Ethics in Computational Communication Science: Between values and perspectives

Knöpfle, Philipp; Haim, Mario; Breuer, Johannes

Preprint / Preprint

Funded by the German Research Foundation (DFG) - Grant No. 3473 within the DFG Priority Program META-REP (SPP 2317)

Preprint – currently under review

Ethics in Computational Communication Science: Between values and perspectives

Authors: Philipp Knöpfle[1a], Mario Haim[1], Johannes Breuer[2,3]

[1] Ludwig-Maximilians-Universität, München, Deutschland;
[2] GESIS - Leibniz-Institut für Sozialwissenschaften, Köln, Deutschland;
[3] Center for Advanced Internet Studies (CAIS), Bochum, Deutschland
[a] Corresponding author

Contact Details:

Philipp Knöpfle
Ludwig-Maximilians-Universität München
Institut für Kommunikationswissenschaft und Medienforschung
Akademiestr. 7
80799 München, Deutschland
E-Mail: Philipp.Knoepfle@ifkw.lmu.de
ORCID-ID: 0000-0003-3575-1229

Prof. Dr. Mario Haim
Ludwig-Maximilians-Universität München
Institut für Kommunikationswissenschaft und Medienforschung
Akademiestr. 7
80799 München, Deutschland
E-Mail: Haim@ifkw.lmu.de
ORCID-ID: 0000-0002-0643-229

Dr. Johannes Breuer
GESIS - Leibniz Institute for the Social Sciences,
Unter Sachsenhausen 6-8
50667 Köln, Deutschland
E-Mail: Johannes.Breuer@gesis.org
ORCID-ID: 0000-0001-5906-7873

Keywords: Computational Communication Science, Content Analysis, Open Science, Research Ethics.

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. Rooted in two opposing lines of philosophical arguments — deontology and consequentialism — we argue that CCS research is particularly difficult to be projected onto this ethical spectrum. Our study aims to empirically assess the nature and prevalence of provided arguments and influencing factors for ethical decision-making in CCS research. Through a manual content analysis of 476 CCS studies, sampled from a corpus of 22,375 collected communication science articles, we shed light on data sharing practices and ethical reflections of CCS researchers. Findings indicate large room for maneuver. The majority of studies (89.50%) chose not to share their data, while 6.93% chose to share their data either full or partially. Only 5.88% of studies explicitly addressed general ethical considerations. Ethical review processes were mentioned by 6.51% of studies, with the majority pointing at ethical procedures such as obtaining informed consent, data anonymization measures, or debriefing. This suggests that researchers in CCS prioritize context-specific ethical procedures in the absence of field-specific standards, emphasizing the importance of flexibility in addressing ethical considerations.

Funding: This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) [Grant No. 3473 within the DFG Priority Program META-REP (SPP 2317)].

1. Introduction[1]

Computational communication science (CCS) is a field characterized by the collection and analysis of large and complex datasets with the goal of exploring various facets of human communication and testing communication theories (van Atteveldt & Peng, 2018). In their work, CCS researchers face a multitude of ethical challenges related to the collection of personal/sensitive or copyrighted media data, the analysis of social behavior, the derivation of personal attributes through algorithmic estimations, and the archiving and sharing of research data. Navigating the associated ethical questions and balancing diverse – and potentially conflicting – values and goals, hence, is an important undertaking for CCS researchers.

Broadly speaking, there are two foundational ethical frameworks that are relevant for empirical research 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 the adherence to specific norms and fundamental values which 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 frameworks (Schlütz & Möhring, 2018), the discrepancy between a-priori decision-making guided by fundamental values (deontology) and a-posteriori consideration of expected outcomes (consequentialism) are likely also the cause of diverging answers to ethical questions within the CCS community.

To illustrate what such differences can mean in practice, consider a study in which browser histories of participants are collected. From a deontological standpoint, researchers may see obtaining informed consent and maximizing data privacy (ideally through data anonymization and restricted access) as essential conditions (*conditio sine qua non*) for research. By contrast, from a consequentialist standpoint, the focus would be on maximizing the scientific knowledge and value obtained from the gathered data. This involves carefully assessing the potential drawbacks and advantages of data collection and sharing, with a heightened emphasis on prioritizing openness. While the latter perspective might thus suggest to openly share collected data, the former perspective strongly opposes this position.

Put more generally, balancing between privacy and transparency while maximizing the benefits of research and minimizing potential harm are critical in addressing ethical challenges in all of empirical research. Particular to CCS, then, are the size and complexity of datasets that make it close-to-impossible to manually verify and guarantee an appropriate ethical approach. As it is exactly these datasets that also allow the most far-reaching derivations, for example in the form of predictive models, it is crucial to understand how the field currently discusses its ethical challenges and how the thereby taken ethical perspectives translate into applied research practices.

To explore the prevalence of issues related to ethical questions around data sharing 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

[1] All data and code underlying the results presented in this study are available from OSF (The OSF.io link is blackened to maintain the anonymity of the blind review process).

science journals published between 2010 and 2021. Our content analysis focuses on types of data, data collection methods, anonymization techniques, methodologies, and researchers' data sharing practices. More specifically, we complement a quantitative content analysis of code and data sharing practices with a qualitative content analysis of mentioned ethical reflections. This includes the discussion of important ethical aspects such as informed consent, privacy protection, ethical review and approval from relevant institutional review boards, and references to underlying/consulted ethical guidelines or frameworks (Leslie, 2023). Our analysis aims to provide insights into the argumentative foundations of ethical reflections employed in CCS, their relation to underlying ethical frameworks, and the associated conceptualizations of CCS.

2. Theoretical framework

2.1 Complexities of CCS research

In the field of CCS, researchers engage in the development and application of computational methods to investigate various aspects of human communication and test theories from communication science, particularly in the context of digital and online communication (Hilbert et al., 2019). 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 their processing and analysis (van Atteveldt & 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 also (semi-)automated analyses of audio, image, and video data (Hilbert et al., 2019). Accordingly, CCS researchers develop and/or utilize a heterogeneous set of software tools, navigating the choices between commercial and open-source solutions, often employing different tools for similar methods, and potentially encountering issues related to software compatibility and long-term sustainability (van Atteveldt et al., 2019).

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 & 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 framework 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 philosophical underpinnings guiding the study of communication in the digital age.

It seems reasonable that many of the ethical considerations applicable to traditional communication research are also relevant to CCS research, while, at the same time, there are some ethical questions particularly pronounced when associated with CCS. One prominent challenge revolves around the collection, use, and sharing of personal, sensitive or copyrighted media data (e.g., digital trace data) for research purposes. The collection of such data raises ethical considerations regarding privacy, informed consent, and the responsible use of information (Lazer et al., 2020). Significantly, ethical dilemmas emerge not only in the collection and sharing of research data but also during the data analysis phase. In this stage, researchers commonly delve into the utilization, interaction, and communication behaviors, often extracting 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).

The range of ethical considerations faced by CCS researchers expands when addressing essential open science practices, including the sharing of code, software, and data. Sharing data and other research products (e.g., code and software) is driven by the ethos of collaboration and transparency inherent to scientific inquiry (Longo & Drazen, 2016). Importantly, however, these resources must be shared and utilized responsibly to mitigate the risk of misuse or unintended consequences. Researchers must 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 materials in science in general and in CCS in particular where data-intensive use as well as the development of novel approaches is common (Alter & 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 within the CCS community. The interdisciplinary nature of CCS adds further complexity to these ethical discourses and conflicts. Hence, it is important for CCS researchers to actively engage in discussions of research ethics and developing procedures to adequately address current and future ethical questions in their work.

2.2 Deontological vs. consequentialist perspectives in 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 frameworks, 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). For example, if a CCS study collects social media data from individuals, a deontological approach emphasizes the importance of consent and data protection as essential requirements (*conditio sine qua non*). Specifically, this could mean that researchers only use data from individuals who provided informed consent, implementing thorough and early anonymization measures to the data, and limiting access to it.

Consequentialism in the context of CCS, on the other hand, translates to evaluating the outcomes and consequences of research decisions. It involves assessing potentially negative with positive impacts (harms and benefits) of decisions. To illustrate, revisiting the previous example, if the researchers consider the collection of social media data from millions of participants, only using data from those that provide explicit consent may become infeasible. Notably, from a legal perspective, within the General Data Protection Regulation (GDPR) in the EU, informed consent is only one of several legal bases for the processing of personal data. Additionally, such large data collections may lead to the emergence of estimable patterns that could reveal sensitive information even beyond participants' consent (and also despite data anonymization efforts). Such results could potentially be disseminated in the form of practical statistical models. From a consequentialist perspective, publishing such insights on individual behavior might thus raise concerns about infringing informed consent agreements and privacy norms. These challenges underscore the need for new perspectives on privacy such as those proposed in the concept of predictive privacy (Mühlhoff, 2021).[2]

When talking about ethics, one of the first things that researchers in the social sciences think of is the protection of participants and their data. However, there also are other ethical considerations that matter for the practical decisions that researchers need to make. Depending on one's perspective, some of those may also be considered as ethical obligations. As outlined before, whether or to what degree these are considered obligations depends on the evaluation of different norms. Besides avoidance of harm, another value that may be considered also as an ethical obligation is being transparent and open. For research to meet these norms and, thus, increase the trustworthiness of its findings, sharing research materials and especially the underlying data is an important aspect. Deciding whether or in what way to share research data necessitates a careful examination of the potential ethical implications of making the data available (Borgman, 2012). This act of sharing research data is integral to promoting transparency, reproducibility as well as replicability, and collaboration between researchers. 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).

However, data sharing introduces significant concerns related to privacy, confidentiality, and the potential misuse of sensitive information (Kirilova & Karcher, 2017). From a deontological perspective, researchers have the obligation to respect the autonomy and rights of research participants, including the protection of personal and sensitive information. In this context, data

[2] The concept of predictive privacy involves ethical considerations in the use of predictive analytics, such as ML, where algorithms and statistical models transition from making inferences about populations to specific predictions for individuals, raising concerns about the ethical implications of projecting aggregate statistical knowledge onto individual cases in situations of incomplete information.

sharing may be decided against or restricted to maximize privacy. In contrast, from a consequentialist perspective, one could argue that the benefits of sharing data (e.g. scientific advancement, reproducibility, and collaboration) outweigh potential harms. The consequentialist viewpoint, hence, may prioritize the greater good for society, the scientific community, and future research over individual privacy concerns. Researchers must navigate the tension between the ethical obligation to protect privacy and the potential positive outcomes of data sharing for scientific progress and knowledge dissemination.

In conclusion, the complex interplay between deontological and consequentialist perspectives in the research practices within CCS research underscores the inherent complexity in ethical decision-making in CCS. In most cases, there is no single best solution. Contemporary CCS research ethics, hence, frequently incorporate elements from both frameworks (Schlütz & Möhring, 2018). As such, various contemporarily applicable frameworks have been provided to translate the seminal ethical perspectives into CCS best practices, thereby drawing on "a consistent, morally solid, and easy-to-apply blend" (Salganik, 2019, p. 303). For example, while (Salganik, 2019), in his four-principles approach, puts an emphasis on human subjects and various past US-based ethical misconducts, (Haim, 2023) applies a more EU-inspired take and suggests to also ethically consider computational resources and data biases. Applicable to both approaches, the deontological roots highlight a clear set of ex-ante decision rules based on explicit norms whereas consequentialist considerations focus more on evaluating the expected outcomes and impacts of research decisions on a situational case-by-case basis.

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 framework explicitly or implicitly are referenced or built upon by researchers.

3. Our study

The aim of our study is to conduct an empirical assessment of the nature and prevalence of factors that can affect ethical decision-making in CCS research. As outlined above, we assumed that most of those would be related to features of the study design and the handling (collection, processing, analysis, and sharing) of research data. We conducted this assessment by means of a systematic review and quantitative content analysis of published CCS literature. To also get a more in-depth picture of the ethical considerations and underlying ethical frameworks, we complemented the latter with a qualitative analysis looking at how ethical reasoning and decision-making is discussed in the existing CCS literature.

3.1 Data collection

Since CCS is a rapidly evolving field and CCS articles are published across different journals in the field of communication science, we first need to create a database containing CCS articles. The classification of communication publications as belonging to the area of CCS is challenging for several reasons. One is the inherently interdisciplinary nature of the field, but also its evolving terminology, diverse research topics, and the common combination of computational methods with other (traditional) methods of empirical communication research (e.g., surveys or interviews). As mentioned earlier, the twofold understanding of CCS as both a methodological toolkit and a paradigmatic perspective further complicate a uniform recognition of suitable studies. The rather abstract nature of the term 'computational' and the absence of standardized classification criteria add to this challenge, necessitating a thorough examination of article content, methodologies, and research objectives for an accurate assessment. Notably, the computational component(s) in a study can span a variety of aspects (Hilbert et al., 2019). These aspects encompass data collection instruments, such as web scraping, API queries, data donations, and tracking tools, along with data analysis techniques, such as ML, NLP, and network analysis (Geise & Waldherr, 2021). Additional computational elements within CCS encompass algorithmic methodologies, which may involve (complex) experimental designs, the generation of research stimuli, and the incorporation of simulations and agent-based modeling. In view of this diversity, we categorized an article as computational when it incorporated computational elements within its study in either the research or experimental design, data collection, data processing, or data analysis.

We thus employed a three-step identification approach, starting with collecting a total of 22,375 English-language communication science articles available in the Clarivate Web of Science database in April 2022. For this, we focused on the top communication journals listed in the top 50 communication science journals according to Google Scholar[3]. We collected all articles from the thereby resulting list of 34 journals.[4] Our sampling period spans the time from January 1, 2010, to December 31, 2021. We chose to commence our data sample in 2010, around one year after the influential paper by Lazer et al. (2009) was published, which significantly catalyzed the development of computational social science (CSS) and, subsequently, also CCS. Our analysis period ends in December 2021, encompassing pre-printed online publications of articles which have been officially fully published in 2022.

Second, we calculated a co-occurrence network model on data extracted from titles, abstracts, and author-tagged keywords of a manually pre-selected CCS corpus consisting of roughly 150 articles. The objective of this step was to derive a specific set of CCS-related keywords that enable us to filter the extensive collection of communication science articles down to explicit CCS research. The keywords identified with this approach were the following: automated content analysis, computational, text mining, automated text analysis, word2vec, doc2vec, corpus analysis, latent semantic analysis, natural language processing, sentiment analysis, topic model, community detection, machine learning, supervised, unsupervised, image detection, network analysis, agent-based model, network data, digital trace data, digital behavioral data, social media data, text data, digital media, web scraping, data mining, API.

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

[4] Our list of journals includes: 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, and Social Media + Society.

Subsequently, we searched the abstract, title, and author-tagged keywords of the initial 22,375 articles for this set of CCS-specific keywords, resulting in a corpus of 6,556 articles.

Third, we manually classified approximately 1,000 out of the previous 6,556 studies as either computational or non-computational, to create a training dataset for a supervised ML model. We did so based on a read of the whole article. Subsequently, we employed a supervised ML approach by means of a Naïve Bayes classifier applied to the text data from the paper abstracts to determine whether the study/studies in a paper were computational or not.[5] Our model achieved an accuracy of 83.82%, a sensitivity of 88.0%, a specificity of 79.81%, and an area under the curve of 88.97%. After applying this model to the remainder of the sample, a total of 2,551 articles were classified as computational. As this amount was too large for a manual content analysis, from this pool, we randomly sampled 500 articles, thereby loosely adhering to the distribution of articles across publication years. Considering that the Computational Communication Research journal (van Atteveldt et al., 2019), launched in 2019, serves as a central outlet for CCS studies, and was not included in the initial rankings, we further integrated all 35 articles available from this journal as of April 2022 into our dataset. After a final manual inspection, we excluded 59 purely qualitative studies without a clear computational component (e.g., in the data collection process) that did not meet our criteria for being computational (see above), resulting in a final sample size for our manual content analysis of 476 articles. Figure 1 summarizes our sampling process.

[5] We pre-processed the abstracts by removing all punctuation, numbers, and symbols. To remove rare terms or outliers which might not contribute significantly to the analysis, we eliminate terms which occurred less than three times across documents. A combined set of standard English and manually maintained stop words were removed. The abstracts were transformed into a corpus, where each document denotes a different abstract. Subsequently, tokenization was applied to break down the text into individual tokens. More information on the pre-processing process can be found in the OSF.io repository.

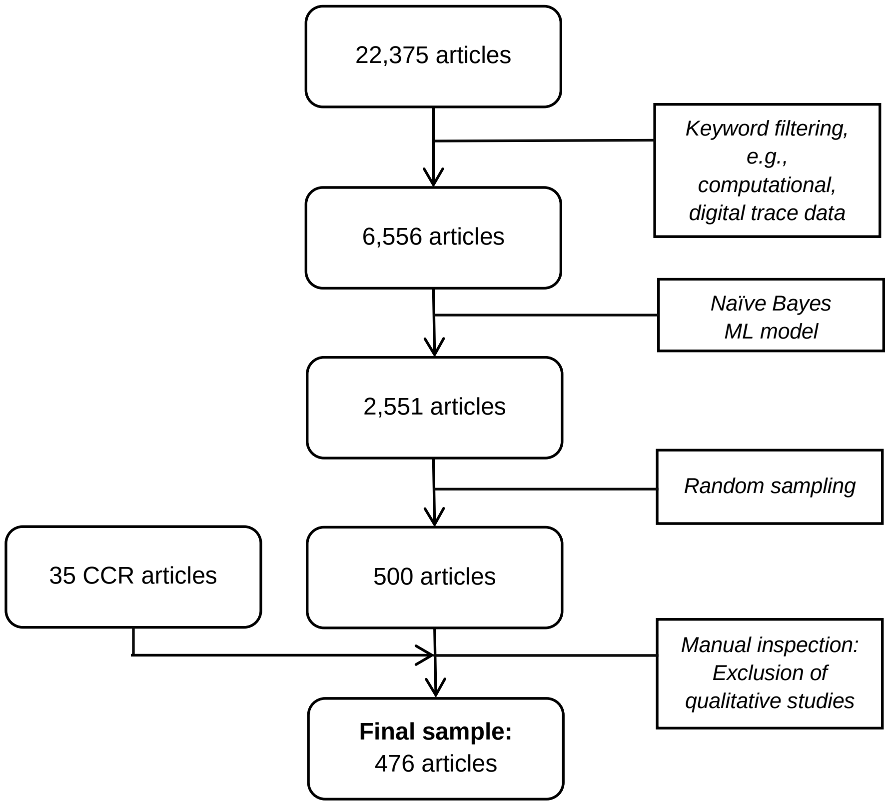

Figure 1 Sample creation process

3.2 Codebook Development

The codebook we employed comprised 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. In addition to basic information, such as article title, author names, year of publication, and journal title, we also captured information about the general type of research (communicator-based, recipient-based, or platform-based). Given that the methodological approach of a study can influence the ethical questions that need to be addressed significantly, we systematically documented the central types of analysis methods in the studies assessed. We defined the central methods of a study as the ones that primarily shape the analysis and produce the key findings of a study. Given the frequent use of multiple analysis methods in articles, we coded information for up to three distinct methods of analysis per article.

The second key category of our codebook were the attributes of the data used in each publication. This includes information about data types, such as media content (e.g., newspaper articles, images, videos, posts, tweets), trace data (e.g., tracking data or sensor data), and self-reported data (e.g., surveys or interviews). The data type determines which analysis methods can be applied but also affects the reproducibility and replicability of research as well as the ethical questions that researchers need to address. We also coded the data source (e.g., Twitter, Facebook, websites) and information about the data origin (primary data, secondary data, or both). Information about both data source and data origin is critical to meet the ethical obligation to conduct sound, transparent, and robust research in CCS and also influences ethical decision-making with regard to the use and sharing of the data. An associated relevant piece of information that we considered is the methodology used for collecting the data such as API access, web scraping, database download, or survey.

Finally, we recorded information about the sample size and the units of analysis (e.g., individuals, websites, articles, tweets, posts). Especially the latter has a strong impact on the required ethical considerations as these naturally differ when human users are the focus of the analysis.

One key dimension of our analysis was also whether the data were made available to other researchers which was coded using the following categories: yes, yes partially, yes available on request, no, not applicable. If the data were made available, we also assessed details about how/where they were shared, such as in an article appendix, a journal online appendix, via an author's website, an institutional repository, a journal repository, a public repository, or by other means. If the data were shared, we also recorded information about the used file format(s) and information about the state in which the data is shared (e.g., full raw data, parts of the data, processed, or aggregated data). This also encompasses capturing whether or how the data have been anonymized or pseudonymized. Moreover, we coded, if the data were documented, we considered factors such as the presence of a codebook, read-me file, structured metadata, or other forms of description.

Finally, we coded information about any data sharing obstacles mentioned by the study authors. Such obstacles may also be particularly driven by ethical considerations. This information was assessed in an open-ended fashion to capture potential reasons for not sharing data, such as copyright restrictions, the proprietary, personal or sensitive nature of the data, and other potential obstacles to data sharing.

Regarding the explicit discussion of ethical procedures and considerations in CCS research, we coded 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 aimed at ensuring the integrity, respectfulness, and fairness of their research. 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 high 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. This approach allows us to comprehensively investigate and categorize the authors' attention to ethical considerations throughout the paper.

Second, we recorded information about mentions of ethical reviews processes. Institutions like 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 like "ethic-" and relevant indicators like "board-," "commit-," "panel-," or "review-."

Third, we also recorded any mentions of other 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 design - 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.[6]

4. Results of the Content Analysis

4.1 Sample Composition

Figure 2 provides an overview of the distribution of the 34 distinct journals within our sample. The distribution of CCS publications across journals exhibits a notable heterogeneity, which can - in parts - be traced back to differences in publishing volumes among journals. As expected, journals with substantial publishing volumes, such as "Information, Communication & Society", "New Media & Society", and the "International Journal of Communication," have the largest representation in our sample, collectively constituting almost a quarter of the total volume (25.42%). Given the interdisciplinary focus of these leading journals, it makes sense that a large share of CCS research is published in the latter. The fourth-largest journal in our sample, "Computational Communication Research", launched at the end of 2019. It specializes in publishing computational research within the field of communication science (van Atteveldt et al., 2019). Overall, our sample spans a broad spectrum of communication journals, encompassing traditional communication research outlets (e.g. “International Journal

[6] Per category values for our intercoder reliability tests can be found in the OSF repository.

of Communication", Communication Research", "Journal of Communication", etc.), more interdisciplinary digital communication and technology journals (e.g. "Journal of Computer-Mediated communication", "Journal of Broadcasting & Electronic Media") more topically focused journals, e.g., in the areas of science communication ("Science Communication"), political communication ("Political Communication"), or health communication ("Journal of Health Communication") . Our sample also includes a substantial representation of journalism journals (e.g. "Digital Journalism", "Journalism Studies", and "Journalism Practice") supporting the argument of a noticeable shift towards computational approaches especially also in journalism research in recent years (Hase et al., 2023).

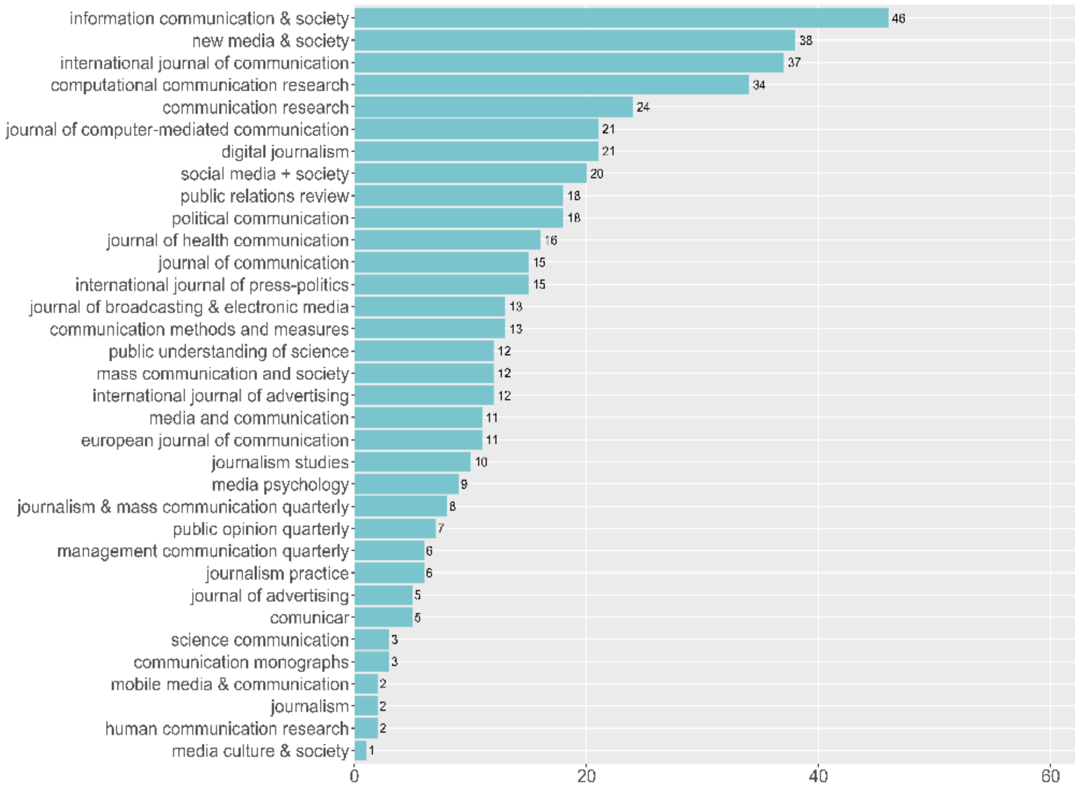

Figure 2 Overview of Journal Titles, n = 476

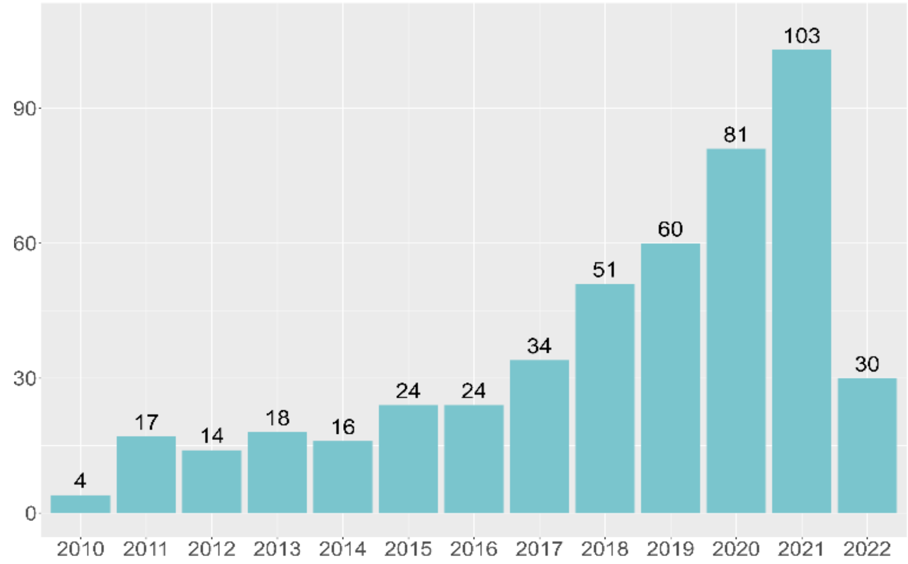

Figure 3 Overview of Publication Dates from 2010 to 2022, n = 476

Figure 3 provides an overview of publication volume across years. Our sample starts one year after the release of Lazer et al.'s (2009) seminal *Nature* article describing the rise of CSS and ends in 2021, as we started collecting data in early 2022. The 30 publications in 2022 represent corrected final publications of previous "advance online publications" from 2021. As can be seen, 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 & Peng, 2018). CCS as a distinct field within both CSS and communication science was still in its early stages of development in the early 2010s (Hilbert et al., 2019) and grew over the years, as computational methods have gained traction among researchers and has CCS received increasing institutional recognition (e.g., in the form of dedicated professorships). Another contributing factor to the rise in CCS publications is the overall increase in communication science publications in the last decade (Rains et al., 2020; Walter et al., 2018).

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

Table 1 Data Collection methods (> 3 obs.), n = 288

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 of the remaining 29.83% was not available. The lack of detailed data collection descriptions violates the norm of research transparency as it, for instance, potentially jeopardizes the possibility to replicate or reproduce the analysis. Table 1 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 underscores the significance of digital data sources in CCS, while also reflecting the field's reliance on digital platforms for data access. Surveys, a main workhorse of quantitative empirical communication science research, also constitute a substantial portion (22.22%) of the data collection methods in the sample.

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

Table 2 Data types, n = 476

Table 2 presents the types of data in the sample. Media content is, by far, the most used analyzed data type in CCS, representing 55.67% of the studies. Media content data encompasses text from social media posts (e.g., Twitter, Facebook, Reddit), news articles, as well as image and video data. This reliance on media content data may introduce ethical challenges related to privacy concerns as well as potential copyright issues. Self-reported data, constituting 31.30%, includes data from surveys, interviews, and experiments (questionnaires). This use of self-reported data raises concerns about the potential inclusion of personal and sensitive information, necessitating careful ethical considerations in research design and implementation. Trace data covers smartphone data, passive tracking data, sensor data, or search engine data. The category labeled "other types of data" includes cases like simulation studies (e.g., data from agent-based models) or metadata. Notably, eight studies in the sample, method as well as tool exhibitions, did not utilize a shareable data set.

Data analysis method	# of obs.	Fraction of sample in %
Content analysis	126	28.31
Network analysis	92	20.67
Regression	79	17.75
Text analysis	31	6.97
Experiment	24	5.39
Method exhibition	21	4.72
Correlation analysis	20	4.49
Tool exhibition	14	3.15
Machine learning	12	2.70
Cluster analysis	10	2.25
Factor analysis	8	1.80
Structural equation modeling	8	1.80

Table 3 Types of Data Analysis Methods (> 4 obs.), n = 445

Table 3 displays the distribution of data analysis methods with more than four observations employed in our sample. Content analysis emerged as the most frequent data analysis method, constituting 28.31% of the sample. This finding aligns with the historical significance of content analyses in communication science (Berelson, 1952; Haim et al., 2023;

Krippendorff, 2018). After all, this is typically considered a genuine method of communication science. Network analysis is also quite widely employed in CCS, representing 20.67% of the sample. Regression, at 17.75%, indicates a significant use of statistical modeling within CCS. Somewhat surprisingly, compared to this, ML approaches only constitute a minor fraction of analysis methods used (2.70%). In sum, content, network and regression analyses seem to be the three most important CCS according to our sample, making up almost two thirds (66.73%) of all identified data analysis methods. Method and tool exhibitions, at 4.72% and 3.15%, respectively, suggest a focus on showcasing methodologies and tools within the CCS community. This may involve presenting new computational techniques or software applications relevant to communication research. Overall, the distribution of data collection methods in CCS is indicative of a methodological diversity reflecting the fundamental interdisciplinary nature of the field. Importantly, similar to the different data types, the various analysis methods are also associated with specific ethical challenges.

4.2 Data Sharing in CCS

Among the 476 studies in our sample, a stunning 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 studies that engaged in some form of data sharing, 26 made their data available in public repositories, such as Open Science Framework (OSF), Zenodo, figshare, or Dataverse, 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 (1.89%) 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 (1.89%) 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 OSF repository, followed by GitHub (4), the Harvard Dataverse (3), and the GESIS Data Archive (1). Additionally, one study made use of the JGSS Daishodai platform, a data sharing platform by the Japanese Ministry of Education, Culture, Sports, Science, and Technology. The preference for such repositories underscores the importance of established platforms for facilitating data sharing in a standardized and accessible manner. Researchers are likely drawn to these platforms due to their established reputation, accessibility, and clear usage guidelines (Rockhold et al., 2019).

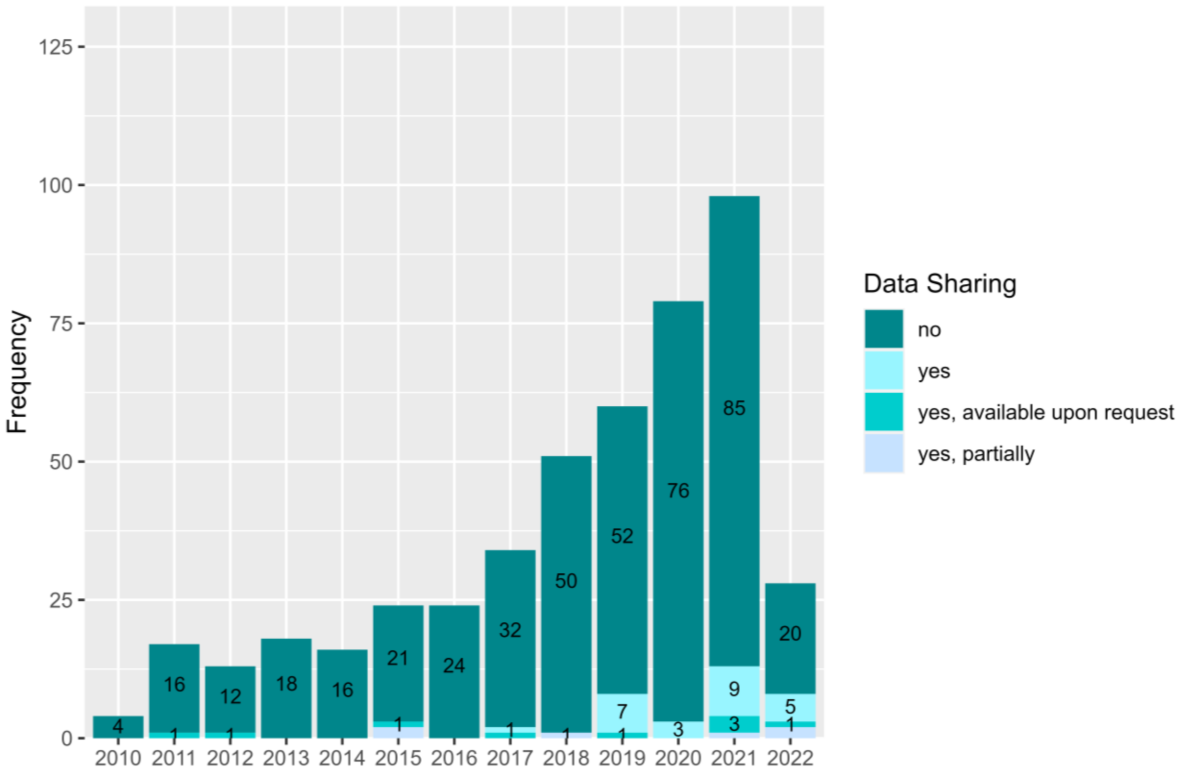

Figure 4 Data sharing over the years from 2010 to 2022

The 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). Put differently, while the concept of data sharing has been existing for a while (Sieber, 1991), it only recently became more common for researchers to share their data, which is largely due to digitalization and the ways in which it facilitates the sharing of research products.

In Figure 4, we analyze data sharing trends in our sample over time (from 2010-2022). We can observe that with an increase in the volume of published CCS articles a positive data sharing trend occurs, particularly in the recent years 2019-2022. 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 ethical 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 & Panofsky, 2023) as well as the social sciences and communication science in particular (Dienlin et al., 2020). 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 research practices. 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 & 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.

4.3 Ethical considerations in the CCS literature

Ethical consideration	# of obs.	Fraction of sample in %
General ethical consideration	28	5.88%
Ethical reviews processes	31	6.51%
Other ethical procedures	69	15.50%

Table 4 Mentions of ethical considerations

Table 4 shows the results of our manual content analysis of the mentioning or discussion of explicit ethical considerations in the considered CCS publications. Among the 476 studies in the sample, only 28 (5.88%) papers explicitly address general ethical considerations. These general ethical considerations encompass a diverse range of topics and principles, generally aimed at upholding the integrity, respectfulness, and fairness of the conducted research. In the following, we showcase a few examples of general ethical considerations from the articles in our sample. For instance, ethical considerations could pertain to researchers emphasizing their ethical commitment to ensuring user privacy when collecting data from participants where no informed consent is asked in advance such as in Urman and Katz (2022, p.17): "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".

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. The authors address ethical issues that resonate across various CCS research endeavors, such as collecting data without explicit consent and safeguarding user privacy in complex data situations. 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. Other examples of general ethical considerations in our sample of articles include mentioning the compliance with ethical guidelines, such as the ones by the

Association of Internet Researchers (Franzke et al., 2019) or the mention of adhering to specific ethical best practices when handling large data volumes (Siapera et al., 2018).

Only 31 (6.51%) of the studies in our sample explicitly mentioned ethical review processes in the sense of getting approval by ethical review institutions. In our sample, the majority of ethical mentions fall into the category of "Other ethical procedures" (15.50%), encompassing a range of different ethical procedures for ensuring the well-being, privacy, and the rights of human participants in their studies. The various ways in which ethical aspects are discussed within the investigated publications suggest that researchers in CCS prioritize different ethical practices, including obtaining informed consent, choosing between opt-in and opt-out designs, and implementing debriefing processes.

5. Discussion

Generally speaking, 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 & Rubenfeld, 2017; Israel & 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 & Antes, 2018). 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 have been significant efforts in the recent past to establish a set of applicable guidelines for research ethics within the CSS community (Engel et al., 2021; Haim, 2023; Herschel & Miori, 2017; Hosseini et al., 2022; Salganik, 2019; Stegenga et al., 2023; Steinmann et al., 2016; Weinhardt, 2020; Zwitter, 2014), as well as relevant professional academic associations, such as the Association of Internet Researchers (AoIR)[7], the American Psychological Association (APA)[8], and the International Communication Association (ICA)[9], there is still a considerable need for discussions and guidance regarding ethical considerations in CCS.

In the context of these ethical considerations, one key finding of our content analysis of the CCS literature was that for a significant majority of CCS studies (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 data collection via database downloads and API access underscores the reliance on digital platforms for data access in CCS. This may introduce intricacies in sharing data due to restrictions imposed by platform terms of service (ToS) or other contractual or license agreements. Additionally, the dominance of media content data raises legal ethical challenges related to privacy and copyright issues. CCS researchers may be hesitant to share such data due to concerns about compromising individual privacy or violating copyright regulations. The use of self-reported data (e.g., surveys) introduces another layer of ethical considerations also regarding the disclosure of personal and possibly sensitive information.

The distribution of data analysis methods shows the methodological diversity in CCS. This diversity may also contribute to particular ethical challenges. CCS researchers may be

[7] https://aoir.org/ethics/.
[8] https://www.apa.org/ethics/code.
[9] https://www.icahdq.org/page/MissionStatement.

cautious about sharing data (and also code) when the analytical techniques could reveal sensitive or personal information. However, we did observe somewhat of a notable positive trend in data sharing over the past years in our sample, which may be ascribed to a cultural shift towards open science practices within the scientific community (Dienlin et al., 2020; Peterson & Panofsky, 2023), institutional and funder mandates (Anger et al., 2022), as well as journal policies (Piwowar & Chapman, 2008; Vasilevsky et al., 2017). One important thing to note in this context is that among the studies which did not share their data, not a single study provided explicit reasons for their decision.

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 studies 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.

Furthermore, only a small fraction (6.51%) of the studies explicitly mentioned the approval from institutional review boards or similar institutions. 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 maximizing the benefit of the research or safeguarding the well-being of participants.

In contrast, a substantial proportion (15.5%) of studies 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 context, debriefing focusses on emphasizing positive outcomes and harm reduction, aligning with consequentialist principles that assess the ethicality of actions based on their overall consequences rather than strict adherence to 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 its 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. 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 in light of 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.

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 (Haim, 2023; Salganik, 2019; Schlütz & 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 in making this research even more robust and, ultimately, even more credible.

References

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), e0273259. 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

Berelson, B. (1952). *Content analysis in communication research* (S. 220). Free Press.

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

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

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. (2020). An Agenda for Open Science in Communication. *Journal of Communication*. 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

Engel, U., Quan-Haase, A., Liu, S. X., & Lyberg, L. E. (2021). *Handbook of Computational Social Science, Volume 1: Theory, Case Studies and Ethics*. 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

Fecher, B., Friesike, S., & Hebing, M. (2015). What Drives Academic Data Sharing? *PLOS ONE*, *10*(2), e0118053. https://doi.org/10.1371/journal.pone.0118053

Franzke, A., Bechmann, A., Zimmer, M., & Ess, C. M. (2019). *Internet research: Ethical guidelines 3.0: Association of internet researchers*.

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*. Springer Fachmedien. https://doi.org/10.1007/978-3-658-40171-9

Haim, M., Hase, V., Schindler, J., Bachl, M., & Domahidi, E. (2023). (Re-)Establishing quality criteria for content analysis: A critical perspective on the field's core method. *Studies in Communication and Media*, *12*(4), 277–288.

Hase, V., Mahl, D., & Schäfer, M. S. (2023). The "computational turn": An "interdisciplinary turn"? A systematic review of text as data approaches in journalism studies. *Online Media and Global Communication*, *2*(1), 122–143. https://doi.org/10.1515/omgc-2023-0003

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 Science: 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

Israel, M., & Hay, I. (2006). *Research Ethics for Social Scientists*. SAGE.

Jürgens, P., Meltzer, C. E., & Scharkow, M. (2022). Age and gender representation on German TV: A longitudinal computational analysis. *Computational Communication Research*, *4*(1).

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

Krippendorff, K. (2018). *Content Analysis: An Introduction to Its Methodology*. SAGE Publications.

Lazer, D. M. J., 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

Lazer, 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

Leslie, D. (2023). The Ethics of Computational Social Science. In E. Bertoni, M. Fontana, L. Gabrielli, S. Signorelli, & M. Vespe (Hrsg.), *Handbook of Computational Social Science for Policy* (S. 57–104). Springer International Publishing. https://doi.org/10.1007/978-3-031-16624-2_4

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

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

Mühlhoff, R. (2021). Predictive privacy: Towards an applied ethics of data analytics. *Ethics and Information Technology*, *23*(4), 675–690. https://doi.org/10.1007/s10676-021-09606-x

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
Piwowar, H., & Chapman, W. (2008). A review of journal policies for sharing research data. *Nature Precedings*, 1–1. https://doi.org/10.1038/npre.2008.1700.1
Rains, S. A., Keating, D. M., Banas, J. A., Richards, A., & 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), Article 2.
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
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
Siapera, E., Boudourides, M., Lenis, S., & Suiter, J. (2018). Refugees and network publics on Twitter: Networked framing, affect, and capture. *Social Media+ Society*, *4*(1), 2056305118764437.
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 Reyes, L., & Lee, L. E. (2023). Qualitative research at the crossroads of open science and big data: Ethical considerations. *Social and Personality Psychology Compass*, *n/a*(n/a), 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 (Hrsg.), *Ethical Reasoning in Big Data: An Exploratory Analysis* (S. 11–27). Springer International Publishing. 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 (Hrsg.), *Ethics, Governance, and Policies in Artificial Intelligence* (S. 97–123). Springer International Publishing. 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
van Atteveldt, W., Margolin, D., Shen, C., Trilling, D., & Weber, R. (2019). A Roadmap for Computational Communication Research. *Computational Communication Research*, *1*(1), 1–11. https://doi.org/10.5117/CCR2019.1.001.VANA
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*.

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.

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

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