PUBLIZISTIK

Key topic or bare necessity? How Research Ethics are Addressed and Discussed in Computational Communication Science

Philipp Knöpfle (iD) · **Mario Haim** · **Johannes Breuer** (iD)

Abstract In Computational Communication Science (CCS) researchers grapple with intricate ethical challenges arising from the collection and analysis of complex data sets, often including sensitive or copyrighted data. Taking into consideration differences between the two main lines of philosophical reasoning in the realm of (research) ethics—deontology and consequentialism—we argue that ethical challenges faced in CCS are multidimensional and, hence, require multiple perspectives and approaches. Our general considerations are complemented by an empirical study that aims to assess the nature, prevalence, and discussion of ethical issues in CCS literature. Through a manual content analysis of 476 CCS publications, we shed light on ethical challenges as well as reflections thereof by CCS researchers. Notably, we find that only 5.88% of studies explicitly address general ethical considerations. Ethical review processes are also only mentioned in 6.51% of the considered studies, with the majority focusing on specific ethical procedures, such as obtaining informed consent, data anonymization measures, or debriefing. This suggests that, in the absence of consensus and field-specific standards, researchers in CCS prioritize context-specific ethical procedures, emphasizing the importance of flexibility in addressing ethical considerations.

✉ Philipp Knöpfle · Prof. Mario Haim
Institut für Kommunikationswissenschaft und Medienforschung, Ludwig-Maximilians-Universität München, Akademiestr. 7, 80799 Munich, Germany
E-Mail: Philipp.Knoepfle@ifkw.lmu.de

Prof. Mario Haim
E-Mail: Haim@ifkw.lmu.de

Dr. Johannes Breuer
GESIS—Leibniz Institute for the Social Sciences, Unter Sachsenhausen 6–8, 50667 Cologne, Germany
E-Mail: Johannes.Breuer@gesis.org

🙋 Springer

Zentrales Thema oder bloße Notwendigkeit? Wie Forschungsethik in der Computational Communication Science behandelt und diskutiert wird

Zusammenfassung In der computergestützten Kommunikationswissenschaft (Computational Communication Science, CCS) stehen Forscher vor komplexen ethischen Herausforderungen, die sich aus der Erhebung und Analyse komplexer Datensätze ergeben, die oft sensible oder urheberrechtlich geschützte Daten enthalten. Unter Berücksichtigung der Unterschiede zwischen den beiden Hauptlinien philosophischer Überlegungen im Bereich der (Forschungs-)Ethik—Deontologie und Konsequentialismus—argumentieren wir, dass diese ethischen Herausforderungen vielschichtig sind und daher mehrere Perspektiven und Ansätze erfordern. Unsere allgemeinen Überlegungen werden durch eine empirische Studie ergänzt, die die Natur, Verbreitung und Diskussion ethischer Fragen in der CCS-Literatur untersucht. Durch eine manuelle Inhaltsanalyse von 476 CCS-Veröffentlichungen beleuchten wir ethische Herausforderungen sowie die Reflexionen darüber durch CCS-Forscher. Nur 5,88 % der Studien sprechen explizit allgemeine ethische Überlegungen an. Allgemeine ethische Überlegungen werden ebenfalls nur in 6,51 % der betrachteten Studien erwähnt, wobei sich die Mehrheit auf spezifische ethische Verfahren wie die Einholung informierter Zustimmung und Maßnahmen zur Datenanonymisierung konzentriert. Dies legt nahe, dass Forscher in der CCS in Ermangelung von fachspezifischen Standards kontextspezifische ethische Verfahren priorisieren, wobei die Bedeutung von Flexibilität im Umgang mit ethischen Überlegungen betont wird.

1 Introduction

In the field of computational communication science (CCS), researchers engage in the development and application of computational methods to investigate various aspects of human communication to test and develop theories within communication science, particularly in the context of digital and online communication (Hilbert et al. 2019; Shah et al. 2015). Two key features that set CCS apart from other fields (within communication science) are the methods and data types that researchers commonly use.

In addition to practical and methodological challenges, the use of such data and methods raises legal and ethical questions for researchers to address. Generally speaking, legal requirements regulate what researchers are (not) allowed to do,

whereas ethical guidelines describe what researchers should (not) do. In that, ethical guidelines or recommendations typically go beyond legal requirements, serving as the moral compass which guides researchers towards responsible decision-making and behavior. As such, while legal and ethical questions are often intertwined, they can both influence practical decisions made by CCS researchers regarding, for example, study design, data collection, processing and analysis steps, and the sharing of data subsequent to study publication.

In this study, we discuss research ethics within CCS and emphasize the significance thereof. This emphasis traces back to the nuanced nature of ethical inquiries, which often lack definitive answers and can legitimately draw from various divergent perspectives or schools of thought, contrasting the more delineated nature of legal matters. Consequently, navigating the ethical questions associated with the data and methods being used is an important undertaking for CCS researchers.

Ethics is inherently multifaceted encompassing a spectrum of dimensions. A classification of the dimensions of research ethics that is often referred to are the so-called Belmont Principles (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1979), namely Respect for Persons, Beneficence, and Justice. Put simply, the aim of research ethics is to minimize risks and harms while prioritizing the value of research. While research ethics and good scientific practice or research integrity are not synonymous, there is some substantial overlap between them (Emmerich 2020). Within the realm of academic research, both research ethics and integrity play pivotal roles in ensuring the credibility and trustworthiness of scholarly endeavors. Research ethics primarily concerns itself with the responsible conduct of research, emphasizing the protection of human participants, the integrity of data, and the avoidance of harm. Integrity, on the other hand, encompasses broader principles of honesty, transparency, and adherence to professional standards throughout the research process. While both frameworks share the overarching goal of upholding the value of research, they diverge in their emphasis on specific aspects. The comparison between research integrity and research ethics already illustrates that there are different values at play. While guidelines on research ethics or integrity usually do not rank these aspects, in practice, they can be evaluated differently and may even conflict in specific scenarios (Iphofen 2020; Israel 2015).

Beyond practical considerations, an approach to weighing ethical dimensions and arriving at decisions for research is the orientation to a general ethical framework. Broadly speaking, there are two philosophical schools of thought that can provide guidance for research ethics considerations in communication and beyond—deontology and consequentialism. While researchers may not be explicitly aware of this, these also shape ethical discourse and practical decisions in CCS research. Put simply, deontological perspectives are rooted in the philosophical thoughts of the Enlightenment and prioritize adherence to specific norms and fundamental values to guide decision-making, while consequentialism, dating back to philosophical roots of Utilitarianism, centers around the evaluation of anticipated outcomes and their ethical implications (Salganik 2019). Although contemporary research ethics in communication science frequently incorporate elements from both perspectives (Schlütz and Möhring 2018), the discrepancy between decision-

making guided by fundamental values (deontology) and considerations of expected outcomes (consequentialism) is likely also the cause of diverging answers to ethical questions within the CCS community.

In this paper, we pursue both a conceptual as well as an empirical exploration of research ethics in CCS. First, we will discuss features of CCS that raise or are related to ethical questions as well as how deontological and consequentialist perspectives can be applied to those. This is followed by an empirical investigation into how topics related to research ethics are addressed in the CCS literature. To explore the prevalence of issues related to ethical questions and ways in which these are addressed, we conducted a content analysis of the CCS literature. We analyze articles from the field of CCS published in highly ranked and relevant communication science journals between 2010 and 2021. Our content analysis focuses on important study attributes and practices that have implications for research ethics as well as the discussion of important ethical aspects by researchers in the publications themselves, such as informed consent, privacy protection, or ethical review. Our combined conceptual and empirical exploration aims to provide insights into ethical decisions and practices in CCS, their prevalence, argumentative foundations, and their relation to underlying ethical frameworks.

2 Research Ethics in CCS

2.1 Complexities of CCS research

CCS is distinctive for its use of large and complex data sets, often consisting of digital traces and other "naturally occurring" data that require computationally intensive solutions for processing and analysis (van Atteveldt and Peng 2018). As a methodological toolkit, CCS represents a diverse set of computational methods employed for the collection, processing, and analysis of data within the realm of human communication. CCS researchers make use of a wide array of data collection methods, such as Application Programming Interfaces (APIs), web scraping, self-reports, experiments, or data donations. The methods of analysis employed in CCS are similarly diverse, ranging from classical statistical analyses and machine learning (ML) techniques applied to tabular data to network analysis, text mining and natural language processing (NLP), and (semi-)automated analyses of audio, image, and video data (Hilbert et al. 2019). Accordingly, CCS researchers develop and utilize a heterogeneous set of software tools.

Due to its transformative nature via its development and use of novel types of methods and data (as well as ways of combining those), CCS can also be comprehended as a paradigmatic perspective. In essence, CCS represents a shift away from traditional communication science methodology towards new research designs and data types. This paradigmatic shift emphasizes the importance of data-driven insights and the focus on digital communication landscapes (Geise and Waldherr 2021). This, in turn, challenges conventional notions of research design, which, accordingly, also requires a new set of research principles and theoretical foundations (Waldherr et al. 2021). As a paradigm, CCS, thus, shapes the methodologies employed as well as the

overarching theoretical frameworks through which researchers understand and analyze communication processes. In sum, CCS can be viewed as both a methodological toolkit, providing practical tools for research, as well as a paradigmatic perspective, shaping the analytical framework and theoretical underpinnings guiding the study of communication in the digital age.

Given these two conceptualizations of CCS, it seems reasonable to assume that many ethical considerations relevant to traditional communication research are also applicable to CCS research. However, there are some ethical considerations that are unique to or at least more pronounced in CCS. One prominent challenge revolves around the collection, use, and sharing of personal, sensitive or copyrighted data (e.g., media data or digital trace data) for research purposes. In addition to legal questions that researchers need to tackle, the collection of such data raises ethical considerations regarding privacy, informed consent, and the responsible use of information (Lazer et al. 2020). Importantly, ethical dilemmas emerge not only in the collection and sharing of research data but also during the data analysis phase. In this stage, CCS often extract personal attributes through algorithmic estimations. When it comes to algorithmically inferred attributes, ethical considerations involve not only the accuracy, fairness, and discriminatory potential of algorithms but also the potential repercussions of such inferences for individuals and communities (Eslami et al. 2017; Tsamados et al. 2021). CCS researchers must grapple with questions of transparency, bias mitigation, and the implications of their algorithmic inferences. Transparency, in this context, can refer to how clearly researchers communicate and document their methods but also the comprehensibility of the algorithms and addressing the uncertainty of their outputs. Relatedly, bias mitigation involves identifying and minimizing any biases present in the data or algorithms used (Mehrabi et al. 2021).

Traditionally, in the social sciences, including communication research, research ethics have largely focused on the interaction between researchers and study participants. For CCS there are, however, two important things to consider in this context. First, as media and digital trace data are commonly used (on a large scale), in addition to researchers and participants, another actor that is involved in the collection and use of data are commercial companies (e.g., news outlets or social media platforms).[1] The relationships between these companies and researchers—as well as, in the case of social media platforms, also the companies and the users whose data researchers collect—are governed by legal regulations, such as Terms of Service (ToS) or other contractual agreements. Nevertheless, these do not regulate all aspects of data usage and the interests of researchers and commercial companies may be conflicting (Breuer et al. 2020). For example, while the privacy of user data may be a shared interest, the interest in exclusivity and profit on the side of the companies may be at odds with researcher interests or obligations with regard to openness and transparency. Also, adherence to ToS may be a subject of ethics reviews by Institutional Review Boards (IRBs) (Halavais 2019). Second, in cases where social media data (or other types of digital traces) are collected via APIs or

---

[1] For a detailed discussion of the stakeholders involved in the ethics of platform research and their roles, see the recent report by Lukito (2024).

web scraping, researchers are normally not in direct contact with the individuals whose data they gather. This means that it may be very difficult or even impossible to obtain informed consent. The fact that individuals do not explicitly consent to and are even not aware of their data being used for research means that the term participants may be considered inappropriate in this context (Breuer et al. 2023). These peculiarities of CCS illustrate that considerations of research ethics (have to) go beyond the relationship between researchers and study participants. These two points also illustrate that both the data types and the data collection methods strongly affect the ethical considerations that researchers need to engage in as well as possible conflicts between various goals and values (of different involved actors).

The range of ethical considerations that CCS researchers need to make, of course, goes beyond data access and also pertains to transparency. Particularly the sharing of data, is driven by the ethos of collaboration and transparency inherent to scientific inquiry and pronounced recently in the discourse around open science (Dienlin et al. 2021; Longo and Drazen 2016). Importantly, however, data must be shared and utilized responsibly to mitigate the risk of misuse or unintended consequences. Researchers have to find a middle ground between promoting the exchange of knowledge and instituting safeguards against potential harm. Data security, the protection of intellectual property, and a conscientious evaluation of the broader societal implications of one's research are important aspects to consider when sharing research data in science in general and in CCS in particular where data-intensive use as well as the development of novel approaches is common (Alter and Gonzalez 2018).

Successfully navigating these ethical concerns is not only a procedural necessity but also a matter of reconciling diverse and sometimes conflicting goals and values. Hence, it is important for CCS researchers to systematically reflect on and actively engage in discussions of research ethics and develop procedures to adequately address current and future ethical questions in their work. For that, it can be helpful to be(come) aware of and make explicit the perspectives that researchers apply.

2.2 Deontological and consequentialist perspectives for CCS research

Although specific decisions in the design and implementation of CCS research are usually influenced by immediate practical considerations, potential ethical dilemmas can, in many cases, be traced back to (often implicit) conflicts in underlying ethical frameworks. Researchers may not be explicitly aware of these frameworks, yet they inform and influence their ethical decision-making, guiding their principles and actions when facing challenging ethical scenarios. Broadly speaking, these conflicts can be conceptually mapped onto two prominent but often conflicting ethical schools of thought, deontology and consequentialism.

A deontological perspective prioritizes the adherence to explicit norms and fundamental values that serve as general guiding principles in decision-making, while consequentialism revolves around assessing the expected outcomes and their ethical implications (Salganik 2019). Consequentialism in the context of research ethics, on the other hand, translates to evaluating the outcomes and consequences of research decisions. It involves assessing and comparing potentially negative effects with positive impacts (harms and benefits) of decisions.

When discussing ethics, researchers in the social sciences likely prioritize safeguarding participants and their data. However, there are additional ethical considerations relevant to the practical decisions researchers must make. Depending on one's perspective, some of these may also be regarded as ethical obligations. As previously discussed, the extent to which these are seen as obligations relies on the assessment of various norms. Apart from preventing harm, another ethical principle that may be viewed as obligatory is transparency and openness. To align with these norms and enhance the credibility of research findings, sharing research materials and particularly the underlying data is crucial. Determining whether and how to share research data requires a thorough examination of the potential ethical implications of data disclosure (Borgman 2012). This act of sharing research data is integral to promoting transparency, reproducibility as well as replicability, and collaboration between researchers (Dienlin et al. 2021).

Notably, this not only facilitates the verification of findings but also encourages the reuse of existing data sets for novel research questions, thereby positively contributing to the collective knowledge of the scientific community (Fecher et al. 2015). With regard to the goal of increasing the value and impact of research—which can also be discussed as a research ethics issue, as we have outlined above—in addition to the question whether the data are shared, another important question is how they are shared. A widely used set of criteria for sharing research data are the FAIR criteria, according to which data should be Findable, Accessible, Interoperable, and Reusable (Wilkinson et al. 2016). The implementation of these principles and data sharing in general not only raises practical and technical questions but also has ethical implications.[2]

As stated before, the consideration of research ethics often involves diverse objectives and values that may be contradictory or challenging to reconcile. In the context of data sharing, the objective that can clash with openness and transparency is data protection. Particularly in the realm of individual-level human subjects' data, data sharing raises concerns regarding privacy, confidentiality, and the potential misuse of (personal or sensitive) information (Kirilova and Karcher 2017). From a deontological perspective, researchers have the fundamental obligation to respect the autonomy and rights of research participants, including the protection of personal and sensitive information. This is, e.g., captured in the *Respect for Persons* dimension of the Belmont principles. In this context, data sharing may be decided against or restricted to ensure or increase privacy. In contrast, from a consequentialist perspective, one could argue that the benefits of sharing data (e.g. scientific advancement, reproducibility, and collaboration) may outweigh potential harms. The consequentialist viewpoint, hence, may prioritize the greater good for society, the scientific community, and future research over individual privacy concerns. Another advantage of data sharing lies in its potential to enhance data utility. By sharing data, redundancy in data collection can be minimized, leading to increased efficiency and resource conservation. Given that academic research is frequently supported by pub-

---

[2] The close connection between data sharing and research ethics is also highlighted by the initiative of complementing the FAIR criteria with the so-called CARE principles that comprise *Collective Benefit, Authority, Responsibility*, and *Ethics* (Carroll et al. 2020).

lic funding, optimizing the scientific and public benefits of research data aligns with broader societal interests. Notably, if optimizing the scientific and public benefit of research (and the underlying) data is understood as a fundamental value, this line of reasoning can also align with a deontologist perspective.

To apply some of the general distinctions between deontology and consequentialism to CCS, consider a study in which researchers want to collect social media data from individuals. From a strict deontological standpoint, researchers may see obtaining informed consent and ensuring data privacy as essential conditions (*conditiones sine qua non*) for research. Hence, they may decide to use a data collection method that allows for obtaining consent, such as data donation, and not share the data or only in a reduced, aggregated, or very restricted form. By contrast, if researchers employ a consequentialist perspective, depending on the topic and platform(s) they want to study, they may weigh the benefits of the research and the value of openness higher than privacy risks and the need for informed consent. Hence, they may decide to collect data via an API or web scraping and share them openly.

In conclusion, the complex interplay between deontological and consequentialist perspectives in the research ethics in general and in common scenarios within CCS research in particular underscores the inherent complexity in ethical decision-making in CCS. As CCS researchers navigate these ethical complexities, it becomes imperative to recognize the implicit influence of these ethical frameworks, fostering a nuanced approach that seeks to balance fundamental principles and anticipated outcomes. Key questions in this regard are how ethical questions are addressed in CCS research, what parts or dimensions of the research process they relate to, and what ethical frameworks explicitly or implicitly are referenced or built upon by researchers.

3 How are research ethics addressed in the CCS literature?

In the empirical segment of this study, we build on data from a systematic literature review and content analysis of the CCS literature. This analysis was conducted as part of a larger project aimed at assessing the replicability of research within CCS. In that project, we systematically coded various attributes of publications, studies, and their underlying datasets. Our focus in this paper centers on elements derived from this analysis that directly pertain to research ethics, namely: 1) types of data, 2) methods of data collection, 3) practices of data sharing, and 4) explicit references to ethical considerations. Augmenting this quantitative approach, we additionally adopted a qualitative methodology to delve deeper into how ethical deliberations and decision-making processes are articulated within CCS literature.[3]

---

[3] The blinded OSF link to access the supplementary materials and data for this scientific paper is: https://osf.io/vn9fr/?view_only=3a8bce6482d24bce893712504f157a39.

3.1 Data collection

To identify CCS articles, we implemented a comprehensive three-step identification strategy. Initially, we compiled a database of 22,375 English communication science articles from the Clarivate Web of Science database of leading journals (i.e., the top 50 communication science journals according to Scimago Communication Rankings[4] and the top 20 communication journals according to Google Scholar[5]) spanning January 2010 to December 2021. Subsequently, we applied a co-occurrence network model to a manually selected CCS corpus of approximately 150 articles to derive CCS-specific keywords. Employing these keywords, we filtered the initial database, resulting in a corpus of 6556 articles. Using manual classification and supervised machine learning techniques, we categorized 2551 articles as "computational", meaning that they used computational methods to collect and/or process and/or analyze data. A subset of 500 articles was randomly sampled from this pool for detailed manual content analysis. Additionally, we integrated 35 articles from the Computational Communication Research journal (van Atteveldt et al. 2019), which was not initially included in the rankings. Following a final manual inspection, 476 articles were selected as a final sample for analysis, excluding purely qualitative studies lacking a clear computational component.

3.2 Codebook

The codebook for the overall project comprised multiple categories. For our analysis here, we use coded information for three key components: general study attributes, characteristics of the data, and ethical aspects.

Our coding of general study characteristics focused on essential metadata about each study included in our analysis. We collect basic study information, such as article title, author names, year of publication, and journal title.

The second key category we consider here are the attributes of the data used in each publication. The specific attributes we look at are data types, and the methodology used for collecting the data. Additionally, we consider whether and, if so, how the data have been shared. The third key category regarding the explicit discussion of ethical procedures and considerations in CCS research subsumes three primary categories loosely based on Leslie (2023). First, we considered explicit general ethical considerations mentioned by the authors of the study. Such explicit general ethical considerations can encompass a range of topics and principles. For example, it could pertain to explicitly addressing a commitment to transparent reporting and openness, the acknowledgment of conflicts of interest, or discussions of how to uphold standards of research integrity. We systematically identified whether a general ethical consideration is present in a text via keyword-based queries in the full text of an article. Specifically, we conducted full-text searches for terms, such as "ethi*" and "mora*" to discern explicit mentions of ethical concepts within the research. Second, we recorded information about mentions of ethical reviews processes. Institutions

---

[4] https://www.scimagojr.com/journalrank.php?category=3315.

[5] https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=hum_communication.

such as Institutional Review Boards (IRBs) or Ethical Research Committees are responsible for evaluating research proposals and protocols to ensure the latter adhere to ethical principles and regulations. Their primary functions include safeguarding the rights, well-being, and privacy of research participants, reviewing research methods to prevent harm and bias, and verifying that informed consent is properly obtained from participants. We systematically ascertained references to ethical institutions by employing a text query in the full text for each study. Our search criteria include terms such as "ethic-" and relevant additional indicators, such as "board*," "commit*," "panel*," or "review*". Third, we also recorded any mentions of specific ethical procedures. This, for example, includes obtaining informed consent from human subjects involved in the study, protecting their privacy by anonymizing or pseudonymizing data, and addressing potential biases (e.g., in algorithmic and ML approaches) to maintain fairness. These procedures collectively encompass the various ethical steps and actions taken by the researchers to ensure the well-being, privacy, and rights of human participants in research studies and reducing the risk of harm or misuse of the research output. For instance, obtaining the informed consent of research participants is typically considered a crucial ethical procedure in research, ensuring that participants voluntarily and explicitly agree to participate in a data collection and are aware of a study's purpose, risks, and potential benefits. Another important ethical procedure is to provide participants with a choice in their level of engagement in a study: Opt-in and opt-out approaches involve participants actively choosing to participate (opt-in) or withdrawing from participation/their data being used (opt-out). Finally, debriefing is a post-study ethical procedure—common especially in experimental designs—through which participants are provided with additional information, clarification, and potentially also pointers to sources of support when particularly sensitive or burdensome topics are covered. We assessed the mentioning of such ethical procedures by conducting a text query using the terms "brief*," "anonym*," "pseudon*," "consent*," and "opt*", again, for the full paper texts.

The manual coding process was conducted by two trained coders from September 2022 to June 2023. Initially, a 10% subset of the sample was coded, followed by iterative refinements to the coding scheme. A second coding was then performed on another 10% subset to calculate intercoder reliability measures. The aggregate intercoder reliability results on the second 10% subsample, averaged across all categories, are as follows: agreement: 95.57%, Krippendorff's Alpha: 0.80, and Cohen's Kappa: 0.77.

4 Results

4.1 Sample description

Our entire sample contains 476 publications from 34 different journals.[6] Our sample starts one year after the release of Lazer et al. (2009) seminal *Science* article de-

---

[6] For a full list of all journals, see the Appendix.

Table 1 Data types, $n = 476$

Data type	# of obs	Fraction of sample in %
Media content	265	55.67
Self-reported data	149	31.30
Trace data	29	6.09
Other types of data	25	5.25
No data set	8	1.68

scribing the rise of CSS and ends in 2021, as we started collecting data in early 2022.[7] The number of CCS publications in our sample increases considerably over the time. This surge of CCS publications in the last decade may be indicative of the field's growing recognition and the increasing integration of computational methods within communication research (van Atteveldt et al. 2019). Another contributing factor to the rise in publication numbers is the overall increase in communication science publications in the last decade (Rains et al. 2020; Walter et al. 2018).

4.2 Data types and collection methods

Table 1 presents the types of data in the sample. Media content is, by far, the most used data type in CCS, being used in 55.67% of the publications we analyzed. This encompasses texts derived from social media posts (e.g., Twitter, Facebook, Reddit), news articles, but also image and video data. The second most frequent data type in the CCS literature are self-reports. This data type was used in 31.30% of the publications in our sample and includes data from surveys, interviews, and experiments, typically in the form of questionnaires. Only 6.09% of publications in our sample use trace data in a narrow sense which encompasses smartphone data, passive tracking data, sensor data, and search engine data. The category labeled "other types of data" includes cases such as simulation studies (e.g., data from agent-based models) or metadata. Notably, eight studies in the sample, mostly method and tool exhibitions, did not utilize a specific data type.

We obtained descriptions of data collection methods for a total of 334 studies. This corresponds to a coverage of 70.17% of our entire sample, indicating that sufficient information on the data collection processes was not available for the remaining 29.83%. Table 2 summarizes the most common data collection methods in our sample with more than three observations. Frequently used methods include database download (25.69%), API access (24.65%), survey (22.22%), web scraping (7.64%), as well as data collection through third-party data collection tools (5.90%). The high representation of database (e.g., LexisNexis) and API (e.g., Twitter) downloads highlights the field's reliance on commercial digital platforms for data access. Surveys, which are used widely in quantitative empirical communication science research in general, also constitute a substantial portion (22.22%) of the data collection methods in our sample.

---

[7] Our sample includes 30 publications in 2022 which represent corrected final publications of previous "advance online publications" from 2021.

Table 2 Data collection methods (>3 obs.), $n = 288$

Data collection method	# of obs	Fraction of sample in %
Database download	74	25.69
Api download	71	24.65
Survey	64	22.22
Web scraping	22	7.64
Third party data collection tool	17	5.90
Experiment	16	5.56
Recording tool	9	3.12
Web crawler	7	2.43
Interview	4	1.39
Manual data collection	4	1.39

4.3 Data sharing

Among the 476 studies in our sample, a large majority of 427 (89.50%) did not share their data. Only for 27 studies (5.67%), full data were shared, while the authors of 6 studies (1.26%) shared parts of the underlying data. Among the 33 studies that engaged in some form of data sharing, 26 made their data available in public repositories, five provided the data via the online supplementary materials option of the journal, one incorporated data in the article appendix, and another made the data available via the personal website of one of the authors. Notably, nine studies included an explicit data availability statement, specifying conditions under which the data can be accessed, typically indicated as "data available upon reasonable request". Another nine studies did not have an underlying dataset, thus, rendering the data sharing category inapplicable to them. Out of the 26 studies which shared their data in public repositories, most studies (17) chose the Open Science Framework (OSF), followed by GitHub (4), and the Harvard Dataverse (3). Other options that were used are the GESIS Data Archive (1), and the JGSS Daishodai platform (1), a data sharing platform by the Japanese Ministry of Education, Culture, Sports, Science, and Technology.

Figure 1 illustrates data sharing trends within our sample spanning from 2010 to 2022. Despite the relatively small number of cases where data sharing occurred (33), we can observe a noticeable upward trajectory in data sharing activity over the past five years. While instances of data sharing are present in the earlier years of our sample, a significant surge in data sharing activities is particularly evident from 2017 to 2022.

4.4 Ethical considerations

Table 3 shows the results of our content analysis of the mentioning or discussion of explicit ethical considerations in CCS publications. Among the 476 publications in our sample, only 28 (5.88%) papers explicitly address general ethical considerations. These general ethical considerations encompass a diverse range of topics and
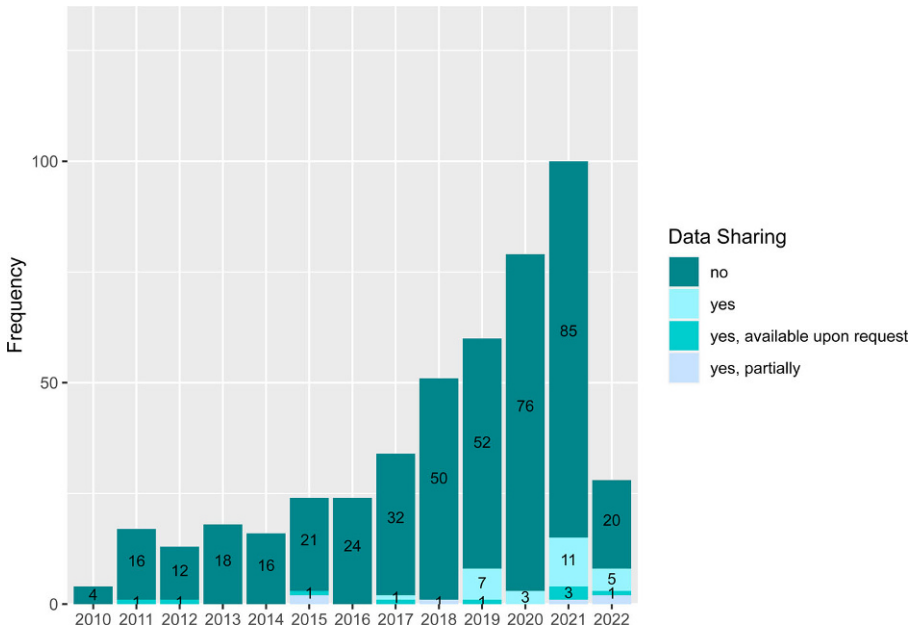
Fig. 1 Data sharing over the years from 2010 to 2022

principles, generally aimed at upholding the integrity, respectfulness, and fairness of the conducted research.

To provide specific examples, researchers may underscore their dedication to safeguarding user privacy during data collection from participants where no prior informed consent is solicited, as exemplified by Urman and Katz (2022): "All the data collected is publicly available to any Telegram user, and, for ethical reasons, in the course of the analysis we relied only on aggregated data without attributing any messages to individual users" (Urman and Katz 2022, p. 17). Other examples include references to specific ethical standards. For instance, Leidig (2019) emphasizes the necessity of data alterations: "[...] to ensure ethical compliance according to the Norwegian Centre for Research Data" (Leidig 2019, p. 86). Similarly, Mc-Cosker (2018) mentions the consideration of ethical guidelines in their research as advocated by the Association of Internet Researchers (AoIR): "The Association of Internet Researchers Ethical processes were considered throughout" (McCosker 2018, p. 6). Siapera et al. (2018) generally acknowledge the notable ethical complexities associated with big data, which have spurred the formulation of specialized ethical protocols: "[...] big data presents considerable ethical challenges, leading to the development of specific codes of ethics and best practices (Zook et al. 2017) [...], in our study, we have strived to closely adhere to these best practices" (Siapera et al. 2018, p. 4).

Another example of a general ethical consideration in our sample comes from McCosker (2018). For his study on digital interventions on a mental health platform, the author emphasizes several aspects related to research ethics, such as continuous communication with the community manager to ensure adherence to ethical prin-

Table 3 Mentions of ethical considerations

Ethical consideration	# of articles	Fraction of sample in %
General ethical considerations	28	5.88
Ethical review processes	31	6.51
Ethical procedures	69	15.50

ciples and guidelines of the platform, as well as the involvement of the platform in the research process from design to publication, including formal processes of ethical reviews: "Communication with beyondblue's [mental health platform] community manager regarding process and ethical oversight was constant, beginning with research design, formal ethical review, and through sign-off for any publications" (McCosker 2018, p. 6). Delving into an extended discussion on the ethical risks linked to face detection technologies, Jürgens et al. (2022, p. 191) exemplify a general ethical consideration in their analysis of age and gender discrimination on German TV with deep learning face recognition: "Precise automated detection and classification of faces is a potentially highly invasive technology with severe ethical implications [...]." Eventually, they discuss their inability to fully publish their research materials for copyright reasons. However, they provide transparency by retaining and sharing the entire code used for the pre-processing and analysis of their dataset.

Another illustration for a general ethical consideration is brought forward by Dambo et al. (2022) and their content analysis related to the Nigerian protests during the EndSARS movement. The ethical discussions in this qualitative analysis of Twitter data revolve around the potential concern for giving away precise user locations through geospatial data revealing coordinates of posted tweets. While the authors argue that Twitter privacy settings addressing data use settle the matter of informed consent, they also adhere to recommendations from previous scholarship, acknowledging difficulties when safeguarding the privacy of Twitter data.

In our sample of 476 publications, 31 instances (6.51%) explicitly reference ethical review processes. These references commonly involve securing approval from designated ethical oversight bodies, such as IRBs at the university level or analogous ethics committees at a national level. Moreover, some publications mention specific measures mandated by the IRB, such as the anonymization of data or the securing of informed consent. However, in our sample, most ethical mentions fall into the category of specific ethical procedures (15.50%), encompassing a range of different practices aimed at ensuring the well-being, privacy, and the rights of human participants. Our closer examination revealed the prevalence of different practices related to research ethics. Debriefing was relatively infrequent, observed in only 8 instances. Informed consent was more commonly reported, appearing in 28 publications. Anonymization techniques for data protection were explicitly addressed in 23 publications. Moreover, data pseudonymization techniques were mentioned in 6 publications. Notably, opt-in sampling methods were only explicitly described as such in 6 publications, reflecting varying approaches to participant recruitment and engagement with research protocols. The various ways in which ethical procedures are discussed within the investigated publications suggest that researchers in
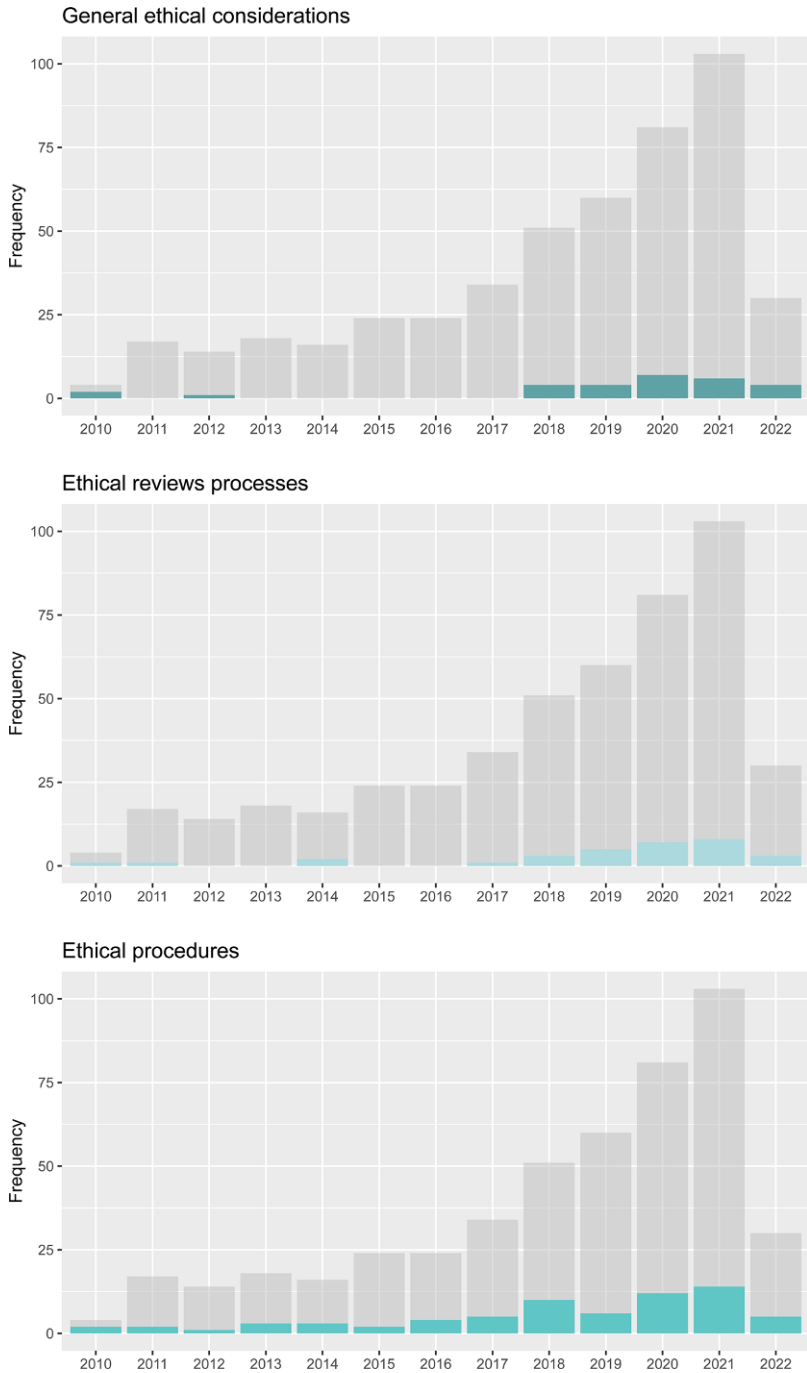
Fig. 2 Mentions of the different types of ethical considerations over time. Note: The *gray columns* in the background represent the full sample (*N* = 476), while the *turquoise columns* in the foreground visualize the absolute numbers of publications mentioning respective considerations per year

CCS prioritize different standardized ethical practices, including obtaining informed consent, choosing between opt-in and opt-out designs, and implementing debriefing processes.

We also assess the extent to which the explicit discussion of ethical considerations changed over time. Figure 2 illustrates the progression of ethical mentions across the three different types of ethical considerations in our sample. Notably, only a small fraction of studies mention, let alone discuss, any form of ethical considerations in their research. There are early mentions of general ethical considerations as well as ethical review processes, but they have only recently seen an increase in mentions over the last five years in our sample. A similar rationale applies to ethical procedures. Though ethical procedures have been somewhat consistently utilized over the years, there is a noticeable increase in mentions for this category over the last five years as well.

5 Discussion

Research ethics serve as a moral compass guiding the conduct of scientific investigations, ensuring that the pursuit of knowledge aligns with principles of integrity, responsibility, and respect (Artal and Rubenfeld 2017; Israel and Hay 2006). It involves a complex interplay of normative ethics, regulatory compliance, social values, and the involvement of various stakeholders, such as researchers, participants, academic institutions, publishers, funding agencies, and the general public (DuBois and Antes 2018; Lukito 2024). While this is true for all and particularly for human-subjects research, some ethical challenges are particularly pronounced for CCS due to the types of data and methods employed in the field. Although there were significant efforts in the recent past to establish a set of applicable guidelines for research ethics within the broader field of computational social science (CSS) (Engel et al. 2021; Haim 2023; Herschel and Miori 2017; Hosseini et al. 2022; Salganik 2019; Stegenga et al. 2024; Steinmann et al. 2016; Weinhardt 2020; Zwitter 2014), communication science in general (Fairfield and Shtein 2014; Roehse et al. 2023; Schlütz and Möhring 2018; Zwitter 2014), as well as relevant professional academic associations, such as the Association of Internet Researchers (AoIR), the American Psychological Association (APA), and the International Communication Association (ICA)[8], there is still a considerable need for discussions and guidance regarding specific ethical considerations in CCS. One key finding of our content analysis of the CCS literature is that for a large majority of CCS publications in our sample (89.50%), researchers opted not to share their data. This low rate of data sharing can be attributed to multiple factors. First, the prevalence of media data and data collection via database downloads and API access means that sharing data is often restricted by platform terms of service (ToS) or other contractual or license agreements. With regard to the (ethical) obligation of research being trans-

---

[8] See https://aoir.org/ethics/ for the Association of Internet Researchers, https://www.apa.org/ethics/code for the American Psychological Association, and https://www.icahdq.org/page/MissionStatement for the International Communication Association respectively.

parent and reliable (including being reproducible and replicable), an opinion piece by Davidson et al. (2023) has recently argued that "social media APIs threaten open science". This is an important aspect to consider for CCS researchers when making choices about data collection methods. In general, the dominance of media content data raises legal ethical challenges related to privacy and copyright issues. For that reason, CCS researchers may be particularly hesitant to share such data. The use of self-reported data through, for example, surveys introduces another layer of ethical considerations also regarding the disclosure of personal and possibly sensitive information. In sum, there are both legal as well as ethical considerations that might make CCS researchers hesitant to share their data. This assumption is supported by findings from a study by Akdeniz et al. (2023) on researchers' experiences with and attitudes towards sharing social media data which found that (perceived) legal and ethical challenges are among the main reasons for not sharing such data.

Of course, discussions about data sharing and its ethical implications in the social sciences are not new. In fact, they have been going on for multiple decades, gaining increasing prominence in recent years (Curty et al. 2016; Zenk-Möltgen et al. 2018). While the concept of data sharing has existed for a while (Sieber 1991), it only recently became more common for researchers to share their data. This recent trend of increased data sharing in CCS can be attributed to a confluence of factors that have collectively shifted the field more towards the principles of openness and transparency. One driver might be the broader cultural shift towards embracing open science practices within the scientific community at large (Peterson and Panofsky 2023) as well as the social sciences and communication science in particular (Dienlin et al. 2021). As scholarly practices and communication evolves, there is a growing recognition of the benefits of making research outputs, including data, openly accessible. Researchers in CCS, also influenced by these changing norms, may be more inclined to share their data to contribute to the collective knowledge base, facilitate reproducibility, and ultimately enhance the credibility of their work. Besides norms in the field, institutional and funder requirements play a pivotal role in shaping sharing practices (Pham-Kanter et al. 2014). Funding agencies and academic institutions are increasingly emphasizing the importance of data sharing as a condition for receiving grants or institutional support (Anger et al. 2022). Journals, as gatekeepers of scholarly communication, have also increasingly started to incorporate data sharing policies (Piwowar and Chapman 2008; Vasilevsky et al. 2017). As these policies become more prevalent and influential, researchers in CCS may be motivated to align their practices with these expectations, leading to a gradual increase in data sharing within the field.

Another interesting finding from our analysis related to data sharing is that among the small share of researchers who share their data, there is a clear preference for public repositories. These platforms offer a reputable venue for data sharing and provide clear guidelines for usage, such as licensing, ethical data management practices, etc.[9] Researchers are likely attracted to these platforms due to their credibility, user-friendly interfaces, and transparent policies, reflecting a collective effort towards ethical conduct in scientific research (Rockhold et al. 2019).

---

[9] See, for example, https://help.osf.io.

In our analysis of explicit mentions or discussions of ethical considerations within the CCS literature, we discovered a multifaceted examination of ethical issues, providing insights that can be interpreted through both deontological and consequentialist perspectives. In particular, our analysis of ethical considerations in the publications revealed several noteworthy patterns. First, only a small fraction (5.88%) of the 476 publications explicitly addressed general ethical considerations. These considerations covered a broad spectrum of ethical subjects, ranging from committing to upholding user privacy to maneuvering regulatory frameworks, and confronting potential biases in their studies. In cases where broader ethical considerations were addressed, the focus was more on a deontological perspective, prioritizing universal values such as respect, fairness, and integrity. The low prevalence of explicit discussions of research ethics in the CCS literature is surprising. However, it is in line with findings from a recent study by Fiesler et al. (2024) which found that, using a very liberal definition, fewer than 200 out of 700 studies using data from Reddit mentioned anything related to research ethics.

Furthermore, only a small fraction (6.51%) of the publications explicitly mentioned the approval from institutional review boards or similar institutions. Again, this mirrors the findings by Fiesler et al. (2024) who discovered that many authors of studies using Reddit data argue that these data are public and, hence, IRB approval is not required. The inclusion of ethical review processes can be interpreted from both deontological and consequentialist perspectives. From a deontological standpoint, the emphasis on ethical review processes underscores a commitment to upholding general ethical standards and principles. From a consequentialist perspective, ethical review processes carry implications focused on the case-by-case evaluation and weighing of outcomes and consequences. By subjecting research proposals to ethical scrutiny, the intention is to prevent and minimize any unforeseen potential harm to participants. In this sense, the mention of ethical review processes aligns with a consequentialist perspective by aiming to achieve positive consequences, such as prioritizing the benefit of the research or safeguarding the well-being of participants.

In contrast, a larger proportion (15.5%) of publications mentioned specific ethical procedures or protocols, covering a range of practices to ensure the well-being, privacy, and rights of human participants. The diversity of ethical procedures requires a more fine-grained analysis to understand which protocol takes on or prioritizes a deontological or consequentialist perspective. For instance, the practice of debriefing typically aligns more closely with a consequentialist ethical framework. Debriefing involves providing participants with transparent information and addressing any concerns or questions after their involvement in a study, aiming to mitigate potential harm and reduce the risk of negative consequences. In this case, debriefing can be understood to focus on achieving positive outcomes and minimizing harm, which aligns with consequentialist principles, evaluating actions based on their overall consequences rather than adhering to a set of predefined principles. Informed consent, on the other hand, can be understood from both deontological and consequentialist perspectives. From a deontological standpoint, informed consent aligns with the principles of autonomy and the ethical duty to uphold individual rights. From a consequentialist perspective, informed consent is warranted as it helps in achieving or increasing ethically positive outcomes. Ensuring that participants are

adequately informed about the research aims and their personal rights is a means to prevent potential harm and enhance comprehensibility. Informed consent can, thus, be seen as incorporating elements of both deontological ethics and consequentialist frameworks. The practices of data anonymization and pseudonymization align more closely with a deontological ethical framework due to their emphasis on adhering to ethical principles and rules for protecting individual privacy and confidentiality. Options for opt-in and opt-out are more in line with a consequentialist ethical framework. The choice between these designs is often at least in parts driven by considerations related to the potential outcome on participant response rates and data quality. Overall, the diversity in ethical procedures and discussions suggests that CCS researchers prioritize a flexible and context-specific stance to address ethical considerations, particularly considering the complex methodologies and data environments they work in. It has to be noted, however, that there are also some blind spots in ethical discussions within the CCS literature. For example, ethical implications of the use of substantial computational resources has, so far, not been discussed in the analyzed CCS studies.

In all three ethical consideration categories, we found a recent increase of mentions. This trend over the last five years could be attributed to several potential factors. First, there may be an increasing emphasis on ethical training and awareness among researchers and institutions, leading to greater attention to ethical procedures in research studies. Additionally, the evolving landscape of regulations and guidelines concerning research ethics may have prompted researchers to incorporate ethical considerations more diligently into their studies (Lukito 2024). Moreover, heightened scrutiny and public awareness of ethical issues in research, particularly in fields involving human participants, could also contribute to the observed recent increase in mentions of ethical considerations across categories.

Overall, our analyses illustrate that ethical discussion and decision-making in CCS revolves around a dynamic interplay between deontological and consequentialist ethical considerations. This underscores the necessity for a flexible and context-specific approach in navigating the ethical dimensions of CCS research, as has been highlighted recently by several scholars (e.g., Haim 2023; Salganik 2019; Schlütz and Möhring 2018). Through balancing different ethical perspectives and obligations, mirroring the complexity of data and methods in the field, this development and refinement of ethical guidelines can aid researchers in designing and conducting their studies, in sharing the products of their research, but also in reviewing other CCS endeavors. While there often is no single correct answer to ethical questions, being aware of potentially conflicting principles and values and explicitly addressing the underlying ethical frameworks certainly contributes to guiding CCS researchers in hands-on ethical decision-making and, thus, working towards better research practices and more consensus on these practices (Lukito 2024).

6 Appendix

6.1 Full list of journals in the sample

- Communication Methods and Measures
- Communication Monographs
- Communication Research
- Comunicar
- Digital Journalism
- European Journal of Communication
- Human Communication Research
- Information Communication & Society
- International Journal of Advertising
- International Journal of Communication
- International Journal of Press-Politics
- Journal of Advertising
- Journal of Broadcasting & Electronic Media
- Journal of Communication
- Journal of Computer-mediated Communication
- Journal of Health Communication
- Journalism
- Journalism & Mass Communication Quarterly
- Journalism Practice
- Journalism Studies
- Management Communication Quarterly
- Mass Communication and Society
- Media and Communication
- Media Culture & Society
- Media Psychology
- Mobile Media & Communication
- New Media & Society
- Political Communication
- Public Opinion Quarterly
- Public Relations Review
- Public Understanding of Science
- Science Communication
- Social Media + Society

References

Akdeniz, E., Borschewski, K.E., Breuer, J., & Voronin, Y. (2023). Sharing social media data: the role of past experiences, attitudes, norms, and perceived behavioral control. *Frontiers in Big Data*, *5*, 971974. https://doi.org/10.3389/fdata.2022.971974.

Alter, G., & Gonzalez, R. (2018). Responsible practices for data sharing. *American Psychologist*, *73*(2), 146–156. https://doi.org/10.1037/amp0000258.

Anger, M., Wendelborn, C., Winkler, E.C., & Schickhardt, C. (2022). Neither carrots nor sticks? Challenges surrounding data sharing from the perspective of research funding agencies—A qualitative expert interview study. *PLOS ONE*, *17*(9), e273259. https://doi.org/10.1371/journal.pone.0273259.

Artal, R., & Rubenfeld, S. (2017). Ethical issues in research. *Best Practice & Research Clinical Obstetrics & Gynaecology*, *43*, 107–114. https://doi.org/10.1016/j.bpobgyn.2016.12.006.

van Atteveldt, W., & Peng, T.-Q. (2018). When communication meets computation: opportunities, challenges, and pitfalls in computational communication science. *Communication Methods and Measures*, *12*(2–3), 81–92. https://doi.org/10.1080/19312458.2018.1458084.

van Atteveldt, W., Strycharz, J., Trilling, D., & Welbers, K. (2019). Towards open computational communication science: a practical roadmap for reusable data and code. *International Journal of Communication*, *13*.

Borgman, C.L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, *63*(6), 1059–1078. https://doi.org/10.1002/asi.22634.

Breuer, J., Bishop, L., & Kinder-Kurlanda, K. (2020). The practical and ethical challenges in acquiring and sharing digital trace data: negotiating public-private partnerships. *New Media & Society*, *22*(11), 2058–2080. https://doi.org/10.1177/1461444820924622.

Breuer, J., Weller, K., & Kinder-Kurlanda, K. (2023). The role of participants in online privacy research: ethical and practical considerations. In S. Trepte & P.K. Masur (Eds.), *The Routledge handbook of privacy and social media* (pp. 314–323). Routledge. https://www.taylorfrancis.com/chapters/oa-edit/10.4324/9781003244677-35/role-participants-online-privacy-research-johannes-breuer-katrin-weller-katharina-kinder-kurlanda.

Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J., & Hudson, M. (2020). The CARE principles for indigenous data governance. *Data Science Journal*, *19*, 43. https://doi.org/10.5334/dsj-2020-043.

Curty, R., Yoon, A., Jeng, W., & Qin, J. (2016). Untangling data sharing and reuse in social sciences. *Proceedings of the Association for Information Science and Technology*, *53*(1), 1–5. https://doi.org/10.1002/pra2.2016.14505301025.

Dambo, T.H., Ersoy, M., Auwal, A.M., Olorunsola, V.O., & Saydam, M.B. (2022). Office of the citizen: a qualitative analysis of twitter activity during the Lekki shooting in Nigeria's #endSARS protests. *Information, Communication & Society*, *25*(15), 2246–2263. https://doi.org/10.1080/1369118X.2021.1934063.

Davidson, B.I., Wischerath, D., Racek, D., Parry, D.A., Godwin, E., Hinds, J., Van Der Linden, D., Roscoe, J.F., Ayravainen, L., & Cork, A.G. (2023). Platform-controlled social media APIs threaten open science. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-023-01750-2.

Dienlin, T., Johannes, N., Bowman, N.D., Masur, P.K., Engesser, S., Kümpel, A.S., Lukito, J., Bier, L.M., Zhang, R., Johnson, B.K., Huskey, R., Schneider, F.M., Breuer, J., Parry, D.A., Vermeulen, I., Fisher, J.T., Banks, J., Weber, R., Ellis, D.A., & De Vreese, C. (2021). An agenda for open science in communication. *Journal of Communication*, *71*(1), 1–26. https://doi.org/10.1093/joc/jqz052.

DuBois, J.M., & Antes, A.L. (2018). Five dimensions of research ethics: a stakeholder framework for creating a climate of research integrity. *Academic medicine : Journal of the Association of American Medical Colleges,*, *93*(4), 550–555. https://doi.org/10.1097/ACM.0000000000001966.

Emmerich, N. (2020). A professional ethics for researchers? In R. Iphofen (Ed.), *Handbook of research ethics and scientific integrity* (pp. 751–767). Springer. https://doi.org/10.1007/978-3-030-16759-2_34.

Engel, U., Quan-Haase, A., Liu, S. X., & Lyberg, L. E. (2021). *Theory, case studies and ethics*. Handbook of computational social science, Vol. 1. Routledge.

Eslami, M., Vaccaro, K., Karahalios, K., & Hamilton, K. (2017). "Be careful; things can be worse than they appear": understanding biased algorithms and users' behavior around them in rating platforms. *Proceedings of the International AAAI Conference on Web and Social Media*, *11*(1), Article 1. https://doi.org/10.1609/icwsm.v11i1.14898.

Fairfield, J., & Shtein, H. (2014). Big data, big problems: emerging issues in the ethics of data science and journalism. *Journal of Mass Media Ethics*, *29*(1), 38–51. https://doi.org/10.1080/08900523.2014.863126.

Fecher, B., Friesike, S., & Hebing, M. (2015). What drives academic data sharing? *PLOS ONE*, *10*(2), e118053. https://doi.org/10.1371/journal.pone.0118053.

Fiesler, C., Zimmer, M., Proferes, N., Gilbert, S., & Jones, N. (2024). Remember the human: a systematic review of ethical considerations in Reddit research. *Proceedings of the ACM on Human-Computer Interaction*, *8*, 1–33. https://doi.org/10.1145/3633070.

Geise, S., & Waldherr, A. (2021). *Computational communication science: Lessons from working group sessions with experts of an emerging research field [Preprint]*. SocArXiv. https://doi.org/10.31235/osf.io/hkc5a.

Haim, M. (2023). *Computational Communication Science: Eine Einführung. [Computational Communication Science: An Introduction]*. Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-40171-9.

Halavais, A. (2019). Overcoming terms of service: A proposal for ethical distributed research. *Information, Communication & Society*, *22*(11), 1567–1581. https://doi.org/10.1080/1369118X.2019.1627386.

Herschel, R., & Miori, V. M. (2017). Ethics & big data. *Technology in Society*, *49*, 31–36. https://doi.org/10.1016/j.techsoc.2017.03.003.

Hilbert, M., Barnett, G., Blumenstock, J., Contractor, N., Diesner, J., Frey, S., González-Bailón, S., Lamberso, P., Pan, J., Shen, C., Smaldino, P. E., Van Atteveldt, W., Waldherr, A., Zhang, J., & Zhu, J. J. H. (2019). Computational communication acience: a methodological catalyzer for a maturing discipline. *International Journal of Communication*, *13*, 3912–3934.

Hosseini, M., Wieczorek, M., & Gordijn, B. (2022). Ethical Issues in social science research employing big data. *Science and Engineering Ethics*, *28*(3), 29. https://doi.org/10.1007/s11948-022-00380-7.

Iphofen, R. (2020). An introduction to research ethics and scientific integrity. In R. Iphofen (Ed.), *Handbook of research ethics and scientific integrity* (pp. 3–13). Springer. https://doi.org/10.1007/978-3-030-16759-2_62.

Israel, M. (2015). *Research ethics and integrity for social scientists: beyond regulatory compliance* (2nd edn.). SAGE. https://doi.org/10.4135/9781473910096.

Israel, M., & Hay, I. (2006). *Research ethics for social scientists: between ethical conduct and regulatory compliance*. SAGE.

Jürgens, P., Meltzer, C. E., & Scharkow, M. (2022). Age and gender representation on German TV: a longitudinal computational analysis. *Computational Communication Research*. https://doi.org/10.5117/CCr2022.1.005.JUrg.

Kirilova, D., & Karcher, S. (2017). Rethinking data sharing and human participant protection in social science research: applications from the qualitative realm. *Data Science Journal*, *16*, 43–43. https://doi.org/10.5334/dsj-2017-043.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Life in the network: the coming age of computational social science. *Science*, *323*(5915), 721–723. https://doi.org/10.1126/science.1167742.

Lazer, D., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M. J., Strohmaier, M., Vespignani, A., & Wagner, C. (2020). Computational social science: obstacles and opportunities. *Science*, *369*(6507), 1060–1062. https://doi.org/10.1126/science.aaz8170.

Leidig, E. C. (2019). Immigrant, nationalist and proud: a twitter analysis of Indian diaspora supporters for Brexit and Trump. *Media and Communication*, *7*(1), 77–89. https://doi.org/10.17645/mac.v7i1.1629.

Leslie, D. (2023). The ethics of computational social science. In *Handbook of computational social science for policy* (pp. 57–104). Cham: Springer.

Longo, D. L., & Drazen, J. M. (2016). Data sharing. *New England Journal of Medicine*, *374*(3), 276–277. https://doi.org/10.1056/NEJMe1516564.

Lukito, J. (2024). Platform research ethics for academic research. Center for Media Engagement. https://mediaengagement.org/research/platform-research-ethics. Accessed 13 May 2024.

McCosker, A. (2018). Engaging mental health online: insights from beyondblue's forum influencers. *New Media & Society*, *20*(12), 4748–4764. https://doi.org/10.1177/14614448187843.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, *54*(6), 115:1–115:35. https://doi.org/10.1145/3457607.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979). The Belmont report: Ethical principles and guidelines for the protection of human subjects of research. https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html. Accessed 13 May 2024.

Peterson, D., & Panofsky, A. (2023). Metascience as a scientific social movement. *Minerva*, *61*(2), 147–174. https://doi.org/10.1007/s11024-023-09490-3.

Pham-Kanter, G., Zinner, D. E., & Campbell, E. G. (2014). Codifying collegiality: recent developments in data sharing policy in the life sciences. *PLOS ONE*, *9*(9), e108451. https://doi.org/10.1371/journal.pone.0108451.

Piwowar, H., & Chapman, W. (2008). A review of journal policies for sharing research data. *Nature Precedings*. https://doi.org/10.1038/npre.2008.1700.1.

Rains, S. A., Keating, D. M., Banas, J. A., Richards, A. S., & Palomares, N. A. (2020). The state and evolution of communication research: a topic modeling analysis of 20,000 journal article abstracts from 1918–2015. *Computational Communication Research*, *2*(2), 203–234. https://doi.org/10.5117/CCR2020.2.004.RAIN.

Rockhold, F., Bromley, C., Wagner, E. K., & Buyse, M. (2019). Open science: the open clinical trials data journey. *Clinical Trials (London, England)*, *16*(5), 539–546. https://doi.org/10.1177/1740774519865512.

Roehse, E. M., Möhring, W., Zillich, A. F., Schlütz, D., & Link, E. (2023). Forschungsethische Praxis in der Kommunikations- und Medienwissenschaft – ein Einblick in die Sicht der Forschenden. *Publizistik*, *68*(4), 459–489.

Salganik, M. J. (2019). *Bit by bit: social research in the digital age*. Princeton University Press.

Schlütz, D., & Möhring, W. (2018). Between the devil and the deep blue sea: negotiating ethics and method in communication research practice. *SCM Studies in Communication and Media*, *7*(1), 31–58. https://doi.org/10.5771/2192-4007-2018-1-31.

Shah, D. V., Cappella, J. N., & Neuman, W. R. (2015). Big data, digital media, and computational social science: possibilities and perils. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 6–13. https://doi.org/10.1177/0002716215572084.

Siapera, E., Boudourides, M., Lenis, S., & Suiter, J. (2018). Refugees and network publics on twitter: networked framing, affect, and capture. *Social Media + Society*. https://doi.org/10.1177/205630511876443.

Sieber, J. E. (1991). Openness in the social sciences: sharing data. *Ethics & Behavior*, *1*(2), 69–86. https://doi.org/10.1207/s15327019eb0102_1.

Stegenga, S. M., Steltenpohl, C. N., Lustick, H., Meyer, M. S., Renbarger, R., Standiford, R. L., & Lee, L. E. (2024). Qualitative research at the crossroads of open science and big data: ethical considerations. *Social and Personality Psychology Compass*, *18*(1), e12912. https://doi.org/10.1111/spc3.12912.

Steinmann, M., Matei, S. A., & Collmann, J. (2016). A theoretical framework for ethical reflection in big data research. In J. Collmann & S. A. Matei (Eds.), *Ethical reasoning in big data: an exploratory analysis* (pp. 11–27). Springer. https://doi.org/10.1007/978-3-319-28422-4_2.

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2021). The ethics of algorithms: key problems and solutions. In L. Floridi (Ed.), *Ethics, governance, and policies in artificial intelligence* (pp. 97–123). Springer. https://doi.org/10.1007/978-3-030-81907-1_8.

Urman, A., & Katz, S. (2022). What they do in the shadows: examining the far-right networks on telegram. *Information, Communication & Society*, *25*(7), 904–923. https://doi.org/10.1080/1369118X.2020.1803946.

Vasilevsky, N. A., Minnier, J., Haendel, M. A., & Champieux, R. E. (2017). Reproducible and reusable research: are journal data sharing policies meeting the mark? *PeerJ*, *5*, e3208. https://doi.org/10.7717/peerj.3208.

Waldherr, A., Geise, S., Mahrt, M., Katzenbach, C., & Nuernbergk, C. (2021). Toward a stronger theoretical grounding of computational communication science: how macro frameworks shape our research agendas. *Computational Communication Research*, *3*(2), 1–28.

Walter, N., Cody, M. J., & Ball-Rokeach, S. J. (2018). The ebb and flow of communication research: seven decades of publication trends and research priorities. *Journal of Communication*, *68*(2), 424–440. https://doi.org/10.1093/joc/jqx015.

Weinhardt, M. (2020). Ethical issues in the use of big data for social research. *Historical Social Research / Historische Sozialforschung*, *45*(3), 342–368.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18.

Zenk-Möltgen, W., Akdeniz, E., Katsanidou, A., Naßhoven, V., & Balaban, E. (2018). Factors influencing the data sharing behavior of researchers in sociology and political science. *Journal of Documentation*, *74*(5), 1053–1073. https://doi.org/10.1108/JD-09-2017-0126.

Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S. P., Goodman, A., Hollander, R., Koenig, B. A., Metcalf, J., Narayanan, A., Nelson, A., & Pasquale, F. (2017). Ten simple rules for responsible big data research. *PLoS computational biology*, *13*(3), e1005399. https://doi.org/10.1371/journal.pcbi.1005399.

Zwitter, A. (2014). Big data ethics. *Big Data & Society*, *1*(2), 2053951714559253. https://doi.org/10.1177/2053951714559253.

Philipp Knöpfle is a research associate at the Department of Media and Communication at Ludwig-Maximilian University of Munich (LMU). He works at the Computational Communication Research unit under the supervision of Professor Haim. His primary research interests encompass meta-science, open science, and computational methods.

Mario Haim is a full professor for Communication Science with a focus on Computational Communication Research at the Department of Media and Communication at LMU Munich, Germany. His research interests circle around algorithmic influences, such as in political communication, journalism, health communication, or media use, as well as on (computational) methods and meta-science. More information: https://haim.it.

Johannes Breuer is a senior researcher and leader of the team Digital Society Observatory in the Department Computational Social Science at GESIS—Leibniz Institute for the Social Sciences in Cologne, Germany. His research interests include the use and effects of digital media, digital trace data and computational methods, as well as open science and meta-science. More information: https://www.johannesbreuer.com.