From Epistemology to Practice:

How Datafication and AI Shape Inequities in Journalism

Universität Hamburg

Fakultät für Wirtschafts- und Sozialwissenschaften

Dissertation

Zur Erlangung der Würde einer Doktorin der

Wirtschafts- und Sozialwissenschaften

„Dr. phil." (gemäß der PromO vom 18. Januar 2017)

vorgelegt von Nadja Schaetz

aus Berlin (Geburtsort)

Hamburg, den 31.01.2024

A dissertation submitted for the degree of Doctor of Philosophy (Dr. phil.) at Universität Hamburg, Faculty of Business, Economics and Social Sciences, by Nadja Schaetz.

Erstgutachterin (First Supervisor)

Prof. Dr. Juliane Lischka
Universität Hamburg
Fakultät für Wirtschafts- und Sozialwissenschaften Sozialwissenschaften
Journalistik und Kommunikationswissenschaft

Zweitgutachterin (Second Supervisor)

Prof. Dr. Alexa Robertson
Stockholm University
Department of Media and Communication Studies

Vorsitzende (Chairperson of the Doctoral Committee)

Jun.-Prof. Dr. Judith Möller
Universität Hamburg
Fakultät für Wirtschafts- und Sozialwissenschaften Sozialwissenschaften
Journalistik und Kommunikationswissenschaft

Disputationsdatum (date of the doctoral defense): 15.08.2024

Summary

Datafication and artificial intelligence have changed the ways journalists gather information, decide what stories they should pursue, how to produce and distribute news, and commodify products. Technological innovations have been used in journalism to hold power to account, expose injustices, and understand audiences' diverse information needs and interests. However, they also have implications for power relationships in newsrooms, between news organizations, between news organizations and tech corporations, and between journalists and their audiences. Against this backdrop, this dissertation asks how datafication and AI structure journalism's *ways of knowing* (epistemology) and *ways of doing* (practice), and what their implications for equity are. The central argument is that inequalities and socio-technological change in journalism are not separate issues. Rather, socio-technological change should be understood as shaping power dynamics, as it restructures journalism's way of *knowing* and *doing*. In some contexts, this has created new opportunities to address and reduce inequities in journalism. In others, it has amplified existing power asymmetries. The dissertation contributes to a more thorough understanding of the underlying logics and conditions of these processes. Empirically, it comprises four studies, using semi-structured interviews, ethnographic observations, as well as qualitative and quantitative textual analysis. It draws on experiences and perspectives of 87 news professionals situated in 27 countries in Europe, the Americas, and the APAC region. Applying a critical data perspective, the work explicitly foregrounds questions of power and discusses socio-technological change in journalism on three levels: the individual, the organizational, and the socio-institutional.

Zusammenfassung

Datafizierung und künstliche Intelligenz haben die Art und Weise verändert, wie Journalist:innen Informationen sammeln, wie sie entscheiden, welchen Themen sie sich widmen und wie Nachrichten produziert und verbreitet werden. Technologische Innovationen werden im Journalismus eingesetzt, um Machtmissbrauch mit Hilfe von investigativen Recherchen aufzudecken oder um unterschiedliche Informationsbedürfnisse und Interessen zu verstehen und zu decken. Technologischer Wandel hat jedoch auch Auswirkungen auf Machtverhältnisse in den Redaktionen, zwischen Nachrichtenorganisationen und Technologieunternehmen, sowie zwischen Journalist:innen und ihrem Publikum. Vor diesem Hintergrund geht diese Dissertation der Frage nach, wie Datafizierung und künstliche Intelligenz "*ways of knowing"* (Epistemologie) und *"ways of doing"* (Praxis) im Journalismus strukturieren und welche Auswirkungen dies auf Machtverteilungen hat. Das zentrale Argument ist, dass Ungleichheiten im Journalismus und sozio-technologischer Wandel keine getrennten Themen sind. Vielmehr sollte sozio-technologischer Wandel als Umstrukturierung von Machtdynamiken verstanden werden, die mit Veränderungen in den journalistischen Arbeits- und Wissensweisen verbunden ist. In manchen Kontexten hat dies bestehende Machtasymmetrien verstärkt, während es gleichzeitig neue Möglichkeiten geschaffen hat, Ungleichheiten im Journalismus zu adressieren und zu mindern. Die Dissertation trägt zu einem tieferen Verständnis der zugrunde liegenden Logiken und Bedingungen dieser Prozesse bei. Empirisch umfasst die Dissertation vier Studien, in denen halbstrukturierte Interviews, ethnografische Beobachtungen sowie qualitative und quantitative Textanalysen durchgeführt wurden. Sie stützt sich auf die Erfahrungen und Perspektiven von 87 Journalist:innen in 27 Ländern in Europa, Amerika und der APAC-Region. Unter Anwendung einer "critical data" Perspektive stellt die Arbeit explizit Machtfragen in den Vordergrund und erörtert sozio-technologischen Wandel im Journalismus auf drei Ebenen: der individuellen, der organisatorischen und der sozio-institutionellen.

List of Publications

The following publications are referred to throughout the dissertation by their Roman numerals.

Study I

Nadja Schaetz (2023) Journalism & Audience Datafication: *How Audience Data Practices Shape Inequity,* Digital Journalism, DOI: 10.1080/21670811.2023.2171894

Study II

Nadja Schaetz, Juliane A. Lischka & Laura Laugwitz (2023) *Datafication of Journalism: How Data Elites and Epistemic Infrastructures Change News Organizations,* Digital Journalism, DOI: 10.1080/21670811.2023.2235601

Study III

Nadja Schaetz & Anna Schjøtt Hansen (forthcoming). *Journalism and AI: Macro Epistemics, Inequities & A Global Journalistic Coalition.*

Study IV

Nadja Schaetz, Laura Laugwitz & Juliane A. Lischka (2023). S*olving journalism with data: Silicon Valley's influence on the Fourth Estate.* Journalism, 0(0). https://doi.org/10.1177/14648849231176110

Acknowledgements

I owe debts of gratitude to both of my advisors, Juliane Lischka and Alexa Robertson. Thank you, Juliane, for your generosity, encouragement, and guidance. I am grateful for the academic freedom and the opportunities you have created for me, without which this dissertation would not have been possible. Thank you, Alexa, for your unwavering support, and for carefully engaging with my thinking and writing before I even started my PhD. This work began years ago in Stockholm, in her office and at her kitchen table. After my time in Sweden, I moved back home to Berlin, where I worked at the Weizenbaum Institute for the Networked Society and Freie Universität Berlin. Thank you to Martin Emmer, to whom I am tremendously appreciative for this opportunity. My former colleagues at the Weizenbaum Institute shaped my thinking on technology in formative ways, for which I am especially grateful. At the Universität Hamburg, where my dissertation eventually found its academic home, I was fortunate to work in a rich environment of journalism scholars. I owe all of you my sincere gratitude and would like to thank the New Frontiers research seminar, for offering a space to discuss my work at preliminary stages. Thank you also to Jannie Møller Hartley, Ida Willig, and David Cheruiyot for providing feedback on individual papers during the writing process, and for the anonymous reviewers who have provided numerous inputs that have sharpened my focus. Thank you to Yanthe Zebregs and María de la Huerga for research assistance, and Annika Schütz and Corinna Kienle for administrative help. Further, I would like to express my gratitude to Florencia Enghel. While we are not living across the street from each other in Stockholm anymore, I am grateful for all the times you have shown me support from afar. Moving from Scandinavia to the Sonoran Desert, I owe special thanks to Jacob Nelson for inviting me to the Walter Cronkite School of Journalism and Mass Communication as a visiting scholar. This dissertation has greatly benefited from my time in Phoenix and your input, and I feel lucky to have been welcomed by you and your family with such open arms. For being thoughtful, supportive colleagues and friends, I dearly thank (in no particular order) Emilija Gagrčin, Hanna Hallin, Anna Schjøtt Hansen, and Laura Laugwitz. This journey has been much more rewarding because of you. Anna and Laura, I am especially grateful for your contributions to this dissertation as co-authors, and for the joy of thinking and working with you. For caring, strategizing, writing next to me, and most importantly, all the

years of friendship, thank you to Jonas Fegert. My deepest gratitude goes to the many journalists who have spent their time in conversation with me and trusted me to adequately represent their experiences and perspectives. While I will not mention you by name to protect your anonymity, you know who you are. For walking alongside me on this journey (the Rilkean way), thank you to Vicky. Finally, for everything, thank you to my father, Gregor, and my mother, Gabriele.

Table Of Contents

Introduction 1

On Datafication 3

Different Academic Approaches to Studying Datafication 10

Ways of Knowing 12

Ways of Doing 14

Connecting Datafication to AI 17

A Critical Data Perspective 20

A Critical Data Studies Perspective in Journalism Studies 23

Methodological Approach 27

Levels of Analysis 29

The People, Organizations, and Worlds in Focus 31

Reflections on the Methodological Implications of Qualitative Research during the COVID-19 Pandemic 35

General Discussion 37

Pathways to Mitigate Inequities 40

Contributions 42

Avenues for Future Research 46

References 49

Appendix I: Erklärung 73

Appendix II: Eidesstattliche Versicherung 74

Appendix III: Selbstdeklaration bei kumulativen Promotionen 75

Empirical Studies 77

Introduction

Data and AI practices have been linked to the exacerbation of social inequalities (Couldry & Mejias, 2019; Eubanks, 2017; Noble, 2018). Biases encoded in technology may have discriminating effects, compounding existing power asymmetries with implications for various domains of life. A biased data set may lead to unjust policing of racial minorities (Sankin & Mattu, 2023) or disparages in mortgage approvals (Martinez & Kirchner, 2021). Algorithmic systems can create segregating effects in education (Lecher & Varner, 2021) and further propagate societal discrimination in healthcare (Norori et al., 2021). The risks of datafication and AI have been shown to overwhelmingly fall on resource-poor and marginalized groups in society (Benjamin, 2019; D'Ignazio & Klein, 2023; Eubanks, 2017). However, overarching concerns are collective, such as the erosion of privacy through processes of datafication that implicate human dignity and autonomy (Nissenbaum, 2010; Renieris, 2023), and the creation of power asymmetries between "Big Data rich" corporations that own and control information and the "Big Data poor", who lack the means to extract, analyze and profit from datafication (boyd & Crawford, 2012).

This also has implications for journalism. As news organizations are adopting technological tools for journalistic pursuits, they face ethical issues regarding accountability, transparency, and privacy (Dörr & Hollnbuchner, 2017; Helberger et al., 2022; Paik, 2023; Porlezza, 2023) that risk amplifying pre-existing inequalities. Facial recognition, for instance, has been used to aid investigative reporting, analyze visual content and determine the identity of sources, but presents a significant threat to privacy rights, and moreover, may be built on algorithms that already exhibit bias and thus lead to false interpretations (van Ess, 2023). At the same time, data-driven journalism can be used to hold power to account. Reporting such as the Buzzfeed News investigation identifying a vast infrastructure built by the Chinese government for the mass detention of Uyghurs, would not have been possible without computational efforts enabling the team to analyze thousands of satellite images (Rajagopalan et al., 2020; Tactical Tech, 2022). AI tools can also be used to give voice to underrepresented communities, for example, by creating and analyzing transcripts of local civil council meetings that are not covered on site by reporters, but can, through these means, reemerge in newsrooms again (Associated Press, 2023).

On a practical level, it is crucial for journalists to differentiate between socially beneficial and harmful applications and practices and develop strategies for ethical guardrails. However, separations between "good" and "bad" are not always clear-cut. Such distinctions are often difficult to maintain, obfuscate the lines between what counts as productive or destructive (Aradau & Bunz, 2022), and distract from the broader, underlying assumptions and logics of datafication and AI that might underpin the reinforcement of inequity. Because AI is not just a technology — but is restructuring worldviews and epistemologies (Katz, 2020) — it requires an examination from a wider, societal context.

With this in mind, this dissertation is concerned with the overarching question: How does datafication and AI structure journalism's ways of *knowing (epistemology)* and *doing (practice),* and what are its implications for equity? When using the term equity, I have multiple sites of equity in mind. I think of inequality among news professionals, but also of power asymmetries between journalists and their audiences. I recognize journalism as a site for voice (Couldry, 2010), in which communities and their needs can be recognized or marginalized. Lastly, when speaking of equity, I am interested in power asymmetries between journalism and external actors, specifically tech corporations.

At the core of this dissertation are four empirical studies. Study I uses audience data practices as an exemplary case to examine ways of *doing*. The primary question is: *How are data practices in journalism linked to both the reproduction and mitigation of social inequities?* In Study II, I turn my attention to journalism's way of *knowing* and examine the epistemic infrastructures underpinning datafication in journalism. The study asks: *how do data materialities, interlinkages, and paradigms shape epistemic infrastructures in journalism?* With this question, I aim to further our understanding of power asymmetries between editorial technologists and editorial staff. In Study III, I examine a global journalistic training programme to look at both power asymmetries between news organizations, as well as between news organizations and Big Tech, and ask: *what are the hierarchies of power in the field of journalism and AI, and what are its implications for the promotion and validation of epistemic practices in journalism?* Lastly, in Study IV, I interrogate the power asymmetries between Big Tech and journalism from a historical

perspective, asking: *How do Silicon Valley and Fourth Estate ideals converge in contemporary journalistic places?*

In the following sections I situate the dissertation in existing scholarship and outline research on datafication from different disciplines (with a main focus on sociology, media, and communication research) and connect them to the journalistic context to motivate my analytical focus on "ways of knowing" and "ways of doing". I argue for the relevance of a critical data perspective in journalism research before I draw connections between datafication and AI.

On Datafication

Scholars from diverse fields within the social sciences have approached datafication as a contemporary phenomenon from various viewpoints and arrived at different definitions and theoretical conclusions (boyd & Crawford, 2012; Couldry & Hepp, 2017; Mayer-Schönberger & Cukier, 2014; Sadowski, 2019; Zuboff, 2019). In this section, I give an overview of these perspectives, contextualize their scholarly arguments, and position this dissertation within the debates. The term datafication was first introduced by Mayer-Schönberger and Cukier (2013), putting forward a definition that encompasses the collection, databasing, quantification, and analysis of information for knowledge production and economic value-generation. For Mayer-Schönberger and Cukier, datafication is primarily a technological process with economic implications that follows a specific pattern of standardization, typification, personalization, optimization, and, ultimately, commodification of data (ibid.). While datafication precedes digitalization — societies have long been engaged in datafication by registering, analyzing, archiving, and indexing to administer and manage populations — the term now mainly refers to digital technology and big data (Thylstrup, 2021). What makes datafication today different from its historical counterpart is the automation of these historical practices, through digital systems that fuel and intensify data extraction and analysis (Flensburg & Lomborg, 2023).

After Mayer-Schönberger and Cukier (2013) coined the term in their seminal work, scholarly attention to datafication picked up significantly, especially in humanistic and social

sciences (Flensburg & Lomborg, 2023, p. 1452). In response to technological focused conceptualizations, critical scholars argued for an understanding of datafication as a socio-technological process. Building on boyd and Crawford's (2012) article on critical questions for big data studies, van Dijk (2014) suggested that datafication is linked to a mythology and logic of dataism, according to which data is viewed as a raw material, ready to be extracted and profited from. For van Dijk, datafication is tied to a wide-spread privately and publicly held belief in the "objective quantification and potential tracking of all kinds of human behavior" through digital systems (p. 198). In this line of thought, datafication has ideological underpinnings that proclaim that every aspect of our lives can and should be translated into data, very often for economic profit (ibid.).

That is, a socio-technological approach to datafication differs from perspectives that view datafication primarily as a technological process, by acknowledging that it is also a social phenomenon that has implications for the way people act and generate knowledge. Conceptualizations of datafication can further be understood as a continuation of other academic debates and discourses, such as mediatization, capitalism, colonialism, and quantification. I will briefly review this scholarship to show how different conceptual approaches allow for different theoretical boundaries that help contextualize the approach of this dissertation.

For Couldry and Hepp (2017), datafication is a late stage of mediatization, which they have coined "deep mediatization". Mediatization as a theoretical lens is used by social sciences and cultural studies to encapsulate the interconnected and dialectical changes within media and society. In this line of thought, transformations in media can be sorted into historical waves of mediatization. Deep mediatization, specifically, signifies the contemporary integration of the analysis of algorithms, data and digital infrastructures. It underscores the accelerated nature of these transformations, through automated data processes, and emphasizes the significance of media in shaping societies and cultures.

Academic discourses concerned with capitalism have further theorized datafication as a form of digital or platform capitalism, informational capitalism, or surveillance capitalism (Cohen, 2015; Fuchs, 2010; Liang et al., 2022; Sadowski, 2019; West, 2019). What these lines of scholarly inquiry have in common, is their foundational argument that datafication is a central element of

contemporary capitalism today that promotes the extraction and accumulation of data from all spheres of life and turns them into capital. According to this line of thought, datafication can be understood as a political economic regime (Sadowski, 2019). Relatedly, scholars have coined concepts such as "platform capitalism" (Srnicek, 2017), "informational capitalism" (Fuchs, 2010), and "surveillance capitalism" (Zuboff, 2019) to highlight different dynamics and societal implications of datafication. While informational capitalism underscores the broader significance of information as a commodity (Arvidsson & Colleoni, 2012), the concept of platform capitalism relates to the specific role of digital platforms as central players in mediating economic activities and social interactions (Törnberg, 2023).

Zuboff's (2019) conceptualization of datafication focuses on the role of surveillance. Surveillance capitalism as a theoretical lens, underscores the profound shift in the logic of accumulation, where companies intentionally engage in extensive and systematic monitoring of individuals' online behavior without their full awareness or informed consent. It highlights the primary objective, which is to extract, analyze, and utilize vast amounts of personal data to predict and influence human behavior and preferences. This predictive power serves as a means to generate revenue and exert control over markets. For West (2019), commercial surveillance is enacted through a logic of data capitalism and justified by narratives that highlight the social and political benefits of networked technologies. West, like critical scholars such as the aforementioned Sadowski (ibid.) and Zuboff (ibid.), is concerned with datafication enabling unequal distributions of power that is weighted toward the actors who have access and the capability to use data for their benefit (West, 2019).

In response to prevalent notions of datafication, as exemplified in the works of scholars such as Zuboff (2019) and Sadowski (2019), Delfanti (2021) challenges an understanding of datafication as the "expropriation" or "extraction" of data from passive users. The author contends that these perspectives overlook the active engagement of individuals involved in the processes subjected to datafication and subsequent incorporation into technological systems. His analysis highlights the influence of various forces compelling individuals to participate in and collaborate with the data collection processes (Delfanti, 2021, p. 40). While Delfanti emphasizes the role of individuals in datafication, Burrell and Fourcade (2021) point to the broader societal shifts

influenced by datafication and automation. Similarly to West (2019), they are concerned with societal power asymmetries and logics of capitalism, arguing that societies under datafication and automatisation are divided into coding elites and cyberteriat. Datafication here, is the fundament of the emergence of a new occupational class that consolidates power through technical control over the means of digital production and labor extraction from a marginalized cybertariat (Burrell & Fourcade, 2021).

Shedding light on longer, historical roots of datafication and global power asymmetries, some critical data scholars have argued for an understanding of datafication as "data colonialism" (Couldry & Mejias, 2019; Mumford, 2022; Thatcher et al., 2016). By highlighting connections between the historical colonialism and the contemporary digital practices of data extraction and domination, data colonialism describes the ways in which powerful entities, typically from the West, extract, control, and exploit data for economic benefit and social control. In this view, datafication as a social phenomenon normalizes the exploitation of human life through data similarly to how historic colonialism appropriated territory and resources and governed subjects for profit.

Scholarship on mediatization, capitalism, and colonialism aid in orienting datafication within historical processes through different theoretical lenses. Mediatization scholarship is helpful in elevating the role of media in processes of datafication, while scholarship on data capitalism and data colonialism foregrounds how political and economic forces implicate social dimensions of datafication and produce power asymmetries. Against this backdrop, I turn to sociologies of quantification by Espeland & Stevens (2008) to outline how previous conceptualizations of quantification can add to our understanding of datafication today. These are, in my view, helpful in understanding the datafication of an industry or profession, such as journalism, on an individual, organizational, and societal level.

Espeland & Stevens (2008) define quantification broadly as the production and communication of numbers. Building on social science and humanistic scholarship, they suggest five sociological dimensions of quantification: work, reactivity, discipline, authority, and aesthetics. In what follows, I discuss these dimensions, with a particular focus on work, authority,

and reactivity, which I consider particularly relevant in the context of this dissertation, and I connect them with journalism scholarship.

Concerning the dimension of *work*, Espeland & Stevens suggest that quantification *produces* work. Because "it is work that makes other work possible", quantification requires "quantitative laborers" (Espeland & Stevens, 2008, p. 411). This matters, as it changes the makeup of industries. News organizations have increasingly sought out talent with computational skills in a response to technological advances (Guo & Volz, 2019; Usher, 2016). Many journalists developed new skills to adapt to changes in the industry, and, at the same time, many professionals with a tech-background and no experience in journalism entered the profession. As a result of this, new professional groups have emerged and challenged previous conceptions of journalism. Working at the intersection of traditional journalistic positions and technologically-intensive work that were once separate (Kosterich, 2020), these professionals have taken on a variety of tasks and job titles, such as data specialists, developers, or designers. While there is often a lack of cohesiveness in how emergent positions in journalism are defined and how they operate (Ferrer-Conill & Tandoc, 2018, p. 436), previous work has identified this novel interdisciplinary group in journalism and conceptualized them as "editorial technologists" (Lischka et al., 2022). That is, datafication requires a significant amount of labor, and with it, the expertise of professionals. It is work that is "embedded in larger social projects" (Espeland & Stevens, 2008, p. 411). Meaning, it fosters the growth of certain professions in industries and signals their legitimacy (ibid.). This connects to the following dimension, authority.

For Espeland and Stevens, *authority* of quantification may be vested, for example, in our sense of the accuracy or validity of quantification "as representations of some part of the world" and "in their usefulness in solving problems" (2008, p. 417). In the journalistic context, the authority of numbers lies in their perceived accuracy, validity, and usefulness to provide meaningful solutions for journalism. This, in theory, would grant professionals working with data authority and agency. Prior work suggests that new actors in journalism that have computational skills struggle for authority in news organizations and strive for more agency (Christin, 2020; Petre, 2015; Lischka, 2021; Kosterich, 2020). Recent work, however, demonstrates that as journalistic processes in news organizations evolve (e.g. news distribution shifts from audience

segments to individual personalization via algorithmic recommender systems) "much power and agency changes hands, as decisions concerning the future of news are ultimately moved from the newsroom and out of the hands of the editors and journalists into the design room and the hands of data scientists, developers and algorithmic models" (Schjøtt Hansen & Hartley, 2021, p. 13). Relatedly, studies examining which skills journalists themselves value found that while traditional journalistic stills are still respected, "digital" and "technology skills" increasingly gain importance (Royal & Kiesow, 2021), resulting in pressure for journalists to adapt digital skills (Koivula et al., 2023; Min & Fink, 2021). This suggests that new emergent groups in journalism with computational skills might initially struggle with the "internalized logic that editorial requests prevail", while their perspectives are automatically subordinated. However, they establish authority over time and gain increasing agency once their practices are normalized, as innovation is more and more accepted (Lischka et al., 2022, p. 1037). I will elaborate on these questions of authority in Study II. This leads us to the next dimension: reactivity.

Reactivity in Espeland and Stevens' sociology of quantification relates to the notion that "measurement intervenes in the social worlds it depicts. Measures are reactive; they cause people to think and act differently" (2008, p. 412). In the journalistic context, the quantification of audiences' and journalists' engagement with audience data is one example of reactivity. A preeminent theme in the literature on audience analytics suggests that journalists' relationship towards audiences have changed with technologies that have granted news workers an unprecedented capacity to monitor, assess, and quantify audience online behavior and engagement with news products (Cohen, 2015, p. 108; Tandoc, 2014; Zamith, 2018). Many journalists say that these affordances have enhanced their understanding of how to align with audience interests and preferences, utilizing audience data as an indicator of public opinion (Beckers et al., 2021, p. 13). Others argue that real time analytics of audiences' individual preferences are often difficult to disregard for news professionals, even when this form of instant feedback is not always welcomed (Belair-Gagnon et al., 2019; Petre, 2021). On one hand, this has caused concerns about conflations and disconnects between actual information needs of audiences and the interpretation of datafied representations of audiences (Livingstone, 2019), especially given the lack of transparency in news platforms that make audience composition opaque (Carlson, 2018, p. 408; Christin, 2020). One line of research has argued that audience analytics incentivise news practices to gain audience

attention for reader revenue (Myllylahti, 2018, 2020, p. 567; T. Wu, 2017). However, other scholarship suggests that access to audience analytics has altered previous journalistic perspectives of audiences as mostly passive, and narrowed approaches to "objectivity" that privileges elite voices and often biases marginalized communities, activists, and non-official sources (Nelson, 2021, pp. 3–4). Whether or not audiences have ultimately benefited from these changes depends on different layers of economic, cultural and political concerns, as I showcase in Study I. Collectively, these studies underscore Espeland and Stevens' (2008) notion of reactivity. Data has changed how journalists understand their audiences, which has implications for their practice.

Discipline, for Espeland and Stevens, refers to the idea that "quantification permits remote parties to check on people and things they wish to control" (2008, p. 415). It makes formerly invisible people, objects, and actions visible and enables scrutiny and judgment (ibid). In the journalistic context, discipline is relevant to both the labor discipline of journalists and audiences' privacy and control over their data. Regarding the former, Petre (2021) demonstrates how newsroom analytics tools to measure journalistic performance can be considered a new form of managerial surveillance and labor discipline. Specifically, Petre explains how discursive strategies and design elements embedded in analytic tools engineer journalists' consent to discipline in a way that obfuscates, rather than highlights, managerial influence in the newsroom. In regards to audiences, Kammer's (2021) study on resource exchanges between mobile news apps and third-parties aids in understanding how quantification permits remote parties to observe audiences. In this way, his work connects Espeland and Stevens' notion of discipline with privacy concerns, by showcasing how news users can hardly avoid that their data is ending up in the hands of companies such as Google, Amazon, Facebook, and Apple (ibid.). We know little about the degree to which infringement of audiences' privacy results in audiences changing their behavior. Previous studies give us some indication, suggesting that the influence is limited. For example, users' privacy concerns do not necessarily deter them from using platforms for news use (for an overview see: Barth & De Jong, 2017), and they are not a significant factor influencing user attitudes toward news personalization (Bodó et al., 2019). That is, audiences are using platforms for news consumption, *despite* their privacy concerns (Schaetz et al., 2023). I will elaborate more on the implications of quantification and commodification of audience data in Study I, and provide further insights into news professional's thoughts on audience privacy in Study III.

Lastly, Espeland and Stevens demonstrate that quantification has an important *aesthetic* aspect that follows two ideals: clarity and parsimony: "Regardless of temporal and disciplinary context, people who make pictures with numbers typically prize representations whose primary information is easily legible (clarity), and which contain only those elements necessary and sufficient for the communication of this primary information (parsimony)" (p. 423). What the authors call "numerical pictures", such as graphs, tables or other visualizations of data, matter to journalistic pursuits because they make large phenomena quickly accessible (p. 428). Data visualizations can give forms and language to things that are difficult to understand. However, they might also displace audience attention from editorial choices, when they are perceived as objective, scientific, and neutral (D'Ignazio & Klein, 2023, p. 82). Data visualizations necessarily highlight some things, while obscuring others, and editorial choices embedded in them inherently have a political dimension (ibid.). It is beyond the scope of this dissertation to explore the role of *aesthetics* in the datafication of journalism. However, the value of clarity and parsimony can inform our conceptualization of how datafication can generate value, which I will contextualize in Study I by demonstrating how data can be understood as an economic, cultural, and political good.

In conclusion, quantification theory can be translated back into contemporary journalistic questions around datafication. This helps us to understand the continuation of 1), scholarly debate, and 2) the process of what was formerly conceptualized as quantification and is now widely discussed as datafication. By extension, this orientation should also be useful in our understanding of AI (a link I will elaborate on further in section "Connecting Datafication to AI"). In what follows, I will briefly discuss different approaches to studying datafication and clarify the approach of this dissertation, which is closely connected to Espeland and Stevens' (2008) notion of reactivity.

Different Academic Approaches to Studying Datafication

Different strands of datafication research address the phenomenon in a variety of ways reflecting different scholarly traditions. Prietl and Houben (2018), for instance, offer a sorting of datafication

scholarship into three categories: 1) the technical artifacts and infrastructures under study (which, in the context of this dissertation, I often refer to as materialities), 2) the meaning of data, and 3) the cultural knowledge structure of society. Similarly, Flensburg & Lomborg's (2023) literature review identified two overarching streams of datafication scholarship: 1) a focus on material technological infrastructure (e.g., the implementation and use of data-driven systems and data mining techniques in organizations) and 2) critical, humanistic, and social science-based studies that scrutinize the uses, imaginaries, and social consequences of datafication. This dissertation is mainly rooted in the latter approach, *focusing on how datafication shapes ways of knowing and doing.* However, specifically with Study II, which examines epistemic data infrastructures, I aim to connect both perspectives by examining technological materialities in conjunction with social processes. As such, the dissertation is motivated by the belief that we benefit from blurring boundaries between approaches. Our knowledge of technological conditions is strengthened by acknowledging people's experiences and perceptions. Reversely, social consequences are better understood by recognizing how material conditions enable and constrain actions (Flensburg & Lomborg, 2023 p.1464-5).

Pellegrino et al. (2019) critique a pattern that they observe in datafication scholarship. In their view, previous research reflects a constant dichotomic and oppositional binary thinking that produces two lines of work: 1) involving a top-down, "apocalyptic" perspective in which technology is enforced "from above" and used to exert and impose power, specifically for surveillance and 2) a more bottom-up approach, the authors consider as more integrated view on datafication as a locus of new opportunities, benefits and progressive futures "from below" (Pellegrino et al., 2019, p. 90). This is not necessarily my experience. Other works that resist this binary thinking come to mind, such as Petre's (2021) work on data-driven journalism that demonstrates in a nuanced way how digital metrics can both be used for managerial surveillance and discipline in news organizations. While, and at the same, metrics can be leveraged by journalists to advocate for their professional worth and autonomy. Nevertheless, Pellegrino et al.'s (2019) point is taken.

This dissertation, hopefully, provides another example of research that does not fall within the lines of this dichotomy. In my view, research that engages with everyday realities from a

bottom-up approach can still connect with larger power asymmetries that structure those realities (which I aim to showcase in Study I and III). This dissertation is in many ways motivated by a concern about the imposition of power — whether it is the influence of Big Tech on journalistic work, or infringement of audiences' right to privacy under datafication. Nevertheless, this is not to say that I believe that datafication and technological developments inevitably and irreversibly have negative consequences and harm journalism or audiences. Particularly in Study I and III, I highlight positive, innovative data practices and journalists' efforts to work collaboratively to mitigate inequities within the field and engage with AI in a critical and productive manner. New ways to improve journalism should always be welcome, and at the same time, they require critical examination to ensure that innovation aligns with socially responsible ways of doing journalism. Asking how datafication and AI structure ways of *knowing* and *doing* in journalism enables me to approach socio-technological change in a way that allows for both the recognition of technology to be enforced "from above" and used to exert and impose power, as well as to pay attention to new opportunities, benefits, and progressive futures "from below" (Pellegrino et al., 2019, p. 90). In what follows, I review what previous research can tell us about datafication's role in structuring ways of *knowing* and *doing* to contextualize these issues in journalism. I then go on to explain my understanding of the connections between datafication and AI and its theoretical implications.

Ways of Knowing

Datafication structures worldviews and epistemologies (Katz, 2020). For journalism, as a knowledge producing industry with representational power, shifts in journalistic knowledge practices matter, as they can either help or hurt the communities they cover (Carlson, 2023, p. 14). There is a long line of academic scholarship examining how knowledge production changes over time (Kuhn, 2023) and how we can understand the connection between these developments and specific epistemic practices (Latour & Woolgar, 1986). With roots in philosophy, epistemic scholarship has contributed to our understanding of human knowledge, the nature of truth and how we can determine whether a belief corresponds to reality (Audi, 2011; Bittle, 2017; Hikins & Cherwitz, 2011; Sosa et al., 2008). In the context of this dissertation, I am less interested in these classical philosophical debates of epistemology, but rather I deploy a socio-technological perspective to understand how journalistic actors engage specific forms of knowledge production

that datafication enables, how epistemic practices are discussed, negotiated and challenged, and ultimately, whether specific forms of knowledge production are privileged over others. That is, taking a constructivist perspective, I look at the social shaping of knowledge generation, highlighting that epistemic practices are neither natural nor neutral, but formed through ever-evolving social interactions (Carlson, 2023, p. 3). This approach is guided by the assumption that epistemic practices are grounded in normative perceptions, or paradigms that structure what should be known, why it should be known and how it should be known (Carlson, 2023, p. 1). Specifically, this understanding is closely aligned with a sociology of the knowledge of science and technology, that investigates how actors make knowledge claims that are legitimized on a collective level (Knorr-Cetina, 1999). Unpacking these processes matters because it allows us to understand perceptions embedded in data practices. It challenges what Benjamin (2019) calls "imagined objectivity of data and technology", the culturally shared assumption or false belief that datasets and technology used to process them are less partial or even less discriminatory than people. Because data is never neutral, but always the biased output of unequal social, historical, and economic conditions (D'Ignazio & Klein, 2023, p. 39), ways of *knowing* with data thus requires a critical analysis. Moving beyond an investigation of knowledge practices within a specific local or organizational context, this dissertation examines what Knorr-Cetina calls "macro-epistemics" to grasp ways of *knowing* that are not confined within one site, but traverse them, producing "wider networks of knowledge generation" (2007, p. 361). In doing so, it contextualizes larger "structures, processes, and environments that make up specific epistemic settings" (Knorr-Cetina, 1999, p. 8).

In the journalistic context, Lewis and Westlund (2014) introduced, almost a decade ago, four conceptual lenses — epistemology, expertise, economics, and ethics. The framework is used to understand the significance of datafication, including how it shapes ways of *knowing* through then emerging analytic techniques, and mythology about data-driven insights. Since then, there has been a renewed interest in journalistic epistemology with a growing academic body examining interrelationships between knowledge production and emergent technology. Scholarly work has pointed to "epistemic tensions" that arise in journalistic projects in which journalists with different epistemological backgrounds come together, specifically in the context of data-driven approaches to investigative journalism (Parasie, 2015). Other studies highlight dependencies on external actors in datafied knowledge production that might be subtle or hidden but nonetheless impactful (Møller

Hartley et al., 2023). Within data journalism, Camaj et al. (2023), for instance, demonstrated how data access (or lack thereof) structures not only journalistic epistemology but also power relationships that define what news is. Their study showcases that datafied knowledge generation usually involves external actors that provide databases, which not only shape work routines but constitute foundational constraints on journalistic epistemologies. Another line of research has been interested in the epistemic implications of digital news circulation. Pointing to divergences regarding acceptable forms of public knowledge that emerge in this context, Carlson (2020), for instance, argues that legitimation of journalistic knowledge should be understood as hard fought rather than assumed.

Collectively, this research shows that epistemic practices are at times disputed or have to be re-assessed and re-negotiated, as new forms of knowledge generation are made possible through datafication and automation. This is what Carlson (2023) describes as "epistemic contests". As he explains, epistemic contests highlight "efforts to promote or demote appropriate journalistic forms. This definitional work can be considered an exercise in power, as it assigns certain norms and practices as legitimate and therefore producing acceptable knowledge or as illegitimate" (p. 11). Following this line of thought, issues regarding epistemology become questions about epistemic authority, or, in other words, "the legitimate power to define, describe, and explain" (Gieryn, 1999, p. 1). That is, paying attention to the distribution of power among actors involved in these processes is paramount to a nuanced understanding of journalistic knowledge practices. As new professional communities with specialized, computational knowledge and shared epistemological understandings guide practice (Lischka et al., 2022), both ways of *knowing* and *doing* are impacted. Hence, I argue we have to look at different actors, technological materialities, as well as ways of *knowing* and *doing*, in conjunction, in order to understand transformations in journalism under datafication.

Ways of Doing

Data is used as a main resource to optimize, evaluate, curate or make journalism more efficient (Porlezza, 2023, p. 2). As all domains of journalism become subjected to datafication, data becomes the foundation on which the current transformation of journalism is built, affecting the "use and allocation of resources, production workflows, the selection and the placement of topics,

as well as the format and the style of news (ibid.). Innovation in data practices, from datafication to AI, allow journalists to do things that were previously not possible, such as analyzing audience data in real time, and to do things *differently* — quicker, cheaper, at a larger scale, and potentially with fewer errors than any human journalist (Graefe, 2016). It also allows journalists to stop doing things they previously had to, like manually transcribing interviews, or translating sources, and other tedious tasks. For many, these changes mean that journalists are now freed up to do more sophisticated work (S. Wu et al., 2019). As Döpfner, the CEO of the media company Axel Springer SE, argued in an interview with Kara Swisher and Scott Galloway: "AI can do all the boring work and then journalists can go out and do what they've always done, observe the world, go somewhere and do investigative work, go [do] something very unconventional, do the interesting part of journalism. We can allocate money towards that" (Swisher & Galloway, 2023). According to this view, innovation does not lead to an upheaval of journalistic practices but allows journalists to do things they have "always done". However, at this point, we know very little about whether freed up resources are actually reallocated to in-the-field reporting, and investigative journalism, as suggested by Döpfner, or if changes in news production lead to decreased staffing levels or an expectation for journalists to produce news content more quickly and in larger volumes. As New York Magazine writer John Hermann cautions: "In a moment where many newsrooms are suffering crises of distribution and revenue — problems for which generative AI offers no obvious help — the most obvious use for any form of automation is cost reduction." (Herrman, 2023)

How technological advancements restructure ways of doing in journalism is thus not just a question of what is technologically possible, but how news organizations adapt, what news professionals deem important, how they re-allocate resources, change workflows, and change individual practices. It is, like any scholarly inquiry about power and equity, inevitably also a question about agency — who has agency to decide what to do with data, the extent to which agency is challenged or redistributed under datafication (including between humans and non-human actors), and who has the agency to hold corporations accountable for extracting and using data.

Previous research on datafication and automation in journalism tells conflicting stories about journalists' agency. Editors in Wu's et al. (2019) study, for instance, suppose that technological innovation could increase their agency. In their view, cutting down staff and

increasing content output through automation might give them an economic advantage, which, in turn, would make them less dependent on previous revenue streams and third parties. It is a much more optimistic take than what Simon's (2022) analysis on platform companies' power over news organizations concludes. In Simon's view, economic factors are the forces limiting journalists' autonomy. Mapping ownership and various channels of control, including technology and its infrastructure, he argues that recent data-driven innovations allow third parties to extend control from distribution channels to the means of production as the technology increasingly permeates all stages of the news-making process. To complicate the issue further, De Haan et al. (2022) found that journalists describe AI as a relatively autonomous force, which, for Helberger and her co-authors raises even more questions: How can you "develop a sense of agency over something that is seen as an "invisible hand" or "looming threat" over which one has no control" (p. 1613)?

Taken together, these studies highlight an important challenge. We already know much about how datafication enabled new ways of doing in journalism, for instance, by changing the ways journalists select topics (Diakopoulos, 2020; Hamilton, 2016), produce news (Diakopoulos, 2019; Fink & Anderson, 2015; Parasie, 2015; Parasie & Dagiral, 2013), or reach their audiences (Bucher, 2017; Gillespie, 2014, p. 1984). However, our understanding of how these changes relate to (re-)distributions of power is still limited. Further, it proves methodologically challenging to address this issue. Asking editors about their perceived agency might already yield different answers than analyzing how they describe their daily practices, as differences in Wu's et al.'s (2019) and De Haan et al.'s (2022) findings would suggest. It indicates that being specific about the contexts in which we may observe (re)distributions of power matters. Study I of this dissertation, for instance, tries to approach this, by examining different domains of inequity — economic, cultural, and political factors that come into play when news professionals use audience data, by drawing on Fraser's theory of justice (2019). Secondly, it underscores the need for methodological variety to arrive at a more nuanced understanding. This dissertation can only provide a couple of puzzle pieces needed. Nevertheless, it aims to bring in a variety of different voices (from news professionals with little to no data skills, to data experts, from journalists working in small, local newsrooms, to employees of global news organizations), and a variety of different methods (from interviewing news professionals to ethnographic observations, and content analysis of job ads), but more on this in the methods chapter.

In the previous sections, I reviewed datafication scholarship to highlight theoretical underpinnings of this dissertation, introduced different approaches to studying datafication, and highlighted how datafication is connected to ways of knowing and doing. In what follows, I will focus on the connection between datafication and AI, which, so far, has only been assumed but not explicitly addressed.

Connecting Datafication to AI

Throughout this dissertation I refer to both, datafication as well as AI, often by joining both concepts together as "datafication and AI". The main theoretical and empirical focus of this work engagees explicitly with datafication, as supposed to AI, (with the exception of Study III). This is not to make a hierarchical distinction — e.g. datafication is more important to socio-technological transformations in journalism than AI. It is done so as an acknowledgement that datafication is the *key precondition* for AI. That is, both phenomena are not interchangeable, but inextricably intertwined. AI requires extensive datafication, because data is the basis of the logical-mathematical codes and models that drive and guide AI (Lavazza & Farina, 2023). Hence, to understand AI we have to understand datafication. Conversely, because increased use of AI requires the expansion of datafication, AI affects datafication.

"Each way of defining artificial intelligence is doing work, setting a frame for how it will be understood, measured, valued, and governed" (Crawford 2021, p. 7). AI is commonly characterized as the computational emulation of human behaviors and abilities, accomplished through the application of machine-learning methods (Broussard et al. in 2019). AI is also a slippery term that is often used to describe a range of different technologies. In their paper "Look! It's a Computer Program! It's an Algorithm! It's AI!" Langer et al. (2022) show that terminological differences can affect people's perceptions of properties of algorithmic decision-making, thus demonstrating a need for analytical clarity. Nevertheless, I agree with Raley & Rhee (2023), who suggest that using AI as an umbrella term opens up new ways to analyze a phenomenon that we could otherwise have difficulties grasping: "AI, while recognizing the reductive, even absurd aspects of the term AI and the magical thinking it perpetuates, nonetheless allows for a kind of linguistic pragmatism, treating the term metonymically and engaging AI as an assemblage of

technological arrangements and sociotechnical practices, as concept, ideology, and dispositif" (p. 188)

Building on this, I define AI in the realm of this dissertation expansively, drawing on previously established approaches of critical AI scholars (Crawford, 2021; Lindgren, 2023; Raley & Rhee, 2023). This includes making explicit that AI is not artificial, but a complex amalgamation of the materials it is relying on — immense natural resources, human labor, and intricate infrastructures. However, the environmental costs that go along with AI, for example, the millions of gallons of water it requires daily to cool the servers of "thirsty data centers" that are often located in particularly dry environments (Hogan, 2015), are not the topic of this dissertation. Neither is the role of human labor in the development and maintenance of AI, such as the work of cleaning, classifying, and verifying datasets that is usually done either by gig or crowd workers and often outsourced to the Global South (Catanzariti et al., 2021; Chandhiramowuli & Chaudhuri, 2023; Graham et al., 2017). However, keeping these conditions in mind helps us ground our understanding of AI in its material bounds and consequences. It serves as a reminder of the limitations of AI and what can be automatically done without human interference. That is, it can help us arrive at an understanding of AI that does not autonomously "think", but rather mimics human-like responses based on computationally intensive training with large datasets or predefined rules.

Instead of stripping AI down to scientific or statistical practice, this dissertation is more concerned with AI's entanglement with values, interests, and priorities (Sadowski, 2019; Winner, 2001) to examine how cultural, economic and political structures are formed around AI as mechanisms for exercising power in the field of journalism. From this vantage point, we can understand AI as an ideology (a distinct way of perceiving the world), an infrastructure, an industry (a result of highly organized capital backed by vast systems of extraction and logistics), and lastly, as a form of exercising power (Crawford 2021, p. 18). From this perspective, AI systematizes the world according to a "linnaean order of machine-readable tables", grounded in logics of datafication that render the world into what can be quantified, which then raises "questions of knowledge production, classification schema, calculative reasoning, and decision making" (Crawford 2021, p. 221). Building on this, we can recognize conceptualizations of AI not just as

mere technology, but as an epistemology and also a form of governmentality and political economy (Raley & Rhee, 2023, p. 196). Meaning, AI governs in two ways that resemble Espeland and Stevens' (2008) understanding of quantification, as discussed earlier. First, it makes the world legible (Uliasz, 2021). It represents people and objects by reconstructing them into machine-readable entities, or data doubles that can be acted upon (Amoore, 2020; Cheney-Lippold, 2020; Couldry & Mejias, 2019). Second "It admits and excludes, separates and allocates in much the same way that law does" (Raley & Rhee, 2023, p. 196). As such, this approach to AI, that foregrounds its epistemic and ideological underpinnings, is closely connected to van Dijk's (2014) conceptualization of datafication that highlights the ideological nature of doing and knowing with data (as discussed above).

In the journalistic context, AI adoption is part of a larger technological transformation in journalism that has been conceptualized as computational (Broussard et al., 2019), i.e. digital journalism (Steensen & Westlund, 2020). On a practical level, AI is employed across the news value chain (Beckett, 2019; Marconi, 2020; Simon, 2022), by encompassing news gathering through information discovery and using data sources like data leaks (e.g., Panama Papers), public records requests, and open-source intelligence. In the production phase, AI aids in transcription, summarization, event detection, image recognition, and content automation (Nishal & Diakopoulos, 2023; Opdahl et al., 2023). Furthermore, it plays a role in distribution through personalization and recommendation systems (Beam, 2014; Møller, 2022b, 2022a; S. Wu et al., 2019). AI also supports business and audience operations by facilitating content curation and moderation (Stockinger et al., 2023; Wang, 2021), and by using pay- and data walls (Evens & Van Damme, 2016; O'Brien et al., 2020; Sjøvaag, 2016). For the purpose of this dissertation, I draw on Lin & Lewis' (2022) conceptualization of "journalistic AI" as the "use of and orientation toward artificial intelligence in newswork" (p. 1629). The "use" element of that definition points to the importance of practice, or *ways of doing.* The "orientation" dimension of that definition points to how the incorporation of AI in newswork is not merely a technical transition but also a social and cultural one. It refers to expectations, values, and ideas attached to AI, that ultimately also structure *ways of knowing.*

In conclusion, because journalistic AI is not simply a technocentric matter of tool adoption but attached to cultures, values and ideologies, it evokes normative questions about what role it should have in journalism and society at large. To approach these issues, this dissertation draws on a critical data perspective that I will further outline and contextualize below.

A Critical Data Perspective

Critical data studies brings various disciplines, such as information and computer sciences, sociology, media and communication studies, or digital humanities together that offer different theoretical lenses to engage with datafication, including critical theories of political economy, governance, race, or gender (e.g. Burns et al., 2018; C. M. Dalton et al., 2016; Iliadis & Russo, 2016; Kitchin, 2014; Thatcher et al., 2016; Thylstrup, 2021). On an overarching level, critical data scholars examine datafication from a specific vantage point that foregrounds questions of power, agency, responsibility, accountability, and subjectivity, and highlights both actors involved in datafication processes as well as those impacted by it (Iliadis & Russo, 2016). This line of research is characterized by efforts to understand datafication as historically situated, that is, situated in both time and place (C. Dalton & Thatcher, 2014). In the context of this dissertation, for instance, I draw on Usher's (2019) conceptualization of place as 1) material and geographic, 2) a setting for action and lived experiences, and 3) a site for cultural meaning and power, and I expand this lens to a historical dimension (Study IV). Doing so allows me to understand place and time in conjunction, in an effort to further contextualize our understanding of datafication.

Critical data studies primarily critique technological determinism and "data-intensive and positivistic" approaches to datafication and underlying paradigms (Iliadis & Russo, 2016, p. 1).

On a methodological level, Flensburg and Lomborg's (2023) review of datafication studies found that most research on datafication is theoretically oriented, whereas empirical analyses largely apply qualitative approaches and rarely make use of data-driven methods (which I interpret as quantitative and computational methods). The authors state that they find this lack surprising given the new methodological opportunities that come with big data in research (p. 1462 & 1465, citing: Kitchin, 2014). For critical data studies, however, this might be less surprising, given that

research aims of critical data studies tend to be more aligned with qualitative, constructivist and interpretive approaches, because it centers people's experiences, perceptions, and imaginations of, for example, algorithmic recommendation systems, tracking of online behavior, and labeling of data.

Another shared attribute of critical data scholarship is, as the name suggests, a form of critique. Within the context of this dissertation, I am interested in a notion of critique that stems from a form of concern and care, as described by Latour. In Latour's (2004) influential work on scholarly critique, he reviews common approaches within academic criticism that aim to debunk beliefs, powers, and illusions (pp. 237-238). According to Latour, this type of critique represents a matter-of-fact analysis that, rather than getting closer to the realities it aims to discuss, moves further away from it by detaching the object of inquiry from the scholars' social realities. Instead, Latour suggests a different approach to critique, one that functions as an exploration focused on matters of *concern*. In his view, a critic is not someone who undermines the beliefs of others but rather someone who provides spaces for participants to gather and discuss. Accordingly, appropriate critique recognizes that if something is constructed, it is inherently fragile and requires careful consideration (Latour, 2004, p. 246). Instead of debunking, the primary aim should thus be to protect and to care (also see De La Bellacasa, 2012; Haraway, 2014) to arrive at a more productive critique that results in "more ideas than we have received" (p. 248). Echoing these sentiments, critical data scholars Raley and Rhee (2023) have called for a criticality that involves working across disciplines, domains, and specializations. In this way, critiques of critical data studies are not reductive, but aim to offer ways of levering practical knowledge with critical framework, methods, counter-mythologies, and epistemologies (p. 188).

The emerging body of work on data justice (Dencik et al., 2019; Dencik & Sanchez-Monedero, 2022; Heeks & Renken, 2018; Taylor, 2017) serves as one example of critical data studies that focus on matters of concern and accounts for different disciplines and perspectives and is generative through this approach. Taylor (2017), for instance, argues for the need to reconcile perspectives that underscore negative and positive consequences of datafications to advance academic inequity, but also inform regulation and build guardrails for practical implementations. Relatedly, Dencik and Sanchez-Monedero (2022) highlight various groups that use data justice

frameworks in productive ways, such as the "Indigenous Data Sovereignty movement made up of a network of alliances and groups around the world that asserts that indigenous peoples need to be decision-makers around how data about them is collected and used" (p. 11).

In the context of this dissertation, a critical data perspective is aimed to be productive by bridging critical data scholarship with journalism research, enabling an analysis that examines how inequity and injustice are enacted through data practices, but also how data practices can mitigate pre-existing inequities (see Study I). That is, with a critical data perspective, the aim is not to create critique for the sake of being critical, but to examine the contexts in which power asymmetries are deepened, as well as the ways in which data practices enable news professionals to pursue journalistic goals that are beneficial for audiences. The more we understand both dynamics, the more we can work towards creating more equity.

In their Special Issue on data power, Gerrard & Bates (2019) "approach discussions of inequality and injustice through three broad lenses: (1) the tactics people use to confront unequal distributions of (data) power; (2) the access to data that is most relevant and essential for particular social groups, coupled with the changing and uncertain legalities of data access; and, (3) the shaping of social relations by and through data." (Gerrard & Bates, 2019, p. 2). Mostly speaking to the last point, this dissertation approaches datafication from a critique which is not limited to technology itself, but rather is interested in "how datafication features in on-going negotiations of social relations and power dynamics" (Dencik & Sanchez-Monedero, 2021, p. 8).

As mentioned above, the dissertation conceptualizes inequalities based on Fraser's theory of justice (2013) that recognizes three analytically distinct dimensions, accounting for 1) economic injustices (maldistribution), 2) cultural injustices (misrecognition), and 3) political injustices (misrepresentation), as well as its interrelations (the theory is adopted and explained in detail in Study I). Migrating Fraser's theoretical framework into the context of datafication advances critical data scholarship, since it acknowledges the relevance of each dimension of power asymmetry and provides an analytical framework to examine how economic, cultural, and political injustices shape and potentially reinforce each other. To connect critical data studies to journalism studies, I will outline why a critical data perspective is relevant to journalism studies, and how it can be enacted in what follows.

A Critical Data Studies Perspective in Journalism Studies

The application of a critical data studies perspective requires some translation given the professional context of journalism and the specific, normative, societal role journalism is ascribed to through Fourth Estate narratives (Hampton, 2010; Schultz, 1998; Tumber, 2001). The main argument I develop here is that an approach to socio-technological change that privileges an explicit concern for equity and pays attention to the unevenness of implications and experiences of datafication is productive, as it allows us to do what Lin and Lewis (2022, p. 1643) describe as "captur[ing] and critiqu[ing] strengths and weaknesses of emerging technologies for journalism's core contribution to democratic society". Whether through a prism of broader democratic aims, or through an inquiry that focuses on how journalistic work can help or harm specific communities, a critical data perspective can help us to understand the ways in which datafication of journalism contribute or stand in conflict with normative ideas of journalism. That is, I aim to integrate a critical data studies perspective with journalism studies to enable interrogations of how changes, as a result of datafication, affect journalistic goals, and ultimately, help contribute to a future of journalism in which data practices are as much aligned with journalistic values as possible.

Journalistic values and ideals are notoriously fuzzy. Journalists hold different views on occupational practices, standards, and professional ideals, relating, for instance, to ideas regarding public interest, fairness, and neutrality. Ideals are not uniformly adopted but are often contested and do not always translate well across cultures and contexts (Plaisance et al., 2012; S. Waisbord, 2016). Moreover, ethical principles do not necessarily represent dominant practices (S. R. Waisbord, 2000; Wasserman, 2010). Nevertheless, journalism is a profession with a public interest mission and some shared norms that are valued globally, even if they are shaped in local contexts (for an analysis of universal values of more than 25,000 journalists from 67 countries, see Hanusch & Banjac, 2021). In democratic contexts, news professionals bear a specific societal responsibility that is tied to journalism's function as the Fourth Estate (which I outline in detail in Study IV). It is beyond the aim of this dissertation to review different streams of journalistic norms and ideals, and much work on this has been done elsewhere; e.g. how journalists negotiate their professional

ideology: (Deuze, 2005), the viability of traditional journalistic values in early computational journalism (Karlsen & Stavelin, 2014), journalists use of ethical frameworks (Porlezza & Ferri, 2022), and role conceptions under automation (Schapals & Porlezza, 2020). My point of departure is to not take a specific set of values in journalistic practice for granted and analyze journalists ways of *knowing* and *doing* through a fixed framework, but instead to pay attention to the ways in which news professionals encounter and navigate the thorniness of normative thinking, and listen to what journalists care about – what costs they are willing to pay, how they articulate their mission, and what their practices tell us about their alignments with their goals.

Keeping in mind that ideals are contested and not universal, what could be a positive example of a close alignment between data practices and journalistic ideals? Algorithmic accountability reporting can serve as one example. As a form of investigative reporting, algorithmic accountability reporting uses computational journalism to examine the implications of datafication and algorithmic power and their impact on individuals and society (Coddington, 2015; Diakopoulos, 2019; Hintz et al., 2022). Here, data practices are used to enact a core ideal — holding power to account (Fell, 2015). ProPublica's "Machine Bias", a series that investigated software used in the US to predict future criminals which they found was biased against Black people (Angwin et al., 2016), falls under algorithmic accountability reporting. Similarly, the Markup has been committed to examining the ways in which algorithmic systems perpetuate systemic biases and inequalities in society, for instance by building tools to analyze what type of content social media platforms amplify (The Markup, 2020). These are projects that demonstrate that data practices are important tools for journalists to realize their societal role. Nevertheless, investigative reporting is only one domain of journalistic work. On a more general level, Diakopoulos (2019) argues that "AI is a new medium through which journalists can express and exercise their ethical and normative values through the code they implement". However, this presupposes two things. First, to express journalistic ideals in code requires significant AI literacies in news organizations, either gained by hiring data professionals and building interdisciplinary teams that work in close collaboration with editorial staff, or by increasing AI literacies across the newsroom. Second, this idea presupposes that news organizations build their own AI tools, or that they can alter pre-existing tools and shape them according to their needs, in order to exercise their ethical and normative values.

Studying AI literacies, Jones et al. (2022) found a gap between increased adoption of AI tools for news production and the level of understanding amongst journalists (at the BBC), who used guesswork and imagination in the absence of accurate conceptualizations of the systems they were using. The authors found that this disconnect could limit journalists' ability to employ AI tools in a responsible manner and to critically examine their outputs and roles in news production. This limitation may also hinder journalists in adapting and shaping these tools, and affect their ability to produce algorithmic accountability reporting, as described earlier. In light of these findings, examinations of strategies for fostering AI literacies, whether on the institutional, organizational, or professional level, are paramount to journalism research that uses a critical data perspective. This is especially true if they address AI intelligibility issues in a sociocultural way, rather than solely in technical terms, to better address normative considerations (ibid.). In the context of this dissertation, Study III addresses efforts on an institutional level to increase AI literacies of individual professionals with implications for their organizations.

To understand journalists' ability to build in-house tools or adapt and shape pre-existing AI systems to encode their ethical and normative values, we need to understand the contexts in which news organizations adopt AI. Previous literature reviews have mapped main areas of AI adaptations in the media industry (e.g. Chan-Olmsted, 2019) and in journalism, specifically (Kotenidis & Veglis, 2021). Further, Beckett and Yaseen's (2023) global survey examines "what news organizations are doing with AI" and provides a thorough understanding of different tools that are used for various purposes (e.g. AI-powered tools for speech-to-text transcription and automated translation, such as Colibri.ai, SpeechText.ai, Otter.ai, and Whisper for newsgathering (p. 15)). The report further highlights extraordinary cases of AI applications used by news organizations, such as The Reuters News Tracer, that "utilises machine learning algorithms to rapidly identify breaking news stories and verify their credibility" (p. 20). However, to what extent these tools are built in house, or have been modified and built on pre-existing infrastructures and tools is mostly opaque. On an overarching level, Beckett and Yaseen's (2023) survey suggests a general inclination to use third-party tools (p. 16). This is perhaps unsurprising given the extensive resources that AI development requires, which ultimately leaves news organizations with few resources and at a particular disadvantage. However, I am unaware of any scholarly work that

systematically analyzes the extent to which news organizations across the industry are building their own in-house tools versus using off-the shelf tools from third parties, and moreover, to what extent news organizations can and do successfully shape pre-existing tools according to their needs. Accordingly, our understanding of journalists' ability to encode their values into AI tools is limited.

Further, previous studies highlight that news organizations struggle, despite the best of intentions, to integrate AI tools into journalistic workflows and that those tools don't always easily integrate with work routines and values (Gutierrez Lopez et al., 2023). This gives us some indication of the complexity of the matter. That is, without a nuanced understanding of what news organizations actually do with AI, taking differences of various forms of AI adaptation, developments, and modifications into account, our understanding of how journalists can align data practices with their values is rather constrained. Integrating a critical data perspective in journalism studies thus requires empirical attention to the material underpinnings of technological practices. That is, examining what specific hardware and software are actually used and how. Study II of this dissertation aims to contribute to this aim by analyizing which specific material components data professionals are expected to work with, as expressed by news organizations in job advertisements. Analyzing material components also gives us empirical ground to understand ownership issues, associated channels of control (Simon, 2022), or, in different words, the extent of power asymmetries between platform companies and news organizations (as described earlier).

In conclusion, echoing calls from Carlson & Peters (2023), a critical data perspective in journalism studies calls for an examination of how various forces, in this case of socio-technological nature, are either directly challenging journalistic ideals or are potentially restructuring abilities to exert journalistic agency on a macro level. On a meso level, we need to pay attention to how organizations produce an interplay of material and socio-technical structures, to understand what kind of practices and paradigms are fostered as a result. Lastly, on a micro level, it necessitates an analysis of how news professionals' experiences with datafication are shaping normative thought (following Espeland & Stevens' (2008) argument that quantification involves reactivity, as elaborated earlier). Further, a micro perspective may counter imagined objectivity of data and technology and contribute to an understanding of journalistic ways of

knowing and *doing* with data that recognizes that people work with technology from a situated, embodied location.

Navigating normative questions and challenges to boundaries with external actors and actants are not new to journalism, but ingrained in its societal roles (Møller, 2022a). However, it requires news professionals to develop a vision for guiding technological developments and a discerning approach to identifying and reconciling relevant and (potentially) conflicting values. This is at its core a normative project. A critical data perspective is helpful here, by making these normative questions explicit and foregrounding issues of power and inequity, instead of taking them for granted. On an overarching level, a critical data approach helps anchor these discussions in their relationship to historical contexts, social structures, and dominant paradigms that are tied to distributions of power. This approach then, hopefully, leaves us "better equipped to confront the array of forces that push and pull" on what journalism should be (Carlson & Peters, 2023, p. 1064).

Methodological Approach

In the following section I outline the methodological approach of this dissertation by 1) addressing its methods and data, 2) situating the work within a philosophical paradigm that is motivating my approach, 3) defining the levels of analysis that the empirical studies encompass, 4) explaining my sampling strategy, and lastly, 5) offering some reflections on the methodological implications of collecting qualitative data during the Covid-19 Pandemic.

Empirically, the dissertation draws on qualitative interviews (Study I, 3), qualitative and quantitative textual analysis of job advertisements (Study II using a mixed-method approach and Study IV being qualitative), and ethnographic observations (Study III). The primary focus on qualitative methods is motivated by the constructivist outlook reflected in the overarching research question: *How does datafication and AI structure journalism's way of knowing (epistemology) and ways of doing (practice), and what are its implications for equity?* Constructivism posits that people actively construct their knowledge based on their experiences and reflections (Honebein,

1996). That is, constructivism recognizes that *knowing* is not removed from *doing*, but that the way people know and understand the world is tied to actions and previous experiences. Accordingly, this dissertation draws on qualitative methods to center people's "thick" accounts of their sensemaking and to acknowledge the ways in which their experiences are relational, set within social, cultural, and political contexts, and moreover, are rich, complex, and often ambiguous.

Altogether, the dissertation reflects the experiences and perspectives of 87 news professionals situated in 27 countries in Europe, the Americas, and the APAC region, by capturing their experiences and perspectives either through interviews (n=49, Study I & Study III), or through observations (Study III). Semi-structured qualitative interviews have been a particularly useful way to begin the research in Study I. Given its open-ended nature, the interview process provided me with opportunities to be steered in directions by interviewees that I did not previously conceive of. In other words, this method allowed me to listen to interviewees tell me what they considered to be relevant, instead of the other way around. Textual analysis enabled me to capture discursive constructions of meanings circulating in news organizations and provided further context for the individuals' accounts in Study I and III. Lastly, ethnographic observations complement the previous methods by providing implicit aspects and contextual information, which in turn allowed me to gain a deeper understanding of cultural or social settings. I consider the dissertation to be enriched by the building and maintaining of relationships and the earned trust of the participants of this research — a central aspect of qualitative research.

One of the main challenges of this dissertation is to grasp a profession while it is swiftly transforming under technological change. Longstanding journalistic practices and ideals might outlast technological developments, but understanding how they evolve requires theoretical work. Qualitative research is paramount to building and refining theories that remain relevant over the course of socio-technological change (Gagrčin & Butkowski, 2023, p. 4). In this way, the dissertation's commitment to qualitative methods is aimed at yielding both new empirical and theoretical insights.

Lastly, I want to emphasize that the dissertation has benefited greatly from the different academic contexts in which the empirical studies were conducted and from the expertise and input of co-authors and collaborators. The data collection of Study I was conducted within the *Just Information* research project at Stockholm University, under the supervision of Alexa Robertson and with support of student assistants Yanthe Zebregs and María de la Huerga. Study II and IV were conducted within the *Digital Journalism* research team at Universität Hamburg in collaboration with Laura Laugwitz and Juliane Lischka. Study III is the result of a collaboration with Anna Schjøtt-Hansen at the University of Amsterdam.

Levels of Analysis

Taken together, the four empirical studies of this dissertation encompass three levels of abstraction and analysis. Using Reese's (2016) hierarchy levels for influences model as an orientation, this dissertation is organized by a micro perspective of "the individual" (Study I) , meso perspective of "the organizational" (Study II) and what Reese (ibid., p. 7) calls "a more macro, socio-institutional level" (Study III & IV).

The Individual

On the micro level, this dissertation posits that individual considerations and actions matter and that paying attention to them ultimately helps us understand larger journalistic projects (ibid., p. 5). This level of analysis recognizes the relative autonomy of individuals, and situates them in bigger, surrounding structures (such as organizational cultures), which they are both shaped by and contribute to. This may include reflections of personal characteristics, such as the ways in which people's personal backgrounds and demographic features (e.g. gender, race, or socio-ecnomic status) inform their actions, or how professional roles shape their understanding of specific phenomena (ibid.). By taking a micro perspective, Study I draws attention to the role of individuals in either reproducing or mitigating inequities in journalism through data practices. This closer look complicates things in a productive way, recognizing that there is no single way of engaging with data, and that while informal and formal rules shape individual practices, much comes down to individuals' actions.

The Organizational

Analyses on the meso level acknowledge that journalism is produced through organizational structures and cultures (Reese, 2016). Organizational culture is learned and fostered through formalized rules, such as bureaucratic interventions (Etzioni, 1961), and more informal expectations and norms that individuals internalize through socialization (Sewell, 1998). Organizational policies, for instance, may influence journalists' perceptions of their work (Relly & González De Bustamante, 2014), or even constrain it (Molyneux & Nelson, 2023). Hence, a journalist's workplace is where journalistic ideals and the autonomy to realize them are negotiated on a day-to-day basis (Örnebring et al., 2016). Many factors, such as the size, age (Xu & Jin, 2017), audience orientation (Belair-Gagnon et al., 2020), or pace of digitalization of an organization (García Avilés et al., 2004) can contribute to shaping the work of news professionals. Moreover, country-specific factors can play a role in shaping organizational cultures. For example, by comparing newsrooms in the US and France, Christin (2020) found that longstanding tensions between market forces and news outlets in the US, and the historical relative absence of market pressures in France have shaped organizational cultures, respectively. Studies encompassing a meso, organizational focus, link micro levels of journalists' individual experiences and perceptions to the macro level of wider societal currents. Analyzing job ads of news organizations, Study II aims to grasp meta-journalistic discourses (Carlson, 2016) of organizations that give insights into their strategic communication, and, by extension, their organizational culture.

The Socio-Institutional

Finally, the "more macro, socio-institutional level" captures the "inter-organizational field", where news organizations cohere into a larger institution (Reese, 2016 p. 7). This level of analysis encompasses the structured dependency relationships and power asymmetries that are formed with other major actors (such as the State or Big Tech). That is, it accesses the interplay of cultural, economic, and political factors that shape journalism. This perspective further allows for a historical dimension, "which helps explain the emergence of practices and norms as a contingent outcome" (ibid.). Study IV, for instance, traces notions of Silicon Valley and Forth Estate ideals to understand contemporary approaches to data in journalism. At this level of analysis, we can better recognize how power flows within and across national contexts and consider how

organizations, together with other external actors (such as research institutions or tech corporations), form relationships with each other and forge new emerging deliberative spaces in which professional practices and ideals are negotiated (ibid.). In doing so, an analysis on this level may avoid a default centering of the nation as a frame for social reality and approach a more global understanding of how socio-technological cultures and practices converge or affect each other in different parts of the world.

In the context of this dissertation, Study III, which ethnographically follows the global *journalismAI* Academy, focuses on such wider networks of knowledge generations (Knorr-Cetina, 2007). "The Academy" is targeted towards small newsrooms across the world that have very little previous knowledge of AI and fellows are instructed by news professionals that are leading experts and belong to large news organizations. The study is thus able to recognize how differently positioned organizations influence each other on a socio-institutional, global level. Because the Academy is academically housed by the London School of Economics and funded by Google, Study III also captures power dynamics that includes these external actors. In other words, not all journalistic work is exclusively taking place within a newsroom, or even within the context of a single country. Recognizing this, this dissertation takes on a *worlds of journalism* framework (Lewis & Zamith, 2017), which is explained in the next section, followed by an outline of the organizations, countries, and different professional groups at the center of this research.

The People, Organizations, and Worlds in Focus

Rather than applying a country-specific focus, a *worlds of journalism* framework approach (Lewis & Zamith, 2017) foregrounds the collective nature of journalism — its interconnected people, processes, and products that collectively enable and legitimize particular forms of journalism. There are several advantages to applying this framework. It can highlight the ever-changing nature of networks in which journalism takes place, while recognizing how these constellations lend authority to certain professionals, practices, and products. In doing so, it underscores the development and negotiation of values that structure journalistic work through interactions between its members. Lewis and Zamith provide three dimensions: defining worlds of journalism

through collective activity, conventions, and status. By looking at the nature of collective activity, the authors direct empirical attention to "both formal gatherings, such as online learning spaces designed to bring together journalists and technologists" to develop tools (Lewis & Usher, 2016), and more informal "meetups". Study III is paying empirical attention to collective activity, by making a journalistic training initiative, which is conducted through live sessions, the object of analysis.

This notion of worlds of journalism additionally explores the role of shared conventions and status. Conventions "both facilitate and constrain the creation of particular works" (Lewis & Zamith, 2017, p. 124). The focus of this dissertation is not to examine norms or something more akin to Bourdieu's notion of habitus (Bourdieu, 1986). However, I interpret conventions to relate to issues of *knowing* and *doing*, as they reflect a shared understanding of doing things, and, as a consequence, reveal a general schema for organizing knowledge and practice. The ability to reconfigure conventions (Lewis & Zamith, 2017, p. 126) is what I refer to, in the context of this dissertation, as "change agency". Highlighting the role of status, Lewis and Zamith direct attention to who has access to what kind of resources, how those resources may be expended, and the ways in which future resources are likely to be gained by pursuing a particular course (2017, p. 116) – all of which speak to the dissertation's aim to understand power asymmetries in journalism and their implications.

That is, the authors' notion of worlds of journalism brings concepts into play that are closely related to themes at the core of this dissertation, but they will not be explored at length in the scope of this dissertation. What is more of interest is the underpinning idea of their framework — to understand journalism not only through the bounds of organizations, or in reference to nation states, but as a bricolage of actors that are connected through tools and techniques, and to call attention to their collective activities — which motivate the sampling strategy of this dissertation. While journalists increasingly adopt common tools, distribute news on the same platforms, and work in interconnected ways, this does not mean that digital journalism has led to a degree of isomorphism (Laaksonen et al., 2022) that renders cultural contexts meaningless. Comparative work, such as Christin's (2020) aforementioned study, have showcased how cultural differences are reproduced despite technological and even economic convergence. I acknowledge that a more holistic understanding of how journalism is changing under socio-technological influence can be

yielded from a comparative approach. Nevertheless, different forms of journalism, each situated individually, also hold significance "in the aggregate, in the network of distinct but interconnected worlds, meaning-making and shared interpretations." Produced in one place (Lewis & Zamith, 2017, p. 126-27). To add more clarity, I will outline both the organizations and countries and the different professional roles of the people in focus of the studies in what follows.

Organizations & Countries

The four empirical studies of this dissertation follow different sampling strategies relating to the local, national, and global reach of news organizations, the degree of professionalism and digitalization of their respective organizations, or their ties to Fourth Estate ideals. Collectively, they address various factors that might play a role in shaping journalists' relationship with technology, with each study contributing to a more nuanced understanding. Study I's sampling encompasses leading national and global media organizations reporting in English and/or German. Global media organizations include outlets such as the BBC World, Deutsche Welle, the Guardian, Al-Jazeera English, and the New York Times, who all have the capacity to address worldwide audiences. Study II narrows the focus to media landscapes in the United States and the United Kingdom, exemplifying liberal media systems with moderate news professionalism and rapid digitization. This study includes incumbent broadcasters, print legacy outlets, and online-only news platforms. Study III expands the geographic scope and incorporates both the perspectives of small, local newsrooms and larger, national and global news organizations. The former includes 41 participants from 41 news organizations spanning 23 countries in the Asia-Pacific and Americas regions, while the latter includes 15 professionals working in 8 different countries (more detailed information can be found in the Appendix of Study III). Study IV re-focuses on the United States and the United Kingdom, selected for their historical connections to Fourth Estate ideals. Notably, the study includes news outlets with the highest number of investigative journalism awards, having earned what Lewis and Zamith conceptualize as status (2017) which structures their access to resources (as described above). The complementary sampling of the individual studies —ranging from global media giants, award-winning organizations, to small local newsrooms — is aimed to provide a more nuanced perspective on how datafication and AI structure journalism's way of

knowing and *doing*, and how implications for equity may differ according to organizational and local contexts.

Professional Groups

One methodological challenge of this dissertation was to identify the relevant voices and professional groups to understand how datafication and AI shape journalism. With journalistic professions in flux, titles, roles and responsibilities might change over time and differ from organization to organization, making it particularly difficult to determine pertinent actors. To address this challenge, the four studies of this dissertation take different sampling approaches. While all studies are concerned with tech-oriented roles, each study sets a different focus, adding to a more comprehensive understanding.

Study I includes two key groups: audience editors and journalists with a record of reporting on social inequities. Audience editors, often hold titles like "Head of Audience," "Deputy Social Media Editor," or "Director of Audience Development," and play a pivotal role in analyzing and interpreting audience data. Incorporating audience insights into newsrooms, audience editors are critical in understanding the intersection of data practices and journalism. Journalists with expertise in inequities, on the other hand, are especially attuned to power asymmetries, and can give insights into how data practices impact journalists' capacity to address inequities. In Study II, the focus shifts to professionals skilled in data-related roles, such as data scientists, UX researchers, product managers, product designers, app developers, and multimedia editors. This approach draws from Kosterich's (2021) categorization of journalism workers. Study III moves to the intersection of journalism and AI. This study includes AI experts, both in the news industry and research community, as well as journalists with little to no prior knowledge in AI. Accordingly, this sampling aims to further our understanding of AI adoption by examining experiences and perceptions of news professionals with varying levels of AI expertise. Study IV takes an approach based on Guo and Volz's work (2019) to identify individuals with technical skills, which encompass newly-developed computational skills like software development, the application of specific software, and data analysis. These technical skills are integral to the innovative expertise of news professionals (ibid.) and are examined to understand their impact on journalism's technological orientation. Together, the studies create a sample of professionals with different

skills, background and relationships towards technology, each offering a different lens to examine journalism under datafication and AI.

Reflections on the Methodological Implications of Qualitative Research during the COVID-19 Pandemic

I anticipated challenges in the data collection process – journalists are notoriously time-strapped, so convincing them to engage in lengthy interview research presents a significant obstacle. I did not, however, anticipate a global pandemic. The beginning of the data collection for this dissertation coincided with the start of the COVID-19 pandemic, which had unexpected effects. While we all watched our worlds shrink at that time, in a paradoxical way, my empirical world expanded. Maybe this was thanks to quickly shifting work norms that made virtual meetings routine and lowered barriers for me. Perhaps it was due to the lockdowns and other COVID protocols that isolated us all to such a degree that a Zoom meeting with a PhD student was, under these circumstances, seen as a way to connect with others rather than a disturbance in a busy work schedule. Interviews that lasted an hour longer than originally planned was one indication of this, and the research would later support my experience that interviewees were particularly motivated to participate, possibly as compensation for lost social connection (Averett, 2021).

Recruiting interviewees was still a demanding task, but I gained an unforeseen variety of insights. Echoing experiences from other qualitative researchers, virtual interviewing enabled a wider access to geographically dispersed participants, which gave my sample a richer diversity in cultural and socioeconomic terms (Eiguren et al., 2021; Keen et al., 2022; Lathen & Laestadius, 2021) and promoted local voices (Pittaway et al., 2010). While I was sitting each day at the same desk in my home office in Berlin, the people on the other side of my screen were talking to me from places like London, Doha, or Stockholm. Instead of meeting them in their newsrooms, I was invited into frames of their kitchens, living rooms, and make-shift offices. Talking to them from the comfort of their homes made it easier to ask difficult questions, which was particularly relevant when it came to issues relating to power asymmetries. It is less challenging to talk about newsroom

inequities, how your perspective differs from others in the organizations, and the issues you have encountered at work, when there is more than a thin wall separating you from your colleagues.

By the time I conducted the last interviews for this dissertation together with Anna Schjøtt Hansen (Study III) it was 2023. Pandemic lockdowns now felt like a distant memory, and yet, I was still virtually hopping around the globe. One of these days I conducted five hours of interviews switching between Malaysia, the Philippines, the US, Ecuador, and Paraguay. I was experiencing something that could be called place warping. Our participants were geographically so dispersed that the contexts started to blur together, and I had to remind myself that Arcata was a city in California and not to be confused with Asunción, the capital of Paraguay. We made sure to ask questions that situate the interview to counter this context vertigo — *If you were to describe the AI landscape in your local context, how would you describe it? Do you see any unique challenges for AI adoption relating to your specific local context?* Nevertheless, this experience undoubtedly reflects a limitation that resulted from my sampling. Casting a wide net geographically made it impossible to dive deep into context-specific conditions or implications and, the scope of this dissertation did not allow for comparative approaches that would have added analytical clarity. As a consequence, distinctions between global trends and local conditions became less clear. This was an issue that was further amplified by my choice to anonymize interview participants, which makes it more difficult for the reader to contextualize individual quotes but was a conscious decision to provide confidentiality and protect organizations and individuals who agreed to participate. On the other hand, this sampling approach allowed me to move beyond nation frames and conduct research from the bottom up, snowballing across country lines. In this regard, the process felt more true to the material.

Concluding, data collection through qualitative interviews and virtual observations during the Covid-19 Pandemic resulted in unexpected benefits (also see Lobe et al., 2020 for comparison), while also contributing to limitations of this research. Further reflections of the limitations of this research can be found in the respective discussion sections of the empirical studies.

General Discussion

Developing a comprehensive understanding of how AI and data practices can contribute to the societal role of journalism is critical at a time of rapid technological advancement. Without a clear perspective on how to navigate AI-driven solutions and responsible data practices, news organizations become susceptible to technological fixes that may not necessarily improve journalistic products or even cause societal harm. Because of this, it is important to discern contexts in which data and AI practices can meaningfully contribute to journalistic objectives. In practice, this requires journalists to ensure that technological advancements align with core journalistic ideals. Previous work has discussed values and principles that should govern responsible AI in journalism (Dörr & Hollnbuchner, 2017; Helberger et al., 2022; Komatsu et al., 2020; Rojas Torrijos, 2021) and various organizations across the industry have published their own guidelines to make their approaches transparent (for an overview see Becker et al., 2023). This dissertation has put the focus on the underlying assumptions and logics of datafication (Espeland & Stevens, 2008) and AI, to understand their potential to reinforce or mitigate inequities. It has been motivated by the belief that taking a wider, societal perspective to examine how our relationships with technologies are restructuring ways of *knowing* and *doing*, enables us to interrogate, contest, re-evaluate, and improve journalistic processes so that they align more closely with professional ideals. In this respect, the empirical studies of this dissertation offer the following takeaways:

First, **journalistic data practices are inextricably linked to economic, cultural, and political inequities.** On an *economic level*, power asymmetries between platforms and news organizations (Møller Hartley et al., 2023; Nielsen & Ganter, 2022; Papaevangelou, 2023; Simon, 2022) have structured journalism's *ways of doing*, making organizations financially reliant on the extraction, analysis, and (for many) the commodification of audience data (Study I). It further motivated news organizations to orient themselves towards AI practices in hopes that AI tools can cut costs and make processes more efficient (Study III). These findings have implications for the journalism-audience relationship. Study II demonstrates that as news organizations prioritize their ability to use audience data for personalization and commodification, ethical concerns, such as

audiences' right to privacy and protection from surveillance(Adams, 2020; Porlezza, 2023; Renieris, 2023; Schaetz et al., 2023), are often deprioritized. In Study I, I argue that this puts audiences' trust in journalism at risk — a worry that was later expressed by professionals participating in Study III. *On a cultural level,* Studies I, II, and IV highlight how data-positivist approaches (Iliadis & Russo, 2016) in journalism are privileged under logics of datafication, and structure *ways of knowing.* This ultimately results in a value imbalance between more data centric, quantitative approaches and more qualitative forms of knowledge production. In Study I, for example, news professionals expressed the view that audiences' informational needs could be better understood through their data traces than through more traditional qualitative approaches, such as focus groups. This was despite the fact that findings reflected that qualitative approaches were often needed just to accurately interpret audience metrics. On a *political level,* Study I demonstrates how audience data is used to make editorial decisions to prioritize certain audience segments over others. News organizations that invest in data-informed strategies to turn users into paying subscribers tend to amplify issues of misrepresentation by neglecting economically less well positioned audiences. This process further disadvantages traditionally marginalized audiences (Usher, 2021). In summary, journalists' engagement with data does not operate in a vacuum, but is intertwined with pre-existing inequalities. Understanding how journalistic data practices might reinforce these inequalities requires an unpacking of the complex interplay between economic, political, and cultural dimensions of power structures. It also necessitates an understanding of the processes within organizations, the actors at the heart of data centric decision making, and the structures supporting them.

This leads me to the second overall finding of my dissertation: **Data elites** (such as data analysis, engineers, and product developers) **consolidate power through their specialist knowledge and leverage their epistemic authority to support, sustain, and grow datafication in journalism.** Lischka et al. (2022) illustrated that technologists working in the news industry have a collective ambition to improve journalism via technological means and strive for more agency within their organizations (also see Kosterich, 2020). Study II shows how professionals working with data, specifically, are drivers of socio-technological change. The study outlines the structures supporting their agency, which I identify as epistemic infrastructures. This includes materialities (such as data itself, or software), social interlinkages that data agents form with

broader networks, and resulting paradigms (drawing on Tichenor et al., 2022). These paradigms recognize datafication as a governing frame that fosters a positivist approach to knowledge creation and subsequent decision-making that cultivate profit-oriented norms and values. This connects with the finding of Study IV, that demonstrates the construction of datasolutionism in journalism. Datasolutionism, the notion that journalism's problems can primarily be solved by data-driven practices, grants data elites epistemic authority and, as a result, agency to change journalistic processes. Findings of Study III further illustrate the pitfalls of dataolutionism. I observe that some data and AI practices continue to be valued and prioritized over non-tech related approaches even when they stand in conflict with professional norms and ideals, based on the premise that they could provide new solutions to pre-existing problems. This is what I call *sticky epistemics*. I show that sticky epistemics are persistent because they are shored up by pre-existing power asymmetries. That is, they function through and reinforce inequity, as they reinstate epistemic power.

Study IV examines the re-negotiations of ideals from a historical perspective and contextualizes datasolutionism. It offers a third key takeaway: **Journalistic ideals are converging with Silicon Valley ideals under datafication.** That is, similarly to Silicon Valley's notion of technosolutionism — the idea that societal problems can be solved with the means of technology — news organization construct datasolutionism as pivotal to archive Fourth Estate ideals. Along with this, news organizations' self-understanding changed, as they aligned themselves with tech organizations: "The New York Times is a technology company committed to producing the world's most reliable and highest quality journalism. Our ability to do so relies on a talented team of expert technologists who help NYT learn from a tremendous abundance of data unique to this company" (Study IV). As this quote highlights, the convergence of Silicon Valley and Fourth Estate ideals goes hand in hand with an expectation that datafication is a prerequisite for the achievement of quality journalism, as well as a distribution of power and agency to technologists.

Overall, the empirical studies illustrate the formation of power asymmetries across four intersections, in connection to datafication and AI adoption of journalism: inequities between news organizations and platforms (Meese & Hurcombe, 2021; Møller Hartley et al., 2023; Nechushtai, 2018; Nielsen & Ganter, 2022; Simon, 2022), between data professionals and editorial staff

(Schjøtt Hansen & Hartley, 2021; Sirén-Heikel et al., 2023), between news organizations and their audiences (Adams, 2020), and between resource rich and resource poor news organizations (Rinehart & Kung, 2022). However, the studies also give insights into possible pathways to mitigate power asymmetries.

Pathways to Mitigate Inequities

News organizations' ability to extract, analyze, and commodify user data is a reflection of power asymmetries between publishers and audiences. Audiences have little to no control over their data when they want to engage with news (Adams, 2020; Binns et al., 2018; Kammer, 2021). Moreover, as mentioned above, news organizations tend to use audience data to make editorial decisions that favor economically well positioned audiences and disadvantage traditionally underserved groups (Study I). However, Study I also shows that **news professionals who possess relevant cultural and economic resources can mitigate inequities and use audience data for greater recognition and representation of audiences.** That is, more critical and attentive analysis that relies on both quantitative and qualitative approaches that structure ways of knowing can be successful in identifying traditionally underserved communities and their information needs with implications for ways of doing. Notably, the professionals who successfully used data practices to mitigate issues of misrepresentation were women and/or members of traditionally underserved groups, who argued that alleviating newsroom inequities is a critical step towards building newsrooms' capacities to improve journalistic representations of diverse audiences. How data is leveraged is thus not only a question of organizational practices or cultures. Individuals' intentions and actions matter. Reversely, how supported professionals are by organizational structures and resources can crucially impact their ability to effectively use their data skills to mitigate inequities.

This point connects to one of the overall findings: **A more equitable journalism under datafication requires epistemic equality that recognizes and values the contributions of both quantitative, computational methods and qualitative approaches to the generation of knowledge.** Results across the empirical studies of this dissertation highlight how over-valuation of quantitative, technology-centered approaches risk a "fetishization of data" and over-production of data that may lead to ignorance rather than insight (Schwarzkopf, 2020). Study I illustrates how

an ability to critically leverage both quantitative and qualitative methods can improve journalistic practice. However, against the backdrop of a push for AI in journalism (Study III) that further fosters positivistic dispositions, epistemic equality is unlikely without a cultural shift. Practices will not change, unless the cultural imaginaries attached to datafication, specifically datasolutionism, are challenged. What would epistemic equality, and thus a restructuring of ways of *knowing* and *doing*, in journalism look like? It would integrate a critical data perspective (boyd & Crawford, 2012; C. M. Dalton et al., 2016; C. Dalton & Thatcher, 2014; Iliadis & Russo, 2016; Raley & Rhee, 2023), by foregrounding questions of power. In regards to the journalism-audience relationship, it would involve channeling resources into efforts to actively reach out and listen to audiences, establishing trust via dialogue and transparency, and building sustainable relationships across diverse communities. As publishers increasingly adopt AI, they might do so with journalistic humility — the acceptance that journalists can never fully understand or control their audiences (Nelson, 2021, p. 6) — while trying to bridge the gap between them to ensure that socio-technological changes are made with audiences' needs and concerns in mind. In regards to relationships between data elites and editorial staff, epistemic equality would entail an interdisciplinary, collaborative environment where both technologists and editorial staff are viewed as equal partners in the knowledge creation process, as envisioned in approaches that are informed by a critical data perspective. Study III shows that such efforts would be welcomed: journalists with little to no AI expertise want to learn and expand their skill set (also see Koivula et al., 2023), and appreciate interdisciplinary work.

The realization of these values is a multi-stakeholder process and requires exchange between professionals within their own newsrooms and data labs, but also across the industry. As Study III underscores, there is a need to **build new spaces in which news professionals can advance their data literacies while also engaging in critical debates that foster epistemic equity.** The spaces in which journalists do meet for interdisciplinary change are usually funded by Big Tech. As the paper demonstrates, this comes with its own issues and risks strengthening dependencies between platforms and publishers (also see Luchs et al., 2023). Ensuring future journalistic resilience under rapid technological change thus requires building spaces that are independent from Big Tech funding in which journalists and technologists can form alliances across news organizations.

Contributions

The overall contribution of this dissertation lies in its critical examination of the relationship between journalism, technology, and inequities in a time of rapid technological developments and an ongoing economic crisis of the news industry. Datafication and Journalistic AI, as this dissertation shows, bring key challenges in journalism to the surface. They speak to the very nature of what constitutes journalism, where we place value in news production, and how we envision the profession to change for the better. Data elites are at the heart of these debates, driving socio-technological change in newsrooms according to their visions, and in doing so, are restructuring ways of knowing and doing in journalism that impact pre-existing inequalities. Explicitly focusing on power asymmetries from a critical data perspective allows us to not only understand how journalism is reorienting itself amid technological developments but deepens our understanding of the consequences. Engaging with these aims, the four empirical studies of this dissertation, provide shared contributions to theory, methodology, and practice.

On a theoretical level, this dissertation applied and extended Tichenor et al's (2022) framework of epistemic infrastructures. Study II provides additional validation of this framework, which originates in the field of public policy. The study further demonstrates its applicability in journalism studies, which offers a new way of analyzing socio-technological expressions of power that encompass data, by linking 1) technological tools and other materialities with 2) practices and the relations that are formed around them, and 3) paradigms that emerge alongside these processes.

The second theoretical contribution of this dissertation lies in adopting Fraser's theory of justice (2013) in the context of journalism under datafication. Different aspects of Fraser's work on justice have recently been used to understand societal consequences of data practices, most notably within the context of civil society organizations (Gangadharan, 2017) and surveillance capitalism (Cinnamon, 2017). To the best of my knowledge, Fraser's theory of justice (2013) has so far not been integrated into the field of digital journalism. Applying Fraser's theoretical framework and using audience data practices as an exemplary case, Study I demonstrates how data

can be understood as an economic, cultural, and political good in journalism. Findings reflect that the integration of Fraser's (2013) theory of justice framework into the context of journalistic data practices allows for a structural examination of the relationship between data practices and the reproduction of inequities that takes issues of maldistribution, misrecognition, and misrepresentation into account. At the same time, it offers a theoretical lens to examine under which conditions data practices can be used for greater recognition and more adequate representation of audiences. Future scholarship could build on and further validate this integrated framework by applying it to other forms of data practices, such as data used for news production.

Lastly, in an effort to provide a framework that encompasses a historical analysis of how Silicon Valley and Forth Estates ideals have converted, this dissertation builds on Usher's conceptualization of place — as material and geographic, a setting for action and lived experiences, and a site for cultural meaning and power (Usher, 2019) — and expands its lens to a historical dimension. Integrating a historical approach into Usher's (2019) conceptual framework allows us to recognize both place and time in conjunction and contributes to our understanding of place as historically rooted in the complex interplay between materialities, narratives, and cultural meaning.

On a methodological level, this dissertation presents an innovative way to analyze the reinforcement of epistemic paradigms that promote and validate certain AI practices in journalism, by identifying and examining spaces of inter-organizational exchange on a global level (Study III). Specifically, it advances previous methodological approaches by addressing an empirical challenge in journalism studies. Research on journalistic AI tend to depend on top-down sampling approaches, using either nation- or region-based frameworks (e.g. Latin America: Soto-Sanfiel et al., 2022; or the Netherlands: Vergeer, 2020) or focusing on specific news organizations (e.g. the Washington Post: Cools et al., 2023; or the BBC: Jones et al., 2022). While these approaches offer valuable insights (e.g. by offering findings based on comparative analysis), they do not account for the interconnected nature of journalism's engagement with AI. As Study III demonstrates, journalists are not operating within a vacuum of their newsrooms, but their approaches to AI are embedded in a global, macro-epistemic field that is shaped by various actors outside of their respective organizations. By broadening the sampling lens to include global networks and inter-

organizational exchanges, researchers can capture a more holistic understanding of how information, practices, and ideas circulate within and across organizations and cultural contexts. This approach not only recognizes the interconnected nature of *worlds of journalism* (Lewis & Zamith, 2017), but also acknowledges the spaces through which ways of *knowing* and *doing* are formed. Investigating these networks can uncover collaborative patterns, shared challenges, and emergent practices that might be crucial for comprehending the evolving dynamics of journalism in the context of datafication and AI, as Study III demonstrates. Hence, Study III calls for a shift in how AI adoption in journalism is studied, suggesting to identify sites of inter-organizational collaborations and exchange between journalists, technologists, academics, and policy makers, to improve upon existing sampling methods and revisit previous questions.

The study makes further contributions **on an empirical level.** At a time in which AI skills are increasingly demanded in journalism (Koivula et al., 2023; Min & Fink, 2021) and AI literacies are entering journalism education (Kothari & Hickerson, 2020), qualitative research on journalistic AI training initiatives provide an important step to our understanding of the future of journalistic education. There are few studies examining AI training programs, and exceptions either used a quantitative method (Heuer et al., 2021), or walkthrough and discourse analysis of pre-recorded training materials (Luchs et al., 2023). To our knowledge, our study is the first of its kind to observe live sessions, exercises, and discussions of AI training programs, and gives unique insights, not only into how AI was presented, but also how it was received by the participants.

Regarding the news industry, the dissertation showcases how news organizations and individual journalists tend to de-prioritize ethical concerns amid a push for AI. This is not to say that journalists do not approach AI in a critical manner — the findings of this dissertation reflect various concerns of journalists and calls to avoid specific AI practices, either for moral or practical reasons (e.g., Study III). However, an overarching theme across the studies was a willingness to adopt data practices under economic pressure, despite ethical concerns. In doing so, news organizations may risk damaging their relationship with their audiences and fall short on their societal responsibilities. It also demonstrates the need for further deliberation from journalists to decide which standards they want to uphold as they move towards AI adoption. However, as Study III observes, news professionals lack resources for this type of critical exchange — specifically,

time and spaces to come together. To support equitable journalistic AI, inter- as well as interorganizational exchange that can foster critical reflection and form professional norms, is critical. Ideally, this would be in dialogue with academic institutions, which might be able to provide spaces for debate that are independent from Tech funding. This would be important not only to mitigate potential risks, but also to showcase successful approaches, as the ones that emerged in this dissertation.

As Study I demonstrates, well-resourced news organizations are capable of mitigating inequities through practices informed by audience data. Specifically, members of traditionally underrepresented groups in journalism were able to leverage their cultural capital to use data for greater representation of diverse sets of audiences. This should be encouraging for news organizations that aim to improve their relationships with diverse sets of audiences and encourage them to invest in staff that are committed to addressing misrepresentation. Highlighting the interconnection between newsroom equites and issues of representations, these findings call for publishers to value (both in cultural and economic terms) the skills, knowledge, and journalistic intuition of news professionals of traditionally marginalized communities.

In sum, this dissertation makes contributions of theoretical, empirical, and methodological nature that have implications for both scholars and news professionals. This, of course, does not suggest that the dissertation provides final answers to questions about how news organizations should navigate normative issues surrounding AI. Rather, I hope that this work will contribute to a broader conversation about how to support and strengthen attempts to make journalism more equitable in respect to newsroom equality, in relations to diverse audiences, and in their relationship with Big Tech. Looking into the future, there are several important challenges that publishers face that all interconnect with socio-technological changes — economic precarity amid hopes that AI can make news work cheaper and more efficient, audiences trust in journalism that is challenged by an information ecosystem that is not controlled by news institutions, and struggles to place value on journalistic work at a time when it is increasingly being automated. To protect and strengthen journalism, we must ensure that news institutions can grow their capacities to carefully scrutinize technological developments and their impacts on societies, hold power to account, and find new ways to improve their relationships with audiences. Doing so will require

both a commitment from journalists to align technological practices with professional ideals, as well as efforts from academics to provide critical research, looking at the intersection of journalism and technology, with an explicit focus on how power and inequities shape journalism with consequences for society. Some possible avenues are addressed in what follows.

Avenues for Future Research

On an overarching level, I hope that this dissertation encourages future research that aims to integrate a critical data perspective in journalism research. The empirical studies of this dissertation reflect a need for further work that recognizes the mutual shaping of ways of *knowing* and *doing* and considers economic, cultural, and political dimensions of inequities in conjunction. Specifically, research that accounts for the materialities of technologies alongside ideological formations are needed. Likewise, approaches that systematically anchor datafication and journalistic AI within a historical trajectory, that delineates the interconnected evolution of communication technology and society, could provide a more comprehensive understanding of this relationship (Flensburg & Lomborg, 2023, p. 1466). In closing, I briefly touch upon two directions for future research — AI collaborations as possible pathways to mitigate inequities in journalism, and the politics of data loss in journalism.

Studying AI collaborations as possible pathways to mitigate inequities in journalism

News organizations are not on an equal playing field regarding AI adoption, as some with ample resources are able to develop their own in-house AI tools (e.g. Bloomberg developing its own AI model trained on financial data: PYMNTS, 2023), while others, particularly small, local newsrooms struggle to keep up with technological advancements due to resource constraints (Study III). A sustainable and pluralistic media environment demands collaboration, knowledge sharing, and resource exchange, which, as Study III shows, is already underway in practice, despite the fact that research has so far paid little attention to it. Future research could further investigate AI networks as potential pathways to mitigate inequities and explore if initiatives for collaborations lessen the gap between technologically more equipped newsrooms and news

organizations lagging behind socio-technological change. Inter-organizational collaborations can facilitate the pooling of resources within a news organization and ensure that AI tools and data skills are accessible to a broader spectrum of journalists. Equally important are collaborations between news organizations and universities. Universities often serve as hubs of innovation that possess sought-after expertise, that is, especially for news organizations with limited resources, otherwise inaccessible. As the news industry invests further in AI, by building in-house AI tools and the relevant infrastructure to maintain AI projects, collaborations with academic institutions will be critical. By thoroughly investigating the dynamics and outcomes of collaborations, future research can thus provide actionable insights into building a more equitable and sustainable future for journalism, where, hopefully, advancements in AI are more accessible and beneficial across a diverse range of news organizations.

Studying the politics of data loss in journalism

In the scope of this dissertation, I have looked at datafication by focusing on news organization's capacities to extract, produce, and accumulate enormous amounts of data. This perspective is a dominant one in journalism research. However, datafication also comes with unprecedented amounts of data loss — whether those losses are intentional forms of data destructions, planned obsolescence and closure of platforms, or caused by failures and accidents (Bonde Thylstrup, 2022). Data losses are not inherently negative events but can be productive ways of protecting audiences and sources with ethical and political intent. Journalism is a data intensive industry, and news professionals manage various forms of politically sensitive data (Adams, 2020). At the same time, journalism plays a central role in memory culture. Because "journalism continues to function as one of contemporary society's main institutions of recording and remembering, we need to invest more efforts in understanding how it remembers and why it remembers and why it remembers in the way that it does" (Zelizer, 2008). To do so, we also need to understand ways of forgetting. News organizations handle sometimes contradictory needs for preservation and erasure, that are only amplified under datafication. A more fundamental understanding of how datafication structures ways of *knowing* and *doing* in journalism would thus also involve politics data loss. This avenue for future research would open pathways to further integrate critical data studies into journalism scholarship and would likely yield important

contributions beyond these fields, including memory studies, to which journalisms' relations have long been underdeveloped (Olick, 2014).

References

Adams, P. C. (2020). Agreeing to Surveillance: Digital News Privacy Policies. *Journalism & Mass Communication Quarterly*, *97*(4), 868–889. https://doi.org/10.1177/1077699020934197

Amoore, L. (2020). *Cloud ethics: Algorithms and the attributes of ourselves and others*. Duke University Press.

Angwin, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. *ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Aradau, C., & Bunz, M. (2022). Dismantling the apparatus of domination? Left critiques of AI. *RADICAL PHILOSOPHY*, *2*(12), 10–18.

Arvidsson, A., & Colleoni, E. (2012). Value in Informational Capitalism and on the Internet. *The Information Society*, *28*(3), 135–150. https://doi.org/10.1080/01972243.2012.669449

Associated Press. (2023). *Case Study, Newsroom: Michigan Radio*. https://www.ap.org/assets/files/ap-local-ai-michigan-radio-oct-2023.pdf

Audi, R. (2011). *Epistemology: A contemporary introduction to the theory of knowledge* (Third ed). Routledge.

Averett, K. H. (2021). A feminist public sociology of the pandemic: Interviewing about a crisis, during a crisis. *Gender, Work & Organization*, *28*(S2), 321–329. https://doi.org/10.1111/gwao.12616

Barth, S., & De Jong, M. D. T. (2017). The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review. *Telematics and Informatics*, *34*(7), 1038–1058. https://doi.org/10.1016/j.tele.2017.04.013

Beam, M. A. (2014). Automating the News: How Personalized News Recommender System Design Choices Impact News Reception. *Communication Research*, *41*(8), 1019–1041. https://doi.org/10.1177/0093650213497979

Becker, K. B., Simon, F. M., & Crum, C. (2023). *Policies in Parallel? A Comparative Study of Journalistic AI Policies in 52 Global News Organisations* [Preprint]. SocArXiv. https://doi.org/10.31235/osf.io/c4af9

Beckers, K., Walgrave, S., Wolf, H. V., Lamot, K., & Van Aelst, P. (2021). Right-wing Bias in Journalists' Perceptions of Public Opinion. *Journalism Practice*, *15*(2), 243–258. https://doi.org/10.1080/17512786.2019.1703788

Beckett, C. (2019). *New powers, new responsibilities: A global survey of journalism and artificial intelligence.* [Polis, London School of Economics and Political Science]. https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/

Beckett, C., & Yaseen, M. (2023). Generating Change: A global survey of what news organisations are doing with AI. *LSE*. https://static1.squarespace.com/static/64d60527c01ae7106f2646e9/t/656e400a1c23e22da0681e46/1701724190867/Generating+Change+_+The+Journalism+AI+report+_+English.pdf

Belair-Gagnon, V., Nelson, J. L., & Lewis, S. C. (2019). Audience Engagement, Reciprocity, and the Pursuit of Community Connectedness in Public Media Journalism. *Journalism Practice*, *13*(5), 558–575. https://doi.org/10.1080/17512786.2018.1542975

Belair-Gagnon, V., Zamith, R., & Holton, A. E. (2020). Role Orientations and Audience Metrics in Newsrooms: An Examination of Journalistic Perceptions and their Drivers. *Digital Journalism*, *8*(3), 347–366. https://doi.org/10.1080/21670811.2019.1709521

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the New Jim Code*. Polity.

Binns, R., Lyngs, U., Van Kleek, M., Zhao, J., Libert, T., & Shadbolt, N. (2018). Third Party Tracking in the Mobile Ecosystem. *Proceedings of the 10th ACM Conference on Web Science*, 23–31. https://doi.org/10.1145/3201064.3201089

Bittle, C. (2017). *Reality and the mind: Epistemology* (New edition of the 1th edition 1936). Editiones Scholasticae.

Bodó, B., Helberger, N., Eskens, S., & Möller, J. (2019). Interested in Diversity: The role of user attitudes, algorithmic feedback loops, and policy in news personalization. *Digital Journalism*, *7*(2), 206–229. https://doi.org/10.1080/21670811.2018.1521292

Bonde Thylstrup, N. (2022). *Data Loss: Theorizing Digital Disappearance and Remains*. *5*. https://www.dhi.ac.uk/dhc2022/

Bourdieu, P. (1986). The Forms of Capital. In *Handbook of Theory and Research for the Sociology of Education* (pp. 241–258). Greenwood Press.

boyd, D., & Crawford, K. (2012). CRITICAL QUESTIONS FOR BIG DATA: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Broussard, M., Diakopoulos, N., Guzman, A. L., Abebe, R., Dupagne, M., & Chuan, C.-H. (2019). Artificial Intelligence and Journalism. *Journalism & Mass Communication Quarterly*, *96*(3), 673–695. https://doi.org/10.1177/1077699019859901

Bucher, T. (2017). 'Machines don't have instincts': Articulating the computational in journalism. *New Media & Society*, *19*(6), 918–933. https://doi.org/10.1177/1461444815624182

Burns, R., Hawkins, B., Hoffmann, A. L., Iliadis, A., & Thatcher, J. (2018). Transdisciplinary approaches to critical data studies. *Proceedings of the Association for Information*

Science and Technology, *55*(1), 657–660. https://doi.org/10.1002/pra2.2018.14505501074

Burrell, J., & Fourcade, M. (2021). The Society of Algorithms. *Annual Review of Sociology*, *47*(1), 213–237. https://doi.org/10.1146/annurev-soc-090820-020800

Camaj, L., Lanosga, G., & Martin, J. (2023). A Global Perspective on Data Journalism Materiality: Knowledge Production across Public Transparency Infrastructure Environments. *Digital Journalism*, 1–20. https://doi.org/10.1080/21670811.2023.2278040

Carlson, M. (2016). Metajournalistic Discourse and the Meanings of Journalism: Definitional Control, Boundary Work, and Legitimation: Metajournalistic Discourse and the Meanings of Journalism. *Communication Theory*, *26*(4), 349–368. https://doi.org/10.1111/comt.12088

Carlson, M. (2018). Confronting Measurable Journalism. *Digital Journalism*, *6*(4), 406–417. https://doi.org/10.1080/21670811.2018.1445003

Carlson, M. (2020). Journalistic epistemology and digital news circulation: Infrastructure, circulation practices, and epistemic contests. *New Media & Society*, *22*(2), 230–246. https://doi.org/10.1177/1461444819856921

Carlson, M. (2023). Epistemic Contests in Journalism: Examining Struggles over Journalistic Ways of Knowing. *Digital Journalism*, 1–16. https://doi.org/10.1080/21670811.2023.2288392

Carlson, M., & Peters, C. (2023). The Forces Shaping Journalism and Journalism Studies: A Reply to Vos, Craft, and Witschge and Sabbah. *Journalism Studies*, *24*(8), 1061–1068. https://doi.org/10.1080/1461670X.2023.2190824

Catanzariti, B., Chandhiramowuli, S., Mohamed, S., Natarajan, S., Prabhat, S., Raval, N., Taylor, A. S., & Wang, D. (2021). The Global Labours of AI and Data Intensive Systems. *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*, 319–322. https://doi.org/10.1145/3462204.3481725

Chandhiramowuli, S., & Chaudhuri, B. (2023). Match Made by Humans: A Critical Enquiry into Human-Machine Configurations in Data Labelling. *56th Hawaii International Conference on System Sciences (HICCS)*.

Chan-Olmsted, S. M. (2019). A Review of Artificial Intelligence Adoptions in the Media Industry. *International Journal on Media Management*, *21*(3–4), 193–215. https://doi.org/10.1080/14241277.2019.1695619

Cheney-Lippold, J. (2020). *We Are Data: Algorithms and the Making of Our Digital Selves*. New York University Press. https://doi.org/10.18574/nyu/9781479888702.001.0001

Christin, A. (2020). *Metrics at work: Journalism and the contested meaning of algorithms*. Princeton University Press.

Cinnamon, J. (2017). Social Injustice in Surveillance Capitalism. *Surveillance & Society*, *15*(5), 609–625. https://doi.org/10.24908/ss.v15i5.6433

Coddington, M. (2015). Clarifying Journalism's Quantitative Turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*, *3*(3), 331–348. https://doi.org/10.1080/21670811.2014.976400

Cohen, N. S. (2015). From Pink Slips to Pink Slime: Transforming Media Labor in a Digital Age. *The Communication Review*, *18*(2), 98–122. https://doi.org/10.1080/10714421.2015.1031996

Couldry, N. (2010). *Why voice matters: Culture and politics after neoliberalism*. SAGE.

Couldry, N., & Hepp, A. (2017). *The mediated construction of reality*. Polity press.

Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Dalton, C. M., Taylor, L., & Thatcher (Alphabetical), J. (2016). Critical Data Studies: A dialog on data and space. *Big Data & Society*, *3*(1), 205395171664834. https://doi.org/10.1177/2053951716648346

Dalton, C., & Thatcher, J. (2014). What Does A Critical Data Studies Look Like, And Why Do We Care? *Society + Space*. https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care

de Haan, Y., van den Berg, E., Goutier, N., Kruikemeier, S., & Lecheler, S. (2022). Invisible Friend or Foe?How Journalists Use and Perceive Algorithmic-Driven Tools in Their Research Process. *Digital Journalism*, 1–19. https://doi.org/10.1080/21670811.2022.2027798

De La Bellacasa, M. P. (2012). 'Nothing Comes Without Its World': Thinking with Care. *The Sociological Review*, *60*(2), 197–216. https://doi.org/10.1111/j.1467-954X.2012.02070.x

Delfanti, A. (2021). Machinic dispossession and augmented despotism: Digital work in an Amazon warehouse. *New Media & Society*, *23*(1), 39–55. https://doi.org/10.1177/1461444819891613

Dencik, L., Hintz, A., Redden, J., & Treré, E. (2019). Exploring Data Justice: Conceptions, Applications and Directions. *Information, Communication & Society*, *22*(7), 873–881.

https://doi.org/10.1080/1369118X.2019.1606268

Dencik, L., & Sanchez-Monedero, J. (2022). Data justice. *Internet Policy Review*, *11*(1). https://doi.org/10.14763/2022.1.1615

Deuze, M. (2005). What is journalism?: Professional identity and ideology of journalists reconsidered. *Journalism*, *6*(4), 442–464. https://doi.org/10.1177/1464884905056815

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.

Diakopoulos, N. (2020). Computational News Discovery: Towards Design Considerations for Editorial Orientation Algorithms in Journalism. *Digital Journalism*, *8*(7), 945–967. https://doi.org/10.1080/21670811.2020.1736946

D'Ignazio, C., & Klein, L. F. (2023). *Data feminism* (First MIT Press paperback edition). The Mit Press.

Dörr, K. N., & Hollnbuchner, K. (2017). Ethical Challenges of Algorithmic Journalism. *Digital Journalism*, *5*(4), 404–419. https://doi.org/10.1080/21670811.2016.1167612

Eiguren, A., Idoiaga, N., Berasategi, N., & Picaza, M. (2021). Exploring the Social and Emotional Representations Used by the Elderly to Deal With the COVID-19 Pandemic. *Frontiers in Psychology*, *11*, 586560. https://doi.org/10.3389/fpsyg.2020.586560

Espeland, W. N., & Stevens, M. L. (2008). A Sociology of Quantification. *European Journal of Sociology*, *49*(3), 401–436. https://doi.org/10.1017/S0003975609000150

Etzioni, A. (1961). *A Comparative Analysis of Complex Organizations on Power, Involvement, and Their Correlates.* Free Press of Glencoe.

Eubanks, V. (2017). *Automating inequality: How high-tech tools profile, police, and punish the poor* (First Edition). St. Martin's Press.

Evens, T., & Van Damme, K. (2016). Consumers' Willingness to Share Personal Data: Implications for Newspapers' Business Models. *International Journal on Media Management*, *18*(1), 25–41. https://doi.org/10.1080/14241277.2016.1166429

Ferrer-Conill, R., & Tandoc, E. C. (2018). The Audience-Oriented Editor: Making sense of the audience in the newsroom. *Digital Journalism*, *6*(4), 436–453. https://doi.org/10.1080/21670811.2018.1440972

Fink, K., & Anderson, C. W. (2015). Data Journalism in the United States: Beyond the "usual suspects." *Journalism Studies*, *16*(4), 467–481. https://doi.org/10.1080/1461670X.2014.939852

Flensburg, S., & Lomborg, S. (2023). Datafication research: Mapping the field for a future agenda. *New Media & Society*, *25*(6), 1451–1469. https://doi.org/10.1177/14614448211046616

Fuchs, C. (2010). Labor in Informational Capitalism and on the Internet. *The Information Society*, *26*(3), 179–196. https://doi.org/10.1080/01972241003712215

Gagrčin, E., & Butkowski, C. (2023). *Out of Sight, Out of Mind? Qualitative Methods in Political Communication Research*. 6 S. https://doi.org/10.17169/REFUBIUM-39042

Gangadharan, S. P. (2017). The downside of digital inclusion: Expectations and experiences of privacy and surveillance among marginal Internet users. *New Media & Society*, *19*(4), 597–615. https://doi.org/10.1177/1461444815614053

García Avilés, J. A., León, B., Sanders, K., & Harrison, J. (2004). Journalists at digital television newsrooms in Britain and Spain: Workflow and multi-skilling in a competitive environment. *Journalism Studies*, *5*(1), 87–100. https://doi.org/10.1080/1461670032000174765

Gerrard, Y., & Bates, J. (2019). Introduction to the Data Power Special Issue: Tactics, access and shaping. *Online Information Review*, *43*(6), 945–951. https://doi.org/10.1108/OIR-10-2019-396

Gieryn, T. F. (1999). *Cultural boundaries of science: Credibility on the line*. University of Chicago Press.

Gillespie, T. (2014). The Relevance of Algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media Technologies* (pp. 167–194). The MIT Press. https://doi.org/10.7551/mitpress/9780262525374.003.0009

Graham, M., Hjorth, I., & Lehdonvirta, V. (2017). Digital labour and development: Impacts of global digital labour platforms and the gig economy on worker livelihoods. *Transfer: European Review of Labour and Research*, *23*(2), 135–162. https://doi.org/10.1177/1024258916687250

Guo, L., & Volz, Y. (2019). (Re)defining Journalistic Expertise in the Digital Transformation: A Content Analysis of Job Announcements. *Journalism Practice*, *13*(10), 1294–1315. https://doi.org/10.1080/17512786.2019.1588147

Gutierrez Lopez, M., Porlezza, C., Cooper, G., Makri, S., MacFarlane, A., & Missaoui, S. (2023). A Question of Design: Strategies for Embedding AI-Driven Tools into Journalistic Work Routines. *Digital Journalism*, *11*(3), 484–503. https://doi.org/10.1080/21670811.2022.2043759

Hamilton, J. (2016). *Democracy's detectives: The economics of investigative journalism*. Harvard university press.

Hampton, M. (2010). The fourth estate ideal in journalism history. In *The Routledge companion to news and journalism* (pp. 3–12). Routledge.

Hanusch, F., & Banjac, S. (2021). Do Journalists Share Universal Values? In S. J. A. Ward (Ed.), *Handbook of Global Media Ethics* (pp. 71–90). Springer International Publishing. https://doi.org/10.1007/978-3-319-32103-5_5

Haraway, D. J. (2014). Speculative Fabulations for Technoculture's Generations: Taking Care of Unexpected Country. In E. Kirksey (Ed.), *The Multispecies Salon* (pp. 242–261). Duke University Press. https://doi.org/10.1215/9780822376989-016

Heeks, R., & Renken, J. (2018). Data justice for development: What would it mean? *Information Development*, *34*(1), 90–102. https://doi.org/10.1177/0266666916678282

Helberger, N., Van Drunen, M., Moeller, J., Vrijenhoek, S., & Eskens, S. (2022). Towards a Normative Perspective on Journalistic AI: Embracing the Messy Reality of Normative Ideals. *Digital Journalism*, *10*(10), 1605–1626. https://doi.org/10.1080/21670811.2022.2152195

Herrman, J. (2023). How Will Artificial Intelligence Change the News Business? Here are three theories of the case. *Intelligencer, New York Magazine*. https://nymag.com/intelligencer/2023/08/how-ai-will-change-the-news-business.html

Heuer, H., Jarke, J., & Breiter, A. (2021). Machine learning in tutorials – Universal applicability, underinformed application, and other misconceptions. *Big Data & Society*, *8*(1), 205395172110175. https://doi.org/10.1177/20539517211017593

Hikins, J., & Cherwitz, R. (2011). On the Ontological and Epistemological Dimensions of Expertise: Why "Reality" and "Truth" Matter and How We Might Find Them. *Social Epistemology*, *25*(3), 291–308. https://doi.org/10.1080/02691728.2011.578304

Hintz, A., Treré, E., & Owen, N. (2022). *Journalism and Data Justice* (2nd ed). Routledge.

Hogan, M. (2015). Data flows and water woes: The Utah Data Center. *Big Data & Society*, *2*(2),

205395171559242. https://doi.org/10.1177/2053951715592429

Honebein, P. C. (1996). Seven goals for the design of constructivist learning environments. In *Constructivist learning environments: Case studies in instructional design.* Wilson, Brent. G.

Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, *3*(2), 205395171667423. https://doi.org/10.1177/2053951716674238

Jones, B., Jones, R., & Luger, E. (2022). AI 'Everywhere and Nowhere': Addressing the AI Intelligibility Problem in Public Service Journalism. *Digital Journalism*, *10*(10), 1731–1755. https://doi.org/10.1080/21670811.2022.2145328

Kammer, A. (2021). Resource Exchanges Between Mobile News Apps and Third-Parties. *Digital Journalism*, 1–19. https://doi.org/10.1080/21670811.2021.2000455

Karlsen, J., & Stavelin, E. (2014). Computational Journalism in Norwegian Newsrooms. *Journalism Practice*, *8*(1), 34–48. https://doi.org/10.1080/17512786.2013.813190

Katz, Y. (2020). *Artificial whiteness: Politics and ideology in artificial intelligence*. Columbia University Press.

Keen, S., Lomeli-Rodriguez, M., & Joffe, H. (2022). From Challenge to Opportunity: Virtual Qualitative Research During COVID-19 and Beyond. *International Journal of Qualitative Methods*, *21*, 160940692211050. https://doi.org/10.1177/16094069221105075

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 205395171452848. https://doi.org/10.1177/2053951714528481

Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Harvard University Press.

Knorr-Cetina, K. (2007). Culture in global knowledge societies: Knowledge cultures and epistemic cultures. *Interdisciplinary Science Reviews*, *32*(4), 361–375. https://doi.org/10.1179/030801807X163571

Koivula, M., Saari, T., & Villi, M. (2023). "I love learning new things": An institutional logics perspective on learning in professional journalism. *Journalism*, 146488492311677. https://doi.org/10.1177/14648849231167786

Komatsu, T., Gutierrez Lopez, M., Makri, S., Porlezza, C., Cooper, G., MacFarlane, A., & Missaoui, S. (2020). AI should embody our values: Investigating journalistic values to inform AI technology design. *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, 1–13. https://doi.org/10.1145/3419249.3420105

Kosterich, A. (2020). Managing news nerds: Strategizing about institutional change in the news industry. *Journal of Media Business Studies*, *17*(1), 51–68. https://doi.org/10.1080/16522354.2019.1639890

Kosterich, A. (2021). Reengineering Journalism: Product Manager as News Industry Institutional Entrepreneur. *Digital Journalism*, 1–30. https://doi.org/10.1080/21670811.2021.1903959

Kotenidis, E., & Veglis, A. (2021). Algorithmic Journalism—Current Applications and Future Perspectives. *Journalism and Media*, *2*(2), 244–257. https://doi.org/10.3390/journalmedia2020014

Kothari, A., & Hickerson, A. (2020). Challenges for journalism education in the era of automation. *Media Practice and Education*, *21*(3), 212–228. https://doi.org/10.1080/25741136.2020.1787724

Kuhn, T. S. (2023). *Die Struktur wissenschaftlicher Revolutionen* (H. Vetter, Trans.; Zweite

revidierte und um das Postskriptum von 1969 ergänzte Auflage, 27. Auflage). Suhrkamp.

Laaksonen, S.-M., Koivula, M., & Villi, M. (2022). Mediated by the giants: Tracing practices, discourses, and mediators of platform isomorphism in a media organization. *New Media & Society*, 146144482211222. https://doi.org/10.1177/14614448221122220

Langer, M., Hunsicker, T., Feldkamp, T., König, C. J., & Grgić-Hlača, N. (2022). "Look! It's a Computer Program! It's an Algorithm! It's AI!": Does Terminology Affect Human Perceptions and Evaluations of Algorithmic Decision-Making Systems? *CHI Conference on Human Factors in Computing Systems*, 1–28. https://doi.org/10.1145/3491102.3517527

Lathen, L., & Laestadius, L. (2021). Reflections on Online Focus Group Research With Low Socio-Economic Status African American Adults During COVID-19. *International Journal of Qualitative Methods*, *20*, 160940692110217. https://doi.org/10.1177/16094069211021713

Latour, B. (2004). Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern. *Critical Inquiry*, *30*. http://www.bruno-latour.fr/sites/default/files/89-CRITICAL-INQUIRY-GB.pdf

Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts* (1. Princeton paperback print). Princeton Univ. Press.

Lavazza, A., & Farina, M. (2023). Infosphere, Datafication, and Decision-Making Processes in the AI Era. *Topoi*, *42*(3), 843–856. https://doi.org/10.1007/s11245-023-09919-0

Lecher, C., & Varner, M. (2021). *NYC's School Algorithms Cement Segregation. This Data Shows How*. https://themarkup.org/machine-learning/2021/05/26/nycs-school-algorithms-cement-segregation-this-data-shows-how

Lewis, S. C., & Usher, N. (2016). Trading zones, boundary objects, and the pursuit of news innovation: A case study of journalists and programmers. *Convergence: The International Journal of Research into New Media Technologies*, *22*(5), 543–560. https://doi.org/10.1177/1354856515623865

Lewis, S. C., & Zamith, R. (2017). On the Worlds of Journalism. In P. J. Boczkowski & C. W. Anderson (Eds.), *Remaking the News* (pp. 111–128). The MIT Press. https://doi.org/10.7551/mitpress/10648.003.0012

Liang, Y., Aroles, J., & Brandl, B. (2022). Charting platform capitalism: Definitions, concepts and ideologies. *New Technology, Work and Employment*, *37*(2), 308–327. https://doi.org/10.1111/ntwe.12234

Lin, B., & Lewis, S. C. (2022). The One Thing Journalistic AI Just Might Do for Democracy. *Digital Journalism*, *10*(10), 1627–1649. https://doi.org/10.1080/21670811.2022.2084131

Lindgren, S. (2023). *Critical theory of AI*. Polity Press.

Lischka, J. A., Schaetz, N., & Oltersdorf, A.-L. (2022). Editorial Technologists as Engineers of Journalism's Future: Exploring the Professional Community of Computational Journalism. *Digital Journalism*, 1–19. https://doi.org/10.1080/21670811.2021.1995456

Livingstone, S. (2019). Audiences in an Age of Datafication: Critical Questions for Media Research. *Television & New Media*, *20*(2), 170–183. https://doi.org/10.1177/1527476418811118

Lobe, B., Morgan, D., & Hoffman, K. A. (2020). Qualitative Data Collection in an Era of Social Distancing. *International Journal of Qualitative Methods*, *19*, 160940692093787. https://doi.org/10.1177/1609406920937875

Luchs, I., Apprich, C., & Broersma, M. (2023). Learning machine learning: On the political

economy of big tech's online AI courses. *Big Data & Society*, *10*(1), 205395172311538. https://doi.org/10.1177/20539517231153806

Marconi, F. (2020). *Newsmakers: Artificial Intelligence and the Future of Journalism*. Columbia University Press. https://doi.org/10.7312/marc19136

Martinez, E., & Kirchner, L. (2021). The Secret Bias Hidden in Mortgage-Approval Algorithms. *The Markup*. https://themarkup.org/denied/2021/08/25/the-secret-bias-hidden-in-mortgage-approval-algorithms

Mayer-Schönberger, V., & Cukier, K. (2014). *Big data: A revolution that will transform how we live, work, and think* (First Mariner Books edition). Mariner Books, Houghton Mifflin Harcourt.

Meese, J., & Hurcombe, E. (2021). Facebook, news media and platform dependency: The institutional impacts of news distribution on social platforms. *New Media & Society*, *23*(8), 2367–2384. https://doi.org/10.1177/1461444820926472

Min, S. J., & Fink, K. (2021). Keeping Up with the Technologies: Distressed Journalistic Labor in the Pursuit of "Shiny" Technologies. *Journalism Studies*, *22*(14), 1987–2004. https://doi.org/10.1080/1461670X.2021.1979425

Møller Hartley, J., Petre, C., Bengtsson, M., & Kammer, A. (2023). Autonomies and Dependencies: Shifting Configurations of Power in the Platformization of News. *Digital Journalism*, *11*(8), 1375–1390. https://doi.org/10.1080/21670811.2023.2257759

Møller, L. A. (2022a). Between Personal and Public Interest: How Algorithmic News Recommendation Reconciles with Journalism as an Ideology. *Digital Journalism*, *10*(10), 1794–1812. https://doi.org/10.1080/21670811.2022.2032782

Møller, L. A. (2022b). Recommended for You: How Newspapers Normalise Algorithmic News

Recommendation to Fit Their Gatekeeping Role. *Journalism Studies*, *23*(7), 800–817. https://doi.org/10.1080/1461670X.2022.2034522

Molyneux, L., & Nelson, J. L. (2023). "Let's Not Tank the Reputation of This Organization." How Newsroom Social Media Policies Exacerbate Journalism's Labor Crisis. *Journalism Studies*, 1–20. https://doi.org/10.1080/1461670X.2023.2263797

Mumford, D. (2022). Data colonialism: Compelling and useful, but whither epistemes? *Information, Communication & Society*, *25*(10), 1511–1516. https://doi.org/10.1080/1369118X.2021.1986103

Myllylahti, M. (2018). An attention economy trap? An empirical investigation into four news companies' Facebook traffic and social media revenue. *Journal of Media Business Studies*, *15*(4), 237–253. https://doi.org/10.1080/16522354.2018.1527521

Myllylahti, M. (2020). Paying Attention to Attention: A Conceptual Framework for Studying News Reader Revenue Models Related to Platforms. *Digital Journalism*, *8*(5), 567–575. https://doi.org/10.1080/21670811.2019.1691926

Nechushtai, E. (2018). Could digital platforms capture the media through infrastructure? *Journalism*, *19*(8), 1043–1058. https://doi.org/10.1177/1464884917725163

Nelson, J. L. (2021). The next media regime: The pursuit of 'audience engagement' in journalism. *Journalism*, *22*(9), 2350–2367. https://doi.org/10.1177/1464884919862375

Nielsen, R. K., & Ganter, S. A. (2022). *The power of platforms: Shaping media and society*. Oxford University Press.

Nishal, S., & Diakopoulos, N. (2023). Envisioning the Applications and Implications of Generative AI for News Media. *Envisioning the Applications and Implications of Generative AI for News Media*. CHI '23 Generative AI and HCI Workshop, Hamburg,

Germany. https://nishalsach.github.io/pdfs/2023-genaihci-chi.pdf

Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford Law Books, an imprit of Standford University Press.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York university press.

Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. *Patterns*, *2*(10), 100347. https://doi.org/10.1016/j.patter.2021.100347

O'Brien, D., Wellbrock, C.-M., & Kleer, N. (2020). Content for Free? Drivers of Past Payment, Paying Intent and Willingness to Pay for Digital Journalism – A Systematic Literature Review. *Digital Journalism*, *8*(5), 643–672. https://doi.org/10.1080/21670811.2020.1770112

Olick, J. K. (2014). Reflections on the Underdeveloped Relations between Journalism and Memory Studies. In B. Zelizer & K. Tenenboim-Weinblatt (Eds.), *Journalism and Memory* (pp. 17–31). Palgrave Macmillan UK. https://doi.org/10.1057/9781137263940_2

Opdahl, A. L., Tessem, B., Dang-Nguyen, D.-T., Motta, E., Setty, V., Throndsen, E., Tverberg, A., & Trattner, C. (2023). Trustworthy journalism through AI. *Data & Knowledge Engineering*, *146*, 102182. https://doi.org/10.1016/j.datak.2023.102182

Örnebring, H., Lindell, J., Clerwall, C., & Karlsson, M. (2016). Dimensions of Journalistic Workplace Autonomy: A Five-nation Comparison. *Javnost - The Public*, *23*(3), 307–326. https://doi.org/10.1080/13183222.2016.1215833

Paik, S. (2023). Journalism Ethics for the Algorithmic Era. *Digital Journalism*, 1–27. https://doi.org/10.1080/21670811.2023.2200195

Papaevangelou, C. (2023). Funding Intermediaries: Google and Facebook's Strategy to Capture Journalism. *Digital Journalism*, 1–22. https://doi.org/10.1080/21670811.2022.2155206

Parasie, S. (2015). Data-Driven Revelation?: Epistemological tensions in investigative journalism in the age of "big data." *Digital Journalism*, *3*(3), 364–380. https://doi.org/10.1080/21670811.2014.976408

Parasie, S., & Dagiral, E. (2013). Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago. *New Media & Society*, *15*(6), 853–871. https://doi.org/10.1177/1461444812463345

Pellegrino, G., Söderberg, J., & Milan, S. (2019). Datafication from Below: Epistemology, Ambivalences, Challenges. *Tecnoscienza – Italian Journal of Science & Technology Studies*, 89-113 Pages. https://doi.org/10.6092/ISSN.2038-3460/17431

Petre, C. (2021). *All the news that's fit to click: How metrics are transforming the work of journalists*. Princeton University Press.

Pittaway, E., Bartolomei, L., & Hugman, R. (2010). "Stop Stealing Our Stories": The Ethics of Research with Vulnerable Groups. *Journal of Human Rights Practice*, *2*(2), 229–251. https://doi.org/10.1093/jhuman/huq004

Plaisance, P. L., Skewes, E. A., & Hanitzsch, T. (2012). Ethical Orientations of Journalists Around the Globe: Implications From a Cross-National Survey. *Communication Research*, *39*(5), 641–661. https://doi.org/10.1177/0093650212450584

Porlezza, C. (2023). The datafication of digital journalism: A history of everlasting challenges between ethical issues and regulation. *Journalism*, 14648849231190232. https://doi.org/10.1177/14648849231190232

Porlezza, C., & Ferri, G. (2022). The Missing Piece: Ethics and the Ontological Boundaries of

Automated Journalism. *ISOJ Journal*, *12*(1), 71–98.

Prietl, B., & Houben, D. (2018). Einführung. Soziologische Perspektiven auf die Datafizierung der Gesellschaft. In D. Houben & B. Prietl (Eds.), *Datengesellschaft* (pp. 7–32). transcript Verlag. https://doi.org/10.1515/9783839439579-001

PYMNTS. (2023). *Bloomberg Develops Generative AI Model Trained on Financial Data*.

Rajagopalan, M., Killing, A., & Buschek, C. (2020). *China Built A Vast New Infrastructure To Imprison Uighurs*. https://www.buzzfeednews.com/article/meghara/china-new-internment-camps-xinjiang-uighurs-muslims

Raley, R., & Rhee, J. (2023). Critical AI: A Field in Formation. *American Literature*, *95*(2), 264.

Reese, S. D. (2016). The New Geography of Journalism Research: Levels and spaces. *Digital Journalism*, *4*(7), 816–826. https://doi.org/10.1080/21670811.2016.1152903

Relly, J. E., & González De Bustamante, C. (2014). Silencing Mexico: A Study of Influences on Journalists in the Northern States. *The International Journal of Press/Politics*, *19*(1), 108–131. https://doi.org/10.1177/1940161213509285

Renieris, E. M. (2023). *Beyond data: Reclaiming human rights at the dawn of the metaverse*. The MIT Press.

Rinehart, A., & Kung, E. (2022). *Artificial Intelligence in Local News*. AP. https://www.ap.org/assets/files/ap_local_news_ai_report_march_2022.pdf

Rojas Torrijos, J. L. (2021). Semi-automated Journalism: Reinforcing Ethics to Make the Most of Artificial Intelligence for Writing News. In M. Luengo & S. Herrera-Damas (Eds.), *News Media Innovation Reconsidered* (1st ed., pp. 124–137). Wiley. https://doi.org/10.1002/9781119706519.ch8

Royal, C., & Kiesow, D. (2021). From Boundary to Bridge and Beyond: The Path to

Professionalization of Product Roles in Journalism. *Journalism Studies*, *22*(11), 1546–1565. https://doi.org/10.1080/1461670X.2021.1944277

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, *6*(1), 205395171882054. https://doi.org/10.1177/2053951718820549

Sankin, A., & Mattu, S. (2023). Predictive Policing Software Terrible At Predicting Crimes. *The Markup*. https://themarkup.org/prediction-bias/2023/10/02/predictive-policing-software-terrible-at-predicting-crimes

Schaetz, N., Gagrčin, E., Toth, R., & Emmer, M. (2023). Algorithm dependency in platformized news use. *New Media & Society*, 14614448231193093. https://doi.org/10.1177/14614448231193093

Schapals, A. K., & Porlezza, C. (2020). Assistance or Resistance? Evaluating the Intersection of Automated Journalism and Journalistic Role Conceptions. *Media and Communication*, *8*(3), 16–26. https://doi.org/10.17645/mac.v8i3.3054

Schjøtt Hansen, A., & Hartley, J. M. (2021). Designing What's News: An Ethnography of a Personalization Algorithm and the Data-Driven (Re)Assembling of the News. *Digital Journalism*, 1–19. https://doi.org/10.1080/21670811.2021.1988861

Schultz, J. (1998). *Reviving the Fourth Estate: Democracy, Accountability and the Media* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511597138

Schwarzkopf, S. (2020). Sacred Excess: Organizational Ignorance in an Age of Toxic Data. *Organization Studies*, *41*(2), 197–217. https://doi.org/10.1177/0170840618815527

Sewell, G. (1998). The Discipline of Teams: The Control of Team-Based Industrial Work Through Electronic and Peer Surveillance. *Administrative Science Quarterly*, *43*(2), 397. https://doi.org/10.2307/2393857

Simon, F. M. (2022). Uneasy Bedfellows: AI in the News, Platform Companies and the Issue of Journalistic Autonomy. *Digital Journalism*, 1–23. https://doi.org/10.1080/21670811.2022.2063150

Sirén-Heikel, S., Kjellman, M., & Lindén, C. (2023). At the crossroads of logics: Automating newswork with artificial intelligence — (Re)defining journalistic logics from the perspective of technologists. *Journal of the Association for Information Science and Technology*, *74*(3), 354–366. https://doi.org/10.1002/asi.24656

Sjøvaag, H. (2016). Introducing the Paywall: A case study of content changes in three online newspapers. *Journalism Practice*, *10*(3), 304–322. https://doi.org/10.1080/17512786.2015.1017595

Sosa, E., Kim, J., Fantl, J., & McGrath, M. (2008). *Epistemology: An anthology* (Second edition). Blackwell.

Srnicek, N. (2017). *Platform capitalism*. Polity.

Steensen, S., & Westlund, O. (2020). *What is Digital Journalism Studies?* (1st ed.). Routledge. https://doi.org/10.4324/9780429259555

Stockinger, A., Schäfer, S., & Lecheler, S. (2023). Navigating the gray areas of content moderation: Professional moderators' perspectives on uncivil user comments and the role of (AI-based) technological tools. *New Media & Society*, 14614448231190901. https://doi.org/10.1177/14614448231190901

Swisher, K., & Galloway, S. (Directors). (2023). Writers Strike Deal, Betting Against the FTC, and Guest Mathias Döpfner. In *Pivot*. Vox Media.

Tactical Tech. (2022). *Megha Rajagopalan: What I've Learned About Investigative Journalism*. https://gijn.org/stories/megha-rajagopalan-what-ive-learned-about-investigative-

journalism/

Tandoc, E. C. (2014). Journalism is twerking? How web analytics is changing the process of gatekeeping. *New Media & Society*, *16*(4), 559–575. https://doi.org/10.1177/1461444814530541

Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, *4*(2), 205395171773633. https://doi.org/10.1177/2053951717736335

Thatcher, J., O'Sullivan, D., & Mahmoudi, D. (2016). Data colonialism through accumulation by dispossession: New metaphors for daily data. *Environment and Planning D: Society and Space*, *34*(6), 990–1006. https://doi.org/10.1177/0263775816633195

The Markup. (2020). *The Citizen Browser Project—Auditing the Algorithms of Disinformation*. https://themarkup.org/citizen-browser

Thylstrup, N. B. (Ed.). (2021). *Uncertain archives: Critical keywords for big data*. The MIT Press.

Tichenor, M., Merry, S. E., Grek, S., & Bandola-Gill, J. (2022). Global public policy in a quantified world: Sustainable Development Goals as epistemic infrastructures. *Policy and Society*, puac015. https://doi.org/10.1093/polsoc/puac015

Törnberg, P. (2023). How platforms govern: Social regulation in digital capitalism. *Big Data & Society*, *10*(1), 205395172311538. https://doi.org/10.1177/20539517231153808

Tumber, H. (2001). DEMOCRACY IN THE INFORMATION AGE: THE ROLE OF THE FOURTH ESTATE IN CYBERSPACE. *Information, Communication & Society*, *4*(1), 95–112. https://doi.org/10.1080/13691180122542

Uliasz, R. (2021). Seeing like an algorithm: Operative images and emergent subjects. *AI &*

SOCIETY, *36*(4), 1233–1241. https://doi.org/10.1007/s00146-020-01067-y

Usher, N. (2016). *Interactive journalism: Hackers, data, and code*. University of Illinois Press.

Usher, N. (2019). Putting "Place" in the Center of Journalism Research: A Way Forward to Understand Challenges to Trust and Knowledge in News. *Journalism & Communication Monographs*, *21*(2), 84–146. https://doi.org/10.1177/1522637919848362

Usher, N. (2021). *News for the rich, white, and blue: How place and power distort American journalism*. Columbia University Press.

Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, *12*(2), 197–208. https://doi.org/10.24908/ss.v12i2.4776

van Ess, H. (2023). Facial Recofnition Made Easy. *Global Investigative Journalism Network*. https://gijn.org/resource/facial-recognition-made-easy/#:~:text=Accountability%20and%20Holding%20Power%20to,criminal%20activities%20or%20other%20wrongdoing.

Waisbord, S. (2016). Professionalism, Journalistic Role Performance, and Situated Ethics Beyond the West. In *Journalistic Role Performance* (1st ed.). Routledge.

Waisbord, S. R. (2000). *Watchdog journalism in South America: News, accountability, and democracy*. Columbia University Press.

Wang, S. (2021). Moderating Uncivil User Comments by Humans or Machines? The Effects of Moderation Agent on Perceptions of Bias and Credibility in News Content. *Digital Journalism*, *9*(1), 64–83. https://doi.org/10.1080/21670811.2020.1851279

Wasserman, H. (2010). Freedom's just another word? Perspectives on media freedom and responsibility in South Africa and Namibia. *International Communication Gazette*, *72*(7),

567–588. https://doi.org/10.1177/1748048510378145

West, S. M. (2019). Data Capitalism: Redefining the Logics of Surveillance and Privacy. *Business & Society*, *58*(1), 20–41. https://doi.org/10.1177/0007650317718185

Winner, L. (2001). *Autonomous technology: Technics-out-of-control as a theme in political thought* (9. printing). MIT Pr.

Wu, S., Tandoc, E. C., & Salmon, C. T. (2019). A Field Analysis of Journalism in the Automation Age: Understanding Journalistic Transformations and Struggles Through Structure and Agency. *Digital Journalism*, *7*(4), 428–446. https://doi.org/10.1080/21670811.2019.1620112

Wu, T. (2017). *The attention merchants: The epic struggle to get inside our heads* (Paperback edition). Atlantic Books.

Xu, Y., & Jin, J. (2017). The Hierarchy of Influences on Professional Role Perceptions Among Chinese Online Journalists: A multilevel analysis. *Digital Journalism*, *5*(2), 194–212. https://doi.org/10.1080/21670811.2016.1162662

Zamith, R. (2018). Quantified Audiences in News Production: A synthesis and research agenda. *Digital Journalism*, *6*(4), 418–435. https://doi.org/10.1080/21670811.2018.1444999

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power* (First edition). PublicAffairs.

Appendix I: Erklärung

Hiermit erkläre ich, Nadja Schaetz, dass ich keine kommerzielle Promotionsberatung in Anspruch genommen habe. Die Arbeit wurde nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

Hamburg, 31.01.2024 
Ort/Datum Unterschrift Doktorandin

Appendix II: Eidesstattliche Versicherung

Ich, Nadja Schaetz, versichere an Eides statt, dass ich die Dissertation mit dem Titel: *„From Epistemology to Practice: How Datafication and AI are Shape Inequities in Journalism“* selbst und bei einer Zusammenarbeit mit anderen Wissenschaftlerinnen gemäß den beigefügten Darlegungen nach § 6 Abs. 3 der Promotionsordnung der Fakultät für Wirtschafts- und Sozialwissenschaften vom 18. Januar 2017 verfasst habe. Andere als die angegebenen Hilfsmittel habe ich nicht benutzt.

Hamburg, 31.01.2024 

Ort/Datum Unterschrift Doktorandin

Appendix III: Selbstdeklaration bei kumulativen Promotionen

Konzeption / Planung: Formulierung des grundlegenden wissenschaftlichen Problems, basierend auf bisher unbeantworteten theoretischen Fragestellungen inklusive der Zusammenfassung der generellen Fragen, die anhand von Analysen oder Experimenten/Untersuchungen beantwortbar sind. Planung der Experimente/ Analysen und Formulierung der methodischen Vorgehensweise, inklusive Wahl der Methode und unabhängige methodologische Entwicklung. Durchführung: Grad der Einbindung in die konkreten Untersuchungen bzw. Analysen. Manuskripterstellung: Präsentation, Interpretation und Diskussion der erzielten Ergebnisse in Form eines wissenschaftlichen Artikels. Die Einschätzung des geleisteten Anteils erfolgt mittels Punkteinschätzung von 1 – 100 %. Für mindestens einen der vorliegenden Artikel liegt die Eigenleistung bei 100 %.

Für Studie II liegt die Eigenleistung für
das Konzept / die Planung bei 50 %
die Durchführung bei 25 %
der Manuskripterstellung bei 65 %

Für Studie III liegt die Eigenleistung bei
das Konzept / die Planung bei 33%
die Durchführung bei 60 %
der Manuskripterstellung bei 40 %

Für Studie IV liegt die Eigenleistung bei
das Konzept / die Planung bei 50 %
die Durchführung bei 50 %
der Manuskripterstellung bei 80 %

Die vorliegende Einschätzung in Prozent über die von mir erbrachte Eigenleistung wurde mit den am Artikel beteiligten Koautoren einvernehmlich abgestimmt.

Hamburg, 31.01.2024 _____

Ort/Datum Unterschrift Doktorand/in

Volume 6 Number 7 September 2018

DIGITAL JOURNALISM

Routledge

Digital Journalism

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rdij20

Journalism & Audience Datafication: How Audience Data Practices Shape Inequity

Nadja Schaetz

To cite this article: Nadja Schaetz (2023): Journalism & Audience Datafication: How Audience Data Practices Shape Inequity, Digital Journalism, DOI: 10.1080/21670811.2023.2171894

To link to this article: https://doi.org/10.1080/21670811.2023.2171894

Published online: 04 Feb 2023.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rdij20

DIGITAL JOURNALISM
https://doi.org/10.1080/21670811.2023.2171894

Routledge
Taylor & Francis Group

Check for updates

Journalism & Audience Datafication: How Audience Data Practices Shape Inequity

Nadja Schaetz

Universität Hamburg, Journalism and Communication Studies, Hamburg, Germany

ABSTRACT

Datafication is embedded in cultural, economic, and political power structures which reinforce social inequities. Instead of simply providing news professionals with insights on user behavior, datafication may facilitate maldistribution, misrecognition and misrepresentation. Applying justice theory on audience data practices based on $n = 31$ interviews with news professionals working for global and national news organizations (including *BBC World, The Guardian, Al-Jazeera English*, and *The New York Times*), this study examines their experiences and perceptions of how audience data practices mitigate and/or reinforce inequity in journalism. Findings show that maldistribution, misrecognition, and misrepresentation are manifest in journalistic audience data practices, as inequities are reinforced when data is transformed into economic capital. At the same time, news professionals who possess cultural and economic resources can both mitigate inequity and use data for greater recognition and representation. The article thus contributes to the literature by (1) conceptualizing audience data as a cultural, economic and political good, (2) connecting data practices to both the reproduction and mitigation of social inequity, and finally, (3) examining these processes on a global scale.

KEYWORDS

Datafication; audience data; audience analytics; digital journalism; justice; inequalities

Introduction

Data practices are often deeply intertwined with the exacerbation of social inequities (Couldry and Meijas 2019; Eubanks 2018; Noble 2018). On a global scale, processes of datafication are not only rooted in capitalism but draw on longer interrelations between capitalism and colonialism (Couldry and Meijas 2019). This calls for a shift from ahistorical notions of fairness to a model of justice that acknowledges that inequities are historic and ongoing (Costanza-Chock 2020).

News professionals aim to leverage audience data to uncover audiences' informational needs, and consequently, to improve their ability to meet their diverse expectations across time and context — a key precondition for mitigating inequities. That is, data might be used to examine audiences' engagement with the news in order to provide information that has 'proportional relevance' (Costera Meijer 2003) to different

CONTACT Nadja Schaetz nadja.schaetz@uni-hamburg.de

© 2023 Informa UK Limited, trading as Taylor & Francis Group

communities (Kormelink and Meijer 2018, p. 681) and to offer adequate representation. However, as news organizations are "embedded in networks of commercial and political power" (Couldry 2010, p. 148), audience data practices are inevitably intertwined with issues of economic distribution, cultural recognition, and political representation — dimensions of justice previously highlighted by Fraser (2009). By drawing on the experiences and perceptions of news professionals, this study asks, "*what are the ways in which audience data practices mitigate and/or reinforce inequity in journalism?*". As the datafication of news audiences and its consequences of equity unfold globally, this study puts a specific focus on organizations reporting to worldwide audiences. It contributes to the current literature by examining these issues concurrently, and by contextualizing journalistic audience data practices within larger processes of datafication. Specifically, it offers a framework for the interrogation of economic, cultural, and political dimensions of audience data practices and its interrelations. While journalism as an institution has always been marked by a level of inequity, the use of audience data necessitates scholarly evaluation to understand how pre-existing asymmetries have been reinforced or even deepened. The study demonstrates how inequity is reinforced through maldistribution, misrecognition, and misrepresentation, and examines how and under which conditions data practices can be used for greater recognition and more adequate representation of audiences. Finally, consequences for the journalism-audience relationship and the news industry at large are discussed.

Literature Review

Audience Data Practices in Journalism

Editorial teams integrate data-driven insights with traditional news judgment, while working with partners, such as platforms or advertisers — for example to sell user data to third parties (Kammer, 2021). Audience data practices thus involve the collection, analysis, interpretation, and representation of audience data, as well as the data's subsequent translation into short and long-term editorial strategies.

Since data practices are inherently embedded in the infrastructures where data is generated, stored, and processed, they are context specific (Kotliar 2020). Accordingly, audience data practices that involve platform generated data might differ from data practices that use data generated by news organizations themselves (e.g. through cookies via websites and apps). This study distinguishes between quantitative and qualitative practices. Quantitative practices involve measuring and analyzing descriptive data through audience metrics – this includes pageviews, dwell time, shares, A/B testing, and randomization – to draw inferences about users' preferences and to maximize audience engagement (Carlson 2018). Qualitative practices consist of observing, engaging with, and imaging audiences. For instance, by analyzing and interpreting user comments (Coddington, Lewis, and Belair-Gagnon 2021). Additionally, this study considers any approach that supports the analysis and interpretation of quantitative audience data, such as talking to sources to further contextualize data as a qualitative data practice. Lastly, using data to negotiate editorial values and organizational needs vis-a-vis audiences' online behavior and inferred preferences are here also viewed as a qualitative approach.

Audience Data Practices and Inequities

Fraser's theory of justice (2013) recognizes three analytically distinct dimensions: accounting for economic injustices *(maldistribution),* cultural injustices *(misrecognition),* and political injustices *(misrepresentation)*. Different aspects of Fraser's work on justice have recently been used to understand societal consequences of data practices – most notably within the context of civil society organizations (Gangadharan 2017) and surveillance capitalism (Cinnamon 2017).

Data as Economic Good — Maldistribution in Journalism

Critical data scholars argue for an understanding of data as capital. Sadowski (2019), for example, has argued that data is a core component of the contemporary political economy, in which every aspect of social life is viewed as a source of data extraction. As governments and firms treat data like capital data extraction by any means possible, with little regard for consent or compensation is being justified by capitalistic logics. Extensive data commodification ultimately enables power asymmetries "weighted towards those who have the access and the capabilities" to seize it (West 2019). These unequal conditions of datafication create "data-classes", dividing those who create data, those who collect it, and those who can seize and analyze it (Andrejevic 2014). Similarly, Burrell and Fourcade (2021) have argued that technologically skilled professionals have emerged as a new occupational class, consolidating power through technological control over data extraction (Burrell and Fourcade 2021). The capital of this class is not only economic in nature, but also cultural and political (ibid.). Hence, data is not only an information source and an economical commodity, but a social and political instrument (Bigo, Isin, and Ruppert 2019). This is because "[d]ata does not happen through unstructured social practices but through structured and structuring fields in and through which various agents and their interests generate forms of expertise, interpretation, concepts, and methods that collectively function as fields of power and knowledge" (Bigo, Isin, and Ruppert 2019, p. 6).

In the journalistic context, power asymmetries between platforms and news organizations have been forged, with platforms not only controlling data structures, but also profiting the most from the data generated through news use (Lindskow 2014; Willig 2021). These processes also have consequences for the relationship between audiences, who generate data through their digital news engagement, and the news organizations that track and commodify them (Nikunen and Hokka 2020). Despite policy efforts to strengthen users' privacy — such as the EU's GDPR or Brazil's LGPD — scholars have argued that audiences have little power to control how their data is used (Adams 2020). Examining the top one million websites, Englehardt and Narayanan (2016) found that the level of tracking varies considerably between different categories, with news sites being among those with the highest number of third-party trackers. Binns et al. (2018) found that news apps alongside apps targeted to children, contain the highest number of third-party trackers. While research analyzing third-party trackers suggests differences between private and publicly owned websites (Sørensen and Kosta 2019), public news organizations are no universal exception (Sørensen and Van den Bulck 2020). Accordingly, scholars have interrogated the use of audience data within larger economic processes shaping the profession and examined

the economic value of audience data (Nelson and Webster 2016). While some argue that the use of audience metrics directs editorial decision making based on market logics (Cohen 2015), most at least agree that market logics driven by audience metrics often conflict with editorial goals, professional norms, or democratic values (Carlson 2018; Moyo, Mare, and Matsilele 2019; Nelson and Tandoc 2019), all of which have consequences for equity. Accordingly, the analysis of audience data has been dubbed "the new currency of the media economy" (Steensen, Ferrer-Conill, and Peters 2020). Maldistribution in this context thus occurs through the unequal distribution of the ability to collect, process, analyze and economically benefit from audience data.

Data as Cultural Good — Misrecognition in Journalism

While data practices are often misperceived as culturally agnostic (Kotliar 2020), scholars have offered ample empirical evidence that data practices cannot be detached from the culture and power structures they interact with (Benjamin 2019; Eubanks 2018; Noble 2018). Kotliar (2020) has argued that there is a need to understand how *"intercultural encounters"* – that is, encounters between the local culture of algorithmic systems and the culture of those users that practitioners are aiming to understand – affect data and its algorithmic interpretation, analysis, and presentation. Understanding data practices thus requires an examination of socio-technical formations and their overarching power structures. Accordingly, how data is analyzed is just as relevant as the epistemologies, worldviews, and narratives of the people who write algorithms, collect and interpret data (ibid).

There have been notable contributions that have enhanced our understanding of how journalistic data practices interact with culture. Christin (2020) asserts that organizational culture shapes how news professionals ascribe meaning to audience data, driving how news professionals imagine their audiences. Belair-Gagnon and Holton (2018) found web analytics companies are important drivers of cultural change within newsrooms. Audience scholars such as Livingstone (2019) have argued that datafication flattens our understanding of audiences and reduces complex cultural connections to simplistic engagement data.

Against the backdrop of this scholarship, I argue that data practices are not only shaped *by* culture, but that data is in return shaping culture, and is thus a cultural good. Classification systems offer a key example of how data is transformed into a cultural good structuring (mis)recognition. Data classification organizes how people are understood and can represent themselves (Crawford 2021). People's identities are viewed through what can be recognized by classification systems, rendering anything that cannot or is not measured online invisible. Algorithmic identities created using classifying systems are influenced by various cultural aspects, such as social context, human expertise, and messy data flows (Kotliar 2020), with the potential for misrecognition. Classification is thus not only knowledge creation tool, but also as an assertion of power (D'Ignazio and Klein 2020). While journalism research has paid little attention to the connection between audience data and inequity, critical data studies show that quantified data and classification systems have long been used to control, dominate, and exclude people, producing misrecognition. Accordingly, in line with Fraser, data can be considered as a cultural good reflecting and reproducing the societal power

structures they arise from and interact with. As a cultural good, data shapes recognition and structures cultural value and status hierarchies.

This matters for journalism and for the collection, contextualization, and communication of data for journalistic pursuits. News professionals have, for example, the potential to mitigate data-driven misrecognition, by means of what Diakopoulos (2021) calls "algorithmic accountability reporting," through which news professionals scrutinize the biases encoded in data and algorithms using a combination of journalistic and technological skills. Being knowledge-producing institutions, news organizations are important venues for understanding data as a cultural good. Especially, since it is in journalistic practice that technological and cultural representations commingle, as journalists use datafied representations of audiences to make editorial decisions and decide whose interests are represented. In the context of audience data practices, misrecognition occurs when the use of classification systems reinforce a particular positivistic idea of audiences, reducing them to their data traces while reproducing hierarchies and unequal status.

Data as Political Good — Misrepresentation in Journalism

Journalists' access to user data is often thought to positively contribute to journalistic representation. However, audience scholars have criticized the notion of "listening" to audiences' needs via their data traces. This is because not all people "attract news" equally, as algorithms reinforce unequal structures of news exposure (Thorson 2020). While journalists have more resources to understand the needs of audiences who are already algorithmically classified as news-interested (Thorson 2020), their ability to understand the needs of less exposed users is inevitably limited. In other words, news professionals' understanding of audiences is dependent on visibility architectures controlled by platforms. This is particularly relevant, as "visibility itself is not necessarily the key logic in the contemporary moment but rather how visibility is *managed and controlled*" (Banet-Weiser 2018, p. 24). Feminist media scholar Sarah Banet-Weiser (2018) has thus argued for a more nuanced approach to "logics and moralities of both economics and culture as a way to understand how identities are constructed within the economy of visibility, and to ask what is at stake in this kind of construction" (Banet-Weiser 2018, p. 25). Drawing on this, the concept of "economies of visibility" can be expanded to a back-end form of data-visibility, in which economic imperatives, as well as technological and cultural representations are deeply intertwined and politically consequential. As such, cultural and political injustices are closely interconnected and, within datafication, socio-technological forms of visibility can be considered a *key representational precondition* in journalism. What data hides or makes visible ultimately shapes whose informational needs can be seen and considered. Representation then concerns whose needs are considered worthy of attention. Usher's work (2021) serves as a prime example, showing how economic logics interact with historical patterns of inequity, resulting in misrepresentation. Usher argues that since US American news organizations economically rely on a rich, white audience, they consequently prioritize this same audience in their editorial decisions. In other words, journalism is characterized by misrepresentation, driven by economic pressures, and weighted against historically marginalized audiences. While injustice in journalistic representation is an age-

old problem, Usher argues that as news organizations try to build a digital subscription base composed of global elites, the distortions of the digital advertising market and the logics of audience datafication have further aggravated this issue. Misrepresentation in the journalistic context thus occurs, when news professionals make audience-data-informed editorial decisions that serve and prioritize a certain segment of their audience-decisions, often based on hierarchies that are established through misrecognition.

Method

Sampling

This study builds on expert interviews ($n = 31$) with news professionals working for large national and global media organizations (see Table 1), reporting in English and/or German (including *BBC World, Deutsche Welle, the Guardian, Al-Jazeera English,* and *the New York Times)*. Global media organizations are here understood as organizations that market themselves as global and address audiences across the world. For a more thorough discussion on the conceptualization of global media that this study builds on, please refer to Robertson (2021). While previous research has given much attention to national media organizations, there is a need to complement insights from global organizations. It is global organizations that face the unique challenge of meeting audiences' informational needs across countries, media systems, and platforms, while establishing a shared understanding of and shared responses to datafication associated inequities. Moreover, datafication and inequity need to be examined at the scale at which they unfold — national and global. Drawing on critical case sampling (Suri 2011), the sample consists of organizations across a spectrum of financing solutions and relations to political power (Robertson 2021, p. 305). Given their economic, cultural and political resources, global news organizations are particularly well

Table 1. Distribution of News Organizations n (organizations) < n (interviewees), as several interviewees were working for the same organizations.

News Organization	*Type*	*Market Orientation*
Al-Jazeera English	Legacy	Privately funded
AJ+	Digital Only	Privately funded
BBC	Legacy	Publicly funded
BBCW	Legacy	Publicly funded
BuzzFeed	Digital-only	For profit
Channel 4	Legacy	Publicly funded
Deutsche Welle (DW)	Legacy	Publicly funded
FAZ	Legacy	For profit
The Guardian	Legacy	Subscription/donation based
The Independent	Digital-only	For profit
The New York Times	Legacy	For profit
NZZ	Legacy	For profit
The Telegraph	Legacy	For profit
The Times	Legacy	For profit
Vice	Digital-only	For profit
WDR	Legacy	Publicly funded
ZDF	Legacy	Publicly funded

The study distinguishes between the BBC and the BBC World Service, and between Al-Jazeera English and AJ + to provide analytical clarity. The BBCW is focused on providing news to international audiences, while the BBC serves the UK public. AJ + is a digital news platform, while AJE is a TV broadcaster.

positioned to leverage data for the mitigation of inequity. If news professionals that have these resources at their disposal are not developing data practices to do so, it is unlikely that news professionals with fewer resources can. Moreover, economic aspects might structure what news organization prioritize in their data strategies. Privately funded news organizations that are economically well positioned might be less depended on growing a digital subscription base, than news organizations that are solely relying on donations and subscriptions. Similarly, state-owned organizations, might focus on other aspects of data practices, than pro-profit news outlets. Accordingly, the sample ranges from legacy to digital-only news organizations and includes digital newspapers as well as broadcast media (see Table 1), to diversify and include possible outliers in data practices. Participants were recruited via email or contacted via LinkedIn and Twitter.

Two sets of professionals are in focus: audience editors primarily focused on navigating audience data and journalists who have a track record of reporting on social inequities (see Appendix A). While audience data touches every journalistic professional group, it is audience editors who are making digital behavior and informational needs of audiences present in the newsroom (Ferrer-Conill and Tandoc 2018). As emerging technology drives change in journalism and professional roles are in flux, the concept of what Ferrer-Conill and Tandoc (2018) call an "audience-oriented position" is still somewhat unwieldy and placed upon a variety of different professional titles, such as "Head of Audience", "Deputy Social Media Editor", or "Director of Audience Development". Nevertheless, their shared core function of analyzing, interpreting, and acting on audience data unifies this professional group (ibid.). Meanwhile, journalists whose work focuses on social inequities appear to be specifically attuned to issues of maldistribution, misrecognition, and misrepresentation. These professionals can therefore assess to what extent data practices add to or hinder their capacity to address inequity. This unique set of perspectives allows for a more nuanced understanding of how data practices play a role in the reproduction or mitigation of inequities.

Interviews

Data was collected between 2020 and 2021, as a component of a larger research project on information inequality. Interview length varied between 1 h 50 m and 35 m, with an average length of 59 m, and were conducted online or via phone. The majority of interviews have been conducted by the author ($n = 25$); the remaining interviews have been conducted by other members of the research team. All names of the interviewees have been anonymized and the names of their organizations have been omitted. Data was collected and processed according to GDPR guidelines and participants approved the sharing of data between all members of the research team.

The interviews focused on news professionals' relationship with their audiences and the role data practices played in structuring them, as well as professionals' perceptions of and experiences with inequity in journalism. The following topics were discussed: daily work routines that include audience data and how professionals evaluated them, ways in which audience data helped or hindered professionals' aim to understand

audiences' informational needs, how data was discussed and negotiated within the newsroom, as well as efforts to reach and serve traditionally marginalized audiences. Most participants brought up newsroom inequities, how inequities in journalism shaped news professionals' work was thus a recurring topic.

Analysis

The subsequent analysis was structured using the thematic analysis method following Braun and Clarke (2006). Employing an inductive analysis within a constructivist paradigm, I identified patterns and interrelationships in the data by means of thematic codes. As a first step, I employed Fraser's dimensions of justice as sensitizing concepts. Sensitizing concepts are understood as "those background ideas that inform the overall research problem", offering ways of seeing, organizing, and understanding experience (Charmaz 2003). The sensitizing concepts of maldistribution, misrecognition, and misrepresentation formed the conceptual framework of my study, and enabled me to search for the economic, cultural, and political dimensions of data practices.

Through "repeated reading," I then traced the connections between data practices and the reproduction or mitigation of inequities in journalism, and sorted the data into meaningful groups (Tuckett 2005). Connections between audience data practices and maldistribution were easily identified. In contrast, detangling misrecognition and misrepresentation in data practices was more demanding. This was the case because misrepresentation builds on misrecognition, therefore both dimensions are closely interconnected. Assessing how all three dimensions inform each other revealed the analytical distinctions between them.

Findings

Audience Data Practices

Data practices among the interviewees did not only differ in quantity (who is performing which practice how often) but also in perceived relevance. Audience editors of this sample, whose main role is to navigate and act on audience data, naturally were more involved in data practices than editorial staff and valued data practices more than their editorial counterparts. Journalists recognized the value audience data has for news organizations. However, while some journalists confessed, they were impacted by data practices — "I was obsessively looking at [audience metrics]" (NP1), most journalists argued that they were not particularly interested in the information gained via quantitative data — "I get sent it, I ask for it. But I don't really spend much time looking at it." (NP2) — or simply did not consider it their responsibility. "I just let the website team and the digital team get on with it" (NP3). Similarly, audience editors argued that it would be inefficient for journalists to analyze and interpret audience data. "It wouldn't make sense if journalists had to go through the piles of data and figure out what does this mean, how do I do this? What does this mean for my story?" (NP4). However, some news professionals embodied a hybrid role. These professionals understood themselves as journalists, while primarily working on audience data-based projects to improve editorial content and audience engagement. In other words,

distinctions between data practices related roles were fuzzy at times. The variance between professional groups can be grasped in three contexts in which data practices were performed — first, the audience-editorial context, second the editorial context and lastly, a hybrid-context, in which data practices are used to bridge both contexts.

In the audience-editorial context, audience editors described translating knowledge inferred from audience data as a mixed-method process — "We look at all the social media performances from all teams, it's a quantitative and a qualitative assessment" (NP5). Audience editors would mainly use quantified data from audience metrics and combine data with more qualitative and manual approaches. One audience editor, for example, described how their team monitors audiences on social media platforms to identify creators and channels and their informational needs: "we had to develop some manual procedures to find out what specific audiences want. You have to go out there in the flow on Twitter, Instagram, and you have to see what people are talking about, what people are searching for" (NP4).

In the editorial context, journalists described using audience analytics to understand audiences' engagement with their work and thus building mostly on quantitative data practices. When audience editors and journalists reflected on daily data practices, they usually did so without explicitly distinguishing between data generated via platforms or via their own websites and apps. However, since the vast majority of data was generated through platforms, platform related data practices had a specific relevance to news professionals. Accordingly, I only distinguish whether data practices were connected to platform infrastructures, or infrastructures of the news organization, if articulated so by the interviewees.

In the hybrid-context, audience editors described their efforts to translate data into actions for journalists or to use audience data to develop strategies in conversation with the editorial staff: "We also use our data tools to check, if what journalists perceive, or managers or editors perceive, they're doing well. If it also is what we're actually seeing in the data." (NP5) Lastly, audience editors would qualitatively work together with journalists to make data practices more accessible to them: "The first thing that we do is that we share the findings to all the journalists. (…) And then after that, we do workshops, offering the data. We help them to go through the same process, not just showing the findings, but the workshop is for them to arrive at the same conclusions. We find that if you want to help them to improve what they do, they need to take ownership" (NP5). Across these three contexts, audience data practices reveal economic, cultural, and political dimensions, which are addressed in what follows.

Economic Dimensions of Audience Data Practices

Economic aspects of data practices became salient when working with audience data from platforms. News professionals were aware that relying on platforms to reach their audiences is an economic necessity, that, at the same time, comes with substantial economic risks: "they change their algorithms and systems so much that you can't sustain that audience" (NP6). These forms of economic uncertainty highlighted the power asymmetries between platforms and news organizations for audience editors, who struggled with journalism's dependence on platforms: "We need platforms more than

they [need news organizations], because we really want our stuff to be out there being read" (NP7). Economic reliance was problematized, as audience editors reflected on data practices within the audience-editorial context. They argued that platforms "are not reliable partners." (NP4), that "can be helpful, when it's in their interest to be helpful", but whose interests and values are not necessarily aligned with their own. This meant, for example, that what news professionals valued as newsworthy, was not necessarily what platforms decided to elevate through either manual or algorithmic selection. Instead, platforms were described as "just be[ing] in it for money" (NP4), while audience editors tried to negotiate journalistic values amidst economic pressures. This conflict was embedded in data practices news professionals rely on to reach and understand their audiences across all three data contexts. Audience editors described how using audience data to develop strategies together with the editorial staff is hindered by the opaque gatekeeping platforms enforce: "What does the algorithm actually do? Journalists want certainty about what happens. But you […] have to accept that there is a multiverse of things that could happen […] you can't dictate your outcome." (NP8)

As platforms structure how news professionals are connecting with audiences, one interviewee voiced concerns over the reinforcement of power asymmetries in which platforms control audience reach: "my biggest concern is that there's a level of digital imperialism that goes on in news reporting" (NP8). "You know, how [the] Nigerian language [program] focused entirely on Facebook, because, guess what? Facebook Essentials is bundled with all the data packages and it costs them money if they go outside of Facebook. So, of course, they wouldn't browse anything outside of Facebook. So, their entire strategy of connecting with audiences was dictated by the fact that they needed to put their radio program on Facebook, their TV program on Facebook: all their journalism contacts had to go through that platform, and the limitations of that platform." (NP8) Inequities between platforms and news organizations were thus thought to produce inequities on a global scale.

Notably, audience datafication was seen as a fundamental way to respond to the uncertainties and related economic pressures in the audience-editorial context. Audience editors tried to counterbalance their lack of control over news selection and audience reach on platforms by investing in audience data analysis. The idea being, that the better audiences' preferences and behavior was understood through data the better news organizations could respond to the continuously shifting algorithmic conditions of platforms. This was despite the fact that these professionals cautioned that data from platforms may be flawed or unreliable: "[Platforms] provide metrics, but it's not always reliable" (NP6).

Attempting to decrease reliance on platforms, audience editors aimed to use audience analysis to navigate audiences away from platforms — for example by leveraging data to turn audiences into subscribers (more on that later). Audience data was hence considered an essential tool to build sustainable relationships with audiences that would result in economic growth. However, news professionals discussed the commodification of audiences' engagement with news and industry's own role in these processes with ambivalence: "Media companies need to find a way to finance themselves. (…) I can be very ideological and say, no, news organizations shouldn't sell

data, right? And that would be [a] very nice ideological stance for me. (…) There are lucky ones like [the] BBC, [or] like us, who don't have commercial pressures. But many do and… should they be selling data? I mean, it depends what the data is." (NP4) The commodification of audience data was considered, at least to a certain degree, as a sensible economic strategy for private news organizations. While news professionals argued that "personally identifiable data" should not be commodified, they persisted that news organizations should be able to sell "aggregated, anonymized data" of audiences under economic pressures. Such deliberations underscored the tensions between economic pressures and professional values. Treating audience data exclusively as a resource to improve journalism for audiences, was not affordable to all. Resource poor organizations were perceived to be forced to weigh audiences' privacy against economic survival.

Regarding economic aspects, news professionals recognized audience data as an economic good and acknowledged the industry's economic reliance on the commodification of audience data. News professionals linked this reliance first to platform dependency and associated power asymmetries between news organizations and platforms and second, to financial pressures of news organizations. News professionals struggled to identify successful strategies to mitigate journalism's reliance on audience commodification and its overall dependency on platforms. However, news professionals showcased how these pressures were internalized in their practices — for example, by using audience data to identify and focus on coverage that turns audiences into subscribers (as discussed further in below), or simply by selling audience data to third parties.

Cultural Dimensions of Audience Data Practices

Cultural dimension of audience data practices became evident in three forms: first, via the organizational culture of news outlets, second, in the reproduction of the positivist notion that audiences and their informational needs could be understood through quantitative approaches (perhaps even better than through other forms of knowledge production), and third, in how cultural resources (such as the cultural knowledge of individual professionals) was leveraged to interpret audience data within cultural frameworks.

In regard to organizational culture, news professionals described a culture in which editorial decision making is grounded in mostly quantitative data practices — "there's nothing more powerful than showing numbers" (NP5). Quantitative approaches, using audience metrics for sensemaking, were often valued over journalistic intuition and qualitative methods, such as engaging in direct dialogue with audiences. This credence was traced back to transformations within organizations: "we have organizationally built capacity and not just capacity, but acceptance [of data-informed editorial decision-making]" (NP4).

However, as journalists' capacity to understand their audiences is shaped by platforms' data infrastructures, news professionals expressed that they are struggling with the limitations of platforms' classification systems in the audience-editorial context:

"It's hard, because data doesn't always give you a breakdown of who is actually consuming your work (…) Particularly looking at Bangladesh, and Pakistan, I really

want to increase the number of women journalists engaged in those two countries. Twitter makes it almost impossible to figure out this gender breakdown of who is following (…)". (NP6) Similarly, a journalist explained how their team made sense of platform-provided constructions of algorithmic identities working with audience-editors (hybrid context): "These [gender constructs] are not necessarily correct. If Facebook tells me you are 70% female and 30% male, or YouTube tells me that, it doesn't mean that [you are] 70/30, but that the platform interprets your click behavior […] It is always said that their AI or their algorithm is extremely good and can actually read that, but we don't know 100% whether that is so. […] We can't read sexual identities, we can't read your race, so to speak." (NP10). In other words, news professionals conceptualized audiences based on algorithmically driven classifications, that construct identities in ways that elevate some (perceived) dimension of audiences' identities, while erasing other dimensions altogether. Like with all forms of quantification, processes of classifying and imagining alter societal visibility architectures, illuminating some social realities, while obscuring others — a process that has consequences for how audiences are understood and represented.

Across data practice contexts, news professionals expressed awareness of the disconnect between audiences' lived realities and their data informed perception of audiences and their information needs: "there are "*huge* blind spots that are completely unanswered by the metrics that we gather […] it's like we know nothing and yet we are learning" (NP6). Hence, there is a tension between news professionals' reliance on quantitative audience data for editorial decision making, and their acknowledgement of the limits of understanding audiences based on their quantified online behavior. In response to this, news professionals in the audience-editorial context argue that there is "no one magic metric" that news organizations can "commission or publish by" (NP11). Instead, data practices are a form of cultural meaning-making, in which news professionals are "putting a narrative around" the data and are asking themselves "what is the story here?" (NP11). Audience editors further argue that their own positionality shapes interpretations of audience data: "Just because you come from a different background or have a different gender or ideas and you'll see other things that your male colleagues or other colleagues won't see." (NP5). Cultural knowledge was considered essential in professionals' ability to contextualize data within appropriate cultural frameworks for the interpretation of audience data.

In conclusion, news professionals across data practice contexts use audience data and classification systems to draw inferences about audiences' behavior and preferences. Data is thus used as a cultural good that shapes news professionals' conceptualizations of audiences. Even though limitations of data regarding reliability and validity are discussed by news professionals across data contexts, data practices are considered to have transformed organizational cultures. This has consequences for the political dimensions of audience data practices, which I outline as follows.

Political Dimensions of Audience Data Practices

Political dimensions of audience data practices became salient when news professionals used audience data in an aim to understand informational needs of different audiences and thus aimed to improve their representation in journalism, and second,

when data was used to make editorial decisions to prioritize certain audience segments over others.

Audience editors reported using data in the audience-editorial context to understand differences in news consumption across socio-demographics and explained that moving beyond a popularity focus is needed to recognize audiences' diverse needs: "there are some articles that look like they had a very, very low average dwell time. Because our core readership didn't engage with it. But when we look at [...] how engaged female readers were, sometimes it's very, very high within that cohort." (NP11) Audience editors argued that data analysis enabled them to put traditionally underserved audiences into focus. They pointed to different projects set up by their organization to improve coverage and to increase engagement of specific audience groups by the means of audience development. One audience editor, for example, explained how their team changed the coverage of a news bulletin after realizing that their previous coverage did not attract their target audience — audiences based in Somalia — but the Somali diaspora in Minnesota, USA:

"[...] we found that when we presented individual stories as single videos, the proportion of [the] audience from Somalia that was young and female increased. And our proportion of older Somali women from outside Somalia decreased. After a few months we came to this conclusion: OK, so we have two different audiences. We have a Diaspora audience who relies on what we provide in the bulletin as a news-from-home service. So, some of them might just like binge-watch an entire week of bulletins. Twenty-five-minute bulletins back-to-back, and they watch them. [...] And we did a bit of sort of talking to the team. [...] basically, they have a call with the family at the weekend. On Friday they watch all the news bulletins so that they know what to talk about when they phone up." (NP8) Audience data could only give the team insights into engagement, but in order to make sense of the data, the team needed to "investigate the cultural framework": "[We] ended up accepting that different audiences wanted different things from us. If you live in America and Minnesota, you've got broadband, you can watch half an hour of video. But if you live in Mogadishu, you probably don't really have half an hour in your day to sit down and watch a news bulletin. You might listen to the radio. You might just watch one story that interests you. [...] audience engagement taught us to not see the audience as homogenous in their need." (NP8)

Audience data analysis that aimed to enhance reach and engagement of traditionally marginalized audiences was driven by individuals who were attuned to inequities in news coverage *and* the newsroom. More often than not, these professionals were women and/or members of traditionally underserved communities. However, they often struggled across all three data practice contexts with structures in journalism that have historically privileged audiences that are economically well positioned, male — and in the Western context — white and dominantly Anglo-Saxon. A news professional reflected that despite their news organization's global reach and an awareness of global issues that call for adequate representation of those affected, the outlet stopped attending to informational needs of audiences outside the US: "I was hired as part of that expansion of global news coverage. And since then, because of a variety of reasons ranging from the financial health of the company, the editorial priorities,

change in management, […] the reality is we have moved away from considering audiences outside of the US." (NP12)

How economic factors are shaping political injustices, by structuring what sectors of audiences are considered worthy of attention, becomes especially evident when news organizations prioritize subscription rates. An audience editor explains, how data practices are used to tailor news content in ways that turn audiences into subscribers:

"The BBC and non-commercial service broadcasters, because they are built on universality, look for the biggest number. But commercial organizations who have a defined audience look for the most financially important number. If you watch one video a week and you never subscribe to The New York Times, your data points don't particularly matter to them, but they want to know what it is they need to do to convert you. They look for the commonalities. What is it about this that isn't working? Hence why they start new sections that last three months and then disappear. They've seen a pool of users who aren't converting and they're going: 'If we do this, can we convert them? And then: 'Well, that didn't work. Kill it.' Sort of thing. (…) It's quite sort of: you must act on the data. And actually, sometimes you mustn't act on the data because data is trying to persuade you to do something that's not to your benefit." (NP8)

Even when economic structures allow organizations to spend resources to serve traditionally marginalized audiences, news professionals, especially journalists situated within the editorial context, argued that audience data only give insights into engagement rates, but is no indicator for adequate representation. News items that cover issues relating to marginalized groups (such as refugees, or the LGBTQ community) usually draw engagement due to platform logics that amplify polarization. "But whether or not this coverage is fair, accurate, truthful, and can really offer a full picture is another matter." (NP6) Consequently, data practices were not perceived as a panacea. Instead, news professionals pointed to strategies to address inequities in the newsroom: "The best, most efficient, most important thing we can do now is finally have a quota for People of Color in journalism." (NP10)

That is, news professionals acknowledged audience data as a political good by recognizing how data practices were informing editorial decisions to serve and prioritize certain segments of their audience over others, and to tailor news coverage accordingly.

Discussion

This study interrogated audience data practices of news professionals in order to understand the ways in which audience data practices mitigate or reinforce inequity in journalism. The analysis has demonstrated that audience data can be understood as an economic, cultural, and political good, that is used in ways that have a bearing on social inequity. In this section, I address and discuss each dimension of data practices and how they relate to the reproduction or mitigation of inequities in journalism (for an overview see Table 2).

In line with concerns of scholars like Willig (2021), economic pressures and mechanisms of platformization that reinforce power asymmetries between platforms, publishers, and audiences, are experienced across news organizations (Table 2, economic

Table 2. Overview of data practices across contexts and dimensions.

Context	Audience-editorial Context	Editorial Context	Hybrid Context	Outcome
Approach	mixed-method	mainly quantitative	mixed method	—
Economic Dimension	Collecting, processing and analyzing audience data for profit; negotiations with platforms as business partner,	Understanding audiences' engagement with journalistic work through quantitative audience data generated by platforms and in-house	Translating data into editorial strategies, while navigating opaque algorithmic platform architectures	Unequal distribution of the ability to collect, process, analyze and economically benefit from data as a resource
Cultural Dimension	Perceiving audiences through classification systems and platform logics			Reinforcement of a positivist conception of audiences, potentially reproducing hierarchies, and unequal status
Political Dimension	Develop strategies to drive engagement rates and build subscription rates, understanding audiences' heterogeneous information needs	Negotiation of journalistic ideals (e.g. serving underrepresented audiences), vis-à-vis economic pressures and platform logics		Audience data informed editorial decisions to rather serve and prioritize certain segments of their audience over other, potential to tailor news to underserved audiences and their needs

dimension). News professionals problematized data inequities between platforms and news organizations, but also legitimized power asymmetries between news organizations and audiences. News professionals experienced little agency over how organizations can use audience data to control audience reach and argued that their ability to successfully use this data to reach, build, and sustain audiences was complicated or even hindered by the opaque and ever-changing algorithmic architecture of platforms. Although news organizations gained most of their reach through platforms, they found them to be unreliable economic partners and data sources (Table 2, economic dimension). At the same time, interviewees legitimized *maldistribution* between news organizations and their audiences. Audiences' loss of agency, as platforms and news organizations collect, analyze, and economically benefit from their data, was justified on an economic level. Especially financially underprivileged news organizations were argued to be reliant on selling audience data to third parties. This unequal allocation of resources and power was nevertheless acknowledged to stand in conflict with traditional journalistic values. Accordingly, maldistribution was discussed on a procedural, normative, and structural level (Table 2, cultural dimension). Despite this, news professionals voiced a responsibility to use audience data, not only in the service of profit, but to address issues of misrepresentation and to improve access to accurate and independent information for audiences across various socio-political, economic, and technological conditions to mitigate inequity. This suggests, that while audience data can be considered a "currency of the media economy" (Steensen, Ferrer-Conill, and Peters 2020) and that economic interest can stand in conflict with professional norms and values (Carlson 2018; Moyo, Mare, and Matsilele 2019; Nelson and Tandoc 2019), economic logics embedded in data practices not necessarily override news professional's commitment to information equity. Instead, findings indicate that audience data

practices have the potential to make the interests of traditionally marginalized audiences visible to mitigate misrepresentation. However, whether or not this potential is seized upon is a matter of redistribution (economic), and recognition (cultural), as discussed in the following.

As argued by (Christin 2020), news professionals engage with "algorithmic publics" when imagining their audiences through data. On one hand, their perception of audiences is flattened through the positivistic lens audience metrics reinforces. Because audience metrics are essentially a marketing tool, quantified audience data practices encourage news professionals to see audiences as commodifiable through their data traces across contexts (Table 2, cultural dimension). On the other hand, audience data are used to identify heterogeneous information needs of audiences. As such, audience data and platform structures mediate the relationship between news professionals and audiences. What values are assigned to datafied representations of audiences and how professionals describe meaning and act on them is thus particularly important. The implications of binary classification systems that assign certain dimensions of audiences' identities — such as gender — based on their measurable online behavior, while disregarding others, are worth considering. Forms of assigning identity, of rendering visible and invisible are not neutral, but structure and reproduce hierarchies with consequences for journalistic production, that have the potential for *misrecognition*. News professionals acknowledge that the identities that platforms ascribe to audiences do not necessarily match audiences' self-identification and moreover, much of audiences' identities remains unknown, but largely operate with the classification systems platforms provide.

Audience editors argue that their own positionality shapes interpretations of audience data, making news professionals more or less aware of issues of misrecognition and sequential misrepresentation, depending on cultural knowledge and experience gained through their social position. This underscores the importance of socially diverse news staff. It also illustrates the potential of audience data analysis — when data practices are employed through a cultural framework, data-based meaning-making may enable recognition and as a result, representation. Misrecognition may be embedded in data itself (or lack thereof), or in its interpretation. At the same time, audience data can be leveraged by news professionals to counter misrecognition, if news professionals are able to seize upon relevant cultural capital.

Usher (2021) argued that contemporary journalism is biased towards the interests of rich, white and liberal audiences, while historically marginalized communities remain underserved and disengaged. Findings confirm that logics of datafication and commodification are fostering mechanisms that aggravate these issues. Interviewees reflected that for-profit news organizations especially rely on data-informed editorial strategies to build a digital subscription base of readers that are economically well-positioned. Audience data practices may be used to make editorial decisions to rather serve and prioritize certain segments of audiences over others and hence, risk *misrepresentation* (Table 2, political dimension). Narrowing what sectors of audiences and associated informational needs are considered worthy of attention is not only problematic in the sense that reproduces inequities but might also further erode trust in journalism.

Despite these concerns, this study also found ways in which news professionals leveraged data practices to mitigate misrepresentation. More critical and attentive

analysis that relied on both quantitative and qualitative approaches was successful in identifying traditionally underserved groups and their information needs. These efforts were successful in the sense that they drove engagement rates among said groups. As mentioned above, data practices that aimed to mitigate misrepresentation was driven by individuals who were generally committed to addressing issues of misrepresentation and were attuned to inequities in news coverage *and* the newsroom. These individuals were overwhelmingly women and/or members of traditionally underserved communities. They argued that the mitigation of inequity in the newsroom increases the capacity of newsrooms to mitigate misrepresentation in journalism. Data practices might have the potential to make the interests of traditionally marginalized audiences more visible and to mitigate misrepresentation. However, as misrepresentation is interlinked with maldistribution and misrecognition, news professionals' ability to alleviate misrepresentation through data practices depends on the cultural and economic resources available to them.

Regarding limitations, the conducted interviews yielded further findings related to the reproduction of inequity, especially pertaining to newsroom equality, yet a further examination of these findings was beyond the scope of this study. Notably, the design of this study only allowed for an analysis of perceptions and experiences as reflected by news professionals. An important next step would be to use newsroom observations to examine the connections between audience data practices and the reproduction and/or mitigation of inequities — research which this study hopefully encourages. It should be highlighted that given the sample, this study cannot draw conclusions that generalize an entire professional community (e.g. audience editors), nor an entire news organization. Moreover, comparative analysis is needed to reveal distinctions in audience data practices, for example of news professionals working for global and national outlets, or commercial and non-commercial outlets. Lastly, the scope of this article only allowed to allude to the interrelations of maldistribution, misrecognition and misrepresentation. Accordingly, there is a need for further research to focus on the interrelations between economic, cultural, and political dimensions of audience data practices, to examine how these dimensions shape and reinforce each other and, as a consequence, inequities in journalism.

Conclusion

By applying Fraser's (2009) theory of justice, this study analyzed news professionals' experiences with and perceptions of audience data practices in journalism, illustrating how data can be understood as a cultural, economic, and political good within journalism, that structures maldistribution, misrecognition, and misrepresentation. It thus contributes to the field, by demonstrating the need to interrogate economic, cultural, and political consequences of audience data practices *in conjunction*. Inequities reproduced in connection with audience data practices are not only a matter of the commodification of audiences (distribution), *or* biases built into algorithmic systems (recognition), but both, and moreover a matter of how news professionals can understand audiences' diverse information needs and how they respond to these needs (representation).

The paper thus furthers our understanding of the relationship between audience data practices and inequity in journalism, and, by building on Fraser (2009), offers a framework to examine these relations. It explores how news organizations can use data to improve recognition and representation of audiences and shows under what conditions inequity is reinforced through misdistribution, misrecognition, and misrepresentation.

In doing so, this article underscores the need for critical engagement with values ingrained in audience data practices. In its absence, the tools, and systems that news organizations are using to serve audiences, become the systems that set values for newsrooms. That is, news professionals are caught between being negatively impacted by and benefitting from mechanisms of datafication and commodification that amplify maldistribution. Both have implications for news professionals' capacities to adequately act within their professional and societal responsibility towards audiences.

This study documents that well-resourced news organizations are capable of mitigating inequities through audience data informed practices, if they channel resources into data practices that are culturally informed and center the mitigation of inequity. In simple terms this means investing in staff that are committed to addressing unequal representation and recognition of audiences. It also means valuing the skills, knowledge, and journalistic intuition of news professionals of traditionally marginalized communities.

Acknowledgements

I would like to thank Alexa Robertson for her generous support as a PI. Further, I would like to acknowledge that Yanthe Zebregs and María de le Huerga helped to procure, conduct, and transcribe the interviews — thank you to both. I am also very grateful to Juliane A. Lischka, Jacob L. Nelson, David Cheruiyot, Ida Willig, Jannie Møller Hartley, and Karen Khodzhayev who provided valuable feedback on earlier versions of the manuscript. Finally, I am thankful to the anonymous reviewers for their insightful comments and suggestions for improving this article.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Vetenskapsrådet (grant #1054707).

References

Adams, P. C. 2020. "Agreeing to Surveillance: Digital News Privacy Policies." *Journalism & Mass Communication Quarterly* 97 (4): 868–889.

Andrejevic, M. 2014. "The Big Data Divide." *International Journal of Communication* 8: 1673–1689.

Banet-Weiser, S. 2018. *Empowered*. New York, USA: Duke University Press.

Belair-Gagnon, V., and A. E. Holton. 2018. "Boundary Work, Interloper Media, and Analytics in Newsrooms." *Digital Journalism* 6 (4): 492–508.

Benjamin, R. 2019. *Race after Technology*. Cambridge, UK: John Wiley & Sons.

Bigo, D., E. Isin, and E. Ruppert. 2019. *Data Politics*. New York, USA: Routledge Studies in International Political Sociology.
Binns, R., U. Lyngs, M. Van Kleek, J. Zhao, T. Libert, and N. Shadbolt. 2018. "Third Party Tracking in the Mobile Ecosystem." In WebSci '18: 10th ACM Conference on Web Science, May 27–30, 2018, Amsterdam, Netherlands. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3201064.3201089
Braun, V., and V. Clarke. 2006. "Using thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77–101.
Burrell, J., and M. Fourcade. 2021. "The Society of Algorithms." *Annual Review of Sociology* 47 (1): 213–237.
Carlson, M. 2018. "Confronting Measurable Journalism." *Digital Journalism* 6 (4): 406–417.
Catherine, D. 2021. "15. Data Journalism: What's Feminism Got to Do with IT?" In *Data Journalism Handbook*. Amsterdam, Netherlands: Amsterdam University Press.
Charmaz, K. 2003. "Grounded theory: Objectivist and Constructivist Methods." In *Strategies for Qualitative Inquiry*, edited by N. K. Denzin and Y. S. Lincoln, 2nd ed., 249–291. Thousand Oaks, CA: Sage.
Christin, A. 2020. *Metrics at Work*. Princeton, USA: Princeton University Press.
Cinnamon, J. 2017. "Social Injustice in Surveillance Capitalism." *SS* 15 (5): 609–625.
Coddington, M., S. C. Lewis, and V. Belair-Gagnon. 2021. "The Imagined Audience for News: Where Does a Journalist's Perception of the Audience Come from?" *Journalism Studies* 22 (8): 1028–1046.
Cohen, N. S. 2015. "From Pink Slips to Pink Slime: Transforming Media Labor in a Digital Age." *The Communication Review* 18 (2): 98–122.
Costanza-Chock, S. 2020. *Design Justice*. Cambridege, USA: MIT Press.
Costera Meijer, I. 2003. "What is Quality Television News? A Plea for Extending the Professional Repertoire of Newsmakers." *Journalism Studies* 4 (1): 15–29.
Couldry, N. 2010. *Why Voice Matters*. London, UK: SAGE.
Couldry, N., and U. A. Meijas. 2019. *The Costs of Connection*. Stanford, USA: Stanford University Press.
Crawford, K. 2021. *Atlas of AI*. New Haven, USA: Yale University Press.
D'Ignazio, C., and L. Klein. 2020. *Data Feminism*. Cambridge, USA: MIT Press.
Diakopoulos, N. 2021. *The Algorithms Beat: Angles and Methods for Investigation*. Amsterdam, Netherlands: Amsterdam University Press.
Englehardt, S., and A. Narayanan. 2016. "Online Tracking: A 1-Million-Site Measurement and Analysis." In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16). Association for Computing Machinery, New York, NY, USA, 1388–1401. https://doi.org/10.1145/2976749.2978313
Eubanks, V. 2018. *Automating Inequality*. New York, USA: St. Martin's Press.
Ferrer-Conill, R., and E. C. Tandoc. 2018. "The Audience-Oriented Editor." *Digital Journalism* 6 (4): 436–453.
Fraser, N. 2009. *Scales of Justice New Directions in Critical Theory Paperback*. New York, USA: Columbia University Press.
Gangadharan, S. P. 2017. "The Downside of Digital Inclusion: Expectations and Experiences of Privacy and Surveillance among Marginal Internet Users." *New Media & Society* 19 (4): 597–615.
Kormelink, T. G., and I. C. Meijer. 2018. "What clicks Actually Mean: Exploring Digital News User Practices." *Journalism (London, England)* 19 (5): 668–683.
Kotliar, D. M. 2020. "Data orientalism: On the Algorithmic Construction of the non-Western Other." *Theory & Society* 49 (5–6): 919–939.
Lindskow, K. 2014. *Exploring Digital News Publishing Business Models: A Production Network Approach*. Copenhagen, Denmark: Copenhagen Business School. PhD series No. 07.2016.
Livingstone, S. 2019. "Audiences in an Age of Datafication: Critical Questions for Media Research." *Television & New Media* 20 (2): 170–183.

Moyo, D., A. Mare, and T. Matsilele. 2019. "Analytics-Driven Journalism? Editorial Metrics and the Reconfiguration of Online News Production Practices in African Newsrooms." *Digital Journalism* 7 (4): 490–506.

Nelson, J. L., and E. C. Tandoc. Jr. 2019. "Doing "Well" or Doing "Good": What Audience Analytics Reveal about Journalism's Competing Goals." *Journalism Studies* 20 (13): 1960–1976.

Nelson, J. L., and J. G. Webster. 2016. "Audience Currencies in the Age of Big Data." *International Journal on Media Management* 18 (1): 9–24.

Nikunen, K., and J. Hokka. 2020. "Welfare State Values and Public Service Media in the Era of Datafication." *Global Perspectives* 1 (1): 12906. https://doi.org/10.1525/gp.2020.12906

Noble, S. U. 2018. *Algorithms of Oppression*. New York, USA: NYU Press.

Petre, C. 2015. "The Traffic Factories: Metrics at Chartbeat, Gawker Media, and the New York Times." *Tow Center for Digital Journalism, Columbia University School of Journalism*. https://www.cjr.org/tow_center_reports/the_traffic_factories_metrics_at_chartbeat_gawker_media_and_the_new_york_times.php

Robertson, A. 2021. "What's Happened to Global News?" *New Global Studies. De Gruyter*. 15: 303–322.

Sadowski, J. 2019. "When data is Capital: Datafication, Accumulation, and Extraction." *Big Data & Society* 6 (1): 205395171882054.

Sørensen, J. K., and H. Van den Bulck. 2020. "Public service Media Online, Advertising and the Third-Party User Data Business: A Trade versus Trust Dilemma?" *Convergence* 26 (2): 421–447.

Sørensen, J., and S. Kosta. 2019. "Before and After GDPR: The Changes in Third Party Presence at Public and Private European Websites." In The World Wide Web Conference (WWW '19). Association for Computing Machinery, New York, USA, 1590–1600. https://doi.org/10.1145/3308558.3313524

Steensen, S., R. Ferrer-Conill, and C. Peters. 2020. "(against a) Theory of Audience Engagement with News." *Journalism Studies* 21 (12): 1662–1680.

Suri, H. 2011. "Purposeful Sampling in Qualitative Research Synthesis." *Qualitative Research Journal* 11 (2): 63–75.

Thorson, K. 2020. "Attracting the News: Algorithms, Platforms, and Reframing Incidental Exposure." *Journalism* 21 (8): 1067–1082.

Tuckett, A. G. 2005. "Applying thematic Analysis Theory to Practice: A Researcher's Experience." *Contemporary Nurse* 19 (1–2): 75–87.

Usher, N. 2021. *News for the Rich, White, and Blue: How Place and Power Distort American Journalism*. New York, USA: Columbia University Press.

West, S. M. 2019. "Data Capitalism: Redefining the Logics of Surveillance and Privacy." *Business & Society* 58 (1): 20–41.

Willig, I. 2021. "From audiences to Data Points: The Role of Media Agencies in the Platformization of the News Media Industry." *Media, Culture & Society* 44 (1): 56–71. https://doi.org/10.1177/01634437211029861

Zuboff, S. 2019. *Age of Surveillance Capitalism*. London, UK: Public Affairs.

Appendix A. Interview participants

NP1	Journalist	F
NP2	Journalist	M
NP3	Principal Presenter	F
NP4	Head of Audience Development	M
NP5	Head of Social Media	F
NP6	Audience Editor, Journalist	F
NP7	Deputy Social Media Editor	M
NP8	Project Data Manager, Journalist	M
NP9	Journalist	M
NP10	Editor	F
NP11	Head of Audience	F
NP12	Journalist	F
NP13	Senior Producer	F
NP14	Editor, Web-Expert	M
NP15	Editor	M
NP16	Senior Social Media Managing Editor, Journalist	F
NP17	Editor at Large	F
NP18	Journalist	F
NP19	Senior Director of Client Solutions and Analytics	F
NP20	Journalist	F
NP21	Senior Producer	M
NP22	Journalist	F
NP23	Managing Editor	M
NP24	Journalist	F
NP25	Editor	M
NP26	Multimedia Journalist	F
NP27	Creative Diversity Lead (involved in audience data based project), Journalist	F
NP28	Journalist	M
NP29	Journalist	M
NP30	Social Media Manager	F
NP31	Journalist	M

DIGITAL JOURNALISM
https://doi.org/10.1080/21670811.2023.2235601

Routledge
Taylor & Francis Group

Check for updates

Datafication of Journalism: How Data Elites and Epistemic Infrastructures Change News Organizations

Nadja Schaetz, Juliane A. Lischka and Laura Laugwitz

Universität Hamburg, Journalism and Communication Studies, Hamburg, Germany

ABSTRACT

Data has become an increasingly important commodity for news organizations. The capability to extract, store, and analyze data is also central to organizational decision-making. Drawing on the concept of epistemic infrastructures, this study sheds light on organizational datafication in journalism. Analyzing job advertisements of incumbent broadcast, print legacy, and online-only news outlets in the US and UK, this study traces material underpinnings of data architectures, networks of data agents, and epistemic data paradigms in news organizations. We show that a professional data elite of data analysts, engineers, and product developers serves as an epistemic community that consolidates power through epistemic authority and reinforces epistemic paradigms in journalism. We describe epistemic data infrastructures in the journalistic field and highlight how reliance on these infrastructures contributes to ethical challenges that news organizations are ill-equipped to address. We conclude that news organizations are deeply engaged in fostering epistemic infrastructures and that by acknowledging them, we can better analyze socio-technological changes in journalism and its consequences.

KEYWORDS

Computational journalism; data; datafication; epistemology; sociotechnical perspective; technology; technological skills; epistemic infrastructures

Introduction

Data is a key resource for news organizations for innovation and development, managerial and strategic decision-making processes, or as commodities for brokerage (Kammer 2021). While news organizations compete with platform companies over control of user data, they are often in a superior position to collect data as their websites attract heavy traffic in a given country and track user behavior more extensively than other websites do (Sørensen and Kosta 2019; Adams, 2020). As such, journalism as an industry is deeply involved in the datafication and commodification of audiences and their information uses (Willig 2021). Extensive and intransparent tracking, and exploitation of user information for commodification on the data market, risks eroding audiences' trust in journalism (Sørensen et al.

CONTACT Nadja Schaetz nadja.schaetz@uni-hamburg.de
© 2023 Informa UK Limited, trading as Taylor & Francis Group

2020) and poses a wealth of ethical challenges to news organizations (Porlezza and Eberwein 2022). Moreover, as a data-intensive industry, journalism risks the overproduction of data that may lead to organizational ignorance instead of insight (Schwarzkopf 2020). At the same time, data practices can help journalists to understand audiences' information needs (Schaetz 2023), optimize editorial work, and develop personalized recommendations to users (Sørensen et al. 2020). To leverage machine learning (ML) and artificial intelligence (AI) applications, data becomes increasingly relevant for journalism. These practices require staff with data skills with deep expertise in computer coding and statistical analysis. This new occupational class, which we, drawing on Burrell and Fourcade (2021), call the "data elite", holds technical control over the digital means of production. Their role is to use data extraction and analysis to develop "business intelligence" (BI) systems, build tools for news organizations, and support strategic decision-making (Phillips-Wren et al. 2021).

Drawing on the concept of epistemic infrastructures (Tichenor et al. 2022) as an overarching framework, we analyze materialities, interlinkages, and paradigms of data in news organizations. Specifically, we ask, *how do data materialities, interlinkages, and paradigms shape epistemic infrastructures in digital journalism?* Our aim is to illustrate the roles of data and its agents to understand how datafication shapes news organizations and journalism itself. Empirically, we apply an explorative document analysis of job advertisements collected from major news outlets in the UK and US for positions requiring data capabilities (e.g., data analysts, product managers, product designers or editors).

Literature review

The literature review adopts the concept of epistemic infrastructures that Tichenor et al. (2022) developed in the context of global public policy and translates this analytical lens into the journalistic context. Similar to what Tichenor et al. (ibid.) observed, it is argued here that data used for journalistic pursuits—such as audience metrics or other quantified data to monitor journalistic work and success—are not only "tools of governance" (p. 431) but that datafication in journalism is indicative of a data culture, that drives an interplay of material, socio-technical, and organizational structures within which certain epistemic practices are fostered and privileged over others, that structure journalistic work. This study builds on Ekström's (2002, p. 260) definition of journalistic epistemology as the "rules, routines and institutionalized procedures that operate within a social setting and decide the form of the knowledge produced and the knowledge claims expressed (or implied)" and employs it on data processes in the news organization. Epistemic infrastructures are institutionalized through webs formed between human and non-human agents and the knowledge-creation practices that reify them (Tichenor et al. 2022). Tichenor et al. (2022) offer three "orders" of analysis, that are not separate but instead interlink and interact with each other—(1) the materialities of the infrastructure, (2) the interlinkages of such materialities that hold the infrastructure together, and lastly, (3) the emergence of a paradigm—that are laid out and connected to the journalistic context in what follows.

Material Underpinnings

Material objects for data collection and analysis are the "building blocks" for epistemic infrastructures. These material underpinnings comprise data used for journalistic pursuits, the software and tools used to collect and analyze them, or reports and visualizations (such as reports to distribute within the organization to assess targets and goals for journalistic work). As such, these elements of the infrastructure are both "material and semiotic". They might emerge as physical manifestations, "but at the same time, they aim to reflect, but also actively construct the meaning of these phenomena themselves" (Tichenor et al. 2022, p. 435).

From an organizational perspective, material elements can be understood as a resource that helps gain a competitive advantage. Data access and storage architectures, for example, are tangible resources that can be acquired from external sources or developed within organizations (Gupta and George 2016). Sadowski (2019) identifies five major ways in which value is derived from data. First, data is used to profile and target people, for instance for personalized advertisements or to assess a person's financial risk. Second, data can optimize systems by identifying inefficiencies. Third, data is used to manage and control objects or subjects that deliver that data due to a power relationship based on knowledge gaps. Data can model probabilities, for instance, to detect "hot topics". Lastly, using data, digital products are built or existing products are enhanced with data and thereby become more useful.

Material components are ever-evolving, incomplete and open (Tichenor et al. 2022). Methods of using data evolve, and software becomes outdated and replaced, as technology advances. This continuous change is driven by individuals with specialist knowledge, as epistemic elites evaluate new technologies as appropriate for solving a task and then implement new technologies within organizations (Orlikowski 1996). That is, changes in the material components of epistemic infrastructures are not characterized by "dramatic discontinuity" but by "reciprocal variations in practice over time" (Orlikowski 1996, p. 66). It is hence the series of situated adaptations that achieve fundamental and ongoing change in epistemic infrastructures. Within an organization, this change is based on the everyday ordinary action of individual staff members possessing specific skills. Transferred to data specialists, processes of building and maintaining data architectures, analysis, and management may accompany organizational change processes that also affect organizational culture (Phillips-Wren et al. 2021). During these processes, material elements of the infrastructure become taken for granted and only become apparent when they are broken, disrupted, or missing. In the context of information gathering, for instance, de Haan et al. (2022) found that the use of algorithmic tools was so normalized in news production that journalists were hardly aware of their use and role. Lastly, what is lacking in epistemic infrastructures is worth paying attention to. Especially because their exclusion might be indicative of larger cultural, or socio-political motivations (Tichenor et al. 2022).

Interlinkages

The second-order level of epistemic infrastructures links the material components of the infrastructures into "a web of relations, processes, and practices" (Tichenor et al.

2022, p. 436). These linkages are achieved through interdependencies of different agents. Tichenor et al. (ibid., p. 436) argue, that data "—with their ability to simplify, stabilize, and travel—reconfigure relationships, dependencies, and structures of organizations". For journalism, linkages have fostered interdependencies between news organizations, advertisers, web analytic companies, platforms, and other digital intermediaries. These linkages in journalism can be seen, for example, in automated data transfers between news organizations and external third parties. Processes have been shown to build complex webs that increase over time (Kammer 2021). While news organizations benefit from such collaborations, these interdependencies are not equal, with tech companies such as Google being dominant agents in resource-exchange networks (Kammer 2021). Further, interlinkages *within* news organizations emerge, as communities with specialist knowledge, such as editorial technologists (e.g. developers, designers, or data specialists) augment their agency within the journalistic field (Lischka et al. 2022). Such tech-centric occupations represent a coding or data elite that possesses exclusive autonomy over socio-technical architecture for digital society (Burrell and Fourcade 2021). Epistemic communities are not entirely separate, but communities in epistemic infrastructures become expansive (Tichenor et al. 2022), as journalistic skills and expertise are increasingly diverse and tech-oriented. This is evidenced by news organizations' growing demand for computationally savvy staff (Wu et al. 2019), such as developers, designers, and data specialists, and by the fact that positions in news organizations and the competencies of news professionals have changed profoundly (Bro et al. 2016; Guo and Volz 2019; Kosterich 2021).

Epistemic practices are not unidirectional, or top-down, with knowledge travelling from data elite communities to the editorial office. Instead, knowledge is created by and through datafication as a dynamic process that is always a negotiation between the computational and editorial elite (Lischka et al. 2022). In other words, epistemic infrastructures are built through an iterative and "often informal process, one of constant making and remaking of partnerships, connections, and evidentiary claims" (Tichenor et al. 2022, p. 437). However, recent research indicates power differentials, demonstrating power and agency shifts between, as data scientists and developers take "the future of news" into their own hands (Schjøtt Hansen and Hartley, 2021). Despite power differentials, the interdependencies created by these actors all contribute to a common goal—whether it is using audience metrics to grow subscription bases, using AI to build news recommendation systems, or simply commodifying audience data. In doing so, epistemic interlinkages are producing a governing paradigm—the third-order level of the epistemic infrastructure, as laid out in the following section.

Paradigms

As laid out in the previous sections, datafication in journalism is more than the sum of individual practices of quantification (such as measurements of audience engagement). Instead, it can be understood as a paradigm that produces a context of a particular set of ideas that recognizes datafication as a governing frame that is more legitimate than others. In doing so, it privileges some kinds of cultural practices over others and fosters specific types of journalism(s) (Creech and Nadler 2018).

This governing frame, or paradigm, is linked to what has been described earlier as the culture of data, or quantitative turn in journalism (Macintyre 2000), for example, has argued, that there is a culture of data that (1) ascribes data an elevated role of authority, (2) is characterized by a self-evident relationship between data and reality which reframes how knowledge is formed, and (3) normalizes data practices across industries, leading to investments in data staff and data capabilities, believing that data produces greater knowledge and enables better decision making. While the notion of a culture of data is a key concept to understanding data's epistemic role, the concept of epistemic infrastructures is needed to highlight the consequences of datafication to connect epistemologies to the materials, and agents that support, sustain, and grow them into an overarching, taken-for-granted paradigm.

Linking data to epistemic frameworks, Schwarzkopf (2020) argues that the wealth of data is connected to a wealth of power, as data leads to knowledge, knowledge to power, power to credibility, and credibility to sustainability. However, it is a fallacy to assume that more data is related to more insight and knowledge. Instead, data overproduction may result in an inability to analyze data to produce actionable knowledge (ibid.). Hence, datafication of the organization can lead to ignorance rather than insight (ibid.). As such, data's epistemic paradigms bring about various ethical implications and reputational risks. News professionals might miss flawed assumptions and biases that models are built on or oversell the potential of data analysis together. This can lead to significant mistakes in editorial decision-making. As one example may serve the use of predictive analysis, a "technology that quietly drives everything from the ads that people see online to billion-dollar acquisition deals"[1], for the 2016 U.S. presidential election. Based on their analysis, the *NYT* estimated Hillary Clinton's chances of winning at 91% just 21 days before the election. While the *NYT* was not the only news organization that did not foresee Trump becoming president, this example underscores the risk of an epistemic paradigm that oversimplifies the potential of practices and de-prioritises critical awareness of their limitations.

This is particularly problematic because news professionals are largely operating in uncharted waters when it comes to ethical data principles, as widely accepted norms and standards are largely missing in the journalistic field (Porlezza and Eberwein 2022; Sørensen et al. 2020). Moreover, it is complicated by the fact that outside agents or partners, such as web analytic companies, cultivate profit-oriented norms and values by introducing material elements of the epistemic infrastructure that often stand in conflict with journalistic responsibilities (Belair-Gagnon and Holton 2018; Fürst, 2020), and thus produce further ethical challenges.

Method

Applying exploratory document analysis to job advertisements from twelve US and UK news outlets in the fall of 2021, we analyze the demand for data skills and the roles of data that are legitimized in news outlets, 1.5 years in the COVID-19 pandemic over a three-month period. The pandemic brings about a critical moment for accelerating the reconfiguration of journalism (Quandt and Wahl-Jorgensen 2021). As such, our material represents the demand for data skills at a moment of acceleration and points to the legitimate role of data for journalism.

Job ads of news outlets have been analyzed to understand the digital skill demand (Guo and Volz 2019) of journalists or to define journalism (Young and Carson 2018). They represent the public step in staff recruitment and include extensive information on the position, responsibilities, expectations, and organizational identity (Walters and Fage-Butler 2014). They denote a discursive manifestation of sought-after capabilities and thus provide an example of meta-journalistic discourse (Carlson 2016). Moreover, they give practices "an institutional and physical form" (D'Souza et al. 2023). Their analysis can serve as a window into the changing expectations, skills, and values that news organizations seek in their workforce, reflecting broader trends in journalism (Lindén et al. 2021).

Sample

The selected countries of the US and UK represent liberal media systems with moderate levels of news professionalism and a faster pace of digitization (Humprecht et al. 2022). Two incumbent broadcasts (CNN, Fox News (US); BBC News, ITV News (UK)), two print legacy (*New York Times* (*NYT*), *Washington Post* (*WP*) (US); MailOnline, *The Guardian* (UK)) and two online-only news outlets (Yahoo News, BuzzFeed US (US); BuzzFeed UK, The Independent (UK)) are selected within each country based on the highest reach indicated in Newman et al. (2022). We follow a two-step intensity sampling approach to identify an information-rich subsample regarding data capabilities for analysis. First, we collected all job ads from the twelve outlets from their online job pages over a three-month period in the fall of 2021 ($N = 463$). This sampling frame includes content-producing positions such as editors or reporters (77%, $n = 355$), strategic positions such as audience development, product management or data scientists (17%, $n = 78$), and technical positions such as motion designers and software engineers (6%, $n = 30$), drawing on Kosterich's categorization of journalism workers (Kosterich 2021). Next, we identified those job ads that required data skills (access, analysis, decision-making) across all positions (15%, $n = 69$, data capabilities subsample). This subsample includes data scientists, UX researchers, product managers, product designers, app developers or multimedia editor positions. Nine out of twelve news outlets remain in the subsample: CNN (6 job ads), Fox News (6 job ads), *NYT* (28 job ads), *WP* (5 job ads), Yahoo News (1 job ad) (US); BBC News (4 job ads), DMGT/MailOnline (9 job ads), *The Guardian* (7 job ads), and The Independent (3 job ads) (UK). BuzzFeed UK and ITVNews did not publish relevant job ads during the observation period.

The final sample includes two news organizations being crucial in the NSA revelations of PRISM by Edward Snowden (*NYT*, *Guardian*), who thus experienced and reported a critical view on data surveillance (Chadwick 2014). Hence, the sample may include critical cases in the sense that "if it doesn't happen here, it won't happen anywhere" (Patton and Patton 2002, p. 236) regarding critical considerations of datafication.

Material

Job ads serve as a public step in the recruitment process, providing extensive information about the position, responsibilities, expectations, and organizational identity (Walters and Fage-Butler 2014). They not only offer insights into journalistic skills

demand but give practices "an institutional and physical form" (D'Souza et al. 2023), representing a manifestation of meta-journalistic discourse (Carlson 2016). As such, they are a complex genre of recruitment and strategic organizational communication that include detailed information not only on the position but also on the organization. We assume that material elements of epistemic infrastructures that have become taken for granted (Tichenor et al. 2022) become explicit in job ads.

A typical job ad format includes sections such as job title, job description, personnel specification, practical information, and organizational identity (Walters and Fage-Butler 2014). In our sample, job ads generally start with a description of key aspects of the role and its reporting lines, followed by detailed information about responsibilities, requirements and qualifications, working conditions, organizational identity, and application procedures ("about the position," "what you will do," "who you are," "what we offer," "about us," and "how to apply"). On average, job ads in the data capabilities sample ($n = 69$) were 551 words in length, with a maximum of 1397 words for a Senior News Editor position at BBC News and a minimum of 188 words for a Senior Staff Editor at *NYT*. The shorter *NYT* job ad only included information about role descriptions, requirements, and contact information.

For our analysis, role description, responsibilities, and requirements contained rich information. Concerning materialities, the requirements section was most informative as it requests precise software competence (e.g. "2+ years coding experience, Python preferred", Data Scientist, *NYT*). Regarding interlinkages, the job ads often explicate a collaboration situation in the role description or responsibilities section, such as "work in close collaboration with the newsroom, marketing, product, engineering, advertising and finance departments" (Data Analyst, *WP*). For paradigms, role description and responsibilities sections were most insightful, extensively explaining what candidates should do with what data.

Analysis

Our approach involves a mixed-method analysis that follows a sequential path of qualitative, quantitative, and qualitative procedures. First, we identify the materialities of the data, then we establish a collaboration network. Finally, we conduct a qualitative analysis to explore the broader roles and agency of the data.

Quantitative Network Analysis

We traced the collaboration of a position within each function towards other functions as described in the job ads for the data capabilities subsample ($n = 69$). We categorized the data capabilities positions into six groups based on their position titles: (1) social media, (2) data analytics, (3) editorial, (4) visual (e.g., graphics journalist, visual designer), (5) product (e.g., product manager, product designer, UX designer), and (6) engineering (e.g., front-end developer, engineering manager). Three additional groups were identified (corporate level, commercial team, external parties (e.g., agency partners, external suppliers)) with which data-capability agents are expected to collaborate. Therefore, the network has a total of nine nodes (see Figure 1).

The network edges represent the collaboration descriptions between nodes in a directional way. For instance, a data scientist at *NYT* is expected to "collaborate with

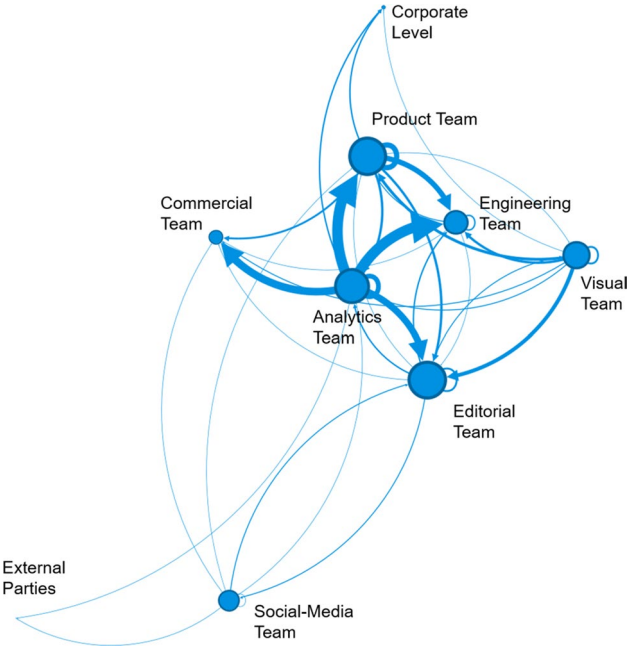


Figure 1. Data-capabilities collaboration network. Collaboration network based on $n = 69$ job ads for roles that require data capabilities. Node size = degree (number of in- and outgoing connections in the network). Visualization using the Force Atlas layout algorithm in Gephi that considers the distance and closeness of nodes proportionally.

both technical and non-technical colleagues in editorial, product management, marketing, and executive leadership groups". We coded directional relations from this analytics position to the editorial, product, commercial, and corporate teams in an adjacency matrix. In this way, we coded all collaborative relationships in the data-capabilities subsample. As a result, all relations in the network represent the expected direction of collaboration from one team to another.

Qualitative Content Analysis

The subsequent analysis was structured using thematic analysis (Braun and Clarke 2006), based on the data capabilities subsample ($n = 69$). Employing an inductive analysis within a constructivist paradigm involved several stages including familiarization with the data, identification of patterns and interrelationships, reviewing and refining themes, and finally, drawing conclusions and interpretations from them. To grasp the material components of data, we searched for references to big data architecture and business intelligence (BI) analytics. In line with Phillips-Wren et al. (2021), we followed material components in a data process flow from data sourcing, preparation, and storage to analysis and visualization. We used Tichenor et al.'s (2022) theoretical lens of epistemic infrastructures to form sensitizing concepts. Sensitizing concepts are understood as "those background ideas that inform the overall research problem", offering ways of seeing, organizing, and understanding experience (Charmaz 2003, p. 259). The sensitizing concepts of materialities, interlinkages, and paradigms

then formed the conceptual framework of this study and enabled us to assess how the notion of epistemic infrastructures could be translated in the journalistic context.

Through "repeated reading" (Braun and Clarke 2006), we then traced the connections between hiring criteria and discourses found in the job and the creation of epistemic infrastructures and sorted the data into meaningful groups (Tuckett 2005). Notably, all dimensions (materialities, interlinkages, and data paradigms) were easily identified, arguably reflecting the pervasive nature of datafication that reproduces across industries.

Findings

First Order Level: Materialities of Data

Material components of data become manifest in references of databases and software skills in the requirements sections of job ads, especially for data analysts and data scientists. While database sourcing and preparation are mentioned rarely (e.g. "Strong knowledge of data migration/ETL [extract, transform, load]" (Senior Data Analyst, MailOnline)), data storing, analysis, and visualization are mentioned regularly. As a standard for storage and processing, extensive experience with or "[m]astery of" (Senior Manager - Metrics and Analysis, Coverage and Audience Analytics, *NYT*) SQL is required.

For data analysis and visualization, analysts are expected to use "BI tools to deliver repeatable insights" (Product Analytics, CNN). The list of analysis and visualization tools includes Tableau, R Shiny, D3 or Power BI (Data Analyst, *WP*), indicating that news organizations are relying on off-the shelf tools of third parties, with which they form interlinkages. Experience with R and Python is requested for data manipulation and analysis (Product Analytics, CNN). A CNN job ad for a product analyst summarizes these three skill areas of processing, analyzing, and visualizing: "querying databases (SQL), leveraging a scripting language (R/Python) to analyze data, and/or using BI tools to deliver repeatable insights", which can be regarded as core materialities of data analytics. Except for user experience (UX) design positions, material components of data management are not explicitly referred to in the remaining job ads. Other roles are requested to make decisions based on data beyond the data analytics process.

Second Order Level: Interlinkages of Data Elites

To grasp the professional community of data elites and their interlinkages within the news organization, we visualize a collaboration network (Figure 1). Overall, the density of the network is .53. In other words, just over half of all possible connections between the teams exist (53%). The analytics team is the central node in this network. This team has the most outgoing relations and is expected to work with a variety of other teams. Most frequently, data analysts are expected to collaborate with the engineering and product teams followed by commercial and editorial teams (Figure 1). Analysts have a frequent relationship with the engineering team, which is expected to support analytics to develop "new, automated, user-friendly tools and dashboards" (Analyst, Data & Insights, *NYT*) and "to bring to life new tools and

forecasts" (Sr. Analyst, Data & Insights (Subscription Growth), *NYT*). The engineering team is needed to implement information infrastructures for data analytics. For the product team, data analysts are expected to provide insights from data. Analysts should "quantify and identify patterns in user engagement and activation, funnel optimization, and retention" (Product Analytics, CNN). Therefore, they develop "KPIs and success measurement plans" (Data Analyst, *WP*) and explore "the commercial use of data" (Analyst, The Independent). Product managers then implement insights from analytics to news products by managing product roadmaps, which are then implemented by the engineering team.

Regarding the commercial team, data analysts are expected to identify and solve "business problems" (Data Scientist, *The Guardian*). Such problems refer to the audience side of the news business "to grow our engaged audience and paying subscriber base" (Senior Manager, Games Analytics, *NYT*). Regarding the advertising market, data analysts provide "business intelligence to support our sales team in generating advertising revenues and retaining clients" (Analyst, The Independent). Eventually, the datafication business itself is mentioned requesting analyses for the "commercial use of data" (Analyst, The Independent).

Analysts have outgoing connections to the editorial team to inform them about the "use data and analytics to support outstanding journalism" (Analyst, Data & Insights, *NYT*). Here, data analysts are requested to "[p]rovide regular analyses for newsroom stakeholders" (Senior Analyst, Data & Insights, *NYT*). There are indications that data agencies might at times stand in conflict with editorial agencies. This becomes particularly evident when news organizations suggest that data analysts at a senior level will be able to "take ownership over reporting" (Senior Data Analyst, MailOnline). Thus, the output of data analytics is fed into procedures at the editorial office, potentially in a hegemonic way. There are fewer relations with the social-media and visual team, the corporate level, and external parties, who are positioned on the periphery of the network.

Third Order Level: Paradigms of Data

Data capabilities are tied to four main goals that are connected to paradigms of data and expected to advance journalism through data practices. Data is envisioned to (1) measure performance as a form of control, (2) strengthen user centrism, (3) improve products, (4) increase profits.

Control

By "establishing meaningful KPIs", data is promised to not only enable professionals to "measure success" but also to "define outcomes for achievement" (Senior Product Manager, MailOnline). In other words, data is thought to give professionals a form of *control,* as it is used for "success measurement plans" (Data Analyst, *WP*), in which journalistic work is quantified and valued based on performance metrics.

User Centrism

As a result of measuring performance based on user behavior, data is deeply interlinked with user centrism. Measured based on quantified audience metrics, users'

interests become central to journalism. News professionals are not only expected to investigate informational preferences, but to be "the voice and advocate of our users" (Senior Product Manager, MailOnline), or to "be an advocate for the customer" (Product Designer in Customer Care, *NYT*). User centrism then also has consequences for how news professionals perceive and relate to audiences. This becomes apparent through language. Audiences are prominently referred to as "users" or even "customers". Further, professionals are rarely asked to use data to examine the informational needs of audiences, but rather to "understand as much as [they] can about [their] customers and their behavior" (Data Analyst, MailOnline), to improve "customer experience" (Product Manager, Growth, *NYT*) in order to drive users "to spend more time with, and more money with" the news organization (Data Analyst, MailOnline). Only one ad explicitly relates to "audience needs"—CNN asks their candidate to focus "relentlessly on audience needs in a time of turbulence and misinformation." (Executive Editor, CNN Breaking News, CNN). However, user centrism is framed to be in favour of audiences. Professionals are envisioned to "be a stalwart champion" of users (Product Manager, Growth, *NYT*), by developing "a deep understanding of how users engage" (Staff Product Designer in CNN Digital Product Design Team, CNN), they are thought to use "data in a way that empowers end users" (Senior Manager, Data & Insights (Subscription Records), *NYT*).

Product Development

Relatedly, job ads focus on improving news products to increase user engagement, instead of centering data use to improve journalistic content. This is reflected in the role of product managers, who are hired to ensure "analytics and learnings are fed into product development", and to "build new features and improve our digital product." (Digital Product Manager, DMGT, startup "the Knowledge"). Data analysts or audience developers who, in theory, might use data to understand shortcomings of journalistic content, are rather asked "to predict and classify reader behaviour", "to anticipate and forecast performance changes" (Data Scientist, *The Guardian*), or in short, "to use research and analytics to identify and capitalize on audience acquisition" (Senior Manager to lead audience development, Fox News).

Profits

This focus ties in with data's ascribed role to increase *profits* for news organizations. Data is, for instance, to be used to "fuel subscription growth" (Sr. Analyst, Data & Insights (Subscription Growth), *NYT*), in order "to help maximize subscription revenue." (Manager, Data & Insights (Standalone Product Growth), *NYT*). To do so, data professionals are expected to grow the value of data "as measured by quality, richness, trust, and accessibility" (Data Quality Analyst, *The Guardian*), "to identify and capitalize on audience acquisition opportunities" (Senior Manager to lead audience development, Fox News) and "to help grow and track all audience development" (Senior Manager to lead audience development, Fox News), in order to "allow the business to get maximum value from our data" (Data Developer, *The Guardian*).

Consequences for Journalism

Data paradigms not only structure specific goals for imagining and doing with data, but they also confer legitimacy and authority on specific epistemic communities and their ideas, norms, and practices. Findings suggest epistemic infrastructures consolidate epistemic authority (1) and change agency (2) in data elites, to support, sustain, and grow datafication in news organizations. Finally, results indicate a de-prioritization of ethical concerns (3), as a consequence of these shifts.

Epistemic Authority

Data's elevated role in knowledge creation and subsequent decision-making is continuously recognized by organizations that require applicants' epistemology to be aligned with notions of datafication. A candidate should "thrive in data-driven decision making", is "leading […] a data-driven approach to product development" (Digital Product Manager, DMGT, startup "the Knowledge"), and "skilled at using data and analytics tools to make decisions" (Executive Editor, CNN Breaking News, CNN). Put differently, candidates should not only be capable of using (almost exclusively quantitative) data but showcase a commitment to data-driven knowledge creation and successive decision-making. In line with this, skills to assess shortcomings in data-based knowledge creation were not sought after and news organizations hardly ever require candidates to use qualitative methods (UX design positions being a notable exception here). Instead, given the epistemic authority assigned to data practices, data is understood as providing answers and solutions. Data scientists are envisioned to have an affinity to "answer hard questions with data" (Data Scientist, *NYT*), while data analysts are asked to "engineer data solutions" (Analyst, Data & Insights, Engagement, *NYT*).

Change Agency

As news organizations assign data a primary role in claims to authority, it is linked to professionals' capacity to drive change. Professionals who are in charge of using data for decision-making are trusted to make structural changes that shape processes of news creation, distribution, and consumption. Using data, professionals are not only offered to follow their own "mission to shape our digital news products." (Manager, Data & Insights, Engagement, *NYT*), but also to start "from the ground up deciphering which platforms are right for [their] audience" (Social Media Lead, MailOnline). Accordingly, data enables professionals to become change agents themselves. They are considered to be "the next innovator" (Data Analyst, *WP*) and promised to "shape the future of [the] business" (Senior UX Researcher, *WP*). Professionals working with data are not only expected to "formulate data-driven insights", but also to "champion them" (Senior Product Analyst, *NYT* Cooking). In other words, they are expected to "promot[e] a data-driven approach" (Digital Product Manager, DMGT, startup "the Knowledge") and "act as a facilitator of change" (News Designer, *WP*) in journalism. As a change agent with epistemic authority, data is envisioned to "transform every aspect of the craft and practice of journalism" (Data Scientists, *NYT*). There are indications that data agencies might at times stand in conflict with editorial agencies. This becomes particularly evident when news organizations suggest that data analysts

at a senior level will be able to "take ownership over reporting" (Senior Data Analysts, MailOnline). Accordingly, furthering datafication is also regarded as a positive advancement in the context of editorial decision-making.

De-Prioritization of Ethical Concerns

Notably, only a single ad listed "data privacy or compliance experience" (Campaign Data Analyst, The Independent) as a hiring criterion. It is worth questioning why skills relating to privacy protection, data ethics, or expertise in data regulations are not sought after when news organizations hire professionals who are working with audience data. While news organizations made passion for data a formalized hiring criterion, the ability to critically navigate the complex set of issues relating to ethical, methodological, and epistemological perspectives, that will inevitably arise when using data, was not formally demanded. This is particularly striking, given the fact job ads recognized journalism's social responsibility and signal their commitment to other values prominently. This absence suggests that data is viewed as an abstract resource that audiences naturally produce for news organizations to leverage in order to optimize processes, structure decision making and increase profits. There are no indications, however, that show that audience data is viewed as something that needs to be protected, that requires consent, responsibility and ethical foresight, as unintended consequences may produce harm to audiences at any point in the future.

Discussion

This study reveals epistemic infrastructures consisting of three analytical orders—data, hardware, and software as material building blocks of the infrastructure; interlinkages that form processes and relations between data elites and other professionals inside and outside of news organizations; and epistemic data paradigms that drive and structure the datafication in journalism. Findings demonstrate how epistemic infrastructures in digital journalism are constructed through these dimensions (Tichenor et al. 2022), as evident in formalized hiring criteria in job ads (Table 1).

First, materialities become apparent in job ads for positions of data analysts as part of requested software experience. Materialities focus on the data storage, processing, analysis, and access stages of the data flow process. These materialities shape the potential of data primarily as a BI tool for news organizations. That is, data solves existing problems and supports decision-making. In this view, more and more data collection and interpretation will lead to greater insights, as suggested by Phillips-Wren et al. (2021). Notably, materialities for data sourcing are not addressed in the job ads. This could be due to a lack of demand during our sampling period or procedures already being in place in the selected news organizations. However, this leaves the ethically relevant stage of data extraction and related considerations and paradigms intransparent. Yet, the material components of storage, processing, analysis, and access are important drivers of more longstanding changes. They are not simply resources for news organizations, but through the creation of data architectures for analysis and management, organizational change may be enacted that has economic as well as cultural implications (Tichenor et al. 2022).

Table 1. Epistemic data infrastructures shaping the datafication of news organizations.

Analytical orders of epistemic infrastructures	Definition	Manifestation in journalism	Exemplary quotes	Outcomes
Materialities	Material and semiotic building blocks of infrastructures	e.g. Data, hardware, software, reports and dashboards	"BI tools to deliver repeatable insights"	(1) Data elites gain in epistemic authority and agency (2) De-prioritization of ethical concerns
Interlinkages	A web of relations, processes, and practices, achieved through interdependencies of different agents	Data elites and data practices connect computational staff with product managers, commercial, editorial teams, and the corporate level.	Data analysts provide "business intelligence to support our sales team in generating advertising revenues and retaining clients"	
Paradigms	A logic and structure that promotes and validates ways of doing journalism	Measure performance as a form of control; Strengthen user centrism; Improve products; Increase profits	"[E]stablishing meaningful KPIs (…) to measure success" "[B]e an advocate for the customer" "[B]uild new features and improve our digital product." "[A]llow the business to get maximum value from our data"	

Therefore, secondly, materialities determine interlinkages between data analysts and other teams in news organizations. For data preparation, data analysts rely on software developers. For processing and analysis, data analytics work autonomously. For providing other teams in the organization access to data, data analysts collaborate with a variety of roles with data competencies. Product managers are the second most important group next to engineers. This triangular relation between data analysts, engineers, and product managers may represent the core components of a datafication elite in news organizations with epistemic authority (Sadowski 2019).

Third, this datafication elite is expected to embrace data culture (Macintyre 2000) by deriving value from data (Sadowski 2019). In line with Schjøtt Hansen and Hartley (2021), findings suggest power differentials between data and editorial elites, with data elites, enabled by the recognition of data's epistemic authority, being granted decision-making powers. Accordingly, we find that data practices not only reconfigure relationships, dependencies, and structures of organizations (Tichenor et al. 2022, p. 6), but when enacted through data's epistemic infrastructure, may produce inequities between journalists and those in control of leveraging data materialities. Data paradigms in turn contribute to the advancement of epistemic infrastructures of the organization. Data paradigms foster a positivist approach to knowledge creation and subsequent decision-making, that cultivate profit-oriented norms and values in the profession. This dynamic suggests that data paradigms in journalism demonstrate a self-enhancing nature, with data elites acting as enabling forces to further data paradigms outside of their epistemic communities.

Overall, findings suggest that these professionals form a data elite who foster datafication within the organization. Their materialities enable them to enact data, ways to analyze data, as well as paradigms related to data, as indicated in Orlikowski's (1996) idea of technology experts that enact new technology for their organization. This is coupled with the fact that news organizations expect applicants' epistemology

to be aligned with notions of BI and datafication. Drawing on their specialist knowledge, news organizations request candidates to pass on data's epistemic authority to other professionals within the organization. They thus are considered important actors in enabling data's change agency within the organizational context. After all, this professional group of data analysts and engineers *datafy* the content, management, and revenue units of the news organization through interlinkages with product managers, editorial staff, the corporate level, and commercial teams.

Schwarzkopf (2020) has cautioned that an over-production of data may lead to ignorance rather than insight. This cautionary view is also applicable to data's paradigm rendering data practices as epistemically superior, while acknowledgement of the ethical dimensions of data work is missing—even in job ads of *NYT* and *The Guardian*, which are news organizations that are sensitized to issues of data privacy and surveillance (Chadwick 2014). Echoing Porlezza and Eberwein (2022), we see no evidence of recognized norms or standards for ethical data practices within the journalistic field. The disconnect between journalism's social responsibility and data paradigms is also reflected in other findings of the study. News organizations are not expressing the aim to use data to, for example, further their understanding of audiences' informational needs, but rather focus on using data to measure performance in an aim to control organizational targets. It should not be dismissed that these two aims might be connected. Investigating user behavior might lead to new insights into audiences' informational needs. Nevertheless, explicitly expressed aims suggest where news organizations place value.

Major limitations of this study concern the purposive sampling, sample size, as well as material. First, the subsampling approach led to a small sample of job ads compared to the selection frame. Due to a lack of demand, the data capabilities of a subset of originally selected news outlets were analyzed. While the sampling procedure allows to quantify demand for data capabilities compared to other roles, future studies should consider a longer observation period to ensure that news outlet diversity is preserved. Second, choosing job ads as material comes with several disadvantages. Job ads only reveal the demand for new positions that are not filled in-house. Our material can thus neither reveal the status of data capabilities in the news organizations studied nor relevant positions filled with staff in-house. Moreover, job ads represent a genre of strategic organizational communication targeted to prospective staff with the goal to generate a large number of applications. Therefore, job ads strive to generate a positive image with prospective staff. In fact, on rare occasions, job ads in our sample went beyond strategic recruiting talk and revealed problems in teams the new role is expected to solve. Also, job ads revealed a variance in explicitly listing materialities in the expectations section. Some job ads rather described the tasks that should be fulfilled with data analytics while others referred to software and tools that are expected to have experience with. Still, job ads communicate responsibilities and expectations quite extensively since they are tools for staff selection that have to persist legally. Because material components of epistemic infrastructures are always evolving and incomplete (Tichenor et al. 2022), findings inevitably reflect a snapshot in time. Third, regarding data fetishism or problems of data overproduction may become transparent only in interviews with data analysts or in field observations of the work of data analysts. In line with Schwarzkopf (2020,

p. 211) assuming that "[a] good number of [data] professionals are acutely aware of the problematic nature of data (over-)production and the tenuous relationship between data and actionable knowledge", future research may gain valuable insight into the work of data analysts in single or multiple case studies. Finally, when considering the data flow process starting with data sourcing and ending with data usage, the first step of sourcing does not become transparent in job ads. To understand the initial steps of data extraction, future research would need to investigate the data departments of news organizations in depth.

Conclusion

This study has demonstrated that digital journalism is structured through epistemic data infrastructures and in doing so, investigated its materialities, networks, and governing paradigms. Data's epistemic infrastructures in journalism are forged through wider economic and socio-technological expressions of power, strengthening interdependencies between news organizations, platforms, advertisers, and audiences, while capitalizing on data as discussed in Willig (2021). In line with Tichenor et al. (2022) and Ekström (2002), we conclude that data epistemologies are structured by architectures of materialities, reinforced through social interlinkages with a broader network of data agents, and institutionalized as paradigmatic setting that produce knowledge meant to solve organizational problems. This occupational class, which we, drawing on Burrell and Fourcade (2021), call journalism's "data elite", holds technical control over the data-related means of processes in news organizations. Data elites are consolidating epistemic agency thanks to their specialist knowledge and are able to leverage their epistemic authority to support, sustain, and grow datafication in their news organization and, eventually, on an industry level, in journalism. That is, data practices and related skills are not only a tool for news organizations but are institutionalized as the very core of paradigms that are restructuring journalism, and shaping the way in which news professionals view their profession. As datafication continues to influence news organizations, there may be a growing emphasis on data-driven decision-making and performance measurement. This shift could lead to a more business-oriented approach to journalism, prioritizing profitable products over serving diverse sets of audiences. Therefore, we argue that efforts to leverage technology for journalistic pursuits need to be guided by epistemic diversity and that news organizations need to channel resources into various ways of news-making to avoid "fetishization of data" and over-production of data that may lead to ignorance rather than insight (Schwarzkopf 2020). In line with Schaetz (2023) we argue that more qualitative and traditional approaches to understanding audiences' information needs are necessary to improve audiences' relationship with the news media. Moreover, we need a better understanding of how computational and more qualitative approaches can complement each other.

Our findings suggested that news organizations are de-prioritizing ethical concerns of datafication, which is indicative of a larger problem of journalism reproducing a culture in which data is viewed as a natural resource to be leveraged, that falls out of journalists' traditional understandings of social responsibility. Our findings raise concerns over a disconnect between journalism's social responsibility and data

paradigms that prioritize user centrism over data privacy and surveillance issues. This is especially troubling as journalism increasingly uses massive amounts of data for AI practices. We believe equitable journalism under datafication requires news organizations to value—culturally as well as economically—data ethics, privacy, and regulatory expertise, and to elevate these skills to a formalized hiring criterion in job ads for data professionals. Lastly, we hope to encourage future research to use the framework of epistemic infrastructures, as a theoretical lens to understand the complex interplay between human actors, data, technology, and their ethical and cultural implications in shaping the future of journalism.

Note

1. Lohr, S. (2016, November 10). How Data Failed Us In Calling An Election (Published 2016). The New York Times. https://www.nytimes.com/2016/11/10/technology/the-data-said-clinton-would-win-why-you-shouldnt-have-believed-it.html

Disclosure Statement

No potential conflict of interest was reported by the author(s).

ORCID

Nadja Schaetz http://orcid.org/0000-0001-6692-3471
Juliane A. Lischka http://orcid.org/0000-0002-4845-0946
Laura Laugwitz http://orcid.org/0000-0001-8527-2504

References

Schaetz, N. 2023. "Journalism & Audience Datafication: How Audience Data Practices Shape Inequity." *Digital Journalism*. Advance online publication. https://doi.org/10.1080/21670811.2023.2171894.

Lischka, J. A., N. Schaetz, and A.-L. Oltersdorf. 2022. "Editorial Technologists as Engineers of Journalism's Future: Exploring the Professional Community of Computational Journalism." *Digital Journalism*. Advance online publication. https://doi.org/10.1080/21670811.2021.1995456.

Adams, P. C. 2020. "Agreeing to surveillance: Digital News Privacy Policies." *Journalism & Mass Communication Quarterly* 97 (4): 868–889. https://doi.org/10.1177/1077699020934197

Belair-Gagnon, V., and A. E. Holton. 2018. "Boundary Work, Interloper Media, And Analytics In Newsrooms: An Analysis of the Roles of Web Analytics Companies in News Production." *Digital Journalism* 6 (4): 492–508. https://doi.org/10.1080/21670811.2018.1445001

Braun, V., and V. Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77–101. https://doi.org/10.1191/1478088706qp063oa

Bro, P., K. R. Hansen, and R. Andersson. 2016. "IMPROVING PRODUCTIVITY IN THE NEWSROOM?: Deskilling, Reskilling and Multiskilling in the News Media." *Journalism Practice* 10 (8): 1005–1018. https://doi.org/10.1080/17512786.2015.1090883

Burrell, J., and M. Fourcade. 2021. "The Society of Algorithms." *Annual Review of Sociology* 47 (1): 213–237. https://doi.org/10.1146/annurev-soc-090820-020800

Carlson, M. 2016. "Metajournalistic Discourse and the Meanings of Journalism: Definitional Control, Boundary Work, and Legitimation: Metajournalistic Discourse and the Meanings of Journalism." *Communication Theory* 26 (4): 349–368. https://doi.org/10.1111/comt.12088

Chadwick, A. 2014. Boundary-Drawing Power and the Renewal of Professional News Organizations: The Case of The Guardian and the Edward Snowden National Security Agency Leak. 22.

Charmaz, K. 2003. "Grounded Theory: Objectivist and Constructivist Methods." In *Strategies for Qualitative Inquiry*, edited by N. K. Denzin and Y. S. Lincoln, 2nd ed., 249–291. Thousand Oaks, CA: Sage.

Creech, B., and A. M. Nadler. 2018. "Post-Industrial Fog: Reconsidering Innovation in Visions of Journalism's Future." *Journalism* 19 (2): 182–199. https://doi.org/10.1177/1464884916689573

de Haan, Y., E. van den Berg, N. Goutier, S. Kruikemeier, and S. Lecheler. 2022. "Invisible Friend or Foe?How Journalists Use and Perceive Algorithmic-Driven Tools in Their Research Process." *Digital Journalism* 10 (10): 1775–1793. https://doi.org/10.1080/21670811.2022.2027798

D'Souza, R. A., J. S. Rauchberg, and B. Nieto-Fernandez. 2023. "The Labor of Diversity in the 2020–2021 U.S. communication Job Market." *Journal of Applied Communication Research*. Advance online publication. https://doi.org/10.1080/00909882.2023.2173532

Ekström, M. 2002. "Epistemologies of TV Journalism: A Theoretical Framework." *Journalism* 3 (3): 259–282. https://doi.org/10.1177/146488490200300301

Fürst, S. 2020. "In the Service of Good Journalism and Audience Interests? How Audience Metrics Affect News Quality." *Media and Communication* 8 (3): 270–280. https://doi.org/10.17645/mac.v8i3.3228

Guo, L., and Y. Volz. 2019. "(Re)Defining Journalistic Expertise in the Digital Transformation: A Content Analysis of Job Announcements." *Journalism Practice* 13 (10): 1294–1315. https://doi.org/10.1080/17512786.2019.1588147

Gupta, M., and J. F. George. 2016. "Toward the Development of a Big Data Analytics Capability." *Information & Management* 53 (8): 1049–1064. https://doi.org/10.1016/j.im.2016.07.004

Humprecht, E., L. Castro Herrero, S. Blassnig, M. Brüggemann, and S. Engesser. 2022. "Media Systems in the Digital Age: An Empirical Comparison of 30 Countries." *Journal of Communication* 72 (2): 145–164. https://doi.org/10.1093/joc/jqab054

Kammer, A. 2021. "Resource Exchanges Between Mobile News Apps and Third-Parties." *Digital Journalism* 9 (10). https://doi.org/10.1080/21670811.2021.2000455

Kosterich, A. 2021. "Reengineering Journalism: Product Manager as News Industry Institutional Entrepreneur." *Digital Journalism*. Advance online publication. https://doi.org/10.1080/21670811.2021.1903959

Lindén, C.-G., K. Lehtisaari, M. Grönlund, and M. Villi. 2021. "Journalistic Passion as Commodity: A Managerial Perspective." *Journalism Studies* 22 (12): 1701–1719. https://doi.org/10.1080/1461670X.2021.1911672

Macintyre, A. 2000. "Ethics and the Culture of Data: The Help and Hindrance of Modern Political Philosophy." 13. Political Studies Association, Media and Politics Group Conference: Media and Politics in the Age of Crisis and Change.

Newman, N., R. Fletcher, A. Schulz, S. Andı, C. T. Robertson, and R. K. Nielsen. 2022. Reuters Institute Digital News Report 2021 10TH EDITION. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2021

Orlikowski, W. J. 1996. "Improvising Organizational Transformation Over Time: A Situated Change Perspective." *Information Systems Research* 7 (1): 63–92. https://doi.org/10.1287/isre.7.1.63

Patton, M. Q., and M. Q. Patton. 2002. *Qualitative Research and Evaluation Methods* (3 ed). Thousand Oaks: Sage Publications.

Phillips-Wren, G., M. Daly, and F. Burstein. 2021. "Reconciling Business Intelligence, Analytics and Decision Support Systems: More Data, Deeper Insight." *Decision Support Systems* 146: 113560. https://doi.org/10.1016/j.dss.2021.113560

Porlezza, C., and T. Eberwein. 2022. "Uncharted Territory: Datafication as a Challenge for Journalism Ethics." In M. Karmasin, S. Diehl, & I. Koinig (Eds.), *Media and Change Management* (pp. 343–361). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-86680-8_19

Quandt, T., and K. Wahl-Jorgensen. 2021. "The Coronavirus Pandemic as a Critical Moment for Digital Journalism: Introduction to Special Issue: Covering Covid-19: The Coronavirus Pandemic as a Critical Moment for Digital Journalism." *Digital Journalism* 9 (9): 1199–1207. https://doi.org/10.1080/21670811.2021.1996253

Sadowski, J. 2019. "When Data is Capital: Datafication, Accumulation, and Extraction." *Big Data & Society* 6 (1): 205395171882054. https://doi.org/10.1177/2053951718820549

Schjøtt Hansen, A., and J. M. Hartley. 2021. "Designing What's News: An Ethnography of a Personalization Algorithm and the Data-Driven (Re)Assembling of the News." *Digital Journalism* 10 (7). https://doi.org/10.1080/21670811.2021.1988861

Schwarzkopf, S. 2020. "Sacred Excess: Organizational Ignorance in an Age of Toxic Data." *Organization Studies* 41 (2): 197–217. https://doi.org/10.1177/0170840618815527

Sørensen, J. K., H. Van den Bulck, and S. Kosta. 2020. "Stop Spreading The Data: PSM, Trust, and Third-Party Services." *Journal of Information Policy* 10 (1): 474–513. https://doi.org/10.5325/jinfopoli.10.1.0474

Sørensen, J. K., and S. Kosta. 2019. "Before and After GDPR: The Changes in Third Party Presence at Public and Private European Websites." *The World Wide Web Conference on* - WWW *'19*, 1590–1600. https://doi.org/10.1145/3308558.3313524

Tichenor, M., S. E. Merry, S. Grek, and J. Bandola-Gill. 2022. "Global Public Policy in a Quantified World: Sustainable Development Goals as Epistemic Infrastructures." *Policy and Society* 41 (4): 431–444. https://doi.org/10.1093/polsoc/puac015

Tuckett, A. G. 2005. "Applying Thematic Analysis Theory to Practice: A Researcher's Experience." *Contemporary Nurse* 19 (1-2): 75–87. https://doi.org/10.5172/conu.19.1-2.75

Walters, N., and A. Fage-Butler. 2014. "Danish Job Advertisements: Increasing in Complexity." *Communication & Language at Work* 3 (3): 38. https://doi.org/10.7146/claw.v1i3.16556

Willig, I. 2021. "*From Audiences to Data Points: The Role of Media Agencies in the Platformization of the News Media Industry*." *Media, Culture & Society* 44 (1): 56–71. https://doi.org/10.1177/01634437211029861

Wu, S., E. C. Tandoc, and C. T. Salmon. 2019. "When Journalism and Automation Intersect: Assessing the Influence of the Technological Field on Contemporary Newsrooms." *Journalism Practice* 13 (10): 1238–1254. https://doi.org/10.1080/17512786.2019.1585198

Young, S., and A. Carson. 2018. "What is a Journalist?: The View from Employers as Revealed by Their Job Vacancy Advertisements." *Journalism Studies* 19 (3): 452–472. https://doi.org/10.1080/1461670X.2016.1190665

Journalism and AI: Macro Epistemics, Inequities & A Global Journalistic Coalition

Nadja Schaetz, Journalism and Communication Studies, Universität Hamburg Germany, nadja.schaetz@uni-hamburg.de

Anna Schjøtt Hansen, Faculty of Humanities, University of Amsterdam, a.s.hansen@uva.nl

Abstract

Journalistic AI training programs shape the profession through the introduction of new skills and knowledge construction. They are also a site where inequities become apparent. Ethnographically following the *JournalismAI's Academy for Small Newsrooms* (situated at the LSE and funded by the Google News Initiative), we ask whether training programs can contribute to the mitigation or amplification of power asymmetries between resource rich and smaller news organizations, or possibly even reduce journalism's dependence on platform businesses. To do so, the study contextualizes the training initiative (that encompasses 41 news organizations, situated in 23 countries) within a framework of macro-epistemics. The Academy was perceived as a potential pathway to mitigate inequities by providing access to cultural, political, and economic dimensions of epistemic power. However, tensions emerged between instructors' efforts to technologically emancipate small newsrooms and promoted practices that reinforce Big Tech's position of power. Against this backdrop, we offer the concept of *"sticky epistemics,"* to describe how practices conflicting with professional ideals persist due to pre-existing power asymmetries. Lastly, we highlight the role of training programs in the formation of a global AI network, and the potential to build a coalition that fosters epistemic agency for news professionals and challenges power asymmetries in the field.

Keywords: epistemics, AI, journalism education, skills, inequalities, power, Big Tech

Introduction

Over the last decade, artificial intelligence (AI) applications have increasingly moved into newsrooms across the globe and, as a result, have become an integral part of news discovery, production, and distribution (Beckett, 2019; Broussard et al., 2019; de-Lima-Santos & Ceron, 2021; Marconi, 2020; Simon, 2022). However, journalistic AI adoption has not been universally accessible. Scholars and practitioners expressed concerns that especially smaller, local news outlets lack the resources to keep up with technological advancements, resulting in diminished competitiveness (Beckett, 2019; Rinehart & Kung, 2022). Further, research has highlighted how local geographic context produces unique challenges for AI adoption, in Pakistan (Jamil, 2021), in the Global South (Soto-Sanfiel et al., 2022), or in South Africa (Munoriyarwa et al., 2023). As a response, a plethora of new training programs have emerged to make AI practices more accessible to journalists around the world, such as the Google News Initiative (GNI) 'Introduction to Machine Learning', the Future Journalism Today's training on AI for media makers, and Associated Press' Local AI Initiative self-paced curriculum. Heuer et al. (2021) and Luchs et al., (2023) have highlighted the importance of educational sites in producing the epistemological foundations of the field of machine learning and computer science (see also Mackenzie, 2017). In the context of AI and news, these training initiatives are equally becoming a new and important space in which fundamental understandings and conceptualizations of what AI is, and should be, are represented and negotiated.

This study draws on interviews and ethnographic observations conducted during the 2022 program of the *JournalismAI's Academy for Small Newsrooms* (from here on referred to as the Academy), a training programme aimed to "help small newsrooms leverage the power of artificial intelligence" (JournalismAI, 2023). Funded by the Google News Initiative (GNI), the Academy is situated within larger economic interrelationships between the news industry and tech corporations (Hermida & Young, 2023; Meese & Hurcombe, 2021; Nechushtai, 2018; Nielsen & Ganter, 2022). That is, while the educational content of the Academy was designed by the program managers and instructors with no direct involvement of Google (Interview, 2023),

the Academy can still be seen as an extension of platform activities aimed at producing dependencies between media and the tech corporations, such as via funding (Hermida & Young, 2023; Papaevangelou, 2023) or infrastructural dependencies (Nechushtai, 2018; Simon, 2022), as observed by Luchs et al. (2023). To contextualize these processes, the study draws on the concept of "macro epistemics" — the larger structures of actors that shape epistemic cultures in different fields (Knorr-Cetina, 2007).

Mapping hierarchies within the macro-epistemics of journalism and AI, the study asks whether training programs can contribute to the mitigation or amplification of power asymmetries between resource rich and smaller news organizations or possibly even reduce journalism's dependence on platform businesses. The study contributes to the field by identifying training initiatives as important actors in furthering an epistemic culture that advance a specific AI paradigm in journalism. While the Academy strives for the technological emancipation of news organizations and values critical use of AI that is aligned with journalistic ideals, the analysis shows those aims often stand in conflict with the ubiquity of the promoted tools and the directives legitimized by this paradigm (Luchs et al., 2023). Finally, we discuss the role of training programs in the formation of a global AI network, and possibilities for practitioners to form a coalition that supports small newsrooms and challenges the growing power asymmetries in the field.

Studying Journalistic AI Training Programs

Previous literature shows that news work increasingly demands technological skills (Koivula et al., 2023; Min & Fink, 2021; Royal & Kiesow, 2021). Relatedly, research on journalism innovation often suggests that news organizations are not equipped to innovate or that journalists are unwilling to adapt to technological changes (see Koivula et al., 2022). However, recent studies demonstrate that journalists value technological competencies and seek out learning opportunities (Koivula et al., 2023), reaffirming journalists' "strong capacity for adaptation and mitigation of new technology" (Linden, 2017, p. 136). Despite this, little empirical attention has been paid to AI training initiatives as a site of study. Prior studies on technologically oriented

courses are predominantly limited to student experiences and curriculars of journalism education (e.g. Kothari & Hickerson, 2020; Nguyen & Lugo-Ocando, 2016). The training or upskilling of professional journalists, however, remains "a strikingly under-researched area" (Porcu, 2020, p. 1556). While online courses are only a small part of journalism's engagement with AI, studying them can offer insights into the constellation of actors that are involved in the "use of and orientation toward artificial intelligence in newswork" (Lin & Lewis, 2022, p. 1629).

The Academy differs from the general courses on AI, explored in Luchs et al. (2023) and Heuer et al. (2021), as the course is not designed to teach the participants how to design AI applications, but focuses on giving the fellows an overarching understanding of what AI tools and practices are relevant for journalism, how to approach and negotiate potential risks, and how to implement an AI project. That is, studying the Academy gives insights into what understandings of AI are being constructed within journalistic contexts, what epistemic practices are valued and promoted (Tichenor et al., 2022), and how they shape journalistic practice. Further, the Academy not only provides training resources online but is a synchronous program that attracts professionals around the globe with a common interest in AI, and understands itself as a community (Interview, 2023). It connects journalists that have very limited computational skills with technologically skilled professionals in the news industry. The latter group can be considered pioneers of journalistic transformation (Hepp & Loosen, 2021) that carry new types of epistemic practices, beliefs, and motivations into journalism (Lischka et al., 2022). As such, the program fosters interdisciplinary exchanges and provide organizational learning opportunities through external expertise. A core aim of the training initiative is to minimise the growing competitional AI gap between large, resource rich and small newsroom around the globe (Beckett, 2019; Rinehart & Kung, 2022). To gain a better understanding of the actors and structures in which the Academy is embedded, we use a framework of macro-epistemics (Knorr-Cetina, 2007).

Macro Epistemics

Much work at the intersection of journalism and AI focuses on the level of epistemic cultures and how they are transformed by technological developments, e.g., via the introduction of new

practices (Ekström et al., 2022; Matheson, 2004), by examining tensions in epistemic values and practices (Ekström et al., 2022; Parasie, 2015), or by interrogating epistemic authority (Carlson, 2017; Usher, 2020; Wahl-Jorgensen, 2022). Less attention has been paid to how these changes are embedded in larger, macro-epistemic structures. Macro-epistemics describe "wider networks of knowledge generation" (Knorr-Cetina, 2007, p. 361). At the core of these networks are macro-epistemic actors that drive specific forms of knowledge constructions, provide ways of validating knowledge, and set the standards for knowledge-based practices (Knorr-Cetina, 2007). Those actors are often higher education institutions, research organizations, tech corporations, and what can be understood as intermediary or hybrid institutions, such as innovation centers, technical communities, or business and university associations (ibid.). Intermediary actors play an elemental role in developing epistemic cultures by facilitating both formal and informal links of collaboration between different types of epistemic actors (ibid.). Analysing how pioneer journalism communities reimagine journalistic epistemology, Anderson (2023), for instance, has demonstrated how pioneer communities extend their traditional epistemic role as facilitators of knowledge and adopt a semi-political, knowledge-based advocacy role.

Tichenor et al. (2022) conceptualizes relations between such actors as interlinkages in epistemic infrastructures. Drawing on their conceptualization (ibid.), we understand epistemics as complex, multi-dimensional processes that involve a diverse set of *materialities* (such as data and software), *interlinkages* (such as networks and communities), and the emergence and circulation of *paradigms* (such as new *ways of doing*). Accordingly, we understand knowledge to not only flow down from macro-epistemic actors to individual organizations. Instead, we regard knowledge as formed by a web of relations, processes, and practices, and the negotiation between intra- and inter-organizational relationships (ibid). These networks are not static or passive but are ever evolving through "constant making and remaking of partnerships, connections, and evidentiary claims" (Tichenor et al., 2022, p. 7), in which interdependencies are built between actors from different positionalities. These processes can be understood through two, distinct but interacting domains — an organizational domain and an interpersonal domain. That is, individuals are not anonymous actors but people with personal backgrounds, individual skill sets, perspectives, experiences, motivations, and beliefs, and that epistemic practices are shaped by the relationships between them. Epistemic processes neither operate in a vacuum nor

on a level playing field. Instead, epistemic systems can replicate or reinforce other systems (e.g., economic systems) and their associated power structures (Knorr-Cetina, 2007, p. 368). With this in mind, we argue that there is a need to interrogate the larger economic, cultural, and political contexts in which macro epistemics function, to understand how interdependencies and inequities between actors are structured. To do so, this study integrates Fraser's theory of justice (2010), as a way to analyze the dimensions of inequity that are shaped within macro-epistemic infrastructures.

Macro epistemics of AI in Journalism and Dimensions of Inequity

Fraser's (2010) theory of justice distinguishes between three dimensions through which power asymmetries occur — economically, through *maldistribution*, culturally, through *misrecognition*, and politically, through *misrepresentation*. Concerned with maldistribution, or the unequal distribution of resources, previous research has examined the economic nature of asymmetries between tech corporations and news organizations and argued that tech corporations leverage infrastructures to further consolidate their position of power, for economic profit, and often at the expense of journalistic autonomy (Nechushtai, 2018; Simon, 2022). Maldistribution, here, manifests itself in their technological control over AI infrastructures, including hardware, software, abilities to hire leading AI professionals, and the means to economically benefit from AI (Dyer-Witheford et al., 2019, p. 43). Prior research further indicates that AI courses can reinforce the positions of power of the companies that fund them (Luchs et al., 2023). As mentioned earlier, maldistribution also occurs between smaller newsrooms and large news organizations that have the economic means to foster computational capacities (Beckett, 2019; Rinehart and Kung, 2022).

On a cultural level, the production of hierarchies and unequal status may produce *misrecognition* (Fraser, 2010). In the context of journalism and AI, valuing and prioritizing AI practices over other epistemic practices may risk the promotion of "data-intensive and positivistic" (Iliadis & Russo, 2016) journalistic culture, data-induced ignorance (Schwarzkopf, 2020) or journalistic datasolutionism (Study IV). Put differently, misrecognition may result in

over-emphasizing the importance and possibilities of data-intensive AI practices. Lastly, political matters of *misrepresentation* structure whose needs are considered worthy of attention, who can make justice claims, and how such claims are adjudicated (Fraser, 2010), as epistemic paradigms become constructed and legitimized (Tichenor et al, 2022). AI adoption in journalism and evokes various ethical and professional questions (Dörr & Hollnbuchner, 2017; Helberger et al., 2022; Monti, 2018; Paik, 2023), and challenges to journalistic ideals such as the pursuit of critical rationality, accountability, and transparency (Waisbord, 2013, p. 225), which make matters or representation relevant. Relatedly, previous research has cautioned that AI adoption in journalism is so far mostly driven by technocentric ideas and "random acts of innovation", rather than a long-term strategy that is guided by a clear understanding of how AI can contribute to the societal role of journalism (Helberger et al., 2022; Posetti, 2018, p. 7; Thurman et al., 2019). To arrive at a vision similar to what these authors describe, we need to address a more fundamental issue — the power asymmetries embedded in the field of journalism and AI, that enable certain forms of journalism, while constraining others. With a nuanced understanding of these conditions, our view on what is needed for a more equitable, socially responsible integration of AI in journalism will be clearer.

Methodology

The paper uses the *Journalism AI Academy for Small Newsrooms* as a case for studying macro epistemics of the field of journalism and AI. The Academy is part of the global initiative *JournalismAI*, which is situated at the London School of Economics (LSE) and funded through the Google News Initiative. Empirically, the study builds on ethnographic observations (Spradley, 1997), conducted during the online training, and combines them with semi-structured interviews.

Case

The Academy was organized in eight themes with two weekly sessions for two cohorts, one for the Americas and one for the APAC region. The first session is centered around a lecture from the instructor(s) on a given topic. The second session is a workshop format during which

participants would work together on an assignment prepared by the instructor(s). During the first week practicalities of the course were explained, and instructors and participants were introduced to each other. The following six weeks were organized under the following themes: 'Key concepts and definitions', 'Applications of AI in journalism', 'Data and Technology', 'People and Culture', 'Risks and Concerns', and 'From Ideation to Implementation'. The themes were the same across the two cohorts, but the instructors differed, due to time zone considerations, but also to include instructors from the local contexts who could better address local concerns voiced by participants of the given cohort. In the final, eighth week, participants presented the ideas they had been working on during the course and received feedback from the two program managers.

Sample

Each of the authors participated in, and observed, the two weekly 1,5h long online training sessions in one of the two cohorts, respectively in the Americas and Asia-Pacific (APAC) region. This amounted to a total of 32 sessions that were conducted from September until the end of October 2022. The two cohorts included a total of 41 participants working for 41 news organizations that are situated in 23 countries. For a detailed breakdown see Appendix 1. The sessions were organized by two program managers from the Academy and instructed by 13 experts working at the intersection of journalism and AI, either in the industry or as part of the research community (including the Guardian, the Markup, the Wall Street Journal, the Associated Press). The observational data was complemented with semi-structured interviews (n=18) with the program managers, participants, and instructors, carried out after the Academy ended (at the end of 2022 and beginning of 2023). The selection of interviewees followed a convenience sampling, however gender balance (male n=9, female n=9), and diversity in geographic location was ensured (countries n=11). Please note that we refrain from revealing further information about interviewees to protect anonymity.

Material

Fieldnotes were produced during the observations by both authors. Recordings of the online sessions, slide decks, and further training materials (such as exercise sheets) were also accessible and reviewed again by the authors during the analysis stage based on the initial notes. All observations of the Academy were carried out under the Chatham House Rules, meaning

participants are referenced without identifying the speaker. The interview guide was based on insights gained in the observational part of the study, following up on identified patterns of interest. The topics included participants' initial motivations to join the academy, perceptions of the possibilities of upskilling, as well as ethical challenges regarding the use of AI, and strategies to mitigate platform dependency. The data (fieldnotes, audio files, and transcriptions of interviews) were stored on an encrypted drive to ensure confidentiality and consent was obtained from the participants, following GDPR standards.

Analysis

All empirical data was iteratively and inductively analyzed by both authors to discover and verify themes, using thematic analysis method following Braun and Clarke (2006). The process of thematic analysis involved several stages, including familiarization with the data, coding, and identification of themes, reviewing and refining themes, and finally, drawing conclusions and interpretations from the themes. Familiarization with the data involved reviewing fieldnotes, repeated readings of the interview transcripts, and rewatching the recordings of the sessions when clarification was needed. This stage also involved sorting and organizing the data into categories based on their content. Coding was done inductively, and the codes were used to identify and extract key themes from the data. Once the themes were identified, they were reviewed and refined to ensure that they accurately reflected the data. This involved checking the themes against the original data to ensure that they were supported empirically. The thematic analysis revealed a hierarchy and associated power asymmetries between the following epistemic actors through their interlinkages: (1) Big Tech, (2) large, resource rich news organizations, (3) small newsrooms, and lastly, (4) intermediary networks that were envisioned to mitigate said power asymmetries. We identified the Academy as an intermediary actor in the macro epistemics of journalism and AI, and lastly, discerned epistemic paradigms that emerged through knowledge constructions, validations, and standard settings.

Findings

The first section addresses an overarching epistemic paradigm that promotes and validates AI practices in the industry. Sequentially, we map hierarchies and associated inequities as experienced and perceived by fellows and instructors, and lastly, we highlight how epistemic paradigms, and their linked practices may stand in conflict with journalistic ideals and how news professionals respond to this.

A Push for AI

Over the course of the program and in the following interviews, fellows and instructors expressed that they perceive a push for AI that is exhilarating in growth and speed and is inevitably affecting the entire news industry. Participants describe the need to "embrace this change" to "remain in the game of newsgathering, dissemination and be relevant to their audiences", as the "the old guard is going to be demolished with everything that we do around us" (PAR14). Implementing AI across entire organizations was described as "just the future", as something that news professionals "have to do" (PAR 8). They expressed a need to match "all the changes which keep coming up", for which the fellows turned to training initiatives: "We keep attending all these workshops and trying to know more about how all these [things that] are changing or evolving." (PAR11). Fellows hoped that acquiring AI knowledge would enable them to "leverage [their] skills and have more exponential productivity." (PAR4) But towards which ends this growth in productivity would be leveraged was not entirely clear: "I mean, it surely will make our jobs more efficient, and we can take more holidays, probably, or not, I don't know" (PAR14).

Notably, participants explained this push as an "external drive", promoted by the tech community, and Big Tech in particular: "It's not journalists themselves who are driving this force. It's really more about developers and software companies and yeah, maybe big organizations with the resources to do this" (PAR7). This "driving force" was particularly tangible through funding infrastructures that reinforce inequities: "[It's] also a matter of opportunities available for organizations such as grants and journalism projects that will increasingly involve more machine learning and AI. And many of those organizations won't be

able to apply because they don't have the capabilities to do those sorts of projects" (PAR7). Similarly, another participant observed: "I always see some festivals from journalism in the region, in Colombia, in Mexico, in Argentina. And the winners usually lately always have some use of data combined with a big story" (PAR6). As part of this drive for AI, the Academy was described to facilitate a cultural change in journalism. As one instructor explained: "I think [the Academy is] a very good thing because it's more about awareness and culture change, basically, rather than building a data science team from scratch, like hiring 20 data scientists and doing something crazy" (INS1). Fellows echoed this sentiment: "there was a change in perspective [...] The biggest takeaway I had was, you should think about the issues that you currently have and consider if AI is something that can help you" (PAR10). With this new perspective, fellows were then expected to further foster this culture in their own organizations. They "should feel like an ambassador" (INS2, Fieldnotes) or 'flag bearers' (INS1, Fieldnotes) and help their newsroom adapt to these techno-cultural changes. This notion was accepted by participants: "it's my work, or I would like it to be my work, to help these people [other professionals at their organization] understand and be a part of building AI tools" (PAR9).

That is, the described "push" for AI was an initial motivating factor for fellows to participate in the Academy, which was then further validated in the Academy. Promoting a cultural shift in journalism, the Academy affirmed participants' perceptions that journalists need to adapt to technological change, by fostering a culture that promotes AI practices in journalism. In this context, power differentials between epistemic actors, as mapped out below, became apparent.

Hierarchies in Journalism and AI

Big Tech corporations

Fellows of the Academy largely described Big Tech corporations as aspirational, due to their epistemic power in terms of tech skills, talent, epistemic materialities, and as a result, the means to economically benefit from AI: "it's always better to be a tech company than a anything else

company" (PAR8). For small newsrooms, the resulting power differentials are particularly stark: "We are a small newsroom for us, it's almost like somebody from, let's say a kid from a village just got to see the huge city with all the skyscrapers. That's the feeling we all get when we visit all these places, because your tech is at some other level, Google's tech is at some other level." (PAR11). In line with this, fellows discussed their reliance on AI tools provided by Big Tech: "The concentration of these tools is in the hands of large tech companies. [...] That's the reality. I don't think we'll be able to develop anything in house, so we'll have to shop around" (PAR7). Acquiring epistemic power through AI adoption was presented as a pathway to mitigate these asymmetries, at least to some degree. As one fellow argued: "The media needs to adopt AI and other technologies and start thinking of itself as a product and a platform and content creators rather than going to these organizations and relying on them entirely." (PAR8).

However, not everyone thought bridging this epistemic gap was possible: "It's a more fundamental problem because we are not big tech companies. We are publishers and we struggle to even store the data, to clean the data, to make sure that our data is in good shape, etc. And obviously, when you have these kinds of struggles, you struggle to build any proper data solutions. And these big companies, they have this monopoly so they [...] are much better positioned to build these models and solutions." (INS1) That is, technological self-sufficiency was generally not considered feasible. While fellows and instructors acknowledged a need to gain epistemic agency through AI adoption, reliance on Big Tech, as the most powerful epistemic actor in the field of journalism and AI, was deemed inevitable.

Large, resource rich news organizations

Fellows described large, resource rich news organizations, such as the New York Times, as journalism' pioneers in AI, thanks to their economic and technological capacities to try out tools and lead a cultural change in the profession: "The big names, the big guns dominate. Journalists by the New York Times doesn't come to us to tell them what they can do with AI, obviously we learn from them." (INS5) However, instructors argued that being part of a large organization comes with its own challenges, making it sometimes easier for small newsrooms to experiment

with AI tools: “You will literally just read a random article that AP, Reuters, New York Times and BBC were doing these super cool things and it just sounded so futuristic [...] Now you have examples all over the world of smaller newsrooms and smaller outlets doing very cool things and showing the way and showing how being small and more nimble sometimes allows you to do things that the big names struggled to do just because they are big machines or they travel [through] different departments and or the bureaucracy that comes with it and so on” (INS5). Despite this, instructors expressed that it is unrealistic for small newsrooms to develop AI tools from scratch: “I’m very aware that [small] newsrooms probably will never be able to build a data science team and build their own models and do something in this space.” (INS 1). “[For that] you need to be a really large and well-funded newsroom” (INS3). Because of that, news professionals working for small newsrooms stated that large organizations have a specific responsibility in the field. That is, fellows argued that the industry’s pioneers should leverage their position to become powerful epistemic agents in their own right and mitigate dependencies: “Google Pinpoint is super helpful, but I would like to live in a world where The New York Times made that instead of Google [...] If The New York Times had developed a transcription service and then they charged us a few dollars a month to use it, I think my newsroom would sign up for it. We trust the New York Times. (PAR8)”.

Small newsrooms

Despite their limited economic capacities, fellows argued that AI implementation was an important goal for their newsroom: “The intention is that there’s going to be a ton of resources and money being fed into getting the AI game right.” (PAR14). This was not only motivated by the possibilities of AI. Highlighting both inequities between smaller and bigger newsrooms as well as geographic inequities, one fellow explained that they were partially motivated by fear: “it’s a threat for small newsrooms and newsrooms in other developed countries that if they don’t jump on this train, they will be left behind” (PAR7). Because small newsrooms lack technologically skilled staff, fellows were advised to use off-the-shelf tools, collaborate, or work with students (Fieldnotes, 2022). Instead of teaching participants how to code, instructors aimed to help fellows navigate epistemic questions, both addressing internal questions regarding their own organization, as well as questions about external partners such as third-party vendors, such

as: “How will I know if I can trust this third-party vendor? How will I know what conversations to have in my newsroom? Whether we’re comfortable with [a specific AI tool] or not?” (INS3). However, participants still worried that using third party tools present particularly high risks for small newsrooms:

> “A site might get shut down and we are done, because there’s nothing to negotiate when such things happen. If a server gets banned, there’s no one to negotiate [with] and the company cannot wait for a settlement [...] All the small newsrooms, they all are aware of that. Even one mistake and we won’t have a chance to save the company. Which is also one more reason people are skeptical of AI. [...] because we can’t legally or even properly verify the ethical nature of all the tools. Where is it pulling information from? Is it the information which is being pulled? Does it take all the agreements or all the disclaimers which needs to be crossed? Or maybe it’s overstepping. So somehow the tool gets banned, and all the sites use the tool, get penalized, and we won’t even know where to go”. (PAR11)

Relying on off-the shelf AI tools was thus not considered a “one size fits all” solution to adapt to technological advancement. Instead, their use presents severe challenges for small newsrooms who are not only ill-equipped when it comes to technological capacities, but also lack the legal and economic resources to shoulder the consequences of unintended malpractice.

Intermediary Networks - A JournalsimAI Coalition

In response to these challenges, fellows and instructors emphasized the importance of an epistemic network that could support small newsrooms. This network was envisioned as a space to “share knowledge related to AI” (PAR12), and foster collaboration among participants. As one fellow expressed, the Academy gave fellows access to such a support system: “I found that now we [the fellows] have a network. I mean, I know you, I know [Program Manager1], I know [Program Manager2]. I am now a part of the group” (PAR12). This “group” had been carefully curated by the program managers: “We’ve been very intentional in making it as varied and as

global and diverse as we can [...] I've had calls with all the shortlisted participants to actually meet the person behind the application and getting to know them and getting a deeper understanding of why and how they can bring something to the academy [...] You want to select the best candidates, but also you want to look at the cohorts and make sure that it is diverse enough in terms of gender, in terms of race, in terms of the countries represented" (INS4). Networking within the Academy was not limited to horizontal connections among participants but also included vertical connections between different newsrooms. This enabled participants to realize that newsrooms facing similar challenges in different parts of the world could learn from one another, as one program manager explained: "[...] a lot of time mitigating the inequalities is also allowing the small newsroom from Colombia to realize that a small newsroom from Kenya is struggling with the same challenges and they can learn together" (INS5).

As such, the Academy acted as a trusted intermediary, facilitating the establishment of this network. Its association with its academic institution, LSE, added credibility and made it a worthwhile investment for participants: "It was LSE, this is for the lack of a better word, like a trusted brand. It will not be a waste of my energy and time" (PAR13). The program managers' personal relationships with fellows and instructors played a crucial role in building trust. Over the years, these relationships had developed, which influenced the instructors' willingness to participate in the Google-sponsored journalism training. As one elaborated: "I would 100% not have done a Google sponsored journalism [training], if I hadn't seen the kind of intention they were bringing to the program and the people they were bringing to it. And like, I saw it over the years." (INS2)

Connecting fellows to instructors was thought to make "the training and their learnings more sustainable across time" (INS4). Further, collaboration was encouraged: "That's one thing I really liked about the Academy, there was no talk about competition, right? Like it was all about collaboration." (PAR3) That is, the academy fostered links between newsrooms, and different epistemic communities to "build bridges" between editorial staff and the tech community (PAR3).

The diverse community of journalists from around the world brought together by the Academy was considered vital for discussions on tech accountability and reporting. As one

instructor argued: “I think it was great to have such a broad coalition of journalists all over the world in the same room discussing these things. I think that’s important in general for tech accountability or tech reporting, because so much of this stuff happens in a multinational context” (INS2). As such, the Academy was envisioned as a facilitator for a greater coalition in journalism, that would not only leverage AI for journalistic pursuits, but also be able to hold power to account: “I think [the Academy] is a useful kind of node in the network. And I think you also need other nodes in the network, some that might be more adversarial, some that might be less adversarial [towards big tech corporations], some that might be looking at this more from a government regulation.” (INS2)

However, the reliance on Google funding for the training program raised concerns. The program managers specifically highlighted the potential biases that could arise when tech companies dictate the goals and direction of training programs: “Google is very excited to fund the Academy and has been able to say, through *JournalismAI*, we train these amount of journalists in the basics of AI, they are not as excited to give money to journalism to build this type of network because what do they get? [...] Their own biases in what their KPIs are will always influence what type of training programs exist and what they exist to do.” (INS5)

Accordingly, fellows and instructors not only expressed the need for an epistemic network for knowledge sharing and collaboration to advance AI practices in journalism, but to create independent spaces that are detached from tech funding. These spaces were also needed to jointly address the various ethical challenges that journalists encountered in their engagement with AI. How ethical challenges were negotiated is discussed further in what follows.

Epistemic paradigms competing with journalistic ideals

Fellows as well as instructors voiced concerns that AI practices would stand in conflict with news professionals’ social and moral responsibility to prioritize the public good over commercial, as well as core ethical ideals, such as the protection of traditionally marginalized communities. Recurring concerns were: 1) infringing on audiences’ privacy when using off the shelf-tools that require to share audiences’ data with third parties, 2) damaging audiences’ trust through “over-personalization”, and 3) unintentionally building biases into AI models that harm

communities — and traditionally marginalized communities, in particular. As one instructor put it: "[We need to] assure our audience that we are neutral, that we are reporting on facts, and that we have no agenda that we're trying to push. And the more you personalize, the more that you collect data on your users and deploy that data kind of against them, the more you could potentially feed into that narrative that undercuts the importance of media and our position as a neutral territory." (INS3)

Participants argued that their precarious position as a small newsroom forced them into epistemic practices that violated their professional norms. Giving Google access to audience data to use audience analytics, participants expressed, for example: "I'm not totally comfortable with it because I really believe that privacy should be something that we could also have in the digital space [...] but do we have some other option? [Alternatives] may be way too expensive and we don't have any choice" (PAR1). Journalists on one hand stated that they have a specific responsibility to "help define a standard" in AI practices (INS2). On the other hand, participants expressed resignation, reflecting the domination of an epistemic paradigm that overrides norms and pursuits of equity: "I feel like eventually there will be no such thing as data privacy because the technology will become more important". (PAR4) This resignation was fostered through engagement with AI over time: "I'm becoming more cynical about these types of things as I work with data and work with AI more, [it's] very uncomfortable." (INS3). However, not everyone was convinced that journalists lacked options. One instructor, for example, argued that collection of demographic data through tools, such as Google Analytics, was not necessary to serve audiences: "for marketing this data is absolutely gold, because it's much easier to go to a company and we have, I don't know, 18-, 25-year-olds, reading our content here. [...] But you don't need this. I know that because we had "[retracted]" and we indeed asked people for age and gender and even postcodes. And I personally have worked on a couple of models incorporating this data, trying to understand where we have gaps. But it didn't reveal anything audience research would reveal." (INS1) The instructor further argued that using demographic audience data for journalistic pursuits took up too many resources to comply with regulations: "It's a minefield, to be honest. So, I think that at some point, from a financial point of view, it was easier just to say we don't use this data" (INS1). This form of epistemic delinking not only brought other forms of knowledge generation — here qualitative audience research — to the

foreground, but further underscored the option to opt-out of certain epistemic practices. However, opting out or disengagement were rarely promoted in the Academy nor considered by the participants. Instead, many fellows stated that there would not be any AI tools or epistemic practices that they would, unequivocally resist, even if they conflicted with professional norms and ideals. Describing facial recognition as the "biggest invasion of privacy", one of the fellows contemplated: "But then what if it's the norm in the future? Like everyone is using it? [...] We are going to use it in everything, including in journalism. Are you going to say no to that?" (PAR4).

Paradoxically, instructors and participants also argued that "journalists have to be extra careful" regarding AI practices, given their responsibility towards society at large. That is, news professionals were concerned that, in comparison with tech companies, their practices could have "more serious repercussions." (PAR7). Journalists thus struggled to align their ideals with ethically challenging AI practices, which they often viewed as unavoidable or as an impending industry norm, resulting in inconsistent attitudes towards AI adoption.

Discussion

The macro epistemics in the field of journalism and AI, as narrated by fellows and instructors of the Academy, is characterized by power asymmetries. Participants describe stark hierarchies — with Big Tech corporations being the most influential macro-epistemic actors (Knorr-Cetina, 2007) thanks to their economic, cultural, and political capital. Fellows further described a gap between large, resource rich news organizations that have the computational capacities and human resources to leverage AI, and small newsrooms that are economically disadvantaged and culturally ill-equipped to adapt to technological developments (Beckett, 2019; Rinehart and Kung, 2022). The Academy was considered a pathway to mitigate these inequities. Connecting professionals working for small newsrooms with a low level of technological skills with AI experts was thought to be a first step for small newsrooms to understand and leverage AI technology. The Academy thus functioned as an intermediary actor in macro epistemics that facilitates both formal and informal links between different types of epistemic communities (Knorr-Cetina, 2007).

The program aimed to drive an AI-forward cultural change by advancing and affirming AI-based forms of knowledge creation. To further foster this culture in their respective newsrooms, fellows were expected to become AI "ambassadors". This echoes findings of Anderson's (2023) study, showing how pioneer communities extend their traditional epistemic role as facilitators of knowledge and adopt a knowledge-based advocacy role. Hence, the Academy can be regarded as an important driver that distributes and reinforces a specific epistemic culture and a paradigm that promotes and validates AI practices in journalism on a global scale.

These processes have consequences for dependencies between news organizations and platforms (Hermida & Young, 2023; Meese & Hurcombe, 2021; Nechushtai, 2018; Nielsen & Ganter, 2022). The promoted AI practices necessitate certain materialities, mainly free, off-the shelf AI tools provided by tech companies. That is, the Academy contributes to epistemic processes that deepen the news industry's asymmetrical power relationship with technology companies. This resonates with Luchs' et al. (2023) study that indicates that AI training programs offer opportunities for tech companies to consolidate and expand their position of power. Adding to Luchs et al. (2023), we argue that this is operationalized through the reinstatement of epistemic power and economic maldistribution. This is not to say that instructors or fellows were naive or uncritical of Big Tech. Rather, there was a tension between instructors' motivations to mitigate power differentials by emancipating small newsrooms through up-skilling, and the practices that they promoted (which required the use of Big Tech software and thus reinforced their position of power).

This was further complicated by the fact that AI practices discussed in the Academy often presented ethical challenges (Dörr & Hollnbuchner, 2017; Helberger et al., 2022; Monti, 2018; Paik, 2023) and that adoption of related epistemic practices stood in conflict with journalistic ideals. It should be noted here that the aim of this study was not to assess AI practices and sort them into warranted, productive, or ethical ways of knowledge generation. However, we want to highlight that the form of AI practices discussed at the Academy varied vastly, ranging from automation of audio transcription, to using AI to analyze racial bias in policing (Fieldnotes, 2022). This, in fact, is a feature of this epistemic paradigm, in which presented AI practices and their potential benefits — whether they be economical, ethical, or reputational — are so varied

and often untransparent, that they can easily be interpreted as solution and problem at the same time. The ability to uphold journalistic ideals, such as prioritizing the public good over commercial interests and safeguarding marginalized communities (Waisbord, 2013), was then contingent upon the economic resources of newsrooms. Driven by economic concerns (maldistribution), fellows expressed alignment with this epistemic AI paradigm even when it promoted misrecognition (Fraser, 2010) via data-intensive and positivistic practices (Iliadis & Russo, 2016), which ultimately contradicted professional ideals. This contradiction is what we offer to call *sticky epistemics.* We speak of sticky epistemics when epistemic practices are valued and prioritized over other forms of knowledge generation, and continue to be, even if they stand in conflict with professional ideals. Sticky epistemics are persistent, because they are supported by pre-existing power asymmetries — structured through maldistribution, misrecognition, and/or misrepresentation (Fraser, 2010). That is, sticky epistemics function through and reinforce inequity, as they reinstate epistemic power.

While these findings reflect that the Academy was a site where inequities deepen through the reinstatement of epistemic power, fellows and instructors also tried to use the Academy to challenge power asymmetries. They envisioned that a global coalition in the field of journalism and AI could challenge power asymmetries, especially in relation to Big Tech. The Academy was regarded as a steppingstone to create this coalition that is both epistemically diverse by design and joined by shared journalistic ideals. Here, different epistemic actors were considered to have different stakes and responsibilities. Newsrooms that, otherwise, due to their lack of computational staff and economic resources, felt they were missing out on technological developments, considered the training to be an important resource (Luchs et al., 2023), because it granted them access to a global network that helped them gain agency in the field of AI. They hoped these connections would last, so that newsrooms could support each other and their AI projects over time. Moreover, according to the fellows, large, resource rich news organizations have a responsibility to leverage their capacities — e.g., by building their own AI systems that can be shared with other news organizations, or simply by spending time and resources to help small newsrooms adapt to technological developments.

Relatedly, findings have underscored that epistemic structures are not only contextual and culturally specific, but socio-relational and deeply dependent on certain actors (Tichenor et al., 2022). That is, the efficacy of the Academy relies on the relationships between program managers, instructors, and participants. This is not unproblematic. While these relationships grant some newsrooms access to this network, only few professionals can benefit from initiatives such as the Academy. This raises the question who will be left behind, and at what cost? Secondly, the positive values that are embedded in the design of the Academy and fostered through the implementation of the training program are dependent on individuals. Given the precarious nature of the funding of these initiatives (Hermida & Young, 2023), it is unclear what will replace them if they lose funding.

Regarding the limitations of this study, we want to highlight that while the cross-country sample of instructors and fellows allowed for insights into a global structure, it may not capture the full diversity and variations within the APAC and Americas regions. Moreover, the sample only reflects the experiences and perspectives of individuals who are invested in the advancement of AI adoption in journalism. While we observed critical attitudes, a sample that would allow for individuals that are uninterested in or resistant towards AI practices might yield further nuances. The time span of data collection presents another limitation to the study. Future research that looks at macro epistemics from a longitudinal perspective is needed to assess the lasting impact of training courses and further contextualize how power asymmetries unfold over time. Observing how learnings and collaborations of fellows translates into lasting changes in their respective newsrooms would be specifically valuable.

Conclusion

Using a framework of macro epistemics, this study demonstrated that training programs, such as the *JournalismAI* Academy, are important drivers of macro epistemics in the field of journalism and AI. They facilitate the reinforcement of epistemic paradigms, which promote and validate AI practices in journalism, through institutional and interpersonal domains. Fellows and instructors identified clear hierarchies in the field with Big Tech corporations having most epistemic power,

followed by resource-rich news organizations, risking small news organizations to be left behind in addressing technological advancement. By connecting professionals from small newsrooms with low technological skills to AI experts, the Academy provides fellows access to cultural, political, and economic dimensions of epistemic power. That is, those most disadvantaged in this context are newsrooms outside of AI networks. The Academy was seen as a steppingstone toward creating a global coalition, that could foster epistemic agency for news professionals and challenge power asymmetries. However, this form of alliance is dependent on the willingness of resource-rich professionals, with computational capacities to invest in small newsrooms.

The study further highlights the contradictory nature between instructors' and fellows' aims to use AI for the mitigation of power asymmetries, and their willingness to fall back on AI practice that conflicted with professional ideals. Against this backdrop, this study offered the concept of *sticky epistemics*, in which certain epistemic practices continue to be valued and prioritized over others, even though they stand in conflict with professional norms and ideals of the given field. Sticky epistemics are persistent, because they are supported by pre-existing power asymmetries — structured through maldistribution, misrecognition, and/or misrepresentation. That is, sticky epistemics function through and reinforce inequity, as they reinstate epistemic power. The study underscores the need for journalism AI training initiatives that are independent from funding of Big Tech companies and foster epistemic diversity to critically mitigate competitive gaps and in doing so, journalistic resilience in a media environment that is characterized by inequity.

References

Anderson, B. (2023). “Join the Coalition”: How Pioneer Journalism Communities Reimagine Journalistic Epistemology from the Periphery. *Digital Journalism*, 1–22. https://doi.org/10.1080/21670811.2023.2278043

Beckett, C. (2019). *New powers, new responsibilities: A global survey of journalism and artificial intelligence.* [Polis, London School of Economics and Political Science]. https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Broussard, M., Diakopoulos, N., Guzman, A. L., Abebe, R., Dupagne, M., & Chuan, C.-H. (2019). Artificial Intelligence and Journalism. *Journalism & Mass Communication Quarterly*, *96*(3), 673–695. https://doi.org/10.1177/1077699019859901

Carlson, M. (2017). *Journalistic Authority: Legitimating News in the Digital Era*. Columbia University Press. https://doi.org/10.7312/carl17444

de-Lima-Santos, M.-F., & Ceron, W. (2021). Artificial Intelligence in News Media: Current Perceptions and Future Outlook. *Journalism and Media*, *3*(1), 13–26. https://doi.org/10.3390/journalmedia3010002

Dörr, K. N., & Hollnbuchner, K. (2017). Ethical Challenges of Algorithmic Journalism. *Digital Journalism*, *5*(4), 404–419. https://doi.org/10.1080/21670811.2016.1167612

Dyer-Witheford, N., Mikkola Kjøsen, A., & Steinhoff, J. (2019). *Inhuman power: Artificial intelligence and the future of capitalism*. Pluto Press.

Ekström, M., Ramsälv, A., & Westlund, O. (2022). Data-driven news work culture: Reconciling tensions in epistemic values and practices of news journalism. *Journalism*, *23*(4), 755–772. https://doi.org/10.1177/14648849211052419

Helberger, N., Van Drunen, M., Moeller, J., Vrijenhoek, S., & Eskens, S. (2022). Towards a Normative Perspective on Journalistic AI: Embracing the Messy Reality of Normative Ideals. *Digital Journalism*, *10*(10), 1605–1626. https://doi.org/10.1080/21670811.2022.2152195

Hepp, A., & Loosen, W. (2021). Pioneer journalism: Conceptualizing the role of pioneer journalists and pioneer communities in the organizational re-figuration of journalism. *Journalism*, *22*(3), 577–595. https://doi.org/10.1177/1464884919829277

Hermida, A., & Young, M. L. (2023). Google's Influence on Global Business Models in Journalism: An Analysis of Its Innovation Challenge. *Media and Communication*, *12*. https://doi.org/10.17645/mac.7562

Heuer, H., Jarke, J., & Breiter, A. (2021). Machine learning in tutorials – Universal applicability, underinformed application, and other misconceptions. *Big Data & Society*, *8*(1), 205395172110175. https://doi.org/10.1177/20539517211017593

Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, *3*(2), 205395171667423. https://doi.org/10.1177/2053951716674238

Jamil, S. (2021). Artificial Intelligence and Journalistic Practice: The Crossroads of Obstacles and Opportunities for the Pakistani Journalists. *Journalism Practice*, *15*(10), 1400–1422. https://doi.org/10.1080/17512786.2020.1788412

Knorr-Cetina, K. (2007). Culture in global knowledge societies: Knowledge cultures and epistemic cultures. *Interdisciplinary Science Reviews*, *32*(4), 361–375. https://doi.org/10.1179/030801807X163571

Koivula, M., Laaksonen, S.-M., & Villi, M. (2022). Practical, Not Radical: Examining Innovative Learning Culture in a Public Service Media Organization. *Journalism Studies*, *23*(9), 1018–1036. https://doi.org/10.1080/1461670X.2022.2065339

Koivula, M., Saari, T., & Villi, M. (2023). "I love learning new things": An institutional logics perspective on learning in professional journalism. *Journalism*, 146488492311677. https://doi.org/10.1177/14648849231167786

Kothari, A., & Hickerson, A. (2020). Challenges for journalism education in the era of automation. *Media Practice and Education*, *21*(3), 212–228. https://doi.org/10.1080/25741136.2020.1787724

Lin, B., & Lewis, S. C. (2022). The One Thing Journalistic AI Just Might Do for Democracy. *Digital Journalism*, *10*(10), 1627–1649. https://doi.org/10.1080/21670811.2022.2084131

Linden, C.-G. (2017). Decades of Automation in the Newsroom: Why are there still so many jobs in journalism? *Digital Journalism*, *5*(2), 123–140. https://doi.org/10.1080/21670811.2016.1160791

Lischka, J. A., Schaetz, N., & Oltersdorf, A.-L. (2022). Editorial Technologists as Engineers of Journalism's Future: Exploring the Professional Community of Computational Journalism. *Digital Journalism*, 1–19. https://doi.org/10.1080/21670811.2021.1995456

Luchs, I., Apprich, C., & Broersma, M. (2023). Learning machine learning: On the political economy of big tech's online AI courses. *Big Data & Society*, *10*(1), 205395172311538. https://doi.org/10.1177/20539517231153806

Mackenzie, A. (2017). *Machine Learners: Archaeology of a Data Practice*. The MIT Press. https://doi.org/10.7551/mitpress/10302.001.0001

Marconi, F. (2020). *Newsmakers: Artificial Intelligence and the Future of Journalism*. Columbia University Press. https://doi.org/10.7312/marc19136

Matheson, D. (2004). Weblogs and the Epistemology of the News: Some Trends in Online Journalism. *New Media & Society*, *6*(4), 443–468. https://doi.org/10.1177/146144804044329

Meese, J., & Hurcombe, E. (2021). Facebook, news media and platform dependency: The institutional impacts of news distribution on social platforms. *New Media & Society*, *23*(8), 2367–2384. https://doi.org/10.1177/1461444820926472

Min, S. J., & Fink, K. (2021). Keeping Up with the Technologies: Distressed Journalistic Labor in the Pursuit of "Shiny" Technologies. *Journalism Studies*, *22*(14), 1987–2004. https://doi.org/10.1080/1461670X.2021.1979425

Monti, M. (2018). Automated Journalism and Freedom of Information: Ethical and Juridical Problems Related to AI in the Press Field. *Opinio Juris in Comparetione*, *1*.

Munoriyarwa, A., Chiumbu, S., & Motsaathebe, G. (2023). Artificial Intelligence Practices in Everyday News Production: The Case of South Africa's Mainstream Newsrooms. *Journalism Practice*, *17*(7), 1374–1392. https://doi.org/10.1080/17512786.2021.1984976

Nechushtai, E. (2018). Could digital platforms capture the media through infrastructure? *Journalism*, *19*(8), 1043–1058. https://doi.org/10.1177/1464884917725163

Nguyen, A., & Lugo-Ocando, J. (2016). The state of data and statistics in journalism and journalism education: Issues and debates. *Journalism*, *17*(1), 3–17. https://doi.org/10.1177/1464884915593234

Nielsen, R. K., & Ganter, S. A. (2022). *The power of platforms: Shaping media and society*. Oxford University Press.

Paik, S. (2023). Journalism Ethics for the Algorithmic Era. *Digital Journalism*, 1–27. https://doi.org/10.1080/21670811.2023.2200195

Papaevangelou, C. (2023). Funding Intermediaries: Google and Facebook's Strategy to Capture Journalism. *Digital Journalism*, 1–22. https://doi.org/10.1080/21670811.2022.2155206

Parasie, S. (2015). Data-Driven Revelation?: Epistemological tensions in investigative journalism in the age of "big data." *Digital Journalism*, *3*(3), 364–380. https://doi.org/10.1080/21670811.2014.976408

Porcu, O. (2020). Exploring innovative learning culture in the newsroom. *Journalism*, *21*(10), 1556–1572. https://doi.org/10.1177/1464884917724596

Posetti, J. (2018). *Time to step away from the 'bright, shiny things'? Towards a sustainable model of journalism innovation in an era of perpetual change*. Reuters Insitute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-11/Posetti_Towards_a_Sustainable_model_of_Journalism_FINAL.pdf

Rinehart, A., & Kung, E. (2022). *Artificial Intelligence in Local News*. AP. https://www.ap.org/assets/files/ap_local_news_ai_report_march_2022.pdf

Royal, C., & Kiesow, D. (2021). From Boundary to Bridge and Beyond: The Path to Professionalization of Product Roles in Journalism. *Journalism Studies*, *22*(11), 1546–1565. https://doi.org/10.1080/1461670X.2021.1944277

Simon, F. M. (2022). Uneasy Bedfellows: AI in the News, Platform Companies and the Issue of Journalistic Autonomy. *Digital Journalism*, 1–23. https://doi.org/10.1080/21670811.2022.2063150

Soto-Sanfiel, M. T., Ibiti, A., Machado, M., Marín Ochoa, B. E., Mendoza Michilot, M., Rosell Arce, C. G., & Angulo-Brunet, A. (2022). In Search of the Global South: Assessing Attitudes of Latin American Journalists to Artificial Intelligence in Journalism. *Journalism Studies*, *23*(10), 1197–1224. https://doi.org/10.1080/1461670X.2022.2075786

Spradley, J. P. (1997). *The ethnographic interview*. Wadsworth group, Thomson learning.

Thurman, N., Lewis, S. C., & Kunert, J. (2019). Algorithms, Automation, and News. *Digital Journalism*, *7*(8), 980–992. https://doi.org/10.1080/21670811.2019.1685395

Tichenor, M., Merry, S. E., Grek, S., & Bandola-Gill, J. (2022). Global public policy in a quantified world: Sustainable Development Goals as epistemic infrastructures. *Policy and Society*, puac015. https://doi.org/10.1093/polsoc/puac015

Usher, N. (2020). News cartography and epistemic authority in the era of big data: Journalists as map-makers, map-users, and map-subjects. *New Media & Society*, *22*(2), 247–263. https://doi.org/10.1177/1461444819856909

Wahl-Jorgensen, K. (2022). Local Knowledge and Epistemic Authority in Entrepreneurial Journalism. *Digital Journalism*, 1–15. https://doi.org/10.1080/21670811.2022.2128388

Waisbord, S. R. (2013). *Reinventing professionalism: Journalism and news in global perspective*. Polity.

Appendix 1

Please note that professional titles are listed here as stated by the participants of the study.

Fellows of the Academy (n=41)

Professional Titles	Countries (n=23)	Gender (male n=21, female n=20)
Director of Development & Innovation	Cuba	Male
Performance Editor	Brazil	Male
CEO & Editor	Mexico	Female
Chief Editor	Paraguay	Male
Special Editor	Brazil	Male
Founder & Citizen Observer	US	Male
Editor	US	Male
Data Journalist	Argentina	Female
Audiovisual Director	Cuba	Male
Product Director	Brazil	Female
Executive Director	Unknown	Female
Journalist	Ecuador	Male

Analysis & Community Editor	Columbia	Female
Senior Data Reporter	US	Male
Chief Innovation Officer	US	Male
Product Manager	US	Female
CEO, Founder & Chief Editor	Venezuela	Female
Project Manager	Bolivia	Male
Editor	Colombia	Female
Journalist	El Salvador	Female
News Editor	Malaysia	Male
Data Journalist	New Zealand	Male
Reporter	Nepal	Female
Journalist	Mongolia	Female
Deputy Editor	India	Male
Head of Online Verification	Philippines	Female
Special Producer	Indonesia	Male
South Asia Editor	India	Female
Managing Editor	India	Male

Reporter and Digital Strategy	Korea	Female
Deputy Editor	Sri Lanka	Female
Co-Founder	Mongolia	Female
Co-Founder & Producer	Malaysia/Singapore	Female
Head of Tech, Digital & New Initiatives	India	Male
Head of Growth & Marketing	India	Male
Technology Journalist & Co-Founder	India	Female
Co-Founder & Editor	Sri Lanka	Male
Editor	India	Male
Chief Editor	India	Female
Head of Digital	India	Female
Editor-in-Chief	China	Male

Instructors (n=13) and Program Managers of the Academy (n=2)

Professional Roles	**Countries (n=8)**	**Gender (male n=6, female n=9)**
Program Manager	Italy	Female
Program Manager	UK	Male
Professor of Journalism and Communication/ Audience Editor	Brazil	Male
Program Manager	US	Female
Product Manager	US	Male
Senior Data Engineer	US	Male
Research and Training Manager	Argentina	Female
Director of Data Science	US	Female
Senior Director, Strategic Initiatives	US	Female
Vice President, Audience Growth	US	Female
PhD Candidate in Media and Communication Studies	Sweden	Female

Head of Insights, AI & Innovations	India	Male
Head of Data	China	Male
Head of the AI + Automation Lab	Germany	Female
Lead Data Scientist	UK	Female

Check for updates

Special Issue: Digital News Histories in Europe

Journalism
2023, Vol. 0(0) 1–19
© The Author(s) 2023

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/14648849231176110
journals.sagepub.com/home/jou


Solving journalism with data: Silicon Valley's influence on the Fourth Estate

Nadja Schaetz
Universität Hamburg, Journalism and Communication Studies, Hamburg, Germany

Laura Laugwitz
Universität Hamburg, Journalism and Communication Studies, Hamburg, Germany

Juliane A Lischka
Universität Hamburg, Journalism and Communication Studies, Hamburg, Germany

Abstract

The historically ambiguous relationship between journalism and Big Tech can be traced back to the late 20th century, when news organizations started to recognize the potential of the Internet as a distribution platform. While a growing body of literature is concerned with power asymmetries between Big Tech and journalism, the role of *place* in shaping the layered histories of journalism remains underexplored. This study uses a framework of place—conceptualized as material and geographic, a setting for action and lived experiences carrying the legacies of their past, and a site of accumulating histories of cultural meaning and power (Usher, 2019)—to examine how Silicon Valley and Fourth Estate ideals converge. Empirically, the study analyzes job advertisements of four US and UK print-legacy news outlets serving as a window into shifting expectations, skills, and values that news organizations seek in their employees, reflecting broader trends in journalism. Findings show that journalism draws on Silicon Valley ideals, merging datasolutionism with Fourth Estate narratives of audience access. Some news organizations are not only tech-oriented but frame themselves as tech companies with a Fourth Estate mission. While both Silicon Valley and Fourth Estate narratives promote ideals of equal power distribution, findings indicate the reinforcement of hegemonic power structures in the news industry. We conclude that the influence of Silicon Valley on journalism is one of consolidating power through location, action, and cultural meaning, as news organizations construct datasolutionism as a pivotal novel layer to achieve long-standing Fourth Estate

Corresponding author:
Nadja Schaetz, Universität Hamburg, Journalism and Communication Studies, Hamburg 20148, Germany.
Email: nadja.schaetz@uni-hamburg.de

ideals. This analysis contributes to our understanding of the historical context and evolving nature of the relationship between journalism and Big Tech, highlighting the significance of place in shaping the dynamics between these two increasingly intertwined industries.

Keywords
Silicon Valley, Fourth Estate, datafication, journalism, place, ideals, inequality

Introduction

Journalism and Silicon Valley have had an ambiguous friend-and-foe relationship. As platforms have gained power over news distribution and audience reach through their channels and dominated the digital advertising industry's underlying infrastructure, news organizations have become increasingly commercially reliant on Silicon Valley (Nielsen and Ganter, 2022). Platform companies distanced themselves from being editorial organizations (Napoli and Caplan, 2017) and have provided journalism with services, technological training, and innovation funds (Papaevangelou, 2023; Simon, 2022), while news organizations have expanded their technologically skilled staff and embraced a more tech-oriented culture in newsrooms (Lischka et al., 2022). As a result, platforms and news organizations have built complex connections that are characterized by interdependencies and power asymmetries, with well-resourced outlets being better positioned to mitigate their dependencies on platforms than smaller newsrooms (Meese and Hurcombe, 2021; Myllylahti, 2021).

Silicon Valley and journalism are historically grounded in different ideals. Silicon Valley's technosolutionism, as part of the "Californian ideology" (Ferrari, 2020), and journalism's Fourth Estate (Hampton, 2010) reach back to the 1960s in the US and the 18th century in Europe and the US, respectively. Despite the fact that news organizations aim to resist Silicon Valley's impact (Lindblom et al., 2022), they have become increasingly entangled with Silicon Valley platform companies. This might also involve a convergence of both ideals that can be traced over time.

Measured over too short a span, change becomes indiscernible and temporary conditions are perceived as permanent. Solnit compares this to navigating your way on a phone. Looking at a map "you see either a picture too small to provide detail or details too close up for context," or you blindly follow algorithmic directions, "and never fully grasp where you are."[1] In this article, we aim to zoom out by grounding our analysis in a cultural historical approach, and zoom in by drawing on contemporary empirical data. Similarly to Solnit, we think through change via place. Drawing on Usher's theoretical framework of place in journalism (2019), we connect place and inequity to the changes the industry has faced and ask, *How do Silicon Valley and Fourth Estate ideals converge in contemporary journalistic places*? to understand how Silicon Valley is shaping practice, cultural meaning, and location in journalism.

Specifically, we qualitatively analyze $n = 55$ job advertisements of four high-status Fourth Estate legacy news outlets in the US and UK (*The News York Times*, *Washington Post*; *The Guardian*, *The Daily Mail/Sunday Mail*), two countries with similar roots in the Fourth Estate but differing in their geographical proximity to Silicon Valley. Job ads not only offer insights into journalistic practice, but they give practices "an institutional and physical form" (D'Souza et al., 2023), signal organizational identity as a site of cultural meaning and power, and bind them to physical manifestations and geographic locations. This place-based approach helps to display power asymmetries that are both material as well as socio-cultural, and "limit and enable particular kinds of journalism" (Usher, 2021: p. 13). Analyzing job ads can serve as a window into the shifting expectations, skills, and values that news organizations seek in their workforce, mirroring wider trends in journalism (Lindén and Grönlund, 2021). The study contributes to the field by identifying manifestations of Silicon Valley and Fourth Estate ideals that are historically embedded and shape contemporary journalism through their convergence. Demonstrating that news organizations construct datasolutionism as pivotal to achieve Fourth Estate ideals, the study furthers the field by highlighting how existing power asymmetries in journalism are reinforced.

Theory

In the following sections we employ the theoretical concept of *place,* which is historically rooted in the sense that places carry the weight of their past, accumulating layers of history, culture, and meaning over time (Massey, 1995). We show how Silicon Valley encroaches on three conceptual dimensions of place in journalism as offered by Usher (2019)—place as a site of cultural meaning and power, as the geographic and material setting of journalism, and as lived, where action and lived experiences are made—concluding that Silicon Valley shapes journalism with consequences for inequity.

Place as a site of cultural meaning and power

Place conceived as a site of cultural meaning and power situates journalism in its intersection with resources, economic imperatives, flows of capital and cultural influence in a historical context (Usher, 2021). In what follows we first trace a historical account of Silicon Valley's ideals, followed by journalism's conception as the Fourth Estate, to examine their production of meaning and power.

Silicon Valley. Geographically situated in the southern San Francisco Bay Area, Silicon Valley, subsumes a mixture of values, promises and concerns (Jones and Sudlow, 2022). As society's resistance to military endeavors gained traction in the 1960s, engineers in California became increasingly resistant to military sponsorship of their academic research (Burrell and Fourcade, 2021). Hippies and hackers found each other to co-create a "techno-utopian social vision" opposing the Cold War's "unfeeling bureaucracy, a calculating, mechanical form of social organization that had brought humankind to the edge of nuclear annihilation" (Turner, 2006: p. 237). They created a counterculture of

collaborative workstyles, hedonistic lifestyles, and a commitment to the circulation of information (Barbrook, 1998; Burrell and Fourcade, 2021; Turner, 2006). Over time, these aspirations were incorporated into neo-liberal politics and further evolved into technosolutionism, the idea that social problems can and should be solved by technology, not policy (Barbrook, 1998; Morozov, 2014). Disruptive innovation, then, became Silicon Valley's narrative to mark both the necessity as well as the inevitability of a technological revolution (Geiger, 2020). And as these companies gained authority (Moore and Tambini, 2018), the rhetoric changed once again: "While the Californian Ideology of the 1990s was an insurgent message, Silicon Valley's technological imaginary of the 2010s is a discourse which legitimizes an established dominant industry, capable of influencing political actors" (Ferrari, 2020). This shift is also illustrated by who has access to the social, cultural and economic capital of Silicon Valley. Prestigious and well-paid positions are overwhelmingly filled by white economically well situated men, whereas historically marginalized people, especially Black people and People of Color, are mainly found in lower-paid service jobs (Tomaskovic-Devey and JooHee, 2018). While dominant discourses about Silicon Valley are idealistic, its actualization is embedded in the same power structures it once aimed to counter (Couldry and Mejias, 2019).

Journalism and the Fourth Estate. The idea of an independent press representing a fourth power besides legislative, judicative, and executive powers formed its roots during the Glorious Revolution, the French Revolution, and the War of Independence in the UK, France, and the United States, respectively (Burrowes, 2011). The independent press "emerged out of a social rupture that resulted in the reorganization of state power" (Burrowes, 2011: p. 52) and was bound to the ideals of a democratic state. In the reorganized state, the British press had permission to publish reports on parliamentary proceedings in the second half of the 18th century (ibid.). The Fourth Estate first represented the Reporter's Gallery, a physical place indicating where reporters sat in the House of Commons. Therefore, the press not only connected political elites to their constituencies but determined the message. This control over discourse has remained a central part of journalists' professional identity (Sherwood and O'Donnell, 2018).

To maximize readership, news outlets positioned themselves beyond a specific party ideology and thus independent from party funding (Burrowes, 2011). Enabled by this political independence, the Fourth Estate represented a commitment to hold those in power accountable, facilitate public debate, represent the public opinion, and link those in power to civic society (Schultz, 1998). These ideals have revived in the digital age, when journalism is expected to facilitate a responsible public debate and enable citizens to participate in democratic processes (Tumber, 2001). Thus, the Fourth Estate is meant to compensate for misconduct in other fields and eventually to improve society. At the same time, in relation to society, the ideals of a Fourth Estate are a source of institutional legitimacy for journalism (Sherwood and O'Donnell, 2018).

Yet, the Fourth Estate is no mere imagination. Examples of award-winning investigative journalism indicate an ability to fulfill journalism's ideals (Hamilton, 2016), as demonstrated by award-winning news outlets such as *The New York Times* and the *Washington Post* in the US and *The Guardian* and *The Daily Mail* with *The Mail on*

Sunday in the UK (Carson and Farhall, 2018). Hence, the Fourth Estate can be regarded as a professional narrative that reflects journalistic ideals but can also reveal conflicts that arise through techno-cultural changes the industry is going through, which may best be localizable in award-winning news outlets.

Place as the geographic and material setting of journalism

Place is imbued with materialities, objects and physical structures that function as a location for action. Connecting people to the places they live in by "providing them with critical news and information as well as a sense of rootedness and belonging" is a core function of journalism (Usher, 2021: p. 5). This function can be understood as tied to the historical dimensions of place, as journalism helps to shape and interpret the evolving narratives and collective memories that define the character and identity of a location. While newswork has moved beyond the Reporter's Gallery, *where* journalism takes place still matters, as it shapes what journalists know and whose information needs they can and do pay attention to. Scholars have long been concerned with geographic concentration of news, and the emergence of news deserts (Toff and Mathews, 2021). This concentration is tied to a divide between large, resource rich news organizations with headquarters in metropolitan areas and smaller, local newsrooms. The techno-cultural change that the industry is going through has only amplified these asymmetries, as small newsrooms are lacking the resources (economically and in form of technologically skilled staff) to keep up and fully leverage new tools and technologies (Rinehart and Kung, 2022).

News organizations' partnerships and the materialities used for journalistic work are locally bound as well. While tech companies, often with the support of national or more local governments, have created new Silicon Valleys, for example in the UK, Germany, France, and Romania (Jones and Sudlow, 2022), the corporations providing the top and middle layers of the infrastructure of digital journalism, are predominantly based in the US (Winseck, 2017). This includes platforms for news distribution and curation, but also tools to track, measure, and physically store data about audiences news consumption, as well as the digital means to leverage this data and deliver ads to consumers (Simon, 2022).

While digital tools and data can seem abstract and immaterial, locating their material conditions enables us to consider their cost—in terms of energy, labor, and capital—but also in terms of how they shape newswork and potentially reinforce inequity. Just like where news organizations are geographically based, these material conditions enable and limit action and experiences in newswork and determine the news that get produced.

Place as lived: Where action and lived experiences are made

The final dimension of place refers to the ways in which news professionals, audiences, and their institutions interact with their environments, build news routines and shape professional ideals and imaginaries. As news organizations hire more technologically skilled staff (Wu et al., 2019), journalists' work is increasingly informed by a new occupational elite (Burrell and Fourcade, 2021). The interaction between this professional group and the journalistic field has produced a new demand for technologically driven

newswork and thereby changed professional roles (Ferrer-Conill and Tandoc, 2018) and news routines (Schjøtt Hansen and Hartley, 2021). While technologically skilled staff are not necessarily formally tied to Silicon Valley (e.g., through former work experience), they are united by the intent and ability to transform journalism through technology—a transformative agency that is inspired by a Silicon Valley techno-centric culture (Lischka et al., 2022; Wu et al., 2019). This meeting of two epistemic cultures—one grounded in traditional understandings of newsmaking and one rooted in technological expertise and new conceptions of newsmaking—has created a renegotiation of journalistic norms, values, and practices in journalism (Lischka, Schaetz and Oltersdorf, 2021). Russell (2019), for example, found that journalists adapted values, that is, the belief that technological solutions can be found for societal problems and that information should be freely accessible. Relatedly, Zotto et al. (2023) caution that contemporary technological development not only creates the demand for technologically skilled staff to transform journalism, but also risks to shift news organizations' attention and resources from their social responsibility to technological-centered marketing activities. Guided by quasi-religious beliefs in technology as facilitator of innovation and disruption (Chen, 2022; Geiger, 2020), Silicon Valley's influence on action and lived experiences in journalism is reflected in what value is assigned to specific epistemic practices, skills, and other forms of labor. Investigating funding intermediaries, Papaevangelou (2023) for example, pointed out that one of the core objectives of Silicon Valley's financial commitment to the news industry is to stimulate innovation. Rao and Neely (2019) demonstrated that a passion schema has become a Silicon Valley marker for structuring hiring and promotion within organizations. Passion here is used as a proxy for professionals' intellectual fit, devotion, and commitment to Silicon Valley ideals. This passion scheme has also increased in journalistic job ads over time (Lindén and Grönlund, 2021). This way, Silicon Valley ideals not only direct who is being hired, but creates homogenization in the workforce. This can take the form of a lack of diversity in race or class, but also in epistemics. It matters who gets to be in a newsroom, and how epistemically diverse organizations are, because professional's backgrounds and associated norms and practices influence what news organizations pay attention to and value, which ultimately shapes the coverage that we see.

Method

Applying a qualitative content analysis on $n = 55$ job advertisements of four award-winning news outlets in the US and the UK, we analyze narratives of Silicon Valley and Fourth Estate ideals in job ads. They represent the public step in staff recruitment and include extensive information on the position, responsibilities, expectations, and organizational identity (Walters and Fage-Butler, 2014).

Sample

We follow a three-step critical-case sampling approach to identify information-rich news outlets and job ads for analysis. First, we selected the US and UK representing two

countries with a historical connection to the development of Fourth Estate ideals. The US is the country in which Silicon Valley is geographically located. Hence, the country selection represents an intensity sampling of the Fourth Estate, while differing in geographical proximity to Silicon Valley.

Second, we selected two news outlets per country with the highest number of awards for investigative journalism according to Carson and Farhall (2018). These outlets are *The New York Times* (*NYT*) and *Washington Post* (*WP*) in the US and *The Guardian* (*Guardian*) and *Daily Mail* with *The Mail on Sunday* (Daily Mail and General Trust, DMGT) in the UK. We collected all job ads from the websites of these news organizations over the period of 3 months in fall 2021, 1.5 years after the beginning of the Covid pandemic ($N = 185$; of those *NYT* $n = 105$, *WP* $n = 32$, *Guardian* $n = 12$, DGMT $n = 36$). The pandemic has been described as an accelerator of the digitalization of newsrooms (Quandt and Wahl-Jorgensen, 2021) and has sparked several funding initiatives from Big Tech (Papaevangelou, 2023), which may indicate that Silicon Valley ideals gained legitimacy in journalism during that period.

Third, we identified those job ads that required technical skills, which is a subset of innovative expertise of news professionals according to Guo and Volz (2019), using content analysis. The coding scheme for technical skills included "newly-developed, machine-related skills" (Guo and Volz, 2019: p. 1301) such as software development, application of specific software, and data analysis. Job ads were coded by two trained coders reaching good intercoder reliability for identifying technical skills (Krippendorff's alpha = 0.84). Across all job ads, 30% included at least one technical skill ($n = 49$, final sample). Typical job positions in the final sample of job ads encompass data analyst, product manager, product designer, multimedia editor or engineering manager. Hence, the skill-based sampling approach includes information-rich, critical cases in a sense that "if it doesn't happen here, it won't happen anywhere" (Patton and Patton, 2002: p. 236) regarding the convergence of Silicon Valley and Fourth Estate ideals.

Material

Job ads are a complex genre of recruitment and strategic organizational communication that include detailed information on the position and the organization. Job ads are material in which notions of place in the sense of Usher (2019) become manifest. Job ads are themselves socially constructed to express expectations of future personnel. By describing the position and its tasks, they provide insight into potential actions and experiences, signal organizational identity as a site of cultural meaning and power, and inform about the workplace as geographic location. Beyond indicating demand for staff, job ads represent a text genre in which professional norms and ideals are observable, revealing what news organizations value and prioritize (Guo and Volz, 2019; Lindén and Grönlund, 2021). The job ads in our sample typically begin with a description of key aspects of the role and its reporting lines ("about the position"), followed by a detailed description of responsibilities ("what you will do"), requirements and qualifications ("who you are") and end with information on working conditions ("what we offer"),

organizational identity ("about us"), and application procedures ("how to apply"). On average, job ads were 532 words in length with a maximum of 1247 words for a Visual Designer (*NYT*, ID 372) and a minimum of 188 words for a Senior Staff Editor at *NYT* (ID 484)). This short *NYT* job ad addressed only role description, requirements, and contact information. For our analysis, role description, responsibilities, and working conditions, and organizational identity contained rich information.

Analysis

The analysis was structured using the thematic analysis method following (Braun and Clarke, 2006). First, we search for core indicators of Silicon Valley and Fourth Estate ideals. Regarding Silicon Valley ideals, we identify disruptive innovation, technosolutionism, and diversity themes. Fourth Estate aspects include narratives about holding those in power accountable, facilitating public debate, representing the public opinion, and linking those in power to civic society as well as independence and signals of Fourth Estate performance through prestigious awards. Second, we noted variations of narratives and themes that are related to Silicon Valley or Fourth Estate narratives, considering how these variations contributed to the overall understanding of the convergence between the two sets of ideals. Third, we differentiated references to sites of meaning-making, action and experience, as well as physical work places. We aimed to understand how these elements were intertwined with the narratives and themes identified earlier. Lastly, through repeated reading and axial coding, we traced the connections between narratives and themes found in the job ads and sorted them according to their relation to each other. We organized connections based on their relationships to each other, allowing for a comprehensive understanding of the interplay between Silicon Valley and Fourth Estate ideals in the context of the job market.

Findings

Place as geographic and material

Describing the location of the workplace, ads detailed the surrounding areas, signaling status to attract candidates: "Our Kings Place offices overlook Regents Canal, with the newly-developed Coal Drops Yard, Granary Square and St Pancras Square shops, restaurants and pop-ups" (Data Scientist, *Guardian*). Place was connected to a certain lifestyle and the possibilities that open up with it. If geographic location was discussed in this manner, it exclusively referred to metropolitan areas that are often characterized by gentrification and belong to the most expensive parts of the city: "You'll be working in the Heart of Kensington, great for shops, bars and restaurants and Hyde Park is just a stone's throw away" (Senior Data Analyst, DMGT). It should be noted that advertisements highlighted lifestyle perks, while journalism-related considerations, that is, proximity to civic structures to facilitate reporting, did not play a role. Similarly, physical proximity to other news organizations (which would have been the case in Kensington) was not mentioned. Lastly, some job postings promoted their "remote-friendly culture" to attract

competitive candidates: "we are proud to employ incredible people across the country" (Analyst, Data & Insights, Wirecutter, *NYT*).

The job ads also revealed materialities in terms of software use in journalism. The adoption of specific software and tools for various aspects of news production, data analysis, and distribution indicates the demand for proficiency in content management systems, data analysis tools like Power BI, data visualization tools like Tableau and D3.js, as well as database management systems like SQL. Moreover, the job ads shed light on the infrastructures supporting news production, such as cloud-based services for content storage and collaboration, and content delivery networks for efficient distribution of multimedia content. The growing use of software and digital tools in journalism indicates a deepening interdependence between the technology and journalism sectors.

Place as location for action and lived experiences

Passion structuring action. Passion emerged as a dominating principle that structures action and lived experiences in journalistic places. It was found to be an established hiring criterion across organizations and positions that is tied to three functions. First, passion is used to assess the intellectual fit of the professional. The ideal candidate "loves to develop solutions for difficult problems," "loves designing through collaboration" (Senior Product Designer/Developer, *WP*), or is generally "a passionate person who loves technology" (App Developer, *Guardian*). This function serves as a precondition for innovation, which is tied to a data analysis capacity. In order to innovate with data, passion is related to data culture, especially in the context of the *NYT*. Professionals should be "passionate about data" (Manager, Data & Insights, Engagement, *NYT*), "enthusiastic about data and its applications" (Senior Manager, Data & Insights, *NYT*). The intellectual passion for "answering hard questions with data" (Data Scientist (Algorithmic Targeting), *NYT*) should transfer to developing data-based outcomes by "turn[ing] passion for data and support of world-class journalism into tools and insights." (various positions, *NYT*). As such, future newsworkers are expected to "thrive in data driven decision making, love ambiguity and use research and data to help answer key business and customer questions" (Product Manager, *DMGT*).

A data-focused culture, collaborative and passionate workstyles, and problem solving can be related to aspects of Silicon Valley ideals. Fourth Estate aspects are also tied to passion. Passion is related to the field of journalism and its mission and becomes a proxy for devotion and commitment. As such, applicants should, for example, be "eager for the opportunity to support [the organization's] journalistic mission" (Manager, Data & Insights, Engagement, *NYT*). Hence, passion serves as a precondition for professionals' agency to contribute to the organization and its journalistic mission as a whole via innovation. This is laid out in the following sections—starting with candidates' agency within the organization.

Organizational agency as innovator. News organizations describe candidates' potential agency in the organization. Their intellectual fit and their allegiance to the culture of data are envisioned to enable them to act as a facilitator of change. Data analysts' responsibility

has "a critical and strategic impact on all aspects of the company," "driv[ing] data informed decision making at all levels of the organization" (Data Analyst, *WP*). Data processing capacities of news organizations were listed as incentives, signaling candidates' ability to fully leverage their skills at the outlet: "Work on a truly modern data stack (dbt, Airflow, Fivetran and Tableau)" (Data Developer, *DMGT*). This conception of agency is closely linked to notions of innovation. News designers are, for example, expected to "push the boundaries of visual storytelling" (News Designer, *WP*). Data analysts are elevated to innovators: "we are now hiring the next innovator" (Data Analyst, *WP*). The work of these professionals is expected to "shape the future of our business" (Senior UX Researcher, WP), as organizations are reframing newsrooms as a site for this professional group to exercise agency over and to model according to their vision: "innovation doesn't end in the newsroom" (Data Analyst, *WP*).

Notably, job ads did not always state to what ends innovation should be fostered. Clear goals were usually expressed in connection to economic expectations—for example, to grow subscription bases, or increase revenue. However, calls for boundary pushing or innovation were often open-ended—like an end in itself—asking "how will you impact tomorrow?" (Data Analyst, *WP*).

Individual agency is positioned within a plurality of agents within the organizations that are described as a professional and diverse community, representing Silicon Valley narratives: "some of the brightest and best people in media and beyond: award winning journalists, cutting edge commercial teams, industry leading digital talent and much, much more. We are a vibrant and diverse community with many events, groups and forums to get involved in" (various data positions, Guardian). That is, award-winning journalism as a signal for Fourth Estate performance and innovation capacities are closely linked. What follows outlines how journalism's ideals as a Fourth Estate are established and how candidates are imagined to exercise societal agency via innovation.

Societal agency as Fourth Estate. Findings show clear references to journalism's ideals of the Fourth Estate. Indicating power over the public sphere The New York Times, for instance, describes itself as "respected for its visionary, high-quality journalism, its reach and its influence." (Product Manager, Growth, *NYT*). In their mission statements, ads argue for journalism's role in improving society: "great journalism has the power to make each reader's life richer and more fulfilling, and all of society stronger and more just." (Product Manager, Growth, *NYT*). As a third variant, references to awards as indicators for Fourth Estate performance are drawn: "We are renowned for our groundbreaking independent journalism and our award-winning digital products." (Data Quality Analyst, *Guardian*). Here, the historic Fourth Estate narrative of independence is connected to a commodified product-related excellence.

These Fourth Estate references are often tied to action and experience narratives: "This is a rare opportunity to work at a historic and innovative global media organization dedicated to improving society by creating and distributing high-quality news and information." (Product Manager, Growth, *NYT*). The power of the organization is made accessible to the individual who becomes a powerful contributor to societal improvement. In contrast, some ads focused on economic objectives. Data analysts were expected to

"help deliver the Customer Vision," which was meant to "understand as much as we can about our customers and their behaviour, driving them to spend more time with, and more money with, the Mail brand" (Data Analyst, *DMGT)*. Others suggested compatibility between Fourth Estate performance and commercialization: "You can count on doing work that has a direct impact as we strengthen and grow our business." (Product Manager, Growth, *NYT*). To become a Fourth Estate agent, specifically, "you will work to evolve the most important surfaces via which readers engage with our journalism including the homepage, the newsreader app, the story page" (Analyst, Data & Insights, Engagement, *NYT*), which again ties into innovation agency. Innovation agency, structuring action, is closely connected to dimensions of power and as such, to Usher's final dimension of place.

Place as a site of cultural meaning and power

Converging powers of tech and Fourth Estate. In order to remain an agent of the Fourth Estate, *NYT* is becoming more tech-oriented and describes itself as a tech company: "*The New York Times* is a technology company committed to producing the world's most reliable and highest quality journalism. Our ability to do so relies on a talented team of expert technologists who help *NYT* learn from a tremendous abundance of data unique to this company" (various data positions, *NYT*). This statement ties data to the Fourth Estate. A similar framing could be found at the Washington Post, converging Fourth Estate and Silicon Valley: "The Post is as much a tech company as it is a media company, combining world-class journalism with the latest technology" (Data Analyst, *WP*).

The ability to extract and analyze data, which is personified in the tech team, becomes a direct source for achieving Fourth Estate ideals. Data has become the capital that is meant to attract tech talent, it invites technologists to participate in (social) analysis and innovation in a way that is proposed to strengthen the Fourth Estate. This shift is acknowledged as "a moment in history when the importance of learning from our data is transforming every aspect of the craft and practice of journalism" (various data positions, *NYT*). It recognizes the disruptive power of data for the field, suggesting that datafication fundamentally changes what has formerly been known as journalism practice. As a result, organizations with a legacy journalistic identity are reframing themselves as tech companies. While keeping with their journalistic mission, agency has at least partially been shifted to data analysis. Who benefits from these transformations and what indicators job ads give for the mitigation or reinforcement of inequity in journalism is explored in what follows.

Power distribution and equity in journalism. While disruption narratives were often linked with mission statements to improve society, ads lacked demonstrations of power redistribution within the news organizations that would reflect its normative expectations. Instead of evidence for changes in hegemonic power structures, postings reflected the status quo. This was particularly evident in narratives surrounding newsroom equity, where expressions of commitment to diversity and inclusion were salient: "[C]ommitted to increasing diversity and maintaining an inclusive workplace culture," organizations

welcomed applications "from all qualified candidates regardless of their ethnicity, race, gender, religious beliefs, sexual orientation, age, marital status, or disability" (Visual Design Lead, DMGT). That is, they were aspiring "to inclusive working experiences and an environment that reflects the audience [they] serve" (Data Scientist, *Guardian*). Notably, organizations linked newsroom diversity to the quality of journalistic work. Striving to provide audiences "with high-quality, trustworthy news and information," they argued that this "mission is best served by a diverse, multi-generational workforce with varied life experiences and perspectives" (Data Analyst, *WP*).

However, these statements were rarely grounded in workplace policy that could enhance equity in journalism. Beyond accommodations like flexible working hours and to work remotely, job ads lacked a commitment to an anonymized application process, mentoring structures or offers for paid parental leave. Additionally, pay transparency was largely absent, despite being a core driver of pay equity along the lines of race and gender. Relevant to the US context, most ads did not offer healthcare plans. The Guardian presented a notable exception, offering physical and mental health insurance, up to 18 weeks of parental leave at full pay, and a "Multi-faith prayer room" (Engineering Manager, Guardian) to candidates. Overall, however, findings suggest that the convergence of Silicon Valley and journalistic ideals does not disrupt, but reinforce power structures that marginalize traditionally underrepresented groups in journalism.

Discussion

The analysis of UK and US based job ads suggests that Silicon Valley and Fourth estate ideals shape three dimensions of place—place as geographic and material, as location for action and experience, and as a site for cultural meaning and power (Usher, 2019)—with consequences for inequity in journalism. Tracing the history of the news industry, we can see that the adoption of innovation as a central ideal has roots in both the early days of print journalism and the more recent rise of digital technologies. Findings suggest that the news industry has adopted innovation as a central ideal, but unlike Silicon Valley's conceptualization, innovation is not primarily expected to be disruptive (Geiger, 2020). Rather, it involves the enhancement of the role of data and its agents for journalism (Lewis and Westlund, 2015). By hiring staff that already show a passion for data analysis, news organizations are fostering a data culture aimed at finding innovative solutions for journalism products that efficiently reach news audiences. Passion is not only a proxy for professionals' commitment and intellectual fit, as suggested in Rao and Neely (2019), but also serves as a crucial condition for the strengthening of data culture. In this way, Silicon Valley techsolutionism is being refined into datasolutionism. This positivist stance on datafication is viewed to support journalism's long-standing Fourth Estate mission and increased data analysts' agency in journalism.

The evolved Fourth Estate ideal reflected in our findings centers on audience access and addressing a wider public. As wide audience access serves as a source of financial independence for journalism (Burrowes, 2011), this narrative is used to legitimize the need for data-based innovation. However, instead of referring to multiple dimensions of the Fourth Estate, such as holding power to account (Schultz, 1998) or controlling content

(Sherwood and O'Donnell, 2018), news organizations focus on retaining control over access to society.

As technologically skilled staff increasingly gains importance to fulfill these goals (Wu et al., 2019), the emplacement of the Fourth Estate has expanded from the Gallery to tech team offices. Although places of institutional political power or proximity to political actors remain crucial to journalism, data analysis has opened up new avenues for holding power accountable that are not necessarily limited to a specific location and depend on physical infrastructure for hardware and knowledge production. Nevertheless, journalism continues to center in metropolitan areas, due to economic and cultural powers that offer news professionals an attractive lifestyle and easy access to international networks. This indicates that inequities in terms of where journalism takes place are likely to persist, despite tech workers' abilities to work remotely.

Silicon Valley and Fourth Estate ideals share a commonality in the way that they narrate and create tangible opposition to hegemonic power structures. Originally, Silicon Valley opposed bureaucratic, state, and particularly military power (Burrell and Fourcade, 2021). Likewise, the Fourth Estate has been closely tied to physical locations of institutional political power, opposing opaque and unequal state power (Burrowes, 2011). From its traditional stronghold in the press gallery, journalism has extended to tech team offices and other locations where data analysis and digital innovation take place. Despite journalism's dependency on tech companies for distribution channels and data extraction (Lewis and Westlund, 2015), we find that news organizations position themselves as an authority over their own data. Data analysis has opened up new avenues for holding power accountable that are not necessarily limited to a specific location and depend on physical infrastructure for knowledge production. Moreover, news organizations seek organizational agency via data practices to reach and inform a broader audience, innovate journalism and, ultimately, remain the Fourth Estate. Power is centered in the economic imperative to extract and analyze data, confirming that journalism, too, has accepted data as a form of capital (Sadowski, 2019). An imperative to remain critical of data practices or to be aware of ethical challenges when levering data for economic pursuits could not be identified.

Findings confirm an additional disconnect between ideals to challenge power and their implementation within the news companies (Twine, 2018). Despite aspirations to hire people from diverse backgrounds, we find little evidence of strategies that would enable historically marginalized people to fill these positions and thrive in them.

On a comparative level, findings indicate that geographic proximity to Silicon Valley between the US and the UK does not cause systematic differences in the adoption of its ideals, while the history of journalism in both countries may have shaped their respective approaches to technology and innovation. NYT stands out as an outlet with strong references to Silicon Valley ideals and a comparatively high demand for data analysts. This may be partially attributable to the historical evolution of American journalism, which has embraced innovation as a way to remain competitive. However, the sample size of four news outlets is too small to arrive at generalizable conclusions, particularly given the focus on legacy news organizations that might depend strongly on platforms for audience reach (Pyo, 2022). Future studies might explore news organizations in other

countries that have different relationships to Silicon Valley or understandings of journalism beyond its traditional role as the Fourth Estate. This could provide valuable insights into how technology and journalism intersect in different cultural contexts.

Further limitations of this study concern the intensity sampling, as well as material. The subsampling approach led to a small sample of job ads compared to the selection frame. A subset of job ads that required programming skills of the originally selected news outlets was analyzed. While this procedure allowed us to quantify demand for tech skills compared to other roles, future studies may benefit from a longer observation period to ensure that news outlet diversity is preserved. Second, using job ads as the primary source of material comes with several disadvantages. Job ads only reveal the demand for new positions. Our material thus gives insights into how news organizations imagine and shape the future of the profession, and gives little insight into narratives that are tied to the past. Because the convergence of Silicon Valley and Fourth Estate ideals are ever evolving, findings inevitably only reflect a snapshot in time. Job ads also revealed a variance in explicitly describing geographic dimensions of place. Some job ads only listed locations, while others described where the work was placed in more detail, in an effort to signal status. Still, job ads communicate geographic dimensions of place reliably, demonstrating concentrations of power in affluent, metropolitan areas.

Further, our sampling focus on positions requiring technological skills could have led to a stronger alignment with Silicon Valley ideals. Diverging notions may only become transparent though interviews with editorial staff or in newsroom observations. Similarly, this study does not provide any insights into the content controlling aspects of the Fourth Estate and allows for little analysis of how ideals are enacted in practice. Finally, our sample was based on large, award-winning outlets. Future research with a more diverse sampling would be needed to examine if findings of this study can be extended to less well-resourced organizations.

While job advertisements provide a valuable source of information about the skills and expertise sought by news organizations, deeper analysis of required software skills could shed light on the technical details of news production, distribution, and monetization. This would give a more nuanced understanding of how these changes are manifested in different positions within newsrooms. Furthermore, the efforts of platforms to fund journalism and become the infrastructure of the news industry go beyond the platformization of news production. They also extend to exerting influence over the larger news ecosystem.

Conclusion

This study investigates how Silicon Valley and Fourth Estate ideals converge, by examining these processes through a historical lens and a framework of place—as material and geographic, a setting for action and lived experiences, and a site of cultural meaning and power (Usher, 2019). We qualitatively analyzed job advertisements of four high-status Fourth Estate legacy news outlets in the US and UK—countries with similar roots in the Fourth Estate histories but that differ in their geographical proximity to Silicon Valley.

Historically, both Silicon Valley and the Fourth Estate have been narrated as oppositions to hegemonic powers, sharing ideals of societal transformation. Findings indicate that Silicon Valley and the Fourth Estate converge in location, action and experience, as well as cultural meaning and power. We argue that it is each place's historic narrative that lays the groundwork for convergence. Because both Silicon Valley and Fourth Estate are narrated as oppositions to hegemonic powers, they share ideals of transforming society. However, the power asymmetries and unequal distribution of resources between Big Tech and journalism reinforces hegemonic power structures in journalism. While journalism holds cultural power due to its unique role in society, Silicon Valley is rich in economic capital and consolidates power through its technological control over journalistic production and news distribution.

As a result, journalism has adopted a form of Silicon Valley's technosolutionism, more specifically, a notion of datasolutionism and connected it to Fourth Estate ideals of audience access. Our findings thus support the notion that Silicon Valley ideals homogenize meaning-making and, as a result, practice independently from the geographical proximity to a news organization.

Specifically, we observed that news organizations are orienting themselves towards, or are even framing themselves as, tech companies, emphasizing the importance of data-driven solutions. As journalism has sought to develop its innovative capacities, news organizations now recruit people with a Silicon Valley mindset that represent a new occupational elite (Burrell and Fourcade, 2021). This elite supports the positivist notion, that journalism's problems can primarily be solved by data-driven practices. As such, journalism advances the convergence of Silicon Valley and Forth Estate ideals within organizations by employing staff who foster datasolutionism in their organization. Although journalism remains committed to its traditional role as the Fourth Estate, it is now embracing Silicon Valley's datasolutionism as a new way to achieve that goal. This convergence is resulting in the consolidation of power and the emergence of a transformative agency for data analysts within these organizations.

The concept of place is historically rooted in the complex interplay between historical narratives, cultural meaning, power dynamics, and the evolving relationship between journalism and Silicon Valley. The convergence of their ideals into tech- and datasolutionism can be seen as a response to the historical context of journalism's role in society and the challenges it faces in the digital age. The layers of history, culture, and meaning contribute to the ongoing transformation of journalism and its adaptation to the challenges of the digital era. By considering place from a historical perspective, our analysis provides a more comprehensive understanding of how the convergence of Silicon Valley and Fourth Estate ideals has occurred, and how this convergence shapes the landscape of journalism today.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Nadja Schaetz https://orcid.org/0000-0001-6692-3471

Note

1. Solnit, R. (n.d.). Now And Then. Harper's Magazine. Retrieved April 4, 2023, from https://harpers.org/archive/2017/09/now-and-then-2/

References

Barbrook R (1998) The hi-tech gift economy. *First Monday* 3(12). doi: 10.5210/fm.v3i12.631. Available at: https://doi.org/10.5210/fm.v3i12.631

Braun V and Clarke V (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology* 3(2): 77–101. doi: 10.1191/1478088706qp063oa.

Burrell J and Fourcade M (2021) The society of algorithms. *Annual Review of Sociology* 47(1): 213–237. doi: 10.1146/annurev-soc-090820-020800.

Burrowes CP (2011) Property, power and press freedom: emergence of the fourth estate, 1640–1789. *Journalism and Communication Monographs* 13(1): 1–66. doi: 10.1177/152263791101300102.

Carson A and Farhall K (2018) Understanding collaborative investigative journalism in a "post-truth" age. *Journalism Studies* 19(13): 1899–1911. doi: 10.1080/1461670X.2018.1494515.

Chen C (2022) *Work Pray Code When Work Becomes Religion in Silicon Valley.* New Jersey: Princeton University Press.

Couldry N and Mejias UA (2019) *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating it for Capitalism*. Stanford: Stanford University Press.

D'Souza RA, Rauchberg JS and Nieto-Fernandez B (2023) The labor of diversity in the 2020–2021 U.S. communication job market. *Journal of Applied Communication Research* 1–19. doi: 10.1080/00909882.2023.2173532.

Ferrari E, Scoglio M, Piazza G, et al. (2020) Transcatheter aortic root replacement with chimney grafts for coronary perfusion: a preliminary test in a three-dimensional-printed root model. *Interactive cardiovascular and thoracic surgery* 31(1): 121–128. doi: 10.1093/ccc/tcz051.

Ferrer-Conill R and Tandoc EC (2018) The audience-oriented editor: making sense of the audience in the newsroom. *Digital Journalism* 6(4): 436–453. doi: 10.1080/21670811.2018.1440972.

Geiger S (2020) Silicon Valley, disruption, and the end of uncertainty. *Journal of Cultural Economy* 13(2): 169–184. doi: 10.1080/17530350.2019.1684337.

Guo L and Volz Y (2019) (Re)defining journalistic expertise in the digital transformation: a content analysis of job announcements. *Journalism Practice* 13(10): 1294–1315. doi: 10.1080/17512786.2019.1588147.

Hamilton J (2016) *Democracy's Detectives: The Economics of Investigative Journalism*. Cambridge & London: Harvard University Press.

Hampton M (2010) The fourth estate ideal in journalism history. *The Routledge Companion to News and Journalism*. London: Routledge, 3–12.

Jones H and Sudlow B (2022) A contemporary history of Silicon Valley as global heterotopia: Silicon Valley metaphors in the French news media. *Globalizations* 19(7): 1122–1136. doi: 10.1080/14747731.2022.2034308.

Lewis SC and Westlund O (2015) Big data and journalism: epistemology, expertise, economics, and ethics. *Digital Journalism* 3(3): 447–466. doi: 10.1080/21670811.2014.976418.

Lindblom T, Lindell J and Gidlund K (2022) Digitalizing the journalistic field: journalists' views on changes in journalistic autonomy, capital and habitus. *Digital Journalism* 1–20. doi: 10.1080/21670811.2022.2062406.

Lindén CG, Grönlund M, et al. (2021) Journalistic passion as commodity: a managerial perspective. *Journalism Studies* 22(12): 1701–1719. doi: 10.1080/1461670X.2021.1911672.

Lischka JA, Schaetz N and Oltersdorf AL (2022) Editorial technologists as engineers of journalism's future: exploring the professional community of computational journalism. *Digital Journalism* 1–19. doi: 10.1080/21670811.2021.1995456.

Massey D (1995) Places and their pasts. *History Workshop Journal* 39: 182–192.

Meese J and Hurcombe E (2021) Facebook, news media and platform dependency: the institutional impacts of news distribution on social platforms. *New Media and Society* 23(8): 2367–2384. doi: 10.1177/1461444820926472.

Moore M and Tambini D (2018) *Digital Dominance: The power of Google Amazon, Facebook, and Apple*. New York: Oxford University Press.

Morozov E (2014) *To Save Everything, Click here: The Folly of Technological Solutionism (Paperback 1. Publ)*. United States: PublicAffairs.

Myllylahti M (2021) It's a dalliance! A glance to the first decade of the digital reader revenue market and how the google's and Facebook's payments are starting to shape it. *Digital Journalism* 1–19. doi: 10.1080/21670811.2021.1965487.

Napoli P and Caplan R (2017) Why media companies insist they're not media companies, why they're wrong, and why it matters. *First Monday* 22(5). doi: 10.5210/fm.v22i5.7051.

Nielsen RK and Ganter AS (2022) *The Power of Platforms: Shaping Media and Society*. New York: Oxford University Press.

Papaevangelou C (2023) Funding Intermediaries: Google and Facebook's Strategy to Capture Journalism. *Digital Journalism* 1–22. doi: 10.1080/21670811.2022.2155206.

Patton MQ and Patton MQ (2002) *Qualitative Research and Evaluation Methods*. 3rd edition. London: Sage Publications.

Pyo JY (2022) Different stakes, different struggles, and different practices to survive: news organizations and the spectrum of platform dependency. *New Media and Society* 23: 146144482211232. doi: 10.1177/14614448221123279.

Quandt T and Wahl-Jorgensen K (2021) The coronavirus pandemic as a critical moment for digital journalism: introduction to special Issue: covering Covid-19: the coronavirus pandemic as a critical moment for digital journalism. *Digital Journalism* 9(9): 1199–1207. doi: 10.1080/21670811.2021.1996253.

Rao A and Neely T (2019) What's love got to do with it? Passion and inequality in white-collar work. *Sociology Compass* 13(12). doi: 10.1111/soc4.12744.

Rinehart A and Kung E (2022) *Artificial Intelligence in Local News. AP.* https://www.ap.org/assets/files/ap_local_news_ai_report_march_2022.pdf

Russell FM (2019) The new gatekeepers: an institutional-level view of Silicon Valley and the disruption of journalism. *Journalism Studies* 20(5): 631–648. doi: 10.1080/1461670X.2017.1412806.

Sadowski J (2019) When data is capital: datafication, accumulation, and extraction. *Big Data and Society* 6(1): 205395171882054. doi: 10.1177/2053951718820549.

Schjøtt Hansen A and Hartley JM (2021) Designing what's news: an ethnography of a personalization algorithm and the data-driven (Re)assembling of the news. *Digital Journalism* 1–19. doi: 10.1080/21670811.2021.1988861.

Schultz J (1998) *Reviving the Fourth Estate: Democracy, Accountability and the Media*. 1st edition. Cambridge: Cambridge University Press. Available at: https://doi.org/10.1017/CBO9780511597138

Sherwood M and O'Donnell P (2018) Once a journalist, always a journalist?: industry restructure, job loss and professional identity. *Journalism Studies* 19(7): 1021–1038. doi: 10.1080/1461670X.2016.1249007.

Simon F (2022) Uneasy Bedfellows: AI in the News, Platform Companies and the Issue of Journalistic Autonomy. *Digital Journalism* 1–23.

Toff B and Mathews N (2021) Is social media killing local news? An examination of engagement and ownership patterns in U.S. Community news on Facebook. *Digital Journalism* 1–20. doi: 10.1080/21670811.2021.1977668.

Tomaskovic-Devey D and JooHee H (2018) Is Silicon Valley tech diversity possible now? *Center for Employment Equity. UMass Amherst*. Available at: https://www.umass.edu/employmentequity/silicon-valley-tech-diversity-possible-now-0

Tumber H (2001) Democracy in the information age: THE role of the fourth estate in cyberspace. Information. *Information, Communication and Society* 4(1): 95–112. doi: 10.1080/13691180122542.

Turner F (2006) *From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism*. Chicago: University of Chicago Press.

Twine FW (2018) Technology's invisible women: black Geek girls in Silicon Valley and the failure of diversity initiatives. *International Journal of Critical Diversity Studies* 1(1): 6–116. doi: 10.13169/intecritdivestud.1.1.0058.

Usher N (2019) Putting "place" in the center of journalism research: a way forward to understand challenges to Trust and knowledge in news. *Journalism and Communication Monographs* 21(2): 84–146. doi: 10.1177/1522637919848362.

Usher N (2021) *News for the Rich, White, and Blue: How Place and Power Distort American Journalism*. New York: Columbia University Press.

Walters N and Fage-Butler A (2014) Danish job advertisements: increasing in complexity. *Communication and Language at Work* 3(38): 38. doi: 10.7146/claw.v1i3.16556.

Winseck D (2017) The geopolitical economy of the global Internet infrastructure. *Journal of Information Policy* 7: 228–267. doi: 10.5325/jinfopoli.7.2017.0228.

Wu S, Tandoc EC and Salmon CT (2019) When journalism and automation intersect: assessing the influence of the technological field on contemporary newsrooms. *Journalism Practice* 13(10): 1238–1254. doi: 10.1080/17512786.2019.1585198.

Zotto CD, Omidi A and Norouzi E (2023) Smart exclusion: how may digital platforms hinder inclusivity within news organizations? In: Weder F, Rademacher L and Schmidpeter R (eds). *CSR Communication in the Media*. Cham: Springer international publishing, 93–108.

Author biographies

Nadja Schaetz is a PhD Candidate at the University of Hamburg, where she studies digital journalism, with a focus on power and inequity.

Laura Laugwitz is a PhD Candidate at the University of Hamburg. She focuses on feminist perspectives on and in computer science and communication studies, e.g. in the context of hate speech or news recommender systems.

Juliane A. Lischka is Professor of Journalism and Communication Studies at the University of Hamburg. Her research focuses on digitization strategies of news companies, the role of technology in journalism, and changes in professional standards in journalism.