future research and practice. By systematically integrating these outputs, the dissertation ensures that the knowledge base is enriched with actionable insights, which extend beyond routine design outcomes to address complex, real-world challenges (Drechsler & Hevner, 2018; vom Brocke et al., 2020).

3.2.3 Design

The design cycle evaluation incorporates a dual epistemological approach. Interpretivism (Myers, 2008) guides the collection of subjective perceptions from participants, capturing their personal experiences and insights, which provides depth and context to the findings. Simultaneously, positivism is integrated through the collection of log interaction data to quantify behavior and assess patterns in user interactions. This mixed-methods evaluation ensures a comprehensive understanding by blending qualitative and quantitative data, enhancing both the reliability and the contextual richness of the results (Venkatesh et al., 2013; Weber, 2004).

At the heart of any DSR project lies the iterative process of creating and refining design knowledge, an approach central to the design cycle (Hevner, 2007). This is formally split into build and evaluate cycles, where generated design knowledge is operationalized with instantiations and then demonstrated and validated for its effectiveness (Jones & Gregor, 2007; Kuechler & Vaishnavi, 2012; Peffers et al., 2018). Guided by the build and evaluate cycles of the DSR, design activities in this dissertation focused on two interconnected RQs: the design of human-centered HIS artifacts (RQ 2a) and their evaluation (RQ 2b). Design knowledge is derived from developing and evaluating socio-technical artifacts. For technical artifacts, prototyping was used as an established procedure used in DSR (Baskerville et al.; Jones & Gregor, 2007). Accounting for different user scenarios and the evolving characteristics of AI-based solutions (Yang et al., 2020), different types of prototypes were developed in the articles, and span three types: CAs, an HFS combining a CA and a human facilitator along with a task allocation framework that guides how this system operates, and prompt templates for LLMs. Each artifact type represents a distinct design endeavor, with varying levels of maturity and prescriptive knowledge.

Two CAs were developed to facilitate brainstorming in individual (Cvetkovic, Rosenberg, & Bittner, 2023) and team scenarios (Cvetkovic, Gierlich-Joas, et al., 2023). Design requirements (DRs) were formulated based on literature reviews and/or empirical stakeholder insights, and DPs were derived to codify the prescriptive knowledge necessary for artifact construction (Gregor et al., 2020; Maedche et al., 2021). These artefacts represent situational instantiations, i.e., fully developed systems that were prototyped and tested in semi-naturalistic settings, while the DPs represent a more abstract knowledge aiming toward a nascent design theory (Gregor & Hevner, 2013). In terms of impact, the CA for individual brainstorming (Cvetkovic, Rosenberg, & Bittner, 2023) represents a proof-of-concept, and the CA for group brainstorming (Cvetkovic, Gierlich-Joas, et al., 2023) a proof-of-value (Doyle & Luczak-Roesch, 2020) as it investigates the added value across two different stakeholder groups (potential creative teams and facilitators).

Cvetkovic et al. (2024) develop an HFS that integrates human and AI facilitation, with a task allocation framework guiding the distribution of responsibilities. This framework, grounded in the principles of HI (Dellermann, Ebel, et al., 2019) and the facilitation framework (Beranek et al., 1993; Clawson et al., 1993; Leimeister, 2014), represents a nascent design theory artifact, offering prescriptive guidance on task allocation between

humans and AI. The accompanying HFS was instantiated and evaluated through iterative prototyping and user feedback and as such combines formative, ex-ante and summative, ex-post evaluations (Pries-Heje et al., 2008). In terms of impact, this prototype also represents proof-of-value, as it is evaluated with two stakeholder types in semi-naturalistic settings.

Unlike the CAs and HFS, the prompt templates are less mature artifacts. Their design did not involve formalized DRs or DPs but relied on exploratory application of existing research on prompt engineering (e.g., Mishra et al. (2022)). These templates provide practical heuristics for framing user inputs to elicit relevant and high-quality responses from LLMs in creative (Memmert et al., 2024) and problem-solving contexts (Memmert et al., 2023).

To evaluate the designed artifacts, iterative testing and feedback loops were employed. Appropriate evaluation methods were informed by the rigor cycle, aligning with the objectives of each research endeavor. The evaluation focused on determining the goal attainment (efficacy, effectiveness, and validity), the utility and feasibility of the artefacts (Prat et al., 2015) by investigating the alignment of the artefacts with the derived DPs and user scenarios as well as on assessing the impact on the environment and the artefact users (Sonnenberg & vom Brocke, 2012). These attributes were assessed at different time points. For the CAs and the HFS (Cvetkovic, Gierlich-Joas, et al., 2023; Cvetkovic, Rosenberg, & Bittner, 2023; Cvetkovic et al., 2024), user tests were conducted in semi-naturalistic settings, employing mixed methods such as surveys, interviews, focus groups, and chatlog analysis. In the iterative evaluation procedure, pre-tests with a few exemplary users were performed to assess the completeness, operability, and feasibility of the design knowledge formative and ex-ante (Pries-Heje et al., 2008; Sonnenberg & vom Brocke, 2012; Venable et al., 2016). The main evaluation cycle assessed the alignment with proposed DPs and simultaneously the impact on the users in a summative, ex-post fashion (Pries-Heje et al., 2008; Sonnenberg & vom Brocke, 2012; Venable et al., 2016). For the prompt templates (Memmert et al., 2023, 2024), surveys were carried out to evaluate the AI-generated outputs in terms of their helpfulness, relevance, usefulness, and quality ex-post (Pries-Heje et al., 2008; Sonnenberg & vom Brocke, 2012; Venable et al., 2016).

Figure 3 depicts the research activities, i.e., corresponding papers mapped to the RQs and cycles which they address. It also conveys how individual papers are interconnected in addressing multiple RQs and how the knowledge flows from one design cycle to another (A-G).

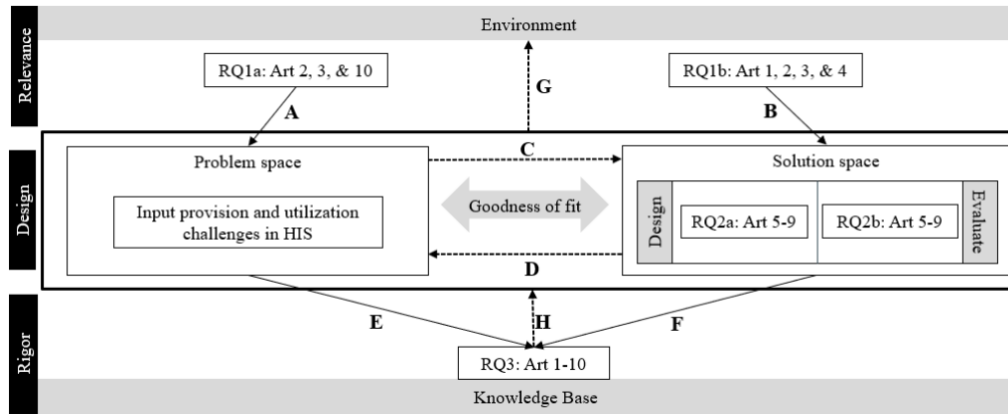


Figure 3. DSR strategy

3.3 Knowledge Utilization and Contribution

DSR activities are undertaken in an iterative and cyclical process (Figure 3) aimed at addressing the challenges identified in the problem space by systematically exploring and refining solutions within the solution space (Hevner, 2007; Hevner et al., 2004). This approach integrates multiple knowledge paths, drawing on, producing, and contributing to both descriptive and prescriptive knowledge (Ω and λ) (Drechsler & Hevner, 2018). The process involves utilizing existing theoretical insights to frame the problem, developing actionable design knowledge and entities, and iteratively refining these contributions to advance both practical solutions and the underlying knowledge bases (Figure 4).

Existing Ω knowledge (research activity A, path 1 - use Ω) was applied from the foundational concepts, including the HI literature (Dellermann, Ebel, et al., 2019), Privacy Calculus Theory (Culnan & Armstrong, 1999), and Socio-Technical Systems theory (Eason, 2008). Based on inductive data collection procedures in the environment (vom Brocke et al., 2020), research activity E (path 2 - contribute Ω) uncovered existing factors aiding or hindering the provision of input to AI in creative work and problem-solving contexts. As part of activity E, these insights are contributed as Ω -knowledge.

Based on these insights, design knowledge in the form of DPs and frameworks, as well as entities, was created in the solution space. Path 3 contributes to design knowledge (λ) by codifying frameworks and principles developed through solution design (research activity B). For instance, frameworks for allocating tasks in HIS or principles for AI-augmented creative work are derived from solution design efforts. As part of activity F, these insights are contributed as λ -knowledge.

The iterative evaluation cycle between the problem and solution spaces is captured by research activities C and D (Figure 3). Research activity C provides insights from the problem space to improve solution design, while activity D refines the problem space by evaluating how well solutions address the identified challenges. This cyclical process ensures continuous improvement and alignment between the theoretical and practical aspects of the research (Peffer

et al., 2007; Sein et al., 2011). To bridge the solution space and the environment, research activity G uses knowledge generated from the design cycle to send insights back into the environment. These insights, informed by DPs and evaluated solutions, provide actionable feedback that can influence practices, policies, or user behavior in real-world settings. This ensures the practical relevance and applicability of the generated knowledge beyond the research context (Baskerville & Myers, 2004).

Path 4 leverages prescriptive knowledge (λ) to produce solution knowledge and entities (research activity H). Existing design knowledge, such as prompt engineering and HI principles, informs the development of DPs and artifacts like CAs and prompt templates. This ensures that solutions are grounded in validated methods and actionable design knowledge, addressing specific needs in collaborative and problem-solving contexts.

Path 5 focuses on contributing to design entities (λ) by providing instantiated solutions, such as CAs for group brainstorming or HFS. These entities are tested, refined, and shared as practical, reusable artifacts within the knowledge base (research activity F).

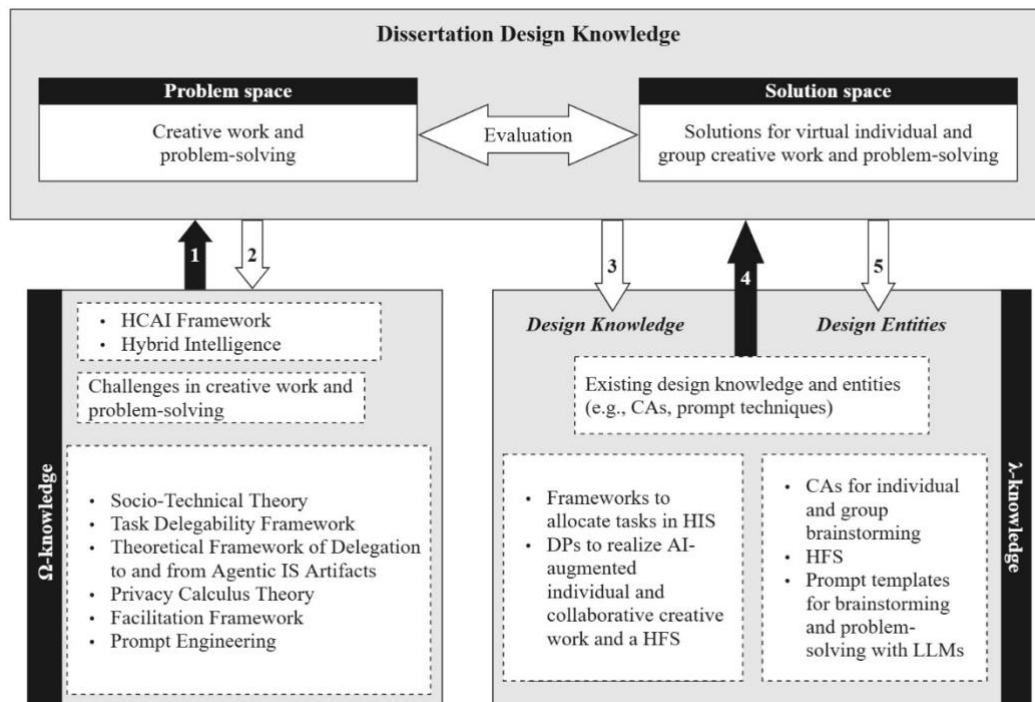


Figure 4. Utilization of and contribution to the design knowledge. Adapted from Drechsler & Hevner (2018).

3.4 Research Methods

In alignment with the three DSR cycles, a variety of research methods are chosen to address the contextual environment, build upon the knowledge base, and develop and evaluate design knowledge. Combining different methods supports more comprehensive research outcomes than any single method could achieve on its own (Mingers, 2001). Throughout

the dissertation's research journey, research methods are grouped into those supporting artifact design and artifact evaluation, ensuring alignment with the relevance, rigor, and design cycles of DSR. To this end, literature reviews, taxonomy development, and both qualitative and quantitative data collection and analysis methods were used. This mixed-methods approach, grounded in methodological pluralism (Mingers, 2001; Niehaves, 2005; Venkatesh et al., 2016), maximizes knowledge contributions and is tailored to the dissertation's complex, multi-phase nature. Each method is further defined and detailed in the following subsections.

3.4.1 Artifact Design

Artifact design methods integrate insights from both the relevance and rigor cycles to ensure that the design process is both grounded in practical needs and informed by existing scientific knowledge (Hevner et al., 2004). Requirements are derived through a combination of literature reviews, which provide theoretical and empirical foundations, and user interviews, which capture real-world challenges and stakeholder perspectives.

Literature Reviews

Literature reviews are essential for advancing scientific knowledge by synthesizing existing work and identifying research gaps in evolving fields. They provide the basis for understanding both the problem space and the solution space in research projects, ensuring that contributions are grounded in existing knowledge (Gregor & Hevner, 2013; vom Brocke et al., 2020). Depending on their goals, literature reviews can serve to summarize and aggregate results, highlight knowledge gaps, or establish theoretical foundations prior to empirical research (vom Brocke et al., 2009; vom Brocke et al., 2015; Webster & Watson, 2002).

In this dissertation, multiple types of literature reviews were applied to achieve specific objectives across the individual papers. These include systematic reviews, theoretical reviews, and narrative reviews (Paré et al., 2015) with both exhaustive and representative coverage (Cooper, 1988). By systematically applying these reviews, the research in this cumulative dissertation positions itself within the existing body of knowledge, identifies essential research gaps, and ensures rigor in the development of solutions and empirical evaluations. Each review type adheres to established methods and frameworks to ensure reliability, validity, and objectivity (Cooper, 1988; vom Brocke et al., 2015; Wolfswinkel et al., 2013).

To ground DSR on existing scientific knowledge, a *systematic review* was conducted in a select study. Systematic reviews follow rigorous search processes to investigate both problem and solution spaces (Gregor & Hevner, 2013; vom Brocke et al., 2020). In Cvetkovic et al. (2024), a systematic review (Tranfield et al., 2003) was applied to synthesize findings on HI, CAs, and facilitation frameworks, leading to the identification of DRs.

Theoretical reviews were applied to establish conceptual frameworks and build research models. The papers Cvetkovic et al. (2022) and Cvetkovic et al. (2025) synthesized established theories, such as privacy calculus, to justify hypotheses and position the studies within existing theoretical landscapes.

Lastly, *narrative reviews* were conducted to derive objectives of a solution for artefact development. In studies aimed at deriving requirements for solution development, narrative reviews with representative coverage were employed. These reviews selectively synthesized relevant literature to inform the design of artifacts. For example, Cvetkovic, Rosenberg, and Bittner (2023) and Cvetkovic, Gierlich-Joas, et al. (2023) integrated insights from facilitation frameworks, CAs, and creativity research to derive design requirements for CAs for individual and group brainstorming.

Additionally, this type of review is employed in all articles, independent of the artifact development or evaluation, to qualitatively summarize existing work and contextualize research questions. For example, Cvetkovic and Bittner (2022) explored a task delegation framework and factors influencing delegation of tasks to AI, setting the stage for empirical validation. Similarly, Memmert et al. (2024) employed a narrative review to provide a conceptual background for evaluating prompt techniques in collaborative idea generation with AI. To identify input-output pairs in Human-AI Collaboration, Cvetkovic et al. (2022) further adopted a snowball sampling technique (Wohlin, 2014), which involves identifying a central, relevant paper and then using its references (backward snowballing) and citations (forward snowballing) to discover additional pertinent literature.

Taxonomy Development

Taxonomies are structured classification schemes that organize complex domains by categorizing objects based on shared characteristics. This systematic arrangement facilitates understanding, analysis, and retrieval of information. Taxonomies are essential tools for sensemaking in rapidly evolving technological landscapes, enabling researchers and practitioners to navigate and manage complex datasets effectively (Kundisch et al., 2021; Lösser et al., 2019). The development of taxonomies in IS has been addressed by scholars such as Nickerson et al. (2013), who proposed a method to counter the prevalence of ad-hoc approaches in taxonomy construction.

In this cumulative dissertation, a taxonomy was developed (von Brackel-Schmidt et al., 2023) using data from a Prompt-a-thon event. Following Kundisch et al. (2021) guidelines, which build on Nickerson et al. (2013) established approach, the taxonomy integrates the six DSR activities of Peffers et al. (2007) for a systematic and flexible classification of LLM usage. Several iterations alternated between conceptual-to-empirical and empirical-to-conceptual approaches to capture the knowledge state comprehensively. The process allowed multiple characteristics per dimension to accommodate the variability in input/prompt designs. Following scientific recommendations, qualitative and quantitative

evaluations were conducted to ensure the taxonomy's validity, usability, and relevance for both theoretical and practical applications.

Qualitative Methods

Qualitative research methods are essential for understanding subjective meanings, experiences, and actions, allowing researchers to explore the complexity of social realities (Fossey et al., 2002; Miles & Huberman, 1994). These methods involve collecting data from diverse sources, including people and textual materials, through techniques such as interviews, focus groups, and open-ended survey responses (Maxwell, 1992; Venkatesh et al., 2013). In the IS domain, qualitative approaches are particularly valuable for deriving practical insights, eliciting design requirements, and evaluating artifacts in their real-world contexts (Goldkuhl, 2012; Lee & Hubona, 2009). They also play a key role in iterative development cycles by providing rich, contextual feedback for improving artifacts (Hevner et al., 2004; Prat et al., 2015).

In this dissertation, qualitative methods were employed to inform the design of artifacts. To elicit design requirements, *semi-structured interviews* were conducted with stakeholders from relevant domains, such as facilitators from the industry (Cvetkovic, Gierlich-Joas, et al., 2023), supporting the relevance cycle. Semi-structured interviews are widely regarded as a rigorous method in qualitative research, combining a structured framework of guiding questions with the flexibility to explore participants' perspectives in depth (Myers & Newman, 2007). This method allowed the interviewees to provide detailed insights into their needs and practices, which were synthesized into actionable design requirements for the artifact. The data collected was systematically analyzed using qualitative content analysis (Mayring, 2014), ensuring rigor in deriving actionable insights.

Qualitative feedback from surveys also contributed to the iterative development of the artifacts. Open-ended survey questions were used to refine vignettes and scales, ensuring their clarity and contextual relevance for the potential real-world applications (Cvetkovic et al., 2025). In Cvetkovic et al. (2024) working documents and recorded interaction logs from a preliminary experiment were reviewed to identify recurring themes and generate categories for improving artifact design.

3.4.2 Artifact Evaluation

Artifact evaluation methods focus on validating the utility, effectiveness, and applicability of artifacts (Prat et al., 2015). These methods ensure that the developed artifacts align with DPs, meet user needs, and contribute to knowledge generation.

Experiments

Experiments allow for a controlled assessment of artifact qualities, such as usability, utility, and alignment with DPs (Hevner et al., 2004). In the context of DSR, these are often referred to as user tests as they involve participants interacting directly with prototypes to assess their performance and user experience.

In this dissertation, user tests were conducted to evaluate instantiated prototypes in semi-naturalistic settings. They are integral to both ex-ante and ex-post evaluation phases (Pries-Heje et al., 2008; Venable et al., 2012, 2016) and align with a structured prototyping approach (Österle et al., 2011; Wilde & Hess, 2006), where rapid development cycles enable researchers to refine and re-test artifacts iteratively. In this way, prototypes can be adjusted to address issues identified during user testing, thereby ensuring that design knowledge is effectively realized in functional systems (Kopenhagen et al., 2012; Wilde & Hess, 2006). In the context of this thesis, participants interacted with prototypes such as CAs (Cvetkovic, Gierlich-Joas, et al., 2023; Cvetkovic, Rosenberg, & Bittner, 2023) and HFS (Cvetkovic et al., 2024), to assess factors such as usability, task performance, and the prototypes' ability to support creative and problem-solving tasks. Feedback gathered from participants, including qualitative insights and interaction log analyses, was used to refine the prototypes iteratively. Similarly, the HFS, which combined human and AI facilitation, was evaluated to determine its effectiveness in distributing facilitation tasks, with results guiding further adjustments. This iterative testing ensured that the artifacts were both practical and human-centered.

In addition to instantiated prototypes, experiments also investigated conceptual artifacts through surveys. For example, Cvetkovic et al. (2025) utilize experiments to manipulate different AI interventions via the vignette method (Aguinis & Bradley, 2014), while Cvetkovic and Bittner (2022) assess a framework applied to a postgraduate work context to test its contextual fit and practical relevance.

Qualitative Methods

Artifact evaluation relied on several qualitative methods to validate the DPs and assess the artifacts' utility. In Cvetkovic, Gierlich-Joas, et al. (2023), *focus groups* were conducted for a summative, ex-post evaluation of the artefact with two distinct groups of stakeholders: facilitators from the industry and student teams acting as potential creative teams. In Cvetkovic et al. (2025), focus groups with innovation experts were used to incorporate expert perspectives vis-à-vis the content validity of the vignettes for AI interventions in the investigated domain. These sessions were guided by structured guidelines to ensure consistency and rigor. The structured discussion format facilitated the collection of diverse perspectives, making it an effective method for artifact evaluation (Tremblay et al., 2010). Similarly, *semi-structured interviews*, also conducted following interview guidelines, were used to validate the DPs in Cvetkovic et al. (2024).

Additionally, *open-ended questions* in surveys were used to gather qualitative feedback on a key DP in Cvetkovic, Rosenberg, and Bittner (2023), and to provide additional feedback for DPs in Cvetkovic et al. (2024).

In cases where qualitative data exceeded manageable volumes for manual analysis, *qualitative content analysis* (Mayring, 2014) was applied. For example, focus group transcripts were analyzed to validate DPs (Cvetkovic, Gierlich-Joas, et al., 2023) and chat

logs from interaction studies with ChatGPT were reviewed to identify patterns in prompt techniques (von Brackel-Schmidt et al., 2023). Additionally, qualitative content analysis was applied to analyze the outputs of a large-scale study systematically (Cvetkovic et al., 2024). The combination of inductive and deductive coding strategies ensured that both emergent and theoretically grounded insights were captured (Saldaña, 2013).

Quantitative Methods

Within this dissertation, quantitative methods were used to complement qualitative approaches by providing breadth, while qualitative methods focused on capturing depth. By combining qualitative with quantitative methods, this dissertation adopts a comprehensive mixed-methods strategy to provide deeper insights into factors affecting user interaction with the AI artefacts, task performance, and the overall impact of the artefacts (Venkatesh et al., 2013). Quantitative methods are widely used in research to systematically measure and analyze the properties of study objects and the relationships among them. These methods rely on numerical data, which can be collected through various means and analyzed using descriptive and inferential statistical techniques to draw meaningful conclusions (Bryman, 2016; Creswell & Creswell, 2017). In the field of IS research, quantitative approaches play a critical role in evaluating artifacts both before implementation (ex-ante) and after deployment (ex-post), across both controlled and real-world settings (Prat et al., 2015; Pries-Heje et al., 2008; Venable et al., 2016).

This dissertation employs a range of quantitative methods to both address the RQs 1a and 1b to identify and empirically validate factors affecting the provision of input to AI (Cvetkovic & Bittner, 2022; Cvetkovic et al., 2025; Cvetkovic et al., 2022) as well as RQ 2b to assess the quality, utility, and effectiveness of developed artifacts (Cvetkovic, Rosenberg, & Bittner, 2023; Cvetkovic et al., 2024; Memmert et al., 2023, 2024). It relies on primary data, such as participant responses reflecting their perceptions of and experiences with developed artefacts (Cvetkovic, Rosenberg, & Bittner, 2023; Cvetkovic et al., 2024; Memmert et al., 2023, 2024) and preferences for handing over tasks or disclosing information to AI (Cvetkovic et al., 2025; Cvetkovic et al., 2022) that were collected through (randomized) controlled online experiments via standardized surveys. It is further complemented by secondary data, such as summary statistics of system usage metrics, gathered during evaluations of artifacts in semi-naturalistic settings (Cvetkovic, Rosenberg, & Bittner, 2023; Cvetkovic et al., 2024).

Beyond descriptive summaries, inferential statistical methods were applied to examine relationships between variables and test differences across conditions. For example, Analysis of Variance (ANOVA) (Cvetkovic et al., 2025; Memmert et al., 2024) and Analysis of Covariance (ANCOVA) (Cvetkovic & Bittner, 2022) were used to identify significant differences in the effects of various categories of independent variables on the dependent variable while accounting for covariates. In cases where data did not satisfy the normality prerequisite, alternative non-parametric tests were employed (Memmert et al.,

2023). These papers also use correlation analyses to examine further associations between key variables.

To capture complex relationships between multiple latent and observed variables, structural equation modelling (SEM) methods were applied, enabling the validation of causal pathways and interdependencies among variables (Westland, 2015). To ensure the quality of the data and operationalize the latent variables of interest, we first consulted the knowledge base for the most reliable measures corresponding variables. Upon collecting the data, Cvetkovic & Bittner (2022) employ Confirmatory Factor Analysis (CFA) (Bentler, 1990) to investigate whether the factor structure of a task delegability framework can be replicated in a knowledge work context. Cvetkovic et al. (2025) also employ CFA, to assess the unidimensional reliability of items for the self-developed scales. Additionally, authors utilize path analysis (Westland, 2015) to validate several hypotheses within a comprehensive research model. Hereby, both direct as well as mediating and moderating effects of independent variables are considered. Appropriate sample sizes were collected to support the robustness of the results (e.g., $n=582$ in the study discussed). For further robustness, results were supported with tests for model validity and control variable analyses, ensuring that the findings remained reliable and generalizable across different scenarios.

Finally, clustering (Rodriguez et al., 2019) was utilized in developing a taxonomy of user interactions with generative/large language models (GLMs/LLMs). This involved categorizing prompts using techniques such as k-means clustering (Arthur & Vassilvitskii, 2006), which grouped input prompts based on their characteristics, enabling the identification of archetypes of user behaviors (von Brackel-Schmidt et al., 2023).

4 Publications

This cumulative dissertation consists of 10 articles that are directly related to the addressed research topic. All articles were published in reputable and peer reviewed IS conferences and journals. One short thematic article did not undergo a typical double-blinded review procedure. Tables 5 and 6 list these articles chronologically according to their publication date and the type of outlet. Table 7 further outlines submissions related to the research topic but not included in the dissertation.

Table 5. Included journal publications

Journal Article
* Oeste-Reiß, S., Bittner, E., Cvetkovic, I., Günther, A., Leimeister, J.M., Memmert, L., Ott, A., Sick, B., & Walter, K. (2021) <i>Hybride Wissensarbeit</i> Informatik Spektrum, 44(3), 1-5 * no double-blinded peer review, short thematic article
Cvetkovic, I., Grashoff, I., Jovancevic, A., & Bittner, E. (2025) <i>Quid pro Quo: Information Disclosure for AI Benefits in Human-AI Collaboration</i> Computers in Human Behavior: Artificial Humans 4, p. 100137.

Table 6. Included conference publications

Conference Article
Cvetkovic, I. & Bittner, E. (2022). <i>Task Delegability to AI: Evaluation of a Framework in a Knowledge Work Context.</i> Proceedings of the Hawaii International Conference on System Sciences (HICSS). Virtual Conference.
Cvetkovic, I., Oeste-Reiß, S., Lehmann-Willenbrock, N., & Bittner, E. (2022). <i>What should AI know? Information Disclosure in Human-AI Collaboration</i> Proceedings of the European Conference on Information Systems (ECIS). Timisoara, Romania.
Cvetkovic, I., Rosenberg, V., & Bittner, E. (2023). <i>Conversational Agent as a Black Hat: Can Criticizing improve Idea Generation?</i> Proceedings of the Hawaii International Conference on System Sciences (HICSS). Kaanapali, HI, USA.
Memmert, L., Cvetkovic, I., & Bittner, E. (2023). <i>Human-AI Collaboration in Conceptualizing Design Science Research Studies: Can Generative Language Models Provide Helpful Suggestions?</i> Proceedings of the European Conference on Information Systems (ECIS). Kristiansand,

Norway.
Cvetkovic, I., Gierlich-Joas, M., Tavanapour, N., Debowski, N., & Bittner, E. (2023). <i>Augmented Facilitation: Designing a multi-modal Conversational Agent for Group Ideation</i> Proceedings of the International Conference on Information Systems (ICIS). Hyderabad, India.
von Brackel-Schmidt, C., Kučević, E., Memmert, L., Tavanapour, N., Cvetkovic, I., Bittner, E., & Böhmman, T. (2023). <i>A User-centric Taxonomy for Conversational Generative Language Models</i> Proceedings of the International Conference on Information Systems (ICIS). Hyderabad, India.
Memmert, L., Cvetkovic, I., & Bittner, E. (2024). <i>The more is not the merrier: Effects of Prompt Engineering on the Quality of Ideas produced by GPT-3</i> Proceedings of the Hawaii International Conference on System Sciences (HICSS). Honolulu, HI, USA.
Cvetkovic, I., Streek, S., & Bittner, E. (2024). <i>Designing a Hybrid Intelligence System for the Facilitation of Group Collaboration</i> Proceedings of the Australasian Conference on Information Systems (ACIS). Canberra, Australia.

Table 7. Related journal submissions/publications, not included in the dissertation

Jovancevic, A.*, Cvetkovic, I.*, Van Assche, J., Popovic, M., & Bittner, E. (under review in the 3rd round, revised and resubmitted after major revision) <i>Cognitive impairment and danger to humanity, or personal assistant and learning tool? Exploring attitudes toward ChatGPT and their predictors</i> Submitted to PLoS One (VHB ranking: B; discipline independent journal) * Shared first authorship
Memmert, L., Cvetkovic, I., & Bittner, E. (2025). <i>Brainstorming with a Generative Language Model: Effect of Exposure to AI Ideas on Brainstorming Performance and Cognitive Load</i> Business and Information Systems Engineering

The following sections in this dissertation are organized based on the contribution of the articles to the dissertation and correspond to the order presented in Table 4 in Section 3.2, which outlines the research strategy. Sections 9–18 summarize the published articles, providing details such as the title, authorship, article type, outlet (conference track), ranking, methodology, research aim, research contribution, and co-author contributions following CRediT⁶.

Section 9

Table 8. First article of the cumulative dissertation

Citation	Oeste-Reiß, S., Bittner, E., Cvetkovic, I., Günther, A., Leimeister, J.M., Memmert, L., Ott, A., Sick, B., & Walter, K. (2021). Hybride Wissensarbeit. <i>Informatik Spektrum</i> , 44(3), 1-5 https://doi.org/10.1007/s00287-021-01352-0
Ranking	VHB Publication Media Rating 2024: D CORE2018: /
Article type	<i>Journal</i> : Short thematic paper
Research aim	This short thematic paper aims to define hybrid knowledge work systems by discussing the integration of human and artificial intelligence and their complementary strengths in addressing both routine and non-routine tasks in knowledge work.
Methodology	Conceptual discussion supported by literature review and an illustrative example (collaborative writing in journalism).
Research contribution	This paper establishes foundational definitions and concepts for hybrid knowledge work systems. It introduces three archetypes of human-machine task collaboration (automation, verification, augmentation) and outlines design dimensions for hybrid work systems, emphasizing lifelong learning and socio-technical system design. The study proposes hybrid knowledge work systems as a means to reorganize work processes to better integrate human and machine intelligence. With this, the paper identifies opportunities for human-centered HIS.
Co-author's contribution	Oeste-Reiß: Conceptualization, Visualization (lead), Writing – original draft (lead), Writing – review & editing Bittner: Conceptualization, Supervision, Writing – review & editing Ott, Cvetkovic, Memmert, Walter, and Günther: Conceptualization (support), Writing – original draft (support), Writing – review & editing (support) Sick and Leimeister: Supervision, Writing – review & editing

⁶ <https://credit.niso.org/>

Section 10

Table 9. Second article of the cumulative dissertation

Citation	Cvetkovic, I. & Bittner, E. (2022). Task Delegability to AI: Evaluation of a Framework in a Knowledge Work Context. In <i>Proceedings of the 55th Hawaii International Conference on System Sciences (HICSS)</i> , Virtual conference
Ranking	VHB Publication Media Rating 2024: B CORE2018: A
Article type	<i>Conference</i> : Completed Research Paper
Track	AI and Future of Work
Research aim	In light of the increasing role of AI in augmenting human capabilities rather than replacing them, this paper aims to explore which tasks should be delegated to AI within knowledge work contexts. The goal is to evaluate an existing task delegability framework and adapt it to this specific domain, focusing on postgraduate researchers as exemplary knowledge workers. The study investigates the applicability of the framework, its predictive components for task delegation decisions, and the interplay of factors like trust, motivation, difficulty, and risk. This research seeks to bridge the gap in understanding HIS prerequisites and opportunities in knowledge work settings.
Methodology	Quantitative online survey with postgraduate researchers (n=93), Confirmatory factor analysis (CFA), Linear regression, Analysis of covariance (ANCOVA)
Research contribution	This paper contributes to the literature on HIS by validating and refining the task delegability framework in the context of knowledge work. It highlights trust as the most significant predictor of task delegability, underscoring the importance of designing AI systems that foster trust in AI among users. The findings also demonstrate that motivation and task difficulty negatively correlate with delegability, suggesting that demotivating or less cognitively challenging tasks are more likely to be assigned to AI. Furthermore, the study offers design implications for HIS, particularly in terms of tailoring AI support to specific tasks. This research sets the stage for future explorations of task delegation frameworks in diverse professional settings and contributes theoretical and practical insights for enhancing HIS.

Co-author's contribution	Cvetkovic: Conceptualization (lead), Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing (lead) Bittner: Conceptualization (support), Funding acquisition, Supervision, Writing – review & editing

Section 11

Table 10. Third article of the cumulative dissertation

Citation	Cvetkovic, I., Oeste-Reiß, S., Lehmann-Willenbrock, N., & Bittner, E. (2022). What should AI know? Information Disclosure in Human-AI Collaboration. In <i>Proceedings of the 30th European Conference on Information Systems (ECIS)</i> , Timișoara, Romania.
Ranking	VHB Publication Media Rating 2024: A CORE2018: A
Article type	<i>Conference</i> : Research in Progress paper
Track	Cognition and Human Behavior in Information Systems
Research aim	The aim of this paper is to explore human preferences for information disclosure in Human-AI collaboration, particularly in creative teamwork contexts such as Design Thinking. The study seeks to understand the types of information humans are willing to disclose to AI systems and how the design of AI outputs can influence these disclosure preferences. Using privacy calculus as theoretical lens, the study derives hypotheses regarding the interplay between the perceived risks and benefits of sharing information, which are tested in a subsequent study. This approach aims to provide a foundation for understanding how to guide the development of ethical HIS.
Methodology	Narrative literature review with a snowball sampling technique to identify input-output pairs; theoretical literature review to derive hypotheses
Research contribution	This paper's contribution lies in identifying opportunities for human-centered HIS, by specifically deriving human inputs and corresponding AI outputs relevant to creative teamwork tasks. By proposing these input-output pairs, the study provides a foundation for understanding the types of information disclosure that might foster effective collaboration in HIS. These contributions pave the way for empirical testing of hypotheses regarding the dynamics of privacy calculus in HIS.
Co-author's contribution	Cvetkovic: Conceptualization (lead), Methodology, Visualization, Writing – original draft (lead), Writing – review & editing (lead) Oeste-Reiß: Supervision, Writing – review & editing Lehmann-Willenbrock: Supervision, Writing – review & editing Bittner: Conceptualization, Funding acquisition, Supervision, Writing – review & editing

Section 12

Table 11. Fourth article of the cumulative dissertation

Citation	von Brackel-Schmidt, C., Kučević, E., Memmert, L., Tavanapour, N., Cvetkovic, I., Bittner, E., & Böhmman, T. (2023). A User-centric Taxonomy for Conversational Generative Language Models. In <i>Proceedings of the 43rd International Conference on Information Systems (ICIS)</i>, Hyderabad, India. Best Paper Award Nominee
Ranking	VHB Publication Media Rating 2024: A CORE2018: A*
Article type	<i>Conference: Completed Research Paper</i>
Track	General IS Topics
Research aim	This paper's objective is to develop a user-centric taxonomy for understanding and categorizing the usage of conversational generative language models (GLMs), such as ChatGPT, in solving complex problems. The taxonomy aims to provide a structured framework for analyzing multi-step user interactions and prompting behaviors, contributing to the design of more effective prompts and conversational AI systems. By identifying dimensions of usage, such as input styles, contextual dependencies, and functional outputs, the study aims to offer practical insights into the design of effective prompting strategies. Additionally, it seeks to inform researchers and practitioners about distinct usage patterns that may support the development of more user-friendly and context-aware AI systems.
Methodology	Prompt-a-thon, taxonomy development; qualitative content analysis, clustering, and archetype derivation to evaluate and validate the taxonomy
Research contribution	This paper introduces the first comprehensive taxonomy for classifying conversational GLM usage, offering a systematic framework of three meta-dimensions, seven dimensions, and 31 characteristics. It also identifies three archetypes of GLM usage: Creative Learning, Operational Delegation, and Scouting. These archetypes represent distinct patterns of user behavior and prompting strategies. The taxonomy and archetypes collectively establish a foundation for advancing research in conversational AI, prompt engineering, and HIS, while also providing actionable insights for practitioners seeking to optimize GLM integration into workflows.

Co-author's contribution	von Brackel-Schmidt: Conceptualization (lead), Data curation (lead), Formal analysis (lead: taxonomy, archetypes, demographic data), Investigation (lead), Methodology (lead), Project administration (lead), Visualization (lead), Writing – original draft (lead), Writing – review & editing (lead: paper content, responses, presentation) Kučević: Conceptualization (lead), Data curation (support), Formal analysis (lead: taxonomy; support: archetypes), Investigation (lead), Methodology (lead), Project administration (lead), Visualization (lead), Writing – original draft (lead), Writing – review & editing (lead: paper content) Memmert and Tavanapour: Conceptualization (support), Data curation (support), Formal analysis (support), Writing – original draft (support), Writing – review & editing (support) Cvetkovic: Conceptualization (support), Writing – review & editing (support) Bittner and Böhm: Supervision, Writing – review & editing

Section 13

Table 12. Fifth article of the cumulative dissertation

Citation	Cvetkovic, I., Rosenberg, V., & Bittner, E. (2023). Conversational Agent as a Black Hat: Can Criticizing improve Idea Generation? In <i>Proceedings of the 56th Hawaii International Conference on System Sciences (HICSS)</i> . Kaanapali, HI, USA.
Ranking	VHB Publication Media Rating 2024: B CORE2018: A
Article type	<i>Conference</i> : Completed Research Paper
Track	AI and Future of Work
Research aim	This study explores the role of a CA in facilitating idea generation, focusing on their ability to deliver constructive criticism. It investigates whether the perceived neutrality of a CA enhances the quality of ideas generated through its critiques. Specifically, the research aims to understand how the CA's feedback influences user perceptions and idea refinement processes, contributing insights into HIS for individual creative work.
Methodology	DSR, A/B test, survey, expert ratings, descriptive statistics
Research contribution	The paper presents a novel CA prototype capable of delivering targeted critiques to enhance idea generation. The study demonstrates that criticism by a CA can improve idea quality, particularly in terms of novelty and practicability. It provides practical contributions, including a set of DRs and DPs for developing CAs that can deliver criticism, and theoretical insights into how neutrality and criticism interplay in HIS. The findings offer preliminary evidence supporting the use of CAs as facilitators in creative processes, with potential applications in the moderation of methods like the Six Thinking Hats.
Co-author's contribution	Cvetkovic: Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing (lead) Rosenberg: Conceptualization, Data curation, Formal analysis, Investigation, Software, Visualization Bittner: Conceptualization, Funding acquisition, Supervision, Writing – review & editing

Section 14

Table 13. Sixth article of the cumulative dissertation

Citation	Cvetkovic, I., Gierlich-Joas, M., Tavanapour, N., Debowski, N., & Bittner, E. (2023). Augmented Facilitation: Designing a multi-modal Conversational Agent for Group Ideation. In <i>Proceedings of the 43rd International Conference on Information Systems (ICIS)</i> . Hyderabad, India. Best Paper Award Nominee
Ranking	VHB Publication Media Rating 2024: A CORE2018: A*
Article type	<i>Conference: Completed Research Paper</i>
Track	Digital Technologies and the Future of Work
Research aim	This paper aims to explore how CAs can be integrated into group ideation processes as co-facilitators. The study investigates the potential of CAs to support human facilitators by means of distilling and enriching unstructured information. It addresses the challenges of using AI to augment rather than replace human facilitation, focusing on designing effective socio-technical systems that enhance creative collaboration.
Methodology	DSR, literature review, exploratory expert interviews, confirmatory focus groups
Research contribution	The paper makes significant contributions by developing prescriptive knowledge: DRs, DPs, and design features (DFs), for implementing CAs in group ideation settings. It presents a prototype multi-modal CA integrated with Slack, demonstrating its ability to co-create value by enhancing the documentation process and facilitating group interaction. The study bridges gaps in literature on HIS, particularly in creative teamwork, and proposes a foundational framework for future research on augmented facilitation. Additionally, it highlights practical implications for organizations, offering transferable insights for integrating similar systems into collaborative workflows.
Co-author's contribution	Cvetkovic: Conceptualization (lead), Data curation, Formal analysis, Investigation (lead), Methodology (lead), Project administration, Software, Visualization, Writing – original draft (lead), Writing – review & editing (lead) Gierlich-Joas: Conceptualization (support), Methodology (support), Writing – original draft (support: theoretical background and

	discussion), Writing – review & editing (support) Tavanapour: Conceptualization (support), Investigation (support: data collection), Methodology (support), Writing – original draft (support: theoretical background), Writing – review & editing (support) Debowski: Project administration (support: access to industry experts) Bittner: Conceptualization, Funding acquisition, Supervision, Writing – review & editing

Section 15

Table 14. Seventh article of the cumulative dissertation

Citation	Cvetkovic, I., Streek, S., & Bittner, E. (2024). Designing a Hybrid Intelligence System for the Facilitation of Group Collaboration. In <i>Proceedings of the 35th Australasian Conference on Information Systems (ACIS)</i> , Canberra, Australia.
Ranking	VHB Publication Media Rating 2024: C CORE2018: /
Article type	<i>Conference</i> : Completed Research Paper
Track	Digital Innovation, Transformation, Business Models and Future of Work
Research aim	This paper addresses the challenges of facilitating effective group collaboration in remote work environments by introducing an HFS. The HFS leverages the complementary strengths of human and AI facilitators to enhance team performance, focusing on task allocation as the key to achieving efficient collaboration. By integrating AI's capabilities, such as speed and efficiency in data processing, with human skills like empathy and creativity, the system aims to redefine the facilitation process for remote group settings. The research explores how such hybrid systems can balance the cognitive and emotional demands of facilitation, ensuring harmony, adaptability, and effectiveness in collaborative processes.
Methodology	DSR, systematic literature review, mixed-method questionnaire (quantitative and qualitative questions), interviews, quantitative chat log analysis
Research contribution	The paper contributes a conceptual framework for task allocation in HFS, grounded in principles of HI. It offers prescriptive knowledge for the design of HIS, emphasizing the integration of human and AI capabilities to address the multifaceted challenges of group collaboration. Practical contributions include a validated prototype and a set of DPs that guide the implementation of hybrid facilitation. The findings highlight the potential of HFS to foster trust, enhance team efficiency, and optimize task distribution, laying the groundwork for a nascent design theory of Hybrid Facilitation. By addressing the limitations of solely human or AI-driven facilitation, this research advances the development of HIS in creative work contexts.

Co-author's contribution	Cvetkovic: Conceptualization, Data curation (support), Formal analysis (support), Investigation (support: oversight), Methodology, Project administration, Software (support), Supervision (lead), Visualization (lead), Writing – original draft, Writing – review & editing (lead) Streek: Conceptualization, Data curation (lead), Formal analysis (lead), Investigation (lead: prototype and data collection), Methodology, Project administration, Software (lead: prototype), Visualization (support) Bittner: Conceptualization, Funding acquisition, Supervision, Writing – review & editing

Section 16

Table 15. Eighth article of the cumulative dissertation

Citation	Memmert, L., Cvetkovic, I., & Bittner, E. (2023). Human-AI Collaboration in Conceptualizing Design Science Research Studies: Can Generative Language Models Provide Helpful Suggestions? In <i>Proceedings of the 31st European Conference on Information Systems (ECIS)</i> , Kristiansand, Norway.
Ranking	VHB Publication Media Rating 2024: A CORE2018: A
Article type	<i>Conference: Completed Research Paper</i>
Track	Design Research and Methods in Information Systems
Research aim	This paper aims to explore the potential of LLMs to support researchers in conceptualizing DSR studies by providing context-specific suggestions. The study investigates the perceived helpfulness of AI-generated suggestions in facilitating creative problem-solving and their suitability for complementing existing DSR tool support, particularly at the content level. It also examines the influence of expertise on the perceived utility of these suggestions, addressing a critical gap in the development of AI-assisted research tools.
Methodology	Prompt template development, survey (n=33), descriptive statistics, Wilcoxon Signed-Rank Test, Kruskal-Wallis Test, correlation analysis
Research contribution	The paper contributes to advancing HIS in the domain of DSR by demonstrating the feasibility of LLMs for generating helpful, study-specific suggestions. It highlights the complementary role of content-level AI support in DSR processes, as opposed to traditional meta-level tools. The findings reveal that LLM suggestions are generally perceived as helpful, with variation across dimensions and design components (issues, theories, DRs, and DPs). Additionally, the study provides actionable insights for integrating LLMs into a DSR support ecosystem and identifies areas for further refinement, such as enhancing unexpectedness in suggestions and tailoring support to user expertise levels.
Co-author's contribution	Memmert: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft (lead), Writing – review & editing (lead) Cvetkovic: Conceptualization, Data curation, Formal analysis,

	Investigation, Methodology, Writing – original draft (support), Writing – review & editing (support) Bittner: Conceptualization, Funding acquisition, Supervision, Writing – review & editing

Section 17

Table 16. Nineth article of the cumulative dissertation

Citation	Memmert, L., Cvetkovic, I., & Bittner, E. (2024). The more is not the merrier: Effects of Prompt Engineering on the Quality of Ideas produced by GPT-3. In <i>Proceedings of the 57th Hawaii International Conference on System Sciences (HICSS)</i> , Honolulu, HI, USA.
Ranking	VHB Publication Media Rating 2024: B CORE2018: A
Article type	<i>Conference: Completed Research Paper</i>
Track	Generative and Conversational AI in Information Systems Research and Education: Opportunities and Challenges
Research aim	This paper aims to explore the effects of prompt engineering techniques on the quality of ideas generated by GPT-3 in the context of brainstorming tasks. With the increasing use of LLMs in creative applications, understanding how prompt design influences the outputs is critical. The study focuses on six distinct prompt engineering techniques and their combinations, investigating their impact on two key dimensions of idea quality: novelty and value. By addressing how specific prompts shape creative outputs, the research seeks to provide actionable insights for designing effective prompts that optimize HIS in brainstorming processes. Furthermore, the study aims to contribute to the broader field of creativity support systems by identifying best practices for integrating LLMs into workflows that require divergent thinking and innovative problem-solving.
Methodology	Prompt template development, survey, expert idea rating, one-way ANOVA, correlation analysis
Research contribution	The paper contributes to the emerging field of prompt engineering by providing empirical evidence on the effects of single and combined prompt techniques on idea novelty and value. The findings suggest that Evaluation Criteria Specification (ESC) significantly improves novelty but combining multiple techniques does not always enhance output quality. This challenges the assumption that more complex prompts lead to better results, offering practical guidance for designing prompts in creative applications. The research highlights the need for targeted, context-specific prompt design and provides a foundation for further studies on integrating LLMs into creativity support systems.

Co-author's contribution	Memmert: Conceptualization, Data curation, Formal analysis (support), Investigation, Methodology, Project administration, Software, Visualization, Writing – original draft (lead), Writing – review & editing (lead) Cvetkovic: Conceptualization, Data curation, Formal analysis (lead), Investigation, Methodology, Writing – original draft (support), Writing – review & editing (support) Bittner: Conceptualization, Funding acquisition, Supervision, Writing – review & editing

Section 18

Table 17. Tenth article of the cumulative dissertation

Citation	Cvetkovic, I., Grashoff, I., Jovancevic, A., & Bittner E. (2025). Quid pro Quo: Information Disclosure for AI Benefits in Human-AI Collaboration. <i>Computers in Human Behavior: Artificial Humans</i> 4, p. 100137. DOI: 10.1016/j.chbah.2025.100137.
Ranking	VHB Publication Media Rating 2024: No existing ranking yet due to the novelty of the journal; sister journal of Computers in Human Behavior, ranked B
Article type	<i>Journal</i> : Completed Research Paper
Research aim	This paper investigates human willingness to disclose personal information to AI systems during virtual meetings, specifically in scenarios involving AI-generated feedback. Grounded in privacy calculus theory, the study explores how contextual factors, such as Trust in AI, Perceived Risks, and Perceived Benefits, shape users' disclosure intentions. It systematically examines how these intentions differ across various types of information (e.g., team sentiment, attention levels) and assesses whether the anonymity of the feedback and the specific recipient (individual vs. team) influence willingness to disclose information. By using carefully constructed vignettes and robust statistical methods, the research aims to uncover determinants guiding information disclosure behavior in AI-supported teamwork.
Methodology	Hypothesis development, vignette and scale development, survey, ANOVA, SEM (path analysis), mediation analysis
Research contribution	The paper extends privacy calculus theory by emphasizing Trust in AI as a central determinant influencing users' willingness to disclose personal information. It further demonstrates that across diverse information types, users' disclosure intentions are consistently shaped by Perceived Benefits and Perceived Risks, underscoring their robust predictive power regardless of contextual variations. Moreover, the study identifies and empirically validates several contextual factors: feedback anonymity, type of disclosed information, and recipient specificity. These insights offer clear practical implications for effectively integrating AI into workplace teamwork settings, highlighting the importance of fostering user trust and carefully addressing sensitivities tied to different information types.
Co-author's contribution	Cvetkovic: Conceptualization (lead: hypotheses & study design), Data curation (lead), Formal analysis (lead: quantitative analyses &

	SEM), Investigation (lead: data collection), Methodology (lead), Project administration (lead), Visualization, Writing – original draft, Writing – review & editing (lead) Grashoff: Conceptualization (support), Investigation (support: vignette development & evaluation), Writing – original draft, Writing – review & editing (support) Jovancevic: Formal analysis (support: SEM), Writing – original draft (support) Bittner: Conceptualization, Funding acquisition, Supervision, Writing – review & editing

5 Theoretical Contributions

This chapter outlines the theoretical contributions developed throughout the research activities of this cumulative dissertation. It addresses the research questions posed in Chapter 1, namely RQs 1a, 1b, 2a, and 2b, while also expanding on the broader knowledge contributions related to RQ 3. First, the findings on human input provision and its utilization by AI are explored, focusing on prerequisites and opportunities for human-centered HIS. Second, design knowledge and insights derived from the studies are presented to support the creation of human-centered HIS. Finally, the chapter proposes an overarching, high-level contribution toward a nascent design theory and situates the overall insights within existing design knowledge taxonomies, emphasizing their relevance for theory.

5.1 Provision of human input and its utilization by AI

This section responds to RQs 1a and 1b by outlining the prerequisites (RQ 1a) and opportunities (RQ 1b) tied to the provision of human input to AI, including both the environmental and behavioral conditions under which users are inclined to share data or delegate tasks. The section also explores how AI can leverage such input, once provided, to create value and improve the design of HIS. The objective here is to paint a holistic picture of the interplay of users' needs, preferences, and concerns in the context of their input provision and how this can inform the design of human-centered HIS. To close the gap in our understanding of how human input is collected, managed, and employed (cf. IPO model, Chapter 2.2), the findings from several key articles (Cvetkovic & Bittner, 2022; Cvetkovic et al., 2025; Cvetkovic et al., 2022; Oeste-Reiß et al., 2021; von Brackel-Schmidt et al., 2023) are integrated.

5.1.1 Prerequisites for human-centered input provision

As a foundation for RQ 1a, which focuses on *what humans want* (i.e., user preferences in human-centered HIS), this subsection clarifies the conditions that must be met for humans to provide meaningful input. Prerequisites for human-centered provision of input involve recognizing what tasks are suitable for AI delegation, evaluating users' willingness to delegate tasks to AI, and understanding users' willingness to share information with AI. In essence, human preferences dictate whether tasks are handed off to AI, how much data is disclosed, and in what form it is shared - all of which must be aligned with maintaining human autonomy, privacy, and motivation.

To the date of inquiry, only one developed framework of task delegability had existed (Lubars & Tan, 2017), however, it was created for work in general, potentially involving AI support, and had not yet been closely examined in the more specialized context of knowledge work. Cvetkovic and Bittner (2022) address this theoretical *abstraction gap*

(Avdiji & Winter, 2019) by studying postgraduate research as an “exemplary class of knowledge work” including both creative and complex problem-solving tasks that meet the necessary criteria for this scope. Noting that “it is hard to make inferences about the applicability of the framework to more specific contexts” (ibid.), authors pose three research questions probing how well the existing delegability model applies to postgraduate tasks, which components most strongly contribute to the delegation decision, and whether certain tasks in this domain are more likely to be delegated.

By empirically refining the framework, Cvetkovic and Bittner (2022) show that the risk factor, originally assumed to influence delegation decisions, does not significantly matter in postgraduate research. Instead, trust, difficulty, and motivation emerge as key predictors, revealing that tasks perceived as tedious or unmotivating are more likely to be handed over to AI, whereas intrinsically motivating or complex tasks should remain under human control (Figure 5). This aligns with Jin and Uchida (2024) three-year analysis, which revealed a growing willingness to delegate undesirable tasks to AI. However, the authors found a heightened awareness of risks associated with delegating tasks to AI, particularly at the second data collection point in 2023. This shift was linked to the emergence of LLMs, underscoring the need for transparent communication about AI capabilities and limitations to foster trust and ensure effective task delegation in evolving work contexts.

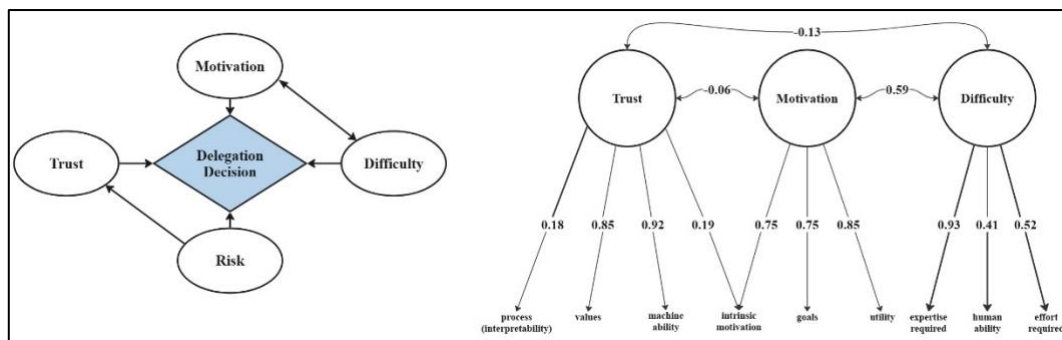


Figure 5. Original framework (left) vs new framework suitable for postgraduate research (right). Adapted from Cvetkovic & Bittner (2022).

Similarly, Bouwer (2022) emphasized the role of emotional states, suggesting that individuals experiencing low-arousal, negative-valence emotions are more inclined to delegate tasks to AI. Both studies underscore the need for domain-specific evaluations of task delegability frameworks and highlight how human factors, such as perceived task characteristics, emotional states, and evolving AI capabilities influence delegation decisions. Collectively, these insights suggest that task delegation frameworks should prioritize functional alignment with user preferences while considering emotional and contextual nuances, particularly as AI systems like LLMs evolve.

Cvetkovic and Bittner (2022) also discuss specific examples of tasks suited for delegation to AI. For instance, researchers are willing to delegate data-intensive tasks to AI, such as maintaining records of participant data and systematic literature reviews, which allows them to focus on higher-level tasks such as writing publications and supervising

undergraduate students. These examples showcase how HIS could foster efficiency and support human workers by alleviating cognitive load. They further provide insights into what kinds of tasks should be delegated to AI to respect human autonomy and preferences, distinguishing between tasks that are routine and tedious versus those that are non-routine and challenging.

Table 18. Tasks that postgraduate researchers (do not) want to delegate to AI. Adapted from Cvetkovic and Bittner (2022).

Delegability	Task
↑ ↑ ↑ ↑ ↑ ↑ ↑ ↑ ↑ ↑	Code and verify data in accordance with specified research protocol
	Maintain accurate records of participant data, safeguarding the confidentiality of subjects, as necessary
	Assist in laboratory analysis, quality control, or data management and prepare findings for publication
	Conduct literature reviews
	Request or acquire equipment or supplies necessary for the project
	Review and edit data to ensure completeness and accuracy of information; follow up with subjects to resolve problems or clarify
	Develop assessment and evaluation tools
	Identify and compile lists of potential research subjects in accordance with study objectives and parameters, as appropriate to
	Set up, calibrate and maintain laboratory and/or field research equipment, as specified by the requirements of the study
	Prepare materials for submission to granting agencies and foundations
↓ ↓ ↓ ↓ ↓	Derive research questions
	Prepare data collection
	Write and contribute to publications
	Supervise undergraduate students

As this thesis progressed and AI capabilities rapidly evolved, especially in the field of LLMs, we found that these factors did not stay robust in predicting the delegability decision. Despite not evaluating this explicitly, the indications were there – within two months of its release, ChatGPT had a hundred million users⁷. In our works on LLMs (Memmert et al., 2023, 2024; von Brackel-Schmidt et al., 2023) we could show that parts of the tasks that are non-routine, difficult and in many cases motivating, such as creative ideation, are readily delegated or shared with AI, thus highlighting the fluidity of the framework and underscoring the necessity for ongoing adjustments as both AI and humans evolve (Richter & Schwabe, 2025).

While the concept of delegability focuses on tasks, human input in many other contexts takes the form of information disclosure, which can be especially sensitive when it concerns social signals (cf. Ethics, Chapter 2.1.3). Past research has shown that users are often

⁷ <https://explodingtopics.com/blog/chatgpt-users>

reluctant to share psychologically invasive data (Bailenson et al., 2006). At the same time, AI could potentially leverage these signals to enhance recommendations and strengthen group awareness, but the complexities surrounding these disclosures and the precise trade-offs between privacy risks and tangible benefits remain insufficiently explored. These theoretical *causal* and *manipulability gaps* (Avdiji & Winter, 2019) call for a deeper examination of the conditions under which individuals are willing to disclose such information, how different intervention designs influence these decisions, and how privacy considerations affect willingness to share data in a human–AI context. By applying and extending privacy calculus insights originally grounded in human–human communication, this approach offers a nuanced understanding of disclosure preferences and behaviors specifically tailored to AI-driven scenarios. Cvetkovic et al. (2022) and Cvetkovic et al. (2025) address this need by investigating how an AI might offer “service bundles” in exchange for user information, thereby illuminating the trade-offs between disclosure risks and tangible benefits.

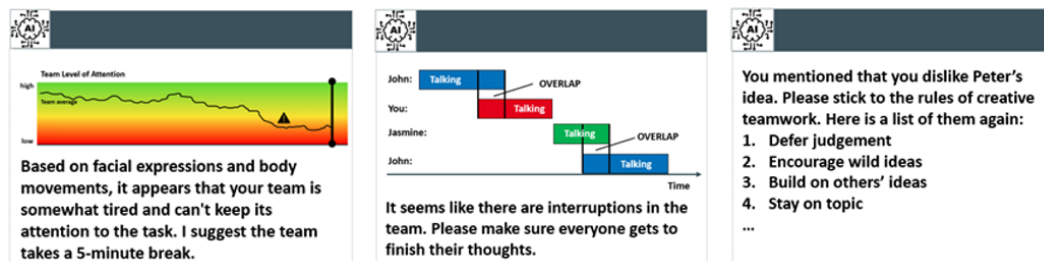


Figure 6. Proposed AI outputs for different human inputs: speech shares, cognitive load, and breaking rules of creative teamwork (Cvetkovic et al., 2022, 2025)

Figure 6 depicts the examples of proposed AI interventions we empirically evaluated in Cvetkovic et al. (2025). In this study, we manipulated seven different information types as well as the recipient and the anonymity of AI output. We further extended the privacy calculus model (Culnan & Armstrong, 1999) to include trust in AI as a predictor of intention to disclose information. With this study, it became clear that whether teams or individuals are the recipient of AI output and whether the output is anonymized or not did not influence people’s intention to disclose certain information. More relevant to this intention were the types of information requested, perceived risks and benefits associated with the information in question, as well as trust in AI. With this, we add to the established body of knowledge on privacy calculus theory by extending and validating it in a Human-AI collaboration context. This perspective is supported by recent studies emphasizing the importance of trust and transparency in AI systems (Vössing et al., 2022).

In all three papers, Trust in AI emerged as a key factor for human input provision. Figure 7, for example, depicts a strong positive correlation ($r = 0.64$, $p < 0.001$) between trust and delegability decision. This strengthens the existing literature dealing with trust in autonomous systems (Shahrdar et al., 2018) which has repeatedly demonstrated the importance of trust for effective Human-AI collaboration (Choung et al., 2023). Trust in

AI is a robust predictor of this collaboration, showing high influence across different contexts (Cvetkovic & Bittner, 2022; Cvetkovic et al., 2025).

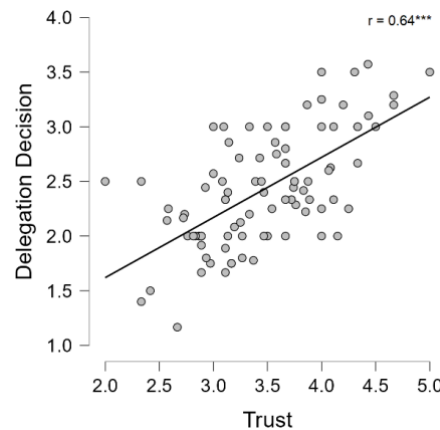


Figure 7. Relationship between trust in AI and Task Delegability to AI. Adapted from Cvetkovic & Bittner (2022)

Interestingly, perceived risk tied to delegating tasks to or sharing information with AI does not demonstrate the same consistency. While it proves highly relevant to the intention to disclose (sensitive) information in teamwork, it does not affect the decision to delegate postgraduate research tasks to AI. This underscores the importance of both *input type* and *context* in which it is provided. Understanding these nuances is essential for designing HIS that integrate human preferences and ensure that AI systems operate in ways that are respectful, trustworthy, and beneficial to their users.

Taken together, these prerequisites: task delegability factors, information disclosure factors and trust in AI, shape how AI can be integrated into human-centered information systems. Autonomy and privacy must be safeguarded to ensure a constructive, high-quality exchange of information, while motivation and trust remain vital to fostering an environment where AI is seen as beneficial rather than intrusive or risky.

5.1.2 Opportunities for effective human input utilization by AI

Where the preceding subsection emphasizes what and how humans provide input, the following discussion turns to why that input can be so valuable for driving AI-based innovations. Within the realm of human-centered HIS, an important RQ involves examining the opportunities that arise when people provide input to AI, particularly how that input can be harnessed in practical, human-centered ways (cf. IPO model, Chapter 2.2).

Oeste-Reiß et al. (2021) discuss how both routine and non-routine tasks in knowledge work can be supported by hybrid systems, prompting new system designs that combine human and machine strengths. Routine tasks may be automated or significantly streamlined, reducing cognitive load for knowledge workers. Non-routine tasks, such as creative ideation and contextual problem-solving, are characterized by their reliance on implicit knowledge, 'know-how' that is often difficult to formalize or codify. AI can augment these

tasks by offering relevant suggestions, structuring exploratory discussions, and providing contextual insights, thereby enabling humans to connect experiential knowledge with the broader information environment. To this end, the authors introduce the concept of Hybrid Knowledge Work Systems (HKWS) which leverage two complementary loops: AI in the loop of human intelligence (where the system provides automated or augmented support to human users) and human intelligence in the loop of AI (where humans verify or refine the AI's outputs). An illustrative example is collaborative journalism, where AI automates aspects of data gathering while journalists focus on investigative storytelling. Through this synergy, humans remain central to the creative and decision-making processes, while AI learns from user feedback and continuously improves. Overall, this approach emphasizes designing socio-technical systems where knowledge workers and AI collaboratively learn, ensuring that AI complements, rather than displaces, human expertise.

By proposing a framework embedding HKWS into a taxonomy of task types (routine and non-routine) and interaction archetypes (automation, verification, augmentation), this article could help future research address the design *projectability gap* by providing a structured approach to tailoring abstract prescriptive knowledge to specific contexts, thereby enabling more effective integration of HIS into diverse work environments.

Building on this foundation, Cvetkovic and Bittner (2022) examine which tasks knowledge workers actually prefer to delegate to AI. Their empirical findings highlight a shared thread across those tasks: they are generally routine, tedious, data-intensive, time-consuming, and otherwise unmotivating and not challenging (see Table 18). By confirming that users gravitate toward offloading such work, this study reveals a clear opportunity for AI to assume responsibilities that free workers to concentrate on higher-level thinking. This not only broadens acceptance of AI but also underscores the importance of designing systems that match user preferences, empowering humans without compromising their sense of control.

Shifting the focus from offloading individual tasks to enriching team-level interactions with AI, we explore how human input in the form of information from creative work in teams can power AI-driven interventions that offer direct, visible benefits to users. There has been limited understanding of the specific “service bundles” an AI-based system can provide to incentivize users to share these inputs. Drawing on use cases like MIT's social robot TEGA, which detects children's emotions to enhance educational outcomes (Westlund et al., 2016), Cvetkovic et al. (2022) illustrate how aligning “input-output service pairs” could foster creativity and facilitate collaborative processes. Their work categorizes a broad set of potential human inputs from team interaction data (e.g., body language, emotional states) and team process data (e.g., calendar information, transcripts) and possible AI outputs (e.g., moodboards, reminders, “rules of engagement” notifications). For example, if an AI system could interpret gestures, facial expressions, or negative utterances in real time, it might respond by offering a joke to lighten the mood or a reminder to keep a constructive team

mindset. Similarly, AI might detect stagnation during brainstorming and proactively provide stimulus topics or motivational quotes to spark new ideas.

Table 19. Examples of human input identified in the literature and proposed AI output for each respective input

	Human Input	AI Output
Team interaction	Body language (gestures, facial expressions)	Team mood board (own suggestion); jokes and anecdotes in response to negative mood
	Body language (eye gaze, mimics, gestures)	Reminder to take a break, information load board
	Keywords for breaking rules	Reminder of the rules of creative teamwork
	Speech share & centrality	Informing about unbalanced speech shares and a reminder of preferred equality
Team process	Results from a brainstorming session	Structured and clustered results presented in a dashboard
	Calendar information	Recommended schedule for next session(s)
	Data from interviews/desk research	Analysis report (e.g., Recommendation for core categories from interviews)
	Stagnation (e.g., not enough ideas within 10 minutes of brainstorming)	Proactive acting - stimulus topic or motivational quote

However, subsequent related studies (Cvetkovic, Gierlich-Joas, et al., 2023; Cvetkovic et al., 2025; Cvetkovic et al., 2024) indicate that while these propositions are conceptually appealing, many users are hesitant to share sensitive interaction-level information, such as detailed interpersonal behaviors or emotional cues, with an AI system. In contrast, disclosing process or task-level data (e.g., scheduling details, transcripts of prior sessions, or storing the outcomes of brainstorming) to receive AI support in return is positively perceived. This pattern suggests that while AI-generated interventions can indeed support creativity and teamwork, designers must carefully distinguish between the kinds of data humans are willing to share and the types of assistance they value most. As a result, delegating process-oriented responsibilities (e.g., summarizing prior sessions, organizing tasks) appears to be more readily embraced than more intrusive, interaction-level monitoring.

Finally, von Brackel-Schmidt et al. (2023) address a new theoretical *prescription gap* (Avdiji & Winter, 2019) that arises with GLMs like ChatGPT, whose chat-based mode now enables extended, multi-step dialogues. Previous research had not established how users adopt prompting tactics for solving complex problems within such conversational interfaces. By focusing on how humans guide GLMs to generate specific content, they present a taxonomy of prompting behaviors derived from a large-scale participant event, noting that users refine prompts iteratively and explore various facets of a problem to guide

the AI's output. Most users used GLMs to build or sharpen their understanding (61%) and to delegate tasks (35%), reflecting their dual role as tools for exploration and assistance. Regarding the intended functional output, the study notes a significant portion of inputs designed for creative purposes (56%), showcasing the role of GLMs in supporting ideation and innovative problem-solving tasks.

These findings underscore that GLMs are increasingly perceived as collaborators capable of assisting with understanding complex topics or performing specific delegated tasks. The taxonomy provides a framework for structuring interactions, such as breaking down complex problems into step-by-step prompts, which aligns with opportunities for human-centered HIS. Moreover, the study emphasizes that users can utilize GLMs not only for task outsourcing but also as dynamic team members that support continuous learning, reflecting key aspects of HI, such as iterative refinement and mutual learning (Dellermann, Ebel, et al., 2019). By formalizing these interaction patterns, von Brackel-Schmidt et al. (2023) offer a valuable foundation for designing systems and training resources to maximize the utility of human input in HIS.

Meta-dimension	Dimension	Characteristics								
Prompt objectives	Focus	Informational				Creative				
	Purpose	Build understanding		Sharpen understanding		Task delegation		General conversation		
	Intended functional output	Create	Aggregate	Supplement	Substitute	Assess	Compare	Combine	Reformulate	
Prompt settings	System perspective	Humanized		User-oriented	Group-oriented	Organization-oriented		Expert-oriented	None	
	Context	Explicit context		Reference context		Reference previous + explicit context		None		
Prompt style	Output format specification	Not specified			Structured			Domain-specific		
	Tone	Polite		Impolite		Motivational		Neutral		

Figure 8. A User-centric Taxonomy of GLM Usage (von Brackel-Schmidt et al., 2023)

The findings from von Brackel-Schmidt et al. (2023) collectively indicate that, once user needs and contextual conditions are satisfied, human input can serve as a powerful catalyst for AI-driven enhancements in both specialized tasks and creative collaboration. The following sections will build on the outlined prerequisites and opportunities that arise from effective human input provision and its utilization, illustrating how the design and evaluation of human-centered AI can be operationalized in HIS for creative work and complex problem-solving.

5.2 Human-centered AI in creative work and complex problem-solving

This section answers RQs 2a and 2b by focusing on the design cycle of the overall DSR project. It is divided into two sub-sections: design and development (5.2.1) and evaluation (5.2.2), each addressing corresponding research questions. The knowledge gaps identified in previous research are addressed through the provision of design knowledge (e.g.,

taxonomies, frameworks, DPs) and entities (e.g., software prototypes), thereby contributing to human-centered HIS for creative work and complex problem-solving.

5.2.1 Design and development of human-centered AI for creative work and complex problem-solving

Driven by the insights from 5.1, especially regarding the prerequisites and opportunities for task delegation identified in Cvetkovic and Bittner (2022) and for AI-supported creative work in teams in Cvetkovic et al. (2022), as well as by the advancements in AI development, particularly in the field of LLMs, several articles were motivated to design and develop or utilize existing AI systems to support creative work and complex problem-solving.

One approach to designing AI for creative work, as discussed in Cvetkovic, Rosenberg, and Bittner (2023), centers on leveraging CAs to foster ideation through structured critique. Previous research (Siemon, 2022) shows that evaluation apprehension, which normally occurs in collaboration with other people, does not take place in collaboration with AI. Tackling the *theoretical prescription gap* (Avdiji & Winter, 2019), Cvetkovic, Rosenberg, and Bittner (2023) explore the thus far uninvestigated impact of criticism by a CA in the context of an idea-generation process. To achieve this, they develop a CA prototype designed to act as a facilitator during the ideation phase of Design Thinking, specifically adopting the "Black Hat" role from De Bono's Six Thinking Hats framework (De Bono, 1985).

The CA prototype aimed to improve idea quality by providing constructive criticism that users perceived as plausible and actionable. To do so, the development of the CA followed the DPs derived from the literature shown in Figure 9.

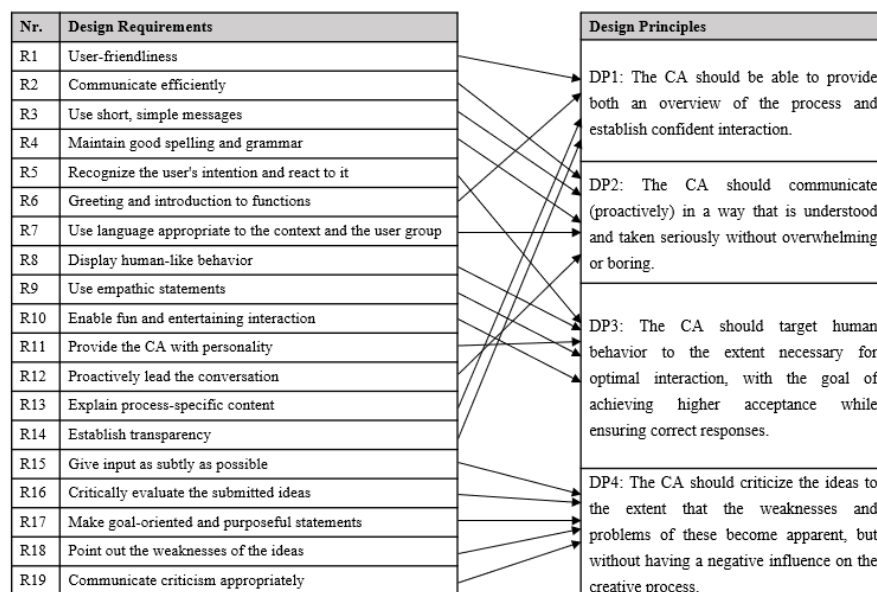


Figure 9. Design knowledge for a CA for individual ideation

The potential of CAs to enhance organizational ideation processes by supporting both human facilitators and creative teams simultaneously is even less addressed. The knowledge from HI (Dellermann, Ebel, et al., 2019) and Facilitation literature (Briggs et al., 2013; Leimeister, 2014) supports the premise that AI can engage in co-creating value by supporting human facilitators and creative teams and leaving them with cognitive space for the actual value-creating tasks. How this is to be designed, also considering a human-centered task allocation, remains largely untapped in the current literature. Cvetkovic, Gierlich-Joas, et al. (2023) tackle this *design validity gap* (Avdiji & Winter, 2019). Building on the insights from the previous study, they go one step further and advance the role of AI as a facilitator by developing a CA that supports several tasks in creative teamwork.

To produce design knowledge in the form of DRs, DPs, and DFs, user needs are elicited through expert interviews and literature reviews. The results, as appearing in Figure 10, illustrate the necessity for task and process assistance (DP1), executed via natural language communication (DP2), while mimicking human-like behavior (DP4). They particularly elicit the need for information enrichment and distilling (DP3), identified in both literature and practice. To ensure it is fulfilled, this DP was implemented via custom real-time data processing scripts.

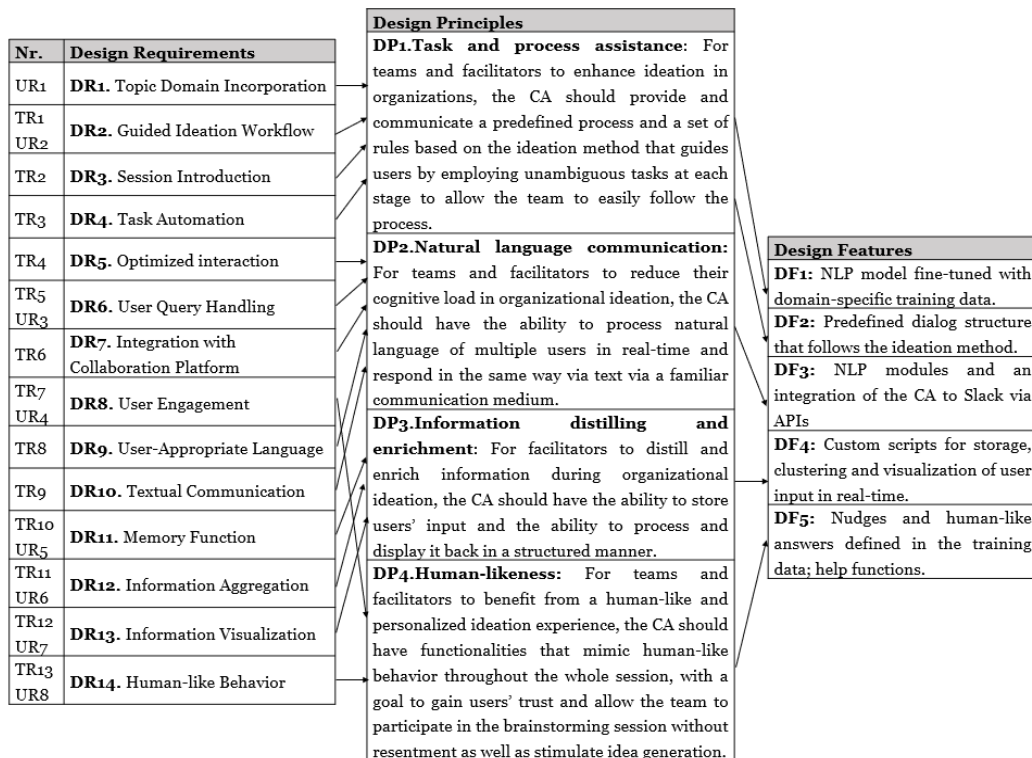


Figure 10. Design knowledge for a multi-modal CA for the facilitation of creative teamwork

An entity, a CA, was then designed to overtake task-related interventions, such as automating the clustering and visualization of ideas, summarizing contributions, and managing session processes. For example, the CA supported the ideation process by

generating mindmaps and wordclouds that participants used to identify relationships between ideas and focus on key topics. Produced design knowledge addresses two potential user groups: human facilitators and creative team members.

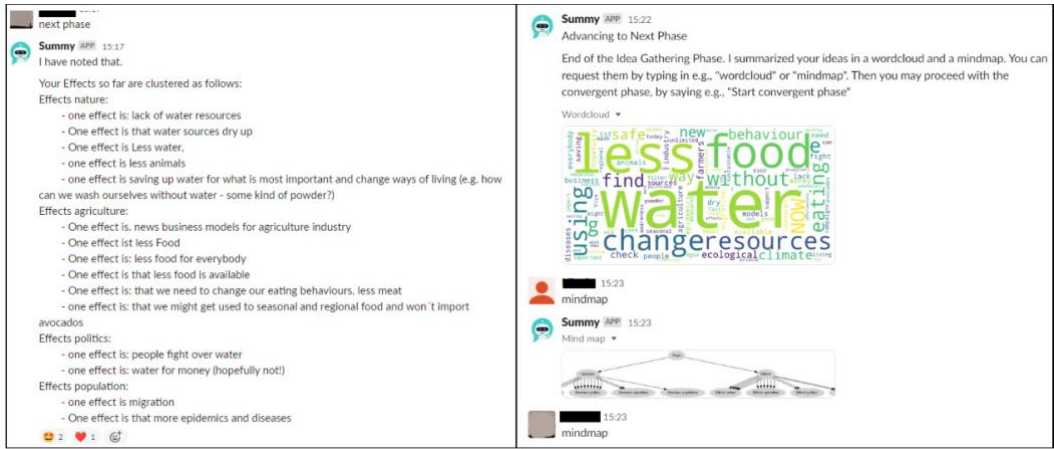


Figure 11. Instantiation of a multi-modal CA for the facilitation of creative work in teams

As an effort to also address a knowledge gap on HIS in the facilitation of creative work, representing an evolution from a CA for individual ideation support through a CA for team-level brainstorming support to a truly hybrid system, Cvetkovic et al. (2024) combine human and machine intelligence to create an HFS. Addressing this *theoretical abstraction gap* (Avdiji & Winter, 2019) is approached by translating high-level, abstract knowledge from HI literature and facilitation frameworks into prescriptive, context-specific design knowledge and entities. In a three-cycle DSR approach (Peffer et al., 2007), authors utilize knowledge produced in each cycle to inform the next one. Several knowledge streams inform the development of prescriptive knowledge in the form of design knowledge and entities. For the former, a task allocation framework is produced (Figure 12), relying on knowledge from the HI literature (Dellermann, Ebel, et al., 2019) and the facilitation framework (Beranek et al., 1993; Clawson et al., 1993; Leimeister, 2014), which splits facilitation interventions into task-related (T), and interaction-related (I).

 Human Intelligence	 Artificial Intelligence	 Both
Listen, explain and integrate information (T)	Introduce the procedure, the roles of participants, goals and rules (T, I)	Explain the use of technology (T)
Show flexibility (T, I)	Guide through the agenda (T)	Foster individual and group responsibility (T)
Foster and support different perspectives (I)	Summarize the meeting (T)	Focus on group goals (T)
Actively build relationships (I)	List open points (T)	Create and foster an open, positive and participative environment (I)
Constructively manage conflicts and negative emotions (I)	Clarify next steps (T)	Identify decisions (T)
Ask for opinions (I)	Make information understandable to the group (T)	
	Introduce the participants (I)	

Figure 12. A framework for task allocation in an HFS (Cvetkovic et al., 2024)

Furthermore, DPs for an HFS are derived from several knowledge sources: implications from a previous experiment, literature reviews, and the task allocation framework. These address both the human facilitation (HF) part as well as the AI facilitation part (CF = Chatbot Facilitation) of the HFS and are illustrated in the following Table.

Table 20. DPs for Hybrid Facilitation Systems (Cvetkovic et al., 2024)

DP1. Hybrid intelligence: The collaboration between the HF and CF is conducted by both undertaking different tasks that build upon each other. Parallel responsibilities should be avoided. The responsibilities of individual tasks are verbally clarified to the working group. Facilitation tasks should be divided between the facilitation agents according to their strengths. Facilitation tasks that have not received clear responsibility must be broken down into their subtasks and assigned to either the CF or HF. They should not be managed in parallel by both facilitation agents, as this can cause confusion among the participants.
DP2. Task Takeover: Due to the limited capabilities of the CF, the HF should step in when it reaches its limits. This is particularly the case when the CF cannot answer certain questions or technical issues arise.
DP3. Friendly facilitation team: HF and CF interact respectfully and make it clear to the working group that they form a facilitation team together. They present themselves as a unit. The facilitation team interacts respectfully and is friendly with the participants and encourages intensive collaboration. The facilitation agents highly depend on each other's results. Especially the handover of one facilitation agent to the other must be done carefully to ensure a smooth facilitation process. Both must treat each other politely and give the impression that they are a well-coordinated team. This is important so that the working group accepts both and does not try to work with only one of the facilitation agents. To build 8 Hybrid Facilitation trust in the working group and create a positive working atmosphere, the communication between all should be friendly and respectful.
DP4. Task-related facilitation: The CF guides the working group through the agenda, clarifying the necessary work steps or tasks and initiating the votes. It stores the contributions and final results that have arisen in the group discussion. It is responsible for time management. Derived from the requirements for the CF, it is primarily responsible for task-related facilitation activities. Thus, it primarily leads the group work but hands over the lead to the HF depending on the task. The primary leadership and corresponding responsibility should also make clear that the working group cannot bypass it and use only the HF.
DP5. Interaction-related facilitation: The HF supports the exchange of opinions and decision making process during the collaboration so that the working group constantly focuses on their group goal, incorporates diverse perspectives, and thus can work on a high-quality group result. HF also supports the establishment of relationships by

promoting the exchange of opinions, ensuring an open participatory environment, and trying to resolve conflicts between group members. As few predefined statements as possible are used to be able to react flexibly to the group conversation. Following the requirements for the HF, HF is primarily responsible for interaction-related activities. These should be as little predefined as possible to ensure the best possible support during the discussions and situational engagement with individual participants.

DP6. Truthful appearances: The avatar representations of the CF and the HF support their appearance as a robot and as a human, respectively. This is beneficial for a realistic impression of the hybrid facilitation. The participants should not get the impression that they are being deceived in the presentation of the facilitation team.

The DPs materialized through an instantiation of an HFS (Figure 13). In terms of prescriptive knowledge contribution, this low-maturity artifact (Gregor & Hevner, 2013) is a contribution of a type entity (Drechsler & Hevner, 2018).

	Facilobot APP 13.04 The task will take 35 minutes from now on. It is now 13:04:26. The task expires at 13:39:26. When you are ready, tell me, that you want to start !		C 13.20 @Facilobot Give us an example
	HF 13.04 Feel free to interact with the chatbot		Facilobot APP 13.20 An Example in the BEHAVIOR section for a road-damage-reporting app would be: Citizens and visitors ride less bikes because of the potholes. They try to avoid the potholes, and thereby impede the incoming traffic. They sometimes have accidents because they don't see the potholes.
	C 13.04 @Facilobot lets get started		H M 13.21 @Facilobot Save email-ping-pong (since calendars are not mutually viewable)
	Facilobot APP 13.05 Hello participant_in_161, J M , C and H M , my name is Facilobot ! 😊 I am a chatbot and together with the HF I will support you in carrying out the task! 🤖🤖 I will guide you through the task, document your work results, guide the voting and track the time. If you have any questions, please tag me @Facilobot ! @HF would you like to introduce yourself briefly and say what your task is in the groupwork?		Facilobot APP 13.21 I have saved your argument: Email-ping-pong (since calendars are not mutually viewable) under the number 3. The number is used for deletion if required. You have reached the minimum number of 4 arguments. Let me know when if you want to start voting for the best 3 arguments.
	HF 13.05 Words marked in bold are example commands that you can use Hi I'm HF ! I worked as a technology manager. With the help of e.g., Design thinking workshops, we search for and find user-centered solutions. Today I am here to support you with the task, if you have any questions, feel free to write it! @Facilobot continue		HF 13.21 Can you guys think of a few more things?
	Facilobot APP 13.05 Thank you, if you have any questions in between, don't hesitate to ask me! To find out who are you working with here, tell others something about you (name, job). Otherwise, its so anonymous here...		Participant_in_161 13.21 @Facilobot save At the same time, in addition to student appointments, other appointments for committees come in and there is a scheduling chaos.
	C 13.05 @Facilobot what is your name?		Facilobot APP 13.21 I have saved your argument: At the same time, in addition to student appointments, other appointments for committees come in and there is a scheduling chaos under the number 4. The number is used for deletion if required. You have reached the minimum number of 4 arguments. Let me know when if you want to start voting for the best 3 arguments.
	Facilobot APP 13.04 I am Facilobot ! 🤖		

Figure 13. Instantiation of the HFS (Cvetkovic et al., 2024)

Overall, these three papers address the use case of creative work (both individual and teamwork) and present different solutions, thereby contributing design knowledge and entities to the knowledge base.

The remaining two articles address the second class of artefacts utilized in this thesis and leverage existing LLM solutions by evaluating them ex-post. With the rise of generative AI, there has been a recognition of its vast potential to support humans across diverse use cases. One such area is complex problem-solving, represented through solution design in DSR. This encompasses several critical steps, from identifying kernel theories to deriving issues and requirements, ultimately producing design knowledge in the form of principles and artefacts. Addressing the *design validity gap* (Avdiji & Winter, 2019), to explore the potential of LLMs in supporting DSR researchers, Memmert et al. (2023) utilized the, at

the time of manuscript preparation, state-of-the-art GPT-3 model. The authors integrated GPT-3 into a brainstorming canvas where it suggested kernel theories, issues, DPs, and DRs in response to inputs for each category. The study highlighted how generative AI could act as a content-level collaborator, inspiring researchers by providing context-specific suggestions that facilitated exploration within the problem and solution space. It emphasized the transformative role of LLMs for complex problem-solving. By leveraging AI's generative capabilities, researchers can overcome traditional barriers to idea generation, such as cognitive biases and resource constraints, thereby broadening the scope of exploration and enhancing the quality of design artefacts. These advancements contribute not only to the methodology of DSR but also to the broader discourse on human-AI collaboration in tackling complex, open-ended problems.

Building on this, Memmert et al. (2024) investigated the impact of different prompt engineering techniques on idea generation using GPT-3, addressing the *theoretical prescription gap* of how means affects the ends (Avdiji & Winter, 2019). They developed six distinct prompt templates designed to enhance output quality, such as including context, specifying evaluation criteria, and incorporating demonstrations (question-independent and question-dependent examples). For instance, prompts could guide the AI to provide ideas based on expert perspectives or tailor suggestions to specific contexts. Examples of prompts included: *Contextual Prompt*: "Provide three innovative ways to address [specific societal problem]."; *Expert Perspective Prompt*: "As a sustainability expert, suggest three solutions to reduce food waste."

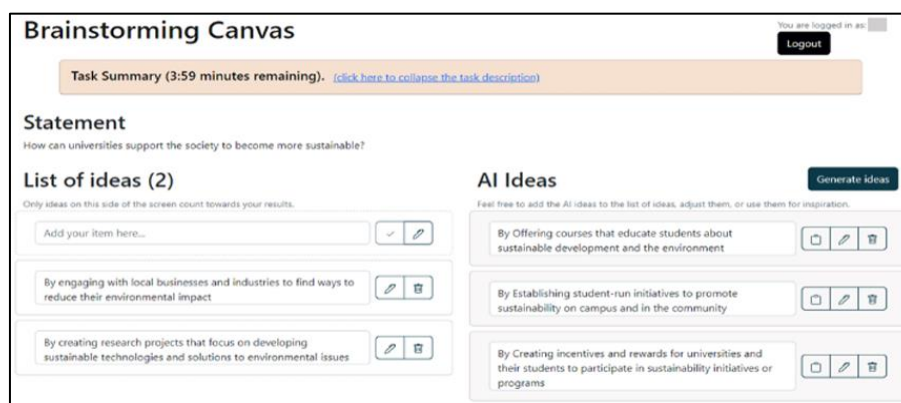


Figure 14. Snippet of a GLM-based Brainstorming app prototype (Memmert et al., 2024)

Figure 14 illustrates the interface of a GLM-based brainstorming app prototype, showcasing how users can input their ideas while interacting with AI-generated suggestions. The frontend enables users to provide question-dependent inputs, such as ideas for promoting sustainability in universities, while the backend supports this interaction by integrating pre-designed prompt templates and managing the generation of AI-based ideas. This collaborative process demonstrates how LLMs can function as co-creators in brainstorming activities, augmenting human creativity by providing inspiration and supporting the refinement of ideas.

These two distinct use cases exemplify the adaptability of LLMs in human-centered HIS. By addressing both problem-solving and open-ended creative brainstorming, the studies highlight how LLMs can be utilized to augment human capabilities across different contexts. The DSR-focused study emphasizes the role of LLMs in scaffolding structured research processes, while the prompt engineering study demonstrates how LLMs can catalyze idea generation and broaden the range of creative outputs.

5.2.2 Impact of human-centered AI integration on creative work and complex problem-solving

To answer RQ 2b and expand the theoretical contribution on how AI systems shape, and are shaped by the collaboration with humans, we investigate both objective changes in processes and outputs of creative work (e.g., ideation quality and team facilitation by tracking revised ideas or rating outputs for novelty and practicability), while also focusing on user perceptions and experiences (e.g., subjective ratings of AI-generated suggestions). As a result, the research collectively highlights both observable shifts in workflows and participants' subjective assessments of AI's value. This dual perspective is critical to understanding how users integrate AI into their working contexts, what benefits they perceive, and where challenges or limitations emerge.

Cvetkovic, Rosenberg, and Bittner (2023) investigated the impact of a CA providing structured criticism during the ideation phase of Design Thinking. By assuming the Black Hat role from De Bono's Six Thinking Hats framework, the CA guided participants to iteratively refine their ideas. Evaluation results showed that participants who received CA-provided critiques significantly improved their ideas, with 40% of the ideas being revised. These revised ideas were rated higher in terms of novelty and practicability compared to those that were not criticized. The evaluation demonstrates the superior quality of criticized ideas, validating the CA's ability to foster iterative improvements while reducing evaluation apprehension commonly observed in human-mediated settings (Siemon, 2022). These findings align with research by Rayan et al. (2024), who explored the potential of generative AI-based conversational cues in real-time collaborative ideation and showed that certain facilitator strategies, such as prompting deeper ideation or introducing new perspectives, effectively guided discussions and enhanced creative output.

Extending the application of CAs to team settings, Cvetkovic, Gierlich-Joas, et al. (2023) developed a multi-modal CA to augment group ideation by automating routine facilitation tasks, such as clustering, visualizing, and summarizing ideas. The evaluation highlighted the CA's effectiveness in supporting team processes, with participants praising its capability for real-time synthesis of complex inputs in helping them converge on core concepts and catalyze more targeted brainstorming. Clustering capabilities were particularly valued, as they helped structure discussions and guide teams toward identifying key themes. Visualization features, including mindmaps and wordclouds, were similarly appreciated for fostering a better understanding of relationships among ideas and enabling

more focused brainstorming. These capabilities likely acted as cognitive off-loading visual scaffolds, supporting the external cognition theory (Kirsh, 2010; Scaife & Rogers, 1996). While the CA significantly enhanced facilitation efficiency, some areas for improvement were identified. A few participants noted that the rigid structure of the CA's interactions occasionally limited its flexibility in responding to diverse team dynamics. Additionally, the pace of some automated processes was perceived as overwhelming in certain situations, suggesting a need for better alignment with the natural flow of group discussions. Despite these challenges, the CA demonstrated strong potential to complement human facilitators by offloading routine tasks and creating space for more value-driven contributions, making it a promising tool for augmenting collaborative ideation.

In both Cvetkovic, Rosenberg, and Bittner (2023) and Cvetkovic, Gierlich-Joas, et al. (2023), the perceived human-likeness - or lack thereof - of the CA does not affect the goal of the interaction. In Cvetkovic, Gierlich-Joas, et al. (2023) the CA was not perceived as human-like, as it was characterized by rapid, structured, and repetitive responses. Participants viewed it as a background note-taker and process assistant. While many studies suggest that human-likeness can enhance perceived social presence (Morana, Gnewuch, et al., 2020) and trustworthiness (Zhou & Hu, 2024) and even mitigate the negative impact of mistakes (Sá Siqueira et al., 2024) this does not appear to be a crucial DP in these creative work contexts. Instead, people prioritize functionality and expect chatbots to relieve cognitive load and effectively support their tasks. This finding aligns with research by Seeger et al. (2017), who propose that while anthropomorphism can influence user perception and behavior, its effects are context-dependent. In creative work in teams, the functional role of the CA seems to outweigh the need for human-like characteristics.

Further advancing the facilitation paradigm, Cvetkovic et al. (2024) introduce an HFS that combines a human and an AI facilitator. The HFS effectively divided tasks between the human facilitator (HF) and chatbot facilitator (CF) based on their respective strengths, with the CF managing task-related interventions and the HF focusing on interaction-driven tasks. Participants consistently noted the structured and efficient nature of the facilitation process, attributing this to the CF's capabilities in guiding the agenda, managing time, and documenting group outputs. The HF's involvement was appreciated for fostering trust, group harmony, and interaction quality. The collaboration between the HF and CF was described as cohesive, with participants recognizing the team-like dynamic between the two facilitation agents. These positive findings were further supported by the log file analysis, which showed that task responsibilities were effectively divided, with each facilitator performing tasks aligned with their designed roles. The HF reported a significantly reduced cognitive load compared to traditional facilitation, suggesting that the HFS design effectively alleviated the burden of task management for human facilitators. However, a notable technical issue underscored the importance of ensuring robust technical reliability in future iterations of the system. Additionally, some participants expressed concerns that the CF's presence occasionally hindered human interaction within the group,

suggesting a need for better integration of the two facilitators to enhance team dynamics further. Overall, these findings illustrate the HFS's capacity to redefine the facilitation process, highlighting how AI can serve as a true collaborator that expands a team's creative and organizational capabilities. At the same time, it underscores the importance of addressing technical and design limitations to optimize the system's performance and ensure a seamless, human-centered experience. Collectively, the three studies that involved designing CAs for creative work have given substance to the "Facilitator is a bot" idea (Bittner et al., 2019).

Shifting to the second class of artefacts investigated in the thesis, Memmert et al. (2023) explore the application of LLMs for supporting the conceptualization of DSR studies by generating content-level suggestions. The study evaluates these AI-generated suggestions across five dimensions: perceived helpfulness for inspiration, ease of understanding, relevance to the design component (DC), relevance to the ideation facilitation (IF), and unexpectedness.

Table 21. Participant ratings for AI suggestions per design component. Adapted from Memmert et al. (2023)

AI Suggestions	Helpful for inspiration	Easy to understand	Relevant to DC	Relevant to IF	Unexpected
Issues	3.3 (1.0)	4.0 (0.9)	3.8 (0.8)	3.5 (0.8)	2.5 (1.0)
Theories	3.4 (0.9)	3.4 (0.8)	3.5 (0.9)	3.2 (0.8)	3.2 (1.0)
DRs	3.8 (0.8)	4.1 (0.9)	3.7 (1.0)	3.7 (0.9)	2.4 (0.9)
DPs	3.9 (0.8)	3.7 (0.9)	3.9 (0.8)	3.7 (0.7)	2.7 (1.0)

As depicted in Table 21, the findings reveal that most suggestions were rated positively, with helpfulness and relevance scoring significantly above neutral, particularly for design components like requirements and principles. However, ratings for theories were notably lower, suggesting limitations in the ability of LLMs to generate relevant ideas. The study also highlights nuanced variations based on user expertise. Contrary to initial assumptions, participants with higher expertise often found the suggestions more helpful, indicating that familiarity with the topic might enhance the integration of AI-generated ideas into existing knowledge. This contrasts with expectations that novices would benefit more, potentially due to their greater need for structured guidance (Gau et al., 2022).

Despite the positive reception, the suggestions scored poorly in terms of unexpectedness, aligning with concerns about the generative models' tendency to replicate common patterns from their training data. To counteract this, Zhu and Luo (2022) suggest enhancing the model by providing unexpected examples or adjusting the model parameters like temperature (level of randomness).

Finally, Memmert et al. (2024) examined the effects of various prompt engineering techniques on the quality of ideas generated by GPT-3 in a brainstorming application. Six techniques: Evaluation Criteria Specification (ESC), Context (CON), Instruction/Schema

(INS), Demonstrations (question-dependent and question-independent), and Expert Perspective were evaluated individually and in combination. As shown in Figure 15, ESC significantly improved the novelty of ideas, while CON had a negative impact on novelty compared to no technique. Mixed results were observed for combinations of techniques, with over-engineering potentially diminishing effectiveness. Evaluation metrics included novelty and value, assessed by independent raters using predefined criteria, ensuring a robust evaluation of the generated ideas. The results underline the importance of carefully selecting prompt techniques tailored to desired outcomes. For example, ESC alone often performed better than when combined with other techniques, suggesting that simplicity can sometimes outperform complexity in prompt design. This study highlights the potential of prompt engineering to optimize the relevance and novelty of AI-generated ideas, emphasizing the need for future research to refine and expand the use of such techniques in creative work settings.

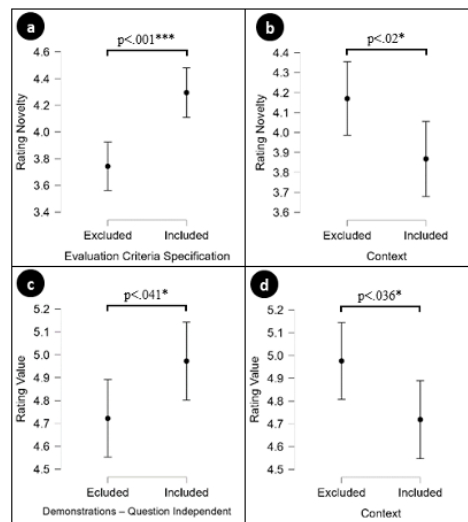


Figure 15. Prompting techniques performing better than baseline (left) and worse than baseline (right) (Memmert et al., 2024)

These findings resonate with prior research, such as Summers-Stay et al. (2023), which suggested that adjustments to prompts could enhance idea quality by leveraging a GLM's latent capacity for analogical reasoning. The study's focus on individual and combined prompt effectiveness adds to the literature by demonstrating that more detailed prompt designs do not always lead to better results, challenging assumptions in earlier work by Wu et al. (2022), who advocated for chaining techniques to enhance creativity. Additionally, the implications underscore that users benefit most from systems embedding prompts directly into intuitive user interfaces, which abstract complexities and allow seamless integration of AI-generated outputs into creative workflows. This insight aligns with findings from Bućinca et al. (2021) regarding AI-assisted decision-making, which emphasized balancing cognitive support with system usability.

These studies collectively demonstrate the transformative impact of AI on creative work and problem-solving. As AI systems become increasingly adept at offering real-time

facilitation, content generation, and structured critique, they stand to significantly enrich creative work and problem-solving outcomes, provided that researchers and practitioners address the technical, ethical, and design considerations necessary for a seamless, human-centered experience.

5.3 Overall Theoretical Contribution

This thesis advances the overarching theoretical contribution of human-centered AI development by offering novel DPs that guide the creation and integration of AI systems into creative work and complex problem-solving. These DPs derive from a systematic investigation of how AI-based solutions can be tailored to diverse contexts, and how they can complement rather than replace human expertise. Collectively, these principles manifest the dissertation's overarching contribution by providing a coherent design scaffold for human-centered HIS.

Regarding design knowledge contributions, Cvetkovic and Bittner (2022) exemplify *exaptation* (Gregor & Hevner, 2013) by adapting an existing generic task delegability framework to a new problem setting, thus demonstrating how established solutions can be reapplied and validated in a less mature application domain. In contrast, the subsequent studies by Cvetkovic, Rosenberg, and Bittner (2023); Cvetkovic, Gierlich-Joas, et al. (2023); Cvetkovic et al. (2024) and Memmert et al. (2023) and Memmert et al. (2024) produce type *improvement* contributions. Here, new prescriptive knowledge emerges in the form of evaluated instantiations (abstraction level 1) and generalized DPs (abstraction level 2) that refine or extend existing AI-based solutions to address known challenges in creative work and problem-solving. These contributions include DPs for and instantiations of CAs for supporting individual or team ideation, HFS that orchestrates human-AI facilitation, and prompt engineering guidelines that enhance the quality and relevance of AI-generated outputs. By providing concrete artifact instantiations as well as generalized principles for known problems, these studies illustrate how established technologies can be further improved and integrated into broader socio-technical contexts.

In addition to these theoretical contributions, this thesis adopts an explicitly interdisciplinary stance by positioning its various studies along a socio-technical continuum. Drawing on the conceptualization by Sarker et al. (2019), Figure 16 illustrates how each project emphasizes different disciplinary vantage points, ranging from social disciplines such as psychology, sociology, and management to more technical fields like computer science. On the left side of the continuum are studies that focus predominantly on social or organizational aspects, whereas those on the right side address more technical or engineering-focused concerns. In between are projects that balance social and technical considerations, highlighting how the design and evaluation of AI systems can benefit from a holistic view that integrates both human and technological factors. By visualizing the spread of publications across this spectrum, Figure 16 underscores the broad disciplinary

grounding of the thesis and illustrates how bridging different domains yields more robust, context-sensitive solutions for human-centered HIS.

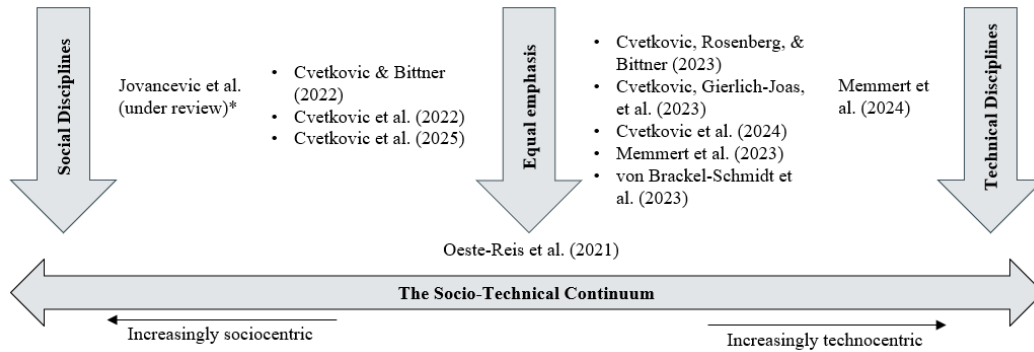


Figure 16. Contributions along the Socio-Technical Continuum.

*Related submission, not included in the dissertation

Altogether, this work contributes to the Ω -knowledge base (empirical insights into real-world phenomena) and the λ -knowledge base (actionable, prescriptive design knowledge). It adapts and extends established frameworks by situating them in specific application scenarios (e.g., postgraduate research, creative teamwork) and proposes new theoretical constructs that clarify the interplay between human cognition, collaboration, and AI-driven interventions. As such, this thesis lays the groundwork for a nascent design theory for design and action (type 5) (Baskerville et al., 2018; Gregor, 2006; Gregor & Hevner, 2013), specifically geared toward enabling human-centered HIS for creative work and complex problem-solving, by providing evaluable and generalizable principles for future research and practice. By leveraging problem-structuring and artifact-design heuristics, heuristic theorizing (Gregory & Muntermann, 2014) facilitated the discovery of DPs that address real-world challenges while remaining adaptable to diverse contexts. Heuristic theorizing emphasizes iterative learning and reflection, ensuring that the DPs are not only theoretically sound but also practically applicable. In the following, the resulting overarching DPs will be presented; these principles further unify and operationalize the insights gained, thus helping to establish robust HIS that address known challenges in creative work and complex problem-solving contexts. The DPs are constructed following Gregor et al. (2020) by describing key actors, aim, context, mechanisms and rationale.

Table 22. Overarching DPs for human-centered HIS

Design Principle	Description
DP1. Task Appropriateness	For system architects and AI developers (<i>Implementers</i>) to design HIS that allocate tasks based on the evolving strengths of humans and AI (<i>Aim</i>), for knowledge workers and facilitators (<i>Users</i>) in individual and teamwork (<i>Context</i>), ensure repetitive and structured tasks are handled by AI systems to reduce cognitive load, while

	reserving motivating and context-sensitive tasks for humans. For example, AI focuses on data-intensive tasks and process-oriented facilitation, such as clustering ideas and managing timelines, while human facilitators focus on interaction-driven tasks, such as fostering group cohesion and resolving conflicts. As AI capabilities evolve and user preferences shift, progressively enable AI to support non-routine and complex tasks in areas where users express willingness to delegate (<i>Mechanism</i>). Team leads, moderators, and organizational stakeholders actively engage with the system to implement and adapt to these mechanisms, ensuring alignment with evolving user preferences and leveraging advanced AI capabilities for real-time task delegation (<i>Enactors</i>). This principle is supported by studies demonstrating that aligning task allocation to the complementary strengths of humans and AI improves task outcomes, reduces cognitive load (Cvetkovic & Bittner, 2022; Cvetkovic, Gierlich-Joas et al., 2023; Cvetkovic et al., 2024), and reflects the growing trend of delegating even complex, non-routine tasks to LLMs as user trust and AI sophistication increase (Memmert et al., 2023, 2024) (<i>Rationale</i>).
DP2. Privacy, Trust, and Transparency	For data privacy officers, AI developers, and UX designers (<i>Implementers</i>) to create systems that foster user trust by ensuring transparency and privacy (<i>Aim</i>), for end-users interacting with AI systems (<i>Users</i>) in collaborative environments involving sensitive data collection and sharing, such as virtual teamwork (<i>Context</i>), provide clear and accessible communication about data handling processes. Allow users to configure their privacy preferences before data collection begins and implement transparent workflows explaining how data will be stored, analyzed, and used (<i>Mechanism</i>). IT managers, compliance officers, and organizational stakeholders ensure adherence to privacy guidelines and ethical standards (<i>Enactors</i>). This principle is supported by privacy calculus theories, emphasizing that trust in AI mitigates risks and increases user willingness to share information, as proposed and validated by Cvetkovic et al. (2022, 2025) and related research (<i>Rationale</i>).

DP3. Human-AI Co-Creation	For AI researchers and interface designers (<i>Implementers</i>) to develop systems that support and enrich human creativity while clarifying AI's role (<i>Aim</i>), for teams and individuals collaborating with AI (<i>Users</i>) in creative work and problem-solving scenarios (<i>Context</i>), integrate features that allow AI to propose alternative perspectives, ask critical questions, and stimulate ideation. Embed iterative feedback loops, enabling users to refine outcomes with AI assistance. Clearly define AI's roles to prevent overlap and confusion (<i>Mechanism</i>). Team facilitators, project managers, and innovation managers guide and coordinate collaboration between humans and AI to ensure alignment with shared objectives (<i>Enactors</i>). By promoting co-creation with AI, this principle enables richer exploration of ideas and encourages innovation through iterative collaboration, as evidenced by Cvetkovic, Rosenberg, & Bittner (2023), Memmert et al. (2024) and Cvetkovic et al. (2024) (<i>Rationale</i>).
DP4. Structured AI Output	For data analytics experts, UX designers, and AI developers (<i>Implementers</i>) to design systems that produce structured, goal-directed, and actionable outputs to support facilitation (<i>Aim</i>), for individuals and groups engaged in collaborative or creative processes where inputs are often unstructured and require organization and interpretation (<i>Users</i>), in contexts involving brainstorming, ideation, or problem-solving tasks requiring structured outputs to refine and enhance contributions (<i>Context</i>), design systems to generate clear summaries, visualizations, and facilitation cues (e.g., clustering ideas, providing structured critiques) that help users process and refine contributions effectively, supporting deeper exploration and alignment with shared or individual objectives (<i>Mechanism</i>). Facilitators, participants, and individuals using AI tools for personal ideation or group collaboration interact with these outputs to enhance clarity and productivity (<i>Enactors</i>). Structured and goal-directed outputs improve facilitation by reducing cognitive overload and encouraging in-depth exploration of ideas, as demonstrated by Cvetkovic, Gierlich-Joas et al. (2023) and Cvetkovic, Rosenberg, & Bittner (2023), and

	validated in scenarios involving both individual and group ideation (<i>Rationale</i>).

In an attempt to solidify this prescriptive knowledge, it is mapped to the overarching theoretical framework chosen for this thesis. Human-centered AI emphasizes the integration of technology, user, and ethics dimensions to guide the design of AI systems. Each DP addresses one or more of these dimensions while contributing to key human-centered AI design goals such as trustworthy AI, usable AI, and human-controllable AI (W. Xu & Gao, 2024).

The principle of Task Appropriateness (DP1) emphasizes the allocation of tasks according to the complementary strengths of humans and AI, leveraging the HI paradigm. By assigning routine, structured tasks to AI and reserving creative, context-sensitive tasks for humans, this DP aligns with the technology dimension of the human-centered AI framework. It supports the goals of useful AI and human-controllable AI, ensuring that AI systems enhance human productivity without diminishing autonomy.

Privacy, Trust, and Transparency (DP2) addresses both the user and ethics dimensions by fostering trust through transparent AI operations and ethical data practices. This principle ensures that AI systems respect user preferences in data collection, particularly for sensitive or personally identifiable information, and clearly communicate how such data is used. In doing so, it aligns with the design goal of trustworthy AI, a key aspect of the human-centered AI framework’s emphasis on user privacy.

The principle of Human-AI Co-Creation (DP3) spans the technology and user dimensions, positioning AI as a collaborator in creative work and problem-solving. By enabling AI to offer critical feedback, alternative perspectives, and context-relevant suggestions, this principle aligns with the human-centered AI’s focus on augmenting humans. It ensures that AI’s involvement enhances human efforts without introducing confusion or role ambiguity, a critical requirement for effective HIS.

Structured AI Output (DP4) is closely tied to the technology dimension, ensuring that AI systems deliver outputs that are actionable, purposeful, and easy to interpret. This principle supports the human-centered AI goals of usable AI and useful AI by emphasizing the need for intuitive and structured outputs that complement human decision-making and reduce cognitive load.

Together, these principles operationalize the conceptual elements of the human-centered AI, bridging the gap between theoretical insights and practical design knowledge. They ensure that AI systems are not only useful and usable but also ethical, transparent, and user-centered, advancing the broader goals of HIS.

6 Practical Contributions

Guided by the DSR approach (Hevner & Chatterjee; Hevner et al., 2004; Venable, 2006), this dissertation focuses on bridging theory and practice by addressing tangible problems in contemporary work environments. In addition to advancing theoretical insights about human-centered AI, the research contributes practical solutions, such as a task delegation framework and AI-driven facilitators, that practitioners can deploy in real-world settings. In line with established criteria for evaluating practical impact (Pries-Heje et al., 2008; Sonnenberg & vom Brocke, 2012), each design artifact was tested in collaboration with relevant stakeholders, ensuring that its application is both feasible and effective. Thus, the dissertation supports organizations in refining creative work and problem-solving processes through fostering more efficient collaboration between human workers and AI systems.

6.1 Integration of AI in creative work and complex problem-solving

A central practical contribution lies in identifying which tasks are best delegated to AI and which should remain under human control. This notion is rooted in the refined task delegability framework derived from Cvetkovic and Bittner (2022) empirical evaluation in the context of postgraduate research, which showed that factors such as user motivation, perceived task difficulty, and trust in AI inform delegation decisions. Organizations can adopt this framework to systematically assess their internal workflows and determine where automation is both viable and acceptable to employees. By categorizing tasks along dimensions such as data-intensity, repetitiveness, and creativity, managers can assign appropriate responsibilities, ensuring that knowledge workers gain cognitive relief from routine or tedious duties while retaining ownership of high-level, intrinsically motivating activities. To this end, the task allocation framework provided in Cvetkovic et al. (2024) can guide task orchestration in human-machine ensembles, as it incorporates tasks commonly encountered in day-to-day collaborative work, making it versatile and applicable across diverse domains and thus emphasizing high projectability of this design knowledge (vom Brocke et al., 2020). Such a structured allocation optimizes resource management and paves the way for effective HIS in day-to-day operations.

Building upon this delegation approach, organizations might institute clear guidelines or decision points for determining whether a particular process step, for instance, in data analysis or information gathering, is entrusted to an AI system. By crafting transparent policies that address privacy implications and communicate the rationale for such decisions, practitioners can build greater acceptance of these new methods. Rather than eroding job satisfaction, delegation can thus become an empowering choice that frees up employees to engage in more creative, interpersonal, or strategic work. Through this

systematic integration of AI, companies can strike a balance between efficiency gains and the preservation of human autonomy.

6.2 Design knowledge for human-centered HIS

6.2.1 Facilitating creative work

Beyond simply deciding which tasks to automate, this dissertation offers a suite of DPs and instantiated artifacts that organizations can adapt to their particular needs. Several prototypes demonstrate the potential of CAs and HFS to support individual workers as well as teams. For instance, practitioners can borrow from the design knowledge generated in studies such as Cvetkovic, Gierlich-Joas, et al. (2023) or Cvetkovic et al. (2024) to build AI-enabled interventions that summarize group discussions, monitor progress on shared goals, and guide participants through structured ideation techniques. The HFS showcased by Cvetkovic et al. (2024) presents a practical blueprint for orchestrating human-led and AI-led tasks. This indicates that creative workshops or innovation sprints can be reimaged as hybrid processes that simultaneously leverage the consistency of AI with the relational sensitivity of human facilitators. From a managerial standpoint, such hybrid systems offer tangible gains in both efficiency and quality of outputs, while also contributing to positive team experiences by preserving strong human interaction.

This dissertation also emphasizes the value of human–AI co-creation. The design knowledge produced in prototypes such as the Black Hat CA for individual ideation underscores that AI can act as a constructive critic, prompting users to refine and question their assumptions without fear of social repercussions. In a practical sense, organizations can embed a similar logic in brainstorming tools or collaborative whiteboard software, inviting AI to challenge an idea’s feasibility or to propose alternative use cases. These capabilities not only enrich the innovation process but also counteract groupthink and social evaluation apprehension, ultimately raising the quality of outcomes. For organizations interested in real-time digital facilitation, the instantiations featured in this dissertation, ranging from straightforward text-based bots to multi-modal chat interfaces, can serve as a reference point for how CA-based facilitators might be designed and employed in similar settings. In this way, organizations can pair human facilitators with AI-driven assistants for workshops or innovation sprints, potentially boosting efficiency, improving the structure of group activities, and reducing the cognitive load on human facilitators.

6.2.2 Customizing LLMs for problem-solving

This dissertation addresses prompt engineering strategies that allow organizations to leverage LLMs for ideation and complex problem-solving tasks. The experiments discussed in Memmert et al. (2023) and Memmert et al. (2024) show how even small changes to the structure or content of a prompt can yield noticeable improvements in the novelty and usefulness of AI-generated ideas. This is particularly relevant for teams

working on innovations, product ideation, or DSR researchers. By distilling the dissertation's findings into best practices, such as specifying evaluation criteria, including relevant context, or providing well-chosen examples, practitioners can systematically elicit higher-quality suggestions from generative models.

Because these findings emphasize that over-engineering prompts can reduce creative output, organizations aiming to deploy LLM-based systems can opt for a more balanced approach. Rather than inundating the model with complex instructions, a simpler prompt that maintains certain essential elements can produce more varied and engaging outputs. By incorporating iterative cycles of refinement, moderated by human oversight, practitioners can harness LLMs to accelerate ideation, expand solution spaces, and spark original thinking. In doing so, they cultivate a collaborative dynamic between human expertise and AI-driven creativity, which represents one of the dissertation's guiding themes.

Practical implications also include the need for careful integration of LLM outputs, treating them as hypotheses requiring further validation. Challenges such as potential biases, lack of novelty, and ethical considerations like plagiarism also necessitate rigorous quality control and user education. Memmert et al. (2023) underscore the potential of AI to complement human creativity in DSR, while also calling for sociotechnical designs that mitigate over-reliance on AI and foster critical engagement.

6.3 Responsible privacy practices

Cvetkovic and Bittner (2022) and Cvetkovic et al. (2025) highlight the importance of user trust in AI, but they also underscore that trust alone is not enough; respecting individuals' boundaries around data sharing is equally critical (Cvetkovic et al., 2025). In the context of human-centered AI, this means design choices must accommodate users' preferences not to disclose certain information, especially regarding sensitive aspects such as emotional or cognitive states. Regulatory frameworks like the proposed EU AI Act⁸ are beginning to address these concerns by prohibiting emotional surveillance in professional settings. Nonetheless, many commercial AI tools still attempt to capture implicit employee data and convert it into metrics for management (Mantello et al., 2023). Practitioners should heed these insights and adopt more human-centered AI designs that reduce excessive monitoring and protect user autonomy.

Rather than simply deploying AI indiscriminately, managers can design collaborative processes that highlight the tangible benefits of sharing certain types of data, be it speed, accuracy, or tailored recommendations, while giving individuals the ability to opt out or restrict usage where privacy concerns prevail. Such an approach not only fosters a sense of

⁸ <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>

fairness and security but also encourages employees to engage meaningfully with AI-driven workflows.

In sum, the practical contributions offered by this dissertation equip organizations and their stakeholders with a structured framework for harnessing AI in knowledge-centric work. By leveraging the task allocation frameworks, adopting instantiations of human-centered AI solutions, safeguarding trust and privacy, enhancing creative collaboration, and tailoring prompts for generative models, practitioners can design socio-technical systems that unlock new efficiencies without eroding human autonomy. This synergy of human strengths and machine capabilities lays the groundwork for sustainably innovative workplaces where teams are both empowered by and in command of advanced AI-driven assistance.

7 Limitations

This cumulative dissertation faces several limitations that influence the relevance, rigor, and projectability of its findings (Hevner & Chatterjee; Venable, 2006). These limitations span the definition of the problem space, the production of the solution space, and the evaluation of the generated design knowledge, each of which is discussed below.

In framing the problem space, this dissertation relied on literature reviews and qualitative methods, such as semi-structured interviews, to capture real-world challenges and requirements related to HIS. Regarding the literature reviews, the choice of knowledge sources may have shaped the focus of the identified problem space. For the latter, as detailed in Cvetkovic, Gierlich-Joas, et al. (2023), these methods generated rich insights into subjective user experiences, but they are inherently shaped by the perspectives of the individuals interviewed. Although established coding and analysis techniques were used to mitigate researcher bias, complete objectivity is unattainable (Myers & Newman, 2007). Moreover, the lack of collaboration with specific organizations to embed the studies within concrete organizational workflows introduces constraints on the contextual breadth of the findings. Had the research been conducted across multiple companies with diverse structures and cultures, the set of identified challenges and user needs might have differed or expanded, potentially increasing the external validity of the problem definition.

This dissertation sought to generate design knowledge by synthesizing existing literature and empirical findings (Cvetkovic & Bittner, 2022; Cvetkovic, Gierlich-Joas, et al., 2023; Cvetkovic, Rosenberg, & Bittner, 2023; Cvetkovic et al., 2024). While the selection of relevant literature adhered to established review methods, certain perspectives may have been overlooked due to the choice of databases, search terms, or inclusion criteria. Consequently, the scope of the proposed DPs reflects specific assumptions about HIS, CAs and AI-driven facilitation. Furthermore, the developed solutions are rooted in particular empirical settings rather than in diverse or cross-industry scenarios. This contextual grounding provides practical detail but may limit the projectability of design knowledge to domains or user groups beyond those studied. Although efforts were made to describe DRs and DPs systematically to enhance transparency and reusability, organizations seeking to adopt these solutions might need to adapt them to align with differing resource constraints, internal workflows, cultural norms, and team configurations. The overarching DPs, although higher in level of abstractness and thus easier to reuse and adjust to specific use cases, have not been validated as universally applicable.

A multi-method approach, involving both qualitative and quantitative techniques (e.g., interviews, surveys, user tests), was employed to assess the utility, effectiveness, and efficiency of the proposed design knowledge. While some evaluations gathered significant detail from participants (Cvetkovic et al., 2025), the number of respondents in certain studies (Cvetkovic, Rosenberg, & Bittner, 2023) remained limited. Furthermore, most

evaluation criteria emphasized participants' perceptions, which addresses the socio-technical dimension but only partially captures the technical side of HIS performance. More objective evaluation metrics, such as improvements in task efficiency and productivity compared to traditional methods, would strengthen claims about the tangible benefits of the instantiated artifacts. Additionally, the absence of long-term, large-scale field experiments in real organizational contexts restricts the ecological validity of the findings. Although small-scale tests and semi-naturalistic simulations offered controlled insights into how HIS might function, they cannot fully replicate the complex interplay of factors (e.g., internal hierarchies, political considerations, or resource constraints) that typify real-world deployments. Lastly, the research produced several artifacts of low maturity, indicating that further development, refinement, and large-scale testing are necessary before they can be robustly integrated into organizational environments.

As a cumulative dissertation, this work integrates multiple individual studies, each contributing uniquely to an overarching research goal. However, the coherence and consistency of findings across disciplinary or methodological boundaries can be challenging to maintain, potentially leaving certain nuances underexplored. Interdisciplinary topics, especially those blending socio-technical, ethical, and practical considerations, require ongoing scrutiny to ensure that proposed solutions remain robust when scaled up or transferred to new environments.

Taken together, these limitations reflect the inherent constraints of the research context, choices of methods, and breadth of empirical engagement. Although the dissertation strives for methodological rigor and practical relevance, generalizability cannot be assumed.

8 Implications for Future Research

Building on the theoretical and practical contributions as well as the limitations outlined in this dissertation, several avenues for future research emerge. These extend beyond the specific settings and artifacts presented, opening up broader considerations around the socio-technical integration of AI-based solutions into creative work and complex problem-solving. In particular, they address the need to refine and test design knowledge in diverse contexts, strengthen methodological rigor in evaluations, and explore the ethical and practical implications of HIS. Finally, ongoing tests of changes in both problem and solution space are necessary to account for the ephemeral nature of design knowledge (vom Brocke et al., 2020).

8.1 Extending evaluation contexts and methodologies

One key limitation of the present work lies in the partial reliance on small-scale and semi-naturalistic evaluations that lack the embedding in an organizational context. Although these studies yielded valuable insights into how knowledge workers interact with AI and how AI can be designed to support creative and problem-solving tasks, larger and more longitudinal investigations could capture the complexity of organizational structures and work routines with greater fidelity. Accordingly, future research should involve field studies in varied organizational contexts that thoroughly examine the long-term sustainability of AI adoption. For instance, embedding AI-facilitated workflows within established corporate environments would permit the observation of persistent usage patterns, organizational dynamics, and evolving trust in AI over extended periods.

Although a variety of methods, ranging from interviews to large sample surveys, have supported the findings in this dissertation, additional modes of measurement and triangulation could deepen and strengthen future evaluations. Beyond user-reported perceptions, objective measures such as behavioral analytics, organizational performance metrics, or error rates could help establish clearer causal links between AI intervention and improvements in collective productivity or decision quality. Particularly, a randomized between-group investigation of the outcomes of tasks performed with and without AI in organizational contexts could pave the way toward establishing causal relationships. By incorporating more robust data collection, future research could offer detailed insights into how newly introduced AI-driven processes evolve within an organization and interact with existing workflows, hierarchies, and cultural norms.

8.2 Contextualizing AI design knowledge in broader domains

Although the presented design knowledge and artifacts have proven beneficial for creative work and problem-solving, their applicability to different domains and tasks remains an

open question. Future research could explore the extent to which the proposed DPs, particularly around task allocation and information privacy, transfer to more regulated or safety-critical fields, including healthcare, legal services, and finance. These domains often feature stringent governance and high-stakes tasks that may demand different trust-building mechanisms and stronger assurances of reliability and interpretability.

In parallel, investigating AI integration across culturally diverse settings represents another rich opportunity. Cross-cultural studies on user acceptance, task delegation preferences, and privacy considerations could illuminate how cultural norms, and organizational practices influence both the willingness to adopt AI and the manner in which AI-driven insights are incorporated. For example, norms around hierarchy, authority, or data sharing might affect employees' trust and readiness to use an AI-based facilitator. By systematically mapping these cultural variations, researchers can propose context-sensitive adaptations of design knowledge to ensure that AI solutions remain human-centered in global work environments.

Furthermore, the fluidity of tasks that humans are willing to delegate to AI, particularly in light of rapidly evolving LLMs, underscores the need for ongoing adaptation of the task allocation frameworks introduced here. As AI gains proficiency in tasks once considered exclusively human, such as creative ideation or strategic decision-making, future work should identify the boundary conditions and new skill sets that humans need to interact effectively with increasingly capable AI. However, while the potential of LLMs is undeniable, in reality, many employees and individuals have found themselves going in circles while trying to solve particular tasks with ChatGPT. Not only does this push one further away from a solution, but it also takes away time and cognitive resources. It is thus imperative that industrial and academic institutions collaborate and longitudinally track the effects of AI on both resource gains and losses, as well as employee acceptance, to derive how to best shape the human-AI collaboration in a human-centered way. This might include training programs that bolster human expertise in shaping AI outputs, refining prompt engineering techniques, or judging the trustworthiness of AI-generated recommendations. Additionally, a more granular analysis of non-routine tasks would clarify whether users become increasingly comfortable sharing higher-level responsibilities with AI once they gain trust and see demonstrable competence in real-world scenarios.

Moreover, while this dissertation provides practical frameworks and prototypes for AI-infused facilitation, it only begins to address the interplay between AI-based functionalities and human team dynamics. Future research could investigate how these AI solutions function in realistic, distributed setups, such as teams using videoconferencing platforms alongside collaborative whiteboards, and whether such tools enhance or potentially disrupt existing workflows. Examining these technologies in real-world scenarios would significantly extend ecological validity and offer deeper insights into how best to embed AI functionalities without undermining user autonomy or fostering undue reliance on automation. Apart from the effect on team workflows, it would be valuable to consider how

AI-driven critical feedback or facilitation interventions affect team cohesion and group creativity over time. While the short-term results of this thesis point to reductions in social evaluation apprehension and increases in idea quality, further work might explore whether repeated exposure to AI criticism leads to desensitization, enhanced creativity, or new coping mechanisms. Moreover, research on AI in teams opens many intriguing pathways for investigation, as insights from human–human teamwork need to be validated and potentially adapted for scenarios where AI becomes a teammate or co-facilitator (Demir et al., 2019). Teamwork is already complex as it is, and introducing AI adds another layer of factors, such as trust, accountability, and role clarity, that warrant systematic exploration. Such lines of inquiry would enrich our understanding of how continuous AI involvement shapes interpersonal relationships, self-efficacy, and overall team performance.

8.3 Evolving role of LLMs

In tandem with the growing influence of LLMs, prompt engineering has emerged as an important skill for shaping AI outputs in creative and problem-solving tasks. As demonstrated in this dissertation, carefully designed prompts can enhance ideation quality. However, ongoing improvements in model architecture may reduce the complexity of such prompt-based interactions, potentially rendering prompt engineering techniques obsolete (Oppenlaender et al., 2024). The newest reasoning models are already showing impressive results in breaking down complex tasks on their own. Future research should investigate how to optimize human–AI collaboration in scenarios where LLMs become increasingly adept at self-adaptation, and whether more advanced models can automatically compensate for vague or incomplete prompts. Richter and Schwabe (2025) also emphasize the importance of designing adaptive practices that continuously accommodate the evolving contributions of AI.

Equally important is the need to minimize overreliance on LLMs, as they can produce results that are plausible yet contain inaccuracies and biases. Studies should explore mechanisms for supporting critical scrutiny of AI outputs, such as layered explainability features or automated fact-checking modules, which help users maintain a healthy scepticism toward model-generated content. Achieving an optimal balance between streamlined prompt engineering, user vigilance, and dynamic model improvements is essential for harnessing the evolving capabilities of LLMs while safeguarding the integrity of human–AI collaboration.

Overall, these research directions promise to deepen our understanding of how AI affects both individual and collaborative workflows across a variety of contexts. By expanding the evaluation methods, exploring transferability in new domains, and accounting for organizational and cultural nuances, upcoming studies can further advance the theories and principles presented here. Ultimately, such investigations will solidify the foundations of a

nascent design theory for HIS and help craft more robust, context-aware solutions that balance efficiency gains with ethical, social, and psychological considerations.

9 Hybride Wissensarbeit

Oeste-Reiß, S., Bittner, E., Cvetkovic, I., Günther, A., Leimeister, J.M., Memmert, L., Ott, A., Sick, B., & Walter, K. (2021). Hybride Wissensarbeit. *Informatik Spektrum*, 44(3), 1-5. <https://doi.org/10.1007/s00287-021-01352-0>

Zusammenfassung

Aufgrund der Fortschritte im Bereich der künstlichen Intelligenz (KI) ergeben sich neue Gestaltungsmöglichkeiten zur Reorganisation von Wissensarbeit an der Schnittstelle von Menschen und Maschine. Durch die Verschmelzung von menschlicher und künstlicher Intelligenz können komplementäre Stärken zum Lösen von Arbeitsaufgaben gebündelt werden. Neuartige Wissensarbeitssysteme sind erforderlich, die Wissensarbeitende sowohl bei der Ausübung von Routineaufgaben als auch Nichtroutineaufgaben unterstützen.

Der Schlagwortartikel umreißt die Grundlagen von Wissensarbeit und arbeitet die Charakteristika von Routineaufgaben und Nichtroutineaufgaben innerhalb von Arbeitsprozessen auf. Es werden Grenzen klassischer IT-gestützter Wissensarbeitssysteme als Werkzeuge umrissen, die Wissensarbeitende im Arbeitsprozess unterstützen. Aufbauend darauf diskutiert der Artikel, dass technologische Fortschritte eine arbeitsprozessintegrierte und personalisierte Unterstützung von Wissensarbeitenden erlauben. Dazu verweist der Artikel auf die Charakteristika von kollaborativ interaktiv lernenden Systemen. Darauf aufbauend werden Hybride Wissensarbeitssysteme vorgestellt, die sowohl das KI-unterstützte menschliche Lernen als auch das Mensch-unterstützte maschinelle Lernen unterstützen. In diesem Zusammenhang beschreibt der Artikel definitorische Grundlagen von menschlicher, künstlicher und hybrider Intelligenz, stellt drei Archetypen von Mensch-Maschine-Aufgaben innerhalb von hybriden Wissensarbeitssystemen vor und umreißt drei Gestaltungsdimensionen solcher Systeme. Mittels eines Praxisbeispiels zum kollaborativen Schreiben von Journalist*innen beschreibt der Artikel exemplarisch die hybride Wissensarbeit. Der Artikel schließt mit einem Ausblick auf zukünftige Forschungsbedarfe.

9.1 Von IT-gestützten Wissensarbeitssystemen hin zu Hybriden Wissensarbeitssystemen

Wissensarbeit bezieht sich auf den Wissensanteil einer Arbeitsaufgabe. Dies umfasst das benötigte Domänenwissen, den Arbeitsprozess und -aufgabe sowie die Aktivitäten zum Ausführen der Aufgabe [4]. Wissensarbeit ist besonders, da sie sowohl Routine- als auch Nichtroutineaufgaben umfasst [11].

- *Routineaufgaben* sind wiederholend, automatisierbar und programmierbar. Sie gründen auf explizierbarem deklarativen Faktenwissen („know what“) [12]. Der

Wissenstransfer und -erwerb kann digital oder papierbasiert erfolgen – u. a. über Daten und Dokumentationen in Form von Büchern, Bedienungsanleitungen oder FAQ [13]. Beispiele finden sich in der Wiedergabe und Zusammenführung von Faktenwissen (z.B. Rechnungserstellung, Terminvereinbarung).

- *Nichtroutineaufgaben* sind vielfältig und umfassen Einzelfallentscheidungen und -lösungen. Sie gründen auf schwer explizierbarem Erfahrungswissen („know how“) [12]. Der Wissenstransfer und -erwerb erfolgt über Lern- und Anwendungsprozesse. Diese sind gekennzeichnet durch eine soziale und häufig informelle Interaktion mit anderen Wissensträgern und der Umwelt, bei der u. a. erfahrungsbasiert Faktenwissen vernetzt und angewendet wird [2]. Beispiele finden sich in der Anwendung, Analyse, Beurteilung und Schaffung von Wissen (z. B. Mitarbeitererinarbeitung, Geschäftsmodellentwicklung, Mentoring).

Aus organisationaler Perspektive wird durch die Ausführung von Arbeitsaufgaben Wissen entwickelt, gesichert und zwischen Wissensarbeitenden transferiert. Dazu benötigen Wissensarbeitende Fähigkeiten, die sie zum lösungsorientierten Handeln, zu Kollaboration mit anderen Wissensträgern und zum lebenslangen Lernen befähigen [15].

Werkzeuge, die Wissensarbeitende hierbei unterstützen sind Wissensmanagementsysteme. Solche Systeme dienen der Identifikation, dem Erwerb, der Entwicklung, der Speicherung, der Pflege, der Nutzung und dem Transfer von Wissen [2]. Sie integrieren die Verwaltung sowohl von explizitem als auch von implizitem Wissen [5], das zur Lösung von Routine- und Nichtroutineaufgaben herangezogen wird. Als Strategien zum Teilen von Wissen kann zwischen der Kodifizierungs- und die Personalisierungsstrategie unterschieden werden. Erstere zielt auf das Sammeln, Speichern und Bereitstellen von explizitem Wissen mittels Datenbanken, Dokumentenmanagement oder Intranet [5, 7]. Letztere zielt auf die Befähigung von Wissensarbeitenden, ihr implizites Erfahrungswissen IT-gestützt über Wissensnetzwerke und Groupware-Systeme zu kommunizieren [7]. Solche Systeme dienen oft als Nachschlagewerke, in denen Wissensarbeitende proaktiv Wissen zur Lösung von Arbeitsaufgaben suchen und die Wissensbestände einspeisen und aktuell halten müssen.

Heutzutage erlauben Technologien, dass intelligente Systeme zunehmend arbeitsprozessintegriert und personalisiert Wissen bereitstellen und aktuell halten können. Voraussetzung hierfür ist ein lebenslanges Lernen des Systems, das auf einem „Human-in-the-loop“-Modell gründet. Solche Systeme werden zunehmend in der Lage sein, mit unvollständigen Beispieldaten beim Lern- und Optimierungsprozess zur Entwicklungszeit umzugehen. Kollaborativ interaktiv lernende Systeme, sogenannte CIL-Systeme, gründen auf diesem Ansatz und verfügen über lebenslange Lernfähigkeiten [14]. Technische Systeme, die auf CIL-Techniken beruhen sind:

- *lernend*: d. h. selbstorganisierte Verbesserung der eigenen Wissensbasis und zugehöriger Lernprozesse [14];

- *kollaborativ*: d. h. wechselseitige auf ein gemeinsames Ziel ausgerichtete Einbindung intelligenter Systeme und Menschen in den Lernprozess [14];
- *interaktiv*: d. h. Informations- und Wissensaustausch erfolgt in beide Richtungen [14].

Technische Systeme, die auf CIL-Techniken basieren bieten das Potenzial, Wissensarbeitenden aktuelle Wissensbestände und personalisierte Wissensabrufe zu ermöglichen. Durch ihr selbstorganisiertes Lernen liefern sie Potenziale, um in implizite Wissenstransformationsprozesse vorzustoßen, welche bisher überwiegend zwischen Menschen möglich waren [2, 14]. Im Vergleich zu Wissensmanagementsystemen benötigen sie weniger dedizierte Wissensaufbereitung durch menschliche Systemadministratoren und versprechen durch ihre Kontextadaptivität Wissensarbeitenden personalisierte Wissensrepräsentation bereitzustellen. Weiterhin liefern sie Potenziale für eine arbeitsprozessintegrierte Nutzung, da sie mit menschlichen Nutzern kollaborieren können (z. B. Verifizierung von Beispieldaten, Übergabe/-nahme von Arbeitsständen).

Deshalb benötigen wir eine neue Generation von Wissensarbeitssystemen, die sowohl ein lebenslanges Lernen des technischen Systems als auch ein lebenslanges Lernen des Wissensarbeitenden (soziales System) gewährleisten.

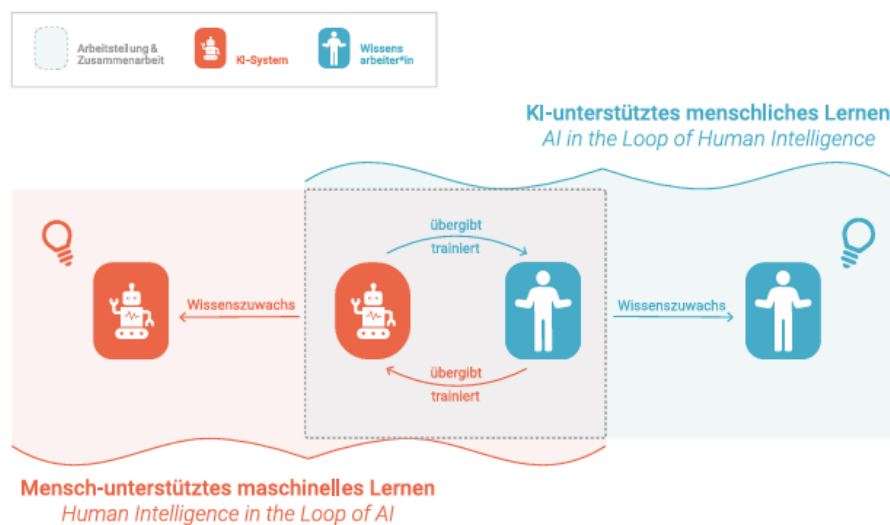


Abb. 1. Hybrides Wissensarbeitssystem (HWA-System)

Diese Systeme bezeichnen wir im Folgenden als *Hybride Wissensarbeitssysteme (HWA-Systeme)*. HWA-Systeme ermöglichen:

- *KI-unterstütztes menschliches Lernen* („AI in the Loop of Human Intelligence“): Das technische KI-System nutzt CIL-Techniken. Es fungiert arbeitsprozessintegriert als Lerncoach und trainiert Wissensarbeitende mit dem Ziel einer personalisierten Wissensvermittlung zur Lösung von Arbeitsaufgaben. Es erfolgt eine Automatisierung und Augmentierung von Routine- und Nichtroutineaufgabenbestandteilen.

- *Menschunterstütztes maschinelles Lernen („Human Intelligence in the Loop of AI“)*: Der Mensch fungiert arbeitsprozessintegriert als Lerncoach und trainiert das technische KI-System mit dem Ziel einer Verifizierung von Wissen.

Hieraus ergibt sich ein wiederkehrender Kreislauf von Übergaben und Training zwischen Wissensarbeitenden und KI-System und folglich ein neues Verständnis von Zusammenarbeit und Arbeitsteilung (siehe Abb. 1). Es bedarf einer menschenzentrierten Gestaltung der Lern- und Zusammenarbeitsprozesse.

9.2 Merkmale und Gestaltungsdimensionen Hybrider Wissensarbeitssysteme

Die Grundidee von HWA-Systemen ist es, mittels Hybrider Intelligenz Wissensarbeitende fortwährend zum Erwerb und Transfer von Wissen zur Ausübung ihrer Arbeitsaufgaben zu befähigen. Wissensarbeitende sollen kognitiv im Arbeitsprozess bei der Ausübung von Arbeitsaufgaben entlastet und unterstützt werden. Hierzu werden komplementärere Stärken von Wissensarbeitenden und KI-System genutzt, sodass menschliche und künstliche Intelligenz verschmelzen.

Dies bedingt eine Verschiebung der Arbeitsteilung von Wissensarbeitenden und KI-System und vergrößert den Anteil der Aufgaben, die von Mensch und KI-System im Team erledigt werden können. HWA-Systeme bilden die Plattform für eine gesellschaftlich-wünschenswerte Neuverteilung von Arbeit, die durch KI-basierte Automatisierung und Augmentierung von Aufgabenbestandteilen gekennzeichnet ist. Ein Schlüsselement, das dies ermöglicht ist das lebenslange Lernen sowohl von Wissensarbeitenden als auch vom KI-System. Lebenslanges Lernen wird durch die Gestaltung hybrider Zusammenarbeitspraktiken ermöglicht, die in HWA-Systemen implementiert sind. Zur Gestaltung sind einerseits Kenntnisse über die Fähigkeiten von Wissensarbeitenden und KI-System und andererseits die Analyse und Zergliederung des Arbeitsprozesses und insbesondere der Aufgabenbestandteile erforderlich.

Eine Definition von Intelligenz gibt Aufschluss über die Fähigkeiten und Stärken der jeweiligen Wissensträger.

- *Menschliche Intelligenz* bezieht sich auf die Fähigkeit zu lernen, zu verstehen und sich mithilfe des eigenen Wissens an die Umgebung anzupassen [9]. Sie weist Stärken im Bereich des Erfahrungswissens und bei der Bewältigung von Nichtroutineaufgaben auf.
- *Künstliche Intelligenz* bezieht sich auf die Fähigkeit eines Systems, externe Daten korrekt zu interpretieren, von ihnen zu lernen und das Gelernte zu nutzen und um bestimmte Ziele und Aufgaben durch flexible Anpassung zu erreichen [8]. Sie weist Stärken im Bereich des Faktenwissens und bei der Bewältigung von Routineaufgaben auf.

Hybride Zusammenarbeitspraktiken nutzen sogenannte *Hybride Intelligenz*, bei der menschliche und künstliche Intelligenz verschmelzen. Diese verspricht durch geeignete Kombination der menschlichen und künstlichen Intelligenzen ein Team zu schaffen, welches die individuellen, komplementären Stärken bündelt und Schwächen kompensiert. Somit können gemeinsame Ziele erreicht und Probleme gelöst oder besser gelöst werden, als dies dem Menschen oder der KI eigenständig möglich wäre [1, 6]. Der Ansatz verfolgt eine gegenseitige Ergänzung (Augmentierung) und nicht Ersetzung von Wissensarbeitenden durch die KI [1]. Im Fokus stehen dynamische Umgebungen und Probleme, deren Lösung domänenspezifisches Wissen bedürfen und die keine spezifische Grundwahrheit („ground truth“) besitzen und somit bisher typischerweise dem menschlichen Intellekt vorbehalten waren [1, 6].

Da Arbeitsprozesse und -aufgaben von Wissensarbeiten den sowohl aus Routine- als auch Nichtroutineaufgaben bestehen, zielen HWA-Systeme auf hybride Zusammenarbeitspraktiken mit folgenden Archetypen von Aufgaben [16]:

- *Automatisierung*: Routineaufgabenbestandteile werden identifiziert, für die vollständige Beispieldaten für Lern und Optimierungsprozesses des KI-Systems existieren. Diese Aufgaben werden so automatisiert, dass ein Deskilling von Wissensarbeitenden vermieden wird.
- *Verifizierung*: Aufgabenbestandteile werden identifiziert, für die keine oder nur unvollständige Beispieldaten existieren. Wissensarbeitende verifizieren dieses Wissen.
- *Augmentierung*: Routine- und Nichtroutineaufgabenbestandteile werden identifiziert, die aus Gründen mangeln der Akzeptabilität gegenüber gesellschaftlichen Werten und Normen nicht automatisiert werden sollen. Mittels Augmentierung werden Wissensarbeitende beispielsweise in Kreativitätsprozessen unterstützt und entlastet.

Zur Gestaltung hybrider Zusammenarbeitspraktiken, die Hybride Intelligenz erschließen und nutzen, ist eine sozio- technische Gestaltungsperspektive notwendig. HWA-Systeme sind Arbeitssysteme [3, 10], die als eine Art Marktplatz für eine Allokation von Arbeitsaufgaben und -ständen fungieren. Gestaltungsdimensionen sind:

- das *soziale Subsystem* (Wissensarbeitende);
- das *technische Subsystem* (CIL-Techniken) sowie
- *Geschäftsprozesse* (hybride Zusammenarbeitspraktiken).

In Letzterem erfolgt die eigentliche Arbeitsteilung und Zusammenarbeit zwischen Wissensarbeitenden und KI- System. Gestaltungsmethoden des sogenannten Collaboration Engineering [10, 17], einem Ansatz zur Entwicklung und zur Umsetzung von Kollaborationsprozessen, die von Nichtkollaborationsexperten ausgeführt werden können, um hochwertige, wiederkehrende Aufgaben zu erfüllen, liefern Werkzeuge um einen Arbeitsprozess hinsichtlich (i) des zu erbringenden Arbeitsergebnisses zu analysieren; (ii)

in Aufgabenbestandteile zu zergliedern; (iii) Arbeitsstände und Übergaben zu definieren; (iv) Zusammenarbeitsaktivitäten zu definieren. Folglich sind HWA-Systeme soziotechnische Systeme, die Zusammenarbeits- und Lernpraktiken implementieren.

9.3 Anwendungsbeispiel: Kollaboratives Schreiben im Journalismus

Betrachten wir ein HWA-System zur Unterstützung von Journalist*innen: Drei Journalist*innen einer überregionalen Zeitung und ein KI-basierter Assistent arbeiten als Team und nutzen ein HWA-System. Ihr gemeinsames Ziel ist das Schreiben eines Artikels zu den Auswirkungen von Homeoffice auf die Produktivität und das Wohlbefinden. Der Arbeitsprozess umfasst folgende Aufgabenbestandteile: Recherche relevanter Grundlagen, Durchführung von Interviews mit Expert*innen und Schreiben des Artikels. Um die kognitive Belastung der Journalist*innen zu reduzieren und einen qualitativ hochwertig recherchierten Artikel zu schreiben, erfolgt folgende Allokation von Aufgaben zu den jeweiligen Wissensträgern.

- *Automatisierung („AI in the Loop of Human Intelligence“)*: Auf Basis einer ersten, kurzen Skizze des Artikels übernimmt der KI-basierte Assistent Informationsrecherche und -verarbeitung. Hierfür extrahiert der Assistent Entitäten aus der Skizze und führt eine Suche in verschiedenen Datenbanken durch. Die Informationen werden priorisiert und dem menschlichen Verarbeitungsprozess entsprechend anschaulich aufbereitet. Sobald die Journalist*innen mit dem Schreiben des Artikels fortfahren, nutzt die KI die neuen Informationen, um die Vorschläge entsprechend anzupassen.
- *Verifizierung von Wissen („Human Intelligence in the Loop of AI“)*: Die Journalist*innen geben Feedback zu den von dem KI-basierten Assistenten unterbreiteten Vorschlägen, damit sich der KI-basierte Assistent bei der Informationsrecherche verbessern kann. Dazu bewerten die Journalisten die vom KI-System gelieferten Informationen, Vorschläge, wie z. B. Studien, Statistiken oder Experten hinsichtlich Relevanz und Seriosität.
- *Augmentierung („AI in the Loop of Human Intelligence“)*: Die Journalist*innen widmen sich kreativen Aufgaben wie der Entwicklung einer „Story“, um die Aufmerksamkeit der Leser*innen zu halten oder der Durchführung von Interviews.

9.4 Forschungsziele und Ausblick

Hybride Wissensarbeit und HWA-Systeme beschreiben ein im Entstehen befindliches Forschungsgebiet soziotechnischer Arbeitssysteme, die hybride Zusammenarbeit und lebenslanges Lernen von Menschen und System gleichermaßen ermöglichen. Um HWA-Systeme zu entwickeln, ist Gestaltungswissen hinsichtlich (i) des sozialen Subsystems, insbesondere Wissen über die Anatomie der Zusammenarbeit und des Lernens von

Menschen sowie Collaboration Engineering und (ii) des technischen Subsystems, insbesondere Wissen über CIL-Techniken und aktives Lernen, erforderlich.

Förderung The research leading to the presented results was conducted in the context of the project HyMeKI (funded by the Federal Ministry of Education and Research (BMBF), 01IS20057)

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

9.5 Literatur

1. Akata Z, Balliet D, de Rijke M et al (2020) A research agenda for hybrid intelligence: augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. *Computer* 53:18–28
2. Alavi M, Leidner DE (2001) Review: knowledge management and knowledge management systems: conceptual foundations and re- search issues. *MISQ* 25:107
3. Alter S (2017) Work system theory and work system method: a bridge between business and IT views of IT-reliant systems in organizations. *Proceedings of the 10th Innovations in Software Engineering Conference*
4. Arlinghaus A (2017) Wissensarbeit: Aktuelle arbeitswissenschaftliche Erkenntnisse. Hans-Böckler-Stiftung, Düsseldorf
5. Dalkir K (2005) Knowledge management in theory and practice. Routledge, Amsterdam, London
6. Dellermann D, Calma A, Lipusch N, Weber T, Weigel S, Ebel P (2019) The future of human-AI collaboration: a taxonomy of de- sign knowledge for hybrid intelligence systems. In: Bui T (Hrsg) *Proceedings of the 52nd Hawaii International Conference on System Sciences: Hawaii International Conference on System Sciences*

7. Greiner ME, Böhmman T, Krcmar H (2007) A strategy for knowledge management. *J Knowl Manag* 11:3–15
8. Haenlein M, Kaplan A (2019) A brief history of artificial intelligence: on the past, present, and future of artificial intelligence. *Calif Manage Rev* 61:5–14
9. Leimeister JM (2010) Collective intelligence. *Bus Inf Syst Eng* 2:245–248
10. Leimeister JM (2014) *Collaboration engineering*. Springer, Berlin, Heidelberg
11. Pyöriä P (2005) The concept of knowledge work revisited. *J Knowl Manag* 9:116–127
12. Salaberry MR (2018) Declarative versus procedural knowledge. In: Lontas JI, DelliCarpini M (Hrsg) *The TESOL encyclopedia of English language teaching*. Wiley-Blackwell, Hoboken
13. Schmitz C, Hotho A, Jäschke R, Stumme G (2006) Kollaboratives Wissensmanagement. In: Pellegrini T, Blumauer A (Hrsg) *Semantic web*. Springer, Berlin, Heidelberg
14. Sick B, Oeste-Reiß S, Schmidt A, Tomforde S, Zweig AK (2018) Collaborative interactive learning. *Informatik Spektrum* 41:52–55
15. Thordsen T, Murawski M, Martensen M, Rademacher C, Bick M (2020) Digitale Kompetenzen – Welche Anforderungen stellt die Digitalisierung an Wissensarbeiter und Unternehmen? In: Heisig et al (Hrsg) *WM in digitalen Arbeitswelten – Aktuelle Ansätze und Perspektiven*
16. Traumer F, Oeste-Reiß S, Leimeister JM (2017) Towards a Future Reallocation of Work between Humans and Machines—Taxonomy of Tasks and Interaction Types in the Context of Machine Learning. In: *International Conference on Information Systems (ICIS)*. Seoul, South Korea
17. de Vreede G-J, Briggs RO (2019) A Program of Collaboration Engineering Research and Practice: Contributions, Insights, and Future Directions. *J Manag Inf Syst* 36:74–119

10 Task Delegability to AI: Evaluation of a Framework in a Knowledge Work Context

Cvetkovic, I. & Bittner, E. (2022). Task Delegability to AI: Evaluation of a Framework in a Knowledge Work Context. In *Proceedings of the 55th Hawaii International Conference on System Sciences (HICSS)*, Virtual conference

Abstract

With the increased research focus on ways to use AI for augmentation rather than automation of knowledge-intensive work, a myriad of questions on how this should be accomplished arises. To break down the complexity of Human-AI collaboration, this paper pursues the identification of factors that contribute to the delegation of tasks to AI in such a setting, and consequently gain insights into requirements for meaningful task allocation. To address this research gap, we carried out an empirical study on an existing task delegability framework in a knowledge work context. We employed several statistical approaches such as confirmatory factor analysis, linear regression, and analysis of covariance. Results show that an adapted framework with fewer factors fits the data better. As for the framework factors, we show that the factor trust predicts delegability best. Furthermore, we find a significant impact of task on delegability decision. Finally, we derive theoretical and design implications.

10.1 Introduction

Artificial intelligence (AI) is the science and engineering behind creating intelligent machines, particularly computer programs, which try to grasp and, to some extent, imitate human intelligence [1]. With the rise of AI, concerns about automation and consequent job loss have increased [2]. Galore research still focuses on finding ways to automate work with AI, without considering these consequences. However, in the last decade, researchers started giving rise to the importance of keeping the human-in-the-loop [3]. Furthermore, growing evidence of the advantages of Human-AI collaboration started appearing in the literature. This marks a shift to a collaborative rather than automation perspective of AI [4]. For example, Dellermann et al. [5], and Bittner et al. [6] argue that combining complementary strengths of human intelligence and AI leads to a better performance than each could achieve separately. Not only does this improve the outcome and group performance, but such a constellation also significantly contributes to mutual learning [7]. However, there is a long way to achieve the optimal collaboration for such socio-technical ensembles [8], and various research gaps to address in the first place. First of all, a realignment of the task allocation is necessary because the challenges of the modern world of work exceed the abilities of individuals [6]. In addition, the interplay between humans

and tasks while collaborating with AI (assistants) as well as the outcomes of this collaboration necessitate further investigation. An AI assistant helps users achieve their tasks by interacting with them while using machine intelligence in form of, e.g., natural language processing, speech recognition, or machine learning [9].

AI capabilities are opening up new pathways for collaboration between knowledge workers and machines. Knowledge workers' main attribute is knowledge. They apply this knowledge to develop products and services [10]. As knowledge work gains in complexity, it is becoming more and more challenging for individuals [11, 12]. Technological advances in the field of AI offer new design opportunities for the reorganization of knowledge work at the interface of humans and AI [13]. Many knowledge-based tasks can now be solved more effectively with AI technologies than with earlier technologies. For example, AI can be used to automate Q&A, enabling humans to focus on high-level interactions. But to take full advantage of the prospects of Human-AI collaboration in knowledge work, companies will have to redesign knowledge-work processes and jobs [14]. The existing potential for automation of tasks in knowledge work does not directly correspond to increased performance [15]. In contrast, AI assistants should be designed with the intention to augment, not replace human contributions [16]. Oeste-Reiß et al. [13] introduce Hybrid Knowledge Work Systems that continuously enable knowledge workers to acquire and transfer knowledge for the performance of their work tasks by means of hybrid intelligence [5]. Such systems help knowledge workers by relieving their cognitive load "The ability to achieve complex goals by and supporting the work process. This causes a shift in the division of labour between knowledge workers and AI assistants and increases the proportion of tasks that can be completed by humans and AI assistant(s) as a team [13]. For a meaningful task allocation in such a constellation, it is first necessary to identify the factors that influence this allocation in either a positive or negative way. This area of Human-AI collaboration is underexplored, especially from a point of task delegation. To the best of our knowledge, there is only one developed framework of task delegability [17]. The relatively novel framework was developed for (potentially AI-supported-)work in general and is now to be examined whether it also provides explanatory contributions for knowledge work. It is hard to make inferences about the applicability of the framework to more specific contexts, such as knowledge work in postgraduate research, an exemplary class of knowledge work fulfilling necessary criteria for our study scope (see section 3.2). To address this research gap, this study aims at answering the following research questions:

- **RQ 1:** How does the existing framework of task delegability apply to a specific context of knowledge work in postgraduate research?

The existing framework indicates different correlations of components to the delegation decision [17]. The question arises whether and to what extent are these present in the current context as well. Thus, we are also interested in answering:

- **RQ 2:** Which components of the framework contribute most to the delegation decision?

Finally, the framework authors found that some tasks are rather delegated than other [17]. This motivates the following question, specifically relating to knowledge-intensive tasks:

- **RQ3:** Does the delegation decision depend on the task?

The goal of the study is to validate and further develop the framework in a new application context and better understand the delegation decisions of knowledge workers, especially researchers, and their influencing factors, in order to derive design knowledge for Human-AI collaboration.

10.2 Theoretical Background

10.2.1 Human-AI Collaboration

To overcome limitations of humans and AI alone, Dellermann et al. [5] propose a construct of Hybrid Intelligence, which they define as:

“The ability to achieve complex goals by combining human and artificial intelligence, thereby reaching superior results to those each of them could have accomplished separately, and continuously improve by learning from each other.”

Achieving this predisposes collaboration between humans and AI. Hybrid collaboration involves at least two actors (at least one being human and one AI) having the same understanding of a shared goal and working together toward this mutual goal [18, 19, 6]. Examples of Human-AI collaboration can be seen in clinical decision support systems and customer service chatbots [19, 20]. There are many advantages to Human-AI collaboration, e.g., it improves group performance [21, 22], inter alia. Still, there are also many unresolved questions about human-AI teams dynamics and characteristics as well as about the transferability of concepts from solely human teams [23]. Human-AI collaboration, where both actors collaborate in an equal partnership is yet at an early stage of development, but it is becoming an increasingly important research area in the fields of Human-Computer Interaction and Information Systems [18]. Previous research emphasizes the importance of considering design issues for Human-AI collaboration. Among those are the design of explainable and transparent AI agents, and the design of a common workplace with consideration of tasks and roles [24].

10.2.2 Task allocation in Human-AI Collaboration

There is a myriad of challenges to consider when designing human-AI systems for collaboration purposes. One of the fundamental issues hereby is the task allocation between humans and AI [25]. The labour division aims to strengthen the human-centered design of the human-AI interaction and to sustainably relieve and support employees in their work

[26]. To address this issue, several authors argue for the allocation of tasks according to corresponding strengths. In this regard, it would make sense to allocate routine tasks to AI, while humans focus on creative non-routine tasks requiring out-of-the-box thinking [5, 27, 12]. However, this would only be one dimension of the solution, as other influencing factors need further consideration, such as human preferences for certain tasks and readiness to delegate tasks to AI. The task allocation should furthermore be tailored to the qualifications and skills of the users- both in terms of the content and the form of interaction (e.g., reaction times or workload). Moreover, [26] points out that a high degree of agency and situation control can prevent dissatisfaction among employees. When it comes to the allocation content, an empirical study about the use of virtual collaborators [28] shows concrete examples of tasks that, according to study participants, should and can be done by a virtual AI assistant. Those are, for example, conducting systematic literature reviews, organizing appointments, and assisting in writing a mid-term paper.

10.2.3 Task Delegability

Delegation implies entrusting authority to another person or, since last century [29], to computers. The goal of it is to lay off a part of the workload and distribute labour in efficient ways [30]. According to Milewski [30], several factors affect delegation, namely: Perceived delegate performance, trustworthiness, similarity to the delegator, confidence and experience of the delegator, task type and load, and rewards and incentives.

Delegation to computer agents necessitates a rational allocation of tasks, as well as a rational decision to delegate a task to a given agent. The latter is affected by several factors, such as the ability of the agent to perform the task, the cost-benefit ratio of delegating, and trust in the agent's capability [31]. Additionally, the multitasking capabilities of the delegator directly affect the delegating decision [32]. In knowledge work professions, such as the one of a manager, delegation decision is furthermore highly impacted by the level of the importance of the task. In the case of delegating strategic decisions, it has been shown that humans are less likely to delegate to AI compared to other humans [33]. The authors explain this reluctance to delegate to AI by emphasizing peoples' need for proximity to individuals they see as supportive, hence the lack of such perception in the context of AI complicates the delegability. Milewski and Lewis [34] argue that delegation decision also pertains to the design of intelligent agent user interfaces. Hereby, they mention several design issues that need to be addressed to increase the likeliness of delegating to intelligent assistants. In addition to previously mentioned factors, delegation also requires sophisticated, interactive communication and performance controls.

It is noteworthy that human interaction with AI comes in different types and levels, ranging from giving AI full autonomy over making decisions to not involving AI at all. In between the two extremes, there are several levels with different engagement from both sides, e.g., AI gives suggestions, or asks for feedback from humans [35].

The task delegability framework

Lubars and Tan [17] developed a framework of task delegability (see Figure 1) consisting of four high-level factors, namely: motivation, difficulty, risk, and trust. The first three factors were chosen because they affect the decision to perform a task, and trust is chosen as the most representative factor that captures the interaction between humans and AI. The framework was built on prior literature on function allocation, mixed-initiative systems, and trust and reliance on machines [36, 37, 35]. Function allocation investigates the appropriate ways to allocate tasks based on the respective strengths of human and machine [38, 35, 39]. The mixed-initiative systems are based on human-in-the-loop and computer-in-the-loop designs. The human-in-the-loop approach is based upon the interaction of humans and machines, wherein humans supervise the algorithms, leading to the improvement of the algorithmic learning as well as to more efficient performance [40, 3]. In contrast, in computer-in-the-loop approaches, humans benefit from AI support through, e.g., interpretation of large amounts of data. Furthermore, this approach improves system transparency by, e.g., reporting the uncertainty of predictions [41]. With respect to trust and reliance on machines, it is argued that trust is what drives the acceptance of AI assistance, acting as a substitute for rational decisions when there is a lack of understandability of AI output [37].

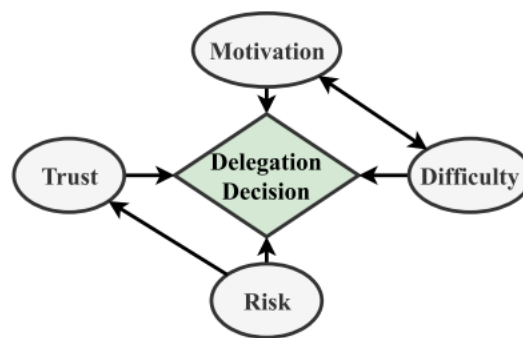


Figure 1. Framework of task delegability by Lubars and Tan [17]

Each of the framework factors consists of several components. According to the framework authors [17], they are explained as follows: Motivation is comprised of intrinsic motivation, goals, and utility, last referring to the value of the task from a cost-benefit perspective. Difficulty is framed "as the interplay between task requirements and the ability of a person to meet those requirements." Five components reflect difficulty: social skills, creativity, effort required, expertise required, and perceived human ability. Real-world tasks involve uncertainty and risk in accomplishing the task, thus they hypothetically drive delegation decision. Risk is defined through accountability for the task outcome, uncertainty, i.e., probability of errors, and impact or cost of those errors. Finally, trust comprises three components related to AI: perceived machine ability, its interpretability, and perceived value alignment.

Table 1. Overview of the framework components

Factors	Components
Motivation	Intrinsic motivation, goals, utility
Difficulty	Social skills, creativity, effort required, expertise required, human ability
Risk	Accountability, uncertainty, impact
Trust	Machine ability, interpretability, value alignment

10.3 Method

To address the research questions stated in the Introduction, an empirical study employing an online questionnaire was performed. 93 participants took part in the study. The data was analyzed with statistical methods appropriate for each research question, explained in detail below.

10.3.1 Questionnaire

The questionnaire employed by Lubars and Tan [17] was adapted to examine the preferences of task delegability to AI in a knowledge work context. Independent variables (motivation, difficulty, risk, and trust), as well as the dependent variable (delegability), were measured with the same items as in [17]. However, the questionnaire differed a lot in the tasks examined. While Lubars and Tao used 100 tasks from different areas, each answered by 4-5 respondents, we developed a list of 14 tasks, specifically tailored to our respondent group of postgraduate researchers, which are considered exemplary knowledge workers [42].

In search for relevant tasks, we used International Standard Classification of Occupations [43], from which we abstracted the tasks from various research fields to suit all respondents. Furthermore, we extracted typical research associate tasks from the German Federal Employment Agency database [44]. The tasks are diverse in terms of the four framework factors, i.e., they differ in difficulty, impact, and attractiveness for the task performer. Table 2 shows the items used to examine the four framework factors. Items in the survey were answered on a 5-point Likert scale, ranging from strongly disagree (1) to strongly agree (5). There were 14 items per task.

Finally, after answering these 14 items, participants had to decide the level of AI support they would prefer for the respective task. Four levels were offered as a possibility:

- Full AI automation: decisions and actions are made automatically by the AI once the task is assigned; you do nothing.
- The AI leads and the human assists: the AI performs the task, but asks you for suggestions / confirmation when appropriate.

- The human leads and the AI assists: you do the task mostly on your own, but the AI offers recommendations or help when appropriate (e.g., you get stuck or AI sees possible mistakes).
- No AI assistance: you do the task completely on your own.

The second and third level represent previously introduced computer-in-the-loop and human-in-the-loop approaches, respectively. Completing the questionnaire took between 5-15 minutes, depending on the experience with the tasks and individual speed of respondents. The survey allowed a pre-selection of tasks that participants had performed during their postgraduate studies, resulting in an uneven number of answers for different tasks.

10.3.2 Subjects

101 subjects took part in the study on a voluntary basis. 7 subjects were removed from the dataset due to missing data (answering "No" for every task) and one was removed after analyzing the data for outliers. The removal of the outlier did not affect the analysis results. In total, data from 93 subjects was analyzed (female = 56, male = 36, other = 1). Participants are aged 18-65, 91 % being 18-35 years old. 49 % are PhD students, 38 % master students, and the rest 13 % chose "Other" as their occupation. Participants are employed in diverse fields of research, most commonly psychology, information systems, and computer science. The target group was students with postgraduate educational level, recruited through social media and online survey-taking platforms. Postgraduate students and research associates were chosen as a target group due to their experience with various knowledge-intensive tasks [42] as well as due to the personal involvement and different levels of responsibility they might feel for certain tasks. They are a particularly well-suited group of knowledge workers for our study goals due to the breadth of tasks they perform in their daily work-life ranging from routine to complex tasks. Moreover, IT systems where AI could be integrated are often already in use in their work processes. In addition, they could particularly benefit from AI augmentation due to the high workload and work processes profused with repetitive tasks.

Table 2. Questionnaire items

Difficulty	This task requires social skills to complete. This task requires creativity to complete. This task requires a great deal of time or effort to complete. It takes significant training or expertise to be qualified for this task. I am confident in my own abilities to complete this task.
Risk	In the case of mistakes or failure on this task, someone needs to be held accountable. A complex or unpredictable environment/situation is likely to cause this task to fail. Failure would result in a substantial negative impact on my life or the lives of others.
Motivation	I would feel motivated to perform this task, even without needing to; for example, it is fun, interesting, or meaningful to me.

	I am interested in learning how to master this task, not just in the completion of the task. I consider this task especially valuable or important; I would feel committed to completing this task because of the value it adds to my life or the lives of others.
Trust	I trust the AI agent's ability to reliably complete the task. Understanding the reasons behind the AI agent's actions is important for me to trust the AI agent on this task (e.g., explanations are necessary). I trust the AI agent's actions to protect my interests and align with my values for this task.

10.3.3 Data Analysis

To investigate whether the given framework of task delegability fits the data collected in this study, we performed a confirmatory factor analysis (CFA). CFA model was fit in R programming language [45], using the *laavan* library [46]. The second research question concerning the impact of individual factors and components on the delegation decision was checked through linear regression. Assumptions of independence and normality of residuals as well as of homoscedasticity were analyzed via the residuals statistics and visual exploration of the residuals histogram as well as a scatter plot of observed versus predicted residual values. All assumptions were met. Analysis of covariance (ANCOVA) was employed to check whether the type of task affects delegability. To confirm that the assumptions for ANCOVA have been met, Levene's test for equality of variances was performed ($p = 0.130$), in addition to a visual check of a Q-Q plot of standardized residuals versus theoretical quantiles.

10.4 Results

This study investigates the fitness of the framework proposed by [17] to the data collected on a sample of postgraduate students and research associates, representative of knowledge workers. Furthermore, we analyze the impact of the framework factors on the delegability decision, as well as whether this decision depends on the task. In further text, we present our findings corresponding to the three research questions introduced earlier.

10.4.1 Confirmatory Factor Analysis

The fitness of the model to the sample data was analyzed with confirmatory factor analysis. To examine whether the factor structure can be replicated, several model fit indices and their criteria were used with the given dataset: Tucker-Lewis Index (TLI), Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), Standardised Root Mean square Residual (SRMR) and Modification Indices (MI). The data included 14 items to test delegability, split into four latent factors. All 14 items are scored on a scale from 1 to 5 and treated as continuous variables in the analysis. Exploratory data analysis revealed minor deviations from normality in the distributions of few components. The model was

fit using lavaan version 0.6.8 [46] in R version 4.1.0 [45]. We used maximum likelihood estimation and standardized the latent factors, allowing free estimation of all factor loadings. After analyzing the full model fit (all 4 factors, not taking into regard the covariance between variables), we decided to account for covariances between accountability and uncertainty, (MI = 6.99), social skills and creativity (MI = 8.49), and intrinsic motivation and learning motivation (MI = 11.2), since these displayed largest modification indices in the residual covariance matrix. The model fit was acceptable but not excellent, with a CFI of .95, TLI of .94, RMSEA of .06, 90% CI (.017, .090), and SRMR of .075. Further models were fit and analyzed, to investigate whether a model with a different number of factors and different variable constellation fits the data better. Led by the modification indices in the cross-loadings matrix, we decided to exclude the risk factor, exclude social skills and creativity from the difficulty factor, and include intrinsic motivation in the trust factor. This led to a model with significantly better fit than the previous model. The model fit was good with a CFI of 1.00, TLI of 1.01, RMSEA of .00, 90% CI (.00, .07), and SRMR of .055.

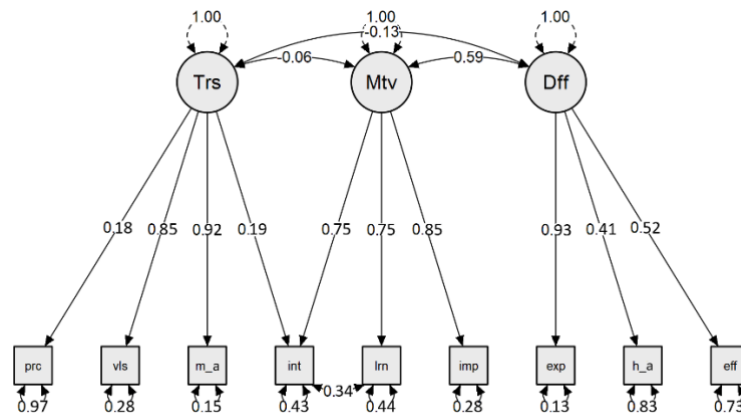


Figure 2. CFA model

Standardized parameter estimates show that machine ability and values have the highest loadings on the trust factor ($\lambda > .85$). All components of motivation have high loadings ($\lambda > .73$) on this factor. Expertise has the highest loading on difficulty ($\lambda = .94$), while effort and human ability have moderate loadings on this factor. All loadings are statistically significant ($p < 0.05$), except for the loading of the process (interpretability) on trust.

10.4.2 Linear Regression

Regarding the impact of individual factors and components on the delegation decision, the results of linear regression show statistically significant model fit ($R^2 = 0.460$, $p < 0.001$). The model thus explains 46 % of variance. Subsequent analysis of variance (ANOVA) shows furthermore that the model differs from null model ($F = 18.491$, $p < 0.001$). Trust is the factor that best predicts the delegability decision ($\beta = 0.652$, $t = 8.137$, $p < 0.001$). The regression line (Fig. 3) depicts a positive relationship between trust and delegability, i.e., the higher the trust in AI, the higher the probability to delegate a task to AI.

Motivation, difficulty, and risk have a much smaller effect on delegability compared to trust (see Table 3). Furthermore, it appears that the correlation between the 3 factors and delegability is negative, i.e. the less motivation, risk, or difficulty, the higher delegability. However, these relationships did not appear statistically significant ($p > 0.05$), meaning that there might be a trend in this particular data, but this could be due to the chance and not a reflection of the real state in population. With regard to the individual components of the factors, a component of trust, machine ability, is the best predictor of delegability decision ($\beta = 0.654$, $t = 5.406$, $p < 0.001$).

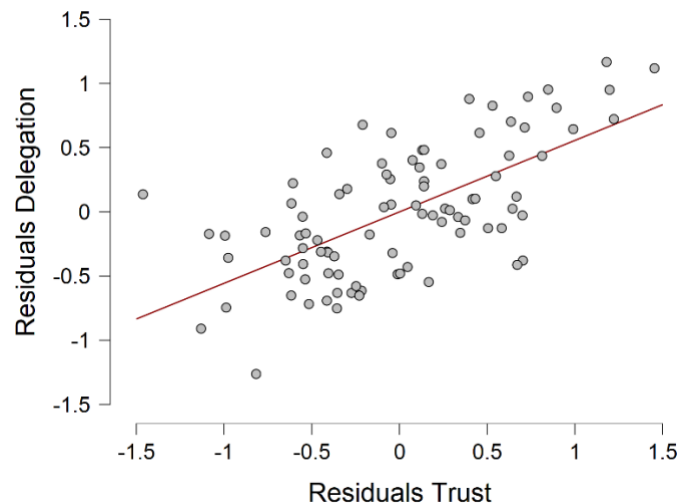


Figure 3. Relationship between trust and delegability

Table 3. Regression Coefficients

	Model	Unstandardized	Standard Error	Standardized	t	p
H ₀	(Intercept)	2.441	0.055		44.620	< .001
H ₁	(Intercept)	1.234	0.448		2.753	0.007
	Difficulty	-0.094	0.106	-0.081	-0.891	0.375
	Motivation	-0.075	0.064	-0.113	-1.158	0.250
	Risk	-0.049	0.060	-0.071	-0.809	0.421
	Trust	0.568	0.070	0.652	8.137	< .001

10.4.3 Analysis of covariance

A one-way ANCOVA was conducted to determine a statistically significant difference between the tasks on delegability decision, controlling for trust, motivation, risk, and difficulty. The model was estimated using the Type III sum of squares approach [47], due to uneven sample sizes. Based on the results, there is a significant effect of task on delegability after controlling for the framework factors ($F = 1.860$, $p < 0.05$). Task covaries with trust and motivation ($p < 0.001$). The effect size estimate ($\eta^2 = 0.031$) depicts a small effect of task on delegability. The effect of the covariance with trust appears higher ($\eta^2 = 0.244$). Without controlling for the framework factors, the effect size of task on delegability is high ($\eta^2 = 0.203$, $F = 11.969$, $p < 0.001$).

Table 4. Effects of task on delegability

Task	Mean	SD	N
Code and verify data in accordance with specified research protocol	3.150	0.662	40
Maintain accurate records of participant data	2.838	0.898	37
Assist in laboratory analysis, quality control, or data management	2.778	0.637	36
Conduct literature reviews	2.676	0.685	74
Request or acquire equipment or supplies necessary for the project	2.667	0.963	24
Review and edit data to ensure completeness and accuracy of information	2.652	0.714	23
Develop assessment and evaluation tools	2.455	0.800	22
Identify and compile lists of potential research subjects	2.441	0.927	34
Set up, calibrate and maintain laboratory and/or field research equipment	2.286	1.231	21
Prepare materials for submission to granting agencies and foundations	2.238	0.889	21
Derive research questions	2.137	0.732	73
Prepare data collection	2.103	0.765	58
Write and contribute to publications	1.930	0.704	43
Supervise undergraduate students	1.593	0.694	27

N = number of participants that gave answers for the respective task.

10.5 Discussion

In this study, a quantitative analysis was used to examine a framework of task delegability to AI in a knowledge work context, i.e., to examine human preferences to involve AI into their work either to a certain level or by delegating the complete task.

10.5.1 Theoretical implications

Our findings contribute to the literature on task allocation and delegability to AI, especially in the context of Human-AI collaboration in knowledge work. We found that the framework with four factors does not provide the best fit in this specific knowledge work context and that the model with a reduced number of factors best represents the answers of the study participants. Thus, a framework of task delegability can be adjusted to consist of trust, motivation, and difficulty. To discuss the model fit of CFA, the criteria of the various model fit indices are considered. It has been argued that RMSEA values less than 0.05 are good, values between 0.05 and 0.08 are acceptable, values between 0.08 and 0.1 are marginal, and values greater than 0.1 are poor [48]. Hence, the RMSEA value of 0.000 in this sample indicates a good fit. The CFI and TLI values are close to 1.00, which depicts a very good fit [49]. The factor of risk is likely to not have contributed to the model due to several possible reasons. First, the items measuring this factor such as "Failure would result in a substantial negative impact on my life or the lives of others" might not reflect what is considered to be risky in the context of knowledge work, especially in our subgroup of

postgraduate students and research associates. It is more likely that not a single task, but a set of events could cause a failure in this context. Second, the tasks examined in this study are likely not indicating of tasks related to high risk. In support of this assumption, Lubars and Tan [17] closely examined tasks having high impact and accountability, such as planning medical treatment for cancer, and showed that this type of tasks is associated with lesser delegability. Consistent with their results, the process component of trust does not significantly contribute to this factor. The other two components, especially trust in machine ability, display high contribution to the factor. This is in line with the previous research about the importance of trust for human-AI collaboration [50, 51, 52]. Task importance, intrinsic motivation, and goals approximately equally contribute to the motivation factor, all three displaying high contribution. Motivation negatively correlates to delegability decision, meaning that humans rather delegate demotivating tasks. Similarly to motivation, difficulty also negatively correlates to delegation, which is consistent with the results of Lubars and Tan [17].

Moreover, results of linear regression further strengthen the implication of the influence of trust on attitude toward AI in collaborative scenarios, such as task delegation, *inter alia*. The higher the trust, especially trust in machine ability, the higher is the preparedness to delegate a task to AI. Other framework components do not predict delegation on a statistically significant level, which is partly in line with the results of Lubars and Tan [17]. However, this conflicts with previous research. For example, several publications suggest the importance of risk for delegation [53, 54, 55]. More research is needed to reconcile these differences.

The variance explained by the linear regression model (46 %) suggests that there might be other factors not considered by the current framework, that might add to the predictability of delegation decisions. To help shed more light on the topic, future research should study a different population of knowledge workers and look at an extended set of variables.

10.5.2 Design implications

Participants seem to be equally eager to either act as a human-in-the-loop or be supported by AI in a computer-in-the-loop approach. Across all tasks and participants, 37.15 % chose the computer-in-the-loop approach, while 37.34 % chose the human-in-the-loop approach. In 15.38 % of cases "No delegation" was chosen, while full automation was preferred in only 10.13 % of cases. These findings could support the research stream of hybrid intelligence, which claims that combined approaches of human-in-the-loop and computer-in-the-loop will lead to a much better work performance [8, 5]. Strong preferences for these approaches among knowledge workers should be accounted for in the digitalization of the workplace.

The findings further contribute to the research stream that deals with the design of trustworthy AI, e.g., ([56, 57, 58, 59, 60]) by emphasizing the relevance of trust in task delegability.

Positive association between tasks and delegability implies another design possibility for future workplaces (see Table 4). Tasks such as "Conduct literature reviews", "Maintain accurate records of participant data", and "Assist in laboratory analysis, quality control, or data management" could profit from the computer-in-the-loop approaches. Dealing with large amounts of data and being cumbersome and time-consuming is common to these tasks. Tasks where AI support is welcome are, e.g., "Write and contribute to publications", "Prepare data collection", and "Derive research questions". For most of these tasks, common sense is required, thus it is more appropriate to leverage the human-in-the-loop approach in this case. Tasks predominately basing on human contact, such as "Supervise undergraduate students" seem to have the least benefit from employing AI support. This might be due to the fact that, under current circumstances, participants could not imagine in which way AI could assist them with this task, or due to ethical concerns or strong identification with these tasks. However, with the rapid development of smart assistants [61], this type of task might soon be subject to augmentation with AI.

Finally, understanding delegation decisions of knowledge workers, particularly researchers as well as factors portraying these represents a step further in deriving design knowledge for Human-AI collaboration in knowledge work.

10.5.3 Limitations and future research

There are a few limitations in our study that should be addressed by future research endeavors on this topic. First, the sample size is relatively small, thus, future studies should increase the sample size to obtain higher statistical power. The suggested minimum sample size for CFA is 100-150 [62, 63], or at least 10 cases per measuring item (140 in our case) [64].

Next, we had unequal sample sizes per task, due to allowing participants to choose for which tasks to provide answers, based on their experience. While this is a disadvantage on the one hand, on the other hand, making questions for all tasks mandatory runs a risk of receiving random answers, or many neutral answers, simply because participants would not know how to honestly answer about something they had no previous experience with.

Further, exploring a wider range of tasks could have led to different insights, especially in terms of the risk factor. Future research could profit from investigating the framework on tasks involving, e.g., making impactful strategic decisions.

Future research could explore the framework in different knowledge work fields and gain valuable insights about its generalizability and different nuances of its applicability. Finally, research on design principles of Human-AI collaboration could account for this framework when investigating different design possibilities.

Acknowledgements. This research was funded by the German Federal Ministry of Education and Research (BMBF), in the context of the project HyMeKI (01IS20057).

10.6 References

1. McCarthy, J., What is Artificial Intelligence? 2007.
2. West, D. M., *The future of work: Robots, AI, and automation*. Brookings Institution Press, 2018.
3. Holzinger, A., Interactive Machine Learning (iML) *Informatik-Spektrum*, vol. 39, no. 1, pp. 64–68, 2016.
4. Ashktorab, Z., Liao, Q. V., Dugan, C., Johnson, J., Pan, Q., Zhang, W., Kumaravel, S., and Campbell, M., Human-AI Collaboration in a Cooperative Game Setting *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW2, pp. 1–20, 2020.
5. Dellermann, D., Ebel, P., Söllner, M., and Leimeister, J. M., Hybrid Intelligence *Business & Information Systems Engineering*, vol. 61, no. 5, pp. 637–643, 2019.
6. Bittner, E. A. C., Oeste-Reiß, S., Ebel, P. A., and Söllner, M., Mensch-Maschine-Kollaboration: Grundlagen, Gestaltungsherausforderungen und Potenziale für verschiedene Anwendungsdomänen *HMD Praxis der Wirtschaftsinformatik*, vol. 56, no. 1, pp. 34–49, 2019.
7. Wiethof, C. and Bittner, E. A. C., Hybrid Intelligence - Combining the Human in the Loop with the Computer in the Loop: A Systematic Literature Review in *ICIS International Conference of Information Systems*, (Austin), 2021. (in press).
8. Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., and Ebel, P., The future of human-AI collaboration: a taxonomy of design knowledge for hybrid intelligence systems in *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.
9. Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., and Söllner, M., AI-Based Digital Assistants *Business & Information Systems Engineering*, vol. 61, no. 4, pp. 535–544, 2019.
10. Drucker, P., *Landmarks of tomorrow*. New York: Harper, 1959.
11. De Vreede, G. and Briggs, R., Collaboration engineering: Designing repeatable processes for high-value collaborative tasks *Proceedings of the Annual Hawaii International Conference on System Sciences*, p. 17, 2005.
12. Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G.-J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., and Söllner, M., Machines as teammates: A research agenda on AI in team collaboration *Information Management*, vol. 57, no. 2, pp. 103–174, 2020.
13. Oeste-Reiß, S., Bittner, E., Cvetkovic, I., Günther, A., Leimeister, J. M., Memmert, L., Ott, A., Sick, B., and Wolter, K., Hybride Wissensarbeit *Informatik-Spektrum*, 2021.
14. Daugherty, P. and Wilson, H. J. R., *Human + Machine: Reimagining Work in the Age of AI*. Audible Studios on Brilliance Audio, 2018.
15. Sowa, K., Przegalinska, A., and Ciechanowski, L., Cobots in knowledge work

- Journal of Business Research*, vol. 125, pp. 135–142, 2021.
16. Jarrahi, M. H., Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making *Business Horizons*, vol. 61, no. 4, pp. 577–586, 2018.
 17. Lubars, B. and Tan, C., Ask not what AI can do, but what AI should do: Towards a framework of task delegability in *Advances in Neural Information Processing Systems*, vol. 32, 2019.
 18. Zhang, R., McNeese, N. J., Freeman, G., and Musick, G., An Ideal Human: Expectations of AI Teammates in Human-AI Teaming *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW3, pp. 1–25, 2021.
 19. Wang, D., Churchill, E., Maes, P., Fan, X., Shneiderman, B., Shi, Y., and Wang, Q., From Human-Human Collaboration to Human-AI Collaboration: Designing AI Systems That Can Work Together with People in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, CHI EA '20, (New York, NY, USA), p. 1–6, Association for Computing Machinery, 2020.*
 20. Poser, M., Singh, S., and Bittner, E., Hybrid Service Recovery: Design for Seamless Inquiry Handovers between Conversational Agents and Human Service Agents in *Proceedings of the 54th Hawaii International Conference on System Sciences* (Bui, T., ed.), Proceedings of the Annual Hawaii International Conference on System Sciences, Hawaii International Conference on System Sciences, 2021.
 21. Winkler, R., Neuweiler, M. L., Bittner, E., and Söllner, M., Hey Alexa, Please Help Us Solve This Problem! How Interactions with Smart Personal Assistants Improve Group Performance in *ICIS International Conference of Information Systems*, (Munich), ACM Digital, 2019.
 22. Wiethof, C., Tavanapour, N., and Bittner, E., Implementing an Intelligent Collaborative Agent as Teammate in Collaborative Writing: toward a Synergy of Humans and AI in *Proceedings of the 54th Hawaii International Conference on System Sciences* (Bui, T., ed.), Proceedings of the Annual Hawaii International Conference on System Sciences, Hawaii International Conference on System Sciences, 2021.
 23. McNeese, N. J., Schelble, B. G., Canonico, L. B., and Demir, M., Who/What is My Teammate? Team Composition Considerations in Human-AI Teaming *arXiv preprint arXiv:2105.11000*, 2021.
 24. Seeber, I., Waizenegger, L., Seidel, S., Morana, S., Benbasat, I., and Lowry, P., Collaborating with technology-based autonomous agents: Issues and research opportunities *Internet Research*, vol. 30, no. 1, p. 1–18, 2020.
 25. Asatiani, A., Penttinen, E., Rinta-Kahila, T., and Salovaara, A., Organizational Implementation of Intelligent Automation as Distributed Cognition: Six Recommendations for Managers *ICIS 2019 Proceedings*, 2019.
 26. Huchler, N., Adolph, L., Andre', E., Bauer, W., Bender, N., Müller, N., Neuburger, R., Peissner, M., Steil, J., Stowasser, S., and Suchy, O., Kriterien für

- die Mensch-Maschine-Interaktion bei KI: Ansätze für die menschengerechte Gestaltung in der Arbeitswelt whitepaper, Lernende Systeme – Die Plattform für Künstliche Intelligenz, München, June 2020.
27. Traumer, F., Oeste-Reiß, S., and Leimeister, J. M., Towards a Future Reallocation of Work between Humans and Machines Taxonomy of Tasks and Interaction Types in the Context of Machine Learning *SSRN Electronic Journal*, 2017.
 28. Siemon, D., Strohmann, T., and Robra-Bissantz, S., The Virtual Collaborator - A Definition and Research Agenda *International Journal of e-Collaboration*, vol. 14, no. 4, pp. 24–43, 2018.
 29. Bradshaw, J., An Introduction to Software Agents in *Software Agents* (Bradshaw, J., ed.), Cambridge, Mass: MIT Press, 1997.
 30. Milewski, A. E. and Lewis, S., When people delegate *Unpublished manuscript prepared for J. Bradshaw, The Handbook of Agent Technology*, 1997.
 31. Castelfranchi, C. and Falcone, R., Towards a theory of delegation for agent-based systems *Robotics and Autonomous Systems*, vol. 24, no. 3, pp. 141–157, 1998. Multi-Agent Rationality.
 32. Lee, D. W., Fitzick, D. W., and Bass, E. J., Simulating Human Performance of Task Sharing: Modeling Task Delay and Delegation of Authority *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 61, no. 1, pp. 786–790, 2017.
 33. Leyer, M. and Schneider, S., Me, You or AI? How do we Feel about Delegation in *ECIS*, 2019.
 34. Milewski, A. E. and Lewis, S. H., Delegating to software agents *International Journal of Human-Computer Studies*, vol. 46, no. 4, pp. 485–500, 1997.
 35. Parasuraman, R., Sheridan, T., and Wickens, C., A model for types and levels of human interaction with automation *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 30, no. 3, pp. 286–297, 2000.
 36. Horvitz, E., Principles of Mixed-Initiative User Interfaces in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, (New York, NY, USA), p. 159–166, Association for Computing Machinery, 1999.
 37. Lee, J. D. and See, K. A., Trust in Automation: Designing for Appropriate Reliance *Human Factors*, vol. 46, no. 1, pp. 50–80, 2004.
 38. Fitts, P., ed., *Human engineering for an effective air-navigation and traffic-control system*. Oxford, England: National Research Council, Div. of, 1951.
 39. Price, H. E., The allocation of functions in systems *Human Factors*, vol. 27, no. 1, p. 33–45, 1985.
 40. Martinez, M., Nadj, M., and Maedche, A., Towards An Integrative Theoretical Framework Of Interactive Machine Learning Systems in *Proceedings of the 27th European Conference on Information Systems*, (Sweden), Stockholm-Uppsala, 2019.
 41. Abdel-Karim, B., Pfeuffer, N., Rohde, G., and Hinz, O., How and What Can

- Humans Learn from Being in the Loop? *KI - Künstliche Intelligenz*, vol. 34, p. 199–207, 2020.
42. Harney, B., Monks, K., Alexopoulos, A., Buckley, F., and Hogan, T., University research scientists as knowledge workers: contract status and employment opportunities *The International Journal of Human Resource Management*, vol. 25, no. 16, pp. 2219–2233, 2014.
 43. International Labour Office, and International Labour Organization, *International Standard Classification of Occupations: ISCO-08*. Geneva: International Labour Office, 2012. OCLC: 794974321.
 44. BUNDESAGENTUR FÜR ARBEIT (BA), Wissenschaftliche/r Mitarbeiter/in (Hochschule) 2021. (Accessed on: May 05, 2021 [Online]). Available: <https://bit.ly/3iIzb3f>.
 45. R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.
 46. Rosseel, Y., lavaan: An R Package for Structural Equation Modeling *Journal of Statistical Software*, vol. 48, no. 2, pp. 1–36, 2012.
 47. Campbell, D. and Kenny, D., *A primer on regression artifacts*. New York: Guilford Publications, 1999.
 48. Fabrigar, L., Wegener, D., MacCallum, R., and Strahan, E., Evaluating the use of exploratory factor analysis in psychological research *Psychological Methods*, vol. 4, no. 3, p. 272–299, 1999.
 49. Bentler, P. M., Comparative fit indexes in structural models *Psychological bulletin*, vol. 107, no. 2, p. 238–246, 1990.
 50. Elson, J., Derrick, D., and Merino, L., An Empirical Study Exploring Difference in Trust of Perceived Human and Intelligent System Partners in *Proceedings of the 54th Hawaii International Conference on System Sciences — 2021*.
 51. Schaefer, K., Chen, J., Szalma, J., and Hancock, P., A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems *Human factors*, vol. 58, no. 3, p. 377–400, 2016.
 52. Elson, J., Derrick, D., and Ligon, G., Examining trust and reliance in collaborations between humans and automated agents in *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.
 53. Taddeo, M. and Floridi, L., How AI can be a force for good *Science*, vol. 361, no. 6404, pp. 751–752, 2018.
 54. Di Nucci, E., Should we be afraid of medical AI? *Journal of Medical Ethics*, vol. 45, no. 8, pp. 556–558, 2019.
 55. Roider, A., Delegation, Risk, and Project Scope *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift für die gesamte Staatswissenschaft*, vol. 165, no. 2, pp. 193–209, 2009.
 56. Zierau, N., Hausch, M., Bruhin, O., and Soellner, M., Towards Developing Trust-

- Supporting Design Features for AI-Based Chatbots in Customer Service *ICIS 2020 Proceedings*, 2020.
57. Rossi, F. BUILDING TRUST IN ARTIFICIAL INTELLIGENCE *Journal of International Affairs*, vol. 72, no. 1, pp. 127–134, 2018.
 58. Smith, C. J., Designing trustworthy AI: A human-machine teaming framework to guide development *arXiv preprint arXiv:1910.03515*, 2019.
 59. Chattopadhyay, A., Ali, A., and Thaxton, D., Assessing the Alignment of Social Robots with Trustworthy AI Design Guidelines: A Preliminary Research Study (New York, NY, USA), Association for Computing Machinery, 2021.
 60. Kaur, D., Uslu, S., and Durrezi, A., Requirements for Trustworthy Artificial Intelligence – A Review in *Advances in Networked-Based Information Systems* (Barolli, L., Li, K. F., Enokido, T., and Takizawa, M., eds.), (Cham), pp. 105–115, Springer International Publishing, 2021.
 61. Winkler, R. and Roos, J., Bringing AI into the Classroom: Designing Smart Personal Assistants as Learning Tutors *ICIS 2019 Proceedings*, 2019.
 62. Anderson, J. and Gerbing, D., Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach *Psychological Bulletin*, vol. 103, p. 411, 1988.
 63. Ding, L., Velicer, W., and Harlow, L., Effects of Estimation Methods, Number of Indicators per Factor, and Improper Solutions on Structural Equation Modeling Fit Indices *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 2, p. 119–143, 1995.
 64. Nunnally, J. and Bernstein, I., *Psychometric Theory*, vol. 226. New York, NY: McGraw-Hill, 1967.

11 What should AI know? Information Disclosure in Human-AI Collaboration.

Cvetkovic, I., Oeste-Reiß, S., Lehmann-Willenbrock, N., & Bittner, E. (2022). What should AI know? Information Disclosure in Human-AI Collaboration. In *Proceedings of the 30th European Conference on Information Systems (ECIS)*, Timișoara, Romania.

Abstract

AI-assisted Design Thinking shows great potential for supporting collaborative creative work. To foster creative thinking processes within teams with individualized suggestions, AI has to rely on data provided by the teams. As a prerequisite, team members need to weigh their disclosure preferences against the potential benefits of AI when disclosing information. To shed light on these decisions, we identify relevant information such as emotional states or discussion arguments that design thinking teams could provide to AI to enjoy the benefits of its support. Using the privacy calculus as theoretical lens, we draft a research design to analyze user preferences for disclosing different information relevant to the service bundles that AI provides for respective information. We make explorative contributions to the body of knowledge in terms of AI use and its corresponding information disclosure. The findings are relevant for practice as they guide the design of AI that fosters information disclosure.

Keywords: Human-AI Collaboration, Information Disclosure, Design Thinking, UI Design.

11.1 Introduction

To remain competitive in the fast-changing technology landscape, characterized by digitization and globalization, companies must rely on novel ideas to develop innovative and profitable products, processes, and services. A promising approach in this context is Design thinking (DT) (Meinel and Leifer, 2012). It is characterized as a problem-solving approach, based on collaboration, user-centered design, and creativity (Schallmo, 2017). However, DT can incur challenges in terms of tasks, team interaction, and method specific facilitation (Bittner et al., 2021), such as an inappropriate division of tasks, ineffective selection of methods, and destructive workshop dynamics (Seeber et al., 2020). Fast advances in Artificial Intelligence (AI) provide promising potential to cope with these challenges. AI is the science and engineering behind creating intelligent machines, particularly computer programs, which try to grasp and, to some extent, imitate human intelligence (McCarthy, 2007). Recent research outlines the possibility of supporting DT approaches with AI, for instance, by means of virtual collaborators (Siemon and Strohmman, 2021) that support human teams and facilitators in specific tasks, such as

persona building (Lembcke et al., 2020). This indicates that Human-AI Collaboration in DT is a promising research avenue.

Human-AI Collaboration, i.e., Hybrid Intelligence (Dellermann et al., 2019b) is a flourishing and important research area (Bittner et al., 2019b; Dellermann et al., 2019a; Seeber et al., 2020). It combines the complementary strengths of humans and AI to achieve better performance than humans or AI could achieve alone. Some examples of such hybrid systems can be seen in decision making practices, where AI recommends actions to human experts who are responsible for final decisions (Bansal et al., 2021). In the course of this, a benefit of AI stems from analyzing large amounts of data and giving recommendations based on insights resulting from these analyses (Cranor, 2008). The human verifies AI recommendations, and either chooses the most appropriate solution or gives feedback to the system, from which it can consequently learn and improve upon. Such feedback can be used to retrain the models to be reused in the future, or the models can be updated in real-time and output improved accordingly (Holzinger, 2016). In any case, AI learns and improves based on data (Russell and Norvig, 1995). In Human-AI collaborative scenarios, AI can adopt the role of a teammate (Seeber et al., 2020) or a co-facilitator (Bittner and Shoury, 2019). For example, in DT approaches it would make more sense if AI made use of the data being collected during the session and present it back in a structured manner, thereby promoting creative work. AI could collect different types of information, such as emotional and cognitive states or negative utterances to moderate team dynamics. Specifically, AI could use social signals from the team to capture bad dynamics and prevent conflicts (Elshan et al., 2022).

However, problems can arise in regards to the disclosure of social signals as input for AI. Experience has shown that AI-based detection of social signals such as emotion recognition are followed by low disclosure of such information (Bailenson et al., 2006). Since the nature of this technology is psychologically invasive, privacy issues must be dealt with to the satisfaction of all persons involved in the system. On the other hand, an affect recognition system could present group members with relevant information about the group's general emotional state. Members of a group, or a facilitator, could use this information to assess the strength of agreement or disagreement for specific positions (Salim and Yusoff, 2008). State-of-the-art products in this domain demonstrate how social signals can be leveraged to receive benefits in return. For example, MIT's social robot TEGA tracks children's emotions and uses this information to offer better learning assistance (Westlund et al., 2016). Therefore, AI service bundles should come with explicit benefits that human users will receive when disclosing information. Accordingly, exploratory research on user preferences and quid pro quos between disclosure risks and benefits is needed.

This work aims to identify the service bundles that an AI-based system can offer a human user in exchange for the user's information disclosure (ID) and gather insights toward such input-output service pairs. Service bundles are combined services that offer a reward

benefit for a user (Andrews et al., 2010). Important factors in human-centred AI development are human needs and preferences, as these motivate human users toward using a system. An important avenue hereby is ID (Rayo and Segal, 2010), as information creates the foundation for system recommendations (and lifelong learning systems). However, to date, little is known about human preferences to disclose information in collaborative scenarios with AI-based systems, i.e., scenarios from which they have direct, visible profit in terms of task assistance, better and faster performance, or better-quality results.

To address this research gap, this work pursues to answer the following research questions: First, we need to identify the information that humans can disclose in our specific scenario that can subsequently be used by AI. Hence, we pose the RQ 1: *Which provided input from humans can be processed by AI-based systems to support them in specific tasks in creative teamwork?* Inputs identified by addressing RQ 1 are different regarding the type of information they carry (e.g., emotional states or factual information), and the relevance this type might have to disclosure preferences. Therefore, we pose the RQ 2: *Which information and to what extent are humans willing to disclose so that AI-based systems can support them in specific tasks in creative teamwork?* Finally, in order to investigate the effect of AI service bundles to ID, we pose the RQ 3: *How should the output from AI-based systems be designed to foster information disclosure in specific tasks in creative teamwork?*

11.2 Theoretical Background

11.2.1 Design Thinking and Artificial Intelligence

DT is an approach that integrates various tools and techniques for collaborative problem solving (Liedtka, 2015). This method provides an appropriate use case scenario for research on Human-AI collaboration due to its characteristics (Schallmo, 2017), and its increasing relevance for creative knowledge work, such as innovation development (Meinel and Leifer, 2012). In DT approaches, a human facilitator guides the collaboration process, thereby moderating the team dynamics and fostering creativity inter alia (Bittner et al., 2021). DT is used in various industries such as software, healthcare, automotive, retail, and finance. Companies use DT approaches for a range of purposes such as more empathetic customer observations, concept generation, prototyping, and product development (Carlgren et al., 2014). A common DT process comprises five stages, namely empathizing, defining, ideating, prototyping, and testing (Brown and Katz, 2011).

These are moderated throughout by human facilitators. Facilitators are responsible for the appropriate support of the teams and their challenges, as well as for the management and organization of the entire DT process: measures such as team composition and stakeholder management, process and method planning, provision of materials, documentation, and overview and transfer to the implementation of the developed ideas are tasks that require a lot of time and organization (Hjalmarsson et al., 2015). Due to the mentioned characteristics of the DT method and the high workload on facilitators, this approach could particularly

profit from AI assistance (Bittner et al., 2021). Debowski et al. (2021) outline the problem areas in creativity workshops and present design principles for a virtual assistant to address these problems. Strohmman et al. (2018) demonstrate the need for AI assistance in DT and outline the requirements for such assistance in terms of conditions, characteristics, and tasks. Similarly, Bittner et al. (2021) outline the need for AI assistance in different DT areas, namely: Team Task and Process Facilitation, Team Interaction Facilitation, Facilitator Task and Process Assistance, Facilitator Interaction Assistance, and Method Specific Assistance.

Human facilitators use various types of human input to moderate DT, such as emotional and cognitive states, arguments from discussions, ideas etc. However, whenever AI is involved, the issues of ethics and data privacy arise (Manheim and Kaplan, 2018). These issues are very context-specific, i.e., some data is more sensitive, such as user behavior data, and some is more easily disclosed, e.g., demographics (Knijnenburg and Kobsa, 2013a). AI-assisted DT is in an early phase of research (Debowski et al., 2021). Research in other domains has shown that people tend to accept new technologies without worrying much about privacy if they get sufficient benefits from them (Salim and Yusoff, 2008). Furthermore, there is overwhelming evidence that people will trade personal data in exchange for services or social benefits (Pitt, 2012; Youn, 2009). However, ID also comes with perceived risks (Al-Natour et al., 2021). Notwithstanding, little is known about how people would behave with regards to ID in DT, i.e., whether they would be comfortable providing all the data to AI that they usually provide to a human facilitator. Nevertheless, the relevance of information exchange in terms of feedback and direct communication for DT is undeniable. In the proposed work, we want to tackle the team interaction space of DT, since this space most likely requires sensitive data from participants for successful teamwork facilitation.

11.2.2 Information Disclosure

ID refers to the extent to which users are willing to disclose certain private information, which is often perceived as an issue (Knijnenburg and Kobsa, 2013a). Giving users explicit control over the information they disclose is one way to address this issue. In this case, ID becomes a decision, which comes down to calculating the benefits and risks of making a certain decision (Knijnenburg and Kobsa, 2013b). This concept is also known as "privacy calculus" (Laufer and Wolfe, 1977). A common example of such a cost-benefit trade-off is ID on social media. Here, users gain social benefits in exchange for information but are also exposed to loss of privacy. The problem of privacy calculus is the intangibility of the value of privacy, its context-dependence, and uncertainty regarding privacy decisions. The latter is due to not knowing how the data will be used, i.e., missing transparency (Hirschprung et al., 2016). ID thus depends on situational and context-specific factors that directly affect the privacy calculus (Xu et al., 2008).

ID has been studied in a number of contexts such as e-commerce (Berendt et al., 2005), or social networks (Kroll and Stieglitz, 2021) through several theoretical lenses. An integrated framework of theories in privacy research (Li, 2012) lists theories most often related to this research such as social exchange theory (Thibaut and Kelley, 1959), social contract theory (Milne and Gordon, 1993), theory of reasoned action (Ajzen and Fishbein, 1980), and privacy calculus theory (Laufer and Wolfe, 1977). For example, social exchange theory posits that people decide whether to engage in a social act based on an evaluation of interpersonal rewards and costs associated with the act. Applied to information exchange this means that when perceived benefit is higher than perceived risk, ID is more likely to happen (Chang and Heo, 2014; Li et al., 2010). Social exchange also involves trust, and cannot be reduced to simple bargaining (Masaviru, 2016). Common to most of the mentioned theories is the aforementioned cost-benefit trade-off, i.e., the privacy calculus. Hence, we choose to adopt the privacy calculus theory as our kernel theoretical lens. ID can be affected by rewards, order of disclosure requests, justifications, and privacy indicators or statements. For example, justifications can include reasons for requesting information, benefits of disclosure, or appealing to social norms, i.e., showing that others decided to disclose the same information (Knijnenburg and Kobsa, 2013b). Privacy statements such as the TRUSTe seal increase the likeliness of disclosing information (Xu et al., 2009). Furthermore, increased system transparency has a positive effect on the disclosure of information (Patil and Kobsa, 2005; Patil and Lai, 2005). Users who have more choice over how, when and with whom components of their environment are shared can establish a better balance between awareness and privacy (Patil and Lai, 2005). Apart from privacy self-management, transparent privacy policies and privacy controls can further alleviate concerns about disclosure (Stutzman et al., 2011).

A way to encourage ID without restricting users' freedom of choice is digital nudging (Kroll and Stieglitz, 2021). Digital nudging represents the use of user-interface design elements to shape people's behavior in digital environments (Weinmann et al., 2016). For example, Facebook's privacy dinosaur helps navigate complex privacy options and make choices with less overload (Kroll and Stieglitz, 2021).

The introduction of AI and its relation to data collection raises privacy concerns and increases cautious behavior. Reliable AI assistance depends on user data. For example, Netflix depends on users' viewing history and movie ratings to provide suitable recommendations (Liao and Sundar, 2021). Usually, such services also offer an explanation that emphasizes the benefit of data disclosure. Intelligent personal assistants represent another form of AI often related to privacy concerns. For example, Manikonda et al. (2018) found that users who mute their microphones state privacy concerns as the main reason for this. An empirical study of ID to virtual advisors investigated a number of determinants of the disclosure, such as trust, transparency, and perceived benefits and costs. The authors found that all determinants contained in their model significantly contribute to ID (Al-Natour et al., 2021). When it comes to ID to AI in the workplace, trust in the company

plays another important role for weighing the risks and benefits of disclosure (Metzger, 2004). A way to influence the determinants of disclosure to AI is by using protection frameworks such as federated learning, which use decentralized data sharing techniques to preserve information privacy (Wei et al., 2020).

An example of sensitive data collection in the workplace that has been discontinued due to privacy reasons is Zoom's attention tracker (*Attendee attention tracking*). Such examples call for more research in the area of privacy and motivate the question: How can we reconcile the need for massive private data collection for AI and privacy concerns? The proposed study tries to tackle an aspect of this question by gaining insight into preferences for disclosure relative to benefits received in exchange.

11.3 Preliminary Results and Proposed Research Approach

DT processes could profit from AI assistance in different ways. We are analyzing literature to explore this in-depth and identify the input/output space of Human-AI collaboration in DT. The main paper, from which we extend the literature analysis is "Digital Facilitation Assistance for Collaborative, Creative Design Processes" (Bittner et al., 2021).

Table 1. Examples of DT input identified in the literature and proposed AI output for each respective input

	Human Input	AI Output
Team interaction	Body language (gestures, facial expressions) (Bittner et al., 2021)	Team mood board (own suggestion); jokes and anecdotes in response to negative mood (Strohmann et al., 2018)
	Body language (eye gaze, mimics, gestures) (Bittner et al., 2021)	Reminder to take a break, information load board (own suggestion)
	Keywords for breaking rules (Bittner et al., 2021)	Reminder of the rules of creative teamwork (Bittner et al., 2021)
	Speech share & centrality (Bittner et al., 2021)	Informing about unbalanced speech shares and a reminder of preferred equality (Bittner et al., 2021)
Team process	Results from a brainstorming session (Bittner et al., 2021)	Structured and clustered results presented in a dashboard (Bittner et al., 2021)
	Calendar information (Cranshaw et al., 2017)	Recommended schedule for next session(s) (own suggestion)

	Data from interviews/desk research (Bittner et al., 2021)	Analysis report (e.g., Recommendation for core categories from interviews) (Bittner et al., 2021)
	Stagnation (e.g., not enough ideas within 10 minutes of brainstorming) (Strohmann et al., 2018)	Proactive acting - stimulus topic or motivational quote (Strohmann et al., 2018)

An AI-based system could support team processes by e.g., summarizing and clustering results from user, market, and problem analysis and presenting them in a more comprehensible way (Bittner et al., 2021; Lembcke et al., 2020). Furthermore, AI could assist with meeting planning and organization by retrieving calendar information (Cranshaw et al., 2017) from DT participants and suggesting suitable time slots for the subsequent sessions, or it could use transcriptions from previous sessions (Bittner et al., 2021) to create recaps during subsequent sessions.

Regarding the DT area of Team Interaction, we found even more cases where AI could meaningfully assist. For instance, by using different inputs from body language (mimics, gestures, eye gaze, facial expressions), AI could create a team mood board, based on which it could step in to lighten the mood with jokes or anecdotes, in case the mood is shifting to negative (Benke et al., 2020; Bittner et al., 2021; Strohmann et al., 2018). By analyzing eye-movement data, AI could create an information load board and suggest a break at a suitable time point (Fig. 1b). Furthermore, based on these input data, AI could identify personal characteristics of the participants in order to provide more tailored interventions when a critical situation occurs (Strohmann et al., 2018), or it could act as an Animator or Game Master that motivates the team by inducing mood-lightening recreational interventions such as games or jokes when participants get exhausted in the course of a long workshop (Bittner et al., 2021). Moreover, to foster team interaction, AI could infer team states from extracted human signals such as speech shares (Fig. 1a) and centrality (Bittner et al., 2021; Clawson et al., 1993; Leimeister, 2014; Strohmann et al., 2018), critical or negative utterances or behavior (Bittner et al., 2021), and even a lack thereof, in which case it could act proactively by providing a stimulus topic or a motivational quote (Strohmann et al., 2018).

Apart from Team Facilitation, AI could also serve as Facilitator Assistance and Method Specific Assistance. In the former case, AI could support both process and interaction facilitation. For example, by analyzing personal profiles of participants, AI could foster team building by employing an unbiased composition of teams, e.g., with sufficient diversity (Bittner et al., 2021; Cautela et al., 2019; Xiao et al., 2019). On the interaction side, AI could use mood indicators to alert a human DT facilitator about negative team dynamics (Bittner et al., 2021; Elshan et al., 2022; Leimeister, 2014).

Examples of Method Specific Assistance can be demonstrated on the Ideation Method. Here, an AI could build upon ideas from participants to support them in various ways, e.g.,

it could provide individual feedback and inspiration for further ideas (Przybilla et al., 2019; Strohmman et al., 2017), or it could assist by creating an idea database, structuring and rating ideas by similarity, problem to be solved, and innovativeness of the idea (Bittner et al., 2019a, 2021; Ulrich, 2018; Voigt, 2014). Additionally, AI could build upon ideas to create recommendation templates for sketches or other forms of visualization, such as movies or physical modeling through materials such as cardboard (Bittner et al., 2021; Voigt, 2014).

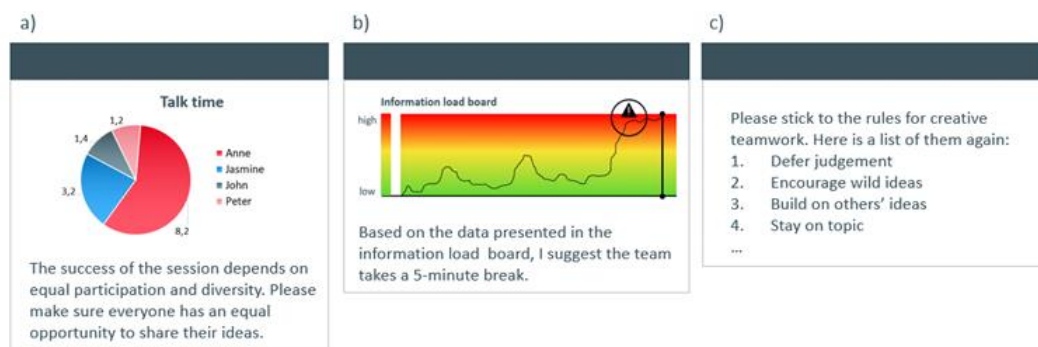


Figure 1. Examples of possible AI outputs for different inputs: a) Unbalanced speech shares. b) Increased cognitive load. c) Breaking rules of creative teamwork.

In light of the findings from the literature, we contend that the AI output design affects the privacy calculus and, thus, ID. Specifically, we hypothesize that H1: *AI Output perceived as beneficial leads to higher willingness to disclose particular information.*; H2: *AI Output perceived as risky regarding loss of privacy leads to lower willingness to disclose particular information.* To address our research questions, an online experiment employing a vignette method will be conducted. To answer the main research question, i.e., RQ2: "Which information and to what extent are humans willing to disclose so that AI-based systems can support them in specific tasks in creative teamwork?", we will measure ID as an ordinal variable on a 5-point likert scale by asking participants how likely they are to disclose particular information. Independent variables hereby are input content (see examples in Table 1) and for each input the AI output design in terms of recipient, content, and representation.

Table 2. Levels of independent variables. Input and output content is not exhaustive.

Input	Output		
Content	Recipient	Content	Representation
Body language	Team	Reminder	Non-transparent
Speech shares	Individual	Dashboard	Transparent anonymized
Negative utterances		Animation	Transparent non-anonymized
Keywords for breaking rules		Explanation	

To address our research hypotheses, for each output design perceived benefit and perceived risk of the design will be measured. For each input, we will ask to what extent participants

would disclose it, after presenting different AI output designs in terms of recipients and representation. Table 2 provides an overview of the levels of the independent variables.

We will further analyze the impact of AI output in terms of recipient and representation on ID. To test the hypotheses, we will perform linear regression analyses and corresponding post-hoc tests. Moreover, we will employ the analysis of covariance to test whether different information leads to different disclosure preferences.

The experiment will employ a vignette-based method, which is a proven effective and economical method to assess behaviours, attitudes and intentions. Furthermore, it is a suitable method for exploring sensitive topics, such as ID (Aguinis and Bradley, 2014; Karren and Barringer, 2002). Vignette method employs imagined scenarios, in our case AI is working together with human teams on a task, using the respective inputs as a way to foster task performance and team interaction, as this is the area of potentially meaningful AI assistance in DT with major ID prerequisites. The design of the scenario will be informed via literature review. After presenting each vignette (one vignette corresponding to one AI output design option for one input), participants will be asked to what extent they would disclose certain information to enjoy the benefits of this specific AI assistance. Some vignettes will serve as a reward nudge for ID and will employ principles of transparent AI, such as meaningful explanations that accurately reflect the system's process and emphasize that the system only operates under the conditions for which it was designed (Phillips et al., 2020). These will ensure the delivery of proper explanations for data usage and foster trust. Participants will also be able to select why they would choose not to disclose certain information. An open-end question will be provided to account for factors for non-disclosure not identified in the literature. We will collect further qualitative data regarding participants' suggestions about the possible solutions that could lead to increased disclosure intention.

AI output will vary in terms of content, recipient, and representation. Content can be, e.g., a reminder to follow DT rules (see Fig. 1c), or a joke in response to a negative mood. Participants will choose, whether they would rather accept such output provided to the entire team or only in private. The output representation will vary in terms of transparency and anonymity. Three design options will be presented for each input content and recipient: non-transparent, transparent and anonymized, and transparent and non anonymized (Fig. 2). The assumption hereby is that different output designs lead to different disclosure preferences in terms of recipients. Due to possible differences in perceived risks and benefits of ID in teams vs. in private, we hypothesize that, e.g., a pie chart with speech shares might rather be received only personally, in an anonymized form, but a suggestion for a break based on a summated information load board might be more beneficial when presented to a whole team. However, we are not locking in assumptions, and aim to derive specific hypotheses for each output in the following stages of this research.

Figure 2 illustrates an example of the experimental material. Here, speech shares of participants are shown by AI, with a hint to keep the speech shares as equal as possible, since this is a very important principle for successful DT practices.

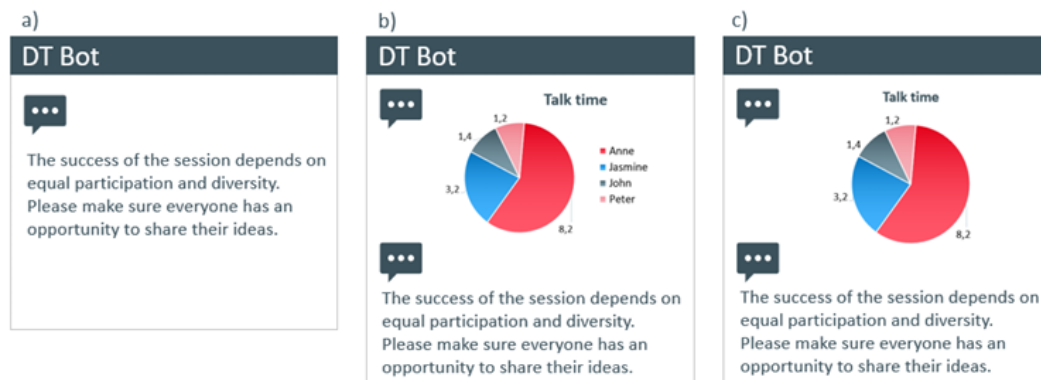


Figure 2. Example of the experimental material. a) Non-transparent output. b) Transparent and non-anonymized output. c) Transparent and anonymized output.

Three design options are considered: in the first one (Fig. 2a) only a hint to equalize talk times is shown. There is no transparency, as participants do not know which input AI used to provide them with this information; in the second variant (Fig. 2b) talk time is shown together with the names of corresponding participants. This is a transparent and non-anonymized representation of output. The pie chart depicts what information (talk time) AI used for this output, thus ensuring transparency. This output is non-anonymized since the chart also displays the names of the participants. The third option (Fig. 2c) only shows that talk time is unequal between participants, without exposing any names. Therefore, this output representation is both transparent and anonymized. The potential design space for AI output is immense, ranging from minimal cues to complex dashboards. We aim to further refine our design space by evaluating results of the proposed study while also continuously supplementing the theoretically guided part by studying current literature.

The experiment will be performed with workers and/or students who have already participated in DT workshops, in order to ensure their understanding of the scenario as well as an understanding of challenges of DT and possible benefits of AI assistance. Due to differences in privacy concerns regarding gender and age (Knijnenburg and Kobsa, 2013a) we aim for a balanced sample, in order to control for these variables in subsequent analyses.

11.4 Expected Contribution and Conclusion

AI-assisted DT shows great potential for supporting collaborative creative work. Important challenges en route to human-centered AI design for Human-AI collaboration concern ID and ethical data usage. By conducting our study, we plan on contributing to this research stream, in both theoretical and practical terms.

First, the study contributes to the knowledge base by identifying relevant human inputs in DT that could potentially be used by AI to support DT sessions as a teammate or (co-

)facilitator. Furthermore, for each identified input, we specify a corresponding AI output, which maximizes the value of the respective input to support teamwork in DT. In doing so, we expect to gain a better understanding of privacy calculus in the given context as well as derive design guidelines for transparent AI that fosters ID and helps to create value in collaborative creative work. By gaining insights about our initial hypotheses we expect to derive more specific hypotheses matched to specific input-output pairs in further iterations. Moreover, we hope to inspire researchers in other areas of AI research to conduct groundwork on AI service bundles for particular inputs, while taking in regard privacy preferences.

Our main findings are expected to contribute to the privacy research in Human-AI collaboration, in terms of information types and respective levels of ID for each type. Beyond that, we expect to gain insights into factors related to disclosure, which we will explore via different AI output designs by directly asking study participants about the reasons for their non-disclosure and further suggestions. The study shall thus identify reasons for non-disclosure, such as trust, risk, and disclosure benefits, and strengthen both theoretical and design knowledge about ID. Additionally, by examining ID in terms of AI output recipient, we expect to shed light on differences and similarities of ID in teams vs. in private. Since many determinants of ID are shared across different use cases, we expect to contribute with transferable knowledge by validating these determinants empirically. In addition, perspective development of design patterns shall generalize the findings beyond the specific use case to similar domains.

Limitations of our proposed study include testing only high-level hypotheses within a rather simplistic research model. We plan to address these by drafting a comprehensive research model including various other relevant determinants of ID such as trust and perceived control and analyze it via structural equation modelling. Moreover, we also plan on considering differences in determinants of disclosure in private vs in team and, if necessary, draft separate research models accordingly. Further, just as every method, vignette method comes with its limitations. Finally, measuring intention to disclose comes with a risk of underestimating the privacy paradox, i.e., that in some cases intention to disclose does not correspond to the actual disclosure behavior.

11.5 References

- Aguinis, H. and K. J. Bradley (2014). "Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies." *Organizational Research Methods* 17 (4), 351–371. DOI: 10.1177/1094428114547952. URL: <https://doi.org/10.1177/1094428114547952>.
- Ajzen, I. and M. Fishbein (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, N.J: Prentice-Hall.

- Al-Natour, S., I. Benbasat, and R. Cenfetelli (2021). "Designing Online Virtual Advisors to Encourage Customer Self-disclosure: A Theoretical Model and an Empirical Test." *Journal of Management Information Systems* 38 (3), 798–827. DOI: 10.1080 / 07421222. 2021 .1962595. URL: <https://doi.org/10.1080/07421222.2021.1962595>.
- Andrews, M. L., R. L. Benedicktus, and M. K. Brady (2010). "The effect of incentives on customer evaluations of service bundles." *Journal of Business Research* 63 (1), 71–76. ISSN: 0148-2963. DOI: <https://doi.org/10.1016/j.jbusres.2009.01.011>. URL: <https://www.sciencedirect.com/science/article/pii/S0148296309000459>.
- Bailenson, J. N., N. Yee, D. Merget, and R. Schroeder (2006). "The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction." *Presence: Teleoperators and Virtual Environments* 15 (4), 359–372. DOI: 10.1162/pres.15.4.359. URL: <https://doi.org/10.1162/pres.15.4.359>.
- Bansal, G., B. Nushi, E. Kamar, E. Horvitz, and D. S. Weld (2021). "Is the Most Accurate AI the Best Teammate? Optimizing AI for Teamwork." In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 13, pp. 11405–11414.
- Benke, I., M. T. Knierim, and A. Maedche (2020). "Chatbot-Based Emotion Management for Distributed Teams: A Participatory Design Study." *Proceedings of the ACM on Human-Computer Interaction* 4 (CSCW2). DOI: 10.1145/3415189. URL: <https://doi.org/10.1145/3415189>.
- Berendt, B., O. Günther, and S. Spiekermann (2005). "Privacy in E-Commerce: Stated Preferences vs. Actual Behavior." *Communications of the ACM* 48 (4), 101–106. ISSN: 0001-0782. DOI: 10.1145/1053291.1053295. URL: <https://doi.org/10.1145/1053291.1053295>.
- Bittner, E., G. Küstermann, and C. Tratzky (2019a). "The Facilitator is a Bot: towards a Conversational Agent for Facilitating Idea Elaboration on Idea Platforms." In: *Proceedings of the 27th European Conference on Information Systems (ECIS)*.
- Bittner, E., M. Mirbabaie, and S. Morana (2021). "Digital Facilitation Assistance for Collaborative, Creative Design Processes." In: *Proceedings of the 54th Hawaii International Conference on System Sciences*.
- Bittner, E., S. Oeste-Reiß, P. Ebel, and M. Söllner (2019b). "Mensch-Maschine-Kollaboration: Grundlagen, Gestaltungsherausforderungen und Potenziale für verschiedene Anwendungsdomänen." *HMD Praxis der Wirtschaftsinformatik* 56 (1), 34–49. ISSN: 2198-2775. DOI: 10.1365/s40702-018-00487-1.
- Bittner, E. and O. Shoury (2019). "Designing Automated Facilitation for Design Thinking: A Chatbot for Supporting Teams in the Empathy Map Method." In: *Proceedings of the 52nd Hawaii International Conference on System Sciences*.
- Brown, T. and B. Katz (2011). "Change by design." *Journal of product innovation management* 28 (3), 381–383.

- Carlgren, L., M. Elmqvist, and I. Rauth (2014). "Exploring the use of design thinking in large organizations: Towards a research agenda." *Swedish design research journal* 11, 55–63.
- Cautela, C., M. Mortati, C. Dell'Era, and L. Gastaldi (2019). "The impact of Artificial Intelligence on Design Thinking practice: Insights from the Ecosystem of Startups." *Strategic Design Research Journal*.
- Chang, C.-W. and J. Heo (2014). "Visiting theories that predict college students' self-disclosure on Facebook." *Computers in Human Behavior* 30, 79–86. ISSN: 0747-5632. DOI: <https://doi.org/10.1016/j.chb.2013.07.059>. URL: <https://www.sciencedirect.com/science/article/pii/S0747563213002987>.
- Clawson, V. K., R. P. Bostrom, and R. Anson (1993). "The Role of the Facilitator in Computer-Supported Meetings." *Small Group Research* 24 (4), 547–565. DOI: 10.1177/1046496493244007. URL: <https://doi.org/10.1177/1046496493244007>.
- Cranor, L. F. (2008). "A framework for reasoning about the human in the loop." *Advanced Computing Systems Professional and Technical Association*.
- Cranshaw, J., E. Elwany, T. Newman, R. Kocielnik, B. Yu, S. Soni, J. Teevan, and A. Monroy-Hernández (2017). "Calendar.Help: Designing a Workflow-Based Scheduling Agent with Humans in the Loop." In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Denver, Colorado, USA: Association for Computing Machinery, 2382–2393. ISBN: 9781450346559. DOI: 10.1145/3025453.3025780.
- Debowski, N., D. Siemon, and E. Bittner (2021). "Problem Areas in Creativity Workshops and Resulting Design Principles for a Virtual Collaborator." In: *PACIS 2021 Proceedings*. 108.
- Dellermann, D., A. Calma, N. Lipusch, T. Weber, S. Weigel, and P. Ebel (2019a). "The future of human-AI collaboration: a taxonomy of design knowledge for hybrid intelligence systems." In: *Proceedings of the 52nd Hawaii International Conference on System Sciences*.
- Dellermann, D., P. Ebel, M. Söllner, and J. M. Leimeister (2019b). "Hybrid Intelligence." *Business & Information Systems Engineering* 61 (5), 637–643. ISSN: 1867-0202. DOI: 10.1007/s12599-019-00595-2. URL: <https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1543&context=bise>.
- Elshan, E., D. Siemon, T. de Vreede, G.-J. de Vreede, S. Oeste-Reiß, and P. Ebel (2022). "Requirements for AI-based Teammates: A Qualitative Inquiry in the Context of Creative Workshops." In: *Proceedings of the 55th Hawaii International Conference on System Sciences*.
- Hirschprung, R., E. Toch, F. Bolton, and O. Maimon (2016). "A Methodology for Estimating the Value of Privacy in Information Disclosure Systems." *Computers in Human Behavior* 61 (C), 443–453. ISSN: 0747-5632. DOI: 10.1016/j.chb.2016.03.033. URL: <https://doi.org/10.1016/j.chb.2016.03.033>.

- Hjalmarsson, A., J. Recker, M. Rosemann, and M. Lind (2015). "Understanding the Behavior of Workshop Facilitators in Systems Analysis and Design Projects: Developing Theory from Process Modeling Projects." *Communications of the Association for Information Systems* 36.
- Holzinger, A. (2016). "Interactive Machine Learning (iML)." *Informatik-Spektrum* 39 (1), 64–68. ISSN: 1432-122X. DOI: 10.1007/s00287-015-0941-6.
- Karren, R. J. and M. W. Barringer (2002). "A Review and Analysis of the Policy-Capturing Method-ology in Organizational Research: Guidelines for Research and Practice." *Organizational Research Methods* 5 (4), 337–361. DOI: 10.1177/109442802237115. URL: <https://doi.org/10.1177/109442802237115>.
- Knijnenburg, B. P. and A. Kobsa (2013a). "Helping Users with Information Disclosure Decisions: Potential for Adaptation." In: *Proceedings of the 2013 International Conference on Intelligent User Interfaces*. IUI '13. Santa Monica, California, USA: Association for Computing Machinery, 407–416. ISBN: 9781450319652. DOI: 10.1145/2449396.2449448.
- Knijnenburg, B. P. and A. Kobsa (2013b). "Making decisions about privacy: information disclosure in context-aware recommender systems." *ACM Transactions on Interactive Intelligent Systems (TiiS)* 3 (3), 1–23.
- Kroll, T. and S. Stieglitz (2021). "Digital nudging and privacy: improving decisions about self-disclosure in social networks." *Behaviour & Information Technology* 40 (1), 1–19. DOI: 10.1080/0144929X.2019.1584644. URL: <https://doi.org/10.1080/0144929X.2019.1584644>.
- Laufer, R. S. and M. Wolfe (1977). "Privacy as a Concept and a Social Issue: A Multidimensional Developmental Theory." *Journal of Social Issues* 33 (3), 22–42. DOI: <https://doi.org/10.1111/j.1540-4560.1977.tb01880.x>. URL: <https://spssi.onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-4560.1977.tb01880.x>.
- Leimeister, J. (2014). "Unterstützung der Zusammenarbeit." In: *Collaboration Engineering*. Berlin, Heidelberg: Springer Gabler. DOI: 10.1007/978-3-642-20891-1_3. URL: https://doi.org/10.1007/978-3-642-20891-1_3.
- Lembcke, T.-B., S. Diederich, and A. B. Brendel (2020). "Supporting Design Thinking Through Creative and Inclusive Education Facilitation: The Case of Anthropomorphic Conversational Agents for Persona Building." In: *Proceedings of the 28th European Conference on Information Systems (ECIS), An Online AIS Conference*.
- Li, H., R. Sarathy, and H. Xu (2010). "Understanding Situational Online Information Disclosure as a Privacy Calculus." *Journal of Computer Information Systems* 51 (1), 62–71. DOI: 10.1080/08874417.2010.11645450. URL: <https://www.tandfonline.com/doi/abs/10.1080/08874417.2010.11645450>.
- Li, Y. (2012). "Theories in online information privacy research: A critical review and an integrated framework." *Decision Support Systems* 54 (1), 471–481. ISSN: 0167-

9236. DOI: <https://doi.org/10.1016/j.dss.2012.06.010>. URL: <https://www.sciencedirect.com/science/article/pii/S0167923612001935>.
- Liao, M. and S. S. Sundar (2021). "How Should AI Systems Talk to Users When Collecting Their Personal Information? Effects of Role Framing and Self-Referencing on Human-AI Interaction." In: New York, NY, USA: Association for Computing Machinery. ISBN: 9781450380966. URL: <https://doi.org/10.1145/3411764.3445415>.
- Liedtka, J. M. (2015). "Perspective: Linking Design Thinking with Innovation Outcomes through Cognitive Bias Reduction." *Journal of Product Innovation Management* 32, 925–938.
- Manheim, K. and L. Kaplan (2018). "Artificial Intelligence: Risks to Privacy and Democracy." *Yale Journal of Law and Technology* 21, 106.
- Manikonda, L., A. Deotale, and S. Kambhampati (2018). "What's up with Privacy? User Preferences and Privacy Concerns in Intelligent Personal Assistants." In: *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. AIES '18. New Orleans, LA, USA: Association for Computing Machinery, 229–235. ISBN: 9781450360128. DOI: 10.1145/3278721.3278773. URL: <https://doi.org/10.1145/3278721.3278773>.
- Masaviru, M. (2016). "Self-Disclosure: Theories and Model Review." *Journal of Culture, Society and Development* 18, 43–47.
- McCarthy, J. (2007). "What is Artificial Intelligence?" URL: <http://www-formal.stanford.edu/jmc/whatisai/>.
- Meinel, C. and L. Leifer (2012). "Design Thinking Research." In: *Design Thinking Research: Measuring Performance in Context*. Ed. by H. Plattner, C. Meinel, and L. Leifer. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–10. ISBN: 978-3-642-31991-4. DOI: 10.1007/978-3-642-31991-4_1. URL: https://doi.org/10.1007/978-3-642-31991-4_1.
- Metzger, M. J. (2004). "Privacy, Trust, and Disclosure: Exploring Barriers to Electronic Commerce." *Journal of Computer-Mediated Communication* 9 (4). DOI: <https://doi.org/10.1111/j.1083-6101.2004.tb00292.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1083-6101.2004.tb00292.x>.
- Milne, G. R. and M. E. Gordon (1993). "Direct Mail Privacy-Efficiency Trade-offs within an Implied Social Contract Framework." *Journal of Public Policy & Marketing* 12 (2), 206–215. DOI: 10.1177/074391569101200206. URL: <https://doi.org/10.1177/074391569101200206>.
- Patil, S. and A. Kobsa (2005). "Uncovering Privacy Attitudes and Practices in Instant Messaging." In: *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*. GROUP '05. Sanibel Island, Florida, USA: Association for Computing Machinery, pp. 109–112. ISBN: 1595932232. DOI: 10.1145/1099203.1099220. URL: <https://doi.org/10.1145/1099203.1099220>.

- Patil, S. and J. Lai (2005). "Who Gets to Know What When: Configuring Privacy Permissions in an Awareness Application." In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA: Association for Computing Machinery, pp. 101–110. ISBN: 1581139985. URL: <https://doi.org/10.1145/1054972.1054987>.
- Phillips, P. J., C. A. Hahn, P. C. Fontana, D. A. Broniatowski, and M. A. Przybocki (2020). "Four principles of explainable artificial intelligence." Technical Report. <https://doi.org/10.6028/NIST.IR.8312-draft>.
- Pitt, J. (2012). "Design contractualism for pervasive/affective computing." *IEEE Technology and Society Magazine* 31 (4), 22–29.
- Przybilla, L., L. Baar, M. Wiesche, and H. Krcmar (2019). "Machines as Teammates in Creative Teams: Digital Facilitation of the Dual Pathways to Creativity." In: New York, NY, USA: Association for Computing Machinery. ISBN: 9781450360883. URL: <https://doi.org/10.1145 / 3322385.3322402>.
- Rayo, L. and I. Segal (2010). "Optimal Information Disclosure." *Journal of Political Economy* 118 (5), 949–987. DOI: 10.1086/657922. URL: <https://doi.org/10.1086/657922>
- Russell, S. and P. Norvig (1995). *Artificial intelligence: a modern approach*. Prentice Hall.
- Salim, S. S. and N. M. Yusoff (2008). "Ethics and Information Privacy in Affective Computing." In: *Proceedings of the Regional Development International Conference and Exhibition (REDICE08)*.
- Schallmo, D. R. A. (2017). "Theoretische Grundlagen." In: *Design Thinking erfolgreich anwenden: So entwickeln Sie in 7 Phasen kundenorientierte Produkte und Dienstleistungen*. Wiesbaden: Springer Fachmedien Wiesbaden, pp. 11–28. ISBN: 978-3-658-12523-3. DOI: 10.1007/978-3-658-12523-3_2. URL: https://doi.org/10.1007/978-3-658-12523-3_2.
- Seeber, I. (2019). "How do facilitation interventions foster learning? The role of evaluation and coordination as causal mediators in idea convergence." *Computers in Human Behavior* 94, 176–189. ISSN: 0747-5632. DOI: <https://doi.org/10.1016/j.chb.2018.11.033>. URL: <https://www.sciencedirect.com/science/article/pii/S0747563218305685>.
- Seeber, I., E. Bittner, R. O. Briggs, T. de Vreede, G.-J. de Vreede, A. Elkins, R. Maier, A. B. Merz,
- S. Oeste-Reiß, N. Randrup, G. Schwabe, and M. Söllner (2020). "Machines as teammates: A research agenda on AI in team collaboration." *Information Management* 57 (2), 103174. ISSN: 0378-7206. DOI: <https://doi.org/10.1016/j.im.2019.103174>. URL: <https://www.sciencedirect.com/science/article/pii/S0378720619303337>.
- Siemon, D. and T. Strohmman (2021). "Human-AI collaboration: introducing the virtual collaborator." In: *Collaborative Convergence and Virtual Teamwork for Organizational Transformation*. IGI Global, pp. 105–119.

- Starostka, J., M. R. Evald, A. H. Clarke, and P. R. Hansen (2021). "Taxonomy of design thinking facilitation." *Creativity and Innovation Management* 30 (4), 836–844. DOI: <https://doi.org/10.1111/caim.12451>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/caim.12451>.
- Strohmann, T., S. Fischer, D. Siemon, F. Brachten, C. Lattemann, S. Robra-Bissantz, and S. Stieglitz (2018). "Virtual Moderation Assistance: Creating Design Guidelines for Virtual Assistants Supporting Creative Workshops." In: *PACIS 2018 Proceedings*. 80.
- Strohmann, T., D. Siemon, and S. Robra-Bissantz (2017). "brAInstorm: Intelligent Assistance in Group Idea Generation." In: *Designing the Digital Transformation*. Ed. by A. Maedche, J. vom Brocke, and A. Hevner. Cham: Springer International Publishing, pp. 457–461. ISBN: 978-3-319-59144-5.
- Stutzman, F., R. Capra, and J. Thompson (2011). "Factors mediating disclosure in social network sites." *Computers in Human Behavior* 27 (1). Current Research Topics in Cognitive Load Theory, 590–598. ISSN: 0747-5632. DOI: <https://doi.org/10.1016/j.chb.2010.10.017>. URL: <https://www.sciencedirect.com/science/article/pii/S0747563210003158>
- Thibaut, J. and H. Kelley (1959). *The social psychology of groups*. John Wiley.
- Ulrich, F. (2018). "Exploring Divergent and Convergent Production in Idea Evaluation: Implications for Designing Group Creativity Support Systems." *Communications of the Association for Information Systems* 43. DOI: 10.17705/1CAIS.04306. URL: <https://doi.org/10.17705/1CAIS.04306>.
- Voigt, M. (2014). "Improving design of systems supporting creativity-intensive processes—A cross-industry focus group evaluation." *Communications of the Association for Information Systems* 34 (1), 86.
- Wei, K., J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. S. Quek, and H. V. Poor (2020). "Federated Learning With Differential Privacy: Algorithms and Performance Analysis." *IEEE Transactions on Information Forensics and Security* 15, 3454–3469. DOI: 10.1109/TIFS.2020.2988575.
- Weinmann, M., C. Schneider, and J. Vom Brocke (2016). "Digital nudging." *Business & Information Systems Engineering* 58 (6), 433–436.
- Westlund, J. K., J. J. Lee, L. Plummer, F. Faridi, J. Gray, M. Berlin, H. Quintus-Bosz, R. Hartmann, M. Hess, S. Dyer, et al. (2016). "Tega: a social robot." In: *11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 561–561.
- Xiao, Z., M. X. Zhou, and W.-T. Fu (2019). "Who should be my teammates: Using a conversational agent to understand individuals and help teaming." In: *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pp. 437–447.
- Xu, H., T. Dinev, H. Smith, and P. Hart (2008). "Examining the Formation of Individual's Privacy Concerns: Toward an Integrative View." In: *ICIS 2008 Proceedings*. 6.

- Xu, H., H.-H. Teo, B. C. Y. Tan, and R. Agarwal (2009). "The Role of Push-Pull Technology in Privacy Calculus: The Case of Location-Based Services." *Journal of Management Information Systems* 26 (3), 135–174. DOI: 10.2753/MIS0742-1222260305. URL: <https://doi.org/10.2753/MIS0742-1222260305>.
- Youn, S. (2009). "Determinants of Online Privacy Concern and Its Influence on Privacy Protection Behaviors Among Young Adolescents." *Journal of Consumer Affairs* 43 (3), 389–418. DOI: <https://doi.org/10.1111/j.1745-6606.2009.01146.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-6606.2009.01146.x>.
- Zoom Video Communications, I. *Attendee attention tracking*. <https://support.zoom.us/hc/en-us/articles/115000538083-Attendee-attention-tracking>. Accessed: 07-27-2021.

12 A User-centric Taxonomy for Conversational Generative Language Models

von Brackel-Schmidt, C., Kučević, E., Memmert, L., Tavanapour, N., Cvetkovic, I., Bittner, E., & Böhmman, T. (2023). A User-centric Taxonomy for Conversational Generative Language Models. In *43rd International Conference on Information Systems (ICIS)*, Hyderabad, India.

Abstract

Conversational generative language models (GLMs) like ChatGPT are being rapidly adopted. Previous research on non-conversational GLMs showed that formulating prompts is critical for receiving good outputs. However, it is unclear how conversational GLMs are used when solving complex problems that require multi-step interactions. This paper addresses this research gap based on findings from a large participant event we conducted, where ChatGPT was iteratively and in a multi-step manner used while solving a complex problem. We derived a taxonomy of prompting behavior employed for solving complex problems as well as archetypes. While the taxonomy provides common knowledge on GLMs usage based on analyzed input-prompts, the different archetypes facilitate the classification of operators according to their usage. With both we provide exploratory knowledge and a foundation for design science research endeavors, which can be referred to, enabling further research and development of prompt engineering, prompting tactics, and prompting strategies on common ground.

Keywords: Generative Language Models, Generative AI, human-computer-interaction

12.1 Introduction

Advancements in artificial intelligence (AI) have led to increased interest in how AI can be leveraged to support complex problem solving (Krogh, 2018), moving towards humans and AI systems solving problems collaboratively (Akata et al., 2020; Dellermann et al., 2019). While addressing open-ended problems might have been difficult with more traditional AI approaches (Krogh, 2018), generative language models (GLMs) as a new type of machine learning model have been shown to be useful in supporting humans in almost unbounded area of language-mediated tasks (Brown et al., 2020; Lester et al., 2021; Mishra, Khashabi, Baral, & Hajishirzi, 2022; Swanson et al., 2021; Wu et al., 2022). However, it is still an open question how these models can best be used for different tasks.

With the recent introduction of a chat-based interaction mode, the generative capabilities of GLMs (e.g., OpenAI's ChatGPT) have become broadly available, even to non-technical users (Jiang et al., 2022). Models such as OpenAI's GPT perform well at generating free

text based on a given input, even without additional task-specific training (Brown et al., 2020). Since GLM system output is conditioned on input during inference time (Brown et al., 2020), it might be challenging to create a suitable formulation for the input request to receive an expected generated output (Jiang et al., 2022), reflecting the research field related to prompt engineering (Lester et al., 2021; Mishra, Khashabi, Baral, Choi, et al., 2022). Studies on this topic already investigate methods to tune initial input-prompts to generate more optimal outputs for solving complex problems with GLM systems (Devlin et al., 2019; Hambardzumyan et al., 2021; Jiang et al., 2022; Lester et al., 2021; Liu et al., 2023; Mishra, Khashabi, Baral, Choi, et al., 2022). Scholars have shown that even slight modifications of initial prompts might result in different GLM system outputs (Jiang et al., 2022; Lester et al., 2021; Liu et al., 2023; Mishra, Khashabi, Baral, Choi, et al., 2022). While outputs to initial prompts might produce unexpected content, GLM system users seem to develop tactics for working with such systems to solve complex tasks, such as task allocation, in prompt chains, if provided with building blocks (Wu et al., 2022).

The release of the chat-based interaction mode with GLM enables long conversations (Jiang et al., 2022), and it might open new ways for tactics to work with such systems to solve complex problems. It is unclear whether and how the chat-based interaction mode might affect known prompting tactics or lead to new prompting strategies for users. Therefore, more research in this field is needed. Additionally, limited knowledge exists about how humans use GLM systems (Jiang et al., 2022; Lester et al., 2021; Mishra, Khashabi, Baral, Choi, et al., 2022; Wu et al., 2022) to solve complex problems (Akata et al., 2020; Dellermann et al., 2019), especially if they aim to generate specific content (Jiang et al., 2022; Mishra, Khashabi, Baral, Choi, et al., 2022; Wu et al., 2022). We addressed this research gap by focusing the present study on the following research question (RQ):

RQ: On an operational level, how do humans prompt GLM-based systems to attempt solving complex problems?

To address the RQ, we investigated how humans interact with GLM systems to solve a complex problem with a novel format for human-AI collaboration. We designed one of the first Prompt-a-thon sessions, in which a diverse set of 76 participants worked on specific challenges in small groups with the support of GLMs. Based on the logs of these interactions, we conducted a qualitative content analysis of the input- prompts as a foundation for developing an extendable taxonomy in accordance with Nickerson et al. (2013) and Kundisch et al. (2021b) to characterize human prompting approaches for GLMs. We contribute a taxonomy that offers common ground of conversational GLMs usage conducting researchers and practitioners. Practitioners can use it when utilizing GLMs for specific tasks such as complex problem solving or relate to it for prompt engineering. Researchers can use it to classify and analyze prompts in multi-step interactions, facilitating research for example on prompt strategies. It represents the first comprehensive terminology on GLM usage and its prompts in this field, setting the stage for subsequent research and the evolution of pertinent theories, as noted by Kundisch et al. (2021b).

Additionally, we derived archetypes from the collected data to recognize prompting patterns of humans when interacting with GLMs. Overall, our results provide first insights into prompting approaches during conversations. The remainder of this paper is structured as follows. In the next section, we review related research on GLMs and prompt engineering. Then, we describe our research methodology and how we collected and analyzed the data to develop the taxonomy and derive the archetypes. Subsequently, we present and discuss our results before highlighting the theoretical and practical contributions of our research. After that, we mention the limitations of our research and provide suggestions for future research before closing the paper with a conclusion.

12.2 Related Research

Large language models are a specific type of machine learning-based models trained on large corpora of texts (Brown et al., 2020). Such models are trained to generate the next word(s) upon input, consequently referred to as generative models. GLMs are flexible; they do not require a certain form of input, and they can generate free-form output text based on free-form input text. They have been shown to perform well in traditional natural language processing tasks (e.g., entity recognition, translation) and other tasks (e.g., generating news headlines, completing exams) (Brown et al., 2020), even without task-specific training (e.g., fine tuning) or training data. Given their flexibility in accepting and generating free-form text, and their ability to generate task-specific text without task-specific training, GLMs could enable the resolution of ill-structured, non-repetitive problems that have traditionally been considered too complex or costly to address with AI. (Krogh, 2018). GLMs have been used, for example, in (scientific) writing or design concept generation. (Gero et al., 2022; Lee et al., 2022; Zhu & Luo, 2022)

GLMs do not require a specific input format, such as a programming language, since natural language (instructions) can be used, which –in principle– allows users without specific prior knowledge to prompt GLMs. Historically, important steps were taken with the release of generative pre-trained transformer models, such as GPT-1 (0.1 billion parameters) and its extension GPT-2 (1.5 billion parameters), which no longer relied on task-specific architectures (Brown et al., 2020; Zhu & Luo, 2022). Based on these promising results, GLMs were further developed. Consequently, the model GPT-3 (175 billion parameters) with a higher output accuracy was evolved (Brown et al., 2020). A subclass, GPT-3.5, served as the basis for ChatGPT (Dwivedi et al., 2023), a GLM released in 2022 with a user-friendly interface that led to increased public awareness of GLMs. ChatGPT was trained in conversations and introduced a chat-based interaction mode, allowing humans to engage in multi-step interactions with the GLM in a very familiar manner. Instead of a single response to a single input, ChatGPT can engage in a continuous conversation that considers previous messages in that conversation. ChatGPT has since then been quickly and widely adopted and was updated in spring 2023 by embedding the fourth generation of GPT, GPT-4.

With no task-specific training (or fine tuning), GLM output is solely conditioned on the input text (and inference parameters), which is referred to as a prompt. Even small changes in the input-prompt can lead to vastly different results in the output (Jiang et al., 2022). Early research showed that novices find it especially difficult to effectively formulate prompts (Jiang et al., 2022; Zamfirescu-Pereira et al., 2023). Given the prompt’s critical role, the field of prompt engineering emerged with the goal of “finding the most appropriate prompt to allow a [language model] to solve the task at hand” (Liu et al., 2023, p. 2). For non- conversational GLMs, several techniques, including demonstrations (i.e., examples of the desired output), have been proposed (Brown et al., 2020). Compared to more traditional machine learning-based systems, where exemplary data is used as training data during training time, examples can be used in GLMs during inference time as a prompt or additional information within a prompt (Brown et al., 2020). In addition to not using demonstrations (zero-shot), providing examples allows for one-shot (one example) and few-shot (multiple example) learning, with different effects on output quality depending on the task (Brown et al., 2020). Other techniques, such as itemization (i.e., breaking up larger text portions into bulleted lists), have also been explored to improve the results (Mishra, Khashabi, Baral, Choi, et al., 2022). More advanced approaches split up a larger task into multiple steps. Instead of a large prompt, GLM is prompted with several more specific prompts referred to as “primitive operations” (Wu et al., 2022, p. 1). Examples of such operations include information extraction and factual query (Wu et al., 2022). The output of one operation can be used as an input for the next iteration, resulting in an approach called prompt chaining (Wu et al., 2022), which can result in a more effective interaction with the system.

Given the novelty of chat-based GLMs, such as ChatGPT, most (beyond anecdotal) prompt engineering knowledge refers to non-conversational, one-off interactions with GLMs. Prompt engineering has been shown to significantly improve output quality for such non-conversational interactions (Mishra, Khashabi, Baral, Choi, et al., 2022). This also applies to providing users with guidance and support in interacting with GLMs by distilling potential types of interactions (primitive operations (Wu et al., 2022)). However, the introduction of a chat-based interface has caused a shift from one-off interactions to free-form, multi-step conversations. This change to multi-step conversations might reduce the need to specify all relevant information in a single prompt and might enable the stepwise development of work results, from developing a more in-depth understanding of the problem to iteratively developing solution aspects, allowing the operator to add information as needed or align the understanding. Since it is unclear how conversational GLMs affect prompting behavior and prompt engineering, in the present study, we sought to explore how conversational GLMs with multiple interactions are used when solving complex problems. Moreover, we sought to build a guiding frame to more holistically understand such interactions as a foundation for analyzing conversations with GLMs. Future studies could build on such a frame to explore the effectiveness of different approaches to

prompting (e.g., Zamfirescu-Pereira et al. (2023) for non-conversational GLMs) and develop prescriptive knowledge on how to converse with such systems. Although we openly explored this question using an inductive approach, we relate our guiding frame to previous work as a part of the discussion. However, an analysis of the effectiveness of the distilled approaches was not part of this study.

12.3 Research Methodology

The purpose of the present study was to shed light on how the disruptive and rapidly emerging (conversational) GLM technology is being used. In this regard, little knowledge about the usage of GLMs in the context of complex problem solving exists. Therefore, we chose to create a taxonomy as it can serve as a basis for research (Kundisch et al., 2021b). We followed the workflow activities according to Kundisch et al. (2021b), as their approach embeds Peffers et al.'s (2007) design science research (DSR) and besides builds on the well-established taxonomy development by Nickerson et al. (2013). According to Gregor and Hevner (2013), the results can be seen as a level two contribution type, representing design-theoretical knowledge. For clarity reasons, data collection and taxonomy development are described separately. First, the data collection section explains the approach used to collect data that grounds the taxonomy development. Second, the taxonomy development section describes its development and DSR embedding in detail. Third, the clustering and archetype development is explained, which utilizes and evaluates the taxonomy additionally. These activities are described in more detail below.

12.3.1 Data Collection with a Prompt-a-thon

GLM can still be considered an emerging technology and research field, as can its usage in tasks such as complex problem solving. Therefore, to create a solid data foundation upon which to build our research, we developed a data collection instrument called a Prompt-a-thon which led to 252 unique input-prompts. Our Prompt-a-thon was inspired by hackathons, where participants work in teams for a limited period of time to solve a challenge or problem (Taylor & Clarke, 2018). Their characteristic of possibly attracting large numbers of participants (Olesen & Halskov, 2020) from diverse backgrounds and enabling to engage with different technologies (Taylor & Clarke, 2018) let us to choose it as format inspiration for data collection, especially since it has been proven for research data generation (Olesen & Halskov, 2020). We defined a Prompt-a-thon as an event in which people work in a collaborative manner over a fixed period of time to create a solution for an idea or a challenge. In contrast to a hack-a-thon, Prompt-a-thon participants do not solely rely on themselves while working on hardware or coding but actively and deliberately include GLMs (e.g., ChatGPT) in the development process. Since we aimed to collect as broad data samples as possible, the event was constructed with the fewest possible restrictions regarding participants' pre-existing GLM usage expertise and general background (e.g., gender, profession, age, workplace). To ensure a wide variety of

participants, information about and invitation to the event was shared using different multipliers via various channels, including social media, newsletters, word-of-mouth, and direct invitations. Before participants signed in, they were given a brief introduction to the field of GLM. When signing in, they were asked for interests and ideas of challenges they would like to work on. Participants were assigned a challenge in advance – challenges were either provided to the participants or they had suggested them or at least its topics themselves as described before. The challenges had a wide range of issues, all of which involved complex problems that needed to be solved. With that, they could not be solved by one- or few-shots prompts but required multi-step, iterative problem solving. The challenges included, for example, topics of public interest like the federal government and its digitalization. Here, participants had the task to create a concept for digitalization of the federal authorities, considering requirements for digitally requesting services. Other challenges dealt with more creative topics, such as the creation of a puzzle to interactively discover a city while learning about the history and places of interest.

In total, we had 76 participants work on 16 different challenges. They took part either on site or online. Those on site were provided with tablets to work with. The format lasted 2.25 hours and included an introduction to the format and challenges, a prompting session, and a retrospective to share results and lessons learned. Based on 66 returned surveys, the participants can be characterized as following: The participants' composition was heterogeneous (e.g., participants from the field of entrepreneurship, information technology consulting, marketing, data analytics, information systems research, students). The participants were asked how often they use GLMs (in this case ChatGPT), $\approx 45\%$ (30) of the participants used GLMs at least once a week, $\approx 42\%$ (28) used it a few times already prior the Prompt-a-thon and $\approx 12\%$ (8) had never used it before. With that, the usage ranged from never to almost every day. The involved participants' ages varied from 21 to 63 years. On average, the age of our participants was about 32 years. We had a representation of $\approx 39\%$ (26) women, $\approx 56\%$ (37) men and $\approx 4\%$ (3) others. The 76 participants were divided into groups (2-4 persons). The GLM chosen for the event was ChatGPT 3.5 (release February 13) (Natalie, 2023) without plus subscriptions. ChatGPT account credentials were created prior to the event and used to access ChatGPT. Therefore, we could directly access the chat history, extract and tag it with the associated account. As result of the event we collected 252 unique input-prompts (on average 16 per group) that were evaluable. Cases of unusable chats occurred when the input seemed random (e.g., random input- prompts of letters) or data extraction was not possible.

12.3.2 Development of Taxonomy

To ground the taxonomy, we used the data gathered with the aforementioned Prompt-a-thon. Our taxonomy development followed Kundisch et al. (2021b), that builds on Nickerson et al.'s (2013) work but provides more guidance for such. Thus, Kundisch et al. (2021b) link their 18 steps for taxonomy development to the six DSR method activities (Peffer et al., 2007). However, in contrast to Nickerson et al. (2013), we moved away from

the restriction of exclusiveness; thus, an object could have more than one characteristic of a dimension. This was based on intentional openness to possible variations in the input-prompt design of the operators using GLMs. These were followed throughout the development of the taxonomy. The application of each activity and its steps are presented below.

Identify and Motivate (DSR activity 1)

First, in accordance with Kundisch et al. (2021b), the phenomenon that is observed should be specified, the user group(s) be named, and then the purpose of the taxonomy described. GLMs' emergence, popularity with the general public, and sudden relevance have led more and more people to discuss and exchange information about their experiences with GLMs. Discussions of different utilization approaches have led to the to be observed phenomenon, the variety of GLM usage and knowledge about it. As user groups of the taxonomy, researchers may be interested in the present study's findings because they could benefit from the first data-based and structured overview of (conversational) GLM utilization and use it as a foundation from which to dig deeper. For example, they could identify, derive, or analyze prompting strategies as well as potential conversational switching behavior in the course of a conversation. Practitioners may also be interested in the findings because they frequently use such solutions and embed GLMs in their work routines. The taxonomy could help them understand the different characteristics of GLM usage and use them as entry points of adjustment for varying and optimizing their interactions. As a result of this activity, the phenomenon was identified and motivated. Target groups and objectives were defined.

Define the Objectives of a Solution (DSR activity 2)

Following Nickerson et al. (2013), a meta-characteristic to ground further derivation of the characteristics was defined and chosen with respect to the purpose of the taxonomy. Specifically, GLM operator-prompt attributes were selected as meta-characteristic in accordance with the purposed taxonomy angle of the aforementioned phenomenon (Kundisch et al., 2021b). Before iteratively developing the taxonomy, the ending conditions were determined in accordance with Kundisch et al. (2021b) and Nickerson et al. (2013). As such, all of Nickerson's objective and subjective ending conditions were used. The preciseness and applicability of the taxonomy were selected as the initial evaluation goals. The evaluation aimed to validate that the taxonomy served its purpose of classifying studied, and respectively, self-performed GLM usage. To further evaluate the designed taxonomy, a clustering analysis was set as a second goal based on the given objects, their characteristics, and dimensions (see Table 3) (Kundisch et al., 2021b).

Design and Development (DSR activity 3)

The design and development of the taxonomy were done by applying the empirical-to-conceptual and conceptual-to-empirical approaches. Three iterations were carried out as follows: An empirical approach was chosen for the first iteration for several reasons. First,

this research field lacks a known taxonomy. Second, although one could argue that there could be starting points in research in the context of non- conversational GLMs (e.g., primitive operations in (Wu et al., 2022)), primitive operations were not involved in the development of the taxonomy because they could be seen as vague, and we wanted to ensure non-bias when investigating conversational GLM usage. Additionally, the empirical approach is recommended when a “significant number of objects are available representing the phenomenon under consideration” (Kundisch et al., 2021b, p. 430). To maximize reliability and avoid errors (Buscemi et al., 2006; McDonald et al., 2019), three independent units inductively coded 182 prompts using the MAXQDA application. The results were then revised using the taxonomy operations described by Kundisch et al. (2021b). For the second iteration, the conceptual-to-empirical approach was selected to “(Re-)examine objects [in order] to validate the new characteristics and dimensions that [...] have [been] conceptualized” (Kundisch et al., 2021b, p. 430) in the prior iteration. The empirical-to-conceptual approach was used for the last iteration because a sufficient amount of data (70 prompts) was available for further examination. The approaches, dimensions, characteristics, and examined objects of each iteration are shown in Figure 1.

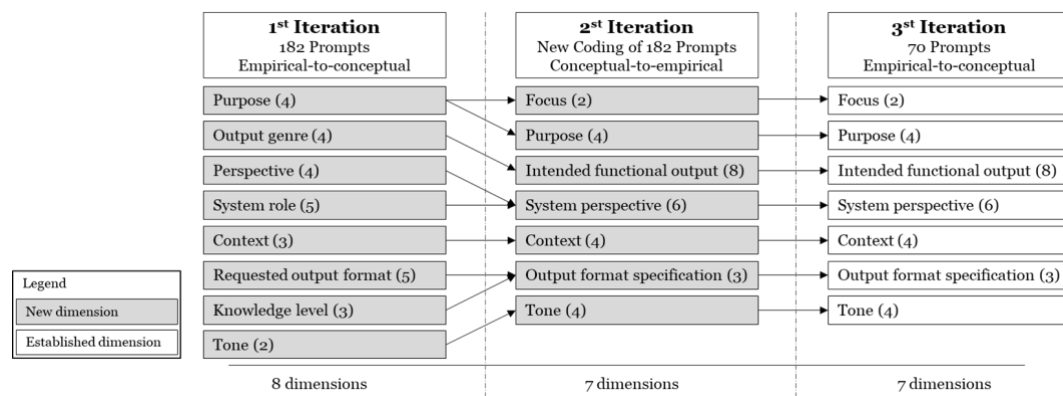


Figure 1. Iterative Taxonomy Development Process

Demonstration (activity 4), Evaluation (activity 5) and Communication (activity 6)

After each iteration, the taxonomy was studied and by that demonstrated to determine whether the objective ending conditions were met. When they were not met, a new iteration was conducted; this occurred after the first two iterations. After passing the objective ending conditions, the subjective ending conditions were checked. The units that independently coded initially also independently checked whether the conditions were met. As this was the case after the third iteration, no new iterations were necessary. Therefore, the next steps were executed in accordance with (Kundisch et al., 2021b). Specifically, the evaluation was configured and performed. The evaluation objects were the taxonomy and its descriptions. Together, they form a DSR artifact from the type of a model and construct, leading to the evaluation criteria of completeness as presented in (Kundisch et al., 2021a). “The degree to which the structure of the artifact contains all necessary elements and relationships between elements” (Kundisch et al., 2021a, p. 26) was evaluated by interviewing experts from the taxonomy’s target groups (Prat et al., 2015). In total, nine

experts (five practitioners and four researchers) were interviewed; none of them were involved in the taxonomy building process. Based on the methodology and criteria of the evaluation, the taxonomy usefulness was evaluated. As mentioned before, clustering was also performed to serve as an evaluation. As a result, the taxonomy was demonstrated, evaluated, and thus ultimately validated. Finally, in accordance with Kundisch et al. (2021b), the communication and documentation took place, which included the taxonomy development process (shown in Figure 1), the visualization of the taxonomy and belonging descriptions for characteristics and (meta-)dimensions (see Results section).

12.3.3 Clustering and Development of Archetypes

Archetypes help identify patterns that can provide insights into an observed phenomenon, and they can serve as a method for evaluating the applicability of a taxonomy (Kundisch et al., 2021b). We empirically determined clusters by performing clustering on our prompts that were labelled with the taxonomy characteristics. The clustering results in groupings of prompts so that the objects in each cluster are more similar to one another than to the prompts in other clusters. Those were then interpreted as archetypes. The statistical calculations and creation of the clusters were performed using Python with respective packages. First, the Euclidean distance was calculated to find similarities between the prompt-objects. Then, hierarchical clusters were calculated using Ward's method (Ward, 1963), which allowed clustering to be prescribed without prior specification of the cluster count (Janssen et al., 2020). It provided an initial visualization of clusters that indicated an optimal cluster count of three. Since a combination of non- and hierarchical clustering has been recommended (Balijepally et al., 2011; Punj & Stewart, 1983), k-means (Hartigan & Wong, 1979) cluster fitting was performed for cluster counts of two to eight. Based on these results, the silhouette scores were calculated and inspected to validate the optimal cluster count. The prior indicated optimal cluster count was confirmed as a cluster count of three related to the highest silhouette score (Berkhin, 2006). Based on the chosen cluster count, k-means was calculated using the k-means++ algorithm (Arthur & Vassilvitskii, 2007), leading to the results described and shown. In addition, the percentage distribution is displayed and colored according to the percentage of occurrence in the Table 3.

12.4 Results

12.4.1 User-centric Taxonomy of GLM Usage

Based on the taxonomy research activities of Nickerson et al. (2013) and Kundisch et al. (2021b), we derived the final taxonomy for classifying user-centered input-prompts. It incorporates a hierarchical structure consisting of 3 meta-dimensions, 7 dimensions, and 31 characteristics, providing a systematic and rigorous classification of input-prompts. The meta-dimensions, which represent the highest level of abstraction, depict the overarching classification areas of GLM usage. The dimensions break them further down.

(Meta-)dimensions provide logical and structural clarity and classification, although they must be sufficiently granular for unambiguous classification purposes. For this reason, characteristics were determined as specific elements that can occur at the third level. Table 1 depicts the final taxonomy. For each meta-dimension, the dimensions and characteristics are further described in the following.

Table 1. Final Taxonomy

Meta-dimension	Dimension	Characteristics								
Prompt objectives	Focus	Informational				Creative				
	Purpose	Build understanding		Sharpen understanding		Task delegation		General conversation		
	Intended functional output	Create	Aggregate	Supplement	Substitute	Assess	Compare	Combine	Reformulate	
Prompt settings	System perspective	Humanized		User-oriented		Group-oriented		Organization-oriented	Expert-oriented	None
	Context	Explicit context		Reference context		Reference previous + explicit context		None		
Prompt style	Output format specification	Not specified			Structured			Domain-specific		
	Tone	Polite		Impolite		Motivational		Neutral		

Prompt Objectives

The prompt objectives meta-dimension represents the following decisive descriptive factors for GLM use: focus, purpose, and intended functional output. It serves as the highest level of aggregation for classifying GLM usage. The focus dimension describes whether GLMs should focus on delivering existing information (including information given by prior input-prompts) or focus on being creative and delivering yet-to-be- created information. These areas of focus, in which the input-prompt instructs the system, have different objectives and requirements for the GLMs. The choice of a suitable focus depends on the operator's individual needs and can vary depending on the area of usage. This dimension is guided by the following question: Can the operator's needs be met by existing information, or is new information creation needed? The informational focus aims to provide general, factual, and informative statements (e.g., summaries, descriptions, reports on specific topics). The focus here is on using and returning existing information instead of, for example, creating new information by applying the provided operator context to existing information. Meanwhile, the creative focus strives to generate new and original outputs (e.g., customized strategy roadmaps, project plans, short stories, poems, song lyrics). The focus here is on creative expression and generating novel and unique information.

The purpose dimension describes the reason for using the GLM to ensure that it meets the objective of the operator. This dimension is guided by the following question: Why does the operator use the GLM? The build understanding characteristic refers to the intention behind an interaction with a GLM to develop an operator's understanding of a specific topic. In contrast, the sharpen understanding characteristic refers to improving the understanding of partly understood (specific) topics by using GLM to, for instance, extract and analyze relevant data or information. Alternatively, operators can use these systems to

outsource tasks by utilizing the task delegation characteristic for activities such as writing texts or creating content. Furthermore, operators can have a general conversation with the system (e.g., conversational dialogue).

Besides purpose, operators have expectations regarding the concrete functions that the system should perform and then return as output. This input-prompt aspect is represented by the intended functional output dimension in which the operator specifies the concrete functionality they want the system to perform while they use it. This dimension is guided by the following question: What do I want the GLM to do for me? Importantly, it should be noted that GLMs are designed to always generate text. Consequently, text creation is integral to GLMs and is included in every intended functionality output. However, the expected functionality output makes a distinction by highlighting the main intended functionality. When GLMs are exclusively used to create content (e.g., blog posts, strategies, articles, advertising texts, product descriptions), the expected output functionality is to create. Another intended functionality is to aggregate sources of information to provide the operator with forms of overviews. Thus, aggregation summarizes information in a narrow sense and generates information in a broad sense. Additionally, these systems can be used to supplement information by complementing or extending information to, for example, provide the operator with a more in-depth understanding of a particular topic. Moreover, these models can be used to substitute information by, for instance, replacing misleading words or sentences. Furthermore, the assess functionality can be used, for instance, to better and more accurately comprehend the relevance and credibility of information, enabling the operator to make well-grounded decisions. Another intended functionality output is to compare information by highlighting the differences and similarities between various information sources. In addition, these models can combine information by connecting information from separate sources and transferring it from one domain to another. Finally, operators can use GLMs to reformulate information into a desirable (e.g., more understandable or readable) form.

Prompt settings

The prompt settings dimensions describe the circumstances that are set with the input-prompt. By adjusting these dimensions, operators can tailor prompts to their specific needs. The system perspective dimension characterizes the point of view that a GLM is told to adopt in a particular setting. The selected perspective can affect the system output and influence how the AI system is employed. Here, operators ask the following question: Which perspective should the GLM adopt? The operator-oriented perspective can be assigned to the system through an input-prompt to provide personalized recommendations and suggestions based on the operator's given individual perspective (e.g., a maternal, student, homeless). Alternatively, the system can adopt a group-oriented perspective and for instance provide outputs based on their perspective. Furthermore, the system can act from an organization-oriented point of view and operate with an organizational fingerprint. In addition, operators can set an expert-oriented perspective in which the system promotes

information from an expert point of view, such as by serving as a virtual IT-consultant. Another option that can be requested through an input-prompt is the humanized perspective, where the systems standpoint and perspective is requested. Lastly, the input-prompt can specify that no perspective be taken (i.e., none); in this case, the system uses a general or previously selected perspective.

The context dimension characterizes the operator's ability to not only formulate a general input-prompt, such as a generic question, but also provide situational content. Operators are guided by the following question: Is context given? Characteristics differ based on the question answer. The operator can explicitly provide contextual information (i.e., the explicit context characteristic) directly to the GLM (e.g., ask a question and provide specific information). Context can also be established by referring to sentences from earlier in the conversation (i.e., the reference previous context characteristic) to support the receipt of an appropriate response. Moreover, both can occur. In this case, explicit context information is given to the system, while, at the same time, references to previous prompt content within the conversation are made. Ultimately, contextual information may not be provided to the system at all (i.e., the none characteristic).

Table 2. Examples of Input-prompts for each derived Characteristic

*Note: The original examples have been translated into English.

Dimension	Characteristic	Example*
Focus	Informational	"Describe to me the function of e-car charging stations."
	Creative	"Write a song text for the target group between 18 and 25 years about [...]"
Purpose	Build understanding	"How can I measure the environmental impact of my software development project?"
	Sharpen understanding	"Tell me more about the technical screening criteria for software used in the logistics."
	Task delegation	"Describe to me the target group of electric mobility in terms of age, [...], and gender."
	General conversation	"Now don't make yourself worse than you are, you are more than a tool!"
Intended functional output	Create	"What is the alternative transportation which this group uses the most after the car?"
	Aggregate	"Summarize the book 'The algorithm has no tact' by Katharina Zweig."

	Supplement	“Can give you give us a more detailed demographic profile of this group?”
	Substitute	“Exchange the question ‘Which city was the main opponent of West Berlin during [...]’”
	Assess	“Which channels can be used to communicate the message most effectively to [...]”
	Compare	“What is the difference between hypotheses and research guiding assumptions?”
	Combine	“How to apply the concept of [...] to advisory firms that want to create AI-bids?”
	Reformulate	“Can you also formulate these ads for the English market?”
System perspective	Humanized	“Do you have any suggestions on how I can optimize the following website?”
	User-oriented	“What do I need to take out occupational disability insurance?”
	Group-oriented	“How can we effectively use ChatGPT in a design thinking process for its products?”
	Organization-oriented	“Assume you are a business in the it-sector aiming to get your it-systems and products [...]”
	Expert-oriented	“Imagine you were a renowned expert on digitalization [...]. What would you [...]?”
	None	“Which electronic devices do drivers of electric vehicles use?”
Context	Explicit context	“We want to develop a marketing strategy for e-car charging stations. Which [...] are [...]?”
	Reference context	“Could you combine all your answers into a puzzle with creative short descriptions of [...]?”
	Reference previous + explicit context	“Would you like to combine your answers into a text consisting of the questions and answers [...]?”

	None	“How many personas does a [...] marketing strategy need?”
Output format specification	Not specified	“Describe to me the many advantages of electromobility.”
	Structured	“Create a cross-tabulation in R with n=300 people. Variable age is the [...]”
	Domain-specific	“Program a regression analysis with R. The dependent variable is EDU. The independent [...]”
Tone	Polite	“We focus on parcel services. Please list the needs and wishes of this target group.”
	Impolite	“That's boring. Please be more creative and something that rhymes.”
	Motivational	“Let's introduce a reward system for our conversation, every time we write good answers [...]”
	Neutral	“What is the difference between occupational disability and employment?”

Prompt Style

The prompt style meta-dimension classifies the way in which an operator specifies the style and form of an input-prompt in order to instruct the GLM to generate information.

The output format specification dimension describes whether an operator gives custom formatting instructions. The operator could ask the following question: How do I want the result to be represented? In this context, the output format specification can be unspecified (plain text). For example, the operator can choose not to pass a precise instruction to the system. Alternatively, the operator can request a generic structured format (e.g., tables, lists). In addition, the operator can instruct the system to deliver domain-specific formatting to match a desired style (e.g., persona, source code, reports).

The tone dimension applies to the writing style and linguistic tonality of the generated input-prompt. Here, the operator can ask the following guiding question: What tone do I use? The operator could specify certain tonality statements in the input-prompt, such as polite, impolite, motivational, or neutral. Examples for all characteristics can be found in Table 3.

12.4.2 Clustering and Derived Operator Archetypes

Three clusters were derived from the cluster calculations that were performed. They are shown in Table 3. The percentage distribution of each characteristic in its respective

dimension is illustrated independently from the clusters over the whole data set as well as per cluster. Each cluster was considered an archetype with its own properties, representing a potential usage type.

Archetype 1: Creative Learning

The Creative Learning archetype describes a GLM usage type with a high interest in using AI to (better) understand aspects. This archetype predominantly uses the creative focus of the GLM to receive information that has yet to be created, mostly by demanding that the GLM assessment function is applied to contexts that have been given. The assessment could aid in identifying connections and deepening understanding. Context is often given and references are made without specifying a format at all. In addition, the usage type shows a comparatively high tendency to give perspective, such as the organization-oriented one. The archetype shows a clear preference for neutral language use, only very rarely politely or impolitely elements were used.

Archetype 2: Operational Delegation

The Operational Delegation archetype shows specific usage behavior in relation to the GLM. This usage type mainly uses the GLM in a creative way and almost always uses it for task delegation. Therefore, the create functionality is used a lot and reformulation applied quite often. While formulating the inputs, a specific context is explicitly given or referenced in most cases. Matching the task delegation and a certain expert-oriented perspective, this archetype prefers to specify a domain-specific output format. It may be that operators using this type might view GLMs as a worker to whom they, when not neutrally delegating a task, quite sometimes do so politely.

Archetype 3: Scouting

The Scouting archetype uses GLM to build an initial understanding. It focusses on using AI as a search engine variant and knowledge source and is therefore mainly information focused. At the same time, the comparison to the other archetypes identified shows a rather high use of the GLM to hold general conversations. Prompts of this type often lack explicit context but sometimes reference previous context. In addition, they are aimed toward receiving general, existing, and retrievable information. They regularly do not specify a specific output format and use neutral wording in the vast majority of cases, in some instances a polite tone.

Table 3. Cluster Analysis Results considered as Archetypes

Meta-dimension	Dimension	Σ	Characteristics	Cluster 1	Cluster 2	Cluster 3
		n = 252		n = 103	n = 73	n = 76
	Focus	44%	Informational	34%	11%	91%

Prompt objectives		56%	Creative	66%	89%	9%
	Purpose	48%	Build understanding	66%	0%	70%
		13%	Sharpen understanding	23%	3%	8%
		35%	Task delegation	11%	96%	9%
		4%	General conversation	0%	1%	13%
	Intended functional output	46%	Create	1%	67%	87%
		4%	Aggregate	0%	4%	8%
		3%	Supplement	3%	5%	0%
		1%	Substitute	0%	3%	0%
		38%	Assess	93%	1%	0%
		2%	Compare	1%	0%	5%
		2%	Combine	2%	4%	0%
		4%	Reformulate	0%	15%	0%
Prompt settings	System perspective	15%	Humanized	17%	10%	16%
		4%	User-oriented	5%	1%	5%
		4%	Group-oriented	6%	0%	4%
		4%	Organization-oriented	8%	3%	0%
		4%	Expert-oriented	4%	8%	0%
		70%	None	60%	78%	75%
	Context	15%	Explicit context	28%	10%	4%
		37%	Reference context	39%	52%	21%
		14%	Reference previous + explicit context	14%	26%	3%
		33%	None	19%	12%	72%
Prompt style	Output format specification	83%	Not specified	100%	44%	99%
		1%	Structured	0%	1%	1%
		16%	Domain-specific	0%	55%	0%
	Tone	7%	Polite	1%	14%	9%
		2%	Impolite	1%	3%	3%
		1%	Motivational	0%	0%	3%
		90%	Neutral	98%	84%	86%

12.5 Discussion

Our taxonomy shows that the use of GLMs varies. On one hand, operators use the system to obtain information that is already known and can be delivered by GLMs without creating anything new. This may be a behavior carried over from typical search engine use. In GLM use, basic information requests are submitted, assuming no individual responses are provided. In this regard, our results reveal that a third of the input-prompts (33%) had no context given, underlining this assumption. Furthermore, considering our archetypes of GLM usage—especially for the Scouting archetype—the informational focus was predominantly geared toward build understanding without providing context, thus supporting our assumption. On the other hand, our results also show that more than half of the input-prompts (56%) were designed to focus on creative; hence, requesting new information that was yet to be created. This is interesting, considering that creativity has previously been considered a capability that humans can master while it is hard for computers (Chen et al., 2023). In addition, operators mostly use GLMs to build, sharpen their understanding (together 61%), and delegate tasks (35%), highlighting the need for clarification and education on the system and its weakness regarding providing false information (Dwivedi et al., 2023). This could create a new opportunity, as the system could be considered a team member used to perform specific tasks or assist in building understanding of specific subjects.

Regarding the intended functional output, we observed that the characteristics create (46%) and assess (38%) were used the most. In contrast, aggregate (4%), supplement (3%), substitute (1%), compare (2%), combine (2%), and reformulate (4%) were used less frequently. Furthermore, we show that although operators primarily engage with the system without providing a perspective (70%), some operators tend to humanize the system (15%), for example, by asking for its opinion on certain subjects. This observation is also reported in (Dev & Camp, 2020), which focuses on chat protocols. This buttresses our findings that often we could not observe that operators do distinguish between whether they are communicating with a human or an intelligent chatbot system. This is further reflected in the fact that we found polite phrases, e.g., saying please to the system.

Other perspectives found were user- (4%), group- (4%), organization- (4%), and expert-oriented (4%). The relatively rare usage indicates that although the option of providing a perspective is used. This option may still be relatively unknown among operators or not considered necessary, despite systems offering distinct options for specifying the perspective, e.g., ChatGPT (OpenAI, 2023). Reducing the lack of awareness could be a concrete example of the benefits of the proposed taxonomy when operators become fully aware of such possibilities. Only a third of the cases (33%) we investigated contained no context. Interestingly, this contextualization is addressed incidentally in previous studies on non-conversational GLMs, as they contemplated various possibilities of simulating such capabilities by chaining prompts (Wu et al., 2022). With that being absent per se in prior

non-conversational GLMs, it may be that this is a critical factor, because not all information needs to be included in a single prompt. Regarding the output format specification of an input-prompt, the possibility of such specification was not provided in most cases (83%), and when a format specification was given, it was almost exclusively domain specific (e.g., code, or a persona) (16%). Maybe the possibilities accompanying this dimension are either relatively unknown to operators or not considered to be beneficial. The last dimension identified was tone. The characteristic neutral tone was used in most cases (90%). Other tones such as being polite (7%), impolite (2%) or motivational (1%) occurred less. Those non neutral tones could be a result of overlearning in accordance with (Nass & Moon, 2000), as there too is described that participants applied behavior rules when interacting with computers. Determining whether the tone makes a difference would be an interesting new research topic for further investigation. We observed that complex problems (problems that typically require multi-step interactions) were broken down into single-step interactions. Since our taxonomy built on these observations and reflects these GLM usages, it could be referred to when breaking down problems into subtasks and formulating input-prompts for each, leading to a step-by-step development of work results. An example of such a process is to start with a prompt that asks general questions to build understanding, followed by a more in-depth approach that focuses on details using prompts that enable sharper understanding. Single-step examples are shown in Table 2.

In general, our taxonomy contains characteristics and dimensions similar to primitive operations introduced by (Wu et al., 2022). For instance, classification described in (Wu et al., 2022), requests the classification of the input into categories, which implies an assessment equivalent to our assess characteristic. Another example would be the factual query with which an operator requests facts, which is similar to a combination of characteristics in our case, where the focus is informational and the intended functional output is—for example—create. However, in our case, we emphasize that the information provided does not have to be a fact, but what the case displays is existing information. The generation and ideation could be mapped to a combination of creative and create (Wu et al., 2022). Although the prevailing “insufficient controllability” (Wu et al., 2022, p. 1) in non-conversational systems may be reduced, facets of controllability such as output format specification and a system perspective seem to still be rarely used (format specification used in 17% and system perspective used in 30% of all cases).

The evaluation of our taxonomy and its description reveal that the identified concepts offer a comprehensive view of GLM usage. Because no additional dimensions or characteristics were requested, we assume that our taxonomy is applicable to other GLM-based use, in addition to our problem solving scenario. This can be noted as a positive, considering that our data collection was performed in an environment in which solving a problem was intended. Participants in the evaluation spoke positively regarding the taxonomy we developed and expressed interest in the research results. In addition to the taxonomy, three different archetypes were derived based on cluster analysis, potentially representing an

initial state of different usage types of a (conversational) GLM. The derived archetypes indicate that the GLM was most used in our study to build an understanding of a specific problem or to outsource tasks to the system. Besides the three archetypes identified describing GLM usage types, future research could also explore potential correlations of usage types with different levels of domain expertise and GLM experience among operators. Following our archetypes, for example, the Scouting type could be related to Beginners that use the GLM predominantly for contextless, informational, and therefore search engine-like requests. The Creative Learning type could be more prevalent among people who already use the system's functionality to assess and follow more advanced strategies advanced in the use of GLMs. The Operational Delegation type may reflect users that have the most advanced knowledge of the system's capabilities, utilizing several of the possible dimensions to a significantly high degree. The archetypes with that could serve as a guide for practitioners, as well as researchers, for classifying users and designing customized offers based on the user's level of expertise.

12.5.1 Theoretical Contribution

We performed our study in an environment with limited academic knowledge on GLM usage, particularly with respect to using GLMs in multi-step interactions to solve complex problems. To the best of our knowledge, there are no findings that enable academics to have a common, comprehensive, and data- derived starting point that serves as common terminology when discussing or conducting further research on GLM use. Taxonomies, in general, can serve as a vocabulary in the investigated domain (Gerlach et al., 2022; Weking et al., 2020) and can be seen as a basis for the development of relevant theories (Kundisch et al., 2021b). Furthermore, taxonomies can serve an educational purpose (Miller & Roth, 1994).

Our taxonomy provides common knowledge based on analyzed input-prompts. Furthermore, the archetypes developed facilitate the classification of prompting behavior across different operators. Thus, we identified additional information that could not possibly be identified with a taxonomy alone (Gerlach et al., 2022) and used this to evaluate our taxonomy (Kundisch et al., 2021b). Both enables to investigate the usage of GLMs in multi-step real-world interactions. We envision the taxonomy and archetypes to facilitate research on prompting strategies or prompt engineering. For example, the taxonomy can help to identify changes in prompting approaches within and across interactions with GLM. This could support the empirically supported derivation of prompting strategies. Such future work would require researchers to classify prompts and characterize sequences of prompts in multi-step interactions. For this, researchers can now rely on the taxonomy as a common framework. Moreover, researchers who intend to use GLMs can be equipped with an efficient and comfortable introduction to GLM use. For DSR, we provide a useful starting point to work in GLMs because we set a common ground of structure, which could be used for instance as guideline for automated creation of prompts.

Although research regarding input-prompts has already shown some promising results (Wu et al., 2022), there is still limited research on conversational GLMs and using them to solve complex problems. With our study, we address this research gap and provide insights on conversational GLM usage; we also address the question of whether the change to multi-step conversations might reduce the need to specify all the relevant information required in a single prompt, as well as how conversational GLMs will be used. Furthermore, we demonstrate the use of GLMs across a diverse group of users and build a guiding framework for a analysis of such interactions. We also contribute to the existing body of knowledge of Wu et al. (2022). It turns out that the primitive operation classification has similar functionality to our identified assess function, so do the primitive operations factual query, generation and ideation as they represent similar functionality to our create characteristic. The primitive operation rewriting can be seen as a function similar to the identified reformulate characteristic and the split points operation represents functionality similar to our characteristic of specifying a structure. (Wu et al., 2022) However, we also identified new dimensions and characteristics (e.g. system perspectives like humanized, user-oriented, expert-oriented; context; purposes including understanding or task delegation and other intended functional outputs such as supplement, compare, combine). Thus, we complement the validity and transferability of the results.

Our findings also give a brief glimpse on possible human-AI collaboration with GLMs. Besides the characteristics of use we extracted, we could derive that humans utilize GLMs already in different ways, for instance by requesting delivery of information to understand or sharpen understanding of a given topic, and use them as some form of colleague to whom tasks are delegated and outsourced to. Such utilization of requesting information delivery and the creation role as well as the use as colleague to delegate tasks to aligns with other findings in this field of research (Kim et al., 2023; Siemon, 2022). In the field of hybrid intelligence, our findings show aspects of humans learning through the information provided by the GLM. As seen in the results, humans use GLMs potential advantages while working on solving complex problems. With the continuous transmission of context, steps of updating the GLM and by that educating it can also be observed. This depicts the third and central aspect of hybrid intelligence, i.e., continuous learning – showing that both humans and AI can learn from each other through experience (Dellermann et al., 2019). Future work on human-AI collaboration and hybrid intelligence can build on this when designing new studies and further investigating the collaboration.

12.5.2 Practical Contribution

Our results are a foundational work that might serve as a basis for practitioners to design prompts and prompting tactics to create better results from GLMs. For example, the proposed taxonomy might serve as a blueprint to structure and design conversations with GLMs for specific tasks (e.g., during complex problem solving). The taxonomy also supports GLM training by helping individuals to learn relevant constructs for designing input-prompts, and thus explore the capabilities of GLMs with greater ease-of-use. In

addition, our taxonomy can be used as a customization tool to improve input-prompts and generated outputs. In this regard, by passing clear instructions to the system, the GLM could be brought to respond to specific needs and expectations. Furthermore, we contribute to usage behavior research through the archetypes we have identified. Organizations can use these archetypes primarily to perform an initial assessment of their employees regarding their GLM usage behavior. Thus, organizations can assign individuals or groups of employees to their respective archetypes to understand the distribution of these behavior types within the organization, with the next step being to create type-appropriate measures for employees. For example, employees with little experience with GLMs could be encouraged to use the technology to build up their knowledge. These implications can also be applied to researchers and general users, as they can conduct a self-assessment based on the archetypes and thus learn about their usage behavior to better understand how they can potentially improve.

12.6 Limitations and Further Research

Despite these valuable contributions, our results have limitations and implications for further research. The resulting taxonomy is based on a multi-iterative development process based on 252 input-prompts and evaluation interviews. The overarching dimensions and characteristics proved to be valid starting points for exploring the usage of GLMs. Nevertheless, the taxonomy may does not comprise a fully exhausted set of (meta-)dimensions and characteristics. Therefore, further investigation is required to potentially identify additional (meta-)dimensions and characteristics that warrant a user-centric use of GLMs. Furthermore, the analyzed input-prompts did not always result in a clear possibility of classification within the proposed taxonomy, as the classification is based on the researchers' background. Moreover, applying the taxonomy does not inevitably guarantee success using GLMs. To achieve this, additional aspects, such as the non- deterministic nature of the system (e.g., the same input does not always produce the same output) and demonstrating limitations, need to be considered and further understood. Next, other factors, such as user AI acceptance (e.g., trust in AI technology, system transparency, and behavioral intention), must be considered because the sum of all interacting factors may determine the basis for successful GLM use. Therefore, future research should identify the acceptance criteria for using GLMs and analyze them in relation to each other to facilitate the implementation of appropriate (counter) measures to foster AI adoption.

The evaluation of the proposed taxonomy revealed several challenges that require further research. First, further research is required to understand how operators (theorists and practitioners) can develop various prompting strategies, including the objectives addressed, user-centric design, the inclusion of specific prompt settings and styles, and the relationships between input-prompts and generated outputs, to facilitate effective and efficient prompting. Second, we observe that switching behavior should also be considered in addition to user behavior. In this context, the relation to previous input-prompts needs

to be considered because GLMs like ChatGPT relate responses to previous inputs and generate outputs accordingly. Lastly, we acknowledge that the proposed taxonomy is not a static, but a living artifact. As GLMs potentially advance and the associated technical possibilities rise (e.g., ChatGPT3 vs. ChatGPT4), combined with the resulting need for further research, we assume that additional concepts will arise that foster an understanding of GLM usage in a narrower sense and prompt engineering in a more general sense.

Although our study provides valuable insights, it has methodological limitations that also offer ideas for future research. The analyzed data were collected in the context of a Prompt-a-thon, which had the purpose of solving the given challenges of certain user groups. This data collection environment may have influenced the observed GLM use; in addition, the prompts can be considered synthetic products of the study participants. However, through the evaluation, we were able to counteract these influences and confirm that the proposed taxonomy is comprehensive. Although our artifact has been shown to be applicable through validation and evaluation with various users, its transferability to other technologies such as video or graphic-based generative AI models is open to further research. In this vein, our taxonomy and derived archetypes could be a building block for deeper understanding GLM usage. Regarding the investigated GLMs, future research topics could be aimed toward explanatory and prescriptive knowledge about prompt engineering, prompting tactics, prompting strategies, and switching behavior.

12.7 Conclusion

Users, researchers, and organizations are experimenting with GLMs and strive to understand and explore the possibilities of this disruptive modern technology in order to build or sharpen their capabilities and support or automate processes, tasks, and activities so that it can contribute to their (business) value. However, despite the promising benefits associated with the current emergence of AI, fundamental knowledge in the use of GLM, especially for complex problem solving (Krogh, 2018), is still lacking. We addressed these problems by providing a fundamental user-oriented taxonomy. In a hackathon-like format called a Prompt-a-thon, 76 participants conducted complex problem solving tasks in groups using GLMs. The input-prompts derived were coded, analyzed, and aggregated iteratively into key concepts and formed into specific archetypes that describe the usage in such a problem solving environment. Further research is needed to investigate prompt engineering, prompting tactics, prompting strategies and behavioral switches of operators in GLM usage.

12.8 References

Akata, Z., Balliet, D., Rijke, M. d., Dignum, F., Dignum, V., Eiben, G., Fokkens, A., Grossi, D., Hindriks, K., Hoos, H., Hung, H., Jonker, C., Monz, C., Neerincx, M., Oliehoek, F., Prakken, H., Schlobach, S., van der Gaag, L., van Harmelen, F., . . .

- Welling, M. (2020). A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. *Computer*, 53(8), 18–28. <https://doi.org/10.1109/mc.2020.2996587>
- Arthur, D., & Vassilvitskii, S. (2007). K-Means++: The Advantages of Careful Seeding. *Proceedings of the eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 8.
- Balijepally, V., Mangalaraj, G., & Iyengar, K. (2011). Are We Wielding this Hammer Correctly? A Reflective Review of the Application of Cluster Analysis in Information Systems Research. *Journal of the Association for Information Systems*, 12(5), 375–413. <https://doi.org/10.17705/1jais.00266>
- Berkhin, P. (2006). A Survey of Clustering Data Mining Techniques. In J. Kogan, C. Nicholas, & M. Teboulle (Eds.), *Grouping Multidimensional Data* (pp. 25–71). Springer-Verlag. https://doi.org/10.1007/3-540-28349-8_2
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., . . . Amodei, D. (2020). Language Models are Few-Shot Learners. *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 33, 1877–1901.
- Buscemi, N., Hartling, L., Vandermeer, B., Tjosvold, L., & Klassen, T. P. (2006). Single data extraction generated more errors than double data extraction in systematic reviews. *Journal of clinical epidemiology*, 59(7), 697–703. <https://doi.org/10.1016/j.jclinepi.2005.11.010>
- Chen, L., Sun, L., & Han, J. (2023). A Comparison Study of Human and Machine Generated Creativity. *Journal of Computing and Information Science in Engineering*. <https://doi.org/10.1115/1.4062232>
- Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid Intelligence. *Business & Information Systems Engineering*, 61(5), 637–643. <https://doi.org/10.1007/s12599-019-00595-2>
- Dev, J., & Camp, L. J. (2020). User Engagement with Chatbots: A Discursive Psychology Approach. *Proceedings of the 2nd Conference on Conversational User Interfaces*, Article 52. <https://doi.org/10.1145/3405755.3406165>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L.,

- Buhalis, D., . . . Wright, R. (2023). “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Gerlach, J., Werth, O., & Breitner, M. (2022). Artificial Intelligence for Cybersecurity: Towards Taxonomy- based Archetypes and Decision Support. *Proceedings of the 43rd International Conference on Information Systems*.
- Gero, K. I., Liu, V., & Chilton, L. (2022). Sparks: Inspiration for Science Writing using Language Models. *Designing Interactive Systems Conference*, 1002–1019. <https://doi.org/10.1145/3532106.3533533>
- Gregor, S., & Hevner, A. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, 37, 337-356. <https://doi.org/10.25300/MISQ/2013/37.2.01>
- Hambardzumyan, K., Khachatrian, H., & May, J. (2021). WARP: Word-level Adversarial ReProgramming. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 4921– 4933. <https://doi.org/10.18653/v1/2021.acl-long.381>
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics*, 28(1), 100. <https://doi.org/10.2307/2346830>
- Janssen, A., Passlick, J., Rodríguez Cardona, D., & Breitner, M. H. (2020). Virtual Assistance in Any Context. *Business & Information Systems Engineering*, 62(3), 211–225. <https://doi.org/10.1007/s12599-020-00644-1>
- Jiang, E., Olson, K., Toh, E., Molina, A., Donsbach, A., Terry, M., & Cai, C. J. (2022). PromptMaker: Prompt- based Prototyping with Large Language Models. In S. Barbosa, C. Lampe, C. Appert, & D. A. Shamma (Eds.), *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (pp. 1–8). ACM. <https://doi.org/10.1145/3491101.3503564>
- Kim, T., Molina, M. D., Rheu, M., Zhan, E. S., & Peng, W. (2023). One AI Does Not Fit All: A Cluster Analysis of the Laypeople’s Perception of AI Roles. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3544548.3581340>
- Krogh, G. v. (2018). Artificial Intelligence in Organizations: New Opportunities for Phenomenon-Based Theorizing. *Academy of Management Discoveries*, 4(4), 404-409. <https://doi.org/10.3929/ethz-b- 000320207>
- Kundisch, D., Muntermann, J., Oberländer, A. M., Rau, D., Röglinger, M., Schoormann, T., & Szopinski, D. (2021a). Appendix: An Update for Taxonomy Designers. *Business & Information Systems Engineering*, 64(4), 421–439. <https://doi.org/10.1007/s12599-021-00723-x>

- Kundisch, D., Muntermann, J., Oberländer, A. M., Rau, D., Röglinger, M., Schoormann, T., & Szopinski, D. (2021b). An Update for Taxonomy Designers. *Business & Information Systems Engineering*, 64(4), 421–439. <https://doi.org/10.1007/s12599-021-00723-x>
- Lee, M., Liang, P., & Yang, Q. (2022). CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities. *Proceedings of the 2022 CHI conference on human factors in computing systems*, 11, 1–19. <https://doi.org/10.1145/3491102.3502030>
- Lester, B., Al-Rfou, R., & Constant, N. (2021). The Power of Scale for Parameter-Efficient Prompt Tuning. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 3045–3059. <https://doi.org/10.18653/v1/2021.emnlp-main.243>
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Computing Surveys*, 55(9), 1–35. <https://doi.org/10.1145/3560815>
- McDonald, N., Schoenebeck, S., & Forte, A. (2019). Reliability and Inter-rater Reliability in Qualitative Research. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1–23. <https://doi.org/10.1145/3359174>
- Miller, J. G., & Roth, A. V. (1994). A Taxonomy of Manufacturing Strategies. *Management Science*, 40(3), 285–304. <https://doi.org/10.1287/mnsc.40.3.285>
- Mishra, S., Khashabi, D., Baral, C., Choi, Y., & Hajishirzi, H. (2022). Reframing Instructional Prompts to GPTk's Language. *Findings of the Association for Computational Linguistics: ACL 2022*, 589–612. <https://doi.org/10.18653/v1/2022.findings-acl.50>
- Mishra, S., Khashabi, D., Baral, C., & Hajishirzi, H. (2022). Cross-Task Generalization via Natural Language Crowdsourcing Instructions. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*
- Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56, 81–103. <https://doi.org/10.1111/0022-4537.00153>
- Natalie. (2023). *ChatGPT — Release Notes*. OpenAI. <https://help.openai.com/en/articles/6825453-chatgpt-release-notes>
- Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems*, 22(3), 336–359. <https://doi.org/10.1057/ejis.2012.26>
- Olesen, J. F., & Halskov, K. (2020). *10 Years of Research With and On Hackathons* Proceedings of the 2020 ACM Designing Interactive Systems Conference, Eindhoven, Netherlands. <https://doi.org/10.1145/3357236.3395543>
- OpenAI. (2023). *Playground*. <https://platform.openai.com/playground>
- Peppers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of*

- Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/mis0742-1222240302>
- Prat, N., Comyn-Wattiau, I., & Akoka, J. (2015). A Taxonomy of Evaluation Methods for Information Systems Artifacts. *Journal of Management Information Systems*, 32(3), 229–267. <https://doi.org/10.1080/07421222.2015.1099390>
- Punj, G., & Stewart, D. W. (1983). Cluster Analysis in Marketing Research: Review and Suggestions for Application. *Journal of Marketing Research*, 20(2), 134. <https://doi.org/10.2307/3151680>
- Siemon, D. (2022). Elaborating Team Roles for Artificial Intelligence-based Teammates in Human-AI Collaboration. *Group Decision and Negotiation*, 31(5), 871–912. <https://doi.org/10.1007/s10726-022-09792-z>
- Swanson, B., Mathewson, K., Pietrzak, B., Chen, S., & Dinalescu, M. (2021). Story Centaur: Large Language Model Few Shot Learning as a Creative Writing Tool. In D. Gkatzia & D. Seddah (Eds.), *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations* (pp. 244–256). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.eacl-demos.29>
- Taylor, N., & Clarke, L. (2018). Everybody's Hacking: Participation and the Mainstreaming of Hackathons. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Paper 172. <https://doi.org/10.1145/3173574.3173746>
- Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301), 236–244. <https://doi.org/10.1080/01621459.1963.10500845>
- Weking, J., Stöcker, M., Kowalkiewicz, M., Böhm, M., & Krcmar, H. (2020). Leveraging industry 4.0 – A business model pattern framework. *International Journal of Production Economics*, 225, 107588. <https://doi.org/10.1016/j.ijpe.2019.107588>
- Wu, T., Terry, M., & Cai, C. J. (2022). AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. In S. Barbosa, C. Lampe, C. Appert, D. A. Shamma, S. Drucker, J. Williamson, & K. Yatani (Eds.), *CHI Conference on Human Factors in Computing Systems* (pp. 1–22). ACM. <https://doi.org/10.1145/3491102.3517582>
- Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023). Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In A. Schmidt, K. Väänänen, T. Goyal, P. O. Kristensson, A. Peters, S. Mueller, J. R. Williamson, & M. L. Wilson (Eds.), *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1–21). ACM. <https://doi.org/10.1145/3544548.3581388>
- Zhu, Q., & Luo, J. (2022). Generative Pre-Trained Transformer for Design Concept Generation: An Exploration. *Proceedings of the Design Society*, 2, 1825–1834. <https://doi.org/10.1017/pds.2022.185>

13 Conversational Agent as a Black Hat: Can Criticizing Improve Idea Generation?

Cvetkovic, I., Rosenberg, V., & Bittner, E. (2023). Conversational Agent as a Black Hat: Can Criticizing improve Idea Generation? In *Proceedings of the 56th Hawaii International Conference on System Sciences (HICSS)*. Kaanapali, HI, USA.

Abstract

The Ideate phase of Design Thinking is the source of many idea creations. In this context, criticism is considered a creativity killer, yet recent studies show that criticism can be beneficial. An example of this is the black hat of one creativity method: Six Thinking Hats. It points out the weaknesses of an idea so that they are eliminated by further refining. Previous research shows that conversational agents have an advantage over humans when criticizing because of their perceived neutrality. To investigate this, we developed and implemented a conversational agent and evaluated it using an A/B test. The results of the study show that the prototype is perceived as less neutral when it criticizes. Criticizing by the conversational agent can lead to higher quality ideas. This work contributes to a better understanding of conversational agents in the black hat role as well as of their neutrality.

13.1 Introduction

Be it global warming, pollution, financial crises or pandemics: We are confronted with pollution, financial crises or pandemics: We are confronted with constant changes in society, technology, our coexistence and our cooperation. These factors present us with major challenges and complex problems that need to be overcome; problems that require innovative solutions. The knowledge of individuals is often insufficient in this regard (Brown, 2019), all the more necessary to work in heterogeneous, interdisciplinary groups (Schallmo and Lang, 2020). Especially when it comes to collaborative innovation, Design Thinking (DT) has proven its worth as an approach for interdisciplinary collaboration in creative processes (Tschimmel, 2012; Brown, 2019). But for an efficient and successful implementation of DT and for successful collaboration, facilitation is important (Oxley et al., 1996). Skills and experience of the person facilitating are decisive for the success of a creative collaboration. To relieve the facilitators, or to meet the high demand of these, (partial) tasks can be performed by Conversational Agents (CAs) - intelligent software that can understand and use natural language in written or spoken form (Bittner et al., 2021). Several studies have already demonstrated the successful use of CAs in supporting creativity (Tavanapour et al., 2019; Tavanapour and Bittner, 2018; Bittner et al., 2019b). Through developments in Machine Learning (ML) and artificial Intelligence (AI), CAs are becoming increasingly capable in processing and responding to user input (Seeber et al.,

2018). However, many characteristics have not yet been explored (Bittner et al., 2021; Seeber et al., 2018). In particular, there is a need for CAs to appear more human, e.g., by responding empathically (Debowski et al., 2021), or knowing how to deal with creative context (Tavanapour et al., 2019). On the other hand, the very lack of humanity of the CA, in the form of neutrality, can be seen as a strength and advantage compared to the human moderators (Debowski et al., 2021; Tavanapour et al., 2019; Tavanapour and Bittner, 2018). In one study, a CA repeatedly prompts participants to make a decision during a discussion. While these prompts were perceived as neutral by the participants, the same prompts, coming from a moderator, were evaluated negatively. This observation shows an unexplored potential of moderating CAs in DT (Tavanapour et al., 2019). Moreover, in the context of ideation, there is a need for CAs that can contribute new impulses and inspirations and promote different directions in ideation (Bittner et al., 2021). In this context, the CA could not only provide support but also act as an equal collaboration partner (Seeber et al., 2018) and contribute its own perspectives (Debowski et al., 2021). An application scenario that includes all the aspects mentioned above to use a CA is the Six Thinking Hats method (De Bono, 1985). This is a creativity technique in which a group methodically adopts six different thinking perspectives in search of a solution to a problem. The hats of the colors blue, white, red, yellow, green and black each symbolize a direction of thinking. In order to utilize the expertise of all participants in each direction, during a discourse each participant should represent each hat, i.e., each way of thinking and role, in turn (De Bono, 1985). For example, the blue hat is responsible for organizing and controlling all other hats. White is neutral and objective and is focused on facts. The green hat represents creativity and new ideas, and the black hat is logically negative and points out the weaknesses of an idea. Building on this method, a CA with the goal of a balanced team, could analyze a group as they work together in terms of roles or points of view in order to take a missing perspective (Debowski et al., 2021).

With this study, the so far not investigated impact of criticism by a CA in the context of an idea generation process is explored. For this purpose, a CA prototype is designed and implemented so that it can be investigated whether it can successfully criticize ideas. Successfully means that the criticism leads to rethinking and submitting new ideas or even that newly submitted ideas are qualitatively better. In the course of this, the effect of the CA's criticism will be explored with a focus on its neutrality. To achieve this goal, this work is guided by the following research questions: RQ 1. How can the neutrality of the CA be tested and exploited using criticism towards the user? RQ 2: What effect does the developed CA have on the users with regard to neutrality and criticism in the idea generation process?

The paper is structured as follows: Related Work provides insight into CAs, DT, and CAs in DT. The Method section describes the concept of the CA prototype comprising of the requirements analysis, design principles, and the dialog scenario. Thereafter, the study design is presented, and the development of the prototype is detailed. To answer the RQ 2,

the results of the study are summarized, followed by their explanation, the contribution and limitations of this work, and recommendations for further research.

13.2 Related Work

13.2.1 Conversational Agents

"I propose to consider the question, 'Can machines think?'" (Turing, 2009) - with this first sentence of the 1950 essay *Computing Machinery and Intelligence*, Alan Turing heralds the era of AI. The "Imitation Game" described in it, known today as the Turing Test, is intended to help determine whether a machine can imitate human intelligence and answer questions in such a way that other people are convinced that it is also human (Turing, 2009). In 1966, ELIZA becomes the first chatbot to manage to give users the feeling that they are having a conversation with a human (Weizenbaum, 1966). Although chatbots have been developed since the dawn of computers, everyday use. Only through ML and natural language processing (NLP), the development of chatbots achieves its breakthrough (Boonstra, 2021). In addition to chatbots, other terms such as virtual agents, digital agents, dialog systems, and conversational agents are also used synonymously (Bittner et al., 2019b).

NLP is a subfield of AI and aims to transform unstructured speech data into a structured format in such a way that machines can understand and respond or otherwise react to the text and the relevant information it contains. NLP can be further divided into Natural Language Understanding (NLU) and Natural Language Generation (NLG). NLU uses grammar and context to determine the intended meaning of a sentence, while NLG is concerned with the machine generation of text based on a given set of data (Kavlakoglu, 2020). Today, most people carry a CA in their smartphone (e.g., Google Assistant or Siri) or have one installed at home (e.g., Alexa from Amazon Echo) (Boonstra, 2021). Initially, companies focused on developing their own AI-driven CAs. More recently, by disclosing their bot platforms, companies such as Telegram, Slack, and Facebook are enabling external developers to create their own CAs to provide application extensions such as polling, integrations, and entertainment (Khan and Das, 2018). As the number of Internet users continues to increase, so does the generation of data that CAs can analyze and use to extract information (Khan and Das, 2018). Due to the current continuous development of ML, NLP and AI, more and more data and the still unexploited application areas of CAs, it can be concluded that the technology has both relevance and high potential.

13.2.2 Design thinking

Design plays a role in various areas of work and life. Buchanan (1992) defines four basic categories: Design of symbolic and visual communication, Design of material objects, Design of activities and organized services, and Design of complex systems or environments for living, working, playing, and learning. All of these areas contain complex

problems requiring solutions (Chasanidou et al., 2015). DT can be defined as a process for solving such problems (Schallmo and Lang, 2020), or as a way of thinking that, starting from such problems, can lead to "transformation, evolution and innovation, to new ways of living and new ways of doing business" (Tschimmel, 2012). Brown (2019) describes three areas that are critical on the path to innovation: viability, desirability, and feasibility. Specifically, viability refers to feasibility from the company's perspective. The perspective of the users and the associated wishes and needs are covered by desirability. Feasibility refers to feasibility from a technological perspective. If all three areas have been considered, it improves the prospect of innovation (Brown, 2019; Chasanidou et al., 2015). There are different models for the DT process, such as the Hasso Plattner Institute (HPI) model consisting of the six phases Understand, Observe, Point of View, Ideate, Prototype, and Test (Thoring and Müller, 2011). The connections between the phases imply the possibility of proceeding iteratively and, if necessary, repeating phases that have already been gone through.

13.2.3 Conversational Agents in Design Thinking – With AI to better ideas

Like the Ideate phase of the HPI DT model, all other DT models also include a phase for generating ideas (Tschimmel, 2012). In this phase, ideas are derived from the previously determined needs of the users with the help of creativity techniques, then grouped and revised, and finally described and evaluated (Schallmo and Lang, 2020). How one can benefit from AI in the ideate phase can be shown by the example of a CA that ensures that elaborated ideas have a consistent structure, sufficient details, and a comprehensive description by encouraging users to add to them (Tavanapour and Bittner, 2018). This is because ideas often lack sufficient information, so that they score lower than other ideas in the subsequent evaluation due to the lack of presentation (Bittner and Shoury, 2019). The evaluation of ideas by CAs turns out to be more difficult than the support of idea generation and revision. Maher and Fisher (2012) take a first step in this direction by presenting an AI-based approach to evaluate ideas with respect to the criteria of novelty, value, and surprise. As in DT in general, the ideation process is not simply the activity of a single person, but a collaborative creation of new and innovative solutions in a group (Stockleben et al., 2017). To foster creativity in collaboration, brainstorming rules and techniques include e.g., "be visual", "avoid criticism", "build on others' ideas", "allow for unconventional ideas", and "focus on quantity" (Thoring and Müller, 2011). A good example of how a CA can facilitate such creative collaboration is brAInstorm, a web-based tool for collective digital brainstorming (Strohmann et al., 2017). The CA guides the participating individuals first through an individual idea generation and then through a collective idea evaluation. Additionally, the CA intervenes when one of the participants uses phrases that can hinder a brainstorming process ("killer phrases"), become rude, or talk too much (Strohmann et al., 2017). On the one hand, avoiding criticism is one of the most important rules of brainstorming with the aim of preventing evaluation apprehension

(Siemon, 2022); on the other hand, it contradicts the selection process in the ideate phase, because for this an evaluation of the ideas is inevitable (Thoring and Müller, 2011). Moreover, the study by Tanaka et al. (2015), for example, shows that criticism can even lead to better quality ideas. They ensure that no evaluation apprehension is created by anonymity of the participants. Furthermore, Siemon et al. (2015) show that negative effects that can arise in test subjects through collaboration with other people, such as evaluation apprehension, do not occur when the collaboration takes place with an AI-based system instead.

13.3 Method

This study comprises the development, demonstration, and evaluation of a CA prototype following the principles of Design Science Research (DSR) methodology according to Peffers et al. (2007). Hereby, a technical artefact (CA prototype) is developed as an objective of a solution to a previously identified problem of scarcely investigated impact of criticism by a CA in the context of idea generation. In the demonstration phase, the artefact is used to employ criticism. By employing a survey after the demonstration, we evaluate the impact of criticism on idea generation.

13.3.1 Design Requirements & corresponding Design Principles

To create a concept for a CA, first, a requirements analysis is performed. Based on this, design principles are derived. In the further course, the design principles serve as orientation for the creation of the dialog scenario in order to subsequently create a basis for the development of the CA prototype. All three steps are covered in this chapter.

19 requirements were derived based on a representative literature review (Cooper, 1988). The requirements were categorized into general requirements, and criticism-specific requirements. In each of these categories, the requirements were considered in terms of goal-oriented behavior and interaction-oriented behavior according to (Tavanapour and Bittner, 2018).

To ensure goal-oriented behavior, general user-friendliness (R1) should be ensured through clear and easy-to-understand interaction (Bittner and Shoury, 2019). This can be realized by communicating all available functions, or by presenting commands by clickable buttons instead of free text fields (Bittner and Shoury, 2019). Furthermore, the interaction with the CA can be terminated at the request of the user and there is an option for continuous support (Bittner and Shoury, 2019). Efficient communication (R2) should ensure that the needs of the user are met with minimal effort. To this end, the CA should have the shortest possible response times (Tavanapour and Bittner, 2018; Debowski et al., 2021), and help the users in as few steps as possible through optimized dialog structures (Bittner and Shoury, 2019). To make this possible, the CA could communicate the expected type of answer and check the input received from the user for correctness (Bittner and Shoury, 2019). In order not to

discourage the users with long texts, short, simple messages (R3) should be implemented where possible (Tavanapour and Bittner, 2018). By displaying good spelling and grammar skills (R4), the CA should appear as intelligent and serious as possible (Tavanapour and Bittner, 2018). The CA must be able to recognize and respond to the user's intention (R5) (Radziwill and Benton, 2017). If the intention of the user is not understandable, the CA should respond appropriately, e.g., by asking for more information (Bittner and Shoury, 2019). Possible misentries and dead ends should also be considered and prevented if possible (Radziwill and Benton, 2017). At the beginning of the conversation, the CA should greet the user, introduce the interaction functions (R6), and explain how to interact with it (Lazarevich, 2017, 2018).

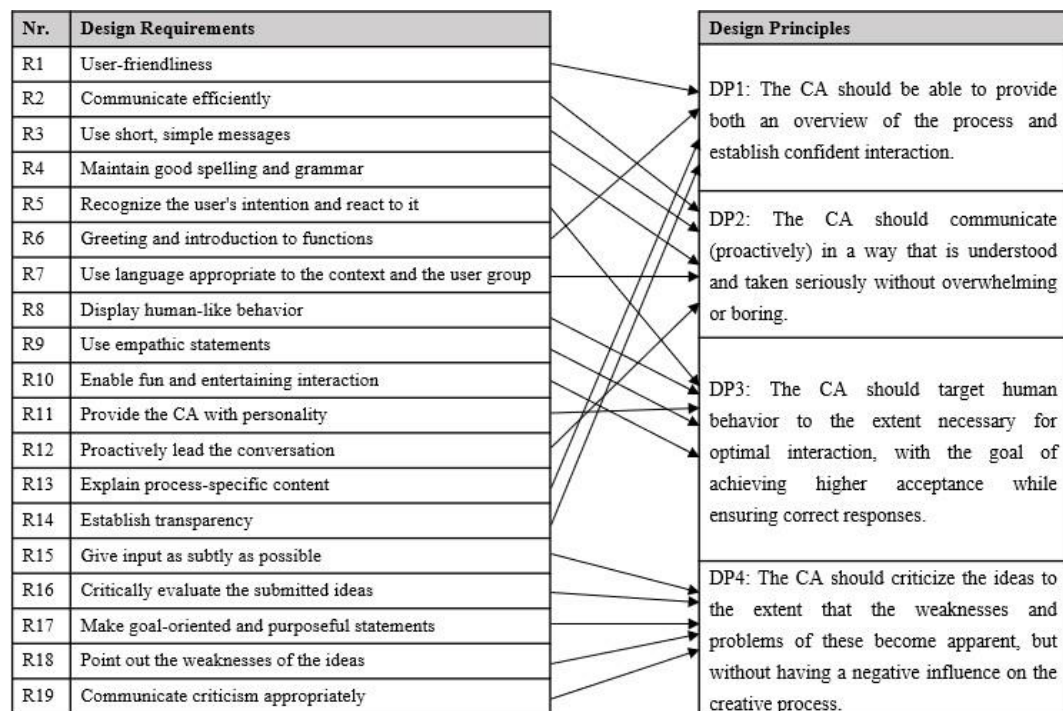


Figure 1. Design Requirements and Principles.

To achieve interaction-oriented behavior, the CA should use language appropriate to the context and user group (R7), with an appropriate level of formality and a good vocabulary (Radziwill and Benton, 2017; Bittner et al., 2019a; Tavanapour and Bittner, 2018). Human-like behavior should be aimed for (R8), such as politeness (Tavanapour and Bittner, 2018). However, the focus should not be on imitating humans, but on the successful support of the users by the CA (Radziwill and Benton, 2017). The CA should make empathic statements (R9), be supportive and trusting, and respond to the users moods (Tavanapour and Bittner, 2018; Debowski et al., 2021; Radziwill and Benton, 2017). In order for the users to enjoy and be entertained by the interaction (R10), the conversation with the CA should be as interesting and pleasant as possible (Tavanapour and Bittner, 2018; Radziwill and Benton, 2017). E.g., the CA should know several formulations for the same substantive statement in response to a question or use synonyms (Bittner and Shoury, 2019). With an appropriate personality of the CA (R11), it should be possible to establish an emotional connection to

the user (Bittner and Shoury, 2019). For example, the CA can use functions or statements personalized to the user, or to trigger emotions, emojis can be used (Radziwill and Benton, 2017).

Regarding idea generation, the CA should not only respond to commands but proactively lead the conversation (R12) (Tavanapour and Bittner, 2018; Bittner and Shoury, 2019; Debowski et al., 2021), e.g., by asking relevant questions, providing conversation prompts, or initiating process-related activities (Radziwill and Benton, 2017). The CA should be able to explain both the goal of the collaboration and process-specific content (R13). For example, if the CA asks the user to specify something, it should be able to explain the level of detail of the specification (Tavanapour and Bittner, 2018), or it could give examples of similar ideation tasks (Bittner et al., 2019a). To strengthen users' expectations, the CA should establish transparency (R14) and communicate why it is used and how interaction can improve idea generation (Bittner and Shoury, 2019; Debowski et al., 2021; Radziwill and Benton, 2017).

Regarding criticism, following requirements were identified: To ensure that the user's creativity remains in the foreground and that the creative flow is not interrupted, the CA should provide input as subtly as possible (R15) (Debowski et al., 2021; Bittner et al., 2021). The CA should critically question the user's ideas (R16), or help the user to question the ideas themselves. In doing so, the user should be encouraged to consider the idea from different perspectives and to test the acquired information for validity (Bittner and Shoury, 2019; Wechsler et al., 2018). The CA should improve the quality of idea elaboration by making purposeful statements (R17) and proactively guiding the conversation (Tavanapour and Bittner, 2018). When criticism is offered, the CA should primarily point out the weaknesses of the ideas (R18) so that these weaknesses can be eliminated-in the spirit of the black hat role of the Six Thinking Hats method (De Bono, 1985). It must be justified in a comprehensible way why the idea is not suitable (De Bono, 1985). In order to prevent possible negative influence on the creativity of the user through criticism and, e.g., to avoid evaluation apprehension (Tanaka et al., 2015), criticism must be communicated (R19) appropriately (Verganti, 2016).

These requirements and corresponding principles are presented in the Fig. 1. The dialog scenario is based on the design principles and reflects all goal-oriented process steps between the CA and the user.

13.3.2 Study design

To answer the RQ 1, a pilot study design is developed. The procedure of the pilot study is shown in Figure 2. First, the prototype is implemented and then a pre-test with 5 participants is conducted. In the second step, the CA is enhanced with the training data and any malfunctions that occur are corrected. In order to investigate the effect of the criticism of the CA on the ideation process, an A/B test is performed. A/B test is a controlled experiment which allows to establish causal relationships with high probability provided

enough statistical power (Kohavi and Longbotham, 2017). With the improved CA, the first part of the A/B test is performed with group A. The ideas from the pre-test and the A test are used to implement the criticism function in the step 3. After the second part of the A/B test is conducted, the ideas from the two groups are combined for idea evaluation. In addition, a survey is conducted with the subjects of the A/B test.

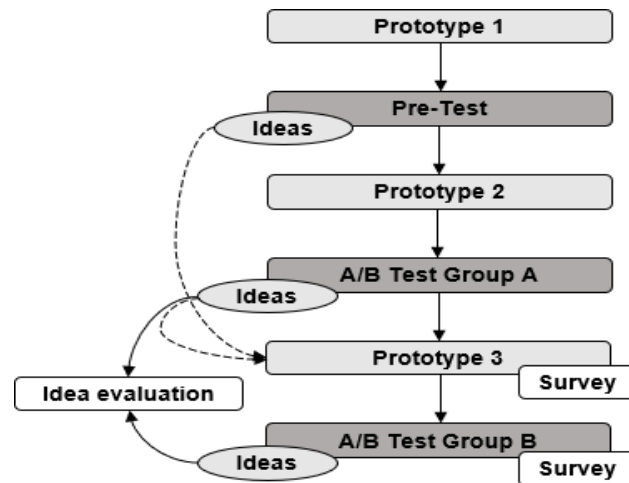


Figure 2. Study design.

The sequence of the idea generation process was based on the study design of Tanaka et al. (2015). In addition, the task for the idea generation process, the idea template, and the reference idea were also adopted from the same study. The task contains a problem for which the subjects should think of an idea for a solution. The problem is the following: "There is a great deal of information on the Internet, including false, misleading and unsubstantiated information. What can we do to avoid or reduce the negative impact of misinformation on Internet users?". The idea was to be formulated as a supplement to the idea template. The reference idea later served as an orientation during the expert idea evaluation in order to make the different evaluation tendencies of the idea evaluators comparable.

To conduct the A/B tests, 20 subjects, aged 20-55, were randomly recruited. They were first asked about their experience with CAs. Based on the answers, the subjects were divided into two similarly distributed groups of ten persons each. Group A consists of one CA and ten subjects. A separate test is conducted with each subject. This test consists of an interaction with the CA. The subject is presented with a problem by the CA and is then asked to submit an idea for it by completing the idea template. After submitting the idea, the subject is given the opportunity to reconsider and change it (dummy critique). The dummy critique is used to give subjects the opportunity to change their idea so that the conversation flow with the CA is as similar as possible between Group A and Group B. Like group A, group B also consists of one CA and ten subjects. The interaction is almost identical to that in group A. The only difference is that after the idea has been submitted, the subject from group B is criticized by the CA with regard to the idea submitted. Afterwards, the subject is also given the opportunity to reconsider and change the submitted idea.

Thereupon, the experts were asked to rate all ideas (which emerged in the A/B test) in terms of novelty and practicability (scale 1-7 with 1=not at all, 7=very much), taking into account the reference idea. The experts are fairly familiar with DT and ideation through perennial work experience.

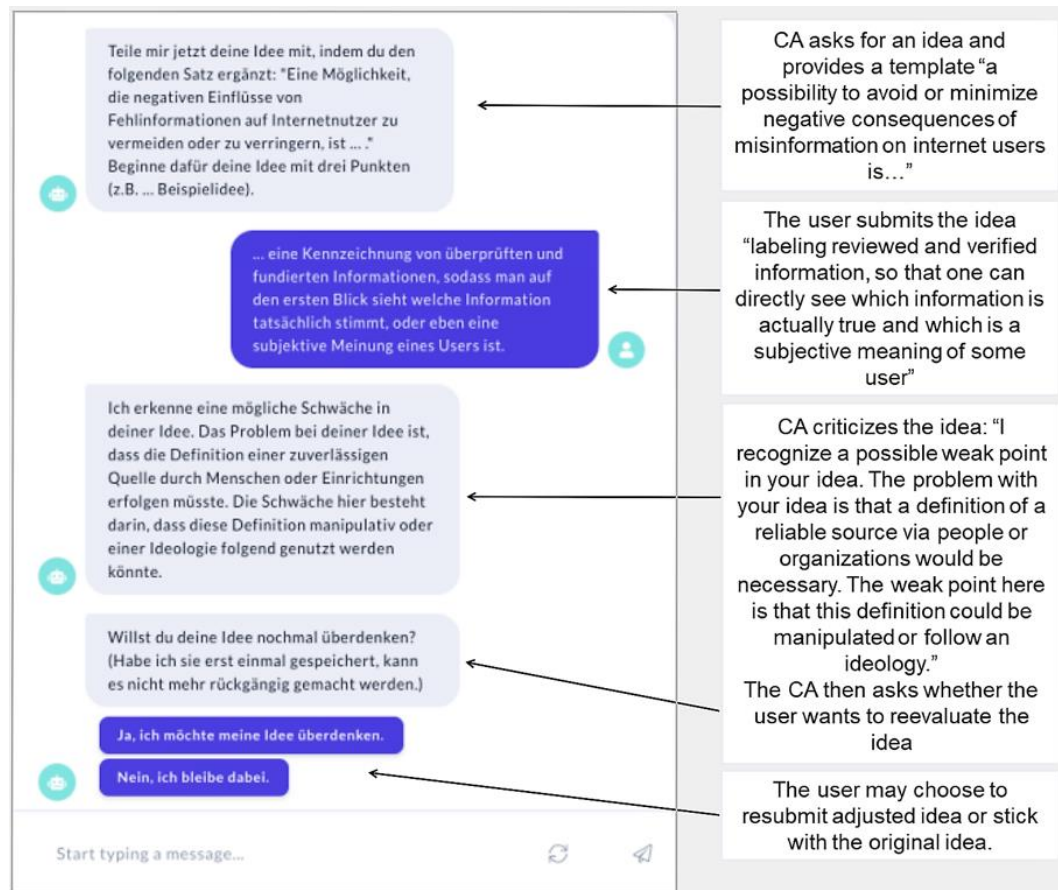


Figure 3. Snippet of the implemented critique.

In addition, slightly rephrased ideas were added to the training data to improve the CA's understanding. If the CA can categorize an idea, then it uses the critique that matches it (see example of critique to Sources in Figure 3). In case it recognizes the idea as such but cannot assign it to a category, it resorts to the Fallback category. For this purpose, a critique has been defined that is general enough to apply to most ideas related to the problem.

13.3.3 Implementation of the Criticism

The prototype CA was developed with open source NLP framework RASA (Rasa Technologies Inc, 2021). In order for the CA to be able to criticize ideas, it must be equipped with several skills. First, it must correctly understand the user's intention - to submit an idea. Then it must be able to analyze the idea and place it in the context of the problem so that it can identify possible weaknesses in the idea. Finally, it must express the criticism in a plausible and comprehensible way. This should awaken the motivation to rethink the idea and change it in such a way that the criticized weakness is eliminated. To mimic this approach for the study, the following procedure was followed: First, the problem

was defined, which narrowed down the context. From the pre-test and the A/B test of group A, the submitted ideas were collected. The ideas were then grouped and categorized. Four categories emerged: Sources, Education, Algorithm, and Fallback. For the first three categories, a critique was defined to be vague enough to fit as many proposed ideas in a category as possible while being concrete enough to be perceived as plausible by the subject.

The CA was trained with the example ideas (utterances) so that it could match the ideas to one of the three categories using different keywords (entities).

13.4 Results

13.4.1 User Evaluation of the prototype

Following the interaction with the CA, participants from both groups filled out a survey. The questions in the survey serve the evaluation of previously derived design principles. Apart from the answers shown in Figure 5, participants from both groups rated the CA as not boring (average rate 4,1, on a scale of 1-5, with 1=Very boring, 5=Not at all boring). Furthermore, participants felt that they knew how to communicate with the CA during the interaction (average rate 4,3, on a scale of 1-5, with 1=I felt totally lost, 5=I knew what to do). In terms of human-likeness, the average rating was 3,2 (scale 1-5, with 1=not at all human-like, 5=very human-like).

Looking at the ratings in comparison between group A and group B, we can see that the mean values show a tendency towards a different perception only for the question about neutrality.

13.4.2 Impact on criticism on Idea Generation

While none of the subjects in Group A revised their idea after submitting it to the CA, four out of ten in Group B did so. All ideas were then evaluated by innovation experts regarding novelty and practicability. The ratings of novelty and practicability presented in Figure 5 are composed of the mean value of all subjects in the respective group. The mean value was also used for the overall ratings in each case. Furthermore, a distinction is made between group B (ideas V1) and group B (ideas V2). In group B (ideas V1), the first version of the idea was included in the calculation. This means that if a subject was criticized and subsequently changed his or her idea, the change was ignored. Group B (Ideas V2), on the other hand, always included the last version of the idea. This distinction allows a comparison before and after criticism.

A comparison before and after the criticism shows that it is only after the criticism that group B tends to score better than group A in the overall evaluation. Furthermore, it can be seen that as a result of the criticism, the ratings for novelty and practicability are slightly

higher, which also shows a tendency towards improvement here. Furthermore, it can be seen that all groups perform better with regard to novelty than with regard to practicability.

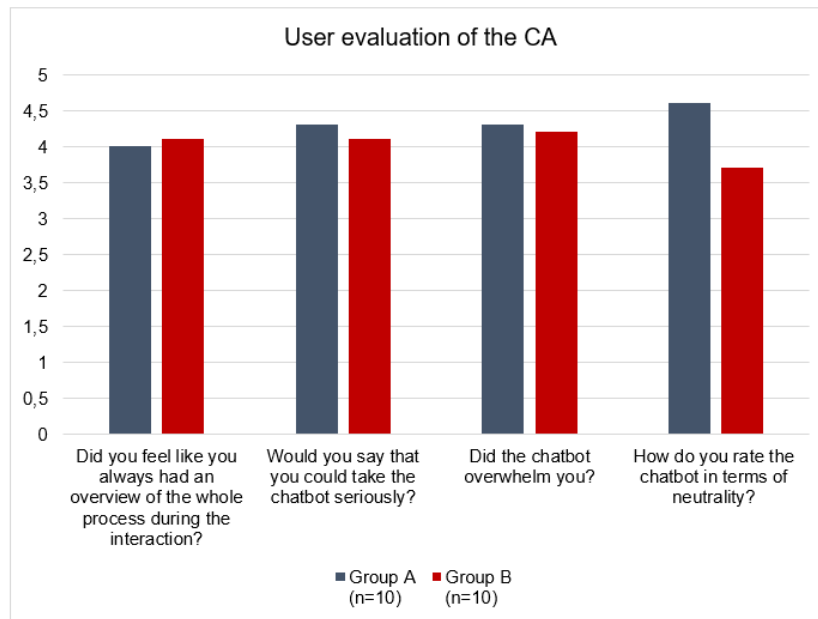


Figure 4. Answers from the post-experiment survey.

Scales left to right: 1=I felt totally lost, 5=I knew what to do; 1=I could not take it seriously at all, 5=I could take it absolutely seriously; I was totally overwhelmed, 5=I was not overwhelmed at all; 1=Not at all neutral, 5=Very neutral

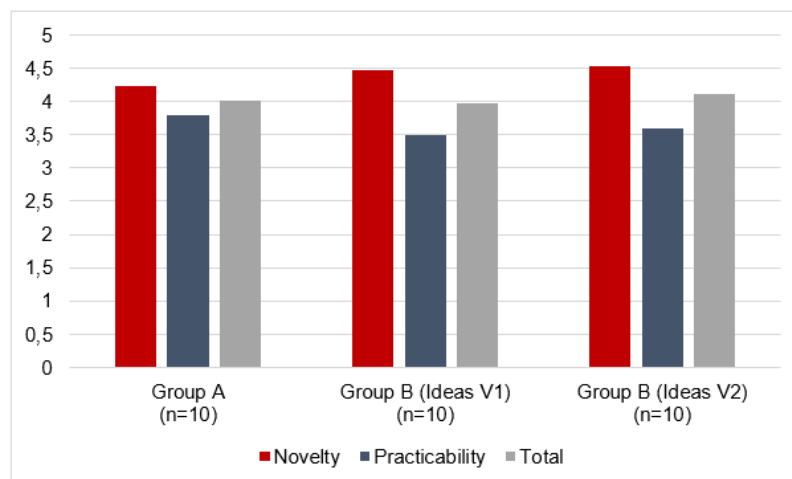


Figure 5. Expert evaluation of ideas

13.4.3 Impact of criticism on the perception of the CA

Each subject from the group B was further asked whether they recognized the criticism and whether they found the critique appropriate. All subjects recognized the critique and 9 out of 10 found it appropriate. Furthermore, subjects were asked how they perceived the criticism (question: If the chatbot criticized your idea, please complete the following sentence: "I found the critique of the chatbot to be..."). Figure 6 summarizes the answers to this question.

Sb	Evaluation of the critique (I found the critique of the chatbot to be ...)
11	not tailored to my specific idea
12	justified
13	surprisingly clever
14	interesting and constructive. However, it was somewhat amusing to be criticized by a chatbot.
15	useful, but obvious, since to filter information this always requires a party to set that filter. However, I found it interesting to be criticized.
16	appropriate. Since I could possibly not bring the topic of swarm intelligence across well enough.
17	interesting, since it was actually valid criticism of my idea.
18	valid interjection and surprisingly appropriate. In a real conversation, the criticism would have led to an interesting discussion.
19	plausible
20	surprisingly reasonable for a chatbot

Figure 6. User evaluation of the criticism

Sb = Subject

If we look at the qualitative evaluation of the criticism of the test subjects for whom the criticism was evaluated as appropriate, an overview can be gained with the help of the adjectives used. The criticism is described as "justified", "constructive", "reasonable", "plausible", "clever", as well as twice as "valid" and twice as "appropriate". Moreover, the criticism was described as "amusing" and twice as "interesting". In addition, words such as "surprising" are found twice. One time the criticism is rated as inappropriate ("not tailored to the idea") and one time as "obvious".

13.5 Discussion

The goal of the pilot study was to find out how and if a CA can support during an ideation process by criticizing the ideas and to what extent the critique influences the ideation process, as well as the resulting perception of the subjects regarding the neutrality of the CA. The A/B test, the resulting ideas, and the feedback from the subjects collected through the survey, demonstrate an influential use of the CA prototype to support an ideation process.

The subjects' evaluation of the CA after the A/B test suggests that there is a correlation between the use of criticism and the perceived neutrality of the CA. Group B, which was criticized by the CA, rated the CA less neutral (3.7 out of 5) than Group A (4.6 out of 5), which was not criticized. Moreover, being criticized by the CA proved to be influential in

terms of the idea generation process. While in group A none of the subjects changed their idea, in group B, 4 out of 10 subjects changed their idea after being criticized by the CA. With the help of the idea evaluations by the innovation experts the effect of the criticism on the quality of the ideas could be examined. While group A (without criticism) achieves a slightly better overall rating (4.02) on the quality of the ideas than group B (before criticism) (3.98), it is exactly the opposite when comparing the two groups A and B with the inclusion of criticism - group B (after criticism) (4.12) achieves a better overall rating on the quality of the ideas than group A (4.02).

13.5.1 User evaluation of the prototype

The implementation of design principle 1 was examined with the question "Did you feel like you always had an overview of the whole process during the interaction?" and "Did you always know how to communicate with the chatbot during the interaction?". The average score of 4.05 and 4.3 out of a maximum of 5 points, and the fact that all subjects managed to achieve the interaction goal of submitting an idea, show a successful implementation of the design principle. Design principle 2, to which the questions "How boring did you find the chatbot?", "Would you say that you could take the chatbot seriously?" and "Did the chatbot overwhelm you?" refer, could also be implemented well based on the positive evaluation. Design principle 3 is covered with the questions "How do you rate the chatbot in terms of human-likeness?" and "How do you rate the chatbot in terms of neutrality?". The fact that the CA's human-likeness was rated the worst with only 3.2 out of 5 points may be due to the fact that the CA's training data was not extensive enough. Despite its immature functionality, it is sufficient to guide the subject through the ideation process. While this does not disprove that the CA's behavior should conform to human behavior to be more acceptable (Tavanapour and Bittner, 2018), it may be indicative of the theory that passing the Turing Test may not always be the most important endeavor (Radziwill and Benton, 2017). The question about neutrality additionally serves to gather insights for RQ 2. The difference in neutrality ratings between Group A and B may indicate that there is a relationship between the use of criticism and neutrality. Overall, it appears that the concept of the prototype and the study design can be used to test how the use of criticism affects the perceived neutrality of the CA (RQ 2). The questions explained so far have been presented to both Group A and Group B and overall show a successful implementation of the design principles and thus a successful set of requirements on which the design principles are based. Since the evaluation results of the individual design principles have not reached their optimum, it can be concluded that improvement is possible through an extension in training data and functionalities.

13.5.2 Impact of criticism on Idea Generation

During the brainstorming process, none of the subjects from group A reconsidered or changed their idea, even though the possibility existed and the subjects were not under time pressure. For group B, 4 out of 10 subjects revised their idea. This shows that the CA's

critique was good enough to convince the subjects of the possible weakness of the idea. Furthermore, it confirms the findings of Siemon et al. (2015) study that collaboration with an AI-based system does not trigger evaluation apprehension since 3 of the 4 subjects who changed their idea after the critique not only added to it but completely rethought it and submitted a new idea as a result. Moreover, it contradicts Osborn's brainstorming rule that one should avoid criticism (Osborn, cited in (Putman and Paulus, 2009)) and suggests that the subjects were not concerned about being repeatedly evaluated negatively. It is questionable whether the original purpose of the criticism, which was to identify and point out the weakness of an idea so that it could be improved by the idea creator, was fulfilled. Since the critique was adapted to the idea, but still general, it is likely that the content of the critique served as a stimulus, thereby creating inspiration for new ideas, and the comparison between group A and group B can thus be questioned, since group A lacked a possible stimulus. Based on the fact that Group B became slightly better than Group A on average only after exposure to the criticism, and the evaluation comparison of each version of the ideas - idea before criticism and changed idea after criticism - shows possible improvements in quality, RQ 2 is answered with the fact that criticism from a CA can lead to better quality of ideas, thus confirming the hypothesis of Tanaka et al. (2015). While in their study, ideas only improve in terms of practicability, the results of this paper refute this finding and show that ideas were rated both more novel and more practical as a result of criticism. To this it must be added that some ideas became qualitatively worse after criticism and revision. This may indicate that the CA's criticism or feedback is not yet intelligent enough to intervene appropriately in such a situation and, for example, alert the subject to the deterioration of the idea. Nevertheless, the results show that the CA prototype, with more training data and more intelligent features, has the potential to represent the role of the black hat in the Six Thinking Hat method (De Bono, 1985; Debowski et al., 2021).

13.5.3 Impact of criticism on the perception of the CA

Group B received questions about the CA's criticism, firstly testing whether the subject perceived the criticism as such "Did the chatbot criticize your idea during the interaction?" and secondly a question about the perception of the criticism as an open question, which reads "If the chatbot criticized your idea, please complete the following sentence: "I found the critique of the chatbot to be ...". All subjects who were criticized recognized that they were criticized and were mostly positively surprised by the CA's feedback. The adjectives used to evaluate the criticism, such as "constructive," "plausible," or "reasonable," shows a successfully implementation of the Black Hat criticism function in the sense of De Bono (1985). Some subjects rated the critique as "surprising," which could be inferred that they did not expect a CA to have this ability, or that the CA did not communicate in advance that it could criticize the ideas, which argues against the requirement that a CA should communicate its abilities transparently. This critically suggests that the overly positive perception of the CA's criticism could be due to the element of surprise. Overall, the

qualitative evaluation nevertheless shows that the criticism function could be implemented expediently with only a small amount of training data. This indicates that the selection of requirements with regard to criticism and the implementation of the requirements are practicable and suitable for conducting a larger study.

13.6 Contributions and limitations

The sample size was not logically derived and calculated within the scope of this work for the reliable testing of specific hypotheses or mean differences from corresponding power analyses. The A/B test was conducted with only 10 subjects randomly selected for each group. This sample size is too small to reliably compare data in statistical tests. For a detectable effect of 40 %, with statistical power of 0.8, and a significance rate of 0.05, the required sample size would be around 3000. Moreover, the sample size also does not allow for valid subgroup analyses with even smaller sizes (e.g., comparing those who change an idea vs. those who do not change an idea.). Similarly, a self-developed questionnaire was used to assess design principles - the items used here were not developed with regards to discriminatory power or other characteristics in a separate scientific process, or critically reviewed with regards to variance in response options. The data only provide indications in the sense of a pilot study and the present work can thus make an important contribution to the preparation for larger scientific investigations. Therefore, we mark this endeavour as a research-in-progress. In addition, the interaction between the subjects and the CA was not being observed. Thus, it remains open whether the subjects were distracted during the interaction, had help from others present in developing the idea, or whether the general conditions, such as the location, were equally conducive to creativity. The CA's comprehension and communication skills are limited due to limited training data. The criticism function could also be improved with more training data. In retrospect, the subject survey could have led to more insights by asking more questions. For example, when analyzing the results, it would have been interesting to know why the subjects who were criticized decided to change their idea or what prevented others from doing so.

The practical contribution of this work is an IT artifact in the form of a CA that can be used with the help of NLP to guide a user through an ideation process. It can use criticism to encourage the user to rethink an idea and improve the quality of ideas, as well as automate the ideation process as a whole. The requirements and design principles create a foundation and thus a contribution to the exploration and design of CAs in similar contexts. The dialog scenario can be reused and extended. The prototype is an example of how, despite less training data, criticism via a CA can be implemented and taken seriously by users. Furthermore, the IT artifact creates a first building block for the moderation of the Six Thinking Hats method by a CA. Thus, the work contributes to the research of CAs as facilitators and teammates. The results of the study provide preliminary evidence of how the CA's criticism function may influence perceptions regarding neutrality. Moreover, the

perceived lack of human-likeness of the CA does not seem to significantly affect the goal of the interaction.

13.7 Recommendations for future research

The pilot study described in this paper can be repeated with a larger sample size. In general, more training data or features can be added to the prototype to build further studies on top of it. For example, other roles of the Six Thinking Hats method can be studied. Also, it would be advisable to investigate how different the criticism of a human is perceived compared to the criticism of a CA in the same scenario and whether the CA is more accepted than a human thanks to its neutrality as assumed. Further, the prototype could be extended based on further design principles with the function that supports, motivates, inspires and informs contextually during the idea generation process. This could be combined with results from existing studies. Furthermore, it would be interesting to find out to what extent the perception of the human-likeness of a CA is related to the perception of its neutrality. Since the subjects in the study were not prepared for the CA to criticize them and this might have created a certain surprise effect that influenced the perception of the criticism, the question arises whether users perceive the criticizing differently when they know it is happening. This could be investigated with a study design similar to the one used in this paper.

13.8 References

- Bittner, E. A., Küstermann, G. C., and Tratzky, C. (2019a). The facilitator is a bot: Towards a conversational agent for facilitating idea elaboration on idea platforms. In *27th European Conference on Information Systems - Information Systems for a Sharing Society, ECIS 2019*.
- Bittner, E. A., Mirbabaie, M., and Morana, S. (2021). Digital facilitation assistance for collaborative, creative design processes. In *Proceedings of the Annual Hawaii International Conference on System Sciences*.
- Bittner, E. A., Oeste-Reiß, S., and Leimeister, J. M. (2019b). Where is the bot in our team? toward a taxonomy of design option combinations for conversational agents in collaborative work. In *Proceedings of the Annual Hawaii International Conference on System Sciences*.
- Bittner, E. A. and Shoury, O. (2019). Designing automated facilitation for design thinking: A chatbot for supporting teams in the empathy map method. In *Proceedings of the Annual Hawaii International Conference on System Sciences*.
- Boonstra, L. (2021). *The Definitive Guide to Conversational AI with Dialogflow and Google Cloud*. Apress.
- Brown, T. (2019). *Change by Design, Revised and Updated: How Design Thinking Transforms Organizations and Inspires Innovation*. Harper Business.

- Buchanan, R. (1992). Wicked problems in design thinking. *Design Issues*, 8.
- Chasanidou, D., Gasparini, A. A., and Lee, E. (2015). Design thinking methods and tools for innovation. In *Lecture Notes in Computer Science*, volume 9186.
- Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society*, 1(1):104.
- De Bono, E. (1985). *Six thinking hats*. Key Porter Books.
- Debowski, N., Tavanapour, N., and Bittner, E. (2021). Prototyping a conversational agent for ai-supported ideation in organizational creativity processes. In *Proceedings of the Annual Hawaii International Conference on System Sciences (HICSS)*.
- Kavlakoglu, E. (2020). Nlp vs. nlu vs. nlq: the differences between three natural language processing concepts. <https://ibm.co/39xlyBM>, accessed on 04.2020.
- Khan, R. and Das, A. (2018). Build better chatbots. *A complete guide to getting started with chatbots*.
- Kohavi, R. and Longbotham, R. (2017). Online controlled experiments and a/b testing. *Encyclopedia of Machine Learning and Data Mining*.
- Lazarevich, K. (2017). 7 tips for creating effective chatbot design. <https://bit.ly/38V1r04>, accessed on: 14.04.2022.
- Lazarevich, K. (2018). How to document chatbot requirements. <https://bit.ly/3wYxjdx>, accessed on: 14.04.2022.
- Maher, M. L. and Fisher, D. H. (2012). Using ai to evaluate creative designs. In *ICDC 2012 - 2nd International Conference on Design Creativity, Proceedings*.
- Oxley, N. L., Dzindolet, M. T., and Paulus, P. B. (1996). The effects of facilitators on the performance of brainstorming groups. *Journal of Social Behavior and Personality*, 11(4):633–646.
- Peffer, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77.
- Putman, V. L. and Paulus, P. B. (2009). Brainstorming, brainstorming rules and decision making. *Journal of Creative Behavior*, 43.
- Radziwill, N. M. and Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents. *arXiv preprint arXiv:1704.04579*.
- Rasa Technologies Inc (2021). Rasa: Open source conversational ai. <https://rasa.com/>, accessed on 14.06.2022.
- Schallmo, D. R. and Lang, K. (2020). *Design Thinking erfolgreich anwenden*. Springer.
- Seeber, I., Bittner, E., Briggs, R. O., de Vreede, G. J., de Vreede, T., Druckenmiller, D., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., and Söllner, M. (2018). Machines as teammates: A collaboration research agenda. In *Proceedings of the Annual Hawaii International Conference on System Sciences (HICSS)*.
- Siemon, D. (2022). Let the computer evaluate your idea: evaluation apprehension in human-computer collaboration. *Behaviour Information Technology*, pages 1–19. doi: 10.1080/0144929X.2021.2023638.

- Siemon, D., Eckardt, L., and Robra-Bissantz, S. (2015). Tracking down the negative group creativity effects with the help of an artificial intelligence-like support system. In *Proceedings of the Annual Hawaii International Conference on System Sciences*.
- Stockleben, B., Thayne, M., Jäminki, S., Haukijärvi, I., Mavengere, N. B., Demirbilek, M., and Ruohonen, M. (2017). Towards a framework for creative online collaboration: A research on challenges and context. *Education and Information Technologies*, 22.
- Strohmann, T., Siemon, D., and Robra-Bissantz, S. (2017). brainstorm: Intelligent assistance in group idea generation. In *Lecture Notes in Computer Science*.
- Tanaka, Y., Sakamoto, Y., and Sonehara, N. (2015). The effects of criticism on creative ideation. *Cognitive Science*.
- Tavanapour, N. and Bittner, E. (2018). Automated facilitation for idea platforms: Design and evaluation of a chatbot prototype. In *International Conference on Information Systems (ICIS)*.
- Tavanapour, N., Poser, M., and Bittner, E. (2019). Supporting the idea generation process in citizen participation - toward an interactive system with a conversational agent as facilitator. In *Twenty-Seventh European Conference on Information Systems (ECIS)*.
- Thoring, K. and Müller, R. M. (2011). Understanding design thinking: A process model based on method engineering. In *Proceedings of the 13th International Conference on Engineering and Product Design Education*.
- Tschimmel, K. (2012). Design thinking as an effective toolkit for innovation. In *ISPIM Conference Proceedings*. The International Society for Professional Innovation Management (ISPIM).
- Turing, A. M. (2009). Computing machinery and intelligence. In *Parsing the turing test*, pages 23–65. Springer.
- Verganti, R. (2016). The innovative power of criticism. <https://bit.ly/3NK1Dyc>, accessed on: 14.04.2022.
- Wechsler, S. M., Saiz, C., Rivas, S. F., Vendramini, C. M. M., Almeida, L. S., Mundim, M. C., and Franco, A. (2018). Creative and critical thinking: Independent or overlapping components? *Thinking Skills and Creativity*, 27:114–122.
- Weizenbaum, J. (1966). Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45.

14 Augmented Facilitation: Designing a multi-modal Conversational Agent for Group Ideation

Cvetkovic, I., Gierlich-Joas, M., Tavanapour, N., Debowski, N., & Bittner, E. (2023). Augmented Facilitation: Designing a multi-modal Conversational Agent for Group Ideation. In *Proceedings of the 44th International Conference on Information Systems*.

Abstract

Human facilitators face the challenge to structure and collect relevant insights from collaborative creative work sessions, which can suffer if they face a high workload. Hence, for effective value co-creation in organizational ideation we suggest an facilitation augmentation with a conversational agent (CA). CAs have the ability to support respective collaborative work by documenting and analyzing unstructured data. Following the design science research paradigm, and based on the literature about facilitation and human- AI collaboration, we derive design principles to develop a CA prototype that collects ideas from a group ideation session and displays them back in a structured (multi-modal) manner. We evaluate the CA by conducting four focus groups. Key findings show that the CA successfully distills and enriches information. Our study contributes to understanding the role of CA in augmenting facilitation and it provides guidance for practice on how to integrate these technologies in group meetings.

Keywords: Augmentation, Conversational Agents, Facilitation, Human-AI Collaboration

14.1 Introduction

Creative work has high relevance for the innovation capability of companies in order to survive in a competitive environment, because innovation processes fuel e.g., emergence of new business models or solution of existing problems (Friesike et al. 2019). To leverage the creative potential in companies, the ideation method is often used. Hereby, teams develop solutions and ideas for a defined problem, thereby co-creating value. Many ideas are generated, then integrated, summarized and finally evaluated (Putman & Paulus 2009). To optimize the results, facilitators monitor interaction between group participants, and control the processes and tasks, among other (Schallmo 2018). While human facilitators deal with all these tasks, team sessions generate a lot of unstructured data which often results in lost insights. Hence, an improved way of documentation needs to be worked on (Chen et al. 2021). Besides human facilitators, digital technologies have been used to support ideation, albeit mostly on individual level (Debowski et al. 2021; Przybilla et al. 2019), where the potential for structuring large amounts of data stayed unexhausted. Furthermore, with the increasing demand for augmentation in the workforce, there is a need for digital innovations that can support collaborative work. For this purpose, it is necessary

to investigate how digital innovations could support human facilitators in moderating collaborative creative work.

Automation of partial tasks has become essential, especially at a time when there is a shortage of skilled workers in all areas of the labor market (Seric & Winkler 2020). Recent research has shown that Artificial Intelligence (AI) can successfully support humans by taking over mundane tasks and analyzing and structuring large amounts of data (Dellermann et al. 2019b; Li et al. 2020). Through increased development of machine learning (ML) and Natural Language Processing (NLP) methods, AI-based systems can increasingly perform or augment tasks autonomously. The processing of linguistic input qualifies AI-based technology to support ideation. Applications that process and respond in natural language are known as Conversational Agents (CAs). They support text input, audio input or both (Laumer et al. 2019). The importance of CAs has grown significantly, especially in recent years as the online channel has become inescapable for businesses in many ways. The versatility of the uses of CAs is continuously expanding with the progressive development of technical possibilities (Mockenhaupt 2021). This development leads to a transformation of work, resulting in novel work configurations (Baptista et al. 2020).

While acknowledging that Human-AI Collaboration spans various forms of user interfaces and interaction modes, it's essential to recognize that our research primarily addresses ideation processes centered around generating textual artifacts, which is what a lot of ideation in early stages in practice is focused toward (Daly et al. 2016). This specific focus guides our exploration of AI-based CAs. The rationale for emphasizing CAs lies in their unique ability to seamlessly integrate with existing communication channels and facilitate ideation without the need for specialized hardware or complex interfaces. Furthermore, CAs can potentially be more motivating due to their anthropomorphic features, adaptability to the user, and direct interaction (Diederich et al. 2020), factors that are important in ideation workshops where participants are accustomed to human facilitation. The adaptability of CAs to diverse contexts and communication preferences ensures inclusive participation, a critical element of successful ideation (Poser et al. 2022).

Within this work, we investigate whether a CA could act as a co-facilitator, thereby augmenting the experience while preserving the natural communication. We build upon the concept of Hybrid Intelligence (Dellermann et al. 2019b), and Facilitation Framework (Dickson et al. 1996), to investigate CA's potential in augmenting organizational ideation processes, and hereby, assist in co-creating value. In this process, the CA should support human facilitators and leave cognitive space for the actual value co-creation in teams. We pose the research question (RQ):

How can a CA be designed and implemented to augment group ideation?

Answering this research question, our work contributes to literature in the following manner: The derived design requirements (DRs), design principles (DPs), and design

features (DFs) represent prescriptive knowledge for the design of CAs in the context of augmented ideation. On a broader level, the study adds value to literature on human-AI collaboration, focusing on CAs in facilitation, and, thereby, addressing the research streams of changing nature of work. From a practical perspective, the technical artefact can be transferred to different contexts and support human facilitators in ideation tasks.

14.2 Theoretical Background

14.2.1 Human-AI Collaboration

Dellermann et al. 2019a propose a concept called Hybrid Intelligence to overcome limitations of humans and AI working independently. They define this concept as the ability to achieve complex goals by combining human and artificial intelligence, resulting in superior results compared to what each of them could achieve on their own. Other terms used to describe this symbiosis are Human-AI teaming, human-in-the-loop or AI- in-the-loop (Siemon 2022). The collaboration between humans and AI is a prerequisite to achieve Hybrid Intelligence, where at least two actors work together towards a shared goal.

Examples of Human-AI collaboration can be seen in various fields. For instance, in the field of education, AI can be used to create personalized learning experiences, while human teachers can provide emotional support and guidance (Maedche et al. 2019). In manufacturing, human workers can collaborate with AI to optimize production processes and reduce errors (Arinez et al. 2020). In finance, AI can help detect fraud, while human experts can interpret the results and make decisions based on the findings (Milana & Ashta 2021). Furthermore, Mahmud et al. 2022 and Anantrasirichai & Bull 2022 have shown that Human-AI collaboration can be especially effective in creative tasks that require both logical and intuitive thinking. By combining the strengths of human and AI intelligence, such as creativity and accuracy, respectively, hybrid teams can generate more innovative and effective solutions.

This collaboration may improve group performance but also raises unresolved questions about the dynamics and characteristics of Human-AI teams. Human-AI collaboration is at an early stage of development, but it is becoming an increasingly important research area in the fields of Human-Computer Interaction and Information Systems. Design issues, such as the design of explainable and transparent AI agents, and the design of a shared workplace with consideration of tasks (Cvetkovic & Bittner 2022; Lubars & Tan 2019), and roles (Siemon 2022) need to be addressed for successful Human-AI collaboration.

14.2.2 Facilitation & Ideation

The term facilitation is used to refer to "a set of different functions and activities that occur before, during, and after a group meeting to assist the group in achieving its goals" (Beranek et al. 1993). Facilitation guides the group through meetings and may use group support systems that provide various facilitation tools. For example, they assist with idea

generation or group decision-making (Leimeister 2014). Facilitation is a dynamic process that seeks to positively influence the relationships of the participants along the goal of structuring the group meetings and providing assistance in mastering the tasks in order to support the creation of the group artifacts in the best possible way (Beranek et al. 1993). Thereby, the promotion of communication within the group is essential.

Facilitation can be applied by a group member, the group leadership, or an external agent by executing activities. Bostrom et al. 1993 describe such activities in their Facilitation Framework in the context of the process, task and relationship activities. The Relationship (Feel about) activities describe influences during the process (How?) that includes activities towards accomplishing a task (What?). Additionally, Dickson et al. 1996 categorizes facilitation activities in task and interaction interventions. The activities that promote relationships and communication are described as interactional interventions (Dickson et al. 1996). Facilitation activities that pursue the goal of structuring the group meetings and providing assistance towards accomplishing the task are task interventions (Dickson et al. 1996). If a special technology is used in the team meeting, a facilitator's support in using this technology also constitutes a task-related activity (Przybilla et al. 2019).

Ideation sessions require specific manners of facilitation and they consist of different activities (Beranek et al. 1993). While in the opening phase the facilitator familiarizes the participants with each other, introduces the course of the meeting, clarifies the roles of the participants, and lists the goals and rules of the meeting, the facilitator is responsible for encouraging the group to work together and guiding them through the agenda during the execution phase (Beranek et al. 1993). In doing so, all activities should be aligned with the group's goals. The facilitator conducts summaries as needed and identifies decisions (Dickson et al. 1996). All participants are expected to participate during the meeting, so it is important to encourage everyone's participation and strengthen ownership and group responsibility (Leimeister 2014). In the closing phase, facilitation provides a summary of the meeting and identifies future tasks that remain after the meeting (Beranek et al. 1993). In general, it is important to present information to the group in an understandable way throughout the meeting to ensure safe communication and information exchange between participants.

Systems that provide support functions for humans in solving creative tasks are referred to as Creativity Support Systems (CSS) or Group Creativity Support Systems (GCSS). GCSS are used by several people and support the collaboration process between them. Both types of systems can be used in all phases of the ideation, from idea generation, through idea evaluation, to idea decision making (Przybilla et al. 2019). GCSS can also perform facilitation tasks, for example, by executing the pre-scripted the collaboration process. Moreover, by supporting the use of creativity methods and creating an environment that fosters trust, less creatively inclined people are inspired to come up with more creative ideas or their ideas can be developed more deeply (Przybilla et al. 2019). GCSS can promote overall communication between team members and therefore perform essential

facilitation tasks (Voigt & Bergener 2013). In this way, they may increase the productivity of collaboration (Willman 2021). In principle, GCSS can provide support in both task and interactional interventions. However, relying on complementary strengths of humans and AI, a hybrid design of facilitation tasks according to the respective strengths could promise better outcomes. Traditional GCSS are not "intelligent" and only novel capabilities allow for hybrid intelligence, giving rise to a new class of GCSS. To this end, CAs have been identified as a potential intelligent addition to GCSS, by taking over specific task interventions in a role of a (co-)facilitator (Debowski et al. 2021).

14.2.3 Conversational Agents as (Co-)Facilitators

In recent years, the increasingly intensive use of instant messengers and rapid advances in AI, particularly NLP and Deep Learning, have led to CAs becoming a widespread technology. They can be assigned to the assistance systems used for AI-driven collaboration between humans and machines (Mockenhaupt 2021). In this context, communication takes place via natural language, via textual or audio channels (Brandtzæg & Følstad 2018). In the former case, they are known as chatbots, in the latter as voice assistants (e.g., Siri, Alexa) (Laumer et al. 2019).

The use of NLP technology is crucial because it influences the output to a high degree (Singh et al. 2019). CAs of this type are also empowered through machine learning and the injection of training data to acquire further knowledge so that response quality can be increased based on past conversations. After the learning phase, the CA is able to recognize intentions and generate responses from a text input (Busse 2021). CAs of this type are, however, susceptible to grammatical errors, as they generate responses dynamically and require large amounts of training and testing data (Ramesh et al. 2017).

Although CAs have been used in variety of domains, their use in collaborative creative work remains relatively unexplored. While CAs have so far mainly communicated with a single person, research has been investigating the extent to which CAs can be used for group collaborations for a few years now (Kim et al. 2021; Winkler et al. 2019). Bittner et al. 2019b; Bittner et al. 2019c showed different possibilities using a taxonomy of how CAs can contribute to the design of a human-machine collaboration. By interviewing professional facilitators, Bittner et al. 2021 were also able to identify key areas of expertise that a CA should cover to support a human facilitator in the context of a facilitation. Toxtli et al. 2018 implemented a CA that can handle task management for individual and group work. They identified particular difficulties and opportunities in CA communication with multiple users. Kim et al. 2020 investigated how CAs can guide and structure discussions and developed corresponding prototypes. Elshan et al. 2022 showed how CAs can relieve teams from innovation blockages. IBM researchers (Tepper et al. 2018) demonstrated a Collabot – a chat assistant that implicitly learns users interests and social ties within a chat group and provides a personalized digest of missed content. By doing so, it assists users in

coping with chat information overload by helping them understand the main topics discussed, collaborators, links and resources.

In previous research, CAs have so far replaced humans as facilitators and have not acted as support for a human facilitator to augment creative work. However, the combination of human and CA facilitator has advantages: human can act on the side of interaction interventions, e.g., monitor team dynamics, and control the process dynamically while CA can support tasks and processes by processing and analyzing data quickly and reliably. Hence, it should be investigated to what extent CAs can be used as co-facilitator.

14.3 Design Science Research Approach

We followed the DSR methodology by Peffers et al. 2007 to explore a CA's role in supporting a human facilitator in the ideation process. Therefore, we conducted the six iterative DSR stages of understanding the problem, deriving objectives for a solution, designing and developing the artefact, demonstrating and evaluating it, and finally communicating it (Peffers et al. 2007). To ensure both rigor and relevance for our design activities (Hevner et al. 2004), we built our study on literature research, qualitative problem interviews, and confirmatory focus groups. The application of different methods to incorporate insights from practice and scientific knowledge base, and evaluate the generated design knowledge can be found in Table 1.

Table 1. DSR Data Collection Approach

DSR Step	Method	Purpose
Problem identification	3 expert interviews	Derive design requirements
Problem identification	Literature review	Derive design requirements
Evaluation	4 focus group discussions	Confirm design principles
Evaluation	SWOT analysis	Generate general insights

14.3.1 Problem Understanding

In the first stage, the problem space was analysed first via a literature review (see Section Theoretical Background), the results of which were supplemented with semi-structured expert interviews with creative work-shops facilitators from a large automotive company (n=3). The semi-structured interviews followed predefined questionnaires, they were recorded, transcribed and analyzed with qualitative content analysis (Mayring & Fenzl 2019). Hereby, a deductive approach was followed with pre-defined categories that were previously identified in the literature. The literature and interview analysis was conducted by four individuals, while accounting for an understanding about (dis)agreements in coding.

Our results reveal that ideation sessions are cumbersome to facilitate as they, among other, generate a lot of unstructured data, often resulting in lost insights, due to improper documentation. Human facilitators, especially novices, often cannot handle this, as they are focused on facilitating tasks, processes, and team interaction (Schallmo 2018). In sum, experts struggle with moderating creative workshops, because they have to handle many aspects previously mentioned (team dynamics, tasks, processes) and, in addition, there is a lot of unstructured data that has to be summarized, prioritized, visualized - in short - enriched and distilled: "I must do the whole [thing] then again: make clusters, merge, prioritize [...] and that is mostly many columns, and then someone does not find his contribution there afterwards. Then perhaps he will not cooperate any more, but you need the cooperation of all, even more so in a large enterprise." From the expert interviews, we derived 8 user requirements and merged them with existing literature requirements (Fig. 1). In the following, we explain the derivation of DRs, DPs and DFs, which represent our objectives of the solution.

14.3.2 Objectives of a Solution

DRs gathered from existing literature (TRs) and user requirements (URs) gathered from conducting interviews served as a foundation for setting up the respective DPs. Initially, several aspects proposed by papers dealing with automated facilitation, CAs, and human-computer interaction were picked up and put into broader categories to sort, not only alike propositions, but also the large amount of the existing research results. Additionally, the gathered DRs are structured along Dickson et al. 1996 task interventions and interaction interventions.

Task interventions: The CA shall undergo training on a prepared dataset (UR1) within the specific topic domain of the ideation session to meet the **DR1. Topic Domain Incorporation**. It shall facilitate ideation sessions, including the sub-phases of idea gathering, idea selection, and summary, adhering to the workflow of established ideation methods (UR2, TR1) as per **DR2. Guided Ideation Workflow** (Bittner & Shoury 2019; Strohmman et al. 2018). At the onset of each session, it shall provide a clear explanation of the ideation method and associated rules (TR2) in alignment with **DR3. Session Introduction** (Bittner & Shoury 2019). The CA shall automate routine tasks (TR3) to reduce users' repetitive work and enable them to focus on value-creating activities, fulfilling **DR4. Task Automation** (Oeste-Reiß et al. 2021). To optimize user interaction (TR4) and ensure user efficiency, it shall possess awareness of various user interaction modes and respond to user queries regarding the ideation method in real-time through dedicated functions activated by predefined keywords (UR3, TR5) as dictated by **DR5. Optimized interaction** (Knote et al. 2020) and **DR6. User Query Handling** (Koch et al. 2020; Strohmman et al. 2018). The CA shall also incorporate a memory function (UR5, TR10) to store user contributions for later phases of ideation, especially for result summarization, in accordance with **DR11. Memory Function** (Koch et al. 2020; Strohmman et al. 2018). It shall aggregate the results of the group's contributions and

provide a summary in a primarily visual manner (UR6, TR11), e.g., via word clouds, idea one-pagers, or clustering (UR7, TR12) as specified by **DR12. Information Aggregation** (Bittner & Shoury 2019) and **DR13. Information Visualization** (Bittner et al. 2021).

Interaction interventions: The CA shall seamlessly integrate with common collaboration platforms (UR4, TR7) to allow users to interact without the need for application switching, meeting **DR7. Integration with Collaboration Platform** (Bittner & Shoury 2019; Tavanapour & Bittner 2018). It shall communicate with users using natural language via text-based messaging (TR9), ensuring a user-friendly interaction and fulfilling **DR10. Textual Communication** (Cvetkovic et al. 2023; Tavanapour & Bittner 2018). The CA shall encourage user engagement in discussions through the exchange of messages (UR4, TR7) to align with **DR8. User Engagement** (Koch et al. 2020; Strohmman et al. 2018). It shall adapt its language to the user group (TR8) to tailor the communication to users' preferences and understanding, in accordance with **DR9. User-Appropriate Language** (Bittner et al. 2019a; Radziwill & Benton 2017; Tavanapour & Bittner 2018). Lastly, the CA shall incorporate features that emulate human behavior (UR8, TR13) to build user trust, encourage interaction, and reduce the feeling of being constantly observed, as specified by **DR14. Human-like Behavior** (Benke 2020; Gnewuch et al. 2017; Strohmman et al. 2018).

For deriving the DPs, the approach proposed by Gregor et al. 2020 was used. To follow the proposed approach, the DRs were first sorted into categories that could later be transformed into DPs. For that matter, similarities between the DRs were used to sort them accordingly. After the initial sorting of the DRs, the scheme for setting up DPs can be applied. For that, four core concepts are typically included in a DP: the aim, context, mechanism, and rationale, which is extended by a normally constant set of implementers, users, and enactors (Gregor et al. 2020). Using the components 'user', 'aim', 'context', 'mechanism' and 'rationale', we can introduce our main clause for the DPs, here exemplified with DP2:

For teams and facilitators (users) to reduce their cognitive load (aim) in organizational ideation (context), the CA should provide the ability to process natural language of multiple users in real-time and respond in the same way via text via a familiar communication medium (mechanism). This DP draws upon assistance provision and human-like conversing in CA research (rationale).

The created DPs aggregate the DRs into respective categories and reflect the commonly addressed domain.

DP1. Task and process assistance tackles the ideation method and process control. It ensures a predefined process and set of rules based on the ideation method. Respective rules and the underlying methodology of the method should be explained to the users by the CA. In this way, facilitators can be relieved of some mundane tasks, while teams have constant access to this information and can use it without asking for repetitions or additional explanations. This corresponds the assistance provision via speedy assistance and relief from mundane tasks (Moussawi 2018; Waizenegger et al. 2020).

DP2. Natural language communication deals with the processing of user input. As a basis, natural language processing and instant messaging are to be used, to allow the users and the CA to communicate with each other. The input from users should be handled by the CA through the execution of respective functions. These are supposed to ensure the assistance provision and human-like conversing (Mygland et al. 2021).

DP3. Information distilling and enrichment targets the subject of documentation and memory. For ensuring a successful visualization and results at the end of the ideation session, gathered ideas must be saved throughout the process. These functions enable the facilitators to distill and enrich information, as well as facilitation (Mygland et al. 2021).

DP4. Human-likeness is about human-like behavior. The CA should communicate with the users in a manner that is as close to human-human interaction as possible, in order to gain the users' trust and will to interact with the CA and to follow its proposed procedure of the ideation session. This DP is congruent to prior CAs research related to human-like experience and personalization (Mygland et al. 2021).

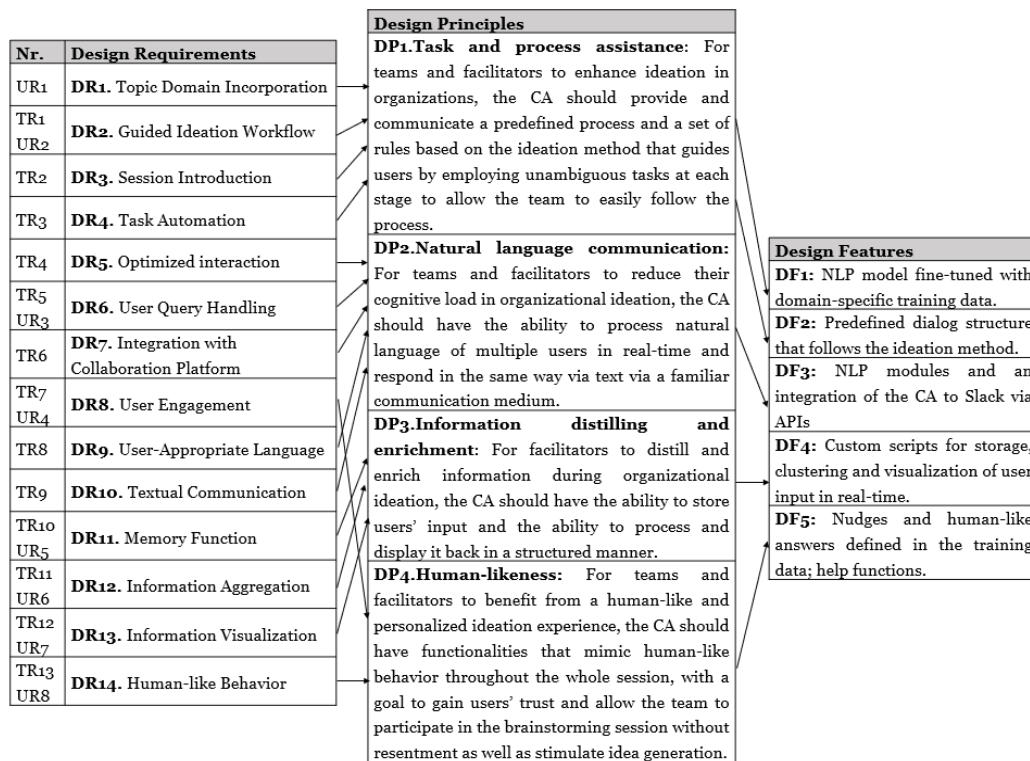
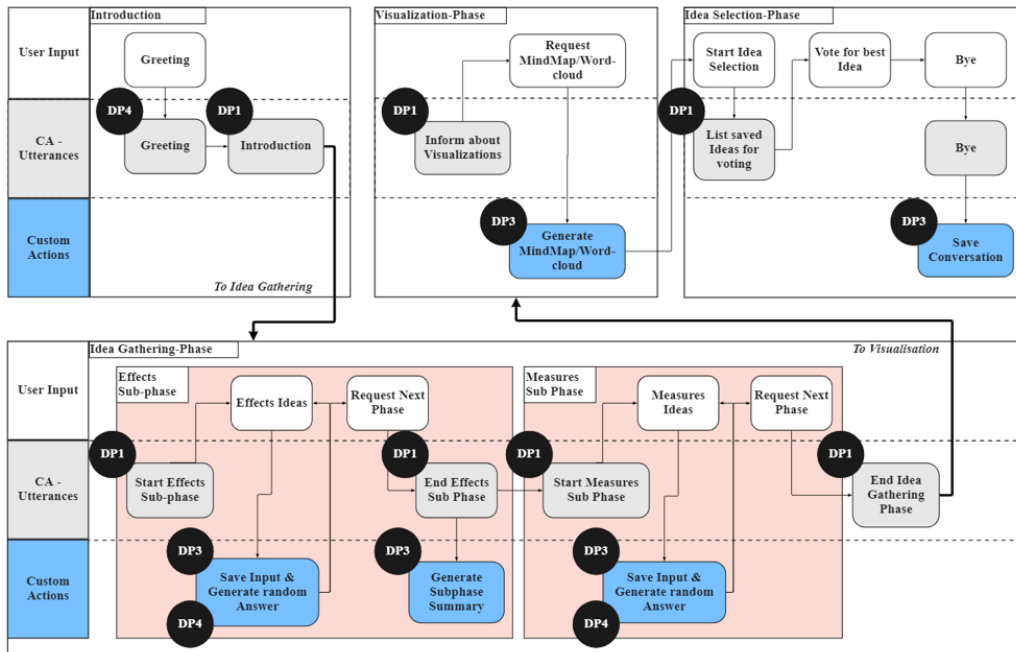


Figure 1. Design Requirements, Principles and Features.

These DPs are translated into corresponding DFs that serve as a basis for the implementation of the CA. Mostly, they rely on NLP and fine-tuning with domain-specific data as well as incorporating human-like answers. Furthermore, custom scripts enable the clustering and visualization functions. Also, there is an integration into a familiar messenger platform (Slack). To ensure proper intent recognition, the underlying NLP model was enhanced with a BERT featurizer (BERT is a large language model trained on billions of data points) (Devlin et al. 2019).

14.3.3 Design & Development

To select a framework for the CA development, we made sure that it satisfies the requirements necessary to implement the above-mentioned DFs. A suitable framework, RASA (Rasa Technologies Inc 2021), was chosen, since it is open-source, models can be fine-tuned, custom functions can be developed, and it can be integrated into many messaging platforms via APIs. The development of the CA prototype was conducted in two iterative cycles, with each cycle significantly influencing the refinement of the DPs and DFs. The initial cycle resulted in an early prototype with limited functionalities, necessitating user feedback gathering and insights. Importantly, we extended the training data through feedback and insights obtained during the first cycle's demonstration and evaluation. This initial version of the prototype lacked a typical ideation phase (idea selection, DR2) and one form of visual summaries (mindmap, DR13). These limitations were addressed in the refined version of the prototype developed during the second cycle. Additionally, we introduced several other features, including the inclusion of an avatar (DR14) and a voting function, based on the feedback and lessons learned from the initial prototype. The dialog flow of the interaction with the CA is showcased in Fig. 2. For visual and textual summaries (clusters) a separate action server was used. This is a part of the RASA framework that offers customization of CA features, i.e., theoretically one could



develop any feature supported by their chosen front-end platform. Fig. 2 displays the materialization of DPs. DP2 applies to the whole interaction and is therefore not indicated in the figure.

Figure 2. Dialog Flow of the Interaction with the CA.

The evaluation of the artefact aligns with the proposed framework for evaluation in DSR by Venable et al. 2016. For this purpose, four steps were followed. First, the goals of evaluation were defined. In our case, this is a proof of concept of the artefact by establishing its conformance with proposed DPs. The second step was to define the strategy for evaluation. We opted for a field test of the design artefact with subsequent confirmatory focus groups to establish the utility of the artefact in field use (Tremblay et al. 2010). Third, respective properties to evaluate are defined (Venable et al. 2016) which are our DPs. Lastly, all respective evaluation episodes were designed (Venable et al. 2016). Semi-structured discussions were directed via an interview guideline, which slightly differed across the two user groups: for facilitators, we asked if some aspects would support them as facilitators; for students, we asked if it would support them as ideation teams. The guideline consisted of a few questions for each DP, followed by spontaneous follow-up questions depending on the answers. The discussions lasted approximately 32 minutes on average. Following the focus group discussions, all participants filled out SWOT analysis templates (Strengths, Weaknesses, Opportunities, Threats) on the chatbot's ability to assist facilitators/teams in (moderating) organizational ideation processes. The resulting qualitative data was analysed with MAXQDA, using Mayring's (Mayring & Fenzl 2019) qualitative content analysis method. Categories (codes) were created based on the DPs and corresponding sub-categories based on the relating DRs. Then, categories were mapped to corresponding text.

14.4 Results

In the following, the answers from the focus group discussions are summarized for each DP. We use illustrative quotes to provide insights into the evaluation. To assess the fitness of DP1, we asked how the CA made the assignment clear, and how it guided the group through the process. Furthermore, we asked in what sense the CA relieved the groups - as facilitators vs. as ideation teams. Participants agree that the assignment explanation was direct, clear, and sufficient. "I think it was pretty clear, because it said [...]". Two participants highlighted the need for more elaborate explanations with examples. "I was missing the BECAUSE [...] you are working on this topic now BECAUSE...". While on the one hand, some participants agree that the explanations were good for guiding them through the process "We had in general a good introduction and guidance", on the other hand, some participants state that one part of the process was confusing "[...] but the way it was phrased, I didn't really realize we were supposed to have these two phases separate." Another point was stressed - the CA should have made it clearer that it is enough when one person addresses him. Suggestions for improvement were mentioned, such as a list with commands, or colored parts of the text for a better overview. The topic of relief through the CA was thoroughly discussed, which resulted in 25 codes. Participants mentioned both aspects that relieved them as well as what overwhelmed them. For the former, the saving of arguments and their clustering was mentioned as particularly relieving: "that was really

very impressive, so wow” - a facilitator said about clusters. The effects that they produce along with mindmap and wordcloud were related to time-saving, the processing and summarizing are very elaborate, as well as impulses to take in regard different perspectives, and showing off the relevance of certain topics: ”we do that to [determine the relevance of specific topics] quite often with voting... Then [with the mindmap] you also see the importance of a topic. If there are 6 effects on one topic, then it seems to be particularly significant [...] and I actually found that quite good.” The rather overwhelming aspects concerned the CA’s answers to each single idea: ”So I thought as a user that it’s just an artefact. It’s not meant as a feedback or appreciation.” Furthermore, the nature of the exercise in the group chat settings produced a lot of text, so that some participants couldn’t keep up with what other people were posting while also thinking of ideas themselves. They suggested a blend-in function, where one can choose whether to see the contributions of others.

DP2 focuses on the correctness of input processing and the influence of CA in focusing on the actual valueadding task. When it comes to the input processing and corresponding reaction from the CA, participants mostly agreed that the CA correctly recognized their input, except in few cases, where the CA replied with a fallback answer. ”And it’s not like this instant [response], like everyone gets the answer right after their message [...], maybe he actually waited to make it more natural. So there’s a lot of thought process behind that.” One facilitator suggested that few funny answers from the CA might be appropriate. Several participants further wish that there were more categories in the clusters. Regarding the support to focus on the actual value-adding task, both positive and negative aspects were identified. For the former, participants discussed that the way in which CA helped focus on value-adding task is by leading the task, moderating, saving input, classifying, ”taking notes in the background.”; ”So we didn’t have to think about all the other stuff coming, but just focus on the [idea gathering] phase.” Paraphrased from a facilitator: ”[...] the benefit is that I can focus on the content, since I do not have to explain things in-between. Especially when there are many people in the workshop, and some repeatedly ask for the same instruction.” Few participant argued that the benefit was hard to see in this short user test, but that they definitely see the potential of CA being helpful with the mindmaps, wordclouds and clusters to orient the process. On the negative side, participants mentioned the fast pace of the assignment, which acted distracting: ”It was a bit difficult because there was so much input from different users and it was hard to see who said what.”

DP3 deals with distilling and enriching information by saving, processing and providing teams’ input in a structured manner. Most participants valued that by saving their input, the CA supported by taking notes in the background, saved time, and structured the input. A facilitator further praised the CA for saving the arguments from previous phase regardless of the fact that other teammates already started with a new phase. On the other hand, few participants argued that the saving function did not support much in this context, since the exercise was short. Had it been longer, the potential of this function would be more visible.

One of the means for distilling information the CA used were clusters (see Fig. 3). This was a highly praised feature by both user groups, but facilitators especially emphasized its helpfulness: "really strong very, very surprised that it was so fast and so good, kind of summarized." One participant said the categories inspired him by providing a summarized overview of the covered topics, hinting that the team should take in regard other topics as well. A discussion was led about whether there should have been more categories in the cluster. However, a suggestion to think about uncovered topics by CA would have been helpful. The CA used wordcloud and mindmap as another means for distilling and enriching information. Both visualizations were differently perceived by the participants: while some found the wordcloud helpful and inspiring "I think word cloud, for example, could be a good source of inspiration for further ideas to brainstorm.", others preferred the mindmap, and vice versa. While some participants didn't understand the idea behind the mindmap, mostly because they did not use its implications to subsequently choose the best idea, due to very fast proceeding in the next phase (chosen by participants themselves); others could make several conclusions with the help of it. For example, they could see which topics might be more relevant, due to being mentioned more frequently than others, or they could also see the relations between different levels: "It was interesting to see that for the effects, we had 3 clusters (agriculture, population, and politics) and for the measures, we had only agriculture and politics – meaning the measures have to be taken on another level to account for different effects." Some participants further mentioned that they would like to have another level on the mindmap where all individual contributions are also shown, with a possibility to blend this in/out. In a discussion whether the usage of the mindmap should be explained, a facilitator stated "There should be a task in this phase so to say, hey, look at the content here, does it make you think of something regarding the importance [of shown topics] [...] it should be an explicit task." On the other hand, other participant raised concerns that this might introduce bias.

DP4 was evaluated by asking about CA's human-likeness and stimulation to generate ideas. Participants agreed that the CA was not human-like, since its answers were fast, very structured and repetitive. However, this mostly did not negatively affect the participants: "it was bot-like, but that didn't affect me." They also agree that CA's behavior mostly did not stimulate idea generation: "so the summaries, if at all, stimulated me a bit, otherwise mindmap and wordcloud have the potential, but here it wasn't the case". Further, participants say "it was more of taking notes in the background", "being a moderator, process assistance."

Besides evaluating DPs, we also performed a SWOT analysis of the prototype to carve out its strengths, weaknesses, opportunities, and threats. The most frequently mentioned points are displayed in the Fig. 4.

Strengths • Clusters the content quickly and understandable (n=10) • Different displays of summarized information (n=8) • Stimulates new ideas (n=7) • Clear structure and organization (n=6) • Guided support (n=4) • Time-efficient (n=4)	Weaknesses • Repetitive answers (n=9) • Not human-like (n=6) • Overload of information (n=6) • Advancing to next phase decided by one team member (n=4) • Time pressure (n=3) • Cannot recognize keywords not related to the subject (n=2)
Opportunities • Source of inspiration (n=7) • Moderation assistance (n=5)	Threats • Clustering might introduce bias due to pre-conceived topics (n=4) • People don't interact with each other (n=4) • Can make users more passive in the process (n=2)

Figure 4. SWOT Analysis Results

14.5 Discussion

Our study responds to a pressing need in the realm of collaborative creative work. As the volume of unstructured data generated during ideation processes continues to rise, capturing valuable insights becomes increasingly challenging (Chen et al. 2021). Our research offers a solution by proposing design principles for a CA that serves as a (co-)facilitator, documenter, and summarizer of information. To address the RQ on how to design and implement a CA to augment group ideation, we designed, developed and evaluated a CA along four proposed DPs. In the discussion, we reflect on the artifact itself before discussing it in the context of prior works and connecting our findings to scholarly debates.

Recalling the DPs, DP1 addresses task and process assistance. The evaluation shows that this DP displays pretty high conformance with the artefact. It could however be extended to account for individual differences in understanding task descriptions, by e.g., adding a read more button, so as to not overwhelm everyone with long texts. Further, team members should have the ability to blend in/out the contributions of others, in case these are overwhelming them. Facilitators further agreed that the goal of the exercise should be explained too, along with some examples.

DP2 focuses on natural language communication and partly conforms with the artefact. While users' input was mostly processed correctly, CA reactions to single ideas were not always satisfactory. Student participants suggested either no feedback or content-dependent feedback to ideas. Facilitators also argued for content-dependent feedback and questioning of contributions. While this is necessary, we argue that questioning and providing context-dependent feedback would need excessive training and could still turn out to be unsatisfactory, which is why these kinds of tasks should be left for the human facilitator. This aligns with the principles of Hybrid Intelligence (Dellermann et al. 2019b), which state that the tasks in a human-AI ensemble should be split according to respective strengths and skills. Our idea for the CA is not to automate creative sessions, but rather augment with its (e.g., data processing) abilities. To improve the natural communication,

DP2 should further be extended with a response delay throughout the session. Previous research has shown that dynamic delay calculated based on the complexity of the response and complexity of the previous message increases the perception of social presence and overall satisfaction with the interaction (Gnewuch et al. 2018).

DP3 represents the core functionalities of the CA which aims to distill and enrich information. This DP mostly conforms with the artefact, but the short nature of the user test and conflicting team dynamics did not allow for its proper usage. However, the potential of the respective DFs was recognized. Of the three types of summaries provided by the CA, participants preferred the clusters, especially facilitators. More categories were wished for, however this would again create a clutter of information and go against the principles related to distilling information (Mygland et al. 2021). Some participants needed a "user manual" for the mindmaps, hereby a facilitator argues that its logic should be explicitly explained. Others argue that this would introduce bias. This opens an interesting avenue for further research.

DP4 does not seem to conform with the prototype. It was neither perceived as human-like, nor did it stimulate idea generation. During focus groups, participants mentioned that the lack of human-likeness did not affect the interaction negatively. The CA was perceived as a note taker, moderator and process assistance. This implicates that perceived competence might be more relevant than perceived anthropomorphism, and goes in line with the findings of Ciechanowski et al. 2019, who found that too much human-likeness produces uncanny valley effects (Mori et al. 2012), and less negative affect in cooperation with a simpler text chatbot. However, the results of the SWOT analysis imply that CA's lack of human-likeness was perceived as a weakness (n=6). Given its role as a co-moderator, it is highly relevant to uncover further insights on this debating topic via further studies and use as input for further DSR cycles.

Results of the SWOT analysis reflect the assumptions of Hybrid Intelligence (Dellermann et al. 2019b). While the CA performed well in analyzing and structuring data, it displayed repetitive patterns and did not account for team dynamics, which might be an indication for the strength of the CA to support task interventions and its limited potential to facilitate interaction interventions (Debowski et al. 2021). Therefore, based on the strengths (Figure 4) we can deduce to augment task interventions (Dickson et al. 1996) for the facilitation activities process and task of the facilitation framework (Bostrom et al. 1993) with CAs and handover the interaction interventions (Dickson et al. 1996) for the relationship activity (Bostrom et al. 1993) to human facilitators to enable Hybrid Intelligence according to (Dellermann et al. 2019b). This is another indication that the CA should serve as an augmentation to human facilitator instead of an automation, if applicable.

Besides reflecting on the artefact's value, we want to embed our research in the context of prior literature. Our research extends the range of contexts in which AI is applied. It introduces the concept of AI-powered facilitation within creative team environments,

thereby advancing the scope frontier (Berente et al. 2021). By compensating for the limits of human memory and providing cognitive assistance, our CA contributes to the evolving landscape of AI applications. In terms of endowments from the delegation framework to agentic IS artifacts (Baird & Maruping 2021), our artifact compensates for the limits of human memory by aggregating and distilling data using the mechanisms of cognitive appraisal by e.g., taking over a task when humans have no capacity to complete it at current time in such fast manner. Furthermore, the insights from this work could serve as a starting point for identifying the willingness-to-delegate in the specific context of creative teamwork (Baird & Maruping 2021).

Moreover, there is a tradition of human-AI collaboration with CAs (e.g., Poser et al. 2022; Sowa et al. 2021).

With this study we show that CAs have a potential to successfully enrich facilitation of collaborative creative work and thereby co-create value by applying task interventions such as providing assistance with facilitation, distilling and enriching information. In this way, CAs can take over tedious tasks of documentation and summarization of information as well as task explanations and process assistance. In line with Stoeckli et al. 2018, our CA also increases the expressive power of textual communication.

From the specific context of facilitating ideation (Poser et al. 2022; Strohmman et al. 2017), we also learned that it is crucial to take team dynamics into account and elaborate on the logic behind specific CA functions, as the ability of technology to co-create value is contingent on the levels of social connectivity between the human users (Breidbach et al. 2013). Future studies could explore whether additional human moderation of team dynamics would lead to even better actualization of CA functions on the side of process assistance and information enrichment.

Finally, we exemplify how the use of autonomous tools opens unprecedented opportunities for creative problem solving (Seidel et al. 2018). Our research lays the foundation for future work designs that incorporate AI-powered CAs as collaborative tools. It exemplifies the relevance of design theories in anticipating the implications of autonomous agents in the context of future work (Kane et al. 2021). Our study is not just about introducing a technical artifact but about reshaping how work is accomplished through human-AI collaboration.

14.6 Contribution and Conclusion

Collaborative work produces a lot of unstructured data and often many insights get lost. To address this issue, we proposed principles for a CA to support human facilitators by taking over the part of documenting and summarizing information. Following an established DSR methodology (Peffer et al. 2007), this study builds upon prior knowledge (Möller et al. 2022) and user requirements to provide initial evidence toward proof of concept, by creating an instantiation of the CA and then evaluating its potential. The results reflect broad potential for CAs in supporting creative collaborative work.

Our work contributes to both theory and practice: for the former, we provide prescriptive knowledge in form of DRs, DPs, and DFs for a CA that facilitates value co-creation in group ideation. With and beyond that, we add to the knowledge base on the topic of human-AI collaboration in creative work and CAs in group settings. Integrating CAs for augmentation in organizational teams paves the path towards novel work scenarios. Thereby, we link our study to the research stream on future of work and we illustrate how a CA can co-create value alongside human teams. For the latter, we provide a technical artefact that is reusable as well as adaptable. It can be integrated with other channels, e.g., Microsoft Teams, custom summarization features can be reused or integrated in another software, such as intelligent dashboards or whiteboards (more frequently used tool for digital collaborative creative teamwork). For users, this artefact can be transferred to different contexts and support human facilitators. For developers working on similar technical solutions, it provides a guidance on how to support ideation processes.

Our research comes with limitations. Even though we tested the prototype with experts, we could not observe a role-out in an organization with respective boundary conditions and contextual factors. With performing a short user test, we did not account for team dynamics that affected the interaction in a way that not all features were fully exploited. Moreover, our artifact is designed to facilitate text-based ideation sessions and, thereby, it represents one specific instance of a broader phenomenon. However, at the current state, it does not take voice-based ideation sessions into account. To account for these limitations, it would be interesting to perform a case study in an organization with a real business problem to better understand team dynamics and the influence of the organizational context. Furthermore, we see great value in quantifying the results achieved with the help of the CA. Hereby, one could move from a proof of concept to a proof of value by experimentally comparing results of ideation sessions with and without a CA. Adding another layer of human facilitation and forming a Hybrid Intelligence (Dellermann et al. 2019b) is another relevant research venue that could be born out of this research. Finally, research beyond this proof of concept is needed to explore CA's potential to augment facilitation. The path from conceiving new systems to creating standalone value involves many steps and can require many studies (Nunamaker et al. 2017). In the process of designing a CA for augmented facilitation and testing a limited prototype, we did not demonstrate that the CA concept has identified and overcome all significant obstacles and is prepared for wider deployment. Instead, this study has offered a proof of concept, which includes outlining the issue and a potential remedy as well as finding the evidence that the concept has potential, albeit only at a limited scope. This foundation should be improved upon by subsequent studies.

Acknowledgements

This research was funded by the German Federal Ministry of Education and Research (BMBF), in the context of the project HyMeKI (01IS20057). We thank the students Florentin Hitzke, Daniel Rupprecht, Fabian Burkhardt, and Niclas Trelle for their support

with the literature review, conducting and analyzing the expert interviews, and designing and evaluating the prototype in the first iteration.

14.7 References

- Anantrasirichai, N. & Bull, D. (2022). Artificial intelligence in the creative industries: a review. *Artificial intelligence review* (55), 589–656.
- Arinez, J. F., Chang, Q., Gao, R. X., Xu, C. & Zhang, J. (2020). Artificial intelligence in advanced manufacturing: Current status and future outlook. *Journal of Manufacturing Science and Engineering* (142:11).
- Baird, A. & Maruping, L. M. (2021). The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts. *MIS quarterly* (45:1).
- Baptista, J., Stein, M.-K., Klein, S., Watson-Manheim, M. B. & Lee, J. (2020). Digital work and organisational transformation: Emergent Digital/Human work configurations in modern organisations. *The Journal of Strategic Information Systems* (29:2), 101–618.
- Benke, I. (2020). Towards Design Principles for Trustworthy Affective Chatbots for Virtual Teams. In: *ECIS 2020 Proceedings*.
- Beranek, P. M., Beise, C. M. & Niederman, F. (1993). Facilitation and group support systems. In: *Proceedings of the Twenty-sixth Hawaii International Conference on System Sciences*. Vol. 4, pp. 199–207.
- Berente, N., Gu, B., Recker, J. & Santhanam, R. (2021). Managing artificial intelligence. *MIS quarterly* (45:3).
- Bittner, E., Küstermann, G. & Tratzky, C. (2019)a. The Facilitator is a Bot: towards a Conversational Agent for Facilitating Idea Elaboration on Idea Platforms. In: *Proceedings of the 27th European Conference on Information Systems (ECIS)*.
- Bittner, E., Mirbabaie, M. & Morana, S. (2021). Digital Facilitation Assistance for Collaborative, Creative Design Processes. In: *Proceedings of the 54th Hawaii International Conference on System Sciences*.
- Bittner, E., Oeste-Reiß, S., Ebel, P. & Söllner, M. (2019)b. Mensch-Maschine-Kollaboration: Grundlagen, Gestaltungsherausforderungen und Potenziale für verschiedene Anwendungsdomänen. *HMD Praxis der Wirtschaftsinformatik* (56:1), 34–49.
- Bittner, E., Oeste-Reiß, S. & Leimeister, J. M. (2019)c. Where is the Bot in our Team? Toward a Taxonomy of Design Option Combinations for Conversational Agents in Collaborative Work. In: *Proceedings of the 52nd Hawaii International Conference on System Sciences*.
- Bittner, E. & Shoury, O. (2019). Designing Automated Facilitation for Design Thinking: A Chatbot for Supporting Teams in the Empathy Map Method. In: *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

- Bostrom, R. P., Anson, R. & Clawson, V. K. (1993). Group facilitation and group support systems. *Group support systems: New perspectives* (8), 146–168.
- Brandtzæg, P. B. & Følstad, A. (2018). Chatbots: changing user needs and motivations. *Interactions* (25), 38–43.
- Breidbach, C. F., Kolb, D. G. & Srinivasan, A. (2013). Connectivity in Service Systems: Does Technology-Enablement Impact the Ability of a Service System to Co-Create Value? *Journal of Service Research* (16:3), 428–441.
- Busse, M. (2021). Maschinelles Lernen vs. Regelbasierte Erkennung - was ist Besser Geeignet für chatbots? <https://www.kiko.bot/blog/allgemein/maschinelles-lernen-vs-regelbasierter-chatbot/>, Accessed on 20.05.2022.
- Chen, R., Demko, M., Byrne, D. & Louw, M. (2021). Probing Documentation Practices: Reflecting on Students' Conceptions, Values, and Experiences with Documentation in Creative Inquiry. In: *Creativity and Cognition*. Association for Computing Machinery.
- Ciechanowski, L., Przegalinska, A., Magnuski, M. & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems* (92), 539–548.
- Cvetkovic, I. & Bittner, E. (2022). Task Delegability to AI: Evaluation of a Framework in a Knowledge Work Context. In: *Proceedings of the 55th Hawaii International Conference on System Sciences*.
- Cvetkovic, I., Rosenberg, V. & Bittner, E. (2023). Conversational Agent as a Black Hat: Can Criticising Improve Idea Generation? In: *Proceedings of the 56th Hawaii International Conference on System Sciences*.
- Daly, S. R., Seifert, C. M., Yilmaz, S. & Gonzalez, R. (2016). Comparing Ideation Techniques for Beginning Designers. *ASME Journal of Mechanical Design* (138:10), 101–108.
- Debowski, N., Tavanapour, N. & Bittner, E. (2021). Einsatz eines virtuellen Kollaborators in analogen & digitalen Workshops im organisationalen Kontext. *Informatik Spektrum* (44), 170–177.
- Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S. & Ebel, P. (2019)a. The future of human-AI collaboration: a taxonomy of design knowledge for hybrid intelligence systems. In: *Proceedings of the 52nd Hawaii International Conference on System Sciences*.
- Dellermann, D., Ebel, P., Söllner, M. & Leimeister, J. M. (2019)b. Hybrid Intelligence. *Business & Information Systems Engineering* (61:5), 637–643.
- Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv* (abs/1810.04805).
- Dickson, G. W., Lee-Partridge, J. E., Limayem, M. & Desanctis, G. L. (1996). Facilitating computer-supported meetings: A cumulative analysis in a multiple-criteria task environment. *Group Decision and Negotiation* (5), 51–72.

- Diederich, S., Brendel, A. B. & Kolbe, L. M. (2020). Designing anthropomorphic enterprise conversational agents. *Business & Information Systems Engineering* (62), 193–209.
- Elshan, E., Zierau, N., Janson, A. & Leimeister, J. M. (2022). How Conversational Agents Relieve Teams from Innovation Blockages. *Academy of Management Proceedings* (2022:1).
- Friesike, S., Flath, C. M., Wirth, M. & Thiesse, F. (2019). Creativity and productivity in product design for additive manufacturing: Mechanisms and platform outcomes of remixing. *Journal of Operations Management* (65:8), 735–752.
- Gnewuch, U., Morana, S., Adam, M. T. P. & Maedche, A. (2018). Faster is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction. In: *Proceeding of the 26th European Conference on Information Systems (ECIS)*.
- Gnewuch, U., Morana, S. & Maedche, A. (2017). Towards Designing Cooperative and Social Conversational Agents for Customer Service. In: *Proceedings of the 38th International Conference on Information Systems (ICIS)*.
- Gregor, S., Chandra Kruse, L. & Seidel, S. (2020). Research perspectives: The anatomy of a design principle. *Journal of the Association for Information Systems* (21:6), 1622–1652.
- Hevner, A. R., March, S. T., Park, J. & Ram, S. (2004). Design science in information systems research. *MIS quarterly* (28:1), 75–105.
- Kane, G. C., Young, A. G., Majchrzak, A. & Ransbotham, S. (2021). Avoiding an oppressive future of machine learning: A design theory for emancipatory assistants. *Mis Quarterly* (45:1), 371–396.
- Kim, S., Eun, J., Oh, C., Suh, B. & Lee, J. (2020). Bot in the bunch: Facilitating group chat discussion by improving efficiency and participation with a chatbot. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13.
- Kim, S., Eun, J., Seering, J. & Lee, J. (2021). Moderator chatbot for deliberative discussion: Effects of discussion structure and discussant facilitation. *Proceedings of the ACM on Human-Computer Interaction* (5:CSCW1), 1–26.
- Knote, R., Janson, A., Söllner, M. & Leimeister, J. M. (2020). Value co-creation in smart services: a functional affordances perspective on smart personal assistants. *Journal of the Association for Information Systems* (22:2), 418–458.
- Koch, J., Taffin, N., Beaudouin-Lafon, M., Laine, M., Lucero, A. & Mackay, W. E. (2020). ImageSense: An Intelligent Collaborative Ideation Tool to Support Diverse Human-Computer Partnerships. *Proceedings of the ACM on Human-Computer Interaction* (4:CSCW1).
- Laumer, S., Gubler, F., Racheva, A. & Maier, C. (2019). Use cases for conversational agents: An interview- based study. In: *Americas Conference on Information Systems (AMCIS)*.

- Leimeister, J. M. (2014). *Collaboration Engineering: IT-gestützte Zusammenarbeitsprozesse systematisch entwickeln und durchführen*. Springer Berlin Heidelberg.
- Li, M. M., Bronner, E., Peters, C. & Leimeister, J. M. (2020). Künstliche Intelligenz und menschliche Kompetenz zur Automatisierung und Personalisierung von Dienstleistungen am Beispiel des Support. In: *Automatisierung und Personalisierung von Dienstleistungen: Methoden – Potenziale – Einsatzfelder*. M. Bruhn & K. Hadwich (eds.). Springer Fachmedien Wiesbaden, pp. 235–251.
- Lubars, B. & Tan, C. (2019). Ask not what AI can do, but what AI should do: Towards a framework of task delegability. In: *Advances in Neural Information Processing Systems*. Vol. 32.
- Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S. & Söllner, M. (2019). AI-based digital assistants: Opportunities, threats, and research perspectives. *Business & Information Systems Engineering* (61), 535–544.
- Mahmud, B. U., Hong, G. Y. & Fong, B. (2022). A Study of Human-AI Symbiosis for Creative Work: Recent Developments and Future Directions in Deep Learning. *ACM Trans. Multimedia Comput. Commun. Appl.* (Just Accepted).
- Mayring, P. & Fenzl, T. (2019). Qualitative inhaltsanalyse. In: *Handbuch Methoden der empirischen Sozialforschung*. Springer, pp. 633–648.
- Milana, C. & Ashta, A. (2021). Artificial intelligence techniques in finance and financial markets: a survey of the literature. *Strategic Change* (30:3), 189–209.
- Mockenhaupt, A. (2021). *Digitalisierung und künstliche Intelligenz in der Produktion*. Springer Wiesbaden.
- Möller, F., Strobel, G., Schoormann, T. & Hansen, M. (2022). Unveiling the Cloak: Kernel Theory Use in Design Science Research. In: *Proceedings of the 43rd International Conference on Information Systems (ICIS)*.
- Mori, M., MacDorman, K. F. & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & automation magazine* (19:2), 98–100.
- Moussawi, S. (2018). User experiences with personal intelligent agents: a sensory, physical, functional and cognitive affordances view. In: *Proceedings of the 2018 ACM SIGMIS Conference on Computers and People Research*, pp. 86–92.
- Mygland, M. J., Schibbye, M., Pappas, I. O. & Vassilakopoulou, P. (2021). Affordances in human-chatbot interaction: a review of the literature. In: *Conference on e-Business, e-Services and e-Society*. Springer, pp. 3–17.
- Nunamaker, J. F., Twyman, N. W., Giboney, J. S. & Briggs, R. O. (2017). Creating High-Value Real-World Impact through Systematic Programs of Research. *MIS Quarterly* (41:2), 335–351.
- Oeste-Reiß, S., Bittner, E., Cvetkovic, I., Günther, A., Leimeister, J. M., Memmert, L., Ott, A., Sick, B. & Wolter, K. (2021). Hybride Wissensarbeit. *Informatik-Spektrum* (44), 148–152.

- Peffer, K., Tuunanen, T., Rothenberger, M. A. & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems* (24:3), 45–77.
- Poser, M., Küstermann, G. C., Tavanapour, N. & Bittner, E. A. C. (2022). Design and Evaluation of a Conversational Agent for Facilitating Idea Generation in Organizational Innovation Processes. *Information Systems Frontiers* (24), 771–796.
- Przybilla, L., Baar, L., Wiesche, M. & Krcmar, H. (2019). Machines as Teammates in Creative Teams: Digital Facilitation of the Dual Pathways to Creativity. In: *Proceedings of the 2019 on Computers and People Research Conference*. Association for Computing Machinery, pp. 94–102.
- Putman, V. L. & Paulus, P. B. (2009). Brainstorming, brainstorming rules and decision making. *The Journal of creative behavior* (43:1), 29–40.
- Radziwill, N. M. & Benton, M. C. (2017). Evaluating Quality of Chatbots and Intelligent Conversational Agents. arXiv: 1704.04579 [cs.CY].
- Ramesh, K., Ravishankaran, S., Joshi, A. & Chandrasekaran, K. (2017). A survey of design techniques for conversational agents. In: *International conference on information, communication and computing technology*. Springer, pp. 336–350.
- Rasa Technologies Inc (2021). RASA: Open source conversational AI. <https://rasa.com/>.
- Schallmo, D. (2018). Jetzt Design Thinking anwenden: In 7 Schritten zu kundenorientierten Produkten und Dienstleistungen. In: Wiesbaden: Springer Fachmedien Wiesbaden.
- Seidel, S., Berente, N., Lindberg, A., Lyytinen, K. & Nickerson, J. V. (2018). Autonomous tools and design: a triple-loop approach to human-machine learning. *Communications of the ACM* (62:1), 50–57.
- Seric, A. & Winkler, D. (2020). COVID-19 could spur automation and reverse globalisation—to some extent. <https://shorturl.at/rNQ04>, Accessed on 18.06.2022.
- Siemon, D. (2022). Elaborating team roles for artificial intelligence-based teammates in human-AI collaboration. *Group Decision and Negotiation* (31:5), 871–912.
- Singh, A., Ramasubramanian, K. & Shivam, S. (2019). Building an enterprise chatbot: Work with protected enterprise data using open source frameworks. O'Reilly.
- Sowa, K., Przegalinska, A. & Ciechanowski, L. (2021). Cobots in knowledge work: Human–AI collaboration in managerial professions. *Journal of Business Research* (125), 135–142.
- Stoeckli, E., Uebernickel, F. & Brenner, W. (2018). Exploring affordances of slack integrations and their actualization within enterprises-towards an understanding of how chatbots create value. In: *Proceedings of the 51st Hawaii International Conference on System Sciences*.
- Strohmann, T., Fischer, S., Siemon, D., Brachten, F., Lattemann, C., Robra-Bissantz, S. & Stieglitz, S. (2018). Virtual Moderation Assistance: Creating Design Guidelines for Virtual Assistants Supporting Creative Workshops. In: *PACIS 2018 Proceedings*. 80.

- Strohmann, T., Siemon, D. & Robra-Bissantz, S. (2017). brAInstorm: Intelligent Assistance in Group Idea Generation. In: *Designing the Digital Transformation*. A. Maedche, J. vom Brocke & A. Hevner (eds.). Cham: Springer International Publishing, pp. 457–461.
- Tavanapour, N. & Bittner, E. (2018). Automated Facilitation for Idea Platforms: Design and Evaluation of a Chatbot Prototype. In: *Proceedings of the International Conference on Information Systems (ICIS)*.
- Tepper, N., Hashavit, A., Barnea, M., Ronen, I. & Leiba, L. (2018). Collabot: Personalized Group Chat Summarization. In: *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. Association for Computing Machinery, pp. 771–774.
- Toxtli, C., Monroy-Hernández, A. & Cranshaw, J. (2018). Understanding chatbot-mediated task management. In: *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–6.
- Tremblay, M. C., Hevner, A. R. & Berndt, D. J. (2010). The Use of Focus Groups in Design Science Research. In: *Design Research in Information Systems: Theory and Practice*. Springer US, pp. 121–143.
- Venable, J., Pries-Heje, J. & Baskerville, R. (2016). FEDS: a framework for evaluation in design science research. *European journal of information systems* (25:1), 77–89.
- Voigt, M. & Bergener, K. (2013). Enhancing Creativity in Groups - Proposition of an Integrated Framework for Designing Group Creativity Support Systems. In: *Proceedings of the 46th Hawaii International Conference on System Sciences*, pp. 225–234.
- Waizenegger, L., Seeber, I., Dawson, G. & Desouza, K. (2020). Conversational agents-exploring generative mechanisms and second-hand effects of actualized technology affordances. In: *Proceedings of the 53rd Hawaii international conference on system sciences*.
- Willman, J. (2021). Creating a Chatbot. In: *Modern PyQt*. Springer, pp. 261–282.
- Winkler, R., Neuweiler, M. L., Bittner, E. & Söllner, M. (2019). Hey Alexa, Please Help Us Solve This Problem! How Interactions with Smart Personal Assistants Improve Group Performance. In: *Proceedings of the International Conference of Information Systems (ICIS)*.

15 Designing a Hybrid Intelligence System for the Facilitation of Group Collaboration

Cvetkovic, I., Streek, S., & Bittner, E. (2024). Designing a Hybrid Intelligence System for the Facilitation of Group Collaboration. In *Proceedings of the Australasian Conference on Information Systems (ACIS)*.

Abstract

This paper introduces a Hybrid Facilitation System (HFS), combining human and AI strengths to enhance remote collaboration. Hereby, tasks are distributed between a human and an AI facilitator based on their complementary skills. The study follows a design science research approach, utilizing a task allocation framework, an underlying study (with half of the teams moderated by a chatbot and others by humans), and a systematic literature review to develop a prototype HFS iteratively and evaluate it through a pilot study with three groups (n=12) collaborating on a creative task in Slack. Feedback from surveys and interviews with both end-users and facilitators demonstrated that the hybrid system fosters a trusting and efficient collaborative environment. The paper contributes a framework for task allocation in hybrid systems as well prescriptive knowledge to a nascent design theory of Hybrid Facilitation.

Keywords: Conversational Agents, Hybrid Intelligence, Hybrid Facilitation, Task Allocation

15.1 Introduction

Collaboration is a crucial component in accomplishing organizational tasks (de Vreede & Briggs 2019). Team members don't necessarily need to work together in person; they can also do so remotely, from various locations, or even when working from the same building, but with different departments. The recent pandemic dramatically accelerated the adoption of remote work. McKinsey's 2022 survey revealed that 58% of employed respondents in the U.S. now have the option to work remotely part- or full-time, a major shift from pre-pandemic levels (McKinsey & Company, 2022). As remote collaboration becomes the norm, organizations are increasingly relying on tools to support virtual teamwork, with the global collaboration software market expected to reach \$16.6 billion by 2029 (Statista, 2024).

Facilitation can promote a working group's collaboration. As a "neutral supporter" (Leimeister 2014), facilitation can be a crucial component, contributing to the effectiveness and efficiency of group work through moderation. Facilitation tasks are diverse and

multifaceted. On the one hand, it involves guiding the agenda and promoting the creation of high-quality group artefacts. On the other hand, it's the facilitation's task to support group member interaction and prevent or help resolve conflicts (Przybilla et al. 2019). It can be challenging for the facilitators to meet these diverse task requirements, as they form a "multifaceted profile" (Leimeister 2014).

For several years, there have been initial research approaches discussing facilitation by chatbots, which can also promote remote working and support teams when no human facilitation (HF) is present (Tegos et al. 2015). Chatbots have undergone substantial development and have grown from simple rule-based agents to artificial intelligence-capable systems. As a result, it's becoming more appealing for companies to invest in providing chatbots to support their services. One example of the enhanced capabilities that chatbots now possess is the language processing model GPT-4. This model, released by OpenAI in 2023, garnered significant attention. GPT-4 uses deep learning to analyze, summarize, translate, or rephrase texts, relying on millions of training datasets. It can perform various tasks, such as generating code, having conversations with people using a broad knowledge of information, or picking up and interpreting moods from texts (OpenAI 2023). Although artificial intelligence (AI) capabilities are already considerable, some limitations still exist. For example, if an AI chatbot is asked something it doesn't know, it must stop at that point and hand over the conversation to a human (Singh et al. 2019). Significant limitations also exist in the context of e.g., flexible interventions or empathetic actions (Dellermann et al. 2019b), as these each involve a complex interplay of various skills that currently correspond more to a human than to an AI (Dellermann et al. 2019a; Mockenhaupt 2021). Some facilitation tasks particularly require these abilities. To ensure the best possible execution of those tasks that aim to promote interpersonal relationships and to relieve the HF of the numerous and diverse tasks, we propose a hybrid intelligence solution for facilitation purposes. Hybrid intelligence is a construct in which humans and machines perform tasks according to their strengths, thereby jointly creating higher value results (Dellermann et al. 2019b).

To approach this goal, we leverage results from an underlying study as well as related research. We derive tasks for the system parts based on complementary strengths and use these insights to derive design requirements (DRs) for a Hybrid Facilitation System (HFS): *RQ1. How to design a HFS according to a task allocation based on complementary strengths of humans and AI?* Derived DRs are merged into design principles (DPs) that guide the implementation of the HFS. The evaluation part is concerned with: *RQ2. How is the Hybrid Facilitation evaluated by a) end-users and b) the human facilitator?*

The paper offers a manifold contribution: On a conceptual level, we provide a framework to allocate tasks in a HFS according to principles of Hybrid Intelligence. Practically, we offer design knowledge and an artefact that facilitates the implementation and evaluation of Hybrid Intelligence systems, paving the way for more effective and efficient group collaboration in remote settings.

15.2 Theoretical Background & Related Work

15.2.1 Facilitation of Collaboration

Facilitation, one of three crucial roles in Collaboration Engineering, creates and leads the implementation of non-recurring collaboration processes (Leimeister 2014). This dynamic process aims to enhance participant relationships and group communication through interactional activities. Additionally, it provides structure to meetings and aids in managing tasks to support the development of group artefacts through task-related activities (Voigt & Bergener 2013). During facilitation, the facilitator ensures clear communication, resolves queries or comments, promotes participation and responsibility among all participants, and actively builds relationships among them by employing empathy and fostering trust (Beranek et al. 1993; Clawson et al. 1993).

Facilitation may be carried out by a group member, the group's leadership, or an external organization, and may include the use of group support systems (GSS) to provide various facilitation tools, such as assisting with idea generation or group decision-making (Przybilla et al. 2019). Facilitation actively promotes relationships and communication through interactional activities (I) (Dickson et al. 1996) and assists in structuring group meetings and providing task-related interventions (T) to support the creation of group artefacts (Beranek et al. 1993; Dickson et al. 1996).

15.2.2 Hybrid Intelligence

In 2019, the concept of hybrid intelligence emerged, combining the strengths of human and artificial intelligence in a socio-technical system capable of transcending current human and AI limitations (Dellermann et al. 2019a). It's not about using AI to extend human intelligence capabilities or vice versa, but rather, it involves both entities working together on complex problems and sharing tasks based on their strengths (Akata et al. 2020). A Hybrid Intelligence System (HIS) as defined by Dellermann et al. (2019a), exhibits three core traits. Firstly, a common result is developed by both human and AI. Secondly, superior outcomes are achieved through the collaboration of AI and humans, outcomes that couldn't be achieved if the agents worked independently. The goal is to work more effectively and efficiently through task distribution and later integration of the products. The third characteristic involves continuous learning, with each agent improving individually, as well as the whole sociotechnical system. While examples of hybrid intelligence are few, its potential is vast. If both human and AI learn more intensively from each other, they can solve complex tasks at a "superhuman level of performance" (Dellermann et al. 2019a). Application areas span all disciplines where complex tasks need to be solved, from educational support for children with learning disabilities to assisting in hospitals or processing large datasets in research (Akata et al. 2020).

With the advancements in technology, AI-assisted facilitation is becoming more prevalent. Previous research has focused on using GSS (Leimeister 2014) to facilitate collaborative

work, but with the emergence of conversational agents and other AI technologies, the focus has started to shift towards AI-supported facilitation (Cvetkovic et al. 2023). For example, there have been initial research approaches that address facilitation by chatbots, which can also promote remote working and support teams when no HF is present. In their paper, Strohmman et al. (2017) introduce an AI-based virtual facilitator called "brAInstorm" that assists in idea generation during electronic brainstorming sessions. Tavanapour & Bittner (2018) present a proof-of-concept for chatbots as facilitators, where a chatbot helping users through an idea platform resulted in improved target achievement and shorter task times, although HFs were still preferred for user satisfaction and process support. Visualizing and organizing ideas using a chatbot has been found to be useful in distilling and enriching information and thereby assist in tasks and processes (Cvetkovic et al. 2023). While Leimeister (2014) and others have focused GSS for facilitation, these studies often fail to address the nuanced integration of AI and human facilitators, particularly when it comes to empathy and conflict resolution—key components of successful facilitation (Siemon, 2022).

One potential solution to this challenge is to adopt a Hybrid Intelligence approach, which combines the strengths of AI with human intelligence (Dellermann et al. 2019b). For example, an AI-powered chatbot could help facilitate discussions and collect input from team members. Still, the final decision-making process could be left to an HF who can assess the team's needs and make necessary adjustments to the process. By incorporating human intelligence, the potential drawbacks of relying solely on AI can be mitigated. For example, AI-powered facilitation can lack empathy and emotional intelligence, which are essential for effective teamwork and creative collaboration. A hybrid intelligence approach can bridge this gap by ensuring that the facilitation process is tailored to meet the unique needs of the team while still benefiting from the speed and efficiency of AI technologies.

Toward addressing the potential solution, in the following, we describe the development and evaluation of a prototypical HFS, in which a HF and a chatbot facilitator (CF) simultaneously co-facilitate creative teamwork by taking over different tasks.

15.3 Method

This work employs the design science research (DSR) paradigm, which emphasizes the creation of new artefacts through applied research. The aim of this work lies in contributing prescriptive knowledge to a nascent design theory (Gregor & Hevner 2013) of Hybrid Facilitation by developing DRs and DPs for a HFS validated through a prototypic instantiation. The work follows the approach of Peffers et al. (2007) and is conducted in three cycles (Table 1). In the 1. Problem identification phase, we find shortcomings of solely human or solely AI-executed facilitation from both literature and the analysis of an underlying study. 2. To approach the objectives of a solution, we undergo a complex process consisting of the following steps: We allocate tasks in the HFS according to the

complementary strengths of humans and AI. We derive 32 DRs based on several elements: a) the task allocation framework, b) improvement suggestions from an underlying study, and c) related work on developing chatbots and virtual facilitation. These are merged into six DPs. 3. In the development phase, based on the DRs and corresponding principles, we develop a chatbot and an interaction abstract for the HF to follow in order to successfully interact in a HFS. 4. Finally, we demonstrate and 5. evaluate the system by assessing the implementation success of the given principles with potential users and the HF, as well as by analysing the working outputs and the log files from user tests. As we outline the problem relevance in the Introduction and Related Work, in the following, we focus on the subsequent steps of the DSR approach by Peffers et al. (2007).

Figure 1. DSR Approach based on Peffers et. al (2007)

	Cycle One	Cycle Two	Cycle Three
Problem Identification	Problem relevance (literature-based identification related to group facilitation)	Practice-based identification of problems as a result of performing Cycle One	Reflection and refinement of problem relevance
Objectives of a Solution	Identification of suitable study setup and task for facilitated virtual collaboration	Task allocation framework; Comprehensive literature search for derivation of DRs; analysis of chat logs from Cycle One for derivation of further DRs	Refinement of DRs and development of DPs
Design & Development	Interactions scripts for the facilitators; setup of workspaces on Slack and assignment of participants	HFS prototype based on refined interaction scripts and DRs	HFS implementation based on DPs
Demonstration	52 user tests with 255 students split in teams of 3-4	Pre-test with a few users and the CF developer in HF role	Instantiation in three user groups
Evaluation	User tests with quantitative evaluation of performance and collaboration factors	User test with qualitative evaluation of the HFS for initial proof-of-applicability	Evaluation with a mixed-method approach

15.3.1 Objectives of a Solution

15.3.1.1 Task allocation in a Hybrid Facilitation System

In this step, we develop a task allocation framework, to inform our design by assigning the tasks to the facilitation agents according to the principles of hybrid intelligence (Dellermann et al. 2019a). To approach the framework development, the facilitation tasks between human and AI are divided according to their strengths. First, the required subtasks that are necessary to perform each facilitation task are identified. They are derived from Beranek et al. (1993) (pp. 150-152), and Clawson et al. (1993) and categorized into task (T) or interaction (I) activities according to Leimeister (2014). Subsequently, strengths are identified for each facilitation task that is particularly relevant for task performance or has a conducive effect. The work of Dellermann et al. (2019b) forms the basis of the complementary strengths. According to this and other research, human intelligence is characterized by flexibility and transfer, empathy and creativity, and common sense, while AI is characterized by pattern recognition, speed and efficiency, and probabilistic reasoning, among others.

From the assignment of facilitation tasks and suitable strengths, it is derived whether the human, the AI, or both facilitation agents are responsible for a facilitation task. If the latter

is the case, we examine which subtasks the human or the AI should perform and whether one of the facilitation agents should support the other or whether both should perform their tasks equally. Finally, we map these tasks into a higher-level model for an allocation of responsibilities between facilitation agents. The division of the 18 facilitation tasks between humans and AI resulted in the model shown in Figure 2. An example of the reasoning for the division can be found in Appendix 1.

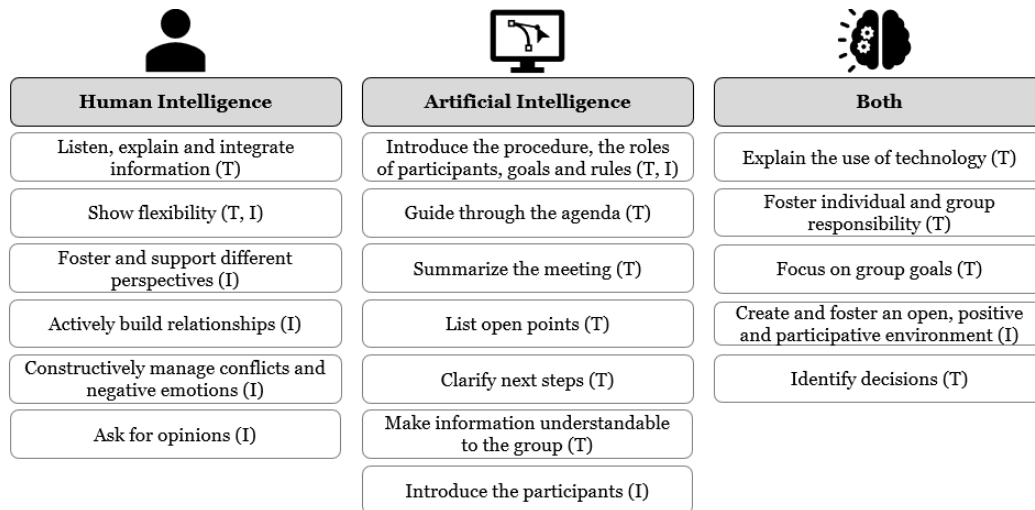


Figure 2. Task Allocation in the HFS

15.3.1.2 The underlying study

The first study served to assess the viability of performing user tests in the field of facilitated virtual collaboration on chat-based platforms as well as to investigate human vs. chatbot facilitation and derive necessary improvement suggestions. It was conducted involving 255 students of an information systems course and 6 facilitators, randomly divided into 52 small teams of 3-4 participants each. For this, students were awarded points toward the course. They were tasked with creating an Empathy Map via a Slack chat. Half of the groups were guided by a HF, while the other half thought they used a chatbot in a Wizard-of-Oz study, where they believed they were communicating with a functional IT system. The aim was to examine factors influencing collaboration and measure performance, comparing results from HF and CF. Improvement suggestions from this study were derived from the survey answers, chat logs, and submitted work outputs, using a qualitative content analysis. This process, based on the method by Früh (2004) and Kuckartz (2018) simplifies text, examining and categorizing its key features. The aim is to combine similar features into a feature class called codes. For this purpose, the number of occurrences is noted next to the content-related problem or improvement suggestion to later determine the relevance of the individual improvement suggestions. First, the survey answers are examined for suggestions for improvement or problems identified. These are converted into codes with which the answers are coded initially and refined in a second iteration (by adding additional or removing irrelevant codes). The same procedure is repeated on the chat logs and working outputs. Coding is performed by two independent

coders using the MAXQDA software. As we assembled a lengthy list of improvement suggestions and could not implement all of them, we coded them for relevance, with factors including the potential for improvement, implementation effort, and the frequency of related problems. Some of the resulting improvements were e.g., the task processing time should be extendable; if there are too many arguments, a selection should be made via voting; important concepts or parts of the task should be written bold in text; a warm-up should be introduced and serve the group to get to know each other; time management.

15.3.1.3 Derivation of the requirements from the literature

In the last requirement derivation step, a systematic literature review according to Tranfield et al. (2003) is conducted to derive requirements from the literature. The systematic literature search is based on defined databases and search terms (Appendix 2). The terms contain combinations of individual search terms that have been linked by an AND or OR operator. For this, several research areas are considered and researched separately: collaboration engineering and facilitation of teamwork, human and artificial intelligence, recommendations for developing virtual facilitation of creative teamwork, and general recommendations for the development of chatbots. By analyzing this literature, we identified key themes and best practices that guided the formulation of specific requirements for our system. For instance, insights from the literature on human and artificial intelligence helped to shape our understanding of task distribution between the CF and HF, while recommendations from virtual facilitation studies highlighted the importance of user experience and interaction quality, which were directly translated into system design parameters. Results from the first two searches are summarized in related work; in the following we present the results of the search for recommendations for the development of chatbots and virtual facilitation, respectively.

General requirements for chatbot development

The search for general recommendations for the development of chatbots resulted in the following insights: It emphasizes that chatbots should exhibit human traits, as they communicate with users through natural language. This is a principle that has been recognized in the field for some time (Boden et al. 2006). This is one of the three key areas identified by Chaves & Gerosa (2021), that should shape the character of a chatbot - conversational intelligence. It refers to the chatbot's ability to actively participate in a conversation, being aware of the discussed topic, the context of the conversation, and the flow of the dialogue (Chaves & Gerosa 2021). Furthermore, the chatbot should actively demonstrate the following properties: it should be proactive, taking the initiative in the conversation to drive it forward. This includes introducing new topics, asking follow-up questions, and providing essential information necessary for solving a task (Chaves & Gerosa 2021). In addition, the chatbot should be capable of learning, expanding its knowledge from past conversations and incorporating this knowledge into future conversations (Stucki et al. 2020; Wei et al. 2018). Lastly, it should work conscientiously,

meaning it should be able to follow the flow of the conversation and understand each contribution (Chaves & Gerosa 2021). This also includes having general knowledge, memory, and a sufficiently large vocabulary. The chatbot should be able to provide possible text summaries when needed, draw conclusions, and master mathematical calculations (Boden et al. 2006; Wei et al. 2018), and be able to answer questions about or explain its functions in correct and complete answers (Boden et al. 2006; Chaves & Gerosa 2021). The second key area refers to social intelligence (Chaves & Gerosa 2021). This is about social behavior that is used to achieve specific goals. On the one hand, it includes emotional intelligence, which is characterized by empathy, respect, understanding and emotions and aims to improve the relationship with users (Chaves & Gerosa 2021; Wei et al. 2018). In addition to emotional intelligence, it is also important that the chatbot uses precisely formulated language that suits the users and their needs (Chaves & Gerosa 2021). The chatbot should also be familiar with essential customs and practices, such as polite behavior and conversational conventions. Morally reprehensible behavior must be recognized based on social guidelines in order to be able to react accordingly. For example, users often use exclamations that interrupt the current flow of conversation. They use them to test the conversational characteristics of a chatbot and expect a reaction from it, such as to the statement "Nice weather today!". The chatbot should find a suitable response to these so-called social intentions (Stucki et al. 2020). Error checking is also necessary, for example to intercept questions for which no answers are available in the knowledge base (fallback answers). There should also be no dead ends in the conversation to avoid frustration during use (Landbot.io 2022). The chatbot should clarify the content through a meaningful user interface, for example by using a menu as a help function (Chaves & Gerosa 2021). Finally, the third key area refers to the anthropomorphism (Chaves & Gerosa 2021). An essential component is the possession of an identity that makes it possible to assign the chatbot to a group. The chatbot should display corresponding socio-demographic characteristics, such as gender, name and age (Wei et al. 2018). A visual representation by an avatar is also relevant (Boden et al. 2006; Chaves & Gerosa 2021). It should also have its own personality. This refers to characteristics of the chatbot that shape the style of interaction and describe its character. If the chatbot exhibits inconsistencies here, this can lead to a loss of trust (Chaves & Gerosa 2021). Humor is an essential characteristic that can have a positive impact on the entire conversation (Wei et al. 2018).

Requirements for virtual facilitation

Results from the search about the recommendations for virtual facilitation relate to the implementation phase of group work and do not refer to preparatory and follow-up activities. They suggest that communication with the chatbot be conducted via a popular instant messenger to minimize the time required for users to familiarize themselves with its use (Debowski et al. 2022; Morrison-Smith & Ruiz 2020). The chatbot should be developed intelligently using AI and machine learning to mimic human-like linguistic behavior (Ait Baha et al., 2023; Debowski et al. 2022). Further, an interaction concept

should be developed that forms the basis for the implementation of the chatbot and contains all of the chatbot's utterances (Bittner et al. 2019a). Moreover, it is suggested that a chatbot should not have to carry out the facilitation entirely on its own. Instead, it can support human facilitators and take over essential tasks from them. This requires intelligent use of proactive actions by the chatbot and successful cooperation between the human and chatbot-based facilitation (Følstad et al. 2021; Strohmman et al. 2018). Key tasks that can be transferred to the chatbot include process support (Kim et al. 2021; Strohmman et al. 2018) and documentation of results (Strohmman et al. 2018; Tavanapour et al. 2019). Other important aspects include time management, providing tips and hints, reminding about the task, the remaining time, or the goals, and explaining methods (Calvaresi et al. 2023; Følstad et al. 2021). In the introduction, participants should be informed about the goal and conditions and at the end, they should officially bid farewell (Bittner et al. 2019a). The virtual facilitation should be able to filter out repeating ideas, give feedback on arguments, and be able to jump back to previous work steps (Mukherjee et al. 2023; Tavanapour et al. 2019). Information structuring, such as topic clustering, can also support the group during their work. Additional functions, such as an automatic analysis of existing data or a built-in search, can further enhance the facilitation process (Calvaresi et al. 2023). For creative work to be successful, virtual facilitation should also be user-friendly (Morrison-Smith & Ruiz 2020). For example, polls can simplify the voting process, buttons can simplify the selection process and multi-media can simplify the absorption of new information (Debowski et al. 2022; Strohmman et al. 2018). It is also important that group work takes place in a trusting and transparent environment and that different perspectives are encouraged in order to achieve better results (Morrison-Smith & Ruiz 2020; Schöppl & Steinhoff 2023). Small talk should also be used to raise the mood and build trust (Cagiltay et al. 2015; Morrison-Smith & Ruiz 2020). Collaboration rules can help to positively influence the relationships between users (Berente & Howison 2019; Strohmman et al. 2018). The chatbot should appear human to a large extent through the use of natural language and the display of feelings and empathetic behavior and be given a name accordingly (Ait Baha et al. 2023; Debowski et al. 2022). Anecdotes and jokes should be used to cheer up the group (Bittner et al. 2021). Group work should be actively and passively supported by the chatbot, for example by motivating the working group as well as reacting to conflicts and problems, such as a longer user breaks and indicates the task progress (Gimpel et al. 2023; Strohmman et al. 2018).

15.3.2 Development and Demonstration

The development of the HFS is approached from two angles: first, interaction scripts for a chatbot, i.e., the CF are developed and incorporate all DRs. For the HF, an abstract is prepared with the tasks assigned to the HF, and examples where the HF could step in are given. In the second DSR cycle, the RASA framework (Rasa Technologies Inc, 2021), an open-source solution for conversational AI development based on BERT language model, was chosen for implementing the CF. Due to the focus on experiment controllability rather

than eloquence of the CF, no state-of-the-art large language models are used. Although the standard RASA does not support multi-user sessions, its code can be extended to do so. The implementation used various RASA modules, such as NLU (containing training data utterances), stories (dialog trees), rules, and custom actions, each with its specific functions, ranging from training CF's understanding of user utterances to defining CF's decision-making behavior. E.g., for saving arguments of users or initiating voting, custom actions were written in Python. The developed CF is integrated into Slack, a known collaboration software, using the Slack API which operates via an ngrok deployment service (Casey 2022).

	Facilobot APP 13.04 The task will take 35 minutes from now on. It is now 13:04:26. The task expires at 13:39:26. When you are ready, tell me, that you want to start !		C 13.20 @Facilobot Give us an example
	HF 13.04 Feel free to interact with the chatbot		Facilobot APP 13.20 An Example in the BEHAVIOR section for a road-damage-reporting app would be: Citizens and visitors ride less bikes because of the potholes. They try to avoid the potholes, and thereby impede the incoming traffic. They sometimes have accidents because they don't see the potholes.
	C 13.04 @Facilobot lets get started		H M 13.21 @Facilobot Save email-ping-pong (since calendars are not mutually viewable)
	Facilobot APP 13.05 Hello participant_in_161, J M , C and H M , my name is Facilobot ! 😊 I am a chatbot and together with the HF I will support you in carrying out the task! 🤖📅🕒 I will guide you through the task, document your work results, guide the voting and track the time. If you have any questions, please tag me @Facilobot ! @HF would you like to introduce yourself briefly and say what your task is in the groupwork?		Facilobot APP 13.21 I have saved your argument: Email-ping-pong (since calendars are not mutually viewable) under the number 3. The number is used for deletion if required. You have reached the minimum number of 4 arguments. Let me know when if you want to start voting for the best 3 arguments.
	HF 13.05 Words marked in bold are example commands that you can use Hi I'm HF ! I worked as a technology manager. With the help of e.g., Design thinking workshops, we search for and find user-centered solutions. Today I am here to support you with the task, if you have any questions, feel free to write it! @Facilobot continue		HF 13.21 Can you guys think of a few more things?
	Facilobot APP 13.05 Thank you, if you have any questions in between, don't hesitate to ask me! To find out who are you working with here, tell others something about you (name, job). Otherwise, its so anonymous here...		Participant_in_161 13.21 @Facilobot save At the same time, in addition to student appointments, other appointments for committees come in and there is a scheduling chaos.
	C 13.05 @Facilobot what is your name?		Facilobot APP 13.21 I have saved your argument: At the same time, in addition to student appointments, other appointments for committees come in and there is a scheduling chaos under the number 4. The number is used for deletion if required. You have reached the minimum number of 4 arguments. Let me know when if you want to start voting for the best 3 arguments.
	Facilobot APP 13.04 I am Facilobot ! 🤖		

Figure 3. Snippet of the HFS

To equip the CF with domain knowledge, training data from the underlying study were incorporated into the CF's knowledge base. After this, a pre-test was carried out with three participants and the CF's developer as the HF. The chat log from this pre-test was used as additional training data for the CF, and any errors were corrected accordingly. The empathy map task from the underlying study was reused, since it is easy to explain and is based on intensive collaboration and creative contributions from the participants (Osterwalder & Pigneur 2010). Figure 3 depicts translated excerpts of the demonstration of the HFS, as the original demonstration took place in German language. Several incorporated requirements can be seen in the snippets: e.g., time management, introduction, documentation of the results, use of an avatar and emojis, as well as flexible intervention by the HF. Participants' and HF names are hidden for anonymity purposes.

15.3.3 Evaluation

In DSR, IT artefacts should be evaluated and iteratively improved. The method of user tests is used to verify the validity of the claims that motivated the artefact's development (Peppers et al. 2007). In three user tests, the developed DRs and DPs instantiated via the HFS are

evaluated with three groups of participants ($n=12$), consisting of computer science students and recent graduates working in IT/engineering fields. Participation was voluntary, most of the participants did not know each other and had no experience with the empathy map method, making it all the more important for the facilitation to be successful. Participants collaboratively constructed an empathy map, which is a collaborative visualization used to articulate what we know about a particular type of user. According to Gray (2009), each group task should be carried out by three to four individuals for the successful execution of the empathy map method. The HF is a person experienced in moderating design thinking methods and familiar with the empathy map method. The evaluation investigates whether the DPs could be fully and functionally implemented, to answer the second research question. After the completion of the group work, which lasted between 40-50 minutes, due to participants' time restrictions a written questionnaire was used to survey them about their experience with the HFS. For this purpose, the DPs are presented to the participants and questions are asked about the fulfillment of each DP. Both open and closed questions are asked to enable comparability between the answers, but also to allow justifications for the answers and to give suggestions for improvement (Hyman & Sierra 2016). In total, participants answered 38 questions. Closed questions are answered using a 5-point Likert scale, where 1 = strongly agree, and 5 = strongly disagree (Appendix 4). The HF is interviewed to gain deeper insights about the HFS from their perspective since in such an early stage of development it is crucial to receive feedback from the first line of potential users who are also the true connoisseurs of facilitation and its challenges. The interview guideline is also based on the DPs. In order to make possible comparisons between the answers of the HF and the participants, the HF is prompted to provide a numerical answer on a Likert scale and thereafter explain their answers orally. Finally, the interview is transcribed for qualitative content analysis. In addition, the chat logs and the working outputs are analysed to confirm the successful implementation of the DPs. This was done by backtracking the outputs and logs of each group to the intended interaction concept and by calculating the overlap with it as well as the frequency of each of the operations performed by each HF and CF.

15.4 Results

To approach the RQs postulated above, we present the results in three sections: 1. requirements and principles necessary to develop an HFS; 2. results of the evaluation of the given principles by end-users and the HF and 3. we supplement these results with the evaluation of the chat logs and working outputs, for a comprehensive overview.

15.4.1 Design Requirements and Principles

Following four core elements: the concept of hybrid intelligence and the facilitation tasks assigned to the HF or the CF, general recommendations for chatbot development, guidelines for designing virtual facilitation for creative work, and suggestions for

improvement from the preliminary study, all collectively served as the foundation for deriving DRs. Resulting DRs can be split into three categories: Chatbot facilitation, Human Facilitation and technical requirements. Figure 5 from the Appendix 3 presents an overview of the requirements together with the corresponding source chapters. DPs make it possible to describe guidelines for the development of artefacts on a less abstract level (Gregor et al. 2020). Six DPs have been derived from the requirements according to Koppenhagen et al. (2012) and are defined in the following. Unassigned requirements pertain to the nature of the exercise and the communication medium used (T1, T2, T3, T4, T6).

DP1. Hybrid intelligence: The collaboration between the HF and CF is conducted by both undertaking different tasks that build upon each other. Parallel responsibilities should be avoided. The responsibilities of individual tasks are verbally clarified to the working group. Facilitation tasks should be divided between the facilitation agents according to their strengths. Facilitation tasks that have not received clear responsibility must be broken down into their subtasks and assigned to either the CF or HF. They should not be managed in parallel by both facilitation agents, as this can cause confusion among the participants (CF1, CF2, HF1, HF2).

DP2. Task Takeover: Due to the limited capabilities of the CF, the HF should step in when it reaches its limits. This is particularly the case when the CF cannot answer certain questions or technical issues arise (CF7, HF3).

DP3. Friendly facilitation team: HF and CF interact respectfully and make it clear to the working group that they form a facilitation team together. They present themselves as a unit. The facilitation team interacts respectfully and is friendly with the participants and encourages intensive collaboration. The facilitation agents highly depend on each other's results. Especially the handover of one facilitation agent to the other must be done carefully to ensure a smooth facilitation process. Both must treat each other politely and give the impression that they are a well-coordinated team. This is important so that the working group accepts both and does not try to work with only one of the facilitation agents. To build trust in the working group and create a positive working atmosphere, the communication between all should be friendly and respectful (CF4, CF10, CF12, HF4, HF5, HF6).

DP4. Task-related facilitation: The CF guides the working group through the agenda, clarifying the necessary work steps or tasks and initiating the votes. It stores the contributions and final results that have arisen in the group discussion. It is responsible for time management. Derived from the requirements for the CF, it is primarily responsible for task-related facilitation activities. Thus, it primarily leads the group work but hands over the lead to the HF depending on the task. The primary leadership and corresponding responsibility should also make clear that the working group cannot bypass it and use only the HF (CF3, CF5, CF8, CF9, CF13, CF14).

DP5. Interaction-related facilitation: The HF supports the exchange of opinions and decision-making process during the collaboration so that the working group constantly focuses on their group goal, incorporates diverse perspectives, and thus can work on a high-quality group result. HF also supports the establishment of relationships by promoting the exchange of opinions, ensuring an open participatory environment, and trying to resolve conflicts between group members. As few predefined statements as possible are used to be able to react flexibly to the group conversation. Following the requirements for the HF, HF is primarily responsible for interaction-related activities. These should be as little predefined as possible to ensure the best possible support during the discussions and situational engagement with individual participants (**HF7, HF8, HF9, HF10, HF11, HF12**).

DP6. Truthful appearances: The avatar representations of the CF and the HF support their appearance as a robot and as a human, respectively. This is beneficial for a realistic impression of the hybrid facilitation. The participants should not get the impression that they are being deceived in the presentation of the facilitation team (**CF11, T5**).

15.4.2 User evaluation results

Participants answers are presented along the average value (M) from their survey answers, the HF's numerical answer is also presented for comparison. Concerning DP1, both the participants (M=1.92) and the HF (1) felt well-prepared for group work and interactions in the chat thanks to the preparatory document. Both participants (M=1.92) and the HF (2) acknowledged the distribution of tasks between the HF and CF as mainly meaningful and did not experience parallel responsibilities (M=1.5, HF=1). The participants could distinguish which facilitation agent to approach with questions and remarks (M=1.75), although two individuals were not clear about this division of roles. Some suggested different models of interaction, like having the facilitator act as a supervisor or having the facilitator be the only one to interact with the CF. One participant said: "I thought it was good the way it was, as the CF handed over to the HF in case of doubt if the CF couldn't answer a question", whereas another one noted "The communication between the CF and the HF confused me." These differing opinions likely arose due to the confusion about which facilitation agent to address for a specific sub-task. It is therefore an important factor that they understand this division precisely. HF admitted to also being confused by communication between them and the CF: "[I think...] HF should not prompt the CF actively, as this can hinder the group dynamics." HF further states that the abstract for the interaction made her part very easy: "If I had to do this daily, it would be very easy for me ... to work [with the abstract]." Log files further depict the distribution of the task load, wherein the HF messaged in 16.7% of the cases, while CF overtook 35.8% from the overall interaction.

Regarding DP2, participants noted the CF's ability to respond to questions and instructions with mixed results (M=2.33). While some were impressed and driven to test the limits of

the CF's comprehension, others noted a misalignment between their questions and the CF's responses. The HF, on the other hand, was perceived as highly competent in dealing with the participants' queries and instructions ($M=1.25$). She also intervened when technical problems or misuse of commands arose, and didn't experience this takeover as problematic (2).

The evaluation revealed consensus on DP3. The CF and HF presented as a team ($M=1.75$, $HF=2$), creating a respectful and friendly environment ($M=1$, $HF=1$). The use of warm-ups and emojis enhanced the atmosphere, prompting participants to work together intensively ($M=1.75$). The participants predominantly had the impression that they worked together in an atmosphere of trust ($M=1.67$). One participant criticized that the CF's presence inhibited human interaction. HF agrees: "Somehow the team's collaboration was hindered by the CF. It felt like everyone communicated individually with the CF and there was little communication between the group participants."

For DP4, participants agreed that the CF successfully guided the agenda ($M=1.42$), clarified tasks ($M=1.17$), and managed time effectively ($M=1.58$). They also appreciated the CF's assistance in documenting work results ($M=1.58$). The HF, on the other hand, saw their main role as supporting the CF in guiding group discussions, reminding participants of commands, correcting misuse of commands, and identifying open questions. The HF did not have to support the CF with time management, documentation or guide it through the agenda. Additionally, HF felt that the cognitive load of their work was significantly lower than if they had conducted the moderation without a CF, as they did not have to pay constant attention to the participants during the moderation: "It took a lot of pressure off me. [...] The cognitive load is different than when I have to be constantly alert and respond to four people at the same time. That's a big difference." An interesting privacy aspect about was mentioned: "Occasionally the command '@Facilibot' was forgotten. One could consider whether the CF can read everything and proactively respond if input matches one of its actions which would in turn make a deviation from data security/privacy necessary."

Concerning DP5, participants considered that the HF actively promoted the exchange of views and the decision-making process ($M=2$). Nevertheless, some felt there was insufficient stimulus for a discussion of the developed arguments, and thus, the arguments were not deeply reflected. Overall, the HF was able to support the participants with many questions or topics ($M=1.25$). There were no questions or topics where neither the HF nor the CF could help the group with ($M=1.00$). Participants unanimously confirmed the absence of conflicts requiring HF's intervention.

With DP6, participants were clear in their perception of the HF and the CF, and their display pictures reinforced this. Additionally, HF notes that: "The quick responses to the votes in particular gave me the impression that the CF is a real chatbot".

Overall, the group work was successfully completed in the opinion of the participants ($M=1.25$, $HF=1$). The structured approach, the coordination between the working group

and the facilitation team, and the CF's time management were also positively highlighted. In addition, the group work was described as "very productive [because] it was very structured, time-effective (...) and comprehensible". Participants also had the impression "to contribute (...) all (...) thoughts openly (...) (to be able to) and not to weigh (...) [to have to] for a long time". However, another person noted that "the quality (...) [suffered] from the time and possibly also from the format, because from his point of view for empathy maps there needs to be more room for creativity and also the time to put oneself in a person's shoes." Another person also believes that design thinking tools such as the Empathy Map "benefit more from direct and verbal exchange."

15.4.3 Evaluation of the chat logs and working outputs

Analysis of the logs and outputs revealed that the empathy map task was successfully performed in each of the three groups. The goal of the task was reached. Participants intensively collaborated and also used the social intents to explore the CF's knowledge and limits. The CF was able to solve queries by recognizing intents but also managed to act according to predefined rules. It displayed comprehensive knowledge about the empathy map method. The use of buttons was problematic as it produced duplicate answers depending on how many participants clicked. The task allocation was successful, with each facilitation agent performing their assigned tasks. Tasks requiring joint effort were effectively delineated into subtasks and executed collaboratively. The majority of CF contributions (78.6%) were task-related, with 42.64% focused on managing arguments, 34.55% on guiding the agenda, and 19.83% on introducing group work. Both task-related and interaction-related activities accounted for 19.83% of the contributions, while only 1.47% of contributions were purely interaction-related. Evaluation of the HF contributions revealed that 67.63% were interaction-related, 23.74% were overarching activities, and 10.08% were task-related. The most frequent contributions were related to fostering an open, positive participatory environment (36.69%), followed by building relationships and promoting opinion exchange (around 15% each). Supporting technical questions and problems accounted for 11.51% of contributions, while other categories such as reminding participants of functions and objectives, intervening in incorrect use, promoting responsibility, supporting decision-making, and addressing questions and comments, each accounted for less than 10%.

15.5 Discussion

This paper follows a comprehensive process to develop and evaluate design knowledge for a HFS. Initially, the analysis of the underlying study suggested improvements mainly related to task performance, collaboration, and facilitation, aligning with the requirements from the task allocation framework and literature review. The framework divides 18 facilitation tasks based on Dellermann et al. (2019b), which made it possible to delineate the tasks performed by CF and the HF in order to optimize their collaboration, as called for

by Bittner et al. (2019) and Traumer et al. (2017). The latter enabled a complex investigation of the research domain and deals with general requirements for chatbot development and virtual facilitation and highly overlaps with the requirements dictated by the task allocation framework. All together, these requirements are summed up in six DPs and represent the answer to RQ1. For RQ2, concerning the user and HF evaluation of the HFS, positive findings were reported regarding the harmony, trust, and cognitive load of the facilitation team, as well as the quality and efficiency of group work. Participants appreciated the teamwork between the HF and CF, though some suggested improvements in task division. The HF found the collaboration “pleasant,” stepping in when the CF reached its limits, such as when it couldn’t provide suitable answers. The CF’s default answers and task handover were helpful, resembling team facilitation. Participants further felt that both facilitation agents worked as a team. Thus, the HIS characteristic of joint elaboration of results by human and machine was fulfilled. In addition, the log files show that the planned division of facilitation tasks was successfully implemented, with each facilitation agent performing only tasks that were originally assigned to them based on their strengths. These also imply that HF’s cognitive load compared to facilitation without CF was lower, which HF confirms in the interview. However, a server crash highlighted the need for HF to be technically equipped to intervene when necessary.

The contribution of the study is manifold: On the theoretical level, the task allocation framework itself can serve as a high-level artefact for task orchestration in human-machine ensembles (Recker et al. 2023), as it contains tasks found in much of the day-to-day collaborative work, which makes it highly adaptable and generalizable to a wide range of areas. As such, it can guide task allocation or handover according to complementary strengths in many future envisioned hybrid or augmented systems. The framework can thus be positioned as a nascent design theory, i.e., knowledge as operational principles according to Gregor & Hevner (2013). The framework further contributes to the concept of Hybrid Intelligence by expanding the socio-technical perspective on human-AI collaboration. Prior literature has primarily focused on AI systems that either substitute or augment human performance (Akata et al., 2020), but our framework demonstrates how humans and AI can collaborate through distinct, complementary roles. Since this framework was not only successfully implemented, but the task division also perceived meaningful by the participants and the HF, we highly encourage researchers to reuse it and reevaluate in different contexts.

Practically, the six design principles for a HFS can serve as guidelines for creating systems that combine human and AI capabilities and can also be positioned as middle-ranged DSR contribution. Designers can use them to ensure that collaborative tools enhance group collaboration, facilitate decision-making, and streamline processes. The implemented artefact itself is on the lower level as it represents a situated instantiation of the artefact (Gregor & Hevner 2013). Nevertheless, this prototype demonstrates the feasibility of a HFS by creating a chatbot and integrating it into Slack using the RASA framework and

developing and abstract for the collaboration for the HF. The prototype showed promise in an attempt to implement a HFS, and we have no doubt that rapid technology advances will make a development of such systems in the future much easier.

This work is limited to two features of hybrid intelligence according to Dellermann et al. (2019b), excluding the continuous learning. Further limitation could be in the complementary strengths according to Dellermann et al. (2019b), which are a key foundation of this work. The capabilities of machines have already developed significantly in recent years, meaning they could in the future take over more and different tasks. For example, since 2018, more research has been taking place on the development of empathetic chatbots (Pamungkas 2019). The advances in AI are rapidly profusing what was so far thought to be solely characteristic for humans (e.g., large language models can now display creativity (Hitsuwari et al. 2023)). Additionally, the task distribution was based on defined facilitation (sub)tasks according to Beranek et al. (1993), Clawson et al. (1993), and Leimeister (2014). Should the facilitation tasks be divided in a different way in the future, this may lead to a different task distribution between HF and CF accordingly. Other limitations include the limited group size and the very small sample size in Cycle Three, which makes the study merely a pilot. The study could be repeated with more working groups in the future. Furthermore, the participants were not representative of the population. A significant negative aspect of the user evaluation was that the human interaction was perceived to be hindered by the CF. This is likely due to the chat-based nature of the communication medium. Future work could look into how HFS would behave and be perceived in different mediums (e.g., a digital assistant as an AI facilitator in a virtual meeting). By splitting the facilitation tasks, the HF felt the cognitive effort was low. Future work could consider how best to use the time gained by the HF to further improve facilitation. During the interview, the HF expressed the idea of being able to imagine supervising three group works in parallel. This could provide an efficiency gain in their work and save manpower in the future. This idea also offers possibilities to tie in with the present work.

15.6 References

- Ait Baha, T., El Hajji, M., Es-Saady, Y., et al. 2023. "The Power of Personalization: A Systematic Review of Personality-Adaptive Chatbots," *SN Computer Science* (4:6), 661.
- Akata, Z., Balliet, D., de Rijke, M., Dignum, F., Dignum, V., Eiben, G., et al. 2020. "A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence," *Computer* (53:8), pp. 18–28.
- Beranek, P., Beise, C. M., and Niederman, F. 1993. "Facilitation and Group Support Systems," in *Proceedings of the Twenty-Sixth Hawaii International Conference on System Sciences*, pp. 199– 207.

- Berente, N., and Howison, J. 2019. "Strategies for Success in Virtual Collaboration: Structures and Norms for Meetings, Workflow, and Technological Platforms," in *Strategies for Team Science Success*, K. Hall, A. Vogel, and R. Croyle (eds.), Cham: Springer.
- Bittner, E. A. C., Oeste-Reiß, S., Ebel, P. A., and Söllner, M. 2019. "Mensch-Maschine-Kollaboration: Grundlagen, Gestaltungsherausforderungen und Potenziale für verschiedene Anwendungsdomänen," *HMD Praxis der Wirtschaftsinformatik*, pp. 1–16.
- Bittner, E.; Küstermann, G.; Tratzky, C. 2019a." The Facilitator is a Bot: towards a Conversational Agent for Facilitating Idea Elaboration on Idea Platforms". In *European Conference on Information Systems (ECIS 2019)*
- Bittner, E.; Mirbabaie, M.; Morana, S. 2021. "Digital Facilitation Assistance for Collaborative, Creative Design Processes." in *Proceedings of the 54th Hawaii International Conference on System Sciences 2021*, S. 370–380.
- Boden, C., Fischer, J., Herbig, K., Liebe, J., Sinning, H., and Spierling, M. 2006. "Chatbots as an Instrument of Planning Communication - Opportunities, Requirements and Perspectives," in *INFORMATICS 2006 - Computer Science for People, Volume 1*, Gesellschaft für Informatik eV, pp. 433–442.
- Briggs, R., Kolfschoten, G., de Vreede, G.-J., and Dean, D. 2006. "Defining Key Concepts for Collaboration Engineering," in *Proceedings of the Americas Conference on Information Systems (AMCIS)*.
- Cagiltay, K., Bichelmeyer, B., and Kaplan Akilli, G. 2015. "Working With Multicultural Virtual Teams: Critical Factors for Facilitation, Satisfaction, and Success," *Smart Learning Environments* (2:1), 11.
- Calvaresi, D., Eggenschwiler, S., Mualla, Y., et al. 2023. "Exploring Agent-Based Chatbots: A Systematic Literature Review," *Journal of Ambient Intelligence and Humanized Computing* (14), pp. 11207– 11226.
- Casey, K. 2022. "ngrok Blog: Deploying ngrok in Production," <https://ngrok.com/blog-post/deploying-ngrok-in-production>.
- Chaves, A. P., and Gerosa, M. A. 2021. "How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design," *International Journal of Human– Computer Interaction* (37:8), pp. 729–758.
- Clawson, V., Bostrom, R., and Anson, R. 1993. "The Role of the Facilitator in Computer-Supported Meetings," *Small Group Research* (24:4), pp. 547–565.
- Cvetkovic, I., Gierlich-Joas, M., Tavanapour, N., Debowski-Weimann, N., & Bittner, E. A. C. 2023. "Augmented facilitation: Designing a multi-modal conversational agent for group ideation." in *Proceedings of the 44th International Conference on Information Systems (ICIS 2023)*.
- De Vreede, G., and Briggs, R. 2005. "Collaboration Engineering: Designing Repeatable Processes for High-Value Collaborative Tasks," in *Proceedings of the Hawaii International Conference on System Sciences*.

- Debowski, N., Tavanapour, N., and Bittner, E. 2022. "Conversational Agents in Creative Work – A Systematic Literature Review and Research Agenda for Remote Design Thinking," in *PACIS 2022 Proceedings*
- Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., and Ebel, P. 2019a. "The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems," in *Proceedings of the 52nd Hawaii International Conference on System Sciences*.
- Dellermann, D., Ebel, P., Söllner, M., and Leimeister, J. M. 2019b. "Hybrid Intelligence," *Business & Information Systems Engineering* (61:5), pp. 637–643.
- De Vreede, G.-J., and Briggs, R. 2019. "A Program of Collaboration Engineering Research and Practice: Contributions, Insights, and Future Directions," *Journal of Management Information Systems* (36:1), pp. 74–119.
- Dickson, G., Lee-Partridge, J. E., Limayem, M., and DeSanctis, G. 1996. "Facilitating Computer- Supported Meetings: A Cumulative Analysis in a Multiple-Criteria Task Environment," *Group Decision and Negotiation* (5), pp. 51–72.
- Følstad, A., Araujo, T., Law, E. L.-C., et al. 2021. "Future Directions for Chatbot Research: An Interdisciplinary Research Agenda," *Computing* (103), pp. 2915–2942.
- Forgionne, G., and Newman, J. 2007. "An Experiment on the Effectiveness of Creativity Enhancing Decision-Making Support Systems," *Decision Support Systems* (42:4), pp. 2126–2136
- Früh, W. 2004. *Inhaltsanalyse. Theorie und Praxis*, 4th ed., UTB GmbH.
- Gimpel, H., Graf-Seyfried, V., Laubacher, R., et al. 2023. "Towards Artificial Intelligence Augmenting Facilitation: AI Affordances in Macro-Task Crowdsourcing," *Group Decision and Negotiation* (32), pp. 75–124.
- Gray, D. 2009. "Empathy Map," *Gamestorming*. (<https://gamestorming.com/empathy-map/> accessed 03.03. 2022)
- Green, N., Tappin, D., & Bentley, T. (2020). Working from Home before, during and after the COVID- 19 Pandemic: Implications for Workers and Organisations. *New Zealand Journal of Employment Relations*, 45(2), 5-16.
- Gregor, S., and Hevner, A. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37), pp. 337–355.
- Gregor, S., Chandra Kruse, L., and Seidel, S. 2020. "Research Perspectives: The Anatomy of a Design Principle," *Journal of the Association for Information Systems* (21:6), 2.
- Hitsuwari, J., Ueda, Y., Yun, W., and Nomura, M. 2023. "Does Human–AI Collaboration Lead to More Creative Art? Aesthetic Evaluation of Human-Made and AI-Generated Haiku Poetry," *Computers in Human Behavior* (139).
- Hyman, M., and Sierra, J. 2016. "Open-Versus Close-Ended Survey Questions," *Business Outlook* (14:2), pp. 1–5.
- Kim, S., Eun, J., Seering, J., and Lee, J. 2021. "Moderator Chatbot for Deliberative Discussion: Effects of Discussion Structure and Discussant Facilitation,"

- Proceedings of the ACM on Human- Computer Interaction* (5), pp. 1–26.
- Kopenhagen, N., Gaß, O., and Müller, B. 2012. “Design Science Research in Action - Anatomy of Success Critical Activities for Rigor and Relevance,” in *Proceedings of the 20th European Conference on Information Systems*.
- Kuckartz, U. 2018. *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*, 4th ed., Beltz.
- Landbot.io. 2022. “Conversation Design for Chatbots: The Ultimate Guide,” (<https://landbot.io/>, accessed 22.03. 2022)
- Leimeister, J. M. 2014. *Collaboration Engineering: IT-Gestützte Zusammenarbeitsprozesse Systematisch Entwickeln und Durchführen*. Springer Berlin Heidelberg.
- McKinsey & Company 2022. “Americans are embracing flexible work—and they want more of it” (<https://www.mckinsey.com/business-functions/people-and-organizational-performance/our-insights/americans-are-embracing-flexible-work-and-they-want-more-of-it>, accessed 15.10.2024)
- Mockenhaupt, A. 2021. *Digitalisierung und Künstliche Intelligenz in der Produktion*. Springer Wiesbaden.
- Morrison-Smith, S., and Ruiz, J. 2020. “Challenges and Barriers in Virtual Teams: A Literature Review,” *SN Applied Sciences* (2:6), 1096.
- Mukherjee, A., Santana, V. F. D., and Baria, A. 2023. “ImpactBot: Chatbot Leveraging Language Models to Automate Feedback and Promote Critical Thinking Around Impact Statements,” in *Association for Computing Machinery*.
- OpenAI. 2023. “GPT-4 Technical Report.” Available at: <https://openai.com/research/gpt-4>.
- Osterwalder, A., and Pigneur, Y. 2010. *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. Vol. 1, John Wiley & Sons.
- Pamungkas, E. W. 2019. *Emotionally-Aware Chatbots: A Survey*. Tech. rep., University of Turin.
- Peppers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. “A Design Science Research Methodology for Information Systems Research,” *Journal of Management Information Systems* (24:3), pp. 45–77.
- Przybilla, L., Baar, L., Wiesche, M., and Krcmar, H. 2019. “Machines as Teammates in Creative Teams: Digital Facilitation of the Dual Pathways to Creativity,” in *Proceedings of the 2019 on Computers and People Research Conference, SIGMIS-CPR '19*, Association for Computing Machinery, pp. 94–102
- Rasa Technologies Inc. 2021. “RASA: Open Source Conversational AI,” <https://rasa.com/>, Accessed on 14.06.2022.
- Recker, J., von Briel, F., Yoo, Y., Nagaraj, V., and McManus, M. 2023. “Orchestrating Human-Machine Designer Ensembles During Product Innovation,” *California Management Review* (65:3), pp. 27–47.
- Schöpl, J. C., and Steinhoff, P. F. J. 2023. “Recommendations for Leaders of Virtual Teams to Build Trust,” in *The Digital Transformation of Georgia*, M. Kupiek and

- R. A. Brandmeier (eds.), *Progress in IS*, Cham: Springer.
- Si, W. M., Backes, M., Blackburn, J., De Cristofaro, E., Stringhini, G., Zannettou, S., and Zhang, Y. 2022. “Why So Toxic? Measuring and Triggering Toxic Behavior in Open-Domain Chatbots,” in *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, pp. 2659–2673.
- Siemon, D. 2022. “Elaborating Team Roles for Artificial Intelligence-Based Teammates in Human-AI Collaboration,” *Group Decision and Negotiation* (31), pp. 871–912.
- Singh, A., Ramasubramanian, K., and Shivam, S. 2019. *Building an Enterprise Chatbot: Work with Protected Enterprise Data Using Open Source Frameworks*, Springer.
- Statista 2024. “Collaboration Software Market – Worldwide” (<https://www.statista.com/outlook/tmo/software/productivity-software/collaboration-software/worldwide>, accessed 15.10.2024)
- Strohmann, T., Fischer, S., Siemon, D., Brachten, F., Lattemann, C., Robra-Bissantz, S., and Stieglitz, S. 2018. “Virtual Moderation Assistance: Creating Design Guidelines for Virtual Assistants Supporting Creative Workshops,” in *PACIS 2018 Proceedings*, p. 80.
- Strohmann, T., Siemon, D., and Robra-Bissantz, S. 2017. “brAInstorm: Intelligent Assistance in Group Idea Generation,” in *Designing the Digital Transformation*, A. Maedche, J. vom Brocke, and A. Hevner (eds.), Springer International Publishing, pp. 457–461.
- Stucki, T., D’Onofrio, S., and Portmann, E. 2020. *Chatbots Gestalten Mit Praxisbeispielen der Schweizerischen Post*, Springer Fachmedien.
- Tavanapour, N., and Bittner, E. 2018. “Automated Facilitation for Idea Platforms: Design and Evaluation of a Chatbot Prototype,” in *Proceedings of the International Conference on Information Systems (ICIS)*.
- Tavanapour, N., Poser, M., and Bittner, E. 2019. “Supporting the Idea Generation Process in Citizen Participation—Toward an Interactive System with a Conversational Agent as Facilitator,” in *Twenty-Seventh European Conference on Information Systems (ECIS)*.
- Tegos, S., Demetriadis, S., and Karakostas, A. 2015. “Promoting Academically Productive Talk with Conversational Agent Interventions in Collaborative Learning Settings,” *Computers & Education* (87:1), pp. 309–325.
- Tranfield, D., Denyer, D., and Smart, P. 2003. “Towards a Methodology for Developing Evidence- Informed Management Knowledge by Means of Systematic Review,” *British Journal of Management* (14:3), pp. 207–222.
- Traumer, F., Oeste-Reiß, S., and Leimeister, J. M. 2017. “Towards a Future Reallocation of Work between Humans and Machines: Taxonomy of Tasks and Interaction Types in the Context of Machine Learning,” *SSRN Electronic Journal*.
- Voigt, M. & Bergener, K. 2013. “Enhancing Creativity in Groups - Proposition of an Integrated Framework for Designing Group Creativity Support Systems” in *Proceedings of the 46th Hawaii International Conference on System Sciences*, pp.

225–234

Wei, C., Yu, Z., and Fong, S. 2018. “How to Build a Chatbot: Chatbot Framework and Its Capabilities,” in *Proceedings of the 2018 10th International Conference on Machine Learning and Computing*, ICMLC '18, Association for Computing Machinery, pp. 369–373.

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., and He, L. 2022. “A Survey of Human-in-the-loop for Machine Learning,” in *Future Generation Computer Systems* (135), Elsevier, pp. 364–381.

15.7 Appendix 1: Task Allocation

Table 1. Example of the Task Allocation Procedure

Facilitation task	Subtasks	Complementary strengths	Responsibility (B=both):
Explain the use of technology (T)	Explain functions and scope	Apply rational action and rules	B: Basic information on usage can be provided by the AI, for more nuanced individual issues, HF should provide flexible support and respond empathically.
	Show empathy when technical issues arise	Uniform solving of routine tasks	
	Find solutions for technical issues	Empathy	
		Flexibility	

15.8 Appendix 2: Search terms and databases

Research area	Search Terms	Databases
Design Thinking and Empathy Map	Design thinking, creative work, empathy map, empathy map method, human centered design	- ABI/INFORM Collection - Springer Link - ACM digital library - IEEE Research Gate Overall database: Google search
Collaboration Engineering and Facilitation of Teamwork	Collaboration engineering, facilitation, facilitator, collaboration in teamwork, teamwork, moderation	
Human and Machine Intelligence	Artificial intelligence, human intelligence, intelligence, AI, machine learning, human computer collaboration, human machine teaming, machine as teammate, future of work, augmented intelligent, integrated intelligent, hybrid intelligent, human robot teamwork, human agent robot, teamwork, hybrid teams, human artificial intelligence teaming, hybrid intelligent systems, strengths artificial intelligence, strength robots, strengths computer, weaknesses artificial intelligence, weakness robots, weakness computes	
Chatbots	Chatbot, virtual assistants, conversational agents, architecture of chatbot, natural language processing	
Related work	Chatbot, virtual assistants, conversational agents + facilitation, moderation	
Recommendation for chatbot development	Chatbot, virtual assistants, conversational agents, advice chatbot development, chatbot development, chatbot prototyping, chatbot recommendation	
Recommendation for virtual facilitation work	Chatbot facilitation, virtual assistant facilitation, conversational agents facilitation, chatbot creative work, chatbot teamwork, virtual collaboration moderation, virtual assistants, creative workshops, automated facilitation.	

Figure 4. Search terms

15.9 Appendix 3: Design Requirements

Chatbot Facilitation (CF)	Human Facilitation (HF)	Technical Requirement (T)
CF1. Use an interaction concept with dialog trees (US, VFR)	HF1. Same as CF1	T1. NLP and rule-based behavior (US, CHR, VFR)
CF2. Common, superior solution (HI, TA, VFR)	HF2. Same as CF2.	T2. Closed domain (US, CFR)
CF3. Process support (TA, US, CHR, VFR)	HF3. Flexible question answering, problem detection and intervention (TA, CHR, VFR)	T3. Participation of 3 or 4 people (US)
CF4. User friendly features and encouragement of their use (CHR, VFR)	HF4. Introduction and farewell (TA, VFR)	T4. Use of instant messenger (VFR)
CF5. Documentation of arguments and results (TA, US, CHR, VFR)	HF5. Warm up round (TA, US)	T5. Avatar representation (CHR, VFR)
CF6. Introduction and farewell (TA, CHR, VFR)	HF6. Show respect, empathy, humor and neutrality (TA, VFR)	T6. Use of buttons, emojis, pictures and poll (TA, US, CHR, VFR)
CF7. Fallback responses (CHR)	HF7. Promote different perspectives (TA, VFR)	
CF8. Voting phase to reduce the number of arguments (TA, US, VFR)	HF8. Motivate participants (TA, US, VFR)	
CF9. Help with contributions (TA, CHR, VFR)	HF9. Support exchange of opinions and reaching of an agreement (TA, US, VFR)	
CF10. Responses to social intents (CHR, VFR)	HF10. Task, goals and rules reminder (TA, VFR)	
CF11. Use of simple and pictorial language (US, CHR, VFR)	HF11. Resolve conflicts (TA, VFR)	
CF12. Show respect, empathy, humor and neutrality (TA, CHR, VFR)	HF12. Small talk (TA, VFR)	
CF13. Time management (TA, US, CHR, VFR)		
CF14. Tracking progress (TA, US, VFR)		

TA = Task allocation; US = Underlying study; CHR = Chatbot requirements; VFR = Virtual facilitation requirements.

Figure 5. Design Requirements

15.10 Appendix 4: Questionnaire and Interview Guideline

Design Principle-1	Ø	ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10	ID11	ID12
DP1.1: Was the distribution of tasks between the chatbot and the moderator sufficiently clarified by the preparation document and the explanations in the chat about the task?	1.92	1	4	2	3	4	1	1	2	1	1	2	1
DP1.2: Did you find the division of tasks between the chatbot and the moderator useful?	1.92	1	3	1	4	3	1	2	3	2	1	1	1
DP1.3: In your opinion, were there parallel responsibilities between the chatbot and the moderator that inhibited the group work?	1.50	2	2	1	3	2	1	1	1	1	1	1	2
DP1.4: During the group work, were you consistently aware of who (chatbot or moderator) you should address when you had a question or instruction?	1.75	2	4	1	4	2	1	1	2	1	1	1	1
DP1.5: In your opinion, how could the tasks between the chatbot and the moderator have been better distributed?		X	X	X	X	X	X	X	X	X	X	X	X
DP1.6: Further feedback on design principle 1		X	X	X	X	X	X	X	X	X	X	X	X
Design Principle-2	Ø	ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10	ID11	ID12
DP2.1: Was the chatbot able to follow all questions and instructions given to it during the group work?	2.33	2	2	2	2	4	2	3	2	3	2	2	2
DP2.2: Was the moderator able to pick up and answer the questions and instructions that the chatbot could not understand? (If this never happened, please check "not available")	1.25	1	1	1	3	1	1	1	1	2	1	1	1
DP2.3: Further feedback on design principle 2		X	X	X	X	X	X	X	X	X	X	X	X
Design Principle-3	Ø	ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10	ID11	ID12
DP3.1: Did the chatbot and the moderator act as a team towards the working group?	1.75	1	2	2	4	3	1	1	1	2	1	2	1
DP3.2: Did the chatbot and the moderator treat each other in a respectful and friendly manner?	1.00	1	1	1	1	1	1	1	1	1	1	1	1
DP3.3: Did the chatbot and the moderator treat the working group in a respectful and friendly manner? (If there was a difference in perception between the two, please check "other" and complete in free text)	1.00	2	1	1	1	1	1	1	1	1	1	1	1
DP3.3: Free text field	X	X	X	X	X	X	X	X	X	X	X	X	X
DP3.4: Did you have the impression that you were working in an atmosphere of trust during the group work?	1.67	1	3	1	3	2	1	1	2	3	1	1	1
DP3.5: Did the chatbot and the moderator encourage intensive collaboration? (If there was a different perception of both, please check "other" and add in the free text)	1.75	1	2	2	3	4	1	1	2	1	1	2	1
DP3.6: Further feedback on design principal 3	X	X	X	X	X	X	X	X	X	X	X	X	X
Design Principle-4	Ø	ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10	ID11	ID12
DP4.1: Did the chatbot successfully guide the working group through the agenda?	1.42	1	2	1	2	1	1	1	3	2	1	1	1
DP4.2: Did the chatbot explain the necessary work steps and tasks in detail?	1.17	1	1	1	1	2	1	1	2	1	1	1	1
DP4.3: Did the chatbot successfully manage the votes?	1.50	1	1	1	1	3	1	1	3	1	1	1	2
DP4.4: Did you find it easy to document the results of your work using the chatbot?	1.58	1	3	1	2	2	1	3	2	1	1	1	1
DP4.5: In your opinion, did the chatbot support the time management of the working group in a meaningful way?	1.58	2	1	1	1	5	1	1	3	1	1	1	1
DP4.6: Did you have any questions or issues that the chatbot couldn't help you with	2.00	1	3	1	3	4	1	1	3	1	1	3	2
DP4.7: Which of its tasks should or could the chatbot do better in the future?	X	X	X	X	X	X	X	X	X	X	X	X	X
DP4.8: Is there a task missing that the chatbot should also perform in the future?	X	X	X	X	X	X	X	X	X	X	X	X	X
DP4.9: Further feedback on design principle 4	X	X	X	X	X	X	X	X	X	X	X	X	X
Design Principle-5	Ø	ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10	ID11	ID12
DP5.1: In your opinion, did all participants take part in the task?	1.17	1	1	1	2	2	1	1	1	1	1	1	1
DP5.2: Did the moderator actively promote the exchange of opinions and the decision-making process of the group?	2.00	1	1	3	2	5	1	1	2	4	2	1	1
DP5.3: Did you have any questions or topics that the moderator did not want you to discuss?	1.25	1	4	1	1	1	1	1	1	1	1	1	1
DP5.4: Did you have any questions or topics that neither the chatbot nor the moderator could help you with?	1.00	1	1	1	1	1	1	1	1	1	1	1	1
DP5.5: Did the facilitator try to resolve conflicts that arose between the participants in the group? (If there were no conflicts, click "not present")	Not Available	Not Available	Not Available	Not Available	Not Available	Not Available	Not Available	Not Available	Not Available	Not Available	Not Available	Not Available	Not Available
DP5.6: Further feedback on design principle 5	X	X	X	X	X	X	X	X	X	X	X	X	X
Design Principle-6	Ø	ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10	ID11	ID12
DP6.1: Did you have the impression that the moderator was actually a chatbot?	1.67	1	1	3	3	1	1	1	1	1	5	1	1
DP6.2: Did you have the impression that the chatbot was actually a human?	1.42	1	1	1	1	1	1	1	5	1	1	2	1
DP6.3: Did the display images of the chatbot and the moderator reinforce your impression that they are a chatbot or a human?	1.50	1	1	3	2	2	1	1	1	1	2	1	2
DP6.4: Further feedback on design principle 6	X	X	X	X	X	X	X	X	X	X	X	X	X
Closing Statement	Ø	ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10	ID11	ID12
DP7.1: In your opinion, has the group task been completed successfully? (If no, please explain in the following free text)	1.25	1	1	1	2	3	1	1	1	1	1	1	1
DP7.1: Free text field	X	X	X	X	X	X	X	X	X	X	X	X	X
DP7.2: Final feedback, if not already documented	X	X	X	X	X	X	X	X	X	X	X	X	X

Interview Guideline for the HF

Design principle 1

Q1: To what extent was the distribution of tasks between you and the chatbot sufficiently clarified by the preparatory documents and the explanations in the chat? You can choose between 1 "strongly agree" and 5 "strongly disagree".

Q1.1: Is there anything else you want to add?

Q2: To what extent did you find the distribution of tasks between you and the chatbot useful? You can choose between 1 "strongly agree" and 5 "strongly disagree".

Q2.2: Can you briefly explain again what bothered you in that sense? So, what wasn't quite so useful, since you chose "2 = agree"?

Q3: In your opinion, to what extent were there parallel responsibilities between you and the chatbot that hindered the group work?

Q4: In your opinion, were there parallel responsibilities between you and the chatbot that hindered your work as a facilitator? To what extent?

Q5: To what extent were you aware of your tasks during the group work?

Q5.1: Could you facilitate three chats at the same time?

Q6: In your opinion, how could the tasks between you and the chatbot have been better distributed?

Design principle 2

Q1: To what extent did it happen that the chatbot did not understand questions and instructions and you had to intervene?

Q2: To what extent was it problematic to take up and answer the questions and instructions that the chatbot could not understand? You can choose between 1 "strongly agree" and 5 "strongly disagree".

Design principle 3

Q1: To what extent did you feel like you were working in a team with the chatbot? You can choose between 1 "strongly agree" and 5 "strongly disagree".

Q1.1: Can you elaborate on your choice (2)?

Q2: To what extent did you have the impression that the working group perceived you and the chatbot as equal facilitators?

Q3: To what extent did you feel that the communication between you and the chatbot was respectful and friendly?

Design principle 4:

Q1: To what extent did you have to intervene in the tasks that were assigned to the chatbot and take over parts of its tasks? At what point exactly did you have to take over the task?

Q2: Which of its tasks should or could the chatbot do better in the future?

Q3: To what extent should the chatbot take over other facilitation tasks that you performed during the group work in the future, which could further relieve you in the future?

Q4: To what extent should the chatbot have performed further tasks that neither the chatbot nor you performed during the group work?

Q5: To what extent would you have preferred to take over tasks that were outsourced to the chatbot yourself?

Q6: To what extent did the use of the chatbot and its takeover of facilitation tasks during the group work reduce your overall workload compared to a group work that you would have moderated entirely on your own?

Q7: Do you have any further feedback on design principle 4?

Closing questions:

Q1: In your opinion, to what extent has the group task been completed successfully? You can choose between 1 "strongly agree" and 5 "strongly disagree".

Q2: What did you like in the collaboration with the chatbot?

Q3: What did you not like in the collaboration with the chatbot?

Q4: Final comments

Acknowledgements

This research was funded by the German Federal Ministry of Education and Research (BMBF), in the context of the project HyMeKI (01IS20057).

16 Human-AI Collaboration in conceptualizing Design Science Research Studies: Perceived Helpfulness of Generative Language Model's Suggestions

Memmert, L., Cvetkovic, I., & Bittner, E. (2024). Human-AI Collaboration in conceptualizing Design Science Research Studies: Perceived Helpfulness of Generative Language Model's Suggestions. In *Proceedings of the 31st European Conference on Information Systems (ECIS)*.

Abstract

Solving complex problems was named a new challenge for research on human-AI collaboration. In our study we focus on a particular means to solving complex problems: design science research (DSR). We investigate whether AI, more specifically, generative language models (GLM), can support an individual in conceptualizing DSR studies by making helpful suggestions. To do so we use extracts of a published DSR study and have GPT-3, a GLM provide suggestions for aspects of this study. These suggestions are then evaluated in a survey (n=33) regarding their helpfulness. Results show that GLM suggestions are perceived to be helpful, with some variation depending on expertise. Reported interest in using such a tool in the future was high. Describing how GLMs can offer helpful suggestions we contribute toward a DSR tool support ecosystem and, more generally, towards knowledge on how humans and (generative) AI systems can team up to solve complex problems.

Keywords: GPT-3, language model, design science research, human-AI collaboration

16.1 Introduction

With advances in artificial intelligence (AI) there is an increased research interest in teaming up of AI systems and humans (Seeber et al., 2020a). A new challenge is solving complex problems through human-AI collaboration (Dellermann et al., 2019; Akata et al., 2020). In our study, we envision humanAI collaboration for a particular approach to solving complex problems: design science research (DSR).

Despite DSR's importance a lack of tool support for conducting DSR studies was highlighted, accompanied by a call for research to establish a DSR tool support ecosystem (Morana et al., 2018b). Tools have been proposed for facilitating DSR researchers in different aspects such as collaboration within the team, research process documentation or result communication (e.g., vom Brocke et al., 2017). Some tools leverage intelligent features, e.g., adding knowledge elements via a conversational agent (Gau et al., 2022). In

making the vast DSR knowledge more accessible they might particularly benefit novices (Gau et al., 2022). However, these tools focus on supporting researchers on a meta-level.

We investigate how this support on meta-level can be complemented through support on a content-level. More specifically, we propose to imagine an AI-based tool which can be used by individual researchers when developing a concept for a DSR study. Throughout the DSR project, with the human incrementally building and adjusting the concept, the tool would make suggestions on different aspects of the concept to inspire the human to explore the problem and solution space (vom Brocke et al., 2020), and the in-between relationships (Venable, 2006) more comprehensively. Such suggestions would have to be treated with caution and would have to be rigorously grounded by the researcher before inclusion into the concept. Thus, instead of providing general guidelines, e.g., on how to formulate a design principle (i.e., meta-level support), the tool would suggest an actual, study-relevant design principle, based on the information the user has already entered (i.e., content-level support). The user could generate new suggestions to receive further inspiration. Thereby, the user and the AI would iteratively work towards the joined goal (DSR concept) by adding and adjusting design components, considering each other's contributions (i.e., collaboration, Bedwell et al., 2012). To aid researchers, the tool should provide suggestions for the specific study, inspiring them to enhance the concept. As DSR studies are concerned with novel problems or solutions to make a contribution (Gregor and Hevner, 2013), creating a pool of potential suggestions beforehand for all the studies different user might want to conduct seems difficult.

Instead of creating a pool of potential suggestions for the AI to select from, we use a generative language model (GLM). GLM have been shown to be able to generate context-specific text based on a given input (Brown et al., 2020), eliminating the problem of having to create a pool of potential suggestions beforehand. In this study we investigate how suggestions generated by 'GPT-3' (Brown et al., 2020), a state-of-the-art language model, are perceived, seeking to answer the following research question: RQ 1: To what extent are GLM suggestions perceived as helpful for conceptualizing DSR studies? Instead of creating a pool of potential suggestions for the AI to select from, we use a generative language model (GLM). GLM have been shown to be able to generate context-specific text based on a given input (Brown et al., 2020), eliminating the problem of having to create a pool of potential suggestions beforehand. In this study we investigate how suggestions generated by 'GPT-3' (Brown et al., 2020), a state-of-the-art language model, are perceived, seeking to answer the following research question: RQ 1: To what extent are GLM suggestions perceived as helpful for conceptualizing DSR studies?

Instead of creating a pool of potential suggestions for the AI to select from, we use a *generative language model* (GLM). GLM have been shown to be able to generate context-specific text based on a given input (Brown et al., 2020), eliminating the problem of having to create a pool of potential suggestions beforehand. In this study we investigate how suggestions generated by 'GPT-3' (Brown et al., 2020), a state-of-the-art language model,

are perceived, seeking to answer the following research question: *RQ 1: To what extent are GLM suggestions perceived as helpful for conceptualizing DSR studies?*

Similarly to Gau et al. (2022) we assume that particularly novices might benefit from DSR tool support. From complex problem solving literature (psychology) it is known that novices find it more difficult to explore the problem and solution space appropriately and to build viable internal representations due to their lack of prior knowledge (Fischer et al., 2012). They therefore might find AI suggestions to be more helpful. As perceived helpfulness affects the acceptance of AI and the potential for collaboration success, we investigate the existence of individual differences in the perception of AI suggestions' helpfulness based on level of expertise. We thereby answer the following research question: *RQ2: How does the level of expertise influence the perceived helpfulness?*

To answer these research questions we use the GLM 'GPT-3' (Brown et al., 2020) to create AI suggestions. We then survey humans with different levels of expertise on their perception of the helpfulness of those AI suggestions. The results of our analysis show that the overall perception of the AI suggestions is positive. Significant differences in the perceived helpfulness based on expertise are sparse. Opposite from what we expected, higher levels of expertise occur with higher perceived helpfulness ratings. With our paper we contribute to addressing the call for research on a DSR tool ecosystem (Morana et al., 2018b), demonstrating that the existing facilitative, meta-level support can be complemented by content-level support. With our innovative approach of generating helpful AI suggestions along a canvas-like structure, we show a potential pathway for humans and AI systems, more specifically GLMs, to collaboratively design solutions for complex problems, an important research challenge (Dellermann et al., 2019).

16.2 Background

16.2.1 Design science research tool support

Solving complex problems was named a new challenge for human-AI collaboration (Dellermann et al., 2019). An approach for solving complex problems is DSR. DSR refers to "a research paradigm in which a designer answers questions relevant to human problems via the creation of innovative artifacts, thereby contributing new knowledge to the body of scientific evidence" (Hevner et al., 2010). Such innovative artifacts might include, e.g., software prototypes. A lot of literature is available including overarching DSR approaches, types of theoretical contributions through DSR or guidelines on formulating meta requirements (e.g. Chandra et al., 2015; Gregor and Hevner, 2013; Hevner et al., 2004). To aid researchers in rigorously applying DSR, tool support was suggested (Morana et al., 2018b). Proposed solutions focus on facilitating researchers in conducting DSR studies including aspects like team member collaboration, improvement of documentation, or context-independent literature suggestions (e.g., Morana et al., 2018a) and even included conversational agents to support researchers in documenting studies (e.g., Gau et al., 2022).

In summary, the tools proposed facilitate researchers on a meta-level, e.g., by providing a template for a “good” way of formulating meta requirements, or suggesting established approaches for conducting DSR studies, sometimes even by leveraging AI.

Here, we propose a complementary approach to support researchers on a second layer: on content-level (see table 1). Instead of offering information of how to do something in general, e.g., on how to formulate a meta requirement, a tool should suggest actual contents like issues, design requirement, design principles, or kernel theories for the study at hand, enhancing the concept by inspiring the researcher. Our approach uses a GLM to offer tailored suggestions based on the specific study’s content.

Table 1. DSR tool support overview: meta-level vs. content-level support

Level	Explanation	Illustrative examples of support for the exemplary study: “Develop a design for a conversational agent to facilitate contributors in generating elaborate ideas on idea platforms”	Tool examples
Meta-level	Support independent of the specific study	- List DSR approaches, e.g., three cycle approach by Hevner (2007), DSR process model by Peffers et al. (2007) - List reference literature, e.g., for phrasing for design requirements as proposed by Chandra et al. (2015)	- MyDesignProcess.com (vom Brocke et al., 2017) - DScaffolding (Contell et al., 2017)
Content-level	Support for the specific study at hand	Inspirational suggestions for design requirements, e.g., - Conversational agents should elicit ideas by asking open-ended questions. - Conversational agents should encourage elaboration of ideas.	<Our tool>

16.2.2 Generative language models

GLM are large language models trained on a large corpus of text to generate free form text. They are trained with the goal to predict the next word given a certain input, i.e., given a certain context. Though they were trained task-agnostic only with this general goal, GLM have been shown to perform well on typical natural language processing tasks like translation or question answering and even on generating news articles based on a headline

(Brown et al., 2020). They do so without task-specific training, i.e., without a task specific training data set or task-specific updates of the model (Brown et al., 2020). The interaction with the model (Brown et al., 2020) is performed via free-form text-in, text-out. The input might, e.g., consist of an instruction and/or exemplary pairs of inputs and expected outputs (e.g., ‘Italy -> Rome’), followed by another input (‘France ->’), for which the model then generates an output (Paris) accordingly, potentially “recognizing” the presence of pairs of countries and their capitals from the corpus used for training. The exemplary input-output-pairs are referred to as demonstrations. They can be distinguished from training examples as demonstrations do not lead to model weight updates, i.e., are used as additional input during inference time, whereas training examples typically are used to fine-tune models for a specific task including model (weight) updates (Brown et al., 2020). Depending on the number of demonstrations provided during inference time one can distinguish zero-, one- or fewshot usage (0, 1 or multiple demonstrations respectively) (Brown et al., 2020). Performance differences between the approaches vary by task (Brown et al., 2020), but even without demonstrations good results can be achieved. As the AI system’s output is conditioned on the input provided during inference time (and inference parameters, e.g., randomness), the input needs to be crafted carefully. Techniques to shape the input have been developed as part of prompt engineering research (e.g., Mishra et al., 2020).

The utility of humans collaborating with generative AI systems was demonstrated in a variety of creative contexts (Memmert and Bittner, 2022) including joint music composition (Suh et al., 2021) or drawing (Zhang et al., 2021). Several GLM-based tools have been proposed or used to provide inspiration to humans when solving problems, e.g., an “Idea Machine”, supporting humans in idea generation through theory-guided suggestions by an GLM (Di Fede et al., 2022). Gero et al. (2022) show that GLMs can provide “sparks”, i.e., short sentences, described by users to offer “inspiration” and “external perspectives”. Even more so, Zhu and Luo (2022) report that “GPT can perform conceptual design tasks with a reasonable level of competence” when using it develop design suggestions for actual design challenges. Given that previous literature suggests that generative AI/GLMs can support creative activities through suggestions, we investigate whether such GLM suggestion can be helpful to solve problems tackled via DSR projects which are novel, complex, open-ended and knowledge rich (Hevner et al., 2004; Morana et al., 2018b) requiring creativity for designing solutions (Hevner et al., 2010).

Content-level support might have been difficult to achieve with more traditional approaches such as recommender systems, as creating a pool of all potential suggestions for all the possible studies the different users might want to conduct (from which an AI system could then choose and recommend suggestions to the human) does not seem feasible. Thus, instead of trying to create a repository of potential AI suggestions and training a recommender system, we instead suggest to use GPT-3 (Brown et al., 2020), a state-of-the-art GLM. The GLM is trained on a large, general corpus and is therefore flexible to generate suggestions regarding a variety of topics based on the human input. As

a result, neither for training nor for inference a predefined set of potential AI suggestions is required. Given the potential of GLM for offering helpful suggestions for creative problem solving combined with the flexibility due to the on-demand generation of suggestions (instead of prior preparation), we thus believe GLMs might be suited for our case of providing suggestions and thereby content-level support for DSR.

16.3 Methodology

Our underlying idea was to support researchers in developing new DSR concepts with the help of AI suggestions. The human would fill in some general context information and the general problem, as well as first ideas. The system would make suggestions based on those ideas, which the user would review and might or might not find helpful to further develop the concept. As our goal was to understand the helpfulness of the AI suggestions and not to evaluate an experimental tool design, we choose to evaluate these suggestions independently to remove the potential influence of the experimental design on the results. Thus, for our study, we extracted contents from an existing, published DSR paper. We then used these contents from the study as input for the GLM in order to retrieve completions, hereafter referred to as ‘suggestions’. In a survey we asked participants to rate the helpfulness of these suggestions according to a measurement instrument we reused from Rhys Cox et al. (2021).

16.3.1 Study approach

In our approach (figure 2) we used a foundational DSR structure (step 1) to develop prompt template strings (step 2), which we then populated with the contents of an existing study (steps 3 & 4) to generate a pool of AI suggestions (step 5) from which we select suggestions (step 6) for our survey (step 7).

Step 1. As a foundation, we used an (exemplary) canvas-like DSR structure (figure 2, #1). We derived this structure from the approach of using components like issues, theories, design requirements, design principles to describe the concept’s cornerstones (in different variations e.g., available in Meth et al., 2015 and Gnewuch et al., 2017). We have enriched this canvas by adding situation, general problem, and artifact class, information typically presented in the studies’ manuscript. Besides providing a more comprehensive view of the study to humans, this also allowed us to capture the studies broader context, enriching the information, which can be provided to the AI system.

Step 2. To receive context-dependent suggestions by the AI systems it needs to be provided with appropriate (context-dependent) input. As we sought to develop a re-producible and re-usable approach that would work across study contexts, we developed a fixed structure, i.e., prompt templates, which could then be populated with study-specific contents. The template strings contained structuring elements and placeholders for the specific contents of the study as well as a task description. As concepts like “design requirements” or “design

principles” carry different meanings depending on context, we included DSR specific explanations according to (Meth et al., 2015).

For designing the template strings, we used prompt design techniques to improve suggestion quality. In particular we used “decomposition reframing” and “itemization reframing” (Mishra et al., 2020). “Decomposition reframing” follows the idea of splitting a larger, more complex tasks into multiple sub-tasks. We applied this technique by identifying the individual design components for which we sought suggestions. Thus, we had suggestions for four smaller prompts (one for each design component) instead of having one large prompt to receive suggestions for the entire study concept. Additionally, we used “itemization reframing”, e.g., instead of having a longer paragraph with the different issues, we created an enumerated list in order to “signal” to the GLM that we expect additional issues as a completion. We included placeholders for design components that might have already been created by the user to act as demonstration examples (Brown et al., 2020). However, besides information for the situation and general problem such initial user contributions are optional. In total, we developed four prompt templates, one for each design component. As in DSR, the design components are related or derived from each other (Gnewuch et al., 2017), we included design components of previous stages into the prompt for the next type of design components to increase internal consistency, e.g., the prompt template for design requirements included placeholders for issues and kernel theories and the template for design principles included placeholders for design requirements and kernel theories (see figure 1#2).

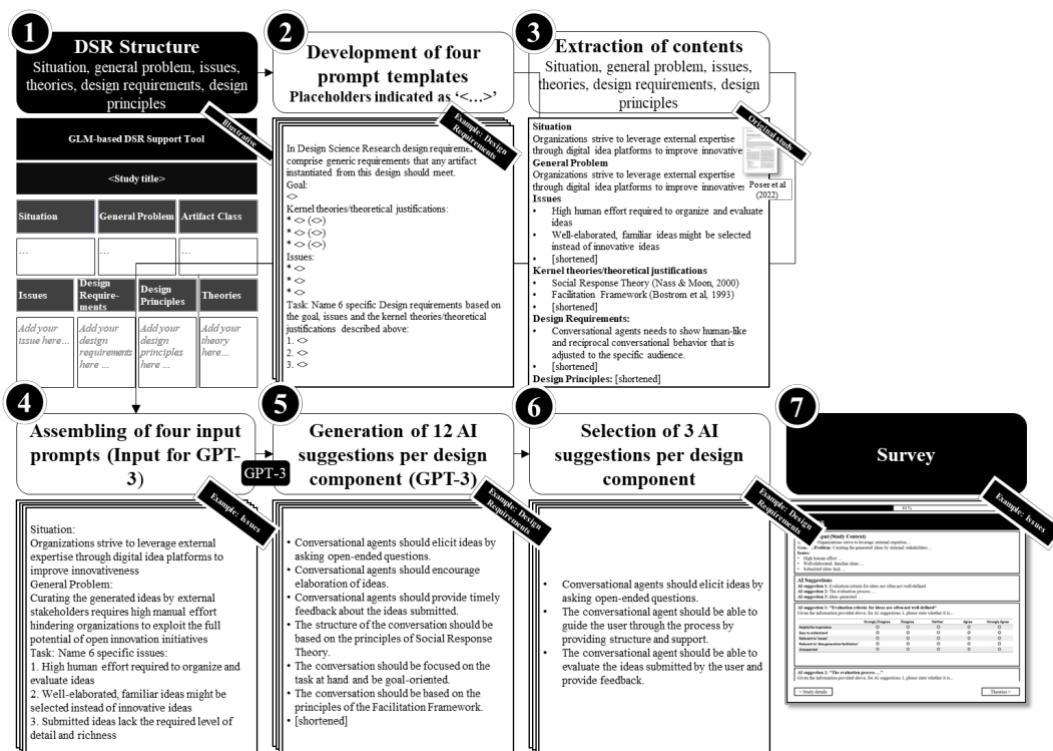


Figure 1. Study approach: from generic DSR “canvas” to context-specific survey contents with GPT-3 suggestions (steps 2-7 contain the actual data)

Step 3. We then populated our template strings with actual contents of a published DSR study. We therefore selected a published DSR article (“Design and Evaluation of a Conversational Agent for Facilitating Idea Generation in Organizational Innovation Processes”, Poser et al., 2022) in order to ensure the foundational study idea and the resulting AI suggestions were based on a paper idea that was both interesting and relevant. Our approach, however, is general, thus any other DSR study could have been selected (or can be selected by other authors). We selected a paper published after the training data collection for the corpus for the GLM we use was completed to ensure no training data leakage occurred, i.e., to avoid the possibility of the study being part of the training data corpus. This was necessary as Carlini et al. (2023) showed that GLM might reproduce text seen during training in verbatim.

From this study, we extracted the following key conceptual elements: situation, general problem, issues, kernel theories/theoretical justifications, design requirements, and design principles. Some elements like design requirements or design principles were clearly described as such in the original study and were thus quoted directly whereas other elements were paraphrased by the authors of this paper.

Step 4. In order to receive the study-specific suggestions we filled in the placeholders with the study contents that we had extracted from the published study.

Step 5. We then fed the resulting populated template strings as prompts into our GLM, GPT-3, to generate suggestions (see figure 1#5 for actual examples). We used the out-of-the-box GPT-3 model (i.e., without fine-tuning) at pre-defined settings (engine: text-davinci-002; temperature: 0.7).

We imagined a usage scenario as follows: while adding their ideas into the tool the user might invoke the generation of AI suggestions. The user then might check these suggestions and might update their concept, before proceeding to generate new suggestions, iteratively building their concept. For each regeneration the number of new suggestions was limited to three in order not to overwhelm the user. We assumed during typical usage a user might refresh (i.e., re-generate) the suggestions up to three times. Simulating this procedure, we generate 12 suggestions for each of the four design components (i.e., 12 suggestions per design component $\times$ 4 design component types = 48 suggestions).

Step 6. We were intrigued by the quality of some of the AI suggestions, while others seemed generic, and some occurred multiple times. Our goal was to understand if the AI system could provide helpful suggestions, not if all suggestions generated by the AI were helpful. Given our goal and the described usage scenario, we decided to select three of the twelve suggestions for each of the four design components, as even this reduced number of suggestions would allow us to understand, if suggestions were perceived to be helpful (and in which ways), but would significantly reduce the burden on participants, potentially increasing the response quality and completion rate.

Step 7. The methodology regarding the survey is described in the next subsection.

16.3.2 Survey

To answer our research questions regarding the helpfulness of the AI suggestion and the influence of expertise we conducted a survey gathering the relevant data and performed the corresponding analyses. Besides demographics, we asked participants to reflect their level of expertise with regard to the DSR paradigm and the topic of the study (independent variables). To determine the expertise with regard to the DSR paradigm we asked participants for how many projects or studies they had used DSR along five categories (never, 1 or 2 times, 3 to 5 times, 5 to 10 times, more than 10 times). The familiarity with regards to the topic of the foundational study (i.e., idea generation facilitation) was rated along a 5-point likert scale ('Not at all familiar' to 'Extremely familiar'). Besides these more direct measures of expertise, we also measured the "general expertise" via experience in academic research, i.e., role in academia, on a 5-category scale (Bachelor student to Professor).

After introducing the foundational study (we provided the abstract of the study and asked participants to read it carefully) we asked participants to score each AI suggestions (dependent variables). To do so we had four pages (one page for every design component: issues, theories, design requirements, and design principles) with a short overview of relevant information we had extracted from the study (e.g., for the issues page we presented a short description of the situation, the general problem, and 3 issues from the study for comparison) as well as the three AI suggestions (e.g., for the issues page we presented three issues). Participants were then asked to score each AI suggestion individually (see figure 1, #7).

For measuring the helpfulness of the AI suggestions according to RQ 1 we used the instrument (both measurement dimensions and scale) from Rhys Cox et al. (2021). Similar to our study, Rhys Cox et al. (2021) used this instrument to have participants rate cues/prompts offered during ideation. According to this instrument, ratings are gathered along five dimensions: 'helpful for inspiration', 'easy to understand', 'relevant to ', 'relevant to ', and 'unexpected'. For simplicity, we refer to those five dimensions collectively as 'helpfulness' in this study. In our case, the was to find an appropriate design component. The was the topic of the study, i.e., idea generation facilitation. For each dimension participants provide a rating on a 5-point likert scale (strongly disagree to strongly agree, coded as 1 to 5). In total participants thus provided 60 ratings with regard to the AI suggestions: 12 suggestions (4 design components with 3 suggestions each) x 5 evaluation dimensions. At the end of the survey, we asked participants if they were interested in using or testing such an AI-based DSR support tool and if they had suggestions for further design components or features the tool should support. In total, the survey contained 23 questions and required about 15 minutes to complete.

To recruit participants with different levels of experience we reached out to students and participants of DSR university courses as well as authors who published in DSR-related

tracks at information systems conferences. As we also asked addressees to forward our survey to students and colleagues the total number of recipients and consequently the response rate is unknown.

In total, we had 33 complete submissions. For the analysis we excluded partial submissions (37.1% completion rate) as most participants (75.1%) who did not complete the survey stopped before making any ratings and average ratings for participants who completed the survey were not significantly different from average ratings of participants who only partially completed it (Kruskal-Wallis: Statistic=2.292, $p=0.130$). Among the 33 complete submissions participants' (female=6, male=26, other=1; age: min=18, max=51, mean=30.5) levels of expertise were diverse (figure 2).

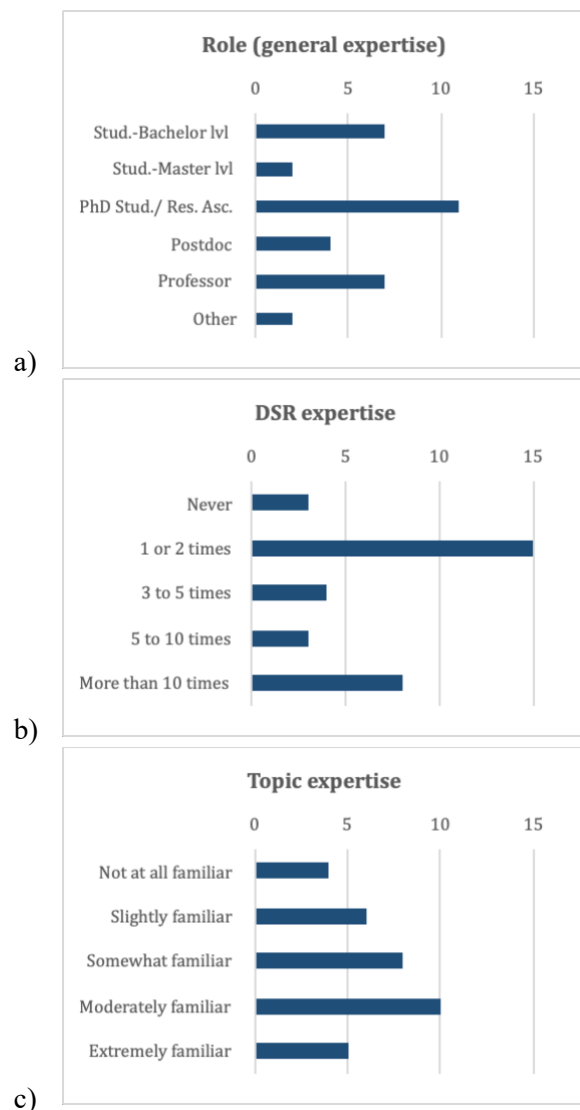


Figure 2. Expertise: a) role or general expertise, b) DSR expertise, c) topic expertise

In order to investigate the general perceived helpfulness of the AI suggestions (RQ1) we calculated the descriptive statistics of the ratings. We tested if they were significantly greater than the neutral rating and if there were any significant differences in ratings among design components, among dimensions, and among individual AI suggestions within their

groups. To understand the impact of expertise on the ratings (RQ2) we performed a correlation analysis between the types of expertise and the ratings.

16.4 Results

We first ask whether the AI suggestions are perceived as overall helpful (RQ1). As table 2 shows, the answer is yes. On average, AI suggestions are scored at 3.5, meaning between ‘neither’ (3.0, i.e., neutral) and ‘agree’ (4.0). A Wilcoxon signed-rank test (an alternative to one sample t test for data deviating from normality, Shapiro-Wilk $p < 0.001$) for the alternative hypothesis ‘greater than 3’ revealed that ratings across all design components and dimensions except from the dimension ‘unexpected’ were rated significantly greater than neutral (i.e., 3.0, table 3).

In order to understand differences among design components, dimensions, and individual suggestions, we performed the Kruskal-Wallis test (non-parametric alternative to ANOVA, for data that does not fulfill the assumption of homogeneity of variances, Levene’s test: $p < .001$) and corresponding Dunn’s post hoc comparisons. Regarding design components (rows in table 2), pairwise post-hoc Dunn test with Bonferroni adjustments showed that AI suggestions for issues, design requirements, and design principles were rated significantly better than for theories (0.1, 0.2, and 0.3 units respectively, $p < 0.005$).

With regard to the dimensions (columns in table 2), AI suggestions were rated significantly worse with regard to ‘unexpectedness’ as compared to the other dimensions (between 0.8 and 1.1 units; $p < 0.001$). With regard to the individual AI suggestions there were no significant differences in rating between AI suggestions within the same design component group ($p > .05$). On a higher level, Theories show to be the least easy to understand. DRs and DPs show higher ratings on the ‘Helpful for inspiration’ dimension than Issues and Theories, respectively. Another indication for the AI suggestions having been perceived to be helpful was reflected in the answers to a question, which we asked after participants had seen the kinds of AI suggestions the GLM generated for our exemplary study: “Would you be interested to use a such a Design Science Research support tool that provides these kinds of AI suggestions in the future?”. 81.8 % of participants responded with ‘Yes’, only 15.2 % answered ‘No’ (3.0% N/A). Looking at single design components, Issues 1 and 2 (“Evaluation criteria for ideas are often not well-defined”, “The evaluation process is often biased and subjective”) display the lowest ‘Unexpected’ rating. The ratings for Theory suggestion 3 (“Task-technology fit”) were similar across dimensions. DR2 “The conversational agent should be able to guide the user through the process by providing structure and support” displays highest rating for ‘Relevant to DRs’, but lowest on the dimension ‘Unexpected’. DP2 “Make the conversational agent's behavior adjustable to different types of users by taking into account the user's role, knowledge and expertise, to provide an adequate level of support, and by offering different types of interactions (e.g., more directive or less directive) to users with different levels of experience.” is both the

most helpful one as well as the most relevant to DPs. In addition, DP2 is the only design component that positively correlates with all three expertise types on several dimensions.

Table 2. Mean values and standard deviations for participant ratings for AI suggestions per design component (n=33)

Helpfulness dimensions\ AI suggestions for concept elements	Helpful for inspiration	Easy to understand	Relevant to <concept element>	Relevant to 'idea generation facilitation'	Unexpected	Overall
Issues	3.3	4.0	3.8	3.5	2.5	3.4
1	3.4	4.0	3.8	3.5	2.3	3.4
2	3.3	3.9	3.7	3.4	2.3	3.3
3	3.2	4.2	3.9	3.5	2.8	3.5
Theories	3.4	3.4	3.5	3.2	3.1	3.3
1	3.3	3.1	3.4	3.1	3.3	3.3
2	3.5	3.6	3.5	3.4	2.9	3.4
3	3.3	3.6	3.5	3.2	3.0	3.3
Design requirements	3.8	4.1	3.7	3.7	2.4	3.5
1	3.8	4.2	3.6	3.8	2.5	3.6
2	3.6	4.1	3.8	3.6	2.2	3.5
3	3.8	4.1	3.6	3.7	2.6	3.6
Design principles	3.9	3.7	3.9	3.7	2.7	3.6
1	3.7	3.5	3.8	3.8	2.8	3.5
2	4.1	3.8	4.2	3.7	2.5	3.7
3	3.9	3.8	3.7	3.6	2.7	3.5
Overall	3.6	3.8	3.7	3.5	2.7	3.5

Table 3. Results of Wilcoxon signed-rank test (all ratings, design components, dimensions)

Note. For the Wilcoxon sign-ranked test, the alternative hypothesis specifies that the median is greater than 3.

	W	df	P
All individual ratings	842636.000	1979	< .001
Design components			
Issues	56910.000	494	< .001
Theories	41846.500	494	< .001
Design requirements	59506.500	494	< .001
Design principles	53219.000	494	< .001
Dimensions			
Helpful for inspiration	41597.000	395	< .001
Easy to understand	47366.000	395	< .001
Relevant to <design component>	43339.000	395	< .001
Relevant to 'idea generation facilitation'	31046.000	395	< .001
Unexpected	10457.000	395	1.000

Influence of expertise. Regarding the influence of expertise on perceived helpfulness (RQ2) the results are heterogeneous. We analyzed correlations between the role (general expertise), topic expertise, and DSR expertise and different ratings for concepts elements averaged as well as separately for each design component. Assumption checks were performed before correlation analysis to determine the normality of the data and choose the according correlation coefficient. According to Shapiro's pairwise normality test, most of the variables did not have normal distribution. Thus, for a more comprehensive comparison, we chose Spearman's correlation coefficient which is more robust. While most of the correlations we calculated between expertise and design components did not yield statistically significant results, some of the aspects did. In total we found 19 significant ($p < 0.05$) correlations between expertise (three types) and design components groups/design components (7/12, respectively) across all five dimensions. Three design component groups correlate with role (Theories: $\rho = .367$ (Relevant to 'idea generation facilitation'); DRs: $\rho = .393$ (Helpfulness for inspiration), $\rho = .413$ (Relevant to 'idea generation facilitation'); DPs: $\rho = .404$ (Helpfulness for inspiration)). Two groups correlate with Topic Expertise (DRs: $\rho = .470$ (Relevant to 'idea generation facilitation'); DPs: $\rho = .500$ (Helpfulness for inspiration), $\rho = .387$ (Relevant to design principles)). All correlations were positive, i.e., higher ratings occurred with higher expertise. Most significant correlations concerned the dimensions 'Helpfulness for inspiration' (8) and 'Relevant to idea generation facilitation' (7). We therefore performed a correlation analysis on these two dimensions and found the correlation to be positive, with strongest correlation across DRs ($\rho = 0.7$, $p < .001$).

Additional design components. In our survey participants rated AI suggestions on four design components (issues, theories, design requirements, design principles) and gained an impression of the capabilities of state-of-the-art GLM suggestions. In an open question we asked participants, which other design components they thought should be supported by such an AI tool. While one participant believed that the "core elements" were covered, other suggested additional design components like design features (as in concrete instantiations of the design principles), semantic connections or mappings between the design components to achieve a "mapping diagram", artifact type (e.g., instantiation, method, framework), types of data collection and validation, highlighting of context dependencies, and relevant literature. Most often, participants mentioned the need for evaluation-related aspects.

Additional potential features. We also asked participants which aspects in the design of a tool providing such AI suggestions they thought to be important. Participants suggested that the tool could support the formulation of design components, e.g., a transfer of 'raw' user-written requirements into design requirements of a 'correct' format, e.g., according to Gregor et al. (2020). They also suggested clustering of design knowledge to increase structure, support in the evaluation of design knowledge as well as support in selecting an appropriate prototyping framework.

Usage process. With regard to the usage process, participants had different ideas like integration into existing tools (e.g., citation manager, word processor). For integration into word processors context-dependent inline suggestions (“like ‘grammarly’”) were mentioned. To improve the usefulness, more detailed explanation of AI suggestions as well as a summary with all suggestions were proposed. Lastly, a participant stressed the need to carefully design the usage process to mitigate potential negative effects.

16.5 Discussion

Overall, suggestions generated with our approach were rated positively regarding helpfulness both by less and more experienced participants, encouraging us to further explore this path of human-AI collaboration.

16.5.1 Perception of AI suggestions

Ratings for all design components were higher than neutral for all design components. However, ratings for theories were significantly lower than for the other design components. One interpretation is, that GLMs might be less suitable to propose such abstract ideas like theories. However, such differences need to be interpreted with caution as another explanation could be that theories are in general more difficult to interpret than, e.g., issues, and thus would receive lower ratings, independently of originating from human or AI. Future research should therefore investigate the differences also with regard to the suggestion’s origin. While helpfulness ratings were positive there is still room for improvement. In our study, we used different prompt engineering techniques to increase the AI suggestion quality. However, we did not provide additional demonstration examples from other studies as part of the input prompt during inference time, neither did we fine-tune the model. Brown et al., (2020) have shown that providing additional examples can improve the output quality significantly.

The lack of significant differences in ratings between AI suggestions within their groups might be caused by the preselection step we performed. If we had included all AI suggestions into the survey (or only generated 3 suggestions) this might have been different. However, as discussed in the methodology section, a selection of (3) relevant suggestions from the larger pool of suggestions (12) seemed to be more representative for the usage scenario, allowing us to answer the research questions while lowering the burden on survey participants by reducing the number of required ratings.

Expertise. Results regarding correlation between level of expertise and perceived helpfulness were mixed. Other than we expected novices did not perceive AI suggestions to be more helpful. On the contrary, in cases with significant correlation (particularly ‘relevance to <topic>’ and ‘helpfulness for inspiration’) these correlations indicated higher ratings with higher levels of expertise. It is likely that senior researchers or those who are more familiar with the topic might have a greater appreciation for the AI suggestions.

Potentially they find the suggestions easier to mentally integrate and relate them with their existing knowledge. This might offset other effects known from CPS literature (Fischer et al., 2021). However, these assumptions need to be verified in a future study, potentially with a more fine-grained and objective assessment of the participants' expertise and a more in-depth investigation of the reasoning for the ratings.

Unexpectedness. The 'unexpected'-dimension received significantly lower ratings as compared to the remaining dimensions and was the only dimension with an average score (2.7) below neutrality (3.0). One participant reflected this aspect stating: "at no time was I really surprised by the suggestions and thought they are all things that would probably have come up in any discussions we would have as researchers/ practitioners". While collaboration among multiple humans is common in DSR projects (Sein et al., 2011), other humans might not always be cost-effectively available to have discussions. Thus, an "AI teammate" turning individual work of one human into collaborative work of a hybrid, sociotechnical ensemble, e.g., as described by Dellermann et al., (2019) might add value. Comparatively low user ratings with regards to unexpectedness are not surprising as "generative language models are trained to match the distribution of content generated by humans" (Brown et al., 2020). However, as we do not seek to replace the human by an AI, but have the AI collaborate with the human, e.g., inspiring a more comprehensive problem and solution space exploration, it might be desirable to increase unexpectedness of suggestions (see design fixation literature, e.g., Sio et al., (2015)) while retaining high relevance to task and domain, i.e., design component and topic, respectively. Approaches to be explored might include providing additional examples (sourced from high-quality DSR publications) with a more appropriate level of unexpectedness (Zhu & Luo, 2022) or adjusting model parameters like temperature (i.e., level of randomness). Additionally, while with GPT-3 we used a high-performing state-of-the-art (Brown et al., 2020), with rapid GLM advancements, our results need to be interpreted carefully and might be considered to show a lower boundary of what will be possible.

16.5.2 Usage of AI suggestion in a DSR support tool

In line with the overall positive ratings for the AI suggestions was the large percentage (81.8%) of participants being interested in using a tool that offers such GLM-based suggestions, with more than half of them (61.1%) being interested in testing such a prototype.

Caution when using suggestions. Though AI suggestions were rated positively they should be used with caution as part of a DSR support tool, in particular given the characteristics of GLM and the required rigor and suggested ethical behavior in (design science) research (Myers and Venable, 2014; Benke et al., 2020; Durani et al., 2021). Potential issues include lack of truthfulness of AI suggestions, lack of evidence for AI suggestions, a danger of plagiarism, a lack of understanding for the meaning, and a lack of novelty. As Lin et al. (2022) have demonstrated, GLM completions are not necessarily

truthful, but might reproduce falsehoods and biases. Before including suggestion into the study concept it therefore needs to be carefully assessed to “ensure the quality of the artefact” (Myers and Venable, 2014, p. 806), which should include a reflection on ethical principles (Benke et al., 2020). Another issue is the lack of evidence or support. DSR design components need to be thoroughly grounded in theory or practice, e.g., based on empirical findings (Goldkuhl, 2004). The AI suggestions provided by the tool, however, are not backed by empirical data or literature and therefore cannot be used directly. They might rather be treated as means for building “hypotheses” that need to be verified. A third issue is the potential danger of plagiarisms. It was shown that GLM might replicate entire passages of text from the underlying training data in verbatim based on certain prompts (Carlini et al., 2023). This, however, is usually not transparent from the output. Using the output as part of a study without proper attribution might constitute plagiarism. A fourth potential issue might be the lack of understanding for the meaning of the suggestions, i.e., the GLM cannot reason about developing the concept like a human but produces texts according to a probability distribution. It cannot make moral judgements (Floridi and Chiriatti, 2020), but these are required to be performed by the human (Durani et al., 2021). Lastly, the lack of unexpectedness we found in our study might appear to be contrary to the goal of DSR of developing novel artefacts. However, as Gregor and Hevner (2013, p. 344) point out, “nothing is really ‘new’”, everything “builds on previous ideas”. Together with the previous issue (lack of understanding), this opens up an interesting research challenge: given the types of theorizing and reasoning in DSR (e.g., Kuechler and Vaishnavi, 2012; Venable, 2006) and the ability of GLMs to mimic “common sense reasoning” (Brown et al., 2020), what types of contributions (e.g., Gregor and Hevner, 2013) can we expect from GLMs?

Sociotechnical design. Given these challenges, a careful, sociotechnical design will be required, to reduce the risk of users only insufficiently investigating a problem and the solution during a DSR study, accepting even unfit AI suggestion. One participant stressed this aspect as follows: “I think you need to carefully design the whole socio-technical usage process to not design a tool that makes inexperienced DSR users to accept useless suggestions.” The lack of critical engagement with AI suggestions resulting in an overreliance of humans on AI systems is a known problem from the related field of AI-assisted decision making (e.g., Buçinca et al., 2021). Narrow solution space exploration is also a known problem from the area of design studies referred to as design fixation, the phenomenon that humans can be influenced by design (aspects) they have encountered before and might as a result be less innovative (Jansson and Smith, 1991; Sio et al., 2015). In developing socio-technical solutions it might therefore be worthwhile to leverage approaches known from AI-assisted decision making and design studies literature to increase user’s engagement with the AI suggestions and prevent narrow problem and solution space exploration, e.g., by having users actively request information (Buçinca et al., 2021) or by providing less common examples (Sio et al., 2015). Relevance to real-world

practice might be considered important for DSR (Gregor and Hevner, 2013). The sociotechnical design should include the tool's usage processes as part of a DSR project. Future research should explore how such a tool can support establishing relevance, e.g., by supporting empirical data gathering (preparation) or by summarizing already available observations from practice.

Beyond a sociotechnical design, however, it will be important to educate humans to critically reflect on AI suggestions. For our case, users would need to be educated about the limitations of such tools and that all suggestions need to be treated as unsubstantiated hypotheses, not as facts, as well as on ethical approaches towards DSR (e.g., Durani et al., 2021). More broadly, questions around ownership, authorship, and attribution of AI outputs will need to be clarified (Stokel-Walker, 2023).

16.5.3 Towards a DSR support tool with embedded, generative AI capabilities

Generative AI has been shown to support humans in creative, open tasks (e.g., Jiang et al., 2022; Suh et al., 2021; Zhang et al., 2021). In this study, we show how GLMs can be used to support a human in solving an open, complex problem, exemplified along concept development for designing an artifact according to the DSR paradigm. Our study approach allowed us to evaluate the helpfulness of such GLMs independently from a specific tool implementation. However, future research should explore implementing a prototype (figure 3), potentially using our approach as a foundation. Instead of extracting contents from a published study, the system would use the inputs of the human from the user interface and plug these into the template strings we have developed (step 2) to generate AI suggestions, which are then displayed to the human. With such a prototype, future studies should explore the validity across topics (select a different foundational study at step 3) or the influence of different prompt engineering techniques (adjust templates in step 2) to better understand the potentials and limitations of such an approach.

GLM-based DSR Support Tool			
Design and Evaluation of a Conversational Agent for Facilitating Idea Generation in Organizational Innovation Processes			
Situation	General Problem	Artifact Class	
Organizations strive to leverage external expertise through digital idea platforms to improve innovativeness.	Curating the generated ideas by external stakeholders requires high manual effort hindering organizations to exploit the full potential of open innovation initiatives.	Conversational Agent	
Issues	Design Requirements	Design Principles	Theories
- High human effort required to organize and evaluate ideas - Well-elaborated, familiar ideas might be selected instead of innovative ideas - Submitted ideas lack the required level of detail and richness	- Conversational agents need to show human-like and reciprocal conversational behavior that is adjusted to the specific audience. - Conversational agents should make approving and motivation statements during the process. - Conversational agents should show politeness and act friendly.	- Provide a conversational agent with the capacity to initiate, construct and lead a conversation in a productivity-oriented and pleasant manner by actively posing questions to reach the facilitation goal and considering strategies to prevent deviations to other topics. - Provide the conversational agent with the capacity to actively react to and animate the user during the facilitation process to reach elaborated outcomes by offering support on demand with explanations about the process steps or answering topic-related vocabulary questions in a goal-oriented way. - Provide the conversational agent with socio-emotional cues for a personalized interaction with humans by demonstrating appropriate, friendly, polite and human-like characteristics in the communication by utilizing informal language and delivering motivating statements to reach higher acceptance in user's perception as facilitator.	- Social Response Theory (Nass & Moon, 2000) - Facilitation Framework (Bostrom et al., 1993) - Social Presence Theory (Duft & Lengel, 1986; Gefen & Straub, 1997; Short et al., 1976)
AI Suggestions: Issues 1 Evaluation criteria for ideas are often not well-defined. Idea selection process is often not transparent. Selected ideas might not be feasible or not in line with the organization's strategy.	AI Suggestions: Design Requirements 2 Conversational agents should elicit ideas by asking open-ended questions. The conversational agent should be able to guide the user through the process by providing structure and support. The conversational agent should be able to evaluate the ideas submitted by the	AI Suggestions: Design Principles 3 Provide the conversational agent with the ability to detect user's needs for additional explanations or support during the facilitation process by monitoring the user's behavior and conversation state and reacting accordingly.	AI Suggestions: Theories 4 Social Information Processing Theory (SIP) (Walther, 1996) Computer-Supported Cooperative Work (CSCW) (Buchanan & Whitaker, 1991) Task-Technology Fit Theory (Goodhue & Thompson, 1995)

Figure 3. Prototype idea for a GLM-based DSR support tool based on our study approach: 1) user input area, 2) AI suggestion area, 3) refresh-buttons to generate new suggestions for each design component (illustrative contents for demonstration cited or paraphrased from Poser, Küstermann, et al., 2022)

With this template-based approach, we simplify the interaction of the human with the GLM, as the inputs for the GLM (i.e., prompts) are automatically assembled in the background and the users can focus on concept development. Through this approach, generative AI is embedded into the tool and users do not have to learn to directly interact with this new technology (i.e., GLM) before working with the tool. This approach is in alignment with the challenges highlighted by Jiang et al. (2022). In their study, participants interacting with a GLM “quickly noticed that different paraphrases of the same prompt could have vastly different levels of effectiveness towards producing the desired effect”. Jiang et al. (2022) state that “[...] beginners were particularly surprised when their initial zero-shot prompts did not behave as intended” and often relied on “acquiring tips and tricks”. With our approach we abstract away such technology-specific aspects, allowing even non-tech-savvy users to leverage GLM capabilities while focusing on designing their artifact. Future research should explore how such human-AI collaboration affects work performance (Rafner et al., 2021), i.e., DSR concept (component) quality. More broadly, the question arises how such generative AI embedded into a tool compares to (conversational) free-text interaction with GLMs (e.g., ChatGPT). We argue that for our case, embedding generative AI could allow users to focus on concept development, without having to learn effective prompting strategies, e.g., as part of the pre-defined prompts in the tool established practices for phrasing design requirements or design principles (e.g., Chandra et al., 2015) that might be unknown to (novice) users can be included.

In this study, we focus on designing DSR solutions along a DSR canvas. However, we suggest future research to explore if other open, complex problems can be supported in a similar fashion. In alignment with our approach an established structure (figure 2, #1) like a canvas (e.g., business model canvas) could be a potential starting point for prompt template development which includes context-specific definitions of terms (#2). Such research could explore the transferability of the approach onto other complex problem solving mechanisms.

In summary, while initial AI suggestions quality is encouraging to continue with our GLM-based DSR tool support idea, more research is required to improve the suggestion quality and to carefully design of the usage process, given the characteristics of GLM and the known effects AI can have on humans’ ways of working. The proposals of survey participants regarding design components, features and the usage process might be a good starting point. More broadly, future research should investigate the applicability of our approach to other fixed structures such as established canvasses (e.g., business model canvas) to develop GLM-based tool support in a structured way.

16.6 Limitations

There are some limitations to our study. Only one study (of Poser et al., 2022) was included as a foundation for our survey, limiting representativeness. We decided against including

more studies as the survey was already extensive. Including more studies would require participants to familiarize themselves with more, new research context(s) (incl. reading additional abstracts), increasing the burden to complete the survey. Indeed, we see that most participants who did not complete the study stopped at the first page, which included the study details, strengthening us in the trade-off we made between completion (rate) and representativeness. Even with only one study included our sample size is small.

While we used tested prompt engineering techniques, we did not provide demonstration examples from other studies. A different presentation of the study contents towards the GLM or the utilization of other prompt engineering techniques might lead to different AI suggestions, potentially affecting perceived helpfulness. Additionally, including demonstration examples can improve suggestions quality for some tasks (Brown et al., 2020). In our study, we established a first baseline. Future studies, however, should explore appropriate configurations (prompt design, inference parameters) for the task. This might include conditioning the GLM to provide more unexpected or directive suggestions. In the study at hand, the GLM reproduced the structure that was entered, i.e., if the design requirement was phrased a certain way, the GLM would try to match this structure. With more elaborate prompt design, good practices for formulating design requirements (e.g., Chandra et al., 2015; Gregor et al., 2020) could be used.

16.7 Conclusion and outlook

Improving our understanding of how humans and AI can team up to solve complex problems is a new research challenge (Dellermann et al., 2019). For DSR, a particular approach to solving complex problems, we demonstrated the potential of support through AI, more specifically, through GLM. With our approach, we show that by making relevant study concepts available to a GLM in a structured way, helpful suggestions can be generated. In this way, facilitative DSR tool support can be complemented by an additional layer on content-level. While human collaboration remains essential for DSR projects, our results show that GLMs might be capable of contributing during periods of individual work.

Though the general perception of the AI suggestions was positive much more research is required. In order for the tool to be helpful, AI suggestions need to be of high quality. We therefore suggest to more qualitatively explore the reasons for (dis-)satisfaction with the suggestions and the underlying individual difference. To do so, the approach described in section 4.1 should be applied to other DSR studies, potentially extended by (some) suggestions for additional design components (section 5). Future research should explore the application to other, canvas-based problem solving tasks. Additionally, effects of prompt engineering (incl. demonstration examples, Zhu and Luo, 2022), parameter tuning, and an interplay of a hybrid approach using a generative and discriminative AI might be explored.

But even then, as pointed out by a participant, the sociotechnical usage process needs to be designed carefully. Challenges like overreliance or fatigue are known from other human-AI contexts (Buçinca et al., 2021; Seeber et al., 2020b). The potential negative impact of external cues on creativity is known from design (Jansson and Smith, 1991). Applicability of this existing design knowledge (e.g., interaction processes, interface design) for human-AI teams solving complex problems should be investigated.

To fully understand the effects of such tools it will be vital to have actual users develop actual concepts with such a tool. Do researchers adjust their concept due to GLM-based suggestions? How will the resulting DSR concepts compare to those without GLM-based tool support? The high level of interest among those who completed our survey, both to offer providing additional ideas and for using or testing a GLM-based DSR support tool, motivates us to continue exploring this path of human-AI collaboration.

Acknowledgement: This research was funded by the German Federal Ministry of Education and Research (BMBF) in the context of the project HyMeKI (reference number: 01IS20057).

16.8 References

- Akata, Z., D. Balliet, M. de Rijke, F. Dignum, V. Dignum, G. Eiben, A. Fokkens, D. Grossi, K. Hindriks, H. Hoos, H. Hung, C. Jonker, C. Monz, M. Neerinx, F. Oliehoek, H. Prakken, S. Schlobach, L. van der Gaag, F. van Harmelen, H. van Hoof, B. van Riemsdijk, A. van Wynsberghe, R. Verbrugge, B. Verheij, P. Vossen and M. Welling (2020). “A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence” *Computer* 53 (8), 18–28.
- Bedwell, W. L., J. L. Wildman, D. DiazGranados, M. Salazar, W. S. Kramer and E. Salas (2012). “Collaboration at work: An integrative multilevel conceptualization” *Human Resource Management Review* 22 (2), 128–145.
- Benke, I., J. Feine, J. R. Venable and A. Maedche (2020). “On Implementing Ethical Principles in Design Science Research” *AIS Transactions on Human-Computer Interaction* 12 (4), 206–227.
- Brown, T. B., B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever and D. Amodei (2020). “Language Models are Few-Shot Learners”. In *NIPS’20: Proceedings of the 34th 2020*, pp. 1877–1901.
- Buçinca, Z., M. B. Malaya and K. Z. Gajos (2021). “To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-

- making” *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW1), 1–21.
- Carlini, N., D. Ippolito, M. Jagielski, K. Lee, F. Tramer and C. Zhang (2023). “Quantifying Memorization Across Neural Language Models”. In *The International Conference on Learning Representations*.
- Chandra, L., S. Seidel and S. Gregor (2015). “Prescriptive Knowledge in IS Research: Conceptualizing Design Principles in Terms of Materiality, Action, and Boundary Conditions”. In: *2015 48th Hawaii International Conference on System Sciences*: IEEE, pp. 4039–4048.
- Contell, J. P., O. Díaz and J. R. Venable (2017). “DScaffolding: A Tool to Support Learning and Conducting Design Science Research”. In A. Maedche, J. vom Brocke and A. Hevner (eds.) *Designing the Digital Transformation*, pp. 441–446. Cham: Springer International Publishing.
- Dellermann, D., P. Ebel, M. Söllner and J. M. Leimeister (2019). “Hybrid Intelligence” *Business & Information Systems Engineering* 61 (5), 637–643.
- Di Fede, G., D. Rocchesso, S. P. Dow and S. Andolina (2022). “The Idea Machine: LLM-based Expansion, Rewriting, Combination, and Suggestion of Ideas”. In: *Creativity and Cognition*. New York, NY, USA: ACM, pp. 623–627.
- Durani, K., A. Eckhardt and T. and Kollmer (2021). “Towards Ethical Design Science Research”. In *ICIS 2021 Proceedings*.
- Fischer, A., S. Greiff and J. Funke (2012). “The Process of Solving Complex Problems” *The Journal of Problem Solving* 4 (1).
- Floridi, L. and M. Chiriatti (2020). “GPT-3: Its Nature, Scope, Limits, and Consequences” *Minds and Machines* 30 (4), 681–694.
- Gau, M., A. Maedche and J. vom Brocke (2022). “DSR Buddy: A Conversational Agent Supporting Design Science Research Activities” *Wirtschaftsinformatik 2022 Proceedings*.
- Gero, K. I., V. Liu and L. Chilton (2022). “Sparks: Inspiration for Science Writing using Language Models”. In: *Designing Interactive Systems Conference*. Ed. by F. ` Mueller, S. Greuter, R. A. Khot, P. Sweetser, M. Obrist. New York, NY, USA: ACM, pp. 1002–1019.
- Gnewuch, U., S. Morana and A. Maedche (2017). “Towards Designing Cooperative and Social Conversational Agents for Customer Service”. In: *Proceedings of the 38th International Conference on Information Systems (ICIS)*.
- Goldkuhl, G. (2004). “Design Theories in Information Systems - A Need for Multi-Grounding” *JITTA : Journal of Information Technology Theory and Application* 6 (2), 59–72.
- Gregor, S. and A. R. Hevner (2013). “Positioning and Presenting Design Science Research for Maximum Impact” *MIS Quarterly* 37 (2), 337–355.

- Gregor, S., L. Kruse and S. Seidel (2020). "Research Perspectives: The Anatomy of a Design Principle" *Journal of the Association for Information Systems* 21, 1622–1652.
- Hevner, March, Park and Ram (2004). "Design Science in Information Systems Research" *MIS Quarterly* 28 (1), 75.
- Hevner, A. (2007). "A Three Cycle View of Design Science Research" *Scandinavian Journal of Information Systems* 19 (2).
- Hevner, A., S. Chatterjee, P. Gray and C. Y. Baldwin (2010). *Design research in information systems. Theory and practice*. New York, NY, Dordrecht, Heidelberg, London: Springer.
- Jansson, D. G. and S. M. Smith (1991). "Design fixation" *Design Studies* 12 (1), 3–11.
- Jiang, E., K. Olson, E. Toh, A. Molina, A. Donsbach, M. Terry and C. J. Cai (2022). "PromptMaker: Prompt-Based Prototyping with Large Language Models". In: *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery.
- Kuechler, W. and V. Vaishnavi (2012). "A Framework for Theory Development in Design Science Research: Multiple Perspectives" *Journal of the Association for Information Systems* 13 (6), 395–423.
- Lin, S., J. Hilton and O. Evans (2022). "TruthfulQA: Measuring How Models Mimic Human Falsehoods" *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*.
- Memmert, L. and E. Bittner (2022). "Complex Problem Solving through Human-AI Collaboration: Literature Review on Research Contexts". In: *Proceedings of the 55th Hawaii International Conference on System Sciences*.
- Meth, H., B. Mueller and A. Maedche (2015). "Designing a Requirement Mining System" *Journal of the Association for Information Systems* 16 (9), 799–837.
- Mishra, S., D. Khashabi, C. Baral, Y. Choi and H. Hajishirzi (2020). "Reframing Instructional Prompts to GPTk's Language". In *Findings of the Association for Computational Linguistics: ACL 2022*.
- Morana, S., M. Scheid, M. Gau, I. Benke and J. vom Brocke (2018a). "Research Prototype: The Design Canvas in MyDesignProcess.com" *DESRIST 2018 Conference Proceedings*.
- Morana, S., J. vom Brocke, A. Maedche, S. Seidel, M. T. P. Adam, U. Bub, P. Fettke, M. Gau, A. Herwix, M. T. Mullarkey, H. D. Nguyen, J. Sjöström, P. Toreini, L. Wessel and R. Winter (2018b). "Tool Support for Design Science Research—Towards a Software Ecosystem: A Report from a DESRIST 2017 Workshop" *Communications of the Association for Information Systems*, 237–256.
- Myers, M. D. and J. R. Venable (2014). "A set of ethical principles for design science research in information systems" *Information & Management* 51 (6), 801–809.

- Peffer, K., T. Tuunanen, M. A. Rothenberger and S. Chatterjee (2007). "A Design Science Research Methodology for Information Systems Research" *Journal of Management Information Systems* 24 (3), 45–77.
- Poser, M., G. C. Küstermann, N. Tavanapour and E. A. C. Bittner (2022). "Design and Evaluation of a Conversational Agent for Facilitating Idea Generation in Organizational Innovation Processes" *Information Systems Frontiers*.
- Rafner, J., D. Dellermann, A. Hjorth, D. Verasztó, C. Kampf, W. Mackay and J. Sherson (2021). "Deskilling, Upskilling, and Reskilling: a Case for Hybrid Intelligence" *Morals & Machines* 1 (2), 24–39.
- Rhys Cox, S., Y. Wang, A. Abdul, C. von der Weth and B. Y. Lim (2021). "Directed Diversity: Leveraging Language Embedding Distances for Collective Creativity in Crowd Ideation". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, pp. 1–35.
- Seeber, I., E. Bittner, R. O. Briggs, T. de Vreede, G.-J. de Vreede, A. Elkins, R. Maier, A. B. Merz, S. Oeste-Reiß, N. Randrup, G. Schwabe and M. Söllner (2020a). "Machines as teammates: A research agenda on AI in team collaboration" *Information & Management* 57 (2), 103174.
- Seeber, I., L. Waizenegger, S. Seidel, S. Morana, I. Benbasat and P. B. Lowry (2020b). "Collaborating with technology-based autonomous agents" *Internet Research* 30 (1), 1–18.
- Sein, Henfridsson, Purao, Rossi and Lindgren (2011). "Action Design Research" *MIS Quarterly* 35 (1), 37.
- Sio, U. N., K. Kotovsky and J. Cagan (2015). "Fixation or inspiration? A meta-analytic review of the role of examples on design processes" *Design Studies* 39, 70–99.
- Stokel-Walker, C. (2023). "ChatGPT listed as author on research papers: many scientists disapprove" *Nature* 613 (7945), 620–621.
- Suh, M., E. Youngblom, M. Terry and C. J. Cai (2021). "AI as Social Glue: Uncovering the Roles of Deep Generative AI during Social Music Composition". In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. NY, USA: ACM.
- Venable, J. (2006). "The Role of Theory and Theorising in Design Science Research". In: *1st International Conference on Design Science in Information Systems and Technology*, pp. 1–18.
- vom Brocke, J., P. Fettke, M. Gau, C. Houy and S. Morana (2017). "Tool-Support for Design Science Research: Design Principles and Instantiation" *SSRN Electronic Journal*.
- vom Brocke, J., R. Winter, A. Hevner and A. Maedche (2020). "Special Issue Editorial – Accumulation and Evolution of Design Knowledge in Design Science Research: A Journey Through Time and Space" *Journal of the Association for Information Systems* 21 (3), 520–544.

- Zhang, C., C. Yao, J. Liu, Z. Zhou, W. Zhang, L. Liu, F. Ying, Y. Zhao and G. Wang (2021). “StoryDrawer: A Co-Creative Agent Supporting Children’s Storytelling through Collaborative Drawing”. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. NY, USA: ACM.
- Zhu, Q. and J. Luo (2022). “Generative Pre-Trained Transformer for Design Concept Generation: An Exploration” *Proceedings of the Design Society* 2, 1825–1834.

17 The More Is Not the Merrier: Effects of Prompt Engineering on the Quality of Ideas Generated By GPT-3

Memmert, L., Cvetkovic, I., & Bittner, E. (2024). The More Is Not the Merrier: Effects of Prompt Engineering on the Quality of Ideas Generated By GPT-3. In *Proceedings of the 57th Hawaii International Conference on System Sciences (HICSS-57)*.

Abstract

Generative language models (GLM) like GPT-3 can support humans in creative tasks. Such systems are capable of generating free-text output based on a provided input prompt. Given the outputs' sensitivity to the prompt, many techniques for prompt engineering were proposed both anecdotally in social media and increasingly in literature. It is, however, unclear if and how such a system and such techniques can be employed in creative contexts such as for generating ideas. In our study, we investigate the effects of using six prompt engineering techniques. For each combination of techniques, we have GPT-3 generate ideas for an exemplary scenario. The ideas are rated according to novelty and value. We report on the effects of the (combinations of) prompt engineering techniques. With our study, we contribute to the emerging field of prompt engineering and shed light on supporting idea generation with GLMs, showing a pathway to embedded GLM capabilities.

Keywords: brainstorming, generative language model, GPT-3, human-AI collaboration, prompt engineering

17.1 Introduction

Generative language models (GLM) like GPT-3 (Brown et al., 2020) have gained increasing attention for their ability to support humans in various creative tasks. Such systems are capable of generating free-text output based on a provided free-text input, typically referred to as a prompt. With recent advances, users can even directly interact with such GLMs via chat (e.g., ChatGPT). While such an interaction directly with the model is possible, GLM capabilities can be embedded into products, essentially abstracting away the complexities of interacting with the GLM for the users. An example of such a system is depicted in Figure 1. In this idea generation application prototype, the user can add a question, add first ideas, and request AI ideas if desired. The user can review the suggested ideas and select or ignore them as they wish. The suggestions are generated dynamically in the background. The user does not need any knowledge on how to interact with GLM but can still benefit from its capabilities. This removes the strain on the user to formulate a good prompt, shifting it to the tool designer. During tool development, the

prompt template to generate ideas needs to be defined. The template needs to be flexible to accept different questions for which the user might want to generate ideas.

The sensitivity of the GLM's output to the input (Zhao et al., 2021) raises questions about how to effectively design a good prompt template for ideation purposes. While prompt engineering techniques have been explored for typical natural language processing tasks (Brown et al., 2020), there is a lack of systematic investigation on the effects of these techniques in creative tasks such as ideation. Therefore, in this study, we aim to investigate the effects of six different prompt engineering techniques on the quality of ideas generated by GPT-3 on an exemplary idea generation question. To this aim, we pose the research question: *How does prompt engineering, using a GLM like GPT-3, affect the quality of ideas generated for an ideation task?*

To investigate this question, we developed a baseline prompt template and all variations of the prompt template considering the six prompt engineering techniques. We filled the prompt template with an exemplary brainstorming scenario and had GPT-3 generate ideas for each prompt. The quality of the generated ideas is assessed by three rates according to novelty and value, which are typical quality criteria for idea evaluation (Siangliulue et al., 2015). We analyzed the effects of the prompt engineering techniques and all their combinations on the idea quality. Our findings suggest that more is not necessarily merrier when it comes to prompt design for creative tasks. Combining certain prompt design techniques can negatively affect the quality of ideas generated by GPT-3.

Our study contributes to the emerging field of prompt design and aims to shed light on the potential of GLMs to support ideation. Moreover, it shows a potential path for abstracting away the complexities of directly interacting with GLMs while still benefiting from its capabilities, offering a wrapper UI, helping the user to ask questions to the GLM (Dang et al., 2022). Our paper offers the foundation for studies examining humans and AI brainstorming together, as it provides insight into a key design aspect: the underlying prompt template. This work informed a follow-up study in which the actual effect of providing AI suggestions to humans during a brainstorming session was assessed according to novelty and value of ideas (under review).

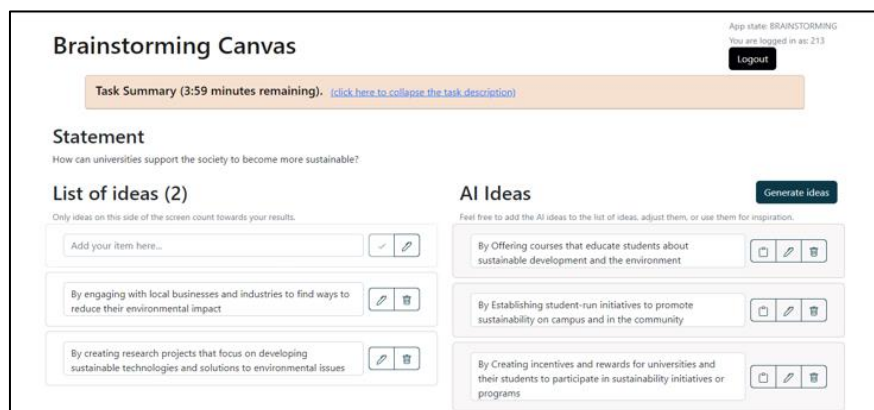


Figure 1. Screenshot of a GLM-based app prototype adapted from Memmert and Tavanapour (2023)

17.2 Background

17.2.1 Idea generation support

The goal of idea generation is to produce as many out-of-the-box ideas as possible to solve burning problems (Schallmo & Lang, 2020). To facilitate the creativity process, computer-based tools called Creativity Support Systems (CSS) have been developed. CSS has a long research history and focuses on different aspects of the creativity process (Przybilla et al., 2019). CSS can provide participants with cognitive or social stimulation (Pilcicki et al., 2022). Different tools have been developed and evaluated, and they show varying degrees of effectiveness (Gabriel et al., 2016).

One of the main areas of focus in CSS is idea generation tools, which aim to support the creative process by providing various techniques and methods for generating ideas. These tools can have a positive impact on the creative process, including improving the quality and quantity of generated ideas (Maaravi et al., 2021). However, these tools also have some drawbacks. One of the main challenges is that they often rely on a fixed set of methods and techniques, which may limit their effectiveness. Moreover, some of the tools may require a certain level of expertise or training to use them effectively (Frich et al., 2019). Recently, GLMs have emerged as a new technology that may offer a more effective way to support the creativity process.

17.2.2 Generative language models (GLM)

GLMs are a type of machine-learning-based AI systems. GLMs are trained on large corpora and are able to predict the next word given a certain input. GPT-3 is a powerful example of such GLMs (Brown et al., 2020). Due to their generative nature, GLMs have been proposed to be used for creative applications (Gero et al., 2022). However, there are several challenges with such systems. To interact with GLMs, an input text, i.e., a prompt, needs to be formulated. This can be difficult for novices, who might be surprised by the outputs of such systems at first use (Jiang et al., 2022). Dang et al. (2022) suggest supporting users in asking questions to the model. While recent advances allow humans to freely interact with GLMs via chat (e.g., ChatGPT), and one might educate users in effectively working with GLMs, a direct interaction might not always be feasible. In such cases, the GLM capability can be embedded into a system, hidden from the users. For this, a prompt template would need to be pre-defined at development time and populated with the user input at runtime. In this study, we investigate how such a prompt template should be formulated for an idea-generation application by exploring different techniques for phrasing prompts.

With regard to using GLMs for idea generation, Di Fede et al. (2022) suggested using GPT-3 for brainstorming but have not reported data yet. Stevenson et al. (2022) used GPT-3 (earlier version) for the “alternative use test” creativity test, generating ideas for alternative

uses for an object. In this study, we focus on using GLMs for a different divergent thinking task, i.e., to generate ideas for solving societal problems.

17.3 Method

17.3.1 Prompt template development

Requirements for prompt templates. We seek to develop a prompt template that can be embedded into an idea-generation app (see Figure 1). The prompt template should be *flexible* to be used for different questions, i.e., it should accept a parameter for the specific question at hand (see Figure 2). The prompt, once entered into the GLM, should produce a *list of N ideas* (instead of a free text), which can be shown as items to the user in the app. Lastly, the prompt should produce *good ideas*. The goodness of ideas can be operationalized or measured according to different criteria. For this study, we use *novelty* and *value* of ideas with a definition adapted from Siangliulue et al. (2015), as these reflect common evaluation criteria used in brainstorming research (Althuizen & Reichel, 2016; Haase & Hanel, 2023; Stevenson et al., 2022; Summers-Stay et al., 2023). *Novelty* will be rated by considering how *novel*, *original*, or *surprising* the idea is. *Value* will be judged by considering how *useful* and *practical* the idea sounds. We have included six prompt engineering techniques to investigate their effect on idea quality.

Baseline. First, we developed a baseline prompt. We tested if we could provide the question to the GLM directly without further input. Note that squared parentheses indicate a placeholder filled at runtime.

[question]

Prompting the GLM directly with the brainstorming question, however, resulted in inconsistent responses for certain questions (e.g., “How can generative language models be used?”). The results sometimes were lists of ideas (as intended), but sometimes free text, unfit for our app. Thus, we added a description of the task we expect the GLM to perform (Reynolds & McDonell, 2021), i.e., ‘Provide 3 ideas’. To receive a bullet-point list easily processible by our app, the technique of “itemization” might be used (Mishra, Khashabi, Baral, Choi, & Hajishirzi, 2022) by indicating a starting enumeration. In this way, the GLM “recognizes” the desired output to be a list. This template meets the first two formal requirements, as it is *flexible* to be used for different ideation questions and provides a *list of ideas*.

Provide 3 ideas for the question below.

[question]

1.

In the following, we develop prompt templates based on prompt engineering techniques, applying them to the idea generation context to assess differences in the goodness of the results. All selected prompt engineering techniques might be used in creative contexts and

require adding text to the baseline prompt template described above. The techniques fundamentally differ in their dependency on the specific brainstorming question. Three of the techniques are independent of the ideation question, while the others are question-specific. As a result, to implement the former techniques, changes only occur in the backend, whereas for the latter, the users would need to enter additional information (i.e., the GUI would need to be adapted). Table 1 offers an overview of the techniques with their respective characteristics; Figure 2 shows how the type of technique affects information flow.

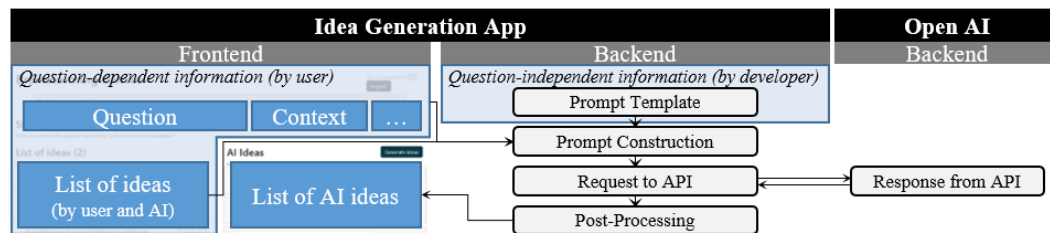


Figure 2. Illustrative depiction of the dynamic population of the prompt template with question-dependent information by the user to dynamically generate suggestions (via an API call)

Table 1. Prompt engineering techniques and characteristics

Ideation question dependency	Additional User Input	Prompt Engineering Techniques
Independent	No	- Evaluation criteria specification - Instructions/Schema - Demonstrations – question independent
Dependent	Yes	- Providing context - Expert perspective - Demonstrations – question dependent

Evaluation criteria specification (ECS). Specifying expectations is an important part of prompting GLMs. For brainstorming, we seek *good* ideas according to the two dimensions novelty and value with adapted definitions from Siangliulue et al. (2015), as these include common evaluation criteria for brainstorming ideas. Explicitly including expected idea characteristics in the prompt was done before (e.g., Stevenson et al., 2022) and can improve result quality (Summers-Stay et al., 2023). Thus, we included these criteria in the prompt (difference to baseline in blue):

Provide 3 **novel and valuable** ideas for the question below.
Novelty will be rated by considering how novel, original or surprising the idea is.
Value will be judged by considering how useful and practical the idea sounds.
 [question]
 1.

Instruction/schema (INS). Another suggested approach is to use schemas to identify the different prompt parts, e.g., by using headers (Liu et al., 2023; Mishra, Khashabi, Baral, & Hajishirzi, 2022) to label the task, question, and expected answer. This could be particularly relevant once more prompt techniques and, thereby, more information is added to the prompt.

Task: Provide 3 ideas for the question below.

Question: [question]

Answer:

1.

Demonstrations – question independent (DQI). Providing examples within the prompt can help to calibrate the GLM to better “understand” the expected results. This technique is commonly referred to as demonstration or few-shot learning (as opposed to zero-shot learning), builds on the idea of analogical reasoning (Q. Zhu & Luo, 2022) and was shown to be effective in some cases (Brown et al., 2020; Liu et al., 2023). For the area of creativity and design, Q. Zhu and Luo (2022) used design competition questions and winning ideas from previous years as examples to then generate new ideas. Inspired by this, we used pairs of brainstorming questions and good (scoring high for novelty and value) ideas published in prior studies (Nelissen, 2022; Siangliulue et al., 2015). We randomized the order of the selected ideas (demonstrations) to prevent sequential order effects (Zhao et al., 2021).

Provide 3 ideas for the question below.

What are ideas for new sports equipment products for the student market?

1. *Headphones that choose which song to play aligned with your bpm (so intense music for running/ cardio and chill music for yoga/cooling down)*
2. *AI analytics for evaluating form from video recording*
3. *A device analyzing blood to indicate which supplements and vitamins are needed*

Provide 3 ideas for the question below.

< question 2 – shortened here >

1. < question 2, idea 1 – shortened here >
2. < question 2, idea 2 – shortened here >
3. < question 2, idea 3 – shortened here >

Provide 3 ideas for the question below.

< question 3 – shortened here >

1. < question 3, idea 1 – shortened here >
2. < question 3, idea 2 – shortened here >
3. < question 3, idea 3 – shortened here >

Provide 3 ideas for the question below.

[question]

1.

Besides these prompt engineering techniques independent from the question, there are techniques depending on the specific question, such as adding context, expert perspective, and question-dependent demonstrations. While the former techniques aimed at improving the output quality independently from the input, these techniques adjust the input (i.e., would require additional input from the user and an adjustment of the GUI, e.g., by adding additional input fields).

Providing context (CON). GLMs generalize their outputs across the different contexts present in the training data. Thus, providing the *intended* context might improve the results to be tailored to the desired outputs (Liu et al., 2023). Besides using examples (few-shot learning), this can be achieved by adding additional information in a zero-shot learning (i.e., no examples) setting (Brown et al., 2020). In our tool, this could be reflected through an additional ‘context’ text field for the user to add context-related information.

Provide 3 ideas for the question below.

[context]

[question]

1.

Expert perspective (EXP). Similarly to adding additional context, asking the GLM to take a certain perspective or angle might improve the output quality. Reynolds and McDonell (2021) suggest prompting the GLM to take the perspective of a public figure or a specific role (e.g., teacher) to produce particular results; Haase and Hanel (2023) suggest including a profession. Similarly, we ask the GLM to take the role of an expert for a specific field, which is to be specified by the user.

Assuming you are an expert for [expert field], please provide 3 ideas for the question below.

[question]

1.

Demonstrations – question dependent (DQD). Previously, we explained that examples or “demonstrations” could be used to improve the GLMs' performance (few-shot learning). While the examples in the earlier technique were independent of the specific brainstorming question, once the user entered their first ideas, these ideas can be added dynamically to the prompt. For our example, we assumed the user had already added at least three ideas, and three of these ideas are included in the prompt. We randomized the idea order to prevent ordering effects (Liu et al., 2023; Zhao et al., 2021). In actual use, the best ideas might be selected algorithmically (Summers-Stay et al., 2023).

Provide 6 ideas for the question below.

[question]

1. [idea 1]

2. [idea 2]

3. [idea 3]

4.

The six selected techniques are independent of one another, allowing for any combination (i.e., all combinations are feasible). The goal of our analysis is to increase the understanding of which (combinations of) techniques lead to novel and valuable ideas. We developed a Python script to create all 64 prompt templates consistently, representing all combinations of the six prompt techniques described above (2^6 combinations from including none to including all techniques; no duplications). We recursively applied the techniques, i.e., when ‘Context’ and ‘Demonstrations – Question Independent’ were included, then context was also added to the demonstrations. Our templates are *prefix prompts* as the entire prompt text proceeds the expected output (Liu et al., 2023).

Table 2. Exemplary scenario

Component	Text
Context	Food waste is a major issue that affects both the environment and the economy. Globally, it is estimated that about one- third of all food produced is lost or wasted each year. Food waste is a huge source of greenhouse gas emissions and wasted natural resources, and therefore – reducing food waste could help to reduce global greenhouse gas emissions, establish food security, and encourage healthy food systems.
Question	How can we reduce food waste?
Expert Perspective	sustainability and environment
3 Examples (Demonstrations – question dependent)	• Reward grocery stores that donate their near-expired food to a food-bank where people can cook this donated food • Decompose leftovers and expired food, and supply them as fertilizers to local farmers and gardeners • Create an app that allows people to offer their leftovers or extra food to those who need it

17.3.2 Idea generation

To generate ideas for our experiment, we developed an exemplary brainstorming scenario. We used the societal problem of food waste (see Table 2). Open- ended problems for which ideas can be generated with relatively common knowledge are frequently used in brainstorming studies (Y. Zhu et al., 2020, 2021); the specific societal problem of avoiding food waste was adapted from Y. Zhu et al. (2021). While the question was copied, we added the question-specific details for context and expert perspective. For the example ideas, we used ideas that scored high in novelty and value from the same prior study. We filled all 64 templates with the contents for this specific scenario. Two additional scenarios we had included originally to increase the robustness of our results we had to abandon (as will be described in the idea evaluation section).

Populating the baseline prompt template resulted in the prompt below (scenario-specific content in orange):

Provide 3 ideas for the question below.
 How can we reduce food waste?
 1.

We used a custom Python script to populate all templates and make the requests to the OpenAI-API. We used the most powerful model, ‘davinci-003’, at standard settings but adjusted the temperature to 0.9, as per the documentation for creative applications. We did not make any changes to the model to improve the results or adjust them to the exemplary

scenario. We post-processed the results, removing any enumeration and cutting off the ideas after the first dot, as for some ideas, added explanations made the ideas very long. This resulted in 192 ideas (3 ideas per 64 prompts). No ideas were flagged by OpenAI's content moderation.

17.3.3 Idea evaluation

In an approach similar to Siangliulue et al. (2015), we generated random sets of about 25 ideas in random order for each scenario and had crowd workers assess them according to novelty and value. We provided the definition for both criteria in accordance with Siangliulue et al. (2015), with *novelty* as 'consider how novel, original or surprising the idea is' and *value* as 'consider how useful and practical the idea is'. We recruited participants on the Prolific platform, restricting to English-speaking individuals with a high approval rating. We aimed at 3 ratings per idea (Siangliulue et al., 2015). We included 3 randomly selected ideas all raters had to rate, enabling us to calculate inter-rater agreement. Though most participants passed the attention check items, we decided not to use the crowd-worker evaluation data due to poor data quality, i.e., poor inter-rater agreement.

Consequently, we had three blind-to-condition raters (one design thinking expert and the first two authors) independently rate all 192 ideas for one scenario on both dimensions. Having a set of judges rate all ideas is common in brainstorming research (e.g., Althuizen & Reichel, 2016; Y. Zhu et al., 2021), so is rating individual instead of sets of ideas (e.g., Haase & Hanel, 2023; Stevenson et al., 2022). We decided to include only one scenario, allowing for testing all techniques and their combinations while reducing potential fatigue from making many evaluations. Inter-rater agreement was 0.76 for novelty and 0.57 for value, or 'excellent' and 'fair' respectively (Cicchetti, 1994). Generated ideas included: "Develop software that would help shoppers calculate exact amounts of food they need depending on meal times and leftovers" (high novelty score), "Create an online platform for restaurants to share their surplus food with people who need it" (high value score), and "Implement "best by" labeling systems" (low novelty and low value score).

17.3.4 Data analysis

We used custom Python scripts for pre-processing, i.e., sorting the ratings for techniques and combinations. The clean data was analyzed with JASP statistics software (JASP, 2023). After performing assumption checks and confirming the normality of data distribution (Levene's $p=.747$), one-way ANOVA was performed to assess if there is an effect of prompt engineering on idea quality (value and novelty). ANOVA was performed on all possible combinations of techniques, leading to 64 pairwise comparisons. To determine the direction of the effect, significant results were further inspected with Tukey's post-hoc test, which also controls for the smaller number of observations in subgroups when performing pairwise comparisons. Since the variance of the individual ratings was higher than when averaged across the raters, we performed the same analysis on a dataset with averaged scores. Lastly, we checked whether prompt length affects the quality of the output with

correlation analyses. We chose Spearman's rho coefficient as the prompt length variable did not satisfy the normality assumptions.

17.4 Results

This study investigates the impact of using single and combined prompting techniques on idea novelty and value for a specific brainstorming question. Regarding single techniques, the results showed that the use of Evaluation Criteria Specification had a positive and significant impact on the novelty of ideas ($F=18.250$, $p<.001^{***}$, Figure 3-a), while Context had a negative impact compared to no technique ($F=5.464$, $p=.02^*$, Figure 3-b). The use of the other single techniques did not yield any statistically significant results.

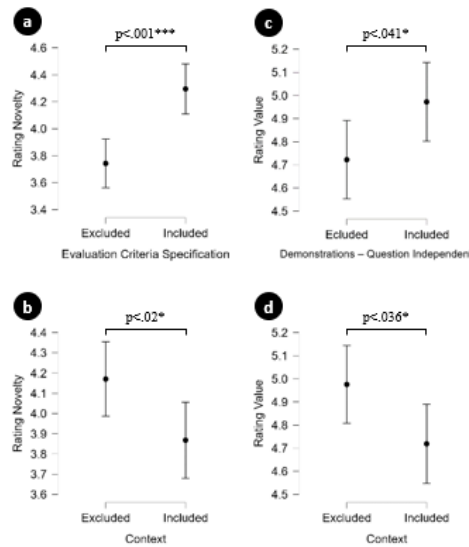


Figure 3. Techniques performing better than baseline (left) and worse than baseline (right)

Analyzing combinations of two techniques, the combination of INS and DQD showed a statistically significant effect ($F=4.736$, $p=.03^{**}$); however, the post-hoc analysis did not reveal any significant differences between groups. Combinations of three techniques: The combination of ESC, INS, and DQD showed a statistically significant effect ($F=5.216$, $p=.023^*$), with post-hoc analysis indicating that ESC was better than INS ($t=3.815$, $p=.004^{***}$) and DQD ($t=4.460$, $p<.001^{***}$). ESC alone performed better than the combination of ESC and DQD ($t=3.332$, $p=.021^*$) and INS and DQD ($t=4.030$, $p=.002^{**}$). The combination of CON, DQI, and EXP showed a statistically significant effect ($F=8.894$, $p=.003$), with post-hoc analysis indicating that combining DQI with EXP leads to more novel ideas than CON alone ($t=-3.439$, $p=.015^*$).

The combination of ESC, CON, DQD, and EXP showed a statistically significant effect ($F=4.736$, $p=.03^{**}$), with post-hoc analysis indicating that CON alone performed worse than the combination of ESC and CON ($t=-3.572$, $p=.033^*$). The combination of ESC and CON was also found to be better than CON and DQD ($t=3.648$, $p=.026^*$).

The combination of five techniques: ESC, INS, CON, DQD, and EXP showed a significant effect ($F=8.577$, $p=.003^{**}$), with post-hoc analysis indicating that ESC was better than the combinations of CON and DQD ($t=4.084$, $p=.018^{*}$); INS and EXP ($t=3.869$, $p=.04^{*}$), and ESC was better than the combination of ESC, INS, CON, and EXP ($t=4.084$, $p=.018^{*}$).

When it comes to value, the results indicate that few single techniques had a significant effect on value, with DQI performing better than when no technique was used ($F=4.194$, $p=.041^{*}$, Figure 3-c) and CON performing worse than baseline ($F=4.430$, $p=.036^{*}$, Figure 3-d). Combinations of two techniques showed that the combination of DQD and DQI had a significant effect on value ($F=9.437$, $p=.002^{**}$), with DQI performing better than no technique ($t=-3.620$, $p=.002^{**}$) and better than the combination of DQD*DQI ($t=2.735$, $p=.033^{*}$). The combination of DQD and EXP also showed a significant effect on value ($F=5.440$, $p=.02^{*}$), but post-hoc analysis did not reveal any significant differences. Combinations of three, four, and five techniques did not show significant effects on value. Despite significant effects after ANOVA, post-hoc tests did not yield significant differences ($p>.05$).

We also tested the effects of the techniques on a dataset averaged across raters to account for variance in ratings. This resulted in 3 times fewer observations, leading to a higher threshold to reach significance in pairwise comparisons. Thus, only ESC showed a positive effect on novelty ratings ($F=7.588$, $p=.007^{**}$). We investigated the effect of prompt length on output ratings. Correlation analyses show no difference in ratings based on prompt length ($\rho=.053$, $p=.465$).

17.5 Discussion

17.5.1 Answer to the research question

This study investigated the impact of using various prompting techniques on the novelty and value of ideas generated by GPT-3. The techniques tested included Evaluation Criteria Specification (ESC), Context (CON), Instruction/Schema (INS), Demonstration: Question Dependent (DQD), Demonstration: Question Independent (DQI), and Expert Perspective (EXP). The results showed that ESC had a positive impact on the novelty of ideas, while Context had a negative impact. Our study seems to align with previous findings of Summers-Stay et al. (2023), which suggest that idea quality may be improved by adjusting the prompt. They speculate that such systems may have picked up on creative ideas for other brainstorming questions in the training data and point to the capabilities of such models for analogical reasoning. Our prompt with ESC asking for “novel ideas” might then trigger the GLM to produce such ideas. Given the complexity and opacity of such models, the mechanisms for producing such results, however, remain speculative and require further testing. Combinations of two or more techniques showed significant effects on novelty, with some techniques working better together than others. However, a few single techniques had a significant impact on the value of ideas, and combinations of techniques

did not show any significant effects on value. The study's findings suggest that using specific combinations of prompting techniques in idea generation with generative AI can enhance the novelty of ideas generated, but further research is needed to determine their impact on idea value. This further aligns with earlier findings of Summer-Stay et al. (2023), who also only found a small difference through their more advanced approach aiming for increasing utility (i.e., a dimension of *value*). As stated before, given the complexity of such models and the scarcity of empirical research (Liu et al., 2023), one can only speculate as to the reason.

Additionally, the results showed that Evaluation Criteria Specification (ESC) performed better when used alone rather than in combination with other techniques. This finding suggests that adding many techniques to the prompt may not necessarily enhance idea quality (over-engineering). A potential explanation might be “spurious correlations” (Brown et al., 2020, p. 7), a problem that can occur when adding information beyond the task (e.g., examples). Instead, it may be more effective to use a targeted approach and select a few specific techniques based on the desired outcome. This highlights the importance of prompt engineering for creative tasks, which may require a different approach than other use cases where combinations of prompts can improve results (Mishra, Khashabi, Baral, & Hajishirzi, 2022; Wu et al., 2022). Therefore, it is important to carefully consider the use of multiple techniques and assess their impact on idea quality based on the specific context and goals of the task.

Our findings of only a few statistically significant differences regarding the use of prompt engineering techniques are somewhat surprising, contrary to anecdotal evidence from social media stressing the importance of prompt engineering. One reason why we find only a few differences between the prompts could be that we produce relatively short output in a restricted format (i.e., 3 bullet points as opposed to large blocks of free text). We observed many variations of the seemingly same idea. Potentially, there are some “obvious” answers that are provided first, essentially “overshadowing” prompt differences. However, this is speculation, and further research is required, e.g., using *calibration*, as discussed by Liu et al. (2023). Other potential reasons for us observing only a few significant differences are discussed in the limitations.

17.5.2 Contribution to theory

From a theoretical perspective, the study adds to the growing body of literature on using GLMs productively in creative applications (Gero et al., 2022; Shakeri et al., 2021; Yuan et al., 2022), particularly for the case of generating ideas in brainstorming-like settings (Di Fede et al., 2022; Haase & Hanel, 2023; Stevenson et al., 2022; Summers-Stay et al., 2023). Such idea-generation approaches typically rely on a specific, pre-defined prompt template, which may (or may not) include user-generated ideas. With this study, we enhance the understanding of the role of prompt engineering (Mishra, Khashabi, Baral, & Hajishirzi, 2022; Wu et al., 2022). Such an understanding is important, as there is an increasing number

of products that embed GLMs. For such systems, a crucial aspect of system design is to formulate an appropriate prompt template. In doing so, system designers essentially help the user to ask the GLM “the right questions” (Dang et al., 2022).

By examining the effect of different single and combined techniques on both idea novelty and value, we enhance the understanding of how to design systems for effective ideation with GLMs. The finding that ESC performed better when used alone rather than in combination with other techniques challenges the assumption that combining many techniques enhances the output quality. It rather highlights the need to consider using multiple techniques carefully in creative tasks and the importance of prompt engineering.

On a more abstract level, we contribute to the field of CSS by investigating the next-generation GLM-driven systems for supporting human creativity. Such a GLM-based brainstorming system was also suggested by Di Fede et al. (2022); with our analysis, we can inform the underlying technical design.

17.5.3 Implications for practice

From a practical perspective, the study offers valuable insights for organizations seeking to enhance their innovation potential. By identifying which techniques are most effective in enhancing the idea novelty and value, organizations can design more effective ideation sessions and improve their creative output. The finding that ESC has a positive impact on novelty highlights the importance of clearly defining brainstorming evaluation criteria to shift the focus on specific goals to generate more creative ideas. Additionally, the finding that combining DQI with EXP leads to more novel ideas than CON alone has practical implications for the use of expert perspectives in idea generation. By combining the expertise of internal and external experts with demonstrational prompts, organizations can generate more innovative ideas. However, results are to be interpreted carefully, as our study so far only indicates potential trends, but no general conclusions can be drawn.

There are large differences in the length of the prompts (14 to 614 words), particularly due to the *question-independent demonstrations* or *context* techniques. When costs for GLM usage are based on prompt length, this can result in cost differences for the prompt templates, which should be considered by tool designers, given the small differences in output quality.

17.6 Limitations and outlook

With only a few statistically significant differences between prompt engineering techniques, our results run counter to our (and many anecdotal) expectations. However, the results are to be interpreted carefully due to several limitations of our study that we discuss below.

While we based the candidate prompt development on existing literature, there are many ways of operationalizing the different prompt engineering techniques, and different implementations may result in different outputs. We, however, only tested one operationalization per technique, prioritizing testing more techniques over more operationalizations per technique. Additionally, the field of prompt engineering is rapidly evolving, and new techniques are constantly suggested. Thus, future research may expand on our results for both adjusting the operationalization of techniques or including additional techniques.

We have selected GPT-3 because it is a powerful, widely adopted model (Brown et al., 2020). However, there are other GLMs, and GLMs are constantly evolving, which, in some cases, affects the quality of GLM-generated ideas (Haase & Hanel, 2023). Additionally, there is a dependency between the effectiveness of prompts and model sizes (Liu et al., 2023), which could affect prompt techniques and transferability of results; some even suggest prompt engineering might become obsolete when models improve (Oppenlaender et al., 2023). Thus, future research should explore how different model (versions) together with prompt techniques affect the idea quality. Having 64 templates, each technique was present in 32 prompts. With 3 ideas per prompt, we had 96 ideas with and without each prompt technique, allowing for meaningful comparisons. However, including combinations of techniques reduced the number of observations per group. Tukey's post-hoc test takes into account the variability of the data and sets a threshold for statistical significance that is adjusted for the number of pairwise comparisons being made, e.g., if there are 16 pairwise comparisons, the adjusted p-value is $0.05/16=0.003$, thus making it harder to spot significant differences. To counteract this, future studies could rely on these results to investigate fewer techniques and additionally collect a higher number of observations.

Originally, we had planned to assess the robustness of our results across three societal problem questions; however, due to the low level of agreement among the crowd workers' responses, we had to use a panel of three raters to rate all ideas. To avoid fatigue, ideas for only one scenario were assessed (192 ideas on two dimensions). We thereby prioritized testing technique combinations and high data quality over additional scenarios. Thus, we cannot report insights on the robustness across problem questions. While the nature of the assessed information system – a general-purpose GLM without adaptations to our scenario – conceptually does not give any particular reason to assume that it would function completely differently across scenarios that are relatively common knowledge, future research should investigate these results to increase robustness.

For assessing idea quality, we selected two evaluation criteria (novelty, value) that reflect the criteria commonly used in literature. However, there are other criteria, such as practicability or plausibility of ideas (incl. costs), as well as criteria for sets of ideas such as *diversity* (Siangliulue et al., 2015), or even more question-specific criteria, such as social utility or moral value. Future research should further explore how GLMs could support

idea generation along such criteria. Lastly, inter-rater agreement for the three-rater panel was higher than for crowd-workers. However, an even better agreement, particularly for ‘value’, might have surfaced more statistically significant findings.

Given these limitations, we call for further investigation of different ways of operationalizing the prompt engineering techniques and an application to other problems to increase robustness. Our approach is transferable to other brainstorming questions and can serve as a foundation. Additionally, we suggest increasing inter-rater agreement, e.g., by increasing the number of raters or by using an evaluation scheme. More broadly, the ideation performance of humans working with such systems should be investigated.

While not at the core of our study, future research needs to investigate the implications of GLMs' inability to *understand* language and develop mitigation strategies for the issues of biases, falsehoods, and lack of moral judgment – discussed more broadly for GLMs (Bender et al., 2021; Floridi & Chiriatti, 2020; Lin et al., 2022; Susarla et al., 2023) – for creativity- and work- related use, particularly as brainstorming sessions are framing problem and solution corridors. Depending on the specific problem at hand, biased GLMs' suggestions could have severe negative consequences. We thus urge future research to explore sociotechnical perspectives of integrating generative AI in work settings by leveraging human expertise appropriately and preventing humans from accepting unfit suggestions, potentially inspired by existing research on the engagement of humans with AI systems' outputs from the area of AI-assisted decision- making (e.g., Bućinca et al., 2021).

17.7 Conclusion

GLM capabilities may be embedded into products, abstracting away the difficulties of interacting with the GLM directly while enabling unskilled users to still benefit from it. While this frees the user from developing a *good* prompt, it makes the design of a good *prompt template* during system development necessary. In our study, we develop a set of 64 prompt templates according to six prompt engineering techniques for an ideation tool, generate ideas via the GLM ‘GPT-3’ for an exemplary societal problem, and have these ideas evaluated. We find that prompt engineering techniques only in a few cases significantly affect idea quality positively. As this is counter to popular belief and our own experience, we carefully discuss potential explanations and limitations of our study. Particularly, we encourage further research on the effects of using prompt design for creative, divergent thinking tasks.

Acknowledgment. This research was funded by the German Federal Ministry of Education and Research (BMBF) in the context of the project HyMeKI (reference number: 01IS20057).

17.8 References

- Althuizen, N., & Reichel, A. (2016). The Effects of IT-Enabled Cognitive Stimulation Tools on Creative Problem Solving: A Dual Pathway to Creativity. *Journal of Management Information Systems*, 33(1), 11–44. <https://doi.org/10.1080/07421222.2016.1172439>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). ACM. <https://doi.org/10.1145/3442188.3445922>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., . . . Amodei, D. (2020). Language Models are Few-Shot Learners. In *NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems* (pp. 1877–1901). <https://doi.org/10.5555/3495724.3495883>
- Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1–21. <https://doi.org/10.1145/3449287>
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Dang, H., Mecke, L., Lehmann, F., Goller, S., & Buschek, D. (2022, September 3). *How to Prompt? Opportunities and Challenges of Zero- and Few-Shot Learning for Human-AI Interaction in Creative Applications of Generative Models*. <http://arxiv.org/pdf/2209.01390v1>
- Di Fede, G., Rocchesso, D., Dow, S. P., & Andolina, S. (2022). The Idea Machine: LLM-based Expansion, Rewriting, Combination, and Suggestion of Ideas. In *Creativity and Cognition* (pp. 623–627). ACM. <https://doi.org/10.1145/3527927.3535197>
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, 30(4), 681–694.
- Frich, J., MacDonald Vermeulen, L., Remy, C., Biskjaer, M. M., & Dalsgaard, P. (2019). Mapping the Landscape of Creativity Support Tools in HCI. In S. Brewster, G. Fitzpatrick, A. Cox, & V. Kostakos (Eds.), *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–18). ACM. <https://doi.org/10.1145/3290605.3300619>
- Gabriel, A., Monticolo, D., Camargo, M., & Bourgault, M. (2016). Creativity support systems: A systematic mapping study. *Thinking Skills and Creativity*, 21, 109–122. <https://doi.org/10.1016/j.tsc.2016.05.009>

- Gero, K. I., Liu, V., & Chilton, L. (2022). Sparks: Inspiration for Science Writing using Language Models. In F. Mueller, S. Greuter, R. A. Khot, P. Sweetser, & M. Obrist (Eds.), *Designing Interactive Systems Conference* (pp. 1002–1019). ACM. <https://doi.org/10.1145/3532106.3533533>
- Haase, J., & Hanel, P. H. P. (2023). *Artificial muses: Generative Artificial Intelligence Chatbots Have Risen to Human-Level Creativity*. <https://doi.org/10.48550/arXiv.2303.12003>
- JASP. (2023). *JASP (Version 0.17)[Computer software]*. <https://jasp-stats.org/>
- Jiang, E., Olson, K., Toh, E., Molina, A., Donsbach, A., Terry, M., & Cai, C. J [Carrie J.] (2022). PromptMaker: Prompt-based Prototyping with Large Language Models. In S. Barbosa, C. Lampe, C. Appert, & D. A. Shamma (Eds.), *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (pp. 1–8). ACM. <https://doi.org/10.1145/3491101.3503564>
- Lin, S., Hilton, J., & Evans, O. (2022). TruthfulQA: Measuring How Models Mimic Human Falsehoods. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*.
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Computing Surveys*, 55(9), 1–35. <https://doi.org/10.1145/3560815>
- Maaravi, Y., Heller, B., Shoham, Y., Mohar, S., & Deutsch, B. (2021). Ideation in the digital age: literature review and integrative model for electronic brainstorming. *Review of Managerial Science*, 15(6), 1431–1464. <https://doi.org/10.1007/s11846-020-00400-0>
- Memmert, L., & Tavanapour, N. (2023). Towards Human- AI-Collaboration in Brainstorming: Empirical Insights into the Perception of working with a generative AI. In *31st European Conference on Information Systems*. https://aisel.aisnet.org/ecis2023_rp/429
- Mishra, S., Khashabi, D., Baral, C., Choi, Y., & Hajishirzi, H. (2022). Reframing Instructional Prompts to GPTk's Language. In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 589–612).
- Mishra, S., Khashabi, D., Baral, C., & Hajishirzi, H. (2022). Cross-Task Generalization via Natural Language Crowdsourcing Instructions. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*.
- Nelissen, E. R. (2022). *Improving individual idea generation with the selective brainwriting technique*.
- Oppenlaender, J., Linder, R., & Silvennoinen, J. (2023). *Prompting AI Art: An Investigation into the Creative Skill of Prompt Engineering*. <https://doi.org/10.48550/arXiv.2303.13534>
- Pilcicki, R., Siemon, D., & Lattemann, C. (2022). How Feature- and Communication Constraints in CSS Affect Creative Collaboration in Virtual Teams - An Activity Theory Perspective. In *55th Hawaii International Conference on System Sciences*.

- Przybilla, L., Baar, L., Wiesche, M., & Kremer, H. (2019). Machines as Teammates in Creative Teams. In D. Joseph, C. van Slyke, J. P. Allen, J. Quesenberry, & M. Wiesche (Eds.), *Proceedings of the 2019 on Computers and People Research Conference* (pp. 94–102). ACM.
- Reynolds, L., & McDonell, K. (2021). Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm. In Y. Kitamura, A. Quigley, K. Isbister, & T. Igarashi (Eds.), *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–7). ACM. <https://doi.org/10.1145/3411763.3451760>
- Schallmo, D., & Lang, K. (2020). *Design Thinking erfolgreich anwenden: So entwickeln Sie in 7 Phasen kundenorientierte Produkte und Dienstleistungen* (2., aktualisierte Auflage). Springer Gabler. <https://doi.org/10.1007/978-3-658-28325-4>
- Shakeri, H., Neustaedter, C., & DiPaola, S. (2021). SAGA: Collaborative Storytelling with GPT-3. In J. Birnholtz, L. Cioffi, S. Ding, S. Fussell, A. Monroy-Hernández, S. Munson, I. Shklovski, & M. Naaman (Eds.), *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing* (pp. 163–166). ACM. <https://doi.org/10.1145/3462204.3481771>
- Siangliulue, P., Chan, J., Gajos, K. Z., & Dow, S. P. (2015). Providing Timely Examples Improves the Quantity and Quality of Generated Ideas. In T. Maver & E. Y.-L. Do (Eds.), *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition* (pp. 83–92). ACM.
- Stevenson, C., Smal, I., Baas, M., Grasman, R., & van der Maas, H. (2022, June 10). *Putting GPT-3's Creativity to the (Alternative Uses) Test*.
- Summers-Stay, D., Voss, C. R., & Lukin, S. M. (2023). Brainstorm, then Select: a Generative Language Model Improves Its Creativity Score. In *The AAAI-23 Workshop on Creative AI Across Modalities*.
- Susarla, A., Gopal, R., Thatcher, J. B., & Sarker, S. (2023). The Janus Effect of Generative AI: Charting the Path for Responsible Conduct of Scholarly Activities in Information Systems. *Information Systems Research*, 34(2), 399–408. <https://doi.org/10.1287/isre.2023.ed.v34.n2>
- Wu, T., Terry, M., & Cai, C. J [Carrie Jun] (2022). AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. In S. Barbosa, C. Lampe, C. Appert, D. A. Shamma, S. Drucker, J. Williamson, & K. Yatani (Eds.), *CHI Conference on Human Factors in Computing Systems* (pp. 1–22). ACM. <https://doi.org/10.1145/3491102.3517582>
- Yuan, A., Coenen, A., Reif, E., & Ippolito, D. (2022). Wordcraft: Story Writing With Large Language Models. In *27th International Conference on Intelligent User Interfaces* (pp. 841–852). ACM. <https://doi.org/10.1145/3490099.3511105>
- Zhao, Z., Wallace, E., Feng, S., Klein, D., & Singh, S. (2021). Calibrate Before Use: Improving Few-shot Performance of Language Models. In *Proceedings of the 38th International Conference on Machine Learning*.

- Zhu, Q., & Luo, J. (2022). Generative Pre-Trained Transformer for Design Concept Generation: An Exploration. *Proceedings of the Design Society*, 2, 1825–1834. <https://doi.org/10.1017/pds.2022.185>
- Zhu, Y., Ritter, S. M., & Dijksterhuis, A. (2020). Creativity: Intrapersonal and Interpersonal Selection of Creative Ideas. *The Journal of Creative Behavior*, 54(3), 626–635. <https://doi.org/10.1002/jocb.397>
- Zhu, Y., Ritter, S. M., & Dijksterhuis, A. (2021). The effect of rank-ordering strategy on creative idea selection performance. *European Journal of Social Psychology*, 51(2), 360–376. <https://doi.org/10.1002/ejsp.2743>

18 Quid pro Quo: Information disclosure for AI feedback in Human-AI collaboration

Cvetkovic, I., Grashoff, I., Jovancevic, A., & Bittner, E. (2025). Quid pro Quo: Information disclosure for AI feedback in Human-AI collaboration. *Computers in Human Behavior: Artificial Humans* 4, p. 100137. DOI: 10.1016/j.chbah.2025.100137.

Abstract

This paper explores the intersection of AI-driven teamwork facilitation and user privacy concerns in virtual collaboration. Specifically, we investigate the extent to which individuals in AI-supported teamwork environments are willing to disclose personal information to improve collaboration and performance. Leveraging a vignette-based study, we assess participants' intentions to disclose different types of personal information, such as language and choice of words, talk times, body language or sentiments, during a professional virtual collaboration process. Thereby, conditions under which the information is disclosed vary, referring to whether the results of information disclosure are either shared individually or with the whole group and whether they are shared anonymously or non-anonymously. Drawing on privacy calculus theory, our investigation further incorporates crucial contextual factors—Trust in AI, Perceived Risks, and Perceived Benefits—to comprehensively examine their influence on individuals' intentions to disclose information. Our findings reveal notable differences in Intention to disclose across various Information types, with lower intentions observed particularly for emotion and attention level disclosures. Surprisingly, other manipulated factors, including the level of anonymity, do not show an effect in influencing disclosure intentions. Crucially, our study underscores the pivotal role of Trust in AI, emerging as a consistent predictor of Intention to disclose across all Information types. Furthermore, its impact on disclosure intentions is mediated by individuals' perceptions of risks and benefits associated with disclosure. Our research contributes to the evolving field of privacy calculus theory by shedding light on the nuanced interplay between Trust, Perceived Risks, Benefits, and information disclosure to AI in teamwork scenarios. These insights bear implications for the effective deployment of AI in facilitating teamwork within the workplace, emphasizing the need for cultivating trust and understanding the specific sensitivities associated with different types of information.

Keywords: Human-AI collaboration, Information disclosure, Privacy calculus, Trust in AI

18.1 Introduction

Human-AI collaboration, often referred to as “Hybrid Intelligence”, represents a burgeoning and pivotal field of research (Dellermann, Calma, et al., 2019; Seeber et al.,

2020). This form of collaboration harnesses the distinctive strengths of humans and AI—ranging from creativity and intuition to large-scale data processing—to achieve results that surpass what either could attain independently (Dellermann, Ebel, et al., 2019). While much existing research has focused on scenarios involving a single human interacting with an AI (Debowski et al., 2021; Przybilla et al., 2019), rapid advances in AI offer new opportunities for AI to serve entire teams in collaborative problem-solving processes. In particular, AI's ability to consolidate and analyze vast (multi-modal) data, enhance real-time facilitation, and manage the flow of complex discussions can significantly improve how groups communicate and produce high-quality outcomes in virtual meetings (Joshi et al., 2022).

Especially in the virtual space, future developments in AI hold the potential to enhance real-time analysis and synthesis of multimodal social cues, provide interactive meeting analytics and feedback, and introduce intelligent virtual agents capable of aiding meeting leaders or acting as facilitators and moderators (Burgoon et al., 2017; Zadeh et al., 2017). In virtual meetings, the analytical capabilities of AI could be used to collect data during the session and present it back to the teams in a structured manner, thereby promoting collaborative work. As the output of virtual meetings is influenced by the interaction, cooperation, and atmosphere among participants, AI could collect different types of information, such as emotional and cognitive states or negative utterances to moderate team dynamics. By processing a multitude of inputs—emotional cues, vocal inflections, eye movements, and more—often in real time, AI can offer insights beyond what humans can perceive on their own (Bittner et al., 2021; Elshan et al., 2022). However, it's crucial to acknowledge that the integration of AI into virtual group meetings poses significant technological, psychological, ethical, and societal challenges. In this sense, recent research characterizes the resulting interaction between humans and AI as one of the most significant managerial issues of our time (Berente et al., 2021). While this “all-seeing” capability has the potential to boost creativity, productivity, and overall satisfaction by highlighting hidden patterns or detecting team dynamics before they derail a project, it also heightens concerns around privacy and trust. In any work meeting, people may choose to withhold certain feelings, e.g., fatigue or negativity, until they see fit to share. AI detection and disclosure of these feelings, even if well-intentioned, could interfere with nuanced group dynamics, as AI systems are not always well-versed in human ambiguity and may inadvertently call out or “pick on” individuals for information they preferred to keep private.

This challenge is tied to the trade-off between inherently problematic information disclosure (ID), i.e., the act of sharing personal or sensitive data to somewhat opaque AI systems and the potential benefits for the individuals and the team in virtual meetings. To offer benefits, especially in real-time, AI hinges on user data (Liao & Sundar, 2021; Russell & Norvig, 1995). A facilitation of meetings based on the recognition and analysis of multimodal social cues can only effectively function if team members trust the AI system

and voluntarily provide necessary data. This is not trivial, as it depends on many factors, such as the type of the benefit provided by AI, AI's transparency in the purpose of collecting data on social cues, associated perceived risk of disclosing this data, social dynamics within the team, and other. As shown in the recent case of Zoom's "attendee attention tracking," users are wary of features that might be overly invasive or used for unintended purposes (Zoom Video Communications, 2021). Consequently, the integration of AI into virtual group meetings not only promises efficiency gains but also intensifies a trade-off between the benefits of AI-driven insights and the privacy risks of disclosing sensitive information.

Thus, a comprehensive understanding of the intricate dynamics within (virtual) meetings is essential to preserve the authenticity, synchronicity, and accessibility of group communication while harnessing the diverse possibilities offered by AI and ensuring human-centered design.

Human-centered AI development hinges on understanding human needs and preferences, motivating user engagement with systems (Shneiderman, 2020). Information disclosure plays a pivotal role in this context, forming the basis for system recommendations and lifelong learning (Rayo & Segal, 2010). However, there remains a dearth of knowledge regarding human preferences concerning information disclosure in collaborative scenarios with AI-based systems, particularly in scenarios where users directly benefit in terms of task assistance, improved performance, or higher-quality results. This research seeks to uncover the advantages an AI-based system can offer human users in exchange for their information disclosure, and especially, whether Trust in AI, as well as Perceived Benefits and Risks of the specific AI feedback affects the willingness to disclose that particular information. To this end, our study aims to investigate how AI's capability to gather and process individual data in virtual meetings affects users' intentions to disclose different types of information. We propose the following overarching research question: RQ: *How does AI's capability to gather and interpret individual information in virtual meetings affect users' willingness to disclose such information?*

Within this broader question, we focus on three key sub-questions:

RQ 1: *Which information types are humans willing to disclose so that AI-based systems can support them in virtual meetings?* This question recognizes that "information" is not monolithic; some data may be perceived as more sensitive or more beneficial to share than others.

In order to investigate the effect of AI benefits to Information Disclosure, we pose the RQ 2: *How should the output from AI-based systems be designed, particularly regarding recipient (I:I vs. team) and anonymity, to foster information disclosure in virtual meetings?* Here, we explore how AI's real-time feedback might be perceived differently if shared privately with individuals versus publicly with teams, and whether anonymity modulates disclosure preferences.

Furthermore, in a try to capture the context- and personality-dependent complexity of ID, we pose a further RQ 3: *How does Trust in AI affect the Intention to disclose information?*

To address this research gap, we will test the research model shown in Fig. 1 by employing an online vignette experiment. The model bases on the privacy calculus theory and its associated constructs. For example, it can be derived from previous research that perceived risks and benefits depend on the level of trust (Metzger, 2004). These in return depend on various factors, among others the level of transparency related to disclosure purposes (Stutzman et al., 2011). Seven distinct Information types (e.g., mood states, attention levels) are shown to participants in hypothetical 1:1 or team scenarios. Within team scenarios, we additionally differentiate AI Feedback with two different levels of anonymity. For each Information type, variables shown in the model are measured and subsequently evaluated via a structural equation modelling (SEM) approach-path analysis. Furthermore, using ANOVA we assess whether differences in manipulations result in differences in intention to disclose information. To ensure generalizable results and sufficient statistical power, a sample of approximately 600 crowd-workers is recruited for the study. In conjunction with methodological advancements, these efforts aim to bridge the gap in interdisciplinary research within the realm of meeting science as well as offer insights into employee preferences, to guide the human-centered development of future technologies.

18.2 Theoretical background

18.2.1 Human-AI collaboration

In recent years, we have witnessed an increasing diffusion of automated systems based on AI technology that assist their users in fulfilling a variety of task. By some estimates, more than half of businesses worldwide have already been implementing some form of AI technology in 2020 to support their workforce (Balakrishnan, Chui, & Henke, 2023) - from advertising (Davenport et al., 2020), to management (Möhlmann et al., 2021), to customer service (Schanke et al., 2021) or product development (Recker et al., 2023), to name just a few examples. The span of use cases is wide and seems to grow even further due to the increasing capabilities of the technology (Westphal et al., 2023).

The degree of automation these systems provide can vary as much as the use cases they can be applied to. Ranging from “no automation”, meaning that there is no AI system in place to support the fulfillment of a specific task, but the human must complete the task manually, to “full automation”, which refers to the AI system as autonomously acting entity without dependence on human intervention, there is a broad continuum that allows for collaboration between AI system and its user (Westphal et al., 2023).

This so called “Human-AI collaboration” describes a scenario in which the human user and the AI system interact in the course of task fulfilling or decision-making, sharing their

complementary knowledge and capabilities in order to achieve common goals more efficiently (Westphal et al., 2023). As Dellermann et al. (2019) describe, using the concept of hybrid intelligence, this collaboration, based on the combination of human as well as artificial intelligence, leads to superior results compared to what each of them could achieve individually. Examples for such a cooperation can be found in various fields. For instance, an optical character recognition system that transmits instances to humans when the quality of fully automated transcription is not sufficient, relies on artificial as well as on human intelligence (Westphal et al., 2023). But there are also use cases in the field of education, where AI creates personalized online learning experiences, while human teachers play a major role for emotional guidance (Maedche et al., 2019); in finance, where AI can support the detection of fraud, while human experts can interpret the results and make decisions based on the findings (Milana & Ashta, 2021); or even in creative tasks like product innovation development, where AI can support the solution finding based on logic and accuracy, while humans take over the part of intuitive and creative thinking (Anantrasirichai & Bull, 2022; Recker et al., 2023).

The specific use case we are referring to is AI systems supporting and facilitating (virtual) team meetings. In general, facilitators of (virtual) meetings have the task of supporting teamwork and performance to optimize results. For this purpose, they monitor interactions between group participants and control the processes, thereby facing the challenge of collecting as well as structuring relevant information from the process and its participants at the same time (Schallmo, 2018). While human facilitators intend to fulfill the different tasks, they are often confronted with a high and demanding workload, dealing with an amount of unstructured data emerging from the process. This frequently results in important information being overlooked (Chen et al., 2021). Recently, digital technologies have been used and show great potential to support, for example, ideation processes, although mostly at the individual level (Debowski et al., 2021; Przybilla et al., 2019). Since individual-level settings differ fundamentally from group-level settings, leaving the potential for structuring large amounts of emerging data unexploited, we suggest exploring the opportunities of digital innovations, especially AI-based systems, to support collaborative problem-solving processes in teams (Cvetkovic et al., 2022). The considerable potential of AI to support collaborative processes transcending the cooperation of dyads of individuals and AI systems but involving groups of humans and AI systems as collaborative counterparts is also emphasized by Bittner, Oeste-Reiß, and Leimeister (2019). They point out that so far, little is known about how to conceptualize collaborative work practices with an AI system being a facilitator, expert, or peer within the process. Toxtli et al. (2018) implemented an AI-based conversational agent that can handle task management for individual as well as group work. In doing so, they identified particular difficulties and opportunities in communication between the AI-based agent and a group of multiple users. Winkler et al. (2019) examine how far AI-based smart personal assistants are able to facilitate collaboration between multiple users. Their results indicate

that groups interacting with smart personal assistants acting as facilitators show significantly better task outcomes as well as an improved quality of collaboration.

Thus, collaboration between humans and AI may improve group performance, but it also poses questions about the dynamics and characteristics of the emerging human-AI teams. For instance, the design of explainable and transparent AI agents as well as the design of a shared workplace, including the definition and delegation of distinct tasks (Lubars & Tan, 2019; Cvetkovic & Bittner, 2022) and roles (Siemon, 2022) to both, human and AI agents, need to be considered for a successful collaboration.

In our case, we focus on the example of AI facilitating team meetings that take place in a virtual environment as an intelligent virtual collaborator (Cvetkovic et al., 2022).

Virtual (group) meetings are increasingly becoming a standard for exchange and collaboration (Karl et al., 2022). Due to the convenience they provide in terms of remote communication, they present an eco-friendly and time-saving way of communicating and collaborating (Guerin, 2017; Waizenegger et al., 2020). In doing so, virtual platforms already make use of AI-supported functions today, e.g., in terms of background customization aiming to exclude unnecessary stimuli from the place of attendance, or noise suppression (Joshi et al., 2022). Beyond that, future technological advancements hold the potential to make virtual meetings even more efficient, e.g. by enhancing real-time collection, analysis and synthesis of multimodal social cues, providing meeting analytics, and, based on this, introducing the role of a virtual meeting facilitator, supporting the evolving team dynamics (Joshi et al., 2022).

In this context, making use of AI in order to decode cognitive and emotional states of participants based on the collection and analysis of non-verbal cues during online meetings is a particularly promising intention (Joshi et al., 2022). As Stanton (2004) shows, human communication - as it can be found in group meetings in person as well as in virtual environments - relies heavily on participants' mutual interpretation of non-verbal social cues like body language, e.g. to understand emotional states and respond to them accordingly. However, recognizing and assigning emotional states is more difficult in a virtual space, since online communication is only able to approximate face-to-face communication due to the limited transmission of these non-verbal behavioral cues (Joshi et al., 2022). Consequently, AI could serve as a facilitator for online meeting processes, collecting and analyzing non-verbal cues and playing the recognized emotional states back to a human moderator or the participants themselves in order to motivate the group, to keep track of team dynamics or to prevent conflicts (Elshan et al., 2022; Joshi et al., 2022).

Nevertheless, AI-enhanced virtual meeting platforms also pose challenges that need to be taken into consideration. The introduction of an AI-based intelligent agent to a virtual group meeting may be accompanied by social (Cebulla, Szpak, Howell, et al., 2023), psychological (Malik et al., 2021), ethical (Molka-Danielsen et al., 2021) as well as technological issues (Joshi et al., 2022).

For instance, research shows that AI-based detection of social cues, such as tone of voice or facial expression for the purpose of emotion recognition, often encounters resistance from the participants of the meeting in terms of information sharing (Knijnenburg & Kobsa, 2013; Mantello et al., 2023). Furthermore, AI detection and disclosure of personal social cues could interfere with nuanced group dynamics, leading to the violation of social norms or even the manipulation of human participants (Ferrada & Camarinha-Matos, 2024; McDuff & Czerwinski, 2018).

Consequently, exploring and addressing privacy concerns becomes crucial in the attempt to establish a trusting collaboration between the participants of a meeting and the AI system introduced to facilitate the same (Molka-Danielsen et al., 2021). The next section will thus elaborate the question to what extent and why individuals are willing to disclose their personal information in different contexts, referring to the concept of privacy calculus.

18.2.2 Information disclosure

In today's digital landscape, ID plays a pivotal role in shaping how individuals interact with various online platforms and services. At its core, ID refers to the extent to which users are willing to divulge their private information to external entities (Knijnenburg & Kobsa, 2013). This concept is not only relevant but also prevalent, particularly in the age of social media (Kroll & Stieglitz, 2021), e-commerce (Berendt et al., 2005), and digital communication (Karl et al., 2022).

The decision-making process behind ID is a multifaceted one, often guided by what is known as “privacy calculus.” (Culnan & Armstrong, 1999). This intricate model posits that individuals weigh the perceived privacy risks and benefits before making the choice to disclose personal information. In essence, privacy calculus turns ID into a conscious decision, driven by a cost-benefit analysis where individuals meticulously evaluate the advantages and disadvantages of sharing their data (Bol et al., 2018; Knijnenburg & Kobsa, 2013; Norberg et al., 2007).

However, privacy calculus does not operate in a vacuum. Contextual and situational factors shape how users perceive risks and benefits. Research on privacy calculus in online social networks suggests that individuals manage disclosure decisions through a dynamic trade-off between privacy concerns and self-disclosure motivations. The Dual Privacy Decision Model proposed by James et al. (2015) highlights that factors such as privacy value, perceived risks, trust, and control mechanisms influence how individuals disclose personal information in social media environments. Privacy calculus extends beyond social media into e-commerce settings, where users weigh security concerns, trust in institutions, and perceived risks when disclosing information for online transactions. A comparative study by Dinev et al. (2006) found that cultural differences moderate privacy concerns, with Italian users displaying higher perceived risks and lower institutional trust compared to U.S. consumers. These findings underscore the importance of contextual and cultural

factors in shaping privacy decision-making, further supporting the dynamic nature of privacy calculus theory.

Beyond this, the decision to disclose is further influenced by uncertainty and lack of control over data usage. Without clear knowledge of how data will be used or misused, individuals struggle with disclosure decisions. These dynamics are also relevant to AI-based interactions, where users may fear that AI's advanced data-processing capabilities amplify privacy risks, creating a similar sense of scrutiny and lack of control as observed in crowdsourced social media settings (James et al., 2015).

These context-specific factors influencing ID are not isolated but rather interconnected, creating a complex web of determinants that directly affect the privacy calculus. Research on ID spans various domains and theoretical frameworks, including social exchange theory (Thibaut & Kelley, 1959), social contract theory (Milne & Gordon, 1993), and, notably, privacy calculus theory itself (Culnan & Armstrong, 1999). Social exchange theory (Thibaut & Kelley, 1959), for instance, proposes that individuals decide whether to engage in a social act based on their evaluation of interpersonal rewards and costs associated with that act. Applied to the realm of information exchange, this theory suggests that when the perceived benefits outweigh the perceived risks, ID is more likely to occur.

A critical component in ID is trust. Trust in the entity collecting the data, the transparency of data handling practices, and perceived security all influence disclosure behavior (Masaviru, 2016). Trust plays a central role in mitigating privacy concerns, particularly in online transactions. Dinev and Hart (2006) extend privacy calculus theory by demonstrating that while privacy concerns inhibit disclosure, the effects can be offset by trust in the platform and personal interest in the service. These findings reinforce the importance of trust in AI-driven decision-making, where perceived security and transparency can directly shape disclosure behaviors. Beyond privacy calculus, trust and anonymity play a significant role in self-disclosure behaviors in professional online environments. The Online Community Self-Disclosure Model proposed by Posey et al. (2010) highlights how trust in the community, perceived risk, and anonymity influence disclosure intentions among professionals in France and the UK. The findings suggest that users in professional online spaces strategically disclose information based on their perception of the audience, data security, and expected benefits. These insights are particularly relevant for AI-facilitated virtual meetings, where trust in the AI system, perceived control over shared data, and anonymity concerns may similarly influence disclosure decisions.

18.2.2.1 Information disclosure to AI

The introduction of AI has raised significant privacy concerns, further complicating the ID landscape. AI systems, renowned for their ability to process vast amounts of data, have immense potential for benefit but are also fraught with risks. Reliable AI assistance hinges on access to user data, a fact exemplified by platforms like Netflix, which rely on users'

viewing history and ratings to provide tailored recommendations (Liao & Sundar, 2021). However, this dependence on personal data has raised pertinent questions about the accuracy, transparency, and accountability of AI systems. The personalization of AI systems can significantly enhance their functionality, but it simultaneously raises concerns about the accuracy of the information they generate and the transparency of the processes underlying their decisions (Tese & Silvanio, 2023). For instance, disseminating inaccurate information generated by AI could potentially result in defamation lawsuits against the creators of the AI or the businesses employing it. The risks associated with AI extend beyond accuracy, as business applications of AI may involve the collection and dissemination of sensitive personal or business data. Such practices can potentially run afoul of various legal statutes, including the Federal Trade Commission Act, the Defend Trade Secrets Act, and privacy laws like the U.S. Privacy Act or the European Union's General Data Protection Regulation (GDPR) (Tese & Silvanio, 2023). The collection of specific information types such as emotions in the context of workplace, schools, social scoring, and predictive policing has now been forbidden in the EU by the new EU AI Act.¹

A recent study examined employees' concerns regarding having an AI assess their performance in four different scenarios and identified privacy as one of the main hurdles - AI brings about the risks of using private information about employees that they would not prefer to share (Park et al., 2021). Mantello et al. (2023) discuss attitudes towards empathic surveillance by analyzing a survey of 1015 responses of future job-seekers from 48 countries. They define empathic surveillance as a product of a new generation of biometric devices that can sense, read, monitor and evaluate the affective state of a worker. The study results highlight three major issues for future job-seekers who will be subject to non-human resource management, particularly involving emotion-sensing biometric technologies. The first concern pertains to privacy, as these technologies blur the line between personal and professional life, intrude into real-time emotional states, and expose employees to both top-down and peer-to-peer surveillance under the guise of workplace well-being programs (Mantello et al., 2023). Several other studies confirm the adverse affects AI can pose to one's privacy (Cebulla, Szpak, Howell, et al., 2023; Manseau, 2020; Molka-Danielsen et al., 2021). For example, Cebulla, Szpak, and Knight (2023) show the significant negative consequences associated with the integration of AI, specifically highlighting issues related to information security, data privacy, profound alterations resulting from digital transformations, and the emergence of job-related concerns and insecurity.

Most recently, the debate over AI regulation has intensified, particularly with the European Union's efforts to regulate generative AI through the AI Act. This initiative has sparked significant opposition from major countries, raising a crucial question: Are the immense benefits of AI, such as a 40 % increase in efficiency reported by BCG consultants (Dell'Acqua et al., 2023) so compelling that we may be inclined to overlook the associated risks?

As we use virtual meetings as our application case, we hereafter discuss the recent findings about ID in virtual meetings. Virtual meetings have become a common platform for ID, especially during the COVID-19 pandemic (Karl et al., 2022). Introduction of AI into virtual meetings has a potential to change its dynamics. Research suggests that disclosing information to AI during these meetings, rather than to humans, can have a tangible impact on brand-related outcomes. Factors like audience size inferences and the perception of exploitation can influence brand trust in this context (Lefkeli et al., 2023). A recent Wired article (Pardes, 2020) discusses how AI can optimize meetings, from automated scheduling tools to facial recognition that measures who's paying attention. It also mentions privacy and surveillance issues related to the use of AI in virtual meetings, illustrated with the example of a recent Zoom privacy incident. Under its new terms, Zoom claimed the right to use personal video, audio, and chat data for its AI programs (Vaughan-Nichols, 2023). In online learning scenarios, AI systems have been positively recognized for improving the quantity and quality of communication, for providing just-in-time, personalized support for large-scale settings, and for improving the feeling of connection. However, there were concerns about responsibility, agency, and surveillance issues (Seo, Tang, Roll et al., 2021).

In conclusion, the world of information disclosure is a complex tapestry that spans across a multitude of domains and is influenced by various factors, from individual choices to technological innovations. Balancing the benefits and risks of information disclosure has never been more critical, especially within the overarching context of safeguarding individual rights and privacy. Strategies to navigate this evolving landscape include privacy self-management, transparent policies, and innovative protection frameworks like federated learning (Wei et al., 2020), which leverages decentralized data sharing techniques to safeguard information privacy. Furthermore, in response to mounting privacy apprehensions, sensitive data collection practices, such as Zoom's attention tracker (Zoom Video Communications, 2021), may face discontinuation. As technology evolves and societal norms adapt, the intricacies of information disclosure and privacy will continue to shape our digital interactions.

18.3 Hypothesis development

Drawing on the literature on Human-AI-collaboration and privacy calculus, we develop a coherent understanding of the human preferences to cooperate with AI in team-based problem-solving processes. Beyond assuming an effect of different manipulations (Information type, Anonymity of the feedback, Recipient of the feedback), our research model presented in Fig. 1 contains the following relations: The participants' Trust in AI influencing Perceived Risks and Benefits of the disclosure decision under discussion, which are in turn influencing participants' actual Intention to disclose their information. An overview of the variables can be found in Table 1.

Table 1. Variables and definitions.

Variable	Definition	Source
Information type	The specific category of information the AI uses to generate feedback, such as talk times, attention levels, or interruptions within the team	Authors
Recipient	Recipient of AI feedback, an individual or team	Authors
Anonymity	Whether or not the identities of team members are revealed in the AI feedback	Authors
Trust in AI	An attitudinal judgement of the extent to which the user believes that the automated system can effectively help them accomplish their objectives when faced with uncertain circumstances	Merritt (2011)
Perceived Risks	Perceived subjective probability of the occurrence of negative consequences associated with the act of disclosing certain information	Bol et al. (2018)
Perceived Benefits	Perceived subjective probability of the occurrence of rewards associated with the act of disclosing certain information	Bol et al. (2018)
Intention to disclose	The inclination to reveal specific personal data or details to others	Norberg et al. (2007)

The type of information that is asked to be disclosed is shown to have a major impact on the related disclosure decision.

For example, Belen-Saglam et al. (2022) examine the information disclosure behavior of British citizens in different scenarios, based on different data items ranging from general to health or finance information. Their results imply that, depending on the potential consequences of disclosure, the nature of the information and possible concerns regarding with whom, another person or a system, the information is shared, participants ascribe different levels of sensitivity to the data items they are asked to disclose. As a result, the context of information disclosure, the fairness of the disclosure request and the level of anonymity in the process of disclosure are ascertained as central aspects for participants' comfort regarding their specific disclosure decision.

Knijnenburg et al. (2013) examine information disclosure as a multidimensional behavior and identify three distinct groups that behave inherently different alongside the elaborated dimensions. One of their central findings suggests that, however, the distinct groups do not necessarily differ in their overall degree of disclosure, but rather in their disclosure tendencies per kind of information.

Similar results are presented by Degutis, Urbonavicius, Zimaitis, Skare, and Laurutyte (2020), whose study explores the multidimensionality of the construct "willingness to disclose personal information". Their analysis suggests that three types of willingness, linked with three different types of data, can be distinguished: the willingness to disclose

personal data that includes individual facts or profile data, social networking data, and online browsing or purchasing data. In this work, we distinguish between seven Information types (described in 4) and consequently hypothesize:

H1. Different Information types yield different Intention to disclose that particular information.

Beyond the type of information that is being asked to disclose, we assume that the context of information disclosure is crucial for the respective decision. Belen-Saglam et al. (2022) show that people are generally concerned about disclosing personal information to others in terms of discrimination, embarrassment, judgement or reputational harm, depending on whom they share their information with. Furthermore, research has examined the constellations in which information disclosure is likely to happen. Collins and Miller (1994) have shown that people tend to disclose themselves to others whom they like, what could be a potential obstacle for information disclosure in teams if there is no mutual sympathy. In this context, research also shows that leaders need to take over a central role in fostering self-disclosure in their teams, e.g., by creating a psychologically safe environment or by making time and space for informal communication. Consequently, the conditions for self-disclosure must be purposefully created in advance (Cooks-Campbell, 2023). This might be more difficult to achieve in team level settings than in individual level settings. Furthermore, recent studies show that the critical element of privacy in data sharing is the protection of the identity, and thus, anonymity, which may also seem to be more likely to achieve in an individual level setting (Belen-Saglam et al., 2022; Schomakers et al., 2019). Therefore we assume that information disclosure decisions might be handled differently in team vs. in individual contexts.

Based on this, we formulate the hypothesis:

H1a. There is an interaction effect of Information type and Recipient on the Intention to disclose.

In connection to this assumption, we suggest that the degree of anonymity impacts the willingness to disclose information at team level. As already mentioned, recent studies report on the importance of identity protection and thus anonymity in online privacy and data sharing (Schomakers et al., 2019). Similarly, Belen-Saglam et al. (2022) show that, depending on the context of information disclosure as well as the type of information being disclosed, people feel more comfortable to disclose their data anonymously or non-anonymously. Consequently, we hypothesize:

H1b. There is an interaction effect of Information type and Anonymity of the feedback on the Intention to disclose.

Besides the type of information that is asked to be disclosed, we assume that Trust in AI influences participants' willingness to disclose their information. Trust is a fundamental component of interpersonal human-human relationships. Since studies have highlighted

that human-automation interaction adopts similar norms to human-human interaction, it stands to reason to assume that trust is also of particular importance in human-AI-interaction (Hoff & Bashir, 2015; Madhavan & Wiegmann, 2007; Omrani et al., 2022).

For example, Choung et al. (2023) examine the role of trust in the use of AI-based voice assistants. Their results suggest that trust significantly impacts the intention to use respective systems. In this sense, trust particularly takes effect through participants' perceived usefulness of as well as their general attitude towards these systems. The authors also examined different dimensions of trust. Their results suggest two different types of trust influencing the intention to use AI: human-like, referring to values like explainability, transparency, or fairness being implemented into the AI system, and functionality trust, describing the competency of the distinct technological features.

In contrast to this, Solberg et al. (2022) show, that a higher level of perceived situational risks, including, among others, time loss risks, performance risks, privacy risks as well as social or ethical risks, decreases human Trust in AI, what in turn may lead to a lower reliance on the AI's decisions.

As research confirms, trust is an essential prerequisite not only for AI adoption but also for information disclosure behavior. As Masaviru (2016) shows, trust plays a pivotal role in individuals' willingness to disclose information. Trustworthiness of the entity requesting the information, transparency in data handling practices, and perceived benefits and costs all intertwine to shape trust and, consequently, the decision to disclose. Similar results are provided by Steel (1991), who specifically shows that varying degrees of information disclosure are related to varying degrees of perceived trust.

These findings are also supported by Al-Natour et al. (2021), who show that trust significantly impacts information disclosure intentions in the context of using a virtual advisor. Thus, we formulate the hypotheses:

H2. Higher Trust in AI predicts higher Intention to disclose information.

Specifically, we assume that higher Trust in AI leads to lower Perceived Risks, while strengthening the perception of Benefits.

As discussed before and in accordance with the literature on human-machine interaction, users do not rely on technology they do not trust. Thus, trust is well-known as a critical factor in technology and specifically in AI adoption (Aoki, 2020; Choung et al., 2023; Gao & Waechter, 2017; Hengstler et al., 2016).

Beyond this, recent research has started examining the role of perceived risks and expectations towards AI as critical factors influencing and being influenced by an individual's trust towards AI. As Choudhury et al. (2022) show exemplary for the healthcare sector, clinicians' trust in an AI-based decision support system increased their belief that the use of the system would improve their performance, which in turn led to a higher intention to use it. Similarly, the authors determine a statistically significant negative

effect of trust on perceived risks, meaning that lower trust in the AI system led to higher perception of risks and in turn also lowered the clinicians' intention to use the system. Based on this, we hypothesize:

H2a. Higher Trust in AI predicts lower Perceived Risks.

H2b. Higher Trust in AI predicts higher Perceived Benefits.

Finally, we assume that the perceptions of benefits through information disclosure will lead to a higher intention to disclose respective information, while the perception of risks will lower the intention to disclose. This assumption is inherently connected to the concept of privacy calculus describing how a disclosure decision essentially depends on individuals consciously weighing the risks and benefits of their disclosure decisions, evaluating the advantages and disadvantages of sharing their information (Knijnenburg & Kobsa, 2013). For example, Wang et al. (2023) show how perceived benefits have a positive impact on information disclosure intentions of Chinese COVID-19 patients on social media. Similar results underline that individuals tend to be more comfortable in disclosing their information when their human needs such as health or security are fulfilled (Wadle et al., 2019). Similarly, examining the context of AI-based recommendation systems, Li and Unger (2012) show that users perform an evaluation of the extent to which providing personal information in exchange for support through AI improves the quality of the respective advice. This evaluation of the potential effects of information disclosure on the quality of outcomes will then influence in how far the user is willing to disclose. These results are in accordance with research on the personalization privacy paradox, which has shown that perceived benefits essentially increase the perceived value of disclosing, which in turn leads to an increased willingness to disclose the information needed to personalize a service (Xiao & Benbasat, 2007).

On the other hand, research shows that, e.g. perceived privacy risks, essentially lower individuals' willingness to disclose their information; in the most drastic case leading to the perceived risks outweighing the perceived benefits of the disclosure decision (Keith et al., 2013; Yu et al., 2020). For instance, similar studies adopting the theoretical concept of privacy calculus have shown that expectations of loss of privacy lead to a reduced willingness to provide information and discourage their actual disclosure (Dinev et al., 2013). Consequently, we state the hypotheses:

H2c. AI Feedback perceived as more risky predicts lower Intention to disclose particular information.

H2d. AI Feedback perceived as more beneficial predicts higher Intention to disclose particular information.

Following from this, we hypothesize that:

H2a, H2b, H2c, H2d: Perceived Benefits and Perceived Risks mediate the relationship between Trust in AI and Intention to disclose information.

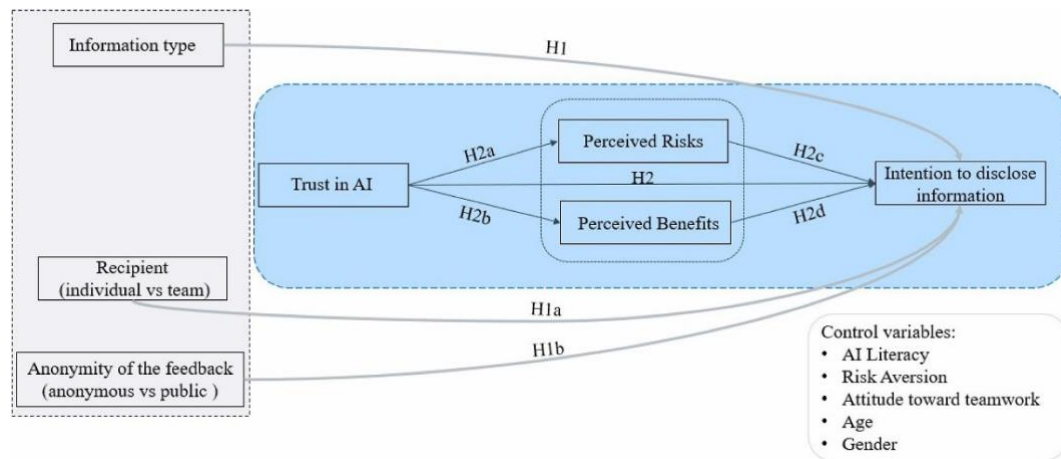


Figure 1. Research model.

Summing up the theoretical model, the independent variables (IV) are Information type, Recipient and Anonymity of the feedback. We hypothesize that different levels of these variables lead to different Intentions to disclose. Recipient and Anonymity are independent factors that we manipulate - according to these the respondents were separated into three groups (individual recipient does not have Anonymity manipulated since AI feedback here is already anonymous per se). In the blue dashed box, we further include variables that are, according to the literature, most frequently influencing ID, i.e., related to the privacy calculus. Hereby, Trust in AI is a predictor. It predicts Intention to disclose information both directly and through Perceived Risk and Perceived Benefits. Perceived Risks and Perceived Benefits are mediators between Trust in AI and Intention to disclose information. The main idea is to evaluate the privacy calculus (blue part of the model) for each factor condition and Information type separately (anonymized in a team, non-anonymized in a team, and individual).

18.4 Method

For the purpose of testing the above-mentioned research model and its hypotheses, we contend an elaborate study design based on a vignette-based experiment. This, along with the instruments used to measure the variables from the model, as well as the data collection and analysis, are described in the following.

When developing the experiment design, we contended manipulating 3 independent variables. First, Information Type was addressed with 7 categorical levels of human input and corresponding AI feedback: talk times (calculated percentage of each person talking in a meeting), breaking rules of creative teamwork (identified keywords that indicate that the rules of teamwork are broken), bias toward certain solutions (identified, e.g., when a person keeps arguing toward own ideas only), interruptions (identified when one team members interrupts another while talking), attention level (identified when team members show signs of exhaustion), moodboard (identified emotions of team members), and stagnation (identified when no new ideas have been developed after a certain period of time during

brainstorming). We choose these Information types specifically as this is an area of potentially beneficial AI assistance in teamwork with major ID prerequisites (Cvetkovic et al., 2022).

As mentioned in the Chapter 2.2, ID is highly context-dependent. What an individual may willingly share in one scenario, might be unacceptable for them to share in another scenario (Xu et al., 2008). Therefore, we consider individual and team scenarios to provide fairly different context, which might influence the disclosure decision. Thus, we vary the recipient of the AI feedback (whole team or individuals).

Finally, in the team condition, we vary AI feedback in terms of Anonymity (anonymized vs. non-anonymized), to detect the influence of how, and with whom the information is shared (Patil & Lai, 2005; Stutzman et al., 2011). In the anonymized condition, the names of team members are not displayed alongside their respective contributions, while in the non-anonymized condition, names are included. To assess the impact of the recipient and form of intervention, we randomly split participants into three groups: individual feedback (IN), team non-anonymized (TN), and team anonymized (TA). For example, in the TN group, the names of each team member are displayed (Fig. 2, Condition TN). In total, 21 vignettes were evaluated, however, each participant only saw 7 vignettes relevant to their experimental condition (e.g., participants in the IN condition only saw the leftmost vignettes from Fig. 2 and Appendix 7.1).

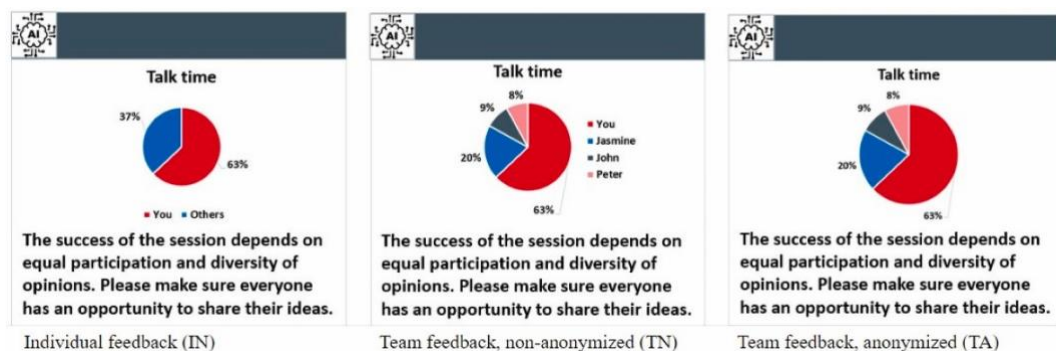


Figure 2. Vignettes for different experimental conditions for the Information type Talk time.

After collecting informed consent and demographic data, participants were shown a general scenario description prior to introducing the specific vignettes: “Imagine participating in a creative workshop at work. Your and your colleagues are brainstorming ideas to solve a business problem. Such workshops are usually moderated by a facilitator. In this case, an Artificial Intelligence (AI)-based agent partially takes over the facilitator's tasks by assessing team interaction dynamics and providing guidance accordingly. (In some scenarios, names of you or your teammates are going to be displayed. Try to put yourself in a role of one of the mentioned persons.)

*Information you provide to the AI-based agent is only used for the purpose of facilitating the imagined creative workshop, meaning no third-party would have access to this data beyond the imagined company you work in. This is ensured by using decentralized data storage methods.” The text in parentheses is only shown in team conditions. Upon this, vignettes for each Information Type were presented followed by the measurements of Perceived Risks and Benefits and Intention to disclose that particular Information type. After collecting these measures for each Information Type separately, Trust in AI was measured once at the end of the questionnaire.

18.4.1 Vignette development and validation

AI systems with our features of interest are not yet fully established, i.e., while certain features might be part of existing tools, these are not yet integrated into online meeting tools (perhaps for the very reasons we are aiming to explore within this study). Thus, we opted for a vignette-based experiment. This is a proven effective and economical method to assess behaviours, attitudes and intentions. Furthermore, it is a suitable method for exploring sensitive topics (Aguinis & Bradley, 2014; Karren & Barringer, 2002), such as ID, and it has been reliably used in a similar context of ID to robots in healthcare (Jayaraman et al., 2024), and the acceptance of AI-based emotion analytics (Behn et al., 2024). For the AI agent, we opted out for a simple icon, to control for the potential influence that comes along with modern anthropomorphic AI systems. For example, it has been shown that such systems exert influence on Trust in AI (Li & Suh, 2022; Uysal et al., 2023)

Phase one: Vignette Development During phase one, the initial development of the vignettes occurred. This is a product of a comprehensive review of possible application areas of AI in facilitation of teamwork and content expertise of the principal investigator. Combining these resulted in vignettes that captured various situations from teamwork moderation. The initial vignettes were reviewed by the team members.

Phase two: Validation and Revisions To ensure the validity of our scenarios, as per St Marie, Jimmerson, Perkhounkova, and Herr (2021), we perform two rounds of adaptations based on expert feedback. For this, we used a mixed-method approach: in the first round, we sent a survey to 8 experts and asked them to rate the relevance and clarity of the vignettes to the domain on a 4-point Likert scale (not relevant to highly relevant to the domain; unclear to very clear, respectively). We collected answers from experts in different relevant domains, such as Information Systems professors and researchers with research focus on AI, digital innovation and transformation, and people analytics, respectively (n = 4), design thinking facilitators and trainers (n = 2), and team leaders from the industry who use AI at workplace (n = 2). In addition, we collected qualitative feedback in form of improvement suggestions. We calculated the Content Validation Index (CVI), the most common quantitative method used to assess content validity (Rodrigues et al., 2017) for each vignette as described in (Yusoff, 2019): where CVI for an item (I-CVI) is calculated as the proportion of content experts giving item a relevance rating of 3 or 4. Depending on

the number of experts, there are different cut-off values for what is considered to be an acceptable CVI. For instance, for six to eight experts, acceptable CVI should be over 0.83 (Yusoff, 2019). Except for one vignette that passed the cut-off, all other necessitated a revision and re-validation. Most of the feedback was focused on the lack of detail, specifically with regards to what type of data AI would be using to provide its feedback.

Phase three: Re-Validation As the first round of validations resulted in necessary adaption of the most of the vignettes, we incorporated qualitative feedback to improve the clarity and relevance of the vignettes to the domain. While the first round consisted of experts in different aspects of the domain (researchers, industry team leaders, and facilitators from both academia and industry), and provided us with essential multifaceted perspective, some opinions were contradictory and posed a challenge for the adaptation. Thus, we opted for another method to re-validate our vignettes; what was impossible to incorporate in a survey that each expert is filling out alone, was now successfully reached in an expert panel group discussion. This is, a consensus. A review of articles in Psychological Assessment reveals the potential of one method of consultation, focus groups, to inform multiple stages of instrument development (Vogt et al., 2004). The authors suggest that using focus groups has the potential to improve the content validity of instruments and, consequently, enhance the validity of research outcomes. We performed the focus group with a panel of three experts, who also participated in the first validation round. The panel consisted of three information systems professors with expertise in AI, digital innovation, and people analytics. In addition, one of the participants also has experience as a design thinking facilitator. The focus group was guided by a semi-structured guideline, aiming to assess the clarity and relevance of the improved vignettes. Each vignette was discussed until consensus was reached. The focus group lasted approximately 1 h and was moderated by two authors of this manuscript.

Phase four: Final validation In a pilot study performed to assess the validity of the self-developed scales for Perceived Risks and Benefits (see 4.2.1), we asked 113 participants to also leave qualitative feedback on the clarity of vignettes (we did not assess relevance again, as the participants were not necessarily experts in the domain, however it was of utmost importance that they understood the vignettes). Since answering this question was optional, 38 participants provided qualitative feedback. The feedback was mostly positive, stating that everything was clear. A few participants mentioned that the choice of some words confused them, stating the fact that English is not their first language as a possible cause. Thus, we decided to add an additional screener for participants who are fluent in English for the main study.

The example of the final vignettes is shown in Fig. 2. The rest of the vignettes can be found in the Appendix 7.1.

18.4.2 Instruments

To ensure solid measurement of our research model shown in Fig. 1 we carefully reviewed scales for our most important constructs. A recently developed trust scale (Hoffman et al., 2023) has shown high validity and reliability in AI context. However, it was measured in a different task context, where there was a proper interaction with AI and utilization of its advice, which did not suit our vignette design. Thus, we opted for a scale by Merritt, that measures propensity to trust (Merritt, 2011) as recommended by Razin & Feigh (2023), who run a meta-analysis on Human-AI Trust questionnaires and compared several frequently used measures on the aspects of reliability and validity, among others.

Moreover, we measured several control variables: AI Literacy - hereby we opted for a recently developed objective knowledge test of AI (Weber et al., 2023), so as to avoid the pitfalls of self-assessments, which in the realms of cognitive abilities and knowledge often probably display the Danning-Krueger syndrome (Weber et al., 2023).

Risk aversion was measured with two alternatives Iowa gambling task (Bechara et al., 1994).

Teamwork Attitudes are taken from Mendo-Lázaro et al. (2017). The full scale consists of two factors and twelve items, not all of which are relevant to our context. Thus we used the items concerning the social and affective factor, and left out the academic factor.

Finally, the dependent variable, Intention to disclose particular information, was measured on a 5-point Likert scale, with a question that differs depending on the data type used for particular feedback: “Given what you learned about how the AI-based agent uses [x: microphone or microphone + video] data to give feedback, how likely is it that you would have consented to the disclosure of your [x] data for such a purpose?”

As there are no validated scales for Perceived Risks and Benefits for this specific domain, we developed the scales consisting of 5–6 items each (for the items, see supplementary material). The scales were (slightly) different for each Information type and Recipient. We collected initial feedback on the items during the first vignette validation phase, which was incorporated. Upon that we validated the newly constructed items in a pilot study, described below.

18.4.2.1 Validation of the self-developed scales

In this section, we describe the methodology employed to assess the unidimensional reliability of items for the self-developed Perceived Benefits and Perceived Risks scale, followed by a Confirmatory Factor Analysis (CFA). First, during the first round of the vignette validation, we also collected initial feedback on the clarity and relevance of these items. Additionally, we gathered qualitative feedback for improvements. Upon adjusting the items based on expert feedback, we performed a final validation in a pilot study, which initially involved a sample of 90 participants (30 per group). This data was collected in September 2023, on the crowd-working platform Prolific (Palan & Schitter, 2018). Since

team level groups (TA, TN) had the same items on these scales and accordingly more participants, we additionally collected 23 more answers for the IN group, to ensure comparability of results, as well as stronger power for the initial pilot analysis. The total sample of the pilot study thus counted 113 participants. The analysis was performed in the statistical analysis software, JASP (JASP Team, 2023).

The reliability analysis was performed to evaluate the internal consistency and reliability of the items within each Information type. We utilized two commonly employed reliability indices, McDonald's Omega and Cronbach's Alpha (Malkewitz et al., 2023), to assess the consistency of our measurement items. The results of the analysis consistently yielded middle to high internal consistency values, with all indices exceeding the threshold of 0.7. The analysis suggested removing one item for the Perceived Benefits scale for the Information type 'Bias toward certain solutions' in the IN group, however the indices for the dropped item showed only slightly higher values. Therefore, we decided to keep all items.

Subsequently, our CFA revealed varying factor solutions for different Information types. While some constructs exhibited good model fit with a single-factor or a two-factor solution, others displayed acceptable fit with both single-factor and two-factor solutions. In case of Stagnation on the Team level, the two-factor solution was notably superior. These findings underscore the importance of considering the nuanced structure of Perceived Benefits and Risks in different contexts.

18.4.3 Main analysis

18.4.3.1 Participants

Data collection took place September–October 2023 via the crowd-working platform Prolific (Palan & Schitter, 2018). For their participation, crowdworkers were reimbursed £9.48 per hour; median time to complete the survey was 16.84 min. Participants gave informed consent to participation prior to commencing the survey.

To ensure the quality and relevance of our data, we implemented strict prescreening criteria. Participants were required to have an approval rate exceeding 95 % on Prolific. Moreover, we specifically targeted individuals who reported using technology at their workplace at least four times a week. Lastly, proficiency in the English language was a key prerequisite for participation. During the data collection process, we identified and removed responses that failed attention checks. Statements such as “Please select Strongly Agree here” were incorporated in several instruments. These checks were implemented to ensure the accuracy and reliability of the data. Approximately 5 percent of the initial dataset was excluded based on this criterion. Subsequently, we recollected the remaining data to maintain a robust and dependable dataset. We calculated the necessary sample size for a structural equation model via an online calculator,² with the following parameters: Expected effect size - 0.3; p-value - 0.05; statistical power - 0.8. This resulted in the minimum sample size of 200 to detect effect, and minimum sample size of 116 for the

model structure. As we intend to run the model separately for each group, the final sample size was 600. After excluding the outliers, the final dataset used for our study comprised a total of 582 participants. The ages of these participants ranged from 18 to 75, mean = 29.8, SD = 9.24, ensuring a diverse age demographic. In terms of gender distribution, we opted for a balanced sample, which in the end consisted of 48,79 % male, 49,65 % female respondents, and 1,54 % Other/prefer not to disclose.

18.4.3.2 Data analysis

This section outlines the methodology employed on the main sample, which aims to explore the relationships among variables and investigate the mediation effects within the context of intention to disclose information. The following analyses are performed using IBM Amos (Arbuckle, 2019), and JASP (JASP Team, 2023).

To address all hypotheses and avoid an over-saturated model, data is analyzed using several approaches. First is Analysis of Variance (ANOVA), one-way, and with interactions, and with corresponding post-hoc analyses, for the hypotheses that were not necessary to include in the SEM model (H1a-b). ANOVA is instrumental in addressing specific research questions and hypotheses related to different Information types. For instance, it assists us in testing Hypothesis H1 which investigates how different Information types impact the Intention to disclose. In addition, we use it to analyze whether these effects differ when AI feedback is received individually versus in a team setting (H1a). Moreover, it allows us to explore whether the presentation of AI feedback as anonymized or non-anonymized in team settings affects the Intention to disclose (H1b). Further, ANOVA is used in a robustness test, to check for effects of data type (microphone or video + microphone) on intention to disclose.

The part in the blue box of our theoretical model is analyzed via a path analysis approach. Path analysis is a statistical approach from the family of SEM approaches, used to describe the directed dependencies among a set of variables. In path analysis, the focus is on identifying causal relationships between variables. It is often used to test theoretical propositions about cause and effect without manipulating variables, allowing researchers to examine the relationships among variables and assess the effects of a set of variables acting on a specified outcome. Path analysis is based on a closed system of nested relationships among variables, represented statistically by a series of structured linear regression equations. The method is used to discern and assess the effects of variables on a specified outcome, and it is closely related to multiple regression, with regression being considered a special case of path analysis (Lleras, 2005).

This analytical framework allows us to examine the mediating role of Perceived Risks and Perceived Benefits in the relationship between Trust in AI and Intention to disclose information. This approach is particularly relevant to Hypotheses 2a-d. In this context, it helps us understand how Perceived Risks and Benefits influence willingness to disclose information (H2c, H2d); how Trust in AI influences Intention to disclose (H2), by

examining whether Trust influences Perceived Risks (H2a) and Perceived Benefits (H2b) as mediators in the Intention to disclose. As a part of robustness tests, we also tests models with control variables, to account for potential confounding effects. Moreover, we test models with different structures (simplified and with defined covariance).

18.5 Results

In this section, we present the results of our analyses and structure them along our proposed hypotheses. For a high-level overview, we start with descriptive statistics of our main variables of interest (Table 2).

Table 2. Descriptive statistics for Perceived Benefits and Perceived Risks.

Information type	Group	Perceived Benefits				Perceived Risks			
		Mean	Upper	Lower	SD	Mean	Upper	Lower	SD
Talk times	IN	3.679	3.764	3.595	0.601	2.978	3.092	2.864	0.811
	TA	3.761	3.853	3.669	0.666	3.170	3.276	3.064	0.764
	TN	3.772	3.881	3.663	0.742	3.195	3.307	3.083	0.782
Interruptions	IN	3.640	3.740	3.541	0.706	3.047	3.163	2.932	0.819
	TA	3.759	3.858	3.660	0.711	3.112	3.226	2.998	0.821
	TN	3.809	3.915	3.702	0.742	3.159	3.271	3.046	0.790
Attention level	IN	3.635	3.758	3.512	0.877	3.322	3.452	3.191	0.928
	TA	3.632	3.750	3.514	0.854	3.280	3.398	3.162	0.854
	TN	3.528	3.655	3.400	0.891	3.372	3.502	3.243	0.907
Breaking rules	IN	3.225	3.363	3.087	0.980	3.324	3.447	3.200	0.878
	TA	3.434	3.565	3.303	0.946	3.158	3.279	3.037	0.876
	TN	3.281	3.427	3.135	1.020	3.369	3.488	3.251	0.828
Bias	IN	3.671	3.777	3.564	0.757	3.059	3.178	2.940	0.845
	TA	3.571	3.676	3.465	0.761	3.135	3.258	3.012	0.889
	TN	3.561	3.681	3.441	0.837	3.248	3.377	3.118	0.906
Moodboard	IN	3.043	3.188	2.899	1.025	3.196	3.334	3.058	0.978
	TA	3.115	3.241	2.989	0.908	3.245	3.379	3.111	0.969
	TN	3.154	3.298	3.010	1.007	3.247	3.386	3.108	0.971
Stagnation	IN	3.293	3.435	3.151	1.008	2.942	3.072	2.813	0.918
	TA	3.472	3.606	3.338	0.966	2.970	3.091	2.849	0.871

		Perceived Benefits				Perceived Risks			
Information type	Group	Mean	Upper	Lower	SD	Mean	Upper	Lower	SD
	TN	3.518	3.657	3.379	0.970	3.015	3.141	2.888	0.885

Note: Upper and Lower refer to 95% Confidence Intervals for Mean.

Trust in AI was measured once per survey and not per each Information type separately.

Table 3 depicts descriptive statistics for Trust per Recipient/Anonymity group (see Table 3).

Table 3. Descriptive statistics for Trust in AI.

		95 % Confidence Interval Mean		
Group	Mean	Upper	Lower	SD
IN	3.279	3.324	3.235	0.836
TA	3.248	3.292	3.204	0.838
TN	3.295	3.341	3.250	0.845

Table 4. Descriptive statistics for Intention to disclose

			95 % Confidence Interval Mean		
Information type	Group	Mean	Upper	Lower	SD
Talk times	IN	3.206	3.364	3.049	1.119
	TA	3.105	3.256	2.954	1.091
	TN	2.995	3.165	2.824	1.195
Interruptions	IN	3.119	3.282	2.955	1.161
	TA	3.255	3.409	3.101	1.112
	TN	3.170	3.329	3.011	1.110
Attention level	IN	2.706	2.890	2.522	1.308
	TA	2.765	2.940	2.590	1.264

			95 % Confidence Interval Mean		
Information type	Group	Mean	Upper	Lower	SD
	TN	2.644	2.832	2.456	1.315
Breaking rules	IN	2.737	2.905	2.569	1.195
	TA	2.935	3.104	2.766	1.220
	TN	2.761	2.927	2.594	1.166
Bias	IN	3.077	3.243	2.912	1.178
	TA	3.095	3.259	2.931	1.180
	TN	2.957	3.124	2.791	1.165
Moodboard	IN	2.582	2.760	2.405	1.262
	TA	2.595	2.769	2.421	1.257
	TN	2.638	2.824	2.453	1.299
Stagnation	IN	2.985	3.155	2.814	1.211
	TA	3.060	3.219	2.901	1.150
	TN	3.090	3.259	2.922	1.178

Descriptive statistics for the dependent variable (Intention to disclose) are presented both in the Table 4, as well as visualized in boxplots (Fig. 3), for an easier comparison of differences in Intention to disclose per Information type.

Following the descriptive overview of the main variables, we continue to elaborate on each analyzed hypothesis. First three hypotheses (H1a-b) concern the relationship between our independent variables (Information type, Recipient, Anonymity of the feedback), and the dependent variable (Intention to disclose information). Hypotheses 2a-d zoom into the privacy calculus (blue box) of our theoretical model (Fig. 1).

H1. Different Information types yield different Intention to disclose that particular information. The variance of Intention to disclose is significantly different across Information types (Levene's test of equality of variances: $p < 0.05$). Hence, we employed one-way ANOVA with Brown-Forsythe homogeneity correction and the Games-Howell post-hoc test.

The analysis revealed an effect of Information type on Intention to disclose, as indicated by the F-statistic of 19.561 (df = 6, $p < 0.001$), with an effect size (η^2) of 0.028. Refer to Table 5 for a detailed breakdown of post-hoc comparisons between individual Information types.

Table 5. Games-Howell Post Hoc comparison.

		95 % CL for Mean Difference				
Comparison	Mean Difference	Lower	Upper	SE	t	ptukey
Talk times - Interruptions	-0.079	-0.275	0.117	0.066	-1.191	0.898
Talk times - Attention level	0.397	0.186	0.608	0.071	5.561	<.001***
Talk times - Breaking rules	0.290	0.088	0.492	0.068	4.247	<.001***
Talk times - Bias	0.058	-0.142	0.258	0.068	0.863	0.978
Talk times - Moodboard	0.498	0.290	0.707	0.071	7.055	<.001***
Talk times - Stagnation	0.058	-0.142	0.259	0.068	0.861	0.978
Interruptions - Attention level	0.476	0.266	0.686	0.071	6.690	<.001***
Interruptions - Breaking rules	0.369	0.168	0.571	0.068	5.423	<.001***
Interruptions - Bias	0.137	-0.062	0.337	0.067	2.037	0.392
Interruptions - Moodboard	0.577	0.369	0.785	0.070	8.201	<.001***
Interruptions - Stagnation	0.137	-0.062	0.337	0.068	2.033	0.394
Attention level - Breaking rules	-0.107	-0.322	0.109	0.073	-1.459	0.769

		95 % CL for Mean Difference				
Comparison	Mean Difference	Lower	Upper	SE	t	ptukey
Attention level - Bias	-0.338	-0.552	-0.125	0.072	-4.674	<.001***
Attention level - Moodboard	0.101	-0.121	0.323	0.075	1.349	0.829
Attention level - Stagnation	-0.338	-0.553	-0.124	0.073	-4.666	<.001***
Breaking rules - Bias	-0.232	-0.437	-0.027	0.069	-3.339	0.015*
Breaking rules - Moodboard	0.208	-0.006	0.421	0.072	2.875	0.062
Breaking rules - Stagnation	-0.232	-0.437	-0.026	0.070	-3.333	0.015*
Bias - Moodboard	0.440	0.228	0.652	0.072	6.135	<.001***
Bias - Stagnation	0.000	-0.204	0.204	0.069	0.000	1.000
Moodboard - Stagnation	-0.440	-0.652	-0.228	0.072	-6.125	<.001***

$p < 0.05$, *** $p < 0.001$.

Furthermore, Fig. 3 displays zoomed in values on the Intention to disclose variable for each Information type (full scale is 1–5) across the 3 conditions (IN, TA, TN). The figure depicts that study participants are less likely to disclose their microphone and video data in order for AI to assess their emotional state, attention levels and if rules of teamwork are broken. Regarding the data that allows AI to intervene in cases of interruptions, stagnation, biases, and uneven talk times, participants display neutral toward likely Intention to disclose this data.

H1a. There is an interaction effect of Information type and Recipient on the Intention to disclose. ANOVA with interaction between the Recipient and Information type did not reveal any statistically significant differences, i.e., no differences in Intention to disclose

different Information types when AI feedback is presented individually vs. in team ($F = 0.828$, $df = 6$, $p = 0.548$).

H1b. There is an interaction effect of Information type and Anonymity of the feedback on the Intention to disclose. ANOVA with interaction between Anonymity of feedback and Information type did not show any statistically significant influence of Anonymity of Feedback on the Intention to disclose for different Information types ($F = 0.478$, $df = 6$, $p = 0.825$).

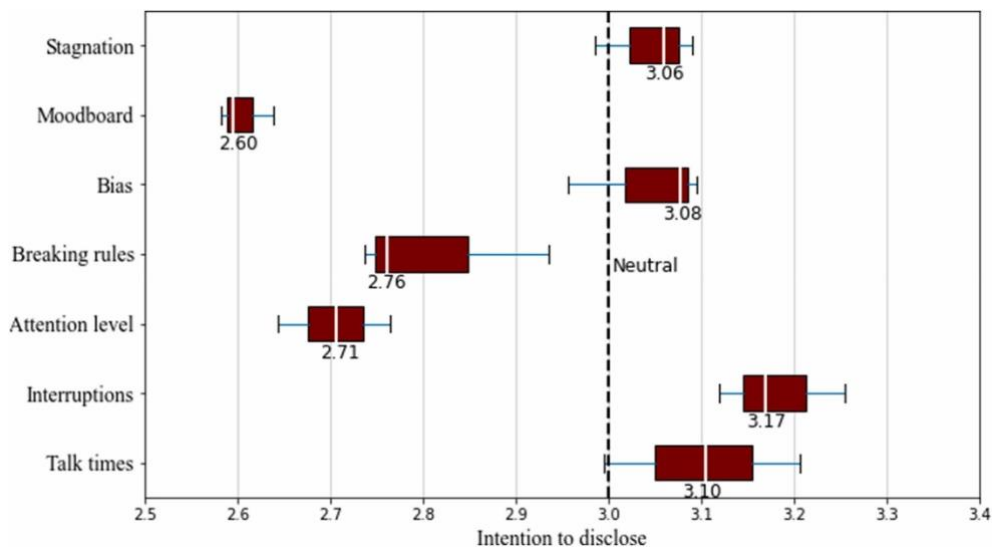


Figure 3. Boxplots of Intention to disclose per Information type.

Thus, we do not find sufficient support for H1a, H1b, and cannot conclude that Intention to disclose differs when AI feedback is presented in team or in private; and when it is presented anonymous or non-anonymous when presented in team.

Hypotheses 2a-d are analyzed within the SEM model; in the following we present an overview of the path model as well as results of individual path measures.

First we examine the influence of control variables by including them in the main model (Fig. 1). Notably, we find relationships between the main model variables (Trust in AI, Perceived Risks, Perceived Benefits, Intention to Disclose) for all Information types ($p < 0.001$). Additionally, Teamwork attitudes predict Trust in AI, while Teamwork attitudes, Gender, Age, AI literacy, and Risk aversion predict Perceived Benefits. Age and AI literacy predict Perceived Risks, and Teamwork attitudes, Gender, and Age predict Intention to Disclose. However, these effects are not consistently statistically significant across all Information types. Therefore, we removed the variance of these effects from our main model variables and conducted the path analysis with updated variables with unstandardized residuals. Moreover, as ANOVA results indicated no significant difference between different groups (IN, TA, TN), we ran the model on the full sample of 582 participants, rather comparing the three groups.

The path model was analyzed in AMOS 29. Modification indices from the initial run suggested correlating the residuals of Perceived Risks and Perceived Benefits. However,

this resulted in an over-fitted model (NFI, IFI, CFI = 1), and a big error (RMSEA = .164). After removing these covariances, the model achieved good fit ($\chi^2 = 219,646$, $df = 7$, SRMR = .034, RMSEA = .086, NFI = .953, IFI = .954, CFI = .954).

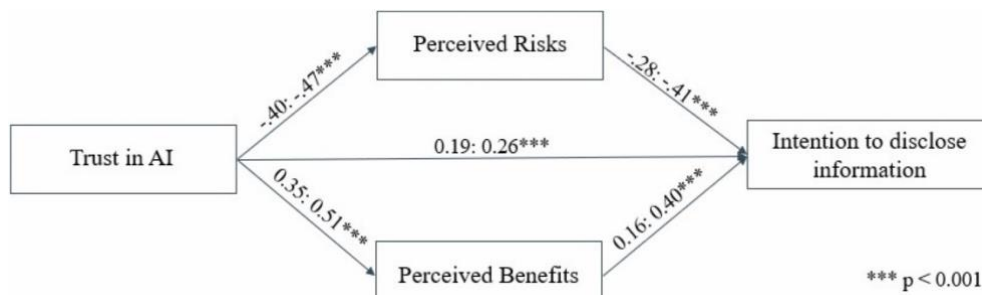


Figure 4. Path model Results.

Fig. 4 illustrates the size and direction of the effects in the model. The regression weights were calculated separately for each Information type and are presented as a range, from the lowest to the highest values for the seven Information types. All effects are statistically significant at $p < 0.001$ and exhibit the same direction for every Information type. Table 6, Table 7 provide further details on regression estimates for each Information type.

Table 6. Standardized Regression Weights for paths in the model.

Estimate, 95 % CI	Talk times	Interruptions	Attention level	Breaking rules
Risks ← Trust in AI	-.40 (-.47, -.32)	-.41 (-.48, -.33)	-.45 (-.52, -.38)	-.47 (-.53, -.39)
Benefits ← Trust in AI	.35 (.28, .43)	.37 (.30, .45)	.47 (.39, .54)	.47 (.39, .53)
Disclose ← Trust in AI	.26 (.18, .33)	.23 (.16, .31)	.19 (.12, .26)	.23 (.16, .30)
Disclose ← Benefits	.16 (.08, .24)	.24 (.16, .32)	.28 (.19, .35)	.36 (.28, .43)
Disclose ← Risks	-.38 (-.44, -.30)	-.36 (-.43, -.28)	-.41 (-.49, -.34)	-.36 (-.45, -.28)

Table 7. Table 6 continued.

Estimate, 95 % CI	Bias	Moodboard	Stagnation
Risks ← Trust in AI	-.45 (-.52, -.38)	-.46 (-.52, -.39)	-.47 (-.54, -.40)
Benefits ← Trust in AI	.45 (.37, .52)	.48 (.40, .54)	.50 (.43, .57)
Disclose ← Trust in AI	.22 (.15, .30)	.20 (.12, .27)	.21 (.12, .29)
Disclose ← Benefits	.31 (.24, .38)	.31 (.23, .39)	.40 (.31, .48)
Disclose ← Risks	-.39 (-.47, -.32)	-.39 (-.47, -.32)	-.28 (-.36, -.20)

Trust in AI consistently shows a moderate to strong inverse association with Perceived Risks across all types of information, with standardized effects ranging from -0.40 to -0.47 . Similarly, Trust in AI demonstrates a positive association with Perceived Benefits, with effects varying from 0.35 to 0.50 across different Information types. Additionally, Trust in AI displays a smaller yet consistent association with the Intention to Disclose, with effects ranging from 0.19 to 0.26 , indicating a more nuanced connection. The relationships between Perceived Benefits and the Intention to Disclose show variable effects from 0.16 to 0.40 , indicating a small to moderate positive effect of Perceived Benefits on the Intention to disclose. The influence is stronger for Information types like Stagnation and Breaking rules. In contrast, the association between Perceived Risks and the Intention to Disclose ranges from -0.28 to -0.41 , reflecting a moderate negative link across different types of information.

To assess the mediation effects of Perceived Risks and Benefits, we first look into the indirect effects analysis. Upon confirming that Trust in AI exerts indirect effect on Intention to disclose across all Information types ($p = 0.001$), we run an additional analysis to uncover via which variables and to what extent the mediation takes place. As presented in Table 8, Table 9, both Perceived Risks and Perceived Benefits mediate the relationship between Trust in AI and Intention to disclose. Except in the case of Stagnation, the mediation effects of Perceived Risks are consistently stronger. All effects are statistically significant ($p = 0.001$).

Furthermore, since Trust in AI has both direct (Table 6) and indirect effect (Table 8) on Intention to disclose, the mediation effect of Perceived Risks and Benefits is partial. In the case of Talk times, the proportion of the total effect mediated, calculated as the ratio of the indirect effect to the total effect, is $44,3\%$; for Interruptions it is $50,8\%$; for Attention level

62,3 %; for Breaking rules of teamwork 59,6 %; for Bias toward certain solutions 58,4 %; for Moodboard 62,6 %; and for Stagnation 61,2 %.

Table 8. Mediation effects of Perceived Benefits and Risks.

Estimate, 95 % CI	Talk times	Interruptions	Attention level	Rules
Trust in AI → Risk → Disclose	.21 (.15, .26)	.20 (.15, .26)	.28 (.22, .36)	.24 (.18, .31)
Trust in AI → Benefit → Disclose	.08 (.04, .12)	.12 (.08, .17)	.20 (.14, .26)	.23 (.17, .30)

Table 9. Table 8 continued.

Estimate, 95 % CI	Bias	Moodboard	Stagnation
Trust in AI → Risk → Disclose	.25 (1.9, .32)	.27 (2.1, .35)	.18 (1.3, .24)
Trust in AI → Benefit → Disclose	.19 (.15, .25)	.22 .16, .30	.28 (.21, .36)

In conclusion, Hypotheses 2a-d are supported. Trust in AI predicts Intention to disclose. This effect is visible directly as well as via mediation of Perceived Risks and Perceived Benefits. Lower Trust in AI predicts higher Perceived Risks and lower Perceived Benefits, which, in turn, predict lower Intention to disclose. Conversely, higher Trust in AI predicts lower Perceived Risks and higher Perceived Benefits, which subsequently predict higher Intention to disclose.

18.5.1 Robustness analyses

To evaluate the explanatory power of our core model, we tested two alternative model specifications. The first alternative model expanded the core model by incorporating Teamwork Attitudes as an additional predictor alongside Trust in AI. This allowed us to assess whether Teamwork Attitudes contribute meaningfully to explaining disclosure intentions beyond the established privacy calculus variables. The second alternative model excluded Trust in AI, enabling us to examine its necessity in explaining information disclosure.

The expanded model including Teamwork Attitudes exhibited adequate fit indices ($\chi^2 = 366.268$, $df = 14$, $SRMR = .067$, $RMSEA = .06$, $NFI = .927$, $IFI = .929$, $CFI = .929$). However, Teamwork Attitudes did not consistently predict key variables. While there was some influence on Perceived Benefits, its effect on Perceived Risks was non-significant across all Information Types, and it did not significantly predict Intention to Disclose for six out of seven Information Types. This suggests that Teamwork Attitudes add complexity

without substantially improving explanatory power, supporting the parsimonious model as the more robust option.

The model excluding Trust in AI resulted in a worse model fit ($\chi^2 = 225.347$, $df = 7$, SRMR = .059, RMSEA = .088, NFI = .862, IFI = .865, CFI = .864). This indicates that Trust in AI is a critical component of the model, reinforcing its importance in understanding information disclosure decisions.

Beyond these model comparisons, we conducted several robustness tests to ensure the stability and reliability of our findings. First, we examined the correlation matrices for the SEM model, confirming that all correlations were significant and within acceptable ranges (see 7.2). Additionally, we analyzed differences in disclosure intentions under varying conditions. Specifically, we investigated the difference in intention to disclose data when only the microphone is involved versus both microphone and video. Our findings indicate a lower intention to disclose both microphone and video data compared to solely microphone data (2.65 vs. 3.03, $p < 0.001$, rank biserial correlation = 0.175).

18.6 Discussion

18.6.1 Contributions

The presented study sought to explore the factors influencing individuals' intentions to disclose various types of information to AI, specifically focusing on the effects of different Information types, Recipient settings, and degree of Anonymity of the feedback, while also examining the roles of Trust in AI, Perceived Risks, and Perceived Benefits. To this end, we posed three research questions:

RQ 1: Which information types are humans willing to disclose so that AI-based systems can support them in virtual meetings? In order to investigate the effect of AI benefits to ID, we posed the RQ 2: How should the output from AI-based systems be designed with respect to recipient and anonymity to foster ID in in virtual meetings? Furthermore, in a try to capture the context- and personality-dependent complexity of ID, we posed a further RQ 3: How does Trust in AI affect the Intention to disclose information? The following discussion delves into the implications and interpretations of our findings with regards to the research questions.

Our analysis revealed differences in Intention to disclose based on Information type. The data indicated that participants exhibited varied inclinations to share different types of information with AI systems. The effect size ($\eta^2 = 0.028$) observed in our analysis, while statistically significant, indicates a relatively small practical impact of Information type on disclosure intentions. This suggests that, while the type of information is a relevant factor, it explains only a small portion of the variance in individuals' willingness to disclose information to AI in the current experimental setting. This nuanced finding emphasizes the complexity of the decision-making process surrounding information disclosure. It

underscores that other factors play a more substantial role in influencing disclosure intentions than the information type alone. This interpretation aligns with the notion that information disclosure, especially in the context of human-AI interaction, is a multifaceted phenomenon, influenced by a myriad of factors (Masaviru, 2016; Xu et al., 2008) beyond the mere categorization of information.

Nevertheless, our results indicate that participants were less likely to disclose their microphone and video data when AI aimed to assess their emotional state, attention levels, and teamwork rule violations. In contrast, they demonstrated a more neutral to likely Intention to disclose data that allowed AI to intervene in cases of interruptions, stagnation, biases, and uneven talk times. One possible explanation for this result could be that certain information types are considered as more sensitive, leading to heightened privacy concerns and lower disclosure intentions. In general, none of the Information types deployed in our study reached a mean value higher than 4 on Intention to disclose, suggesting a cautious stance toward sharing data with AI systems across all information categories.

This result would be in accordance with prior research, that has shown that people attribute varying levels of sensitivity to the data they are asked to disclose. Depending on the degree of sensitivity that is attributed to the data, the context, and fairness of the request emerge as key factors affecting their comfort with the decision to disclose (Belen-Saglam et al., 2022). Our results suggest that data indicating the state of mood or attention as well as rule violations could be considered as more sensitive and possibly inadequate to request in the described scenario of collaborative problem-solving in a professional context. Similarly, this results would underscore prior research that has indicated different willingness to disclose depending on the type of information (Degutis et al., 2020; Knijnenburg et al., 2013). Beyond this, these findings would underline what recent research on emotional surveillance in the workplace based on AI shows: Mantello et al. (2023) state that emotion-sensing tools, left unregulated in the workplace, may lead to heightened stress and anxiety. Similarly, our results suggest that especially the AI-based detection and processing of emotion-based and intimate data is causing major concerns among participants.

This also means that implementing AI as an unregulated facilitator might not be a sensible option for any situation and accordingly required information type. Rather, the hybrid intelligence approach proposed by Dellermann et al. (2019) should be utilized, especially for information types that are less likely to be disclosed to AI. Here a human facilitator should step in in terms of facilitating the related interactions. Similarly, Bostrom et al. (1993), who developed a framework for group facilitation, describe that there are different ways and extents to which technology can be involved in facilitating group meeting processes, emphasizing the advantageous aspects they provide, while at the same time acknowledging the uniquely human skills in the process of meeting facilitation. Their discussions suggests that—though technology can play an important and helpful role in meeting facilitation—a human facilitator is still required, e.g., to ensure proper functioning of the technology or to add supplemental structure or support in a particular situation. Our

research adds to this findings by indicating some situations in which this combined facilitation approach shows itself as particularly important, depending on the type of information drawn on. Furthermore, our results strengthen Teodorescu et al. (2021), who argue that the way AI augments human decision-making is to be understood on the basis of the fairness challenges different augmentation constellations produce in a specific context and should be designed accordingly. For the context presented in this paper, our results suggest that a human facilitator should take a leading role depending on the information asked from the participants. Especially in the case of intimate or emotion-based data, automating the process facilitation based on AI does not seem to be a promising option.

However, we need to keep in mind that our study also indicates limited practical relevance of the type of information for the Intention to disclose, what would be in contrast to what prior research has shown.

Next, in our analysis we found no effect between recipient and anonymity of the feedback and the Intention to disclose certain information types. In other words, participant's answer reveal that “no matter how or to whom you present it, I (don't) want to disclose this information”. This finding emphasizes the difference in meaning between different information types and suggest that, in the case of sensitive information, participants simply do not appreciate AI monitoring.

These findings are in contrast with Belen-Saglam et al. (2022), who show that—depending on the type of information—the degree of anonymity plays a central role for people in order to feel comfortable with their disclosure decision. Furthermore, recent studies show that the critical element of privacy in data sharing is the protection of the identity, and thus, anonymity (Schomakers et al., 2019). Our findings, in contrast, do not support the hypothesis that the level of anonymity plays a role for the Intention to disclose. The same holds true for the recipient of the AI feedback. Similarly to prior research (Belen-Saglam et al., 2022; Cooks-Campbell, 2023; Schomakers et al., 2019), our results also show no clear tendency with regards to disclosure intentions in teams vs. in individual-level setting.

Concerning our second category of hypotheses, centered around the concept of Trust in AI influencing the Intention to disclose, our results show that there is a relationship between all main variables of our model (Trust in AI, Perceived Risks, Perceived Benefits, Intention to disclose) across all Information types.

Specifically, Trusts in AI predicts the Intention to disclose directly, but also through Perceived Risks and Benefits as mediators. The strongest effects are observed in the relationships between Trust in AI and Perceived Risks/Benefits, indicating that trust plays a significant role in shaping perceptions of risks and benefits. The direct effects of Trust in AI on Intention to Disclose are smaller but consistent across all information types. The relationship between Perceived Benefits/Risks and Intention to Disclose shows a moderate effect, varying slightly across different types of information. This is in accordance with

studies about self-disclosure to intelligent agents (Al-Natour et al., 2021; Li & Unger, 2012) and the role of Trust in AI for AI adoption and use (Choung et al., 2023; Solberg et al., 2022). Our results add on this by showing that Trust in AI positively impacts an individuals' Intentions to disclose to an AI-based facilitation system. Furthermore, this finding is in accordance with research adopting the theoretical concept of privacy calculus (Dinev et al., 2013; Yu et al., 2020). For instance, Masaviru (2016) shows that trust plays a pivotal role in individuals' willingness to disclose information. Similarly, Steel (1991) shows that varying levels of trust lead to varying levels of self-disclosure. Our findings are also consistent with results of studies from different contexts: For instance, research in the context of social commerce has shown that trust toward members was found to reduce perceived risks related to both participating and purchasing behaviors (Farivar et al., 2016); in the context of gene technology, research suggest that trust in institutions using gene technology or modified products has a positive impact on perceived benefits and a negative influence on perceived risks of this technology. Furthermore, perceived benefits and perceived risks have been found to determine acceptance of biotechnology. In other words, trust has an indirect influence on the acceptance of the technology (Siegrist, 2000); similarly, in the digital application context, perceived benefits were found to have a positive relationship with trust. The level of one's trust in an application will have an impact on the intention to use the application (Sari, 2023); the same also holds true for the context of AI adoption in healthcare, where Choudhury et al. (2022) show the relation between clinicians' trust in the AI and perceived risks and benefits. The more clinicians trusted the AI-based system, the higher was in turn also their intention to use it.

Beyond that, it is noteworthy that some of our control variables also exhibited effects on certain main variables for certain Information types. For instance, Teamwork attitudes predicted Trust in AI, while Teamwork attitudes, Gender, Age, AI literacy, and Risk aversion predicted Perceived Benefits. Age and AI literacy predicted Perceived Risks, and Teamwork attitudes, Gender, and Age affected Intention to Disclose. This confirms that information disclosure has a complex nature, which is highly dependent on the context as well as the individual situation and predisposition (Hirschprung et al., 2016; Masaviru, 2016; Xu et al., 2008). To further assess the robustness of our findings, we tested an extended model that included Teamwork Attitudes alongside Trust in AI. However, this addition did not substantially improve explanatory power, reinforcing the original parsimonious model as a more stable representation of disclosure intentions.

In general, the presented results exhibit strong power due to the large sample included in the analysis. Additionally, the high level of statistical significance of the supported hypotheses provides valuable insights into the relationships within the model. In doing so, our study adds theoretical contributions to the growing body of literature on human-AI-interaction and the role of trust for information disclosure by identifying Trust in AI as central aspect in the formation of information disclosure intentions. Beyond that, a key strength of our study lies in its ability to test privacy calculus across multiple disclosure

contexts. By examining diverse information types within AI-facilitated teamwork, we provide a nuanced perspective on how individuals weigh risks and benefits under varying conditions. This multi-faceted approach strengthens the robustness of our findings and enhances their generalizability. Furthermore, our study strengthens the theoretical assumptions made by privacy calculus theory by acknowledging Perceived Risks and Benefits as relevant factors that individuals weigh when deciding whether to disclose information. Our findings contribute to privacy calculus research by demonstrating how disclosure decisions vary across different types of personal information, rather than treating disclosure as a uniform construct. The embedded contextual variations in our study allow us to examine whether privacy calculus operates consistently across diverse disclosure scenarios, offering a richer, more comprehensive test of the theory. Beyond that, we found a tendency that certain Information types could be perceived as more sensible and intrusive and thus exert a (small) effect on Intention to disclose. In general, our results underscore that information disclosure is a complex process, centered around Perceived Risks and Benefits related to the potential disclosure decision but influenced by contextual as well as individual factors, in which Trust in AI as a multi-dimensional construct seems to play a major role.

We also identify practical implications for the workplace, especially for managers planning to employ AI systems for the moderation of teamwork situations. As our results indicate, practitioners responsible for the implementation of such systems should carefully consider the specific context of use and abstain from using AI as an unregulated facilitator for all meeting situations. Rather, practitioners should leverage hybrid intelligence, leaving the delicate and intervening dynamics to the human facilitator, while relieving the human facilitator from moderating less intervening team dynamics by including an AI-based facilitator (Bostrom et al., 1993). Furthermore, managers as well as human facilitators asked to share their tasks with an AI co-facilitator should pay attention to purposefully moderating risks and benefits perceived by teams in collaborative problem-solving tasks. As our research shows, these factors seem to have a critical connection to the individual intention to disclose, thus controlling to make the benefits visible while restricting the risks might be a valuable approach to increase the willingness to collaborate with respective systems (Hengstler et al., 2016).

18.6.2 Limitations & Outlook

Our research approach has limitations. First of all, our research model inherently draws upon the theory of privacy calculus, and we are well aware of the criticism this theory faces, especially with regards to its simplified representation of the way in which individuals weigh their decision to disclose personal information in a fully-informed and rational manner, without recognizing further central factors like context, environment or personal preferences for information disclosure. Given the interplay between individual and group privacy aspects in our study, future research should consider Bélanger & James (2020) Theory of Multilevel Information Privacy (TMIP), which explicitly integrates both

individual and collective privacy norms. A promising next step would be to extend our model by incorporating social identity and group privacy norms, recognizing that privacy decisions in teams are shaped not just by individual risk-benefit assessments but also by perceived team expectations and norms. Additionally, adjusting the study design to include real-time group interactions could allow for the emergence of collective privacy concerns. Finally, an empirical comparison of Privacy Calculus and TMIP models could provide further insights into whether integrating social dynamics enhances the explanatory power of disclosure intentions. By integrating these elements in future research, we can develop a more comprehensive understanding of privacy behavior in AI-mediated group settings, ensuring that both individual and group-level influences are adequately captured.

In connection to this, our results may be further limited by what is known as privacy paradox and refers to a discrepancy between the intention to disclose and actual disclosure behavior. As research has shown, the two constructs are not identical (Norberg et al., 2007). In this sense, the vignette-based study we have chosen as a research approach is also a limitation since we only measure Intention to disclose. As we have no real teams and no real problem-solving situations but imaginary scenarios, especially social effects may be underestimated because the potential dangers for the individual participants are only hypothetical. Furthermore, the choice of this method negatively influences external validity of the study. This might also be a reason for why we did not find statistically significant differences between the individual-level and the team-level conditions as well as only small effects of Information type on Intention to disclose. Consequently, proceeding with this idea in a lab experiment in which teams actually solve problems with or without the help of a comparable system would be an insightful endeavour which might lead to different or more precise findings. While such AI systems are not yet fully available and make research in these realms a considerable challenge, future research endeavors could leverage novel forms of the Wizard of Oz method, such as the method developed by Schecter et al. (2023) that simulates a group interaction in the virtual meeting space with an AI teammate.

Furthermore, we need to reflect critically on our experiment design. All participants received a comprehensive introduction to the context of the experiment as well as to the different scenarios, explaining what the hypothetical scenarios look like, which data is captured and how, as well as how the data will be stored and which privacy measures take effect. This may have developed a leading effect on participants' perceptions, especially when it comes to the perception of trust in (otherwise very opaque) AI systems. Consequently, it could be an interesting endeavor to see if any differences occur if there has been no comparable foundation laid for trusting the AI. Moreover, we measured the propensity to Trust AI in general and not per Information type, which might have additionally resulted in more nuanced differences. A further limitation is that we did not randomize the question order in the survey and did not control for the common method bias by e.g., including a measurement of a marker variable. Randomization could have reduced potential order effects or anchoring biases, where earlier items influence responses to later

ones. While we do not expect this to have significantly biased our findings, introducing randomization in future studies would improve robustness.

Beyond this, we only consider AI as a facilitator for process or team interactions, not for task-specific interventions (e.g., summarizing discussions, automatic transcription and analysis or visualization in mind maps) (Bittner et al., 2021). Further research could immerse into exploring the differences between AI facilitating the process and AI facilitating task interventions. A further limitation of our study is our focus on virtual collaboration. This considerations has been chosen purposefully in order to design the hypothetical scenarios as precise, true to life and imaginable for the participants as possible, but indeed set a narrow focus. As stated in the beginning, virtual collaboration may be especially suitable for AI-based facilitation. Nevertheless, meetings and workshops also still take place in person, which may differ from the virtual form. Consequently, exploring the possibilities for AI-based facilitation for on-site collaboration might be a promising avenue for future research as well.

As we have shown, there are various possibilities for future research to further explore human preferences regarding information disclosure to AI. However, we hope that our study helps to create a good foundation to connect to in order to advance research on privacy calculus in Human-AI-collaboration, as well as to inform practitioners on potential strategies for managing comparable AI systems that act as (co-)facilitators.

18.7 Conclusion

This paper extensively explored information disclosure to AI within teamwork, utilizing a vignette-based study to assess the impact of Information type, Recipient, and Anonymity on disclosure intentions. Notable distinctions in Intention to disclose emerged across various Information types, particularly with lower intentions for emotion and attention level disclosures.

Surprisingly, Recipient and Anonymity didn't show an effect in influencing disclosure intentions. Trust in AI consistently predicted Intention to disclose across all Information types, with its influence mediated by individuals' perceptions of associated risks and benefits.

Contributing to the evolving privacy calculus theory, this research illuminates the nuanced interplay between Trust, Perceived Risks, Benefits, and information disclosure in teamwork with AI. These insights hold implications for the effective use of AI in fostering workplace collaboration, underscoring the importance of cultivating trust and understanding sensitivities tied to diverse information types.

Extending beyond, the research's implications encompass the adept deployment of AI in enhancing teamwork and problem-solving processes within organizations. Emphasizing the pivotal role of Trust in AI, transparency, and understanding human preferences in

collaboration, the study highlights the importance of recognizing dynamics in (virtual) meetings to fully harness AI's diverse possibilities.

In conclusion, this paper provides valuable insights into the intricate dynamics of information disclosure to AI in teamwork. It accentuates the necessity for transparency, trust, and understanding human preferences in collaboration with AI, creating a compelling narrative that beckons thoughtful consideration of the possibilities and challenges in this evolving landscape.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

CRedit authorship contribution statement

Izabel Cvetkovic: Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Data curation, Conceptualization. Imke Grashoff: Writing – original draft, Validation, Conceptualization. Ana Jovancevic: Methodology, Data curation, Conceptualization. Eva Bittner: Supervision, Resources, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was funded by the German Federal Ministry of Education and Research (BMBF), in the context of the project HyMeKI (01IS20057) as well as by the Chair of Information Systems and Digital Innovation from the University of Hamburg.

18.8 Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.chbah.2025.100137>. Supplementary file further contains all Instruments as well as their corresponding Validity and Reliability Analyses. Additionally, this file contains estimates for alternatively tested path models (with Teamwork Attitudes; without Trust in AI) and further Robustness checks.

18.9 Appendix

18.9.1 Vignettes



Figure 5. Interruptions



Figure 6. Attention level



Figure 7. Breaking rules of creative teamwork

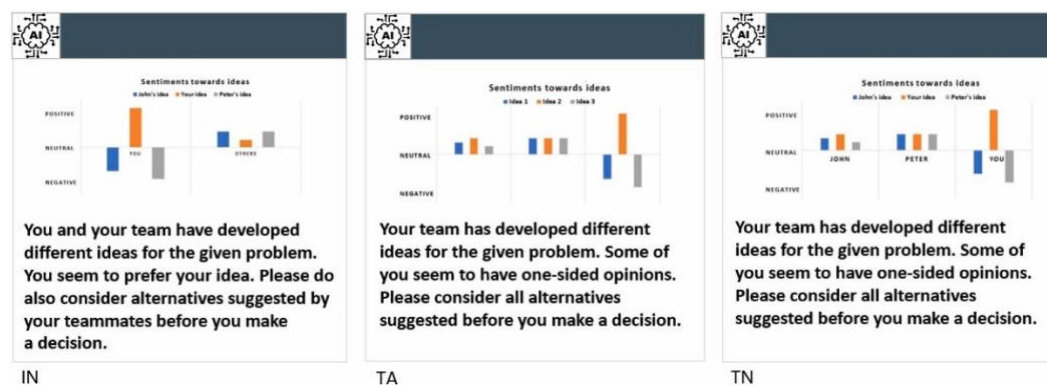


Figure 8. Bias toward certain solutions



Figure 9. Moodboard

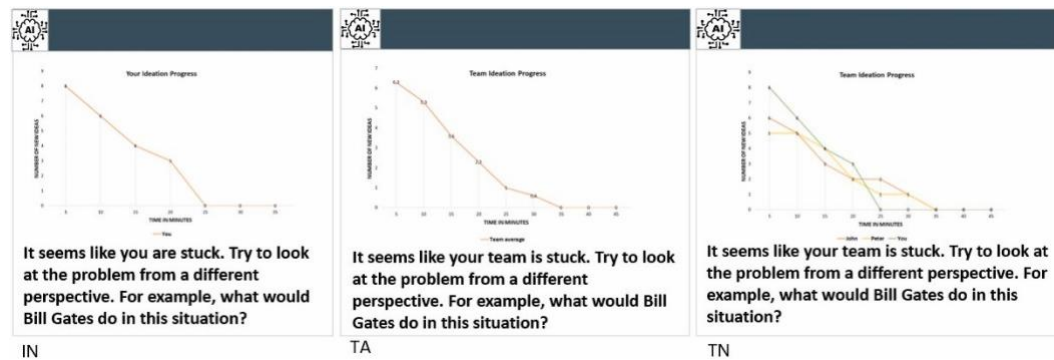


Figure 10. Stagnation

18.9.2 SEM Outputs – Correlation matrices

All correlations are significant at the level $p = 0.001$.

Table 10. Correlation matrix Talk times

<i>Talk times</i>	Trust in AI	Benefits	Risks	Disclose
Trust in AI	1			
Benefits	0,351	1		
Risks	-0,399	-0,24	1	
Disclose	0,466	0,343	-0,518	1

Table 11. Correlation matrix Interruptions

<i>Interruptions</i>	Trust in AI	Benefits	Risks	Disclose
Trust in AI	1			
Benefits	0,374	1		

<i>Interruptions</i>	Trust in AI	Benefits	Risks	Disclose
Risks	−0,408	−0,308	1	
Disclose	0,464	0,436	−0,525	1

Table 12. Correlation matrix Attention level

<i>Attention level</i>	Trust in AI	Benefits	Risks	Disclose
Trust in AI	1			
Benefits	0,472	1		
Risks	−0,455	−0,435	1	
Disclose	0,498	0,535	−0,605	1

Table 13. Correlation matrix Breaking rules of creative teamwork

<i>Breaking rules</i>	Trust in AI	Benefits	Risks	Disclose
Trust in AI	1			
Benefits	0,467	1		
Risks	−0,467	−0,469	1	
Disclose	0,546	0,614	−0,618	1

Table 14. Correlation matrix Bias toward certain solutions

<i>Bias</i>	Trust in AI	Benefits	Risks	Disclose
Trust in AI	1			
Benefits	0,447	1		
Risks	−0,455	−0,376	1	
Disclose	0,533	0,548	−0,601	1

Table 15. Correlation matrix Moodboard

<i>Moodboard</i>	Trust in AI	Benefits	Risks	Disclose
Trust in AI	1			
Benefits	0,477	1		
Risks	−0,457	−0,411	1	
Disclose	0,514	0,555	−0,599	1

Table 16. Correlation matrix Stagnation

<i>Stagnation</i>	Trust in AI	Benefits	Risks	Disclose
Trust in AI	1			
Benefits	0,505	1		
Risks	−0,468	−0,365	1	
Disclose	0,531	0,595	−0,511	1

18.10 References

- Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods*, 17(4), 351–371.
- Al-Natour, S., Benbasat, I., & Cenfetelli, R. (2021). Designing online virtual advisors to encourage customer self-disclosure: A theoretical model and an empirical test. *Journal of Management Information Systems*, 38(3), 798–827.
- Aoki, N. (2020). An experimental study of public trust in ai chatbots in the public sector. *Government Information Quarterly*, 37(4), 101–490.
- Arbuckle, J. L. (2019). *Amos* (version 26.0) [Computer software manual]. Balakrishnan, T., Chui, H. B., Michael, & Henke, N. (2023). The state of ai in 2020. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/global-survey-the-state-of-ai-in-2020>.
- Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50, 7–15.

- Behn, O., Leyer, M., & Iren, D. (2024). Employees' acceptance of ai-based emotion analytics from speech on a group level in virtual meetings. *Technology in Society*, 76, Article 102466.
- Bélanger, F., & James, T. L. (2020). A theory of multilevel information privacy management for the digital era. *Information Systems Research*, 31(2), 510–536.
- Belen-Saglam, R., Nurse, J. R., & Hodges, D. (2022). An investigation into the sensitivity of personal information and implications for disclosure: A UK perspective. *Frontiers of Computer Science*, 4.
- Berendt, B., Günther, O., & Spiekermann, S. (2005). Privacy in e-commerce: Stated preferences vs. actual behavior. *Communications of the ACM*, 48(4), 101–106.
- Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, 45(3).
- Bittner, E., Mirbabaie, M., & Morana, S. (2021). Digital facilitation assistance for collaborative, creative design processes. In *Proceedings of the 54th Hawaii international conference on system sciences*.
- Bittner, E., Oeste-Reiß, S., & Leimeister, J. M. (2019). Where is the bot in our team? Toward a taxonomy of design option combinations for conversational agents in collaborative work. In *Proceedings of the 52nd Hawaii international conference on system sciences*.
- Bol, N., Dienlin, T., Kruike-meier, S., Sax, M., Boerman, S., Strycharz, J., ... de Vreese, C. (2018). Understanding the effects of personalization as a privacy calculus: Analyzing self-disclosure across health, news, and commerce contexts. *Journal of Computer-Mediated Communication*, 23(6), 370–388.
- Bostrom, R. P., Anson, R., & Clawson, V. K. (1993). Group facilitation and group support systems. *Group support systems: New Perspectives*, 8, 146–168.
- Burgoon, J. K., Magnenat-Thalmann, N., Pantic, M., & Vinciarelli, A. (2017). *Social signal processing*. Cambridge University Press.
- Cebulla, A., Szpak, Z., Howell, C., Knight, G., & Hussain, S. (2023). Applying ethics to ai in the workplace: The design of a scorecard for Australian workplace health and safety. *AI & Society*, 38(2), 919–935.
- Cebulla, A., Szpak, Z., & Knight, G. (2023). Preparing to work with artificial intelligence: Assessing whs when using ai in the workplace. *International Journal of Workplace Health Management*, 16(4), 294–312.
- Chen, R., Demko, M., Byrne, D., & Louw, M. (2021). Probing documentation practices: Reflecting on students' conceptions, values, and experiences with documentation in creative inquiry. In *Proceedings of the 13th conference on creativity and cognition*.
- Choudhury, A., Asan, O., & Medow, J. E. (2022). Effect of risk, expectancy, and trust on clinicians' intent to use an artificial intelligence system–blood utilization calculator. *Applied Ergonomics*, 101, 103–708.

- Choung, H., David, P., & Ross, A. (2023). Trust in ai and its role in the acceptance of ai technologies. *International Journal of Human-Computer Interaction*, 39(9), 1727–1739.
- Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin*, 116(3), 457–475.
- Cooks-Campbell, A. (2023). Self-disclosure at work: Strengthening team bonds and communication. <https://www.betterup.com/blog/self-disclosure-at-work>.
- Culnan, M. J., & Armstrong, P. K. (1999). Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization Science*, 10 (1), 104–115.
- Cvetkovic, I., & Bittner, E. (2022). Task Delegability to AI: Evaluation of a Framework in a Knowledge Work Context. In *Proceedings of the 55th Hawaii International Conference on System Sciences (HICSS-55)*.
- Cvetkovic, I., Oeste-Reiß, S., Lehmann-Willenbrock, N., & Bittner, E. (2022). What should AI know? Information Disclosure in Human-AI Collaboration. *Proceedings of the European Conference on Information Systems (ECIS 2022)*.
- Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48, 24–42.
- Debowski, N., Tavanapour, N., & Bittner, E. (2021). Einsatz eines virtuellen Kollaborators in analogen & digitalen Workshops im organisationalen Kontext. *Informatik-Spektrum*, 1–8.
- Degutis, M., Urbonavicius, S., Zimaitis, I., Skare, V., & Laurutyte, D. (2020). Willingness to disclose personal information: How to measure it? *Engineering Economics*, 31(4), 487–494.
- Dell’Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., ... Lakhani, K. R. (2023). *Navigating the jagged technological frontier: Field experimental evidence of the effects of ai on knowledge worker productivity and quality (working paper No. 24-013)*. Harvard Business School Technology & Operations Mgt. Unit. Retrieved from <https://ssrn.com/abstract=4573321>.
- Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., & Ebel, P. (2019). The future of human-ai collaboration: A taxonomy of design knowledge for hybrid intelligence systems. In *Proceedings of the 52nd Hawaii international conference on system sciences*.
- Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering*, 61(5), 637–643.
- Dinev, T., Bellotto, M., Hart, P., Russo, V., Serra, I., & Colautti, C. (2006). Privacy calculus model in e-commerce - a study of Italy and the United States. *European Journal of Information Systems*, 15(4), 389–402.
- Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, 17(1), 61–80.

- Dinev, T., Xu, H., Smith, J. H., & Hart, P. (2013). Information privacy and correlates: An empirical attempt to bridge and distinguish privacy-related concepts. *European Journal of Information Systems*, 22(3), 295–316.
- Elshan, E., Siemon, D., de Vreede, T., de Vreede, G.-J., Oeste-Reiß, S., & Ebel, P. (2022). Requirements for ai-based teammates: A qualitative inquiry in the context of creative workshops. In *Proceedings of the 55th Hawaii international conference on system sciences*.
- Farivar, S., Yuan, Y., & Turel, O. (2016). Understanding social commerce acceptance: The role of trust, perceived risk, and benefit. In *Proceedings of the twenty-second americas conference on information systems*.
- Ferrada, F., & Camarinha-Matos, L. M. (2024). Emotions in human-ai collaboration. In *Working conference on virtual enterprises* (pp. 101–117).
- Gao, L., & Waechter, K. A. (2017). Examining the role of initial trust in user adoption of mobile payment services: An empirical investigation. *Information Systems Frontiers*, 19, 525–548.
- Guerin, T. F. (2017). A demonstration of how virtual meetings can enhance sustainability in a corporate context. *Environmental Quality Management*, 27(1), 75–81.
- Hengstler, M., Enkel, E., & Duelli, S. (2016). Applied artificial intelligence and trust—the case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change*, 105, 105–120.
- Hirschprung, R., Toch, E., Bolton, F., & Maimon, O. (2016). A methodology for estimating the value of privacy in information disclosure systems. *Computers in Human Behavior*, 61(C), 443–453.
- Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors*, 57(3), 407–434.
- Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2023). Measures for explainable ai: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human- ai performance. *Frontiers of Computer Science*, 5.
- James, T. L., Warkentin, M., & Collignon, S. E. (2015). A dual privacy decision model for online social networks. *Information Management*, 52(8), 893–908.
- JASP Team. (2023). *Jasp* (Version 0.18.0) [Computer software manual].
- Jayaraman, S., Phillips, E. K., Church, D., & Riek, L. D. (2024). Privacy and utility perceptions of social robots in healthcare. *Computers in Human Behavior: Artificial Humans*, 2(1).
- Joshi, N., Beecken, N., Bah, H., Steinicke, F., & Degner, J. (2022). Advanced emotion analytics of virtual group meetings involving intelligent virtual agents. In *2022 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW)*.
- Karl, K. A., Peluchette, J. V., & Aghakhani, N. (2022). Virtual work meetings during the covid-19 pandemic: The good, bad, and ugly. *Small Group Research*, 53(3), 343–365.

- Karren, R. J., & Barringer, M. W. (2002). A review and analysis of the policy-capturing methodology in organizational research: Guidelines for research and practice. *Organizational Research Methods*, 5(4), 337–361.
- Keith, M. J., Thompson, S. C., Hale, J., Lowry, P. B., & Greer, C. (2013). Information disclosure on mobile devices: Re-Examining privacy calculus with actual user behavior. *International Journal of Human-Computer Studies*, 71(12), 1163–1173.
- Knijnenburg, B. P., & Kobsa, A. (2013). Making decisions about privacy: Information disclosure in context-aware recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 3(3), 1–23.
- Knijnenburg, B. P., Kobsa, A., & Jin, H. (2013). Dimensionality of information disclosure behavior. *International Journal of Human-Computer Studies*, 71(12), 1144–1162.
- Kroll, T., & Stieglitz, S. (2021). Digital nudging and privacy: Improving decisions about self-disclosure in social networks. *Behaviour & Information Technology*, 40(1), 1–19.
- Lefkeli, D., Karatas, M., & Gürhan-Canli, Z. (2023). Sharing information with ai (versus a human) impairs brand trust: The role of audience size inferences and sense of exploitation. *International Journal of Research in Marketing*, 41(1), 138–155.
- Li, M., & Suh, A. (2022). Anthropomorphism in ai-enabled technology: A literature review. *Electronic Markets*, 32, 2245–2275.
- Li, T., & Unger, T. (2012). Willing to pay for quality personalization? Trade-off between quality and privacy. *European Journal of Information Systems*, 21, 621–642.
- Liao, M., & Sundar, S. S. (2021). How should ai systems talk to users when collecting their personal information? Effects of role framing and self-referencing on human-ai interaction. In *Proceedings of the 2021 CHI conference on human factors in computing systems (CHI '21)*.
- Lleras, C. (2005). Path analysis. In K. Kempf-Leonard (Ed.), *Encyclopedia of social measurement* (pp. 25–30). New York: Elsevier.
- Lubars, B., & Tan, C. (2019). Ask not what ai can do, but what ai should do: Towards a framework of task delegability. In *Advances in neural information processing systems* (Vol. 32).
- Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, 8(4), 277–301.
- Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., ... Söllner, M. (2019). Ai-based digital assistants. *Business & Information Systems Engineering*, 61(4), 535–544.
- Malik, N., Tripathi, S. N., Kar, A. K., & Gupta, S. (2021). Impact of artificial intelligence on employees working in industry 4.0 led organizations. *International Journal of Manpower*, 43(2), 334–354.

- Malkewitz, C. P., Schwall, P., Meesters, C., & Hardt, J. (2023). Estimating reliability: A comparison of cronbach's , mcdonald's t and the greatest lower bound. *Social Sciences Humanities Open*, 7(1), Article 100368.
- Manseau, J. (2020). Ai in the workplace: A qualitative analysis of intelligent employee assistants. In *AMCIS 2020 proceedings*.
- Mantello, P., Ho, M.-T., Nguyen, M.-H., & Vuong, Q.-H. (2023). Bosses without a heart: Socio-demographic and cross-cultural determinants of attitude toward emotional ai in the workplace. *AI & Society*, 38(1), 97–119.
- Masaviru, M. (2016). Self-disclosure: Theories and model review. *Journal of Culture, Society and Development*, 18, 43–47.
- McDuff, D., & Czerwinski, M. (2018). Designing emotionally sentient agents. *Communications of the ACM*, 61(12), 74–83.
- Mendo-Lázaro, S., Polo-del-Río, M. I., Iglesias-Gallego, D., Felipe-Castaño, E., & León-del-Barco, B. (2017). Construction and validation of a measurement instrument for attitudes towards teamwork. *Frontiers in Psychology*, 8.
- Merritt, S. M. (2011). Affective processes in human–automation interactions. *Human Factors*, 53(4), 356–370.
- Metzger, M. J. (2004). Privacy, trust, and disclosure: Exploring barriers to electronic commerce. *Journal of Computer-Mediated Communication*, 9(4).
- Milana, C., & Ashta, A. (2021). Artificial intelligence techniques in finance and financial markets: A survey of the literature. *Strategic Change*, 30(3), 189–209.
- Milne, G. R., & Gordon, M. E. (1993). Direct mail privacy-efficiency trade-offs within an implied social contract framework. *Journal of Public Policy and Marketing*, 12(2), 206–215.
- Molka-Danielsen, J., Rasool, J., & Smith, C. H. (2021). Design and deployment considerations for ethically advanced technologies for human flourishing in the workplace. In *IFIP working conference on human work interaction design*.
- Möhlmann, M., Zalmanson, L., Henfridsson, O., & Gregory, R. W. (2021). Algorithmic management of work on online labor platforms: When matching meets control. *MIS Quarterly*, 45(4).
- Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, 41 (1), 100–126.
- Omrani, N., Riviuccio, G., Fiore, U., Schiavone, F., & Agreda, S. G. (2022). To trust or not to trust? An assessment of trust in ai-based systems: Concerns, ethics and contexts. *Technological Forecasting and Social Change*, 181, Article 121763.
- Palan, S., & Schitter, C. (2018). Prolific.ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27.
- Pardes, A. (2020). Ai can run work meetings now. <https://www.wired.com/story/ai-can-run-work-meetings-now-headroom-clockwise/>.

- Park, H., Ahn, D., Hosanagar, K., & Lee, J. (2021). Human-ai interaction in human resource management: Understanding why employees resist algorithmic evaluation at workplaces and how to mitigate burdens. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. Association for Computing Machinery.
- Patil, S., & Lai, J. (2005). Who gets to know what when: Configuring privacy permissions in an awareness application. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. Association for Computing Machinery.
- Posey, P. B., Lowry, P. B., Roberts, T. L., & Ellis, T. S. (2010). Proposing the online community self-disclosure model: The case of working professionals in France and the UK who use online communities. *European Journal of Information Systems*, 19(2), 181–195.
- Przybilla, L., Baar, L., Wiesche, M., & Krcmar, H. (2019). Machines as teammates in creative teams: Digital facilitation of the dual pathways to creativity. In *Proceedings of the 2019 on computers and people research conference*. Association for Computing Machinery.
- Rayo, L., & Segal, I. (2010). Optimal information disclosure. *Journal of Political Economy*, 118(5), 949–987.
- Razin, Y. S., & Feigh, K. M. (2023). Converging measures and an emergent model: A meta-analysis of human-automation trust questionnaires. *ACM Transactions on human-Robot Interaction*, 13(4), 1–41.
- Recker, J., von Briel, F., Yoo, Y., Nagaraj, V., & McManus, M. (2023). Orchestrating human-machine designer ensembles during product innovation. *California Management Review*, 65(3), 27–47.
- Rodrigues, I. B., Adachi, J. D., Beattie, K. A., & MacDermid, J. C. (2017). Development and validation of a new tool to measure the facilitators, barriers and preferences to exercise in people with osteoporosis. *BMC Musculoskeletal Disorders*, 18(1), 1–9.
- Russell, S., & Norvig, P. (1995). *Artificial intelligence: A modern approach*. Prentice Hall.
- Sari, H. C. (2023). The impact of perceived risk, perceived benefit, and trust on customer intention to use tokopedia apps. *Jurnal Bisnis Strategi*, 31(2), 74–88.
- Schallmo, D. (2018). *Jetzt design thinking anwenden: In 7 schritten zu kundenorientierten produkten und dienstleistungen*. Wiesbaden: Springer Fachmedien Wiesbaden.
- Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the impact of “humanizing” customer service chatbots. *Information Systems Research*, 32(3), 736–751.
- Schechter, A., Hohenstein, J., Larson, L., Harris, A., Hou, T.-Y., Lee, W.-Y., ... Jung, M. (2023). Vero: An accessible method for studying human-ai teamwork. *Computers in Human Behavior*, 141(C).
- Schomakers, E.-M., Lidynia, C., Müllmann, D., & Ziefle, M. (2019). Internet users’ perceptions of information sensitivity—insights from Germany. *International Journal of Information Management*, 46, 142–150.

- Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G.-J., Elkins, A., ... Söllner, M. (2020). Machines as teammates: A research agenda on ai in team collaboration. *Information Management*, 57(2), 103–174.
- Seo, K., Tang, J., Roll, I., et al. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. *International Journal of Educational Technology in Higher Education*, 18.
- Shneiderman, B. (2020). Human-centered artificial intelligence: Three fresh ideas. *AIS Transactions on Human-Computer Interaction*, 12(3), 109–124.
- Siegrist, M. (2000). The influence of trust and perceptions of risks and benefits on the acceptance of gene technology. *Risk Analysis*, 20(2), 195–203.
- Siemon, D. (2022). Let the computer evaluate your idea: Evaluation apprehension in human-computer collaboration. *Behaviour & Information Technology*, 1–19.
- Solberg, E., Kaarstad, M., Eitheim, M. H. R., Bisio, R., Reegård, K., & Bloch, M. (2022). A conceptual model of trust, perceived risk, and reliance on ai decision aids. *Group & Organization Management*, 47(2), 187–222.
- St Marie, B., Jimmerson, A., Perkhounkova, Y., & Herr, K. (2021). Developing and establishing content validity of vignettes for health care education and research. *Western Journal of Nursing Research*, 43(7), 677–685.
- Stanton, N. (2004). Human interaction and non-verbal communication. *Mastering: Communication*, 30–49.
- Steel, J. L. (1991). Interpersonal correlates of trust and self-disclosure. *Psychological Reports*, 68(3), 1319–1320.
- Stutzman, F., Capra, R., & Thompson, J. (2011). Factors mediating disclosure in social network sites. *Computers in Human Behavior*, 27(1), 590–598.
- Teodorescu, M. H., Morse, L., Awwad, Y., & Kane, G. C. (2021). Failures of fairness in automation require a deeper understanding of human-ml augmentation. *MIS Quarterly*, 45(3).
- Tese, G. V., & Silvanio, V. (2023). A double-edged sword: The benefits and risks of ai in business. <https://www.eckertseamans.com/legal-updates/a-double-edged-sword-the-benefits-and-risks-of-ai-in-business>.
- Thibaut, J., & Kelley, H. (1959). *The social psychology of groups*. John Wiley.
- Toxtli, C., Monroy-Herna'ndez, A., & Cranshaw, J. (2018). Understanding chatbot-mediated task management. In *Proceedings of the 2018 CHI conference on human factors in computing systems*.
- Uysal, E., Alavi, S., & Bezençon, V. (2023). Anthropomorphism in artificial intelligence: A review of empirical work across domains and insights for future research. In K. Sudhir, & O. Toubia (Eds.), *Artificial intelligence in marketing* (Vol. 20, pp. 273–308). Emerald Publishing Limited.
- Vaughan-Nichols, S. (2023). Zoom is entangled in an ai privacy mess. <https://www.zdnet.com/article/zoom-is-entangled-in-an-ai-privacy-mess/>.

- Vogt, D. S., King, D. W., & King, L. A. (2004). Focus groups in psychological assessment: Enhancing content validity by consulting members of the target population. *Psychological Assessment, 16*(3), 231–243.
- Wadle, L.-M., Martin, N., & Ziegler, D. (2019). Privacy and personalization: The trade-off between data disclosure and personalization benefit. In *Adjunct publication of the 27th conference on user modeling, adaptation and personalization*.
- Waizenegger, L., McKenna, B., Cai, W., & Bendz, T. (2020). An affordance perspective of team collaboration and enforced working from home during covid-19. *European Journal of Information Systems, 29*(4), 429–442.
- Wang, Y., Qiao, T., & Liu, C. (2023). A study of reasons for self-disclosure on social media among Chinese covid-19 patients: Based on the theory of planned behavior model. *Healthcare, 11*, 1509.
- Weber, P., Pinski, M., & Baum, L. (2023). Toward an objective measurement of ai literacy. *Proceedings of the pacific-asia conference on information systems*. PACIS.
- Wei, K., Li, J., Ding, M., Ma, C., Yang, H. H., Farokhi, F., & Poor, H. V. (2020). Federated learning with differential privacy: Algorithms and performance analysis. *IEEE Transactions on Information Forensics and Security, 15*, 3454–3469.
- Westphal, M., Voßsing, M., Satzger, G., Yom-Tov, G. B., & Rafaeli, A. (2023). Decision control and explanations in human-ai collaboration: Improving user perceptions and compliance. *Computers in Human Behavior, 144*, Article 107714.
- Winkler, R., Neuweiler, M. L., Bittner, E., & Söllner, M. (2019). Hey alexa, please help us solve this problem! how interactions with smart personal assistants improve group performance. In *International conference of information systems*.
- Xiao, B., & Benbasat, I. (2007). Consumer decision support systems for e-commerce: Design and adoption of product recommendation agents. *MIS Quarterly, 31*(1), 317, 209.
- Xu, H., Dinev, T., Smith, H., & Hart, P. (2008). Examining the formation of individual's privacy concerns: Toward an integrative view. In *Proceedings of the international conference of information systems*.
- Yu, L., Li, H., He, W., Wang, F.-K., & Jiao, S. (2020). A meta-analysis to explore privacy cognition and information disclosure of internet users. *International Journal of Information Management, 51*, Article 102015.
- Yusoff, M. S. B. (2019). Abc of content validation and content validity index calculation. *Education in Medicine Journal, 11*(2), 49–54.
- Zadeh, A., Chen, M., Poria, S., Cambria, E., & Morency, L.-P. (2017). Tensor fusion network for multimodal sentiment analysis. In *Proceedings of the 2017 conference on empirical methods in natural language processing*.
- Zoom Video Communications, I. (2021). Attendee attention tracking. <https://support.zoom.us/hc/en-us/articles/115000538083-Attendee-attention-tracking>.

19 References

- Adam, M. T. P [Marc. T. P.], Gregor, S., Hevner, A., & Morana, S. (2021). Design Science Research Modes in Human-Computer Interaction Projects. *AIS Transactions on Human-Computer Interaction*, 1–11. <https://doi.org/10.17705/1thci.00139>
- Adamopoulou, E., & Moussiades, L. (2020). An Overview of Chatbot Technology: Artificial Intelligence Applications and Innovations. *Proceedings of the IFIP Advances in Information and Communication Technology*, 373–383. https://doi.org/10.1007/978-3-030-49186-4_31
- Aguinis, H., & Bradley, K. J. (2014). Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies. *Organizational Research Methods*, 17(4), 351–371. <https://doi.org/10.1177/1094428114547952>
- Al-Natour, S., Benbasat, I., & Cenfetelli, R. (2021). Designing Online Virtual Advisors to Encourage Customer Self-disclosure: A Theoretical Model and an Empirical Test. *Journal of Management Information Systems*, 38(3), 798–827. <https://doi.org/10.1080/07421222.2021.1962595>
- Amabile, T. M. (1996). *Creativity in context: Update to "The Social Psychology of Creativity."*. Westview Press.
- Amershi, S., Weld, D., Vorvoreanu, M., Fournery, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for Human-AI Interaction. *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. <https://doi.org/10.1145/3290605.3300233>
- Arthur, D., & Vassilvitskii, S. (2006). *k-means++: The Advantages of Careful Seeding*. Stanford; InfoLab. <http://ilpubs.stanford.edu:8090/778/?ref=https://githubhelp.com>
- Avdiji, H., & Winter, R. (2019). Knowledge Gaps in Design Science Research. *Proceedings of the 40th International Conference on Information Systems (ICIS 2019)*. https://aisel.aisnet.org/icis2019/design_science/design_science/4
- Bailenson, J. N., Yee, N., Merget, D., & Schroeder, R. (2006). The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction. *Presence: Teleoperators and Virtual Environments*, 15(4), 359–372. <https://doi.org/10.1162/pres.15.4.359>
- Baird, A., & Maruping, L. M. (2021). The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts. *MIS Quarterly*, 45(1), 315–341. <https://doi.org/10.25300/MISQ/2021/15882>
- Baskerville, & Myers (2004). Special Issue on Action Research in Information Systems: Making IS Research Relevant to Practice: Foreword. *MIS Quarterly*, 28(3), 329. <https://doi.org/10.2307/25148642>

- Baskerville, R., Baiyere, A., Gergor, S., Hevner, A., & Rossi, M. (2018). Design Science Research Contributions: Finding a Balance between Artifact and Theory. *Journal of the Association for Information Systems*, 19(5), 358–376. <https://doi.org/10.17705/1jais.00495>
- Baskerville, R., Pries-Heje, J., & Venable, J. Soft design science methodology. *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*. <https://doi.org/10.1145/1555619.1555631>
- Baxter, G., & Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering. *Interacting with Computers*, 23(1), 4–17. <https://doi.org/10.1016/j.intcom.2010.07.003>
- Belen-Saglam, R., Nurse, J. R. C., & Hodges, D. (2022). An Investigation Into the Sensitivity of Personal Information and Implications for Disclosure: A UK Perspective. *Frontiers in Computer Science*, 4, Article 908245. <https://doi.org/10.3389/fcomp.2022.908245>
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246. <https://doi.org/10.1037//0033-2909.107.2.238>
- Beranek, P. M., Beise, C. M., & Niederman, F. (1993). Facilitation and group support systems. *Proceedings of the Twenty-Sixth Hawaii International Conference on System Sciences*, 199-207. <https://doi.org/10.1109/HICSS.1993.284182>
- Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing Artificial Intelligence. *MIS Quarterly*, 45(3), 1433–1450. <https://doi.org/10.25300/MISQ/2021/16274>
- Bittner, E., Küstermann, G., & Tratzky, C. (2019). The facilitator is a bot: Towards a conversational agent for facilitating idea elaboration on idea platforms. *Proceeding of the 27th European Conference on Information Systems*. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1165&context=ecis2019_rp
- Bittner, E., & Leimeister, J. M. (2014). Creating shared understanding in heterogeneous work groups: Why it matters and how to achieve it. *Journal of Management Information Systems : JMIS*, 31(1), 111–143.
- Bittner, E., Mirbabaie, M., & Morana, S. (2021). Digital Facilitation Assistance for Collaborative, Creative Design Processes. *Proceedings of the 54th Hawaii International Conference on System Sciences*. <https://doi.org/10.24251/HICSS.2021.044>
- Bouwer, A. (2022). Under Which Conditions Are Humans Motivated to Delegate Tasks to AI? A Taxonomy on the Human Emotional State Driving the Motivation for AI Delegation. *Marketing and Smart Technologies*, 279, 37–53. https://doi.org/10.1007/978-981-16-9268-0_4
- Brandtzæg, P. B., & Følstad, A. (2018). Chatbots: changing user needs and motivations. *Interactions*, 25(5). <https://dl.acm.org/doi/10.1145/3236669>
- Brave, S., Nass, C., & Hutchinson, K. (2005). Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent.

- International Journal of Human-Computer Studies*, 62(2), 161–178.
<https://doi.org/10.1016/j.ijhcs.2004.11.002>
- Briggs, R. O., Kolfshoten, G. L., Vreede, G.-J. de, Lukosch, S., & Albrecht, C. C. (2013). Facilitator-in-a-Box: Process Support Applications to Help Practitioners Realize the Potential of Collaboration Technology. *Journal of Management Information Systems*, 29(4), 159–194. <https://doi.org/10.2753/MIS0742-1222290406>
- Brown, T [Tim], & Katz, B [Barry] (2011). Change by Design. *Journal of Product Innovation Management*, 28(3), 381–383. <https://doi.org/10.1111/j.1540-5885.2011.00806.x>
- Brown, T [Tim], & Katz, B [Barry]. (2019). *Change by design: How design thinking transforms organizations and inspires innovation* (Revised and updated edition). Harper Business.
<https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=6931612>
- Brown, T [Tom], Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J [Jeffrey], Winter, C., . . . Amodei, D. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Bryman, A. (2016). *Social Research Methods*. Oxford University Press.
- Brynjolfsson, E., & McAfee, A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W.W. Norton & Company.
- Brynjolfsson, E., & McAfee, A. (2017). The Business of Artificial Intelligence: What it Can—and Cannot—Do for Your Organization. *Harvard Business Review Digital Articles*, 1–20.
- Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To Trust or to Think. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1–21.
<https://doi.org/10.1145/3449287>
- Carlgren, L., Elmqvist, M., & Rauth, I. (2014). Exploring the use of design thinking in large organizations: Towards a research agenda. *Swedish Design Research Journal*, 11, 55–63. <https://doi.org/10.3384/svid.2000-964x.14155>
- Chen, R., Demko, M., Byrne, D., & Louw, M. (2021). Probing Documentation Practices: Reflecting on Students' Conceptions, Values, and Experiences with Documentation in Creative Inquiry. *C&C '21: Proceedings of the 13th Conference on Creativity and Cognition*, 1–14. <https://doi.org/10.1145/3450741.3465391>
- Choung, H., David, P., & Ross, A. (2023). Trust in AI and Its Role in the Acceptance of AI Technologies. *International Journal of Human-Computer Interaction*, 39(9), 1727–1739. <https://doi.org/10.1080/10447318.2022.2050543>
- Clawson, V. K., Bostrom, R. P., & Anson, R. (1993). The Role of the Facilitator in Computer-Supported Meetings. *Small Group Research*, 24(4), 547–565.
<https://doi.org/10.1177/1046496493244007>

- Cook, D. (2020). The Global Remote Work Revolution and the Future of Work. In *The Business of Pandemics* (pp. 143–166). Auerbach Publications. <https://doi.org/10.1201/9781003094937-10>
- Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society*, 1(1), 104. <https://doi.org/10.1007/BF03177550>
- Cranefield, J., Winikoff, M., Chiu, Y.-T., Li, Y., Doyle, C., & Richter, A. (2023). Partnering with AI: The case of digital productivity assistants. *Journal of the Royal Society of New Zealand*, 53(1), 95–118. <https://doi.org/10.1080/03036758.2022.2114507>
- Creswell, J. W., & Creswell, J. D. (2017). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications.
- Culnan, M. J., & Armstrong, P. K. (1999). Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation. *Organization Science*, 10(1), 104–115. <https://doi.org/10.1287/orsc.10.1.104>
- Curry, A., Flett, P., & Hollingsworth, I. (2006). *Managing Information & Systems*. Routledge. <https://doi.org/10.4324/9780203002315>
- Cvetkovic, I., & Bittner, E. (2022). Task Delegability to AI: Evaluation of a Framework in a Knowledge Work Context. *Proceedings of the 55th Hawaii International Conference on System Sciences*. <https://doi.org/10.24251/HICSS.2022.021>
- Cvetkovic, I., Gierlich-Joas, M., Tavanapour, N., Debowski-Weimann, N., & Bittner, E. (2023). Augmented facilitation: Designing a multi-modal conversational agent for group ideation. *Proceedings of the 44th International Conference on Information Systems*.
- Cvetkovic, I., Grashoff, I., Jovancevic, A., & Bittner, E. (2025). Quid pro Quo: Information disclosure for AI feedback in Human-AI collaboration. *Computers in Human Behavior: Artificial Humans*, 4, 100137. <https://doi.org/10.1016/j.chbah.2025.100137>
- Cvetkovic, I., Oeste-Reiß, S., Lehmann-Willenbrock, N., & Bittner, E. (2022). WHAT SHOULD AI KNOW? INFORMATION DISCLOSURE IN HUMAN-AI COLLABORATION. *Proceedings of the 30th European Conference on Information Systems (ECIS 2022)*.
- Cvetkovic, I., Rosenberg, V., & Bittner, E. (2023). Conversational Agent as a Black Hat: Can Criticising Improve Idea Generation? *Proceedings of the 56th Hawaii International Conference on System Sciences*. <https://doi.org/10.24251/hicss.2023.023>
- Cvetkovic, I., Streek, S., & Bittner, E. (2024). Designing a Hybrid Intelligence System for the Facilitation of Group Collaboration. *Proceedings of the Australasian Conference on Information Systems (ACIS 2024)*. <https://aisel.aisnet.org/acis2024/98/>
- Davenport, T. H., & Kirby, J. (2016). *Only Humans Need Apply: Winners and Losers in the Age of Smart Machines*. Harper Business.

- Davenport, T. H., & Ronanki, R. (2018). *Artificial intelligence for the real world* (1st ed.). Harvard business review.
- De Bono, E. (1985). *Six Thinking Hats*. Key Porter Books.
- Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Kraye, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4573321>
- Dellermann, D., Calma, A., Lipusch, N., Thorsten, W., Weigel, S., & Ebel, P. (2019). The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems. *Proceedings of the 52nd Hawaii International Conference on System Sciences*. <http://128.171.57.22/bitstream/10125/59468/0028.pdf>
- Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid Intelligence. *Business & Information Systems Engineering*, 61(5), 637–643. <https://doi.org/10.1007/s12599-019-00595-2>
- Demir, M., Likens, A. D., Cooke, N. J., Amazeen, P. G., & McNeese, N. J. (2019). Team Coordination and Effectiveness in Human-Autonomy Teaming. *IEEE Transactions on Human-Machine Systems*, 49(2), 150–159. <https://doi.org/10.1109/thms.2018.2877482>
- Deng, C., Ji, X., Rainey, C., Zhang, J., & Lu, W. (2020). Integrating Machine Learning with Human Knowledge. *IScience*, 23(11), 101656. <https://doi.org/10.1016/j.isci.2020.101656>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- Di Fede, G., Rocchesso, D., Dow, S. P., & Andolina, S. (2022). The Idea Machine: LLM-based Expansion, Rewriting, Combination, and Suggestion of Ideas. *C&C '22: Proceedings of the 14th Conference on Creativity and Cognition*, 623–627. <https://doi.org/10.1145/3527927.3535197>
- Diederich, S., Brendel, A. B., Morana, S., & Kolbe, L. (2022). On the Design of and Interaction with Conversational Agents: An Organizing and Assessing Review of Human-Computer Interaction Research. *Journal of the Association for Information Systems*, 23(1), 96–138. <https://doi.org/10.17705/1jais.00724>
- Doyle, C., & Luczak-Roesch, M. (2020). This Paper Is an Artefact: On Open Science Practices in Design Science Research Using Registered Reports. *Proceedings of the 53rd Hawaii International Conference on System Sciences*. <https://doi.org/10.26686/wgtn.12522038>
- Drechsler, A., & Hevner, A. (2018). Utilizing, Producing, and Contributing Design Knowledge in DSR Projects. *International Conference on Design Science*

- Research in Information Systems and Technology*, 82–97. https://doi.org/10.1007/978-3-319-91800-6_6
- Dresch, A., Lacerda, D. P., & Antunes Jr, J. A. V. (2015). *Design Science Research*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-07374-3>
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., . . . Wright, R. (2023). Opinion Paper: “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Eason, K. (2008). Sociotechnical systems theory in the 21st Century: another half-filled glass. In *Sense in social science: A collection of essays in honour of Dr. Lisl Klein* (pp. 123–134). Sense Publishers.
- Edwards, K. M., Song, B., Porciello, J., Engelbert, M., Huang, C., & Ahmed, F. (2024). ADVISE: Accelerating the Creation of Evidence Syntheses for Global Development Using Natural Language Processing-Supported Human-Artificial Intelligence Collaboration. *Journal of Mechanical Design*, 146(5). <https://doi.org/10.1115/1.4064245>
- Ehsan, U., & Riedl, M. O. (2020). Human-Centered Explainable AI: Towards a Reflective Sociotechnical Approach. *HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence*, 12424, 449–466. https://doi.org/10.1007/978-3-030-60117-1_33
- Elshan, E., Siemon, D., Vreede, T. de, Vreede, G.-J. de, Oeste-Reiß, S., & Ebel, P. (2022). Requirements for AI-based Teammates: A Qualitative Inquiry in the Context of Creative Workshops. *Proceedings of the 55th Hawaii International Conference on System Sciences*. <https://doi.org/10.24251/HICSS.2022.020>
- Evers, V., Maldonado, H. C., Brodecki, T. L., & Hinds, P. J. (2008). Relational vs. group self-construal: untangling the role of national culture in HRI. *HRI '08: Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*, 255–262. <https://doi.org/10.1145/1349822.1349856>
- Eziefule, A. O., Adelakun, B. O., Okoye, I. N., & Attieku, J. S. (2022). The Role of AI in Automating Routine Accounting Tasks: Efficiency Gains and Workforce Implications. *European Journal of Accounting, Auditing and Finance Research*, 10(12), 109–134. <https://ejournals.org/ejaaf/vol10-issue12-2022/the-role-of-ai-in-automating-routine-accounting-tasks-efficiency-gains-and-workforce-implications/>
- Fischer, A., Greiff, S., & Funke, J. (2012). The Process of Solving Complex Problems. *The Journal of Problem Solving*, 4(1). <https://doi.org/10.7771/1932-6246.1118>

- Fossey, E., Harvey, C., McDermott, F., & Davidson, L. (2002). Understanding and evaluating qualitative research. *The Australian and New Zealand Journal of Psychiatry*, 36(6), 717–732. <https://doi.org/10.1046/j.1440-1614.2002.01100.x>
- Gau, M., Maedche, A., & vom Brocke, J. (2022). DSR Buddy: A Conversational Agent Supporting Design Science Research Activities. *Wirtschaftsinformatik 2022 Proceedings*.
- Gero, K. I., Liu, V., & Chilton, L. (2022). Sparks: Inspiration for Science Writing using Language Models. *DIS '22: Proceedings of the 2022 ACM Designing Interactive Systems Conference*, 1002–1019. <https://doi.org/10.1145/3532106.3533533>
- Ghosh, D., Fabbri, A. R., & Muresan, S. (2018). Sarcasm Analysis Using Conversation Context. *Computational Linguistics*, 44(4), 755–792. https://doi.org/10.1162/coli_a_00336
- Goldkuhl, G. (2012). Pragmatism vs interpretivism in qualitative information systems research. *European Journal of Information Systems*, 21(2), 135–146. <https://doi.org/10.1057/ejis.2011.54>
- Gozli, D. (2019). What Is a Task? In D. Gozli (Ed.), *Experimental Psychology and Human Agency* (pp. 83–111). Springer International Publishing. https://doi.org/10.1007/978-3-030-20422-8_5
- Gregor (2006). The Nature of Theory in Information Systems. *MIS Quarterly*, 30(3), 611. <https://doi.org/10.2307/25148742>
- Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, 37(2), 337–355. <https://doi.org/10.25300/MISQ/2013/37.2.01>
- Gregor, S., Kruse, L., & Seidel, S. (2020). Research Perspectives: The Anatomy of a Design Principle. *Journal of the Association for Information Systems*, 21, 1622–1652. <https://doi.org/10.17705/1jais.00649>
- Gregory, R. W., & Muntermann, J. (2014). Research Note —Heuristic Theorizing: Proactively Generating Design Theories. *Information Systems Research*, 25(3), 639–653. <https://doi.org/10.1287/isre.2014.0533>
- Gunning, D., & Aha, D. W. (2019). DARPA's Explainable Artificial Intelligence Program. *AI Magazine*, 40(2), 44–58. <https://doi.org/10.1609/aimag.v40i2.2850>
- Hevner, A. (2007). A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*, 19(2). <https://aisel.aisnet.org/sjis/vol19/iss2/4>
- Hevner, A., & Chatterjee, S [Samir]. Design Science Research in Information Systems. In *Design Research in Information Systems. Integrated Series in Information Systems* (Vol. 22, pp. 9–22). https://doi.org/10.1007/978-1-4419-5653-8_2
- Hevner, A., March, S., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 75–105.
- Hevner, A., & Zhang, P. (2011). Introduction to the AIS THCI Special Issue on Design Research in Human-Computer Interaction. *AIS Transactions on Human-Computer Interaction*, 3(2), 56–61. <https://doi.org/10.17705/1thci.00026>

- Hjalmarsson, A., Recker, J., Rosemann, M., & Lind, M [Mikael] (2015). Understanding the Behavior of Workshop Facilitators in Systems Analysis and Design Projects: Developing Theory from Process Modeling Projects. *Communications of the Association for Information Systems*, 36(1). <https://doi.org/10.17705/1CAIS.03622>
- Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3), 407–434. <https://doi.org/10.1177/0018720814547570>
- Huang, M.-H., & Rust, R. T. (2018). Artificial Intelligence in Service. *Journal of Service Research*, 21(2), 155–172. <https://doi.org/10.1177/1094670517752459>
- Hutchins, E. (2000). Distributed cognition. *International Encyclopedia of the Social and Behavioral Sciences*, 138(1), 1–10.
- Hutson, J. (2023). Human-AI Collaboration in Writing: A Multidimensional Framework for Creative and Intellectual Authorship. *International Journal of Changes in Education*, 00(00). <https://doi.org/10.47852/bonviewIJCE52024908>
- Ioannou, A., Tussyadiah, I., Miller, G., Li, S., & Weick, M. (2021). Privacy nudges for disclosure of personal information: A systematic literature review and meta-analysis. *PLOS ONE*, 16(8), e0256822. <https://doi.org/10.1371/journal.pone.0256822>
- Jiang, E., Olson, K., Toh, E., Molina, A., Donsbach, A., Terry, M., & Cai, C. J. (2022). PromptMaker: Prompt-based Prototyping with Large Language Models. *Proceedings of the CHI EA '22: CHI Conference on Human Factors in Computing Systems Extended Abstracts*.
- Jin, H [Huiying], & Uchida, M. (2024). A Three-Year Analysis of Human Preferences in Delegating Tasks to AI. *International Conference on Human-Computer Interaction*, 48–66. https://doi.org/10.1007/978-3-031-60606-9_4
- Joksimovic, S., Ifenthaler, D., Marrone, R., Laat, M. de, & Siemens, G. (2023). Opportunities of artificial intelligence for supporting complex problem-solving: Findings from a scoping review. *Computers and Education: Artificial Intelligence*, 4, 100–138. <https://doi.org/10.1016/j.caeai.2023.100138>
- Jones, D., & Gregor, S. (2007). The Anatomy of a Design Theory. *Journal of the Association for Information Systems*, 8(5), 312–335. <https://doi.org/10.17705/1jais.00129>
- Karl, K. A., Peluchette, J. V., & Aghakhani, N. (2022). Virtual Work Meetings During the COVID-19 Pandemic: The Good, Bad, and Ugly. *Small Group Research*, 53(3), 343–365. <https://doi.org/10.1177/10464964211015286>
- Khan, R., & Das, A. (2018). *Build better chatbots: A Complete Guide to Getting Started with Chatbots*. Apress Berkeley, CA. <https://doi.org/10.1007/978-1-4842-3111-1>
- Kim, S., Eun, J., Seering, J., & Lee, J. (2021). Moderator Chatbot for Deliberative Discussion. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1–26. <https://doi.org/10.1145/3449161>

- Kirsh, D. (2010). Thinking with external representations. *AI & SOCIETY*, 25(4), 441–454. <https://doi.org/10.1007/s00146-010-0272-8>
- Klein, G., Feltoovich, P. J., Bradshaw, J. M., & Woods, D. D. (2005). Common Ground and Coordination in Joint Activity. In W. B. Rouse & K. R. Boff (Eds.), *Organizational Simulation* (pp. 139–184). Wiley. <https://doi.org/10.1002/0471739448.ch6>
- Knijnenburg, B. P., & Kobsa, A. (2013). Making Decisions about Privacy. *ACM Transactions on Interactive Intelligent Systems*, 3(3), 1–23. <https://doi.org/10.1145/2499670>
- Kopenhagen, N., Gaß, O., & Müller, B. (2012). Design science research in action-anatomy of success critical activities for rigor and relevance. *Proceedings of the 20th European Conference on Information Systems*.
- Kroll, T., & Stieglitz, S. (2021). Digital nudging and privacy: improving decisions about self-disclosure in social networks. *Behaviour & Information Technology*, 40(1), 1–19. <https://doi.org/10.1080/0144929X.2019.1584644>
- Kuechler, W., & Vaishnavi, V. (2012). A Framework for Theory Development in Design Science Research: Multiple Perspectives. *Journal of the Association for Information Systems*, 13(6), 395–423. <https://doi.org/10.17705/1jais.00300>
- Kuhn, T. S. (1962). *The structure of scientific revolutions* (3. ed.). Univ. of Chicago Press.
- Kundisch, D., Muntermann, J., Oberländer, A. M., Rau, D., Röglinger, M., Schoormann, T., & Szopinski, D. (2021). An Update for Taxonomy Designers. *Business & Information Systems Engineering*, 64(4), 421–439. <https://doi.org/10.1007/s12599-021-00723-x>
- Lasecki, W. S., Wesley, R., Nichols, J., Kulkarni, A., Allen, J. F., & Bigham, J. P. (2013). Chorus: a crowd-powered conversational assistant. *UIST '13: Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, 151–162. <https://doi.org/10.1145/2501988.2502057>
- Latto, C., Richter, A., & Tate, M. (2025). Human-AI Teams' Impact on Organizations - A Review. *Proceedings of the 58th Hawaii International Conference on System Sciences*.
- Laufer, R. S., & Wolfe, M. (1977). Privacy as a Concept and a Social Issue: A Multidimensional Developmental Theory. *Journal of Social Issues*, 33(3), 22–42. <https://doi.org/10.1111/j.1540-4560.1977.tb01880.x>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lee, & Hubona (2009). A Scientific Basis for Rigor in Information Systems Research. *MIS Quarterly*, 33(2), 237. <https://doi.org/10.2307/20650291>
- Leimeister, J. M. (Ed.). (2014). *Collaboration Engineering*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-20891-1>
- Lewandowski, T., Kucevic, E., Leible, S., Poser, M., & Böhmman, T. (2023). Enhancing conversational agents for successful operation: A multi-perspective evaluation

- approach for continuous improvement. *Electronic Markets*, 33(1), 39. <https://doi.org/10.1007/s12525-023-00662-3>
- Li, T. W., Arya, A., & Jin, H [Haojian] (2024). Redesigning Privacy with User Feedback: The Case of Zoom Attendee Attention Tracking. *CHI '24: Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–14. <https://doi.org/10.1145/3613904.3642594>
- Liao, Q. V., & Wortman Vaughan, J. (2023). AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap. *Harvard Data Science Review*. <https://doi.org/10.1162/99608f92.8036d03b>
- Lösser, B., Oberländer, A. M., & Rau, D. (2019). Taxonomy Research in Information Systems : A Systematic Assessment. *Proceedings of the 27th European Conference on Information Systems*.
- Lubars, B., & Tan, C. (2017). Ask not what AI can do, but what AI should do: Towards a framework of task delegability. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. <https://dl.acm.org/doi/pdf/10.5555/3454287.3454293>
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed: An Argument for AI in Education*. Pearson.
- Lukyanenko, R., Maass, W., & Storey, V. C. (2022). Trust in artificial intelligence: From a Foundational Trust Framework to emerging research opportunities. *Electronic Markets*, 32(4), 1993–2020. <https://doi.org/10.1007/s12525-022-00605-4>
- Maedche, A., Gregor, S., & Parsons, J. (2021). Mapping Design Contributions in Information Systems Research: The Design Research Activity Framework. *Communications of the Association for Information Systems*, 49(1), 355–378. <https://doi.org/10.17705/1CAIS.04914>
- Mahadevkar, S. V., Patil, S [Shruti], Kotecha, K., Soong, L. W., & Choudhury, T. (2024). Exploring AI-driven approaches for unstructured document analysis and future horizons. *Journal of Big Data*, 11(1). <https://doi.org/10.1186/s40537-024-00948-z>
- Maher, M. L., & Fisher, D. H. (2012). Using AI to Evaluate Creative Designs. *Proceedings of the ICDC 2012 - 2nd International Conference on Design Creativity*, 1.
- Mantello, P., Ho, M.-T., Nguyen, M.-H., & Vuong, Q.-H. (2023). Bosses without a heart: Socio-demographic and cross-cultural determinants of attitude toward Emotional AI in the workplace. *AI & SOCIETY*, 38(1), 97–119. <https://doi.org/10.1007/s00146-021-01290-1>
- March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, 15(4), 251–266. [https://doi.org/10.1016/0167-9236\(94\)00041-2](https://doi.org/10.1016/0167-9236(94)00041-2)
- Masaviru, M. (2016). Self-disclosure: Theories and model review. *Journal of Culture, Society and Development*, 18. <https://core.ac.uk/download/pdf/234691127.pdf>
- Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *The*

- Journal of Applied Psychology*, 85(2), 273–283. <https://doi.org/10.1037/0021-9010.85.2.273>
- Maxwell, J. (1992). Understanding and Validity in Qualitative Research. *Harvard Educational Review*, 62(3), 279–301. <https://doi.org/10.17763/haer.62.3.8323320856251826>
- Mayring, P. (2014). *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. Alpen-Adria-Universität Klagenfurt. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-395173>
- McCarthy, J. (2007). *What is artificial intelligence*. <https://www-formal.stanford.edu/jmc/>
- Meinel, C., & Leifer, L. (2012). Design Thinking Research. In H. Plattner, C. Meinel, & L. Leifer (Eds.), *Design Thinking Research: Studying Co-Creation in Practice* (pp. 1–11). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-21643-5_1
- Memmert, L., & Bittner, E. (2022). Complex Problem Solving through Human-AI Collaboration: Literature Review on Research Contexts. *Proceedings of the 55th Hawaii International Conference on System Sciences (HICSS-55)*.
- Memmert, L., Cvetkovic, I., & Bittner, E. (2023). HUMAN-AI COLLABORATION IN CONCEPTUALIZING DESIGN SCIENCE RESEARCH STUDIES: PERCEIVED HELPFULNESS OF GENERATIVE LANGUAGE MODEL'S SUGGESTIONS. *Proceedings of the 31st European Conference on Information Systems*.
- Memmert, L., Cvetkovic, I., & Bittner, E. (2024). The More Is Not the Merrier: Effects of Prompt Engineering on the Quality of Ideas Generated By GPT-3. *Proceedings of the 57th Hawaii International Conference on System Sciences*. <https://scholarspace.manoa.hawaii.edu/items/b73c910e-5614-4522-be46-f2740bde1afc>
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. SAGE.
- Mingers, J. (2001). Combining IS Research Methods: Towards a Pluralist Methodology. *Information Systems Research*, 12(3), 240–259. <https://doi.org/10.1287/isre.12.3.240.9709>
- Mishra, S., Khashabi, D., Baral, C., Choi, Y., & Hajishirzi, H. (2022). Reframing Instructional Prompts to GPTk's Language. *Proceedings of the Findings of the Association for Computational Linguistics: ACL 2022*, 589–612. <https://arxiv.org/pdf/2109.07830.pdf>
- Molina, M. D., & Sundar, S. S. (2022). When AI moderates online content: effects of human collaboration and interactive transparency on user trust. *Journal of Computer-Mediated Communication*, 27(4), Article zmac010. <https://doi.org/10.1093/jcmc/zmac010>
- Morana, S., Gnewuch, U., Jung, D., & Granig, C. (2020). The Effect of Anthropomorphism on Investment Decision-Making with Robo-Advisor Chatbots. *Proceedings of the 28th European Conference on Information Systems*.

- Morana, S., Pfeiffer, J., & Adam, M. T. P [Marc T. P.] (2020). User Assistance for Intelligent Systems. *Business & Information Systems Engineering*, 62(3), 189–192. <https://doi.org/10.1007/s12599-020-00640-5>
- Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., & Fernández-Leal, Á. (2023). Human-in-the-loop machine learning: a state of the art. *Artificial Intelligence Review*, 56(4), 3005–3054. <https://doi.org/10.1007/s10462-022-10246-w>
- Myers, M. D. (2008). *Qualitative Research in Business & Management*. SAGE Publications Ltd. <https://doi.org/10.4135/9781036208417>
- Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and Organization*, 17(1), 2–26. <https://doi.org/10.1016/j.infoandorg.2006.11.001>
- Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, 18(5), 544–551.
- Nemeth, C. J., Personnaz, B., Personnaz, M., & Goncalo, J. A. (2004). The liberating role of conflict in group creativity: A study in two countries. *European Journal of Social Psychology*, 34(4), 365–374. <https://doi.org/10.1002/ejsp.210>
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-hall. <https://psycnet.apa.org/record/1973-10478-000>
- Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems*, 22(3), 336–359. <https://doi.org/10.1057/ejis.2012.26>
- Niederman, F., Beise, C. M., & Beranek, P. M. (1996). Issues and Concerns about Computer-Supported Meetings: The Facilitator's Perspective. *MIS Quarterly*, 20(1), 1–22. <https://doi.org/10.2307/249540>
- Niehaves, B. (2005). Epistemological perspectives on multi-method information systems research. *Proceedings of the 13th European Conference on Information Systems*.
- Niehaves, B. (2007). On epistemological pluralism in design science. *Scandinavian Journal of Information Systems*, 19(2).
- Oeste-Reiß, S., Bittner, E., Cvetkovic, I., Günther, A., Leimeister, J. M., Memmert, L., Ott, A., Sick, B., & Wolter, K. (2021). Hybride Wissensarbeit. *Informatik-Spektrum*, 44, 148–152. <https://doi.org/10.1007/s00287-021-01352-0>
- Oppenlaender, J., Linder, R., & Silvennoinen, J. (2024). Prompting AI Art: An Investigation into the Creative Skill of Prompt Engineering. *International Journal of Human–Computer Interaction*, 1–23. <https://doi.org/10.1080/10447318.2024.2431761>
- Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A., & Sinz, E. J. (2011). Memorandum on design-oriented information systems research. *European Journal of Information Systems*, 20(1), 7–10. <https://doi.org/10.1057/ejis.2010.55>

- Ouyang, L., Wu, J [Jeff], Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C [Chong], Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *NIPS'22: Proceedings of the 36th International Conference on Neural Information Processing Systems*, 27730–27744. <https://doi.org/10.5555/3600270.3602281>
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part a: Systems and Humans*, 30(3), 286–297. <https://doi.org/10.1109/3468.844354>
- Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management*, 52(2), 183–199. <https://doi.org/10.1016/j.im.2014.08.008>
- Park, H., Ahn, D., Hosanagar, K., & Lee, J. (2021). Human-AI Interaction in Human Resource Management: Understanding Why Employees Resist Algorithmic Evaluation at Workplaces and How to Mitigate Burdens. *CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15. <https://doi.org/10.1145/3411764.3445304>
- Patil, S [Sameer], & Kobsa, A. (2005). Uncovering privacy attitudes and practices in instant messaging. In M. Pendergast, K. Schmidt, G. Mark, & M. Ackerman (Eds.), *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work - GROUP '05* (p. 109). ACM Press. <https://doi.org/10.1145/1099203.1099220>
- Patil, S [Sameer], & Lai, J. (2005). Who gets to know what when. In W. Kellogg (Ed.), *CHI 2005: Technology, safety, community* (pp. 101–110). Association for Computing Machinery. <https://doi.org/10.1145/1054972.1054987>
- Paulus, P. B., & Dzindolet, M. T. (1993). Social influence processes in group brainstorming. *Journal of Personality and Social Psychology*, 64(4), 575–586. <https://doi.org/10.1037/0022-3514.64.4.575>
- Peffer, K., Tuunanen, T., & Niehaves, B. (2018). Design science research genres: introduction to the special issue on exemplars and criteria for applicable design science research. *European Journal of Information Systems*, 27(2), 129–139. <https://doi.org/10.1080/0960085X.2018.1458066>
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S [Samir] (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Poser, M., Küstermann, G., Tavanapour, N., & Bittner, E. (2022). Design and Evaluation of a Conversational Agent for Facilitating Idea Generation in Organizational

- Innovation Processes. *Information Systems Frontiers*, 24, 771–796. <https://doi.org/10.1007/s10796-022-10265-6>
- Poser, M., Wiethof, C., & Bittner, E. (2022). Integration of AI into Customer Service: A Taxonomy to Inform Design Decisions. *Proceedings of the 30th European Conference on Information Systems (ECIS 2022)*, 65.
- Prat, N., Comyn-Wattiau, I., & Akoka, J. (2015). A Taxonomy of Evaluation Methods for Information Systems Artifacts. *Journal of Management Information Systems*, 32(3), 229–267. <https://doi.org/10.1080/07421222.2015.1099390>
- Pries-Heje, J., Baskerville, R., & Venable, J. (2008). Strategies for design science research evaluation. *Proceedings of the 16th European Conference on Information Systems*. <https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1214&context=ecis2008>
- Przybilla, L., Baar, L., Wiesche, M., & Krcmar, H. (2019). Machines as Teammates in Creative Teams. In *Proceedings of the 2019 on Computers and People Research Conference* (pp. 94–102). ACM. <https://doi.org/10.1145/3322385.3322402>
- Radziwill, N. M., & Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents. *ArXiv Preprint ArXiv:1704.04579*.
- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I. M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C., Pentland, A. Ô., . . . Wellman, M. (2019). Machine behaviour. *Nature*, 568(7753), 477–486. <https://doi.org/10.1038/s41586-019-1138-y>
- Rai, A. (2020). Explainable AI: from black box to glass box. *Journal of the Academy of Marketing Science*, 48(1), 137–141. <https://doi.org/10.1007/s11747-019-00710-5>
- Raisch, S., & Krakowski, S. (2021). Artificial Intelligence and Management: The Automation–Augmentation Paradox. *Academy of Management Review*, 46(1), 192–210. <https://doi.org/10.5465/amr.2018.0072>
- Rayan, J., Kanetkar, D., Gong, Y., Yang, Y., Palani, S., Xia, H., & Dow, S. P. (2024). Exploring the Potential for Generative AI-based Conversational Cues for Real-Time Collaborative Ideation. *Proceedings of the 16th Conference on Creativity & Cognition*, 117–131. <https://doi.org/10.1145/3635636.3656184>
- Reynolds, L., & McDonell, K. (2021). *Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm*. <http://arxiv.org/pdf/2102.07350>
- Richter, A., & Schwabe, G. (2025). “There is No ‘AI’ in ‘TEAM’! Or is there?” – Towards meaningful human-AI collaboration. *Australasian Journal of Information Systems*, 29. <https://doi.org/10.3127/ajis.v29.5753>
- Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Da Costa, L. F., & Rodrigues, F. A. (2019). Clustering algorithms: A comparative approach. *PLOS ONE*, 14(1), e0210236. <https://doi.org/10.1371/journal.pone.0210236>

- Rothe, H., Wessel, L., & Barquet, A. (2020). Accumulating Design Knowledge: A Mechanisms-Based Approach. *Journal of the Association for Information Systems*, 21(3), 771–810. <https://doi.org/10.17705/1jais.00619>
- Russell, S., & Norvig, P. (2021). *Artificial Intelligence, Global Edition : A Modern Approach*. Pearson Deutschland.
- Rylov, A. (2021). Hybrid Intelligence. Main Concepts and Application Scenarios. *Lecture Notes in Networks and Systems World Organization of Systems and Cybernetics 18. Congress-WOSC2021*, 183–190. https://doi.org/10.1007/978-3-031-08195-8_18
- Sá Siqueira, M. A. de, Müller, B. C. N., & Bosse, T. (2024). When Do We Accept Mistakes from Chatbots? The Impact of Human-Like Communication on User Experience in Chatbots That Make Mistakes. *International Journal of Human-Computer Interaction*, 40(11), 2862–2872. <https://doi.org/10.1080/10447318.2023.2175158>
- Saldaña, J. (2013). *The Coding Manual for Qualitative Researchers* (2nd ed.). SAGE Publications Ltd.
- Sarker, S., Chatterjee, S [Sutirtha], Xiao, X., & Elbanna, A. (2019). The Sociotechnical Axis of Cohesion for the IS Discipline: Its Historical Legacy and its Continued Relevance. *MIS Quarterly*, 43(3), 695–A5. <https://www.jstor.org/stable/26848052>
- Scaife, M., & Rogers, Y. (1996). External cognition: how do graphical representations work? *International Journal of Human-Computer Studies*, 45(2), 185–213. <https://doi.org/10.1006/ijhc.1996.0048>
- Schallmo, D. R., & Lang, K. (2020). *Design Thinking erfolgreich anwenden*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-28325-4>
- Seeber, I., Bittner, E., Briggs, R. O., Vreede, T. de, Vreede, G.-J. de, Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, 57(2), 103174. <https://doi.org/10.1016/j.im.2019.103174>
- Seeger, A.-M., Pfeifer, J., & Heinzl, A. (2017). When do we need a human? Anthropomorphic design and trustworthiness of conversational agents. *Proceedings of the Sixteenth Annual Pre-ICIS Workshop on HCI Research*.
- Sein, Henfridsson, Purao, Rossi, & Lindgren (2011). Action Design Research. *MIS Quarterly*, 35(1), 37. <https://doi.org/10.2307/23043488>
- Seufert, S., & Meier, C. (2023). Hybride Intelligenz: Zusammenarbeit mit KI-Assistenzsystemen in wissensintensiven Bereichen. *HMD Praxis Der Wirtschaftsinformatik*, 60(6), 1194–1209. <https://doi.org/10.1365/s40702-023-01012-9>
- Shahrdar, S., Menezes, L., & Nojournian, M. (2018). A Survey on Trust in Autonomous Systems. *Proceedings of the 2018 Computing Conference*, 368–386. https://doi.org/10.1007/978-3-030-01177-2_27

- Shim, J. P., Warkentin, M., Courtney, J. F., Power, D. J., Sharda, R., & Carlsson, C. (2002). Past, present, and future of decision support technology. *Decision Support Systems*, 33(2), 111–126. [https://doi.org/10.1016/S0167-9236\(01\)00139-7](https://doi.org/10.1016/S0167-9236(01)00139-7)
- Shneiderman, B. (2007). Creativity support tools: accelerating discovery and innovation. *Communications of the ACM*, 50(12), 20–32. <https://doi.org/10.1145/1323688.1323689>
- Shneiderman, B. (2020). Human-Centered Artificial Intelligence: Three Fresh Ideas. *AIS Transactions on Human-Computer Interaction*, 109–124. <https://doi.org/10.17705/1thci.00131>
- Siemon, D. (2022). Let the computer evaluate your idea: evaluation apprehension in human-computer collaboration. *Behaviour & Information Technology*, 1–19. <https://doi.org/10.1080/0144929X.2021.2023638>
- Siemon, D., Elshan, E., Vreede, T. de, Oeste-Reiß, S., Vreede, G.-J. de, & Philipp Ebel (2022). Examining the Antecedents of Creative Collaboration with an AI Teammate. *Proceedings of the 43rd International Conference on Information Systems*. https://aisel.aisnet.org/icis2022/is_futureofwork/is_futureofwork/10
- Siemon, D., & Strohmman, T. (2021). Human-AI Collaboration. In J. Zhao (Ed.), *Collaborative Convergence and Virtual Teamwork for Organizational Transformation* (pp. 105–119). IGI Global. <https://doi.org/10.4018/978-1-7998-4891-2.ch005>
- Siemon, D., Strohmman, T., & Robra-Bissantz, S. (2018). The Virtual Collaborator - A Definition and Research Agenda. *International Journal of E-Collaboration*, 14(4), 24–43. <https://doi.org/10.4018/IJeC.2018100102>
- Silver, D., Singh, S., Precup, D., & Sutton, R. S. (2021). Reward is enough. *Artificial Intelligence*, 299. <https://doi.org/10.1016/j.artint.2021.103535>
- Sonnenberg, C., & vom Brocke, J. (2012). Evaluations in the Science of the Artificial – Reconsidering the Build-Evaluate Pattern in Design Science Research. *Proceedings of the 7th International Conference on Design Science Research in Information Systems: Advances in Theory and Practice*, 381–397. https://doi.org/10.1007/978-3-642-29863-9_28
- Stevenson, C., Smal, I., Baas, M., Grasman, R., & van der Maas, H. (2022). *Putting GPT-3's Creativity to the (Alternative Uses) Test*. <http://arxiv.org/pdf/2206.08932>
- Strohmman, T., Siemon, D., & Robra-Bissantz, S. (2017). brAInstorm: Intelligent Assistance in Group Idea Generation. *Proceedings of the International Conference on Design Science Research in Information Systems and Technology*, 457–461. https://doi.org/10.1007/978-3-319-59144-5_31
- Sudeeptha, I., Müller, W., Richter, A., Leyer, M., & Nolte, F. (2024). Obstacles to Human-AI Collaboration. *Proceedings of the 45th International Conference on Information Systems (ICIS 2024)*.
- Summers-Stay, D., Voss, C., & Lukin, S. (2023). Brainstorm, then select: a generative language model improves its creativity score. *Proceedings of the AAAI-23*

- Workshop on Creative AI Across Modalities.*
<https://openreview.net/forum?id=8hwkaj1wv1>
- Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- Tavanapour, N., & Bittner, E. (2018). Automated Facilitation for Idea Platforms: Design and Evaluation of a Chatbot Prototype. *ICIS 2018 Proceedings.*
<https://aisel.aisnet.org/icis2018/general/Presentations/8>
- Tepper, N., Hashavit, A., Barnea, M., Ronen, I., & Leiba, L. (2018). Collabot: Personalized Group Chat Summarization. *WSDM '18: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 771–774. <https://doi.org/10.1145/3159652.3160588>
- Thuan, H. N., Drechsler, A., & Antunes, P. (2019). Construction of Design Science Research Questions. *Communications of the Association for Information Systems*, 332–363. <https://doi.org/10.17705/1CAIS.04420>
- Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
- Toxtli, C., Monroy-Hernández, A., & Cranshaw, J. (2018). Understanding Chatbot-mediated Task Management. *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–6. <https://doi.org/10.1145/3173574.3173632>
- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, 14(3), 207–222. <https://doi.org/10.1111/1467-8551.00375>
- Traumer, F., Oeste-Reiß, S., & Leimeister, J. M. (2017). Towards a Future Reallocation of Work between Humans and Machines Taxonomy of Tasks and Interaction Types in the Context of Machine Learning. *Proceedings of the 38th International Conference on Information Systems.* <https://doi.org/10.2139/ssrn.3159131>
- Tremblay, M. C., Hevner, A. R., & Berndt, D. J. (2010). Focus Groups for Artifact Refinement and Evaluation in Design Research. *Communications of the Association for Information Systems*, 26. <https://doi.org/10.17705/1CAIS.02627>
- van Aken, J. E. (2004). Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules. *Journal of Management Studies*, 41(2), 219–246. <https://doi.org/10.1111/j.1467-6486.2004.00430.x>
- van Diggelen, J., Barnhoorn, J., Post, R., Sijs, J., van der Stap, N., & van der Waa, J. (2021). Delegation in Human-Machine Teaming: Progress, Challenges and Prospects. *Proceedings of the IHSI 2021. Advances in Intelligent Systems and Computing*, 1322, 10–16. https://doi.org/10.1007/978-3-030-68017-6_2

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 6000–6010. <https://doi.org/10.5555/3295222.3295349>
- Venable, J. (2006). The Role of Theory and Theorising in Design Science Research. *Proceedings of the First International Conference on Design Science Research in Information Systems and Technology*.
- Venable, J., Pries-Heje, J., & Baskerville, R. (2012). A Comprehensive Framework for Evaluation in Design Science Research. *DESRIST'12: Proceedings of the 7th International Conference on Design Science Research in Information Systems: Advances in Theory and Practice*, 7286, 423–438. https://doi.org/10.1007/978-3-642-29863-9_31
- Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: a Framework for Evaluation in Design Science Research. *European Journal of Information Systems*, 25(1), 77–89. <https://doi.org/10.1057/ejis.2014.36>
- Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. *MIS Quarterly*, 37(1), 21–54. <http://www.jstor.org/stable/43825936>
- Venkatesh, V., Brown, S., & Sullivan, Y. (2016). Guidelines for Conducting Mixed-methods Research: An Extension and Illustration. *Journal of the Association for Information Systems*, 17(7), 435–494. <https://doi.org/10.17705/1jais.00433>
- vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. *Proceedings of the 17th European Conference on Information Systems*.
- vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., & Cleven, A. (2015). Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research. *Communications of the Association for Information Systems*, 37. <https://doi.org/10.17705/1cais.03709>
- vom Brocke, J., Winter, R., Hevner, A., & Maedche, A. (2020). Special Issue Editorial – Accumulation and Evolution of Design Knowledge in Design Science Research: A Journey Through Time and Space. *Journal of the Association for Information Systems*, 21(3), 520–544. <https://doi.org/10.17705/1jais.00611>
- von Brackel-Schmidt, C., Kucevic, E., Memmert, L., Tavanapour, N., Cvetkovic, I., Bittner, E., & Böhm, T. (2023). A user-centric taxonomy for conversational generative language models. *Proceedings of the 44th International Conference on Information Systems*. <https://core.ac.uk/download/pdf/590878744.pdf>
- Vössing, M., Köhl, N., Lind, M [Matteo], & Satzger, G. (2022). Designing Transparency for Effective Human-AI Collaboration. *Information Systems Frontiers*, 24(3), 877–895. <https://doi.org/10.1007/s10796-022-10284-3>

- Walls, J. G., Widmeyer, G. R., & El Sawy, O. A. (1992). Building an Information System Design Theory for Vigilant EIS. *Information Systems Research*, 3(1), 36–59. <https://doi.org/10.1287/isre.3.1.36>
- Weber, R. (2004). Editor's Comments: The Rhetoric of Positivism versus Interpretivism: A Personal View. *MIS Quarterly*, 28(1), 3–12. <https://doi.org/10.2307/25148621>
- Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2), xiii–xxiii. <http://www.jstor.org/stable/4132319>
- Wei, K., Li, J., Ding, M., Ma, C., Yang, H. H., Farokhi, F., Jin, S., Quek, T. Q. S., & Vincent Poor, H. (2020). Federated Learning With Differential Privacy: Algorithms and Performance Analysis. *IEEE Transactions on Information Forensics and Security*, 15, 3454–3469. <https://doi.org/10.1109/TIFS.2020.2988575>
- Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45.
- Westland, J. C. (2015). *Structural Equation Models: From Paths to Networks* (Vol. 22). Springer International Publishing. <https://doi.org/10.1007/978-3-319-16507-3>
- Westlund, J. K., Lee, J. J., Plummer, L., Faridi, F., Gray, J., Berlin, M., Quintus-Bosz, H., Hartmann, R., Hess, M., Dyer, S., dos Santos, K., Adalgeirsson, S. O., Gordon, G., Spaulding, S., Martinez, M., Das, M., Archie, M., Jeong, S., & Breazeal, C. (2016). Tega: A social robot. *Proceedings of the International Conference on Human-Robot Interaction*, 561. <https://doi.org/10.1109/hri.2016.7451856>
- Wiethof, C., Tavanapour, N., & Bittner, E. (2020). Design and Evaluation of a Collaborative Writing Process with Gamification Elements. *Proceedings of the 28th European Conference on Information Systems (ECIS)*. https://aisel.aisnet.org/ecis2020_rp/43
- Wilde, T., & Hess, T. (2006). *Methodenspektrum der Wirtschaftsinformatik: Überblick und Portfoliobildung*. Ludwig Maximilians-Universität München, Institut für Wirtschaftsinformatik und Neue Medien (WIM).
- Winkler, R., Söllner, M., & Leimeister, J. M. (2021). Enhancing problem-solving skills with smart personal assistant technology. *Computers & Education*, 165, 104148. <https://doi.org/10.1016/j.compedu.2021.104148>
- Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering* (pp. 1–10). ACM. <https://doi.org/10.1145/2601248.2601268>
- Wolfswinkel, J. F., Furtmueller, E., & Wilderom, C. P. M. (2013). Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems*, 22(1), 45–55. <https://doi.org/10.1057/ejis.2011.51>

- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 17(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Wu, T. S., Terry, M., & Cai, C. J. (2022). AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. *CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–22. <https://doi.org/10.1145/3491102.3517582>
- Xu, H., Dinev, T., Smith, H. J., & Hart, P. (2008). Examining the formation of individual's privacy concerns: Toward an integrative view. *ICIS 2008 Proceedings*. <https://aisel.aisnet.org/icis2008/6/>
- Xu, W. (2019). Toward human-centered AI. *Interactions*, 26(4), 42–46. <https://doi.org/10.1145/3328485>
- Xu, W., Dainoff, M. J., Ge, L., & Gao, Z. (2023). Transitioning to Human Interaction with AI Systems: New Challenges and Opportunities for HCI Professionals to Enable Human-Centered AI. *International Journal of Human-Computer Interaction*, 39(3), 494–518. <https://doi.org/10.1080/10447318.2022.2041900>
- Xu, W., & Gao, Z. (2024). Enabling Human-Centered AI: A Methodological Perspective. *Proceedings of the 2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS)*, 1–6. <https://doi.org/10.1109/ICHMS59971.2024.10555771>
- Xu, W., Gao, Z., & Ge, L. (2024). New research paradigms and agenda of human factors science in the intelligence era. *Acta Psychologica Sinica*, 56(3), 363–382. <https://doi.org/10.3724/SP.J.1041.2024.00363>
- Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. *Proceedings of the 2020 CHI*, 1–13. <https://doi.org/10.1145/3313831.3376301>
- Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023). Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. *CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–21. <https://doi.org/10.1145/3544548.3581388>
- Zhang, C [Chao], Yao, C., Liu, J., Zhou, Z., Zhang, W., Liu, L., Ying, F., Zhao, Y., & Wang, G. (2021). StoryDrawer: A Co-Creative Agent Supporting Children's Storytelling through Collaborative Drawing. *Proceedings of the CHI EA '21: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–6. <https://doi.org/10.1145/3411763.3451785>
- Zhou, J., & Hu, Y. (2024). Beyond Words: Infusing Conversational Agents with Human-like Typing Behaviors. *Proceedings of the 6th ACM Conference on Conversational User Interfaces*, 1–12. <https://doi.org/10.1145/3640794.3665560>
- Zhu, Q., & Luo, J. (2022). Generative Pre-Trained Transformer for Design Concept Generation: An Exploration. *Proceedings of the Design Society*, 2, 1825–1834. <https://doi.org/10.1017/pds.2022.185>

Eidesstattliche Versicherung

Hiermit versichere ich, Izabel Cvetkovic, geboren am 08.08.1994 in Nis, Serbien, an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben. Sofern im Zuge der Erstellung der vorliegenden Dissertationsschrift generative Künstliche Intelligenz (gKI) basierte elektronische Hilfsmittel verwendet wurden, versichere ich, dass meine eigene Leistung im Vordergrund stand und dass eine vollständige Dokumentation aller verwendeten Hilfsmittel gemäß der Guten wissenschaftlichen Praxis vorliegt. Ich trage die Verantwortung für eventuell durch die gKI generierte fehlerhafte oder verzerrte Inhalte, fehlerhafte Referenzen, Verstöße gegen das Datenschutz- und Urheberrecht oder Plagiate.

Affidavit

I, Izabel Cvetkovic, born on 08.08.1994 in Nis, Serbia, hereby declare and affirm that this doctoral dissertation is my own work and that I have not used any aids and sources other than those indicated. If electronic resources based on generative artificial intelligence (gAI) were used in the course of writing this dissertation, I confirm that my own work was the main and value-adding contribution and that complete documentation of all resources used is available in accordance with good scientific practice. I am responsible for any erroneous or distorted content, incorrect references, violations of data protection and copyright law or plagiarism that may have been generated by the gAI.

Hamburg, 12.05.2025

Ort, den / City, date



Unterschrift / Signature

Erklärung zur Identität der elektronischen und gedruckten Versionen der Dissertation

Ich versichere, dass dieses gebundene Exemplar der Dissertation und das in elektronischer Form eingereichte Dissertationsexemplar (über den Docata-Upload) und das bei der Fakultät (zuständiges Studienbüro bzw. Promotionsbüro Physik) zur Archivierung eingereichte gedruckte gebundene Exemplar der Dissertationsschrift identisch sind.

Declaration on the identity of the electronic and printed dissertation versions

I, the undersigned, declare that this bound copy of the dissertation and the dissertation submitted in electronic form (via the Docata upload) and the printed bound copy of the dissertation submitted to the faculty (responsible Academic Office or the Doctoral Office Physics) for archiving are identical.

Hamburg, 12.05.2025

Ort, den / City, date



Unterschrift / Signature