

Made, Contested, Invisible – A Critical Data Studies Perspective on Reimagining Methodology in a Datafied Communication Science

Annekatriin Bock*

This paper explores how Critical Data Studies (CDS) and communication studies can enrich each other, particularly in the context of the current uptake of methodological approaches supported by large-scale data analytics and “AI.” First, the paper discusses the entanglement of Science and Technology Studies (STS) and CDS as a theoretical and methodological framework. Second, it highlights key assumptions of CDS: (1) the “made-ness” of data, which emphasizes that data are not neutral, but shaped by algorithms and practices; (2) the “interpretative sovereignty over data,” which addresses power relations and access barriers in data analysis; and (3) the “(in-)visibility of data,” which concerns the de- and recontextualization of data and their (in)visibility in research processes. Third, two methodological examples—data journeys and the walkthrough approach—illustrate how CDS approaches are applied in communication studies. The paper concludes by emphasizing the importance of cooperative collaboration between different scientific disciplines and paradigms to develop innovative and critical methodological approaches capable of addressing the challenges of future-oriented communication studies in a data-infused, algorithmically orchestrated society.

Key words: Datafication, Science and Technology Studies, Critical Data Studies, data journeys, walkthrough approach, socio-technical systems, Retrieval-Augmented Generation (RAG)

1. Introduction

Digital technologies, algorithmic systems, and data-driven practices are no longer peripheral to social life—they are constitutive of it (Esposito, 2024; Mayer-Schönberger & Cukier, 2013). From social media platforms to AI-powered content generation, data are produced, collected, and used in ways that reshape how individuals communicate, how institutions operate, and how knowledge is produced. This transformation is not merely technical; it is deeply social, political, and epistemological. As communication research grapples with these changes, a growing body of scholarship calls for a critical re-engagement with data—not as neutral evidence, but as something *made*, *contested*, and *inherently relational* (boyd & Crawford, 2012; van Dijck, 2014; Zakharova, 2022).

Against this background, communication science research is currently called upon not to replace its established methods, but to (re-)position itself methodologically by critically reflecting on its existing strengths and exploring how they can be reimagined in response to a datafied world. Contemporary research, for example, discusses the application of large language models (LLM) in qualitative research (e.g. Franken, 2022; Wiedemann, 2013; Wollin-Giering et al., 2024). Yet, while many scholars highlight the potential of emerging technologies—such as LLMs, and while the implications of retrieval-augmented generation (RAG) remain under debate—to revolutionize data processing and analysis, this enthusiasm often rests on a reductive view of both technology and method (e.g. Binz et al., 2025; Peters

* Prof. Dr. Annekatriin Bock, Universität Vechta, Medienkompetenzzentrum Vechta, Kolpingstraße 25, 49377 Vechta, Germany, annekatriin.bock@uni-vechta.de, <https://orcid.org/0009-0006-9951-4846>.

& Chin-Yee, 2025). The assumption that AI tools can “automate” communication research, for instance, overlooks the fact that data are not raw material but a product of specific practices, power relations, and interpretive choices (Crawford et al., 2014; Pink et al., 2022).

This paper argues that a fundamental challenge for communication science lies not merely in adopting new tools, but in rethinking the very foundations of data and methodology, a task for which Critical Data Studies (CDS) and Science and Technology Studies (STS) are well equipped. Both offer promising starting points for empirically grasping, analyzing, and theorizing the (not always entirely new) developments associated with the increasing datafication of communication and communication science (e.g. Dieter et al., 2019; Light et al., 2018; Troeger & Bock, 2022). Informed by STS, which in turn combine insights from information science and cultural studies, CDS have long been concerned with the datafication of societies driven by the availability and processability of digital data (e.g. Lippert & Mewes, 2021; M. T. Schäfer & Van Es, 2017; Schneider et al., 2021). For example, CDS can be applied to analyze data traces (Breiter & Hepp, 2018), data practices (Decuyper, 2021), and data flows (Jarke et al., 2023) in communication science research.

Drawing on the core tenets of CDS—the *made-ness of data*, *interpretative sovereignty over data*, and the *invisibility of data flows*—this contribution reframes the current methodological debate not as a crisis of obsolescence, but as an opportunity for critical innovation. Data is not a given; it is constructed through decisions about what to collect, how to process it, and which interpretations to privilege. Access to data is unevenly distributed, reinforcing existing power imbalances between corporations, institutions, and independent researchers. Crucially, much of data’s journey—from collection to analysis to publication—is obscured by what scholars call “blackboxing” (Hillman, 2023; Rogers, 2017).

Against this background, the following argumentation is divided into three parts. First, drawing on an understanding of datafication as a social “meta-trend in the context of digitality” (Breiter & Bock, 2023), the entanglement of datafication, communication, and society are problematized. Second, STS and the core assumptions of contemporary CDS are introduced, and their points of reference for communication studies are outlined. Third, two methodological examples are used to reflect on the meaningful interconnection between CDS and research questions in communication science.

The article concludes with an outlook on implications of a CDS perspective for communication science concerned with datafied communication and its socio-political implications. It frames this perspective as an invitation for researchers to follow data not as a linear path, but as a network of decisions, negotiations, and exclusions—revealing what is made visible, what is ignored, and who holds the power to decide. Far from advocating a retreat from technology, this contribution urges communication science to embrace complexity, uncertainty, and reflexivity. In a world where data is made, access is unequal, and processes are often invisible, the researcher’s role is not to automate, but to *interrogate*. Consequently, the future of communication science methodology lies not in faster tools, but in deeper questions.

2. Datafied Society | Communication | Research

“Smart”, “personalizable”, “data-driven” technologies, along with the interwoven practices of collecting, producing, distributing, and using data, are widely considered to be fundamentally transforming society (Breiter & Bock, 2023; Couldry, 2017; Decuyper et al., 2021; Hintz et al., 2019; Mayer-Schönberger & Cukier, 2013). The collection and utilization of data has a long tradition not only in datafied societies but, in particular, in empirical research (M. T. Schäfer & Van Es, 2017). What is remarkable about contemporary datafication, however, is that data are not only analogue but also digital, available in unprecedented

quantities and varieties (Crawford et al., 2014). As a result, data can be generated and processed in real time and flexibly expanded and scaled across heterogeneous data sets, enabling exhaustive compilation and fine-grained analysis (Kitchin, 2014).

Datafication, understood as a “social meta-trend in the context of digitalization”, is therefore a “highly complex social process in which different stakeholders, actor constellations, practices and power structures interact” (Breiter & Bock, 2023). Data-enabled communication has consequently altered the ways in which societies communicate through technology. On the one hand this development has encouraged communication research (e.g. Katzenbach, 2022; Keller & Klinger, 2019; Livingston, 2018; van Es & de Lange, 2020) that examines what it means to engage communicatively with machines that, while lacking genuine understanding, nevertheless act as conversational partners (Pütz & Esposito, 2024). This discourse is further enriched by studies on technology-driven social changes that conceptualizes technologies as “transformers” reshaping human experience (Allert, 2020; Dander et al., 2020).

On the other hand, these developments also entail methodological implications. Currently, a rapid evolution can be observed from large language models (LLMs) toward Retrieval Augmented Generation (RAG), which combines traditional language models with sophisticated information retrieval processes (Sahin, 2025). This integration enables information from extensive text corpora to be accessed and incorporated prior to answer generation, thereby significantly enhancing the contextual awareness and informational depth of generative socio-technical systems compared to earlier models. As such advanced models become increasingly embedded in interpersonal communication, they function not merely as tools but as co-creators of individual interactions, social dynamics, and broader societal norms (Esposito, 2024; Pütz & Esposito, 2024). In an increasingly digital and networked world, shaped by technologies that frame digital autonomy through algorithmic processes, scientific interest in these dynamics of change continues to grow, even as their societal effects remain opaque.

There is already a substantial body of research addressing the datafication of communication and society (e.g. Houben & Prietel, 2018; M. T. Schäfer & van Es, 2017). For instance, Houben and Prietel (2018) introduce the concept of a “data society” to describe a society that continuously produces data, describes itself through data, and depends on data in many respects. Schäfer and van Es also problematize the changing role of data: “Just as electricity changed industrial processes and domestic practices in the nineteenth century, a data-driven paradigm will constitute the core of twenty-first-century processes and practices” (Schäfer and van Es, 2017, S. 11).

For the authors of these contributions, it is therefore consequential that data increasingly move to the center of media and communication research. They conclude that in a society in which many aspects of language, discourse, and culture are datafied, research must also critically examine data itself as an object of inquiry. Scholarship concerned with the datafication of society (Bock et al., 2023; e.g. Houben & Prietel, 2018; M. T. Schäfer & van Es, 2017) shares, among other insights, the view that datafication processes are essential for contemporary social transformations and that the role of data for the reproduction and control of society requires continuous critical reflection.

The following considerations therefore rest on the assumption that communication in a datafied society, in a world shaped by algorithmic decision-making and interwoven with socio-technical systems, is undergoing significant change, and that communication science research must adapt accordingly in order to continue to adequately address emerging objects of study and the associated research questions.

Consequently, communication science has long shared an interest in data and data-based communication processes with other research disciplines (e.g. Breiter & Hepp, 2018; Gentzel, 2017). Currently, however, against the backdrop of ongoing datafication processes and stimulated by the possibilities of socio-technical systems, often simplistically referred to as “AI”, a new entanglement of methodological approaches in (communication) science can be observed. At the same time, many established methods appear to require adaptation, if not to have become partially outdated, in light of technologies based on machine learning, deep learning, or language modeling. For example, who will continue to rely on manual transcriptions of guided interviews or complex text reduction and categorization procedures informed by content analysis when LLMs can process interview data within seconds?

In view of these developments, communication research is increasingly called upon to (re-)position itself methodologically, or rather to reflect more systematically on its existing strengths and integrate them into the available systems. This requires collaboration across different schools of thought and research paradigms in order to find answers for the question of which alternative or innovative approaches might replace established methods or meaningfully complement them. Irina Zakharova points out, that social science research on datafication is “performative” (Zakharova, 2022), meaning that research on datafication can actively shape the phenomena it investigates. She therefore argues for greater reflexivity regarding methodological and conceptual approaches in data research and, following Law (2021) and Law and Lin (2020), calls for “more careful data studies, attuned to empirical, conceptual, and ontological multiplicities (Mol, 2002) of datafication processes” that are simultaneously open to reflection and “(self-)critique” (Zakharova, 2022, p. 20).

As illustrated so far, an emerging body of research promotes a closer interweaving of datafication research and qualitative research approaches (e.g. Wiedemann, 2013; Zakharova, 2022). One field that already brings together technological implications with critical and reflexive inquiry are STS, which have gained considerable prominence in Anglo-American research contexts and have been substantially informed by media and communication research. STS, together with CDS, therefore serve as a theoretical frame of reference for the following considerations.

3. Science and Technology Studies as a theoretical and methodological reference frame

As an interdisciplinary field, STS explore the entanglement of science, technology, and society, seeking to understand how scientific knowledge and technological systems are developed, how they influence society, and how society, in turn, influences them (Sismondo, 2010). STS conceive of science and technology as “thoroughly social activities”, since “scientists and engineers are always members of communities, trained into the practices of those communities and necessarily working within them” (Sismondo, 2010, p. 24).

STS combine insights and methods from various disciplines, including sociology, anthropology, history, philosophy, and political science, to provide valuable perspectives on how scientific knowledge and technological systems emerge and operate within broader social and cultural contexts. The field has developed and become institutionalized since the 1970s (Gentzel, 2017, Lengersdorf & Wieser, 2014). Today, STS encompass a range of interdisciplinary research approaches, including critical data, code, and algorithm studies, which address, for example, the power of algorithmically structured orders and the emergence of big data within data assemblages.

Interdependencies between STS and communication studies have been discussed, for instance, by Carrozza & Pereira, who explore “whether Digital Methods could thus be an area where STS and Communication Studies meet” (Carrozza & Pereira, 2016, p. 231). Based on their literature review, the authors are able to show that “it is not only Communication

Studies and STS that might connect through digital methods; digital methods also emerges as a mode of communication of different scientific practices” (Carrozza & Pereira, 2016, p. 232).

To interrogate the entanglement of communication studies and STS, this article focuses on CDS as a research strand closely linked to STS. Three core assumptions of CDS serve as a starting point for further reflections on intersections and cultivations relevant to communication studies: the “made-ness of data”, “interpretative sovereignty over data”, and the “(in-)visibility of data”.

3.1 *The made-ness of data*

Research that approaches data from a CDS perspective assumes that data are neither neutral nor raw nor value-free; rather, they are always “cooked” and aggregated through algorithms imbued with specific interests and connotations (boyd & Crawford, 2012, p. 667). Data surround us and are generated through our actions and our digital tools. They are collected, evaluated, monetized and, in part, made publicly available (Crawford et al., 2014; Mayer-Schönberger & Cukier, 2013; van Dijck, 2014). While communication studies often treat data as a product or units of calculation, CDS, with its emphasis on the made nature of data, invites a shift in perspective toward data practices—that is, toward *doing data*.

Sergio Sismondo, for example, highlights that conversations about data interpretation in the laboratory rarely become part of the research publication. “Other mechanisms must replace conversations in order to turn data into convincing evidence” (Sismondo, 2010, p. 127). Similarly, in contemporary communication science, data are predominantly treated as proof or fact (“data as evidence”). The negotiation processes before and during data interpretation, before data become data and evidence, are usually absent from published papers. In fact, some journals even require that data analysis be documented in a linear, clear, and comprehensible manner. While data are “made” in the sense understood by STS and CDS and emerge in contexts beyond scientific inquiry, “at the same time digital methods rely heavily on these data produced for other purposes than research in multiple ways” (Zakharova, 2022, p. 59).

For instance, during data collection, researchers may employ surveys, interviews, or automated data collection tools such as web scrapers, each representing a distinct form of data production. Surveys and interviews require active participation from respondents, whose answers are influenced by their interpretations and biases. Web scrapers, on the other hand, collect data passively from online sources, which themselves are shaped by the algorithms and structures of the platforms being scraped.

During data analysis, researchers may clean, transform, and aggregate data, processes that also constitute forms of data production and can introduce additional biases and assumptions. Even in the publication phase, the way data are presented—in charts, tables, or narratives—reflects choices that influence how the data are perceived and interpreted.

For example, consider a study on social media usage patterns. In the data collection phase, researchers might use Application Programming Interfaces (APIs) to gather tweets, making decisions about which keywords to search for and which time periods to include. During data analysis, they might apply sentiment analysis algorithms to categorize tweets as positive, negative, or neutral, embedding assumptions about language and emotion. In the publication phase, they might present their findings in a bar chart, which involve choices about grouping and visual displaying the data. Each of these steps involves a form of data production, introducing biases and assumptions that shape the final results.

The example illustrates the importance of researchers remaining particularly attentive to the constructed nature of data at every stage of the research process. Reflections on the “made-ness of data” give rise to important follow-up questions for negotiating new methodologies or further developing existing ones in communication studies: How is data produced throughout the research process? How do data sets and their informative value change when certain data gaps are addressed through methods such as “data implementation”, while others remain unfulfilled? Which evaluation logics and procedures are taken for granted as standard and unquestioned, and which must still demonstrate their empirical “usefulness” or “correctness”?

3.2 *The interpretative sovereignty over data*

Not only are data „made“, but their analysis and subsequent use are also subject to significant restrictions. For example, the resources required to access and make data available are unevenly distributed. Some can collect digital traces of human interaction “from a multitude of technical systems, such as websites, social media platforms, smartphone apps, or sensors” (Stier et al., 2020, p. 504), yet they are constrained by what information is technically accessible, crawlable, or otherwise collectable. Research that relies on non-public, proprietary data can be prohibitively expensive. Certain data owners may charge substantial fees or permit access only after appropriate legitimization or disclosure of the research design, which can limit academic freedom. At the same time, independent research, such as that conducted in communication science institutes at universities, often lacks the necessary financial resources.

Opportunities to utilize data for decision-making or to participate in the design of datasets and surveys are similarly unevenly distributed (cf. eg. Bock et al., 2023; Eubanks, 2019; Hepp et al., 2022; Kennedy, 2018; Warschauer, 2003). STS-informed work demonstrates that data are not uncontroversial (van Dijck, 2014; Zuboff, 2019), yet they underpin far-reaching decision-making processes across modern societies (e.g. Couldry, 2017), particularly in domains where communication plays a central role. This includes, for example, journalistic newsrooms increasingly publishing AI-generated content or curating content algorithmically for specific audiences, as well as social media platforms that create algorithmically driven filter bubbles, influencing information behavior and communication to varying degrees (Mahrt, 2019).

Questions about the role of bots or hate speech in election campaign communication further illustrate the stakes, particularly in political communication research (Keller & Klinger, 2019; Pentzold et al., 2019). The concept of interpretative sovereignty over data raises critical concerns regarding who holds the power to control, interpret, and leverage data (e.g. von Scherenberg et al., 2024).

Consider, for example, social media platforms such as Facebook, TikTok, or LinkedIn. The platform providers collect vast amounts of user data, which they use to target advertisements and influence user behavior. Access to this data for academic research is often limited, either due to high costs imposed by the platform providers as data owners or because of strict data-sharing policies designed to protect what is considered proprietary data. These restrictions constrain researchers’ ability to study the societal impact of social media on society and to critically examine the algorithms and data practices of these platforms. Consequently, interpretative sovereignty over data largely remains in the hands of corporations, which can shape public discourse and policy decisions according to their own interests.

This example underscores the need to recognize the limitations of data resulting from unequal access and (un-)availability. For communication science research informed by CDS, insights into the interpretative sovereignty over data raise several important questions:

Who possesses the skills or knowledge to locate and analyze openly available research data? And how can research on datafied communication critically engage with its subject matter if access to the necessary data remains restricted or entirely blocked?

3.3 *The (in-)visibility of data*

Data and their entanglements are not always visible or equally empirically tangible. Data are often “decontextualized” and must subsequently be “recontextualized” (Zakharova, 2022; Zakharova & Bock, 2023). In other words, data, figures, and datasets are “made” from actions. These data artifacts then “end-up” in the hands of individuals who recontextualize them—initiating actions, issuing instructions, drawing conclusions, and making decisions based on the available data. The underlying processes and data flows are not always transparent and are often partially obscured, with conscious or unconscious “blackboxing” complicating data analysis (Hillman, 2023; Jarke et al., 2023; Rogers, 2017; Zakharova & Bock, 2023).

The insight from CDS that data can be made visible or invisible generates important methodological considerations for communication science research (Stehle et al., 2024). For example, when using data for research purposes, it is essential to reflect on which data are actually tangible, visible, and observable during collection and analysis. Qualitative and ethnographic research have long established strategies for visualizing the researcher’s own situatedness and the limitations of the field of observation (e.g. situatedness, dealing with going native, follow the data, etc. (e.g. Agar, 2006; Quinlan, 2008)). In these approaches, the impossibility of accessing invisible data is not viewed as a deficiency in scientific quality; rather, it is productively transformed into socio-critical questions about access: Who determines which data are made freely available, and what are the societal implications if some data are emphasized while others are rendered invisible?

Consider, for example, algorithmic decision-making in political communication during election campaigns. Political parties and candidates often use data analytics to target specific voter demographics with tailored messages. However, the data and algorithms underpinning these targeted communications are frequently invisible to both the public and researchers. This invisibility carries significant implications for democratic processes, as it can lead to the manipulation of public opinion and contribute to the spread of misinformation.

For communication science research, several critical questions arise from the considerations on the (in)visibility of data: Which data are invisible, or actively made invisible? Which data collection and analysis methods can be employed to uncover such invisible data and render it researchable? Which decision-making and analytical logics determine which data are visualized and which are overlooked in the process? How do these practices of (in)visualisation affect the interpretation and meaningfulness of the research?

4. **Discussing methodological approaches**

The assumptions of CDS discussed here serve as a theoretical foundation for considering the practical implications of datafication for communication science research, particularly regarding the applicability of CDS approaches in empirical practice. CDS encompass a diverse methodological spectrum aimed at uncovering the social, political, and ethical dimensions of data practices (Decuyper, 2021).

These methods include data ethnographies, which involve immersive observation and participation in data-driven environments; data activism, which focuses, for example, on advocating for transparency and accountability in data use; and data feminism, which examines issues such as gendered aspects of data production and interpretation (e.g. Charles

& Gherman, 2019; Knox & Nafus, 2018; Pink et al., 2018). Additionally, CDS employ approaches such as data visualization critiques, which analyze biases and assumptions embedded in visual representations of data, and data justice frameworks, which assess the fairness and equity of data-driven decision-making processes (e.g. Gerlitz, 2017; Williamson, 2016). Among these methods, data journeys and walkthrough approaches are particularly illustrative and will be discussed in more detail in the following sections.

4.1 *Data journeys*

Juliane Jarke and colleagues (2023), for example, use the visualization of invisible practices as a starting point for their conceptual analysis, employing data journeys as a method to investigate data-driven educational management and administration. Drawing inspiration from Eleftheriou et al. (2018) and the work of Jo Bates and colleagues on data journeys (Bates et al., 2016), they focus on visualizing data flows—tracing where data travel, where interruptions occur, where data leave an institution or silo, and what happens in subsequent decision-making steps.

Jarke and her colleagues pay particular attention to the challenges and breaks in visualization, as well as the lengthy negotiation process required to produce a shared representation of the data flow. What is typically invisible in publications and presentations—the ambivalences, ruptures, and complexities of datafication in administrative organizations—is central to their analysis, making these phenomena empirically tangible.

The authors highlight that data flows in communication processes cannot be neatly visualized; instead, they are messy and fragile. Much of what could provide especially valuable insights into decision-making in management and administrative communication proves impossible to visualize or observe and therefore cannot be captured by traditional methods such as guided interviews or surveys. The example of fragile data journeys thus underscores the limitations of “established” methods, which largely rely on participants’ ability to verbalize their impressions.

4.2 *The walkthrough approach*

Another example of an STS- respectively CDS-informed method applied in communication research is the walkthrough approach, which draws analytical attention to the affordances of socio-technical systems and the practices interwoven with them (Light et al., 2018; M. S. Schäfer & Wessler, 2020; Troeger & Bock, 2022). The socio-technical walkthrough allows researchers to analyze algorithmic personalization of socio-technical systems such as platforms or apps, making system affordances visible and open to reflection (Troeger & Bock, 2022).

Walkthroughs, that utilize a “research persona”, include a systematic exploration of the respective platform, whereby interactions and feedback received by the research persona are documented for each step, each time the platform is approached. “Research personas” and “othering” techniques are employed to visualize, better understand, and critically reflect on individual usage paths and algorithmically personalized feedback. As Light et al. (2018) note, the walkthrough approach is particularly suited “to explore interfaces, including how they ‘script’ the user, by systematically documenting and abstracting interface features from their normative infrastructural settings.”

The socio-technical walkthrough functions as a “border crosser” (Troeger & Bock, 2022, p. 51), combining qualitative and quantitative aspects of research, making it well-suited for multi-method designs and interdisciplinary studies. For communication science, this approach is particularly valuable because it integrates media-analytical and reception-ana-

lytical perspectives, addressing two central questions of the field: What do people do with technology, and what does technology do with people?

In sum, data journeys allow to trace the lifecycle of data from its creation to its use and eventual disposal, revealing the various actors, technologies, and contexts that shape data at each stage (Bates et al., 2016; Eleftheriou et al., 2018; Jarke et al., 2023). This method is particularly valuable for understanding the complex and often hidden processes through which data are produced, circulated, and interpreted. Socio-technical walkthrough approaches, in contrast, systematically explore and document the algorithmic personalization of data-driven platforms and technologies. By simulating user interactions, these walkthroughs can uncover design choices, biases, and power dynamics embedded in data-driven interfaces (e.g. Dieter et al., 2019; Light et al., 2018; M. S. Schäfer & Wessler, 2020; Troeger & Bock, 2022). Focusing on data journeys and walkthroughs in this contribution highlights the dynamic and relational nature of data, as well as the ways in which data practices are embedded in broader social and technological systems. These approaches offer a comprehensive and nuanced understanding of the made-ness, interpretative sovereignty, and (in-)visibility of data, making them particularly relevant for communication studies.

5. Concluding thoughts

This contribution advances a threefold argument: First, it underscores the need for communication science to adapt its research paradigms in response to the datafication of society, emphasizing interdisciplinary collaboration to develop innovative and reflexive methods (e.g. Houben & Prietel, 2018; M. T. Schäfer & van Es, 2017). Second, it demonstrates how STS, and particularly a CDS perspective, provides a productive theoretical lens for understanding data as made, interpreted, and invisibly entangled within socio-technical systems. Third, it highlights methodological approaches such as data journeys (Bates et al., 2016; Jarke et al., 2023) and socio-technical walkthroughs (Troeger & Bock, 2022) as tools that expose the constructed, contingent, and relational nature of data—making them especially suited for critical communication research.

CDS add a valuable perspective to communication science by emphasizing the social, political, and ethical dimensions of data practices (e.g. boyd & Crawford, 2012; Decuyper, 2021; van Dijck, 2014). By adopting a CDS lens, communication scientists can gain a deeper understanding of how data are produced, interpreted, and utilized across different communication contexts. This approach encourages researchers to critically examine the biases, assumptions, and power dynamics embedded in data-driven communication processes, thereby enhancing the rigor and relevance of their analyses (e.g. Gerlitz, 2017; Rogers, 2017; Zakharova & Bock, 2023).

Moreover, CDS methods—such as data journeys and socio-technical walkthroughs—offer practical tools for uncovering hidden aspects of data, making them visible and open to critical scrutiny. Without broader adoption of CDS, communication science risks overlooking the complexities and nuances of data practices, potentially producing oversimplified or biased conclusions. By integrating CDS, communication science can enrich its methodological toolkit, foster interdisciplinary collaboration, and contribute to more informed, reflective, and responsible approaches to data-driven communication practices.

There is already an ongoing discussion within the discipline regarding the futures of research in a datafied society. The computational turn, according to van Es and colleagues (2021), provides tools that greatly expand the possibilities for conducting research and allow researchers to engage directly with the activities that are shaping our digital society. In the context of an “algorithmic turn” (Katzenbach, 2022) or a “computational turn” (van Es et al., 2021, p. 61) which introduce new uncertainties for communication science research, van

Es and colleagues advocate for the development of best practices for tool application. These include standards for the use of data visualizations in publications and training colleagues and students in both the use of the tools and in critical tool literacy. Tool critique is a key component of responsible research, ensuring that computational tools are applied in a reflective and informed manner (van Es et al., 2021).

This contribution builds on these discussions but takes a slightly different direction. These developments shed new light on the so-called “science war” (Parsons, 2003) and invite a closer interlinking of the logics of qualitative and quantitative research—or at least a stronger dialogue between the two approaches. What are the implications if standardized research continues to follow the “drive to replication” (van Es et al., 2021, p. 21) in an era of algorithmic data production and analysis? What “rules” and research criteria should the discipline adopt when reproducibility or traceability cannot be guaranteed due to algorithmic methods?

One possible avenue is to draw on the quality criteria and research logic of qualitative or ethnographically oriented research, which has long emphasized, for example, the explication of the constructedness of data and the situatedness of researchers (e.g. Breidenstein et al., 2020; Hirschauer, 2014). Communication science research already supports this perspective: “In other words, responsible knowledge claims not only make transparent how they have been constructed, they also provide a justification and account as to why such claims were constructed in the way they have” (van Es et al., 2021, p. 56).

In the current landscape of datafied communication and datafied research, there is not yet a satisfactory, fixed “method compendium” with standardized rules (van Es et al., 2021, p. 61), as has long existed for traditional approaches such as observation, interviews, or content analysis. It is also uncertain whether such a standardized set of methodological tools will ever be available in an era of algorithmic data collection and analysis—particularly given the continuing (and potentially exponential) growth in the capabilities of socio-technical systems, exemplified by developments such as Retrieval-Augmented Generation (RAG) in generative AI.

This raises a critical question for the discipline: in light of the rapid acceleration of change in communication science research objects, should the field focus on standardizing individual methods, or—as this article argues—invest more energy in methodological innovation that foregrounds the “made-ness”, “interpretive processes”, and “invisibility” of data?

As an outlook for future communication science interested in datafied communication and its socio-political implications, three key questions emerge, highlighting challenges for further research.

First, if data is “cooked”, “biased”, or otherwise “problematic”, how can researchers critically engage with its use in politics, science, and society? From a CDS perspective, researchers are encouraged to adopt a stance of critical data literacy. This involves interrogating the provenance of data, the methods used to collect it, and its inherent biases. Reflexive engagement with data requires acknowledging its socio-political context and the power dynamics that shape it (e.g. Eubanks, 2019; Zakharova & Bock, 2023). Collaborative efforts with data scientists, ethicists, and domain experts can foster a more nuanced understanding and responsible understanding of data, while promoting transparency and accountability through open data practices and robust documentation can help mitigate these challenges.

Second, if the availability of data for research is highly regulated and concentrated among powerful “*data haves*”, how can researchers who are “*data non-haves*” contribute to shape infrastructures for research? Navigating the power imbalances requires proactive engagement in the development of research infrastructures. Researchers who are “data

non-haves” can advocate for open data policies, data sharing agreements, and collaborative research frameworks that democratize data access.

Engaging with policymakers, funding bodies, and institutional stakeholders can help ensure that infrastructures prioritize equitable data access. Additionally, building alliances and networks with other “data non-haves” can amplify voices and influence and foster more inclusive data infrastructures. Initiatives such as the Communicative AI Research group—with its ComAI research space and a focus on governance at the ZEMKI in Bremen—offer promising platforms to explore and advance these goals.

Third, if data is “travelling”, “invisible”, and “intertwined”, what methodological (re-)adjustments should, can, or must researchers make? Scholars are encouraged to embrace interdisciplinary approaches that integrate qualitative and quantitative methods to trace data journeys and uncover hidden dynamics. Adopting a relational perspective that situates data within its broader ecosystem can illuminate how data are intertwined with people, technologies, and institutions (e.g. Wiedemann, 2013).

Methods such as data ethnography, data provenance analysis, and network analysis can provide insights into the complex lifecycles and entanglements of data. Additionally, fostering a culture of continuous learning and methodological adaptation enables researchers to remain responsive to the evolving landscape of data and its societal implications.

Literature

- Agar, M. (2006). An ethnography by any other name ... *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, Vol 7, No 4* (2006): Qualitative Research in Ibero America. <https://doi.org/10.17169/fqs-7.4.177>
- Allert, H. (2020). Plattformökonomie und Entstaatlichung: Familienorientiert, ortsunabhängig und #freierlernend. In V. Dander, P. Bettinger, E. Ferraro, & K. Rummler (Eds), *Digitalisierung – Subjekt – Bildung. Kritische Betrachtungen der digitalen Transformation* (pp. 183–212). Barbara Budrich. <https://doi.org/10.2307/j.ctvvb7n3h.13>
- Bates, J., Lin, Y.-W., & Goodale, P. (2016). Data journeys: Capturing the socio-material constitution of data objects and flows. *Big Data & Society*, 3(2), 2053951716654502. <https://doi.org/10.1177/2053951716654502>
- Binz, M., Alaniz, S., Roskies, A., Aczel, B., Bergstrom, C. T., Allen, C., Schad, D., Wulff, D., West, J. D., Zhang, Q., Shiffrin, R. M., Gershman, S. J., Popov, V., Bender, E. M., Marelli, M., Botvinick, M. M., Akata, Z., & Schulz, E. (2025). How should the advancement of large language models affect the practice of science? *Proceedings of the National Academy of Sciences*, 122(5), e2401227121. <https://doi.org/10.1073/pnas.2401227121>
- Bock, A., Breiter, A., Hartong, S., Jarke, J., Jornitz, S., Lange, A., & Macgilchrist, F. (Eds). (2023). *Die datafizierte Schule* (1. Auflage 2023). Springer VS. <https://doi.org/10.1007/978-3-658-38651-1>
- boyd, d., & Crawford, K. (2012). Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), Article 5. <https://doi.org/10.1080/1369118X.2012.678878>
- Breidenstein, G., Hirschauer, S., Kalthoff, H., & Nieswand, B. (2020). *Ethnografie: Die Praxis der Feldforschung* (3., überarbeitete Auflage). UVK Verlag.
- Breiter, A., & Bock, A. (2023). Datafizierte Gesellschaft | Bildung | Schule. In A. Bock, A. Breiter, S. Hartong, J. Jarke, S. Jornitz, A. Lange, & F. Macgilchrist (Eds), *Die datafizierte Schule* (pp. 1–35). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-38651-1_1
- Breiter, A., & Hepp, A. (2018). The complexity of datafication: Putting digital traces in context. In A. Hepp, A. Breiter, & U. Hasebrink (Hrsg.), *Communicative Figurations* (pp. 387–405). Springer International Publishing. https://doi.org/10.1007/978-3-319-65584-0_16
- Carrozza, Tiago Santos, & Pereira. Chiarra (2016). Mapping digital methods: Where science and technology studies and communication studies meet? In A. Gradim & C. Moura (Hrsg.), *Communicating and Evaluating Science* (pp. 203–236). Beira Interior University.

- Charles, V., & Gherman, T. (2019). Big data analytics and ethnography: Together for the greater good. In A. Emrouznejad & V. Charles (Eds), *Big Data for the greater good* (Vol. 42, pp. 19–33). Springer International Publishing. https://doi.org/10.1007/978-3-319-93061-9_2
- Couldry, N. (2017). The Myth of Big Data. In M. T. Schäfer & K. van Es (Eds), *The datafied society: Studying culture through data* (pp. 235–239). Amsterdam University Press.
- Crawford, K., Gray, M. L., & Miltner, K. (2014). Big data| Critiquing big data: Politics, ethics, epistemology | Special Section Introduction. *International Journal of Communication*, 8(0), Article 0. <https://ijoc.org/index.php/ijoc/article/view/2167>
- Dander, V., Bettinger, P., Ferraro, E., Leineweber, C., & Rummler, K. (Eds). (2020). *Digitalisierung – Subjekt – Bildung: Kritische Betrachtungen der digitalen Transformation* (1st edn). Verlag Barbara Budrich. <https://doi.org/10.2307/j.ctvvb7n3h>
- Decuyper, M. (2021). The Topologies of Data Practices: A Methodological Introduction. *Journal of New Approaches in Educational Research*, 10(1), Article 1. <https://doi.org/10.7821/naer.2021.1.650>
- Decuyper, M., Grimaldi, E., & Landri, P. (2021). Introduction: Critical studies of digital education platforms. *Critical Studies in Education*, 62(1), Article 1. <https://doi.org/10.1080/17508487.2020.1866050>
- Dieter, M., Gerlitz, C., Helmond, A., Tkacz, N., van der Vlist, F. N., & Weltevrede, E. (2019). Multi-Situated App Studies: Methods and Propositions. *Social Media + Society*, 5(2), 2056305119846486. <https://doi.org/10.1177/2056305119846486>
- Eleftheriou, I., Embury, S. M., Moden, R., Dobinson, P., & Brass, A. (2018). Data journeys: Identifying social and technical barriers to data movement in large, complex organisations. *Journal of Biomedical Informatics*, 78, 102–122. <https://doi.org/10.1016/j.jbi.2017.12.001>
- Esposito, E. (2024). *Kommunikation mit unverständlichen Maschinen*. Residenz Verlag.
- Eubanks, V. (2019). *Automating inequality: How high-tech tools profile, police, and punish the poor* (First Picador edition). Picador St. Martin's Press.
- Franken, L. (2022). Digitale Daten und Methoden als Erweiterung qualitativer Forschungsprozesse. Herausforderungen und Potenziale aus den Digital Humanities und Computational Social Sciences. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 23(2). <https://doi.org/10.17169/fqs-22.2.3818>
- Gentzel, P. (2017). Praktisches Wissen und Materialität. Herausforderungen für kritisch- konstruktivistische Kommunikations- und Medienforschung. *Medien & Kommunikationswissenschaft*, 65(2), 275–293. <https://doi.org/10.5771/1615-634X-2017-2-275>
- Gerlitz, C. (2017). Data point critique. In M. T. Schäfer & K. van Es (Eds), *The datafied society: Studying culture through data* (pp. 241–244). Amsterdam University Press.
- Hepp, A., Jarke, J., & Kramp, L. (Eds). (2022). *New Perspectives in Critical Data Studies: The ambivalences of data power*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-96180-0>
- Hillman, V. (2023). Bringing in the technological, ethical, educational and social-structural for a new education data governance. *Learning, Media and Technology*, 48(1), 122–137. <https://doi.org/10.1080/17439884.2022.2052313>
- Hintz, A., Dencik, L., & Wahl-Jørgensen, K. (2019). *Digital citizenship in a datafied society*. Polity Press.
- Hirschauer, S. (2014). Un/doing Differences. Die Kontingenz sozialer Zugehörigkeiten. *Zeitschrift für Soziologie*, 43(3), 170–191.
- Houben, D., & Prietel, B. (2018). *Datengesellschaft: Einsichten in die Datafizierung des Sozialen*. Transcript.
- Jarke, J., Hartong, S., Raabe, T., Dabisch, V., Breiter, A., Lange, A., & Zakharova, I. (2023). Zur Erfassung und Modellierung der „Hinterbühne“ von Datenflüssen: Das Beispiel Unterrichtsausfall. In A. Bock, A. Breiter, S. Hartong, J. Jarke, S. Jornitz, A. Lange, & F. Macgilchrist (Eds), *Die datafizierte Schule* (pp. 61–91). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-38651-1_3
- Katzenbach, C. (2022). Der „Algorithmic turn“ in der Plattform-Governance. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 74(S1), 283–305. <https://doi.org/10.1007/s11577-022-00837-4>
- Keller, T. R., & Klinger, U. (2019). Social Bots in Election Campaigns: Theoretical, Empirical, and Methodological Implications. *Political Communication*, 36(1), 171–189. <https://doi.org/10.1080/10584609.2018.1526238>
- Kennedy, H. (2018). Living with data: Aligning data studies and data activism through a focus on everyday experiences of datafication. *Krisis I*, 18–30. <https://eprints.whiterose.ac.uk/id/eprint/12995>

- 9/2/Krisis-2018-1-Helen-Kennedy-Living-with-Data-Aligning-Data-Studies-and-Data-Activism-Through-a-Focus-on-Everyday-Experiences-of-Datafication.pdf-18-30.pdf [30.01.2026].
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures & their consequences*. Sage Publications.
- Knox, H., & Nafus, D. (2018). *Ethnography for a data-saturated world*. Manchester university press.
- Law, J. (2021). From after method to care-ful research (a foreword). In C. Addey & N. Piattoeva (Eds.), *Intimate accounts of education policy research* (pp. xvi–xx). Routledge.
- Law, J., & Lin, W. (2020). *Care-ful Research: Sensibilities from STS*. 17.
- Lengersdorf & M. Wieser (Hrsg.), *Schlüsselwerke der Science & Technology Studies* (S. 123-132). Wiesbaden: VS.
- Light, B., Burgess, J., & Duguay, S. (2018). The walkthrough method: An approach to the study of apps. *New Media & Society*, 20(3), 881–900. <https://doi.org/10.1177/1461444816675438>
- Lippert, I., & Mewes, J. S. (2021). Data, methods and writing: Methodographies of STS ethnographic collaboration in practice. *Science & Technology Studies*, 34(3), 2–16. <https://doi.org/10.23987/sts.110597>
- Livingston, S. (2018). Audiences in an age of datafication: Critical questions for media research. *Television and New Media*.
- Mahrt, M. (2019). Beyond filter bubbles and echo chambers: The integrative potential of the Internet. ifpuk – Institute for Media and Communication Studies at FU Berlin. (*Digital Communication Research*; 1). <https://doi.org/10.17174/DCR.V5.0>
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work and think*. Murray.
- Mol, A. (2002). *The Body Multiple: Ontology in Medical Practice*. Duke University Press. <https://doi.org/10.2307/j.ctv1220ncl>
- Parsons, K. M. (Ed.). (2003). *The science wars: Debating scientific knowledge and technology*. Prometheus Books.
- Pentzold, C., Brantner, C., & Fölsche, L. (2019). Imagining big data: Illustrations of „big data” in US news articles, 2010–2016. *New Media & Society*, 21(1), Article 1. <https://doi.org/10.1177/1461444818791326>
- Peters, U., & Chin-Yee, B. (2025). Generalization bias in large language model summarization of scientific research. *Royal Society Open Science*, 12(4), 241776. <https://doi.org/10.1098/rsos.241776>
- Pink, S., Berg, M., Lupton, D., & Ruckenstein, M. (2022). *Everyday automation: Experiencing and anticipating Emerging Technologies* (1st edn). Routledge. <https://doi.org/10.4324/9781003170884>
- Pink, S., Ruckenstein, M., Willim, R., & Duque, M. (2018). Broken data: Conceptualising data in an emerging world. *Big Data & Society*, 5(1), Article 1. <https://doi.org/10.1177/2053951717753228>
- Pütz, O., & Esposito, E. (2024). Performance without understanding: How ChatGPT relies on humans to repair conversational trouble. *Discourse & Communication*, 18(6), 859–868. <https://doi.org/10.1177/17504813241271492>
- Quinlan, E. (2008). Conspicuous invisibility: Shadowing as a data collection strategy. *Qualitative Inquiry*, 14(8), 1480–1499. <https://doi.org/10.1177/1077800408318318>
- Rogers, R. (2017). Foundations of digital methods. Query design. In M. T. Schäfer & K. van Es (Eds.), *The datafied society: Studying culture through data* (pp. 75–94). Amsterdam University Press.
- Sahin, A. (2025, March 5). What is Retrieval-Augmented Generation (RAG) in LLM and How it works? *Medium*. <https://medium.com/@sahin.samia/what-is-retrieval-augmented-generation-rag-in-llm-and-how-it-works-a8c79e35a172> [27.01.2025].
- Schäfer, M. S., & Wessler, H. (2020). Öffentliche Kommunikation in Zeiten künstlicher Intelligenz. *Publizistik*, 65(3), 307–331. <https://doi.org/10.1007/s11616-020-00592-6>
- Schäfer, M. T., & van Es, K. (Eds.). (2017). *The datafied society: Studying culture through data*. Amsterdam University Press.
- Schneider, T., Fuchs, K., & Mayer, S. (2021). Science and Technology Studies: Die Datafizierung von Alltagspraktiken: Datenaktivismus als neue Verantwortung? In A. Henkel (Ed.), *10 Minuten Soziologie: Verantwortung* (pp. 183–196). transcript Verlag. <https://doi.org/10.14361/9783839451120-012>
- Sismondo, S. (2010). *An introduction to science and technology studies* (Second edition). Wiley-Blackwell.

- Stehle, H., Bock, A., Wilhelm, C., Springer, N., Mahrt, M., Lobinger, K., Linke, C., Engelmann, I., Detel, H., & Brantner, C. (2024). In/Visibility in the digital age: A literature review from a communication studies perspective. *International Journal of Communication*, 18(0), 5471–5493.
- Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review*, 38(5), 503–516. <https://doi.org/10.1177/0894439319843669>
- Troeger, J., & Bock, A. (2022). The sociotechnical walkthrough – a methodological approach for platform studies. *Studies in Communication Sciences*, 22(1), 43–52. <https://doi.org/10.24434/j.scoms.2022.01.3064>
- 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 Es, K., & de Lange, M. (2020). Data with its boots on the ground: Datawalking as research method. *European Journal of Communication*, 35(3), 278–289. <https://doi.org/10.1177/0267323120922087>
- van Es, K., Schäfer, M. T., & Wieringa, M. (2021). Tool criticism and the computational turn. A „Methodological moment” in Media and Communication Studies. *M&K Medien & Kommunikationswissenschaft*, 69(1), 46–64. <https://doi.org/10.5771/1615-634X-2021-1-46>
- van Scherenberg, F., Hellmeier, M., & Otto, B. (2024). Data sovereignty in information systems. *Electronic Markets*, 34(1), 15. <https://doi.org/10.1007/s12525-024-00693-4>
- Warschauer, M. (2003). *Technology and social inclusion: Rethinking the digital divide*. MIT Press.
- Wiedemann, G. (2013). Opening up to Big Data: Computer-assisted analysis of textual data in social sciences. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 14(2). Art. 23. <https://doi.org/10.17169/fqs-14.2.1949>
- Williamson, B. (2016). Digital education governance: Data visualization, predictive analytics, and ‘real-time’ policy instruments. *Journal of Education Policy*, 31(2), 123–141. <https://doi.org/10.1080/02680939.2015.1035758>
- Wollin-Giering, S., Hoffmann, M., Höfting, J., & Ventzke, C. (2024). Automatic transcription of English and German Qualitative Interviews. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 25(1) Art. 8. <https://doi.org/10.17169/fqs-25.1.4129>
- Zakharova, I. (2022). *Understanding data studies—A methodological and conceptual inquiry into research on datafication* [Universität Bremen]. <https://doi.org/10.26092/ELIB/1675>
- Zakharova, I., & Bock, A. (2023). Researching in/visibilities and data practices in education. *On_education. Journal for Research and Debate*, 6(18). https://doi.org/10.17899/on_ed.2023.18.4
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile books.



© Annekatrin Bock