SC|M
Studies in Communication and Media

FULL PAPER

Ethical challenges in contemporary quantitative content analysis

Forschungsethische Herausforderungen in der modernen quantitativen Inhaltsanalyse

Annemarie Wiedicke

Annemarie Wiedicke (Dr.), TU Chemnitz, Institute for Media Research, Reichenhainer Straße 41, 09126 Chemnitz, Germany. Contact: annemarie.wiedicke@phil.tu-chemnitz.de

FULL PAPER

Ethical challenges in contemporary quantitative content analysis

Forschungsethische Herausforderungen in der modernen quantitativen Inhaltsanalyse

Annemarie Wiedicke

Abstract: Questions of validity and research ethics are closely linked, with content-analytic research being no exception. Nevertheless, in the past, quantitative content analysis has often been excluded from ethical discussions among communication scholars. Recently, however, there is an increasing discourse on the various ethical challenges arising from conducting quantitative content analysis in the digital age. Drawing from the four principles of ethical research (respect for a person's autonomy, beneficence, nonmaleficence, and justice), this paper seeks to present a systematic overview of the ethical challenges and potential methodological-ethical dilemmas arising throughout the process of contemporary content-analytic research. Where possible, according solutions are offered. As such, this paper contributes not only to the growing literature reflecting ethical aspects of quantitative communication research, but also to the current discourse on quality in contemporary quantitative content analysis.

Keywords: Content analysis, research ethics, principlist ethics, big data, social media.

Zusammenfassung: Validität und der Forschungsethik sind eng miteinander verknüpft, auch in der inhaltsanalytischen Forschung. Dennoch wurde die quantitative Inhaltsanalyse in früheren forschungsethischen Debatten innerhalb der Kommunikationswissenschaft zumeist ausgeblendet. Inzwischen wird jedoch zunehmend über die verschiedenen ethischen Herausforderungen, welche sich aus der Durchführung quantitativer Inhaltsanalysen im digitalen Zeitalter ergeben, diskutiert. Vor diesem Hintergrund liefert der vorliegende Beitrag, ausgehend von den vier Grundsätzen ethischer Forschung (Respekts vor der Selbstbestimmung, Wohltun, Schadensvermeidung und Gerechtigkeit) und unter Berücksichtigung der verschiedenen Schritte im Forschungsprozess, einen systematischen Überblick über bestehende Herausforderungen und mögliche methodisch-ethische Dilemmata moderner quantitativer Inhaltsanalysen. Wo möglich, werden entsprechende Lösungsvorschläge vorgestellt. Auf diese Weise knüpft der vorliegende Beitrag sowohl an die wachsende Forschungsliteratur zur Ethik in der quantitativen Kommunikationsforschung als auch an den aktuellen Diskurs zur methodologischen Qualität moderner quantitativer Inhaltsanalysen an.

Schlagwörter: Inhaltsanalyse, Forschungsethik, Prinzipienethik, Big Data, Soziale Medien.

1. Introduction

Quantitative content analysis, i.e., the systematic and replicable assignment of communication content to categories (coding) and the statistical analysis of relationships involving these categories aiming to both describe communication and infer from communication to social reality (Riffe et al., 1998, p. 20) is often referred to as the only empirical method originally developed in communication research (Rössler, 2017, p. 15). In recent years, concepts and methods from computational science, such as machine learning, natural language processing, and network analysis, are more and more incorporated in media and communication studies (Hase et al., 2022, p. 61; Neuendorf, 2017, p. 143; van Atteveldt & Peng, 2018, p. 1). This "computational turn" (Hase et al., 2022, p. 60) of communication research has led to fundamental changes of quantitative content analysis as a well-established research method (Nelson, 2020, p. 4).

One major change concerns the coding process (Lewis et al., 2013, p. 34). Although *manual* coding is still of significance (Kessler et al., 2023, p. 9), *automatic* approaches have increasingly gained attention in content-analytic research (Hase, 2023, p. 23). Here, in contrast to traditional, manual approaches, coding decisions are not made by humans, but by computer algorithms (Kessler et al., 2023, p. 9). Thus, the analyzed content is not read and understood as one unit, but automatically broken down to its features, e.g., single words or distinctive features of images (Ha et al., 2021, p. 54; Hase, 2023, pp. 23–24). At the same time, most automated content analyses still rely on manual coding, at least to some extent (Hase, 2023, p. 24).

Against this backdrop and drawing on the aforementioned definition by Riffe et al. (1998, p. 2), *contemporary* quantitative content analysis can be understood as the systematic and replicable assignment of communication content to categories *by either human coders, computer algorithms or a combination of both*, and the statistical analysis of relationships between these categories in order to describe communication as well as infer to social reality. In consequence, by incorporating both manual and automatic approaches, contemporary quantitative content analysis enables communication researchers to analyze a wide variety of contents, such as traditional mass media content, but also audio-visual and auditory (big) data from various online sources, including social media (Jünger et al., 2022, p. 1482; Kessler et al., 2023, p. 10).

Currently, contemporary quantitative content analysis faces various discussions concerning quality criteria and standards (e.g., Casas & Williams, 2022, p. 6; Hase, 2023, p. 31; Krippendorff, 2021, p. 165). In particular, scholars have addressed the issue of ensuring *validity* (e.g., Baden et al., 2022, p. 1; Hopp & Weber, 2021; Mahl et al., 2022; Song et al., 2020, p. 550), i.e., asking (1) whether a measuring procedure actually represents the intended concept (*internal validity*, Neuendorf, 2017, p. 122) and (2) if a methodological approach allows for adequate inference to a social reality (*external validity*, Carrig & Hoyle, 2011, p. 136; Zyphur & Pierides, 2017, p. 7). Notably, ensuring validity as a major criterion for research quality has always been important, and sometimes difficult, in content-analytic research (Janis, 2009, p. 358). Nevertheless, the rise of (big) online data and automated approaches have led to new challenges in this context. For instance, there is an ongoing debate on which of the (complex) constructs at the core of communication studies (e.g., frames) can – or even should be – measured by algorithms rather than human coders, and which not (Hase, 2023, p. 24). In addition, the potential bias of (big) data from online sources (e.g., user-generated content or behavioral traces) raises the question whether, and if so, under which circumstances, inference to social reality based on such data is possible (Olteanu et al., 2019, 2–5). These and other issues will be discussed in detail later in the manuscript.

Importantly, inference can neither be separated from the practices it draws on nor their ethical implications (Schlütz & Möhring, 2018, p. 36): Poorly designed research may yield invalid findings and therefore result in misleading knowledge claims (Bond, 2012, p. 102; Vail et al., 2009, p. 85). In that case, those who contribute to research, but also society at large, will have been exposed to risks and burdens of research (such as emotional strain from constantly coding violent media content or financial costs) for no reason, i.e., no clear benefit (Vail et al., 2009, p. 85). At worst, public trust in the integrity of the research process – and even science in general – could be undermined (Bond, 2012, p. 102). Or, in the words of Wassenaar and Mamotte (2013, p. 15), "poor science is unethical," as it wastes resources, including time as well as (public) money.

Still, resolving ethical issues without compromising methodological rigor and, in consequence, internal and external validity as well as a study's overall quality can be challenging to impossible (Hunter, 2007, p. 24; Schlütz & Möhring, 2018, p. 36). This is, for example, illustrated by Bishop and Gray (2018, pp. 174–178) and Sugiura et al. (2017, pp. 189–190) regarding the issue of obtaining informed consent from social media users for analyzing and publishing their data.

So, both methodological *and* ethical considerations are closely linked to a study's validity (Zyphur & Pierides, 2017, p. 4), and their critical assessment thus contributes to the overall quality of research (Heise, 2017, p. 771; Schlütz & Möhring, 2018, p. 34). Nevertheless, the systematic integration of method and ethics in order to optimize research quality seems to be the exception rather than the rule (Panter & Sterba, 2011, pp. 1–2; Schlütz & Möhring, 2018, p. 35). For instance, ethical discussions in quantitative research have often been limited to the treatment of *human subjects*, with communication research being no exception (Cortina, 2020, p. 20; Schlütz & Möhring, 2018, p. 33). In consequence, in the past, communication scholars have seldomly considered ethical implications of quantitative content analysis as a text-based and thus (supposedly) non-reactive method (Heise, 2017, p. 769; Neuendorf, 2017, pp. 130–131).

However, in recent years, there is an increasing debate on ethical issues in content-analytic research, such as, for example, (student) coders experiencing emotional distress due to coding violent or stigmatizing content (Rössler, 2017, p. 231; Schlütz & Möhring, 2018, p. 39; Signorielli, 2008). In this context, some researchers argue that conducting an automated *instead* of a manual content analysis might be the right choice (e.g., Schlütz & Möhring, 2018, p. 39). Still, as I will discuss later in the manuscript, this solution is not as straightforward as it might appear initially. In addition, there is a growing discourse on the various ethical challenges arising from the collection and analysis of (big) data from online sources (e.g., Hollingshead et al., 2022; Hosseini et al., 2022; Sugiura et al., 2017;

Taylor & Pagliari, 2018; Zook et al., 2017). Thus, in line with the current changes in quantitative content analysis (e.g., regarding the coding procedure and the contents analyzed), its ethical implications are changing as well.

Nevertheless, to my knowledge, the emerging debate still lacks (1) an overview of the ethical challenges and potential dilemmas arising from the decisions made throughout the process of contemporary quantitative content-analytic research (Heise, 2017, p. 769), and (2) how they are interrelated with a study's validity and, in consequence, overall quality. Thus, guided by the four principles of ethical research (Beauchamp & Childress, 2019; Wiles, 2012, p. 14), this paper systematically addresses a variety of ethical issues within the research process – from research interest to reporting and publication (Neuendorf, 2017, pp. 40–41). Importantly, it is *not* the objective of this paper to lay out *all* the ethical challenges that may occur during conducting contemporary content analysis. Rather, by discussing issues arising from both manual and automated approaches and in the context of various types of media content, this work seeks to show that *each step* in the research process of contemporary quantitative content analysis poses ethical challenges, and that the (methodological and ethical) decisions we make throughout this process are closely interlinked. As such, this paper seeks to contribute to the growing literature reflecting ethical aspects of quantitative communication research (e.g., Döveling et al., 2016; Heise, 2017; Schlütz & Möhring, 2016; 2018) as well as the current discourse on quality criteria and standards for contemporary quantitative content analysis (e.g., Casas & Williams, 2022; Krippendorff, 2021; Song et al., 2020).

2. Principles of ethical research

The quality of research and research ethics are closely linked. On the one hand, poorly designed research is unethical, simply because unreliable or invalid methods waste resources, and yield invalid and unusable results (Bond, 2012, p. 101; Wassenaar & Mamotte, 2013, p. 15). On the other hand, responsible ethical conduct could also *negatively* affect validity (Döveling et al., 2016, p. 4). However, before addressing such dilemmas in detail, I would like to introduce my theoretical basis of reflecting the ethics of content analysis.

There are various theoretical approaches to research ethics, including, for instance, *ethics of care*, *virtue ethics*, *consequentialist*, and *non-consequentialist ethics* (see Wiles, 2012, pp. 12–15 for an overview). Still, ethical considerations in social sciences in general and communication research in particular frequently refer to *principlist approaches* (Beauchamp & Childress, 2019) which are based on four widely accepted philosophical principles (Wassenaar & Mamotte, 2013, p. 12) – specifically, (1) *respect for people's autonomy*, (2) *beneficence*, (3) *nonmaleficence*, and (4) *justice*.

The (1) *respect for autonomy* is "based on the premise that individuals have an intrinsic right to make decisions for themselves" (Guttman & Thompson, 2010, p. 295). Against this backdrop, it is – in general – our obligation as researchers to make sure that people can decide for themselves whether they want to participate in or contribute to a study, or not (Beauchamp, 2003, p. 269). This includes providing prospective participants (or other contributors, such as coders) with detailed information about the project, and giving them the opportunity to decline the participation or to withdraw from the study at any time, and without adverse consequences (Crow et al., 2006, pp. 83–84). Therefore, this principle relates to ethical standards such as *voluntariness* and *informed consent*. (2) *Beneficence* requires researchers to “provide benefits to others” (Beauchamp & Childress, 2019, p. 217) which includes scientific benefits (e.g., theoretical contributions) as well as societal benefits (e.g., practical implications). In contrast, (3) *nonmaleficence* concerns researchers’ responsibility to “abstain from causing harm to others” (Beauchamp & Childress, 2019, p. 155), such as emotional distress or *informational risk,* i.e., the potential harm from disclosure of information (Salganik, 2018, p. 307). Notably, this principle relates to standards such as *confidentiality* and *anonymity* (Wiles, 2012, p. 14). In sum, *beneficence* and *nonmaleficence* are about weighing up a study’s benefits and risks (including the probability of adversary events as well as their severity), and deciding whether they strike the right balance (Salganik, 2018, p. 296). Finally, (4) *justice* addresses the obligation of benefits and burdens of research being distributed equally (Salganik, 2018, p. 298; Wiles, 2012, p. 14). In other words, “all persons should have access to and benefit from the contributions of science” (Schlütz & Möhring, 2018, p. 40), especially those who mostly contribute to it (such as, for instance, students) or carry the costs. Alternatively, they should at least be compensated adequately (Mark & Lenz-Watson, 2011, pp. 197–198). Please see Table 1 for an overview of the different principles.

Table 1. Four principles of ethical research

Principle	Meaning
Respect for autonomy	Researchers’ responsibility to ensure that people can decide for themselves whether they want to take part in or contribute to a study, including the option to decline or withdraw; refers to standards such as voluntariness and informed consent.
Beneficence	Research should provide scientific and societal benefits, such as a theoretical contribution to the field or specific practical implications.
Nonmaleficence	Researchers’ responsibility to keep potential risks and harms of research as low as possible; refers to standards such as confidentiality and anonymity.
Justice	Burdens and benefits of research should be distributed equally, and everyone should be able to access research.

Note. Principles of ethical research based on the works of Beauchamp and Childress (2019).

According to various scientific associations relevant for the field (e.g., APA, 2017; ICA, 2019; WAPOR, 2017), *respect for autonomy*, *beneficence*, *nonmaleficence*, and *justice* should inform ethical decisions in communication research. They (and the standards arising from them, i.e., *voluntariness*, *informed consent*, *confidentiality*, and *anonymity*) are not only widely accepted in social sciences and applied in various ways to research ethics (Wassenaar & Mamotte, 2013, p. 12), but also offer helpful guidance for addressing ethical challenges and dilemmas (Wiles, 2012, p. 15) for communication researchers (Schlütz & Möhring, 2016, p. 488). Therefore, in the following, I refer to these four principles in order to discuss ethical challenges and dilemmas of contemporary quantitative content analysis.

3. Ethical challenges in the process of contemporary content-analytic research

According to McKee and Porter (2009, p. 19), "ethical considerations are inseparable from methodological considerations, and they occur throughout the entire research process." Thus, from (1) research interest, to (2) sampling and data collection, (3) coding, and, finally, (4) reporting, publication, and data management, every phase of content-analytic research offers specific challenges (Schlütz & Möhring, 2018, p. 37). As mentioned before, adhering to ethical principles will, in general, ensure valid research practices and results (Heise, 2017, p. 771; Schlütz & Möhring, 2018, p. 37). Occasionally, however, ethical and methodological assessments might collide (Schlütz & Möhring, 2018, p. 37), and it is crucial to address and discuss such issues. Thus, when presenting ethical challenges in contemporary content analysis, I am going to contrast ethical and methodological arguments in a transparent manner.

Firstly, I address the ethical implications of purposeful research (in contrast to research without a clear purpose) and the importance of reflecting on the ethical appropriateness of a study *before* it is conducted. A summary of all ethical challenges and methodological concerns discussed in the following can be found in Table 2 at the end of this chapter.

3.1 Research interest

When planning a quantitative content analysis, communication scholars usually ask themselves *what* content to analyze, and *why*; or, more specifically, whether certain theories or concepts indicate that a particular message content is important to study (Neuendorf, 2017, p. 40). In providing answers to such questions, researchers lay the foundation for a coherent, valid, sensible, and purposeful design which contributes to the field both theoretically and practically (McKee & Porter, 2009, p. 142; Zyphur & Pierides, 2017, p. 1). However, this first step is not as straightforward as it seems. For instance, the purpose of contemporary quantitative content analyses, and especially automated approaches, is often not necessarily clear (Hase, 2023, p. 31). More specifically, according to Hase (2023, p. 31), one of the biggest questions scholars need to address in this context is whether measuring "latent constructs such as topics, frames, or sentiment through automated content analysis" does enable researchers to "capture things that are relevant for theories and frameworks within communication science" (p. 31) – or not. Such issues need to be solved, not only because they diminish a study's validity, but because without a clear purpose, both the *beneficence* and *nonmaleficence* of research can hardly be ensured.

That is, a content analysis lacking a clear purpose and thus leading to invalid or irrelevant findings will not provide clear benefits (e.g., a theoretical contribution to the field, practical implications), while still taking up both time and financial resources or carrying potential risks.

Moreover, when choosing a research topic, scholars need to reflect on whether, and under which circumstances, *respect for autonomy*, *beneficence*, *nonmaleficence*, and *justice* – and thus, ethically appropriate research – can be ensured (Schlütz & Möhring, 2018, p. 38). For instance, researchers interested in analyzing (potentially) disturbing media content, including porn, war, or health coverage, should weigh the scientific and societal benefits of their research against its risks (e.g., emotional distress of researchers and coders, reproducing stigma) and decide whether these strike the right balance.

Critically assessing such questions *before* conducting research is even more important when it comes to the analysis of (big) data from online sources, such as messages or images shared on social media (e.g., Hollingshead et al., 2022, 171): This data particularly represents and affects people, and even apparently innocuous data may contain sensitive and private information (Zook et al., 2017, 2). Thus, according to Salganik (2018, p. 309), *all* data are *potentially* sensitive. For example, it is possible to extract information regarding the heart rates of people from YouTube Videos (Zook et al., 2017, 2). In consequence, analyzing Instagram posts, tweets, TikTok videos, or similar contents poses an informational risk (Salganik, 2018, p. 307). This does not mean that communication researchers should abstain from analyzing these contents; but that for an informational risk and other potential harms to be justified, it is crucial that a study provides clear scientific and societal benefits (including valid results). Of course, considering said harms is especially relevant when it comes to highly controversial research[1] or particularly vulnerable groups (*justice*). Nevertheless, we as researchers should keep in mind that data being publicly available (and us being technologically able to analyze it) alone does *not* justify its analysis (Heise & Schmidt, 2014; Stommel & Rijk, 2021, p. 275; Zook et al., 2017, 8), as even seemingly neutral data "can yield discriminatory outcomes, thereby compounding social inequities" (Zook et al., 2017, 2).

At the same time, however, ethical concerns (which are discussed in more detail later) should not generally prevent relevant and important research from happening (Salganik, 2018, p. 281; Zook et al., 2017, 7). For instance, in times of an emerging pandemic, it might be crucial to temporarily concentrate less on questions of individual privacy in order to serve a larger public good (Salganik, 2018, p. 281; Zook et al., 2017, 7–8), such as helping controlling the outbreak as well as the spread of misinformation ("Infoveillance", Chen & Wang, 2021, 4). Referring to the principles of ethical research, one could argue that in this case, the (1) *beneficence* of the overall project outweighs the violation of the (2) principle of *respect* towards a person's *autonomy* – assuming the (3) findings serve as many people as possible (*justice*) and the (4) informational risk is minimized (*nonmaleficence*). So, overall, ethical research is not about refraining from doing a study just because there are ethical concerns, but about critically assessing the study's implications regarding the four principles of ethical research (*respect for autonomy*, *beneficence*, *nonma-*

1 Including, for instance, various recently published studies that sought to predict a person's sexual orientation, political attitudes, or other attributes, such as "criminal tendencies" based on the content they share online. I refrain from offering specific references, as I would prefer to not support such studies by citing them.

leficence, and *justice*), and deciding whether, and how, they can be ensured in a way that is ethically appropriate (Schlütz & Möhring, 2018, p. 38).

As major next steps in the research process, sampling and data collection also come with a variety of methodological and ethical challenges. In the following, I point out the significance of an adequate sample size concerning a study's validity, but also *nonmaleficence* and *beneficence*. Furthermore, I discuss the difficulty of ensuring the *respect for people's autonomy* when analyzing social media data as well as potential solutions. Finally, I address the issue of bias in contemporary quantitative content analysis from an ethical perspective, particularly regarding questions of *justice*.

3.2 Sampling and data collection

When it comes to sampling and data collection for contemporary content analysis, one crucial (and often neglected) aspect is that – from an ethical perspective – samples should be *as big as necessary*, but also *as small as possible* (Maxwell & Kelley, 2011, 159ff.; Salganik, 2018, p. 319). On the one hand, if a study fails to have sufficient *statistical power*, it may lead to unreliable or invalid findings, thus wasting resources. As stated before, such poor science is, due to a lack of scientific and societal benefits (*beneficence*), unethical (Wassenaar & Mamotte, 2013, p. 15). On the other hand, the principle of *nonmaleficence* suggest to keep the potential risks and burden arising from research as small as possible, whilst still achieving the research objective (Salganik, 2018, p. 319). In consequence, especially in the context of big data studies, it is important to critically reflect on adequate sample sizes (Rössler, 2017, p. 228). Or, as Zook et al. (2017) put it, "big does not automatically mean better" (p. 4).

Similarly important is the question of which data is perceived as *public*, and which as *private* (Sugiura et al., 2017, p. 184; Taylor & Pagliari, 2018, p. 3). While contents such as newspaper articles, or radio and TV newscasts are generally considered as public (Rössler, 2017, p. 224), the distinction between public and private data becomes less clear when we look at user-generated content on social media (Rössler, 2017, p. 224; Sugiura et al., 2017, p. 185). We generally might agree that tweets shared by authorities (e.g., politicians and institutions) are public (e.g., Krieger et al., 2014, pp. 210–213), but what about Instagram stories in which people openly discuss their health issues, or Telegram posts in which they address their political attitudes? Just because such information is often openly accessible, this does *not* automatically mean that users are aware of the public status of their contributions, or intended it (Sugiura et al., 2017, p. 193; Williams et al., 2018, pp. 37–38). Moreover, most people who engage in online interactions rarely anticipate that their information might be used in the context of future research projects (Hosseini et al., 2022, p. 12).

Overall, this leads us to the next question: Under which circumstances is it ethically appropriate to collect, and later analyze, such data? Considering the principle of *respect for* people's *autonomy*, it seems desirable to seek people's *informed consent* before mining and examining social media data (Kozinets, 2006, p. 11). However, this is easier said than done, not only with regards to practical, but also methodological implications (Salganik, 2018, p. 306). Let's suggest we try to obtain informed consent for analyzing information shared on social media by posting to communities or platforms: There is no guarantee that everyone whose contributions we would like to examine actually sees these posts, or is still active in the community respectively on the platform (Sugiura et al., 2017, pp. 191–192). Moreover, people might change the way they act in these spaces when they are aware of their contents being examined, which would lead to invalid findings (Sugiura et al., 2017, p. 190). Against this backdrop, Heise and Schmidt (2014, pp. 9–14) recommend to critically assess the (1) accessibility and sensitivity of contents and the (2) aim of a study in order to decide whether users' informed consent to collect and analyze social media data is necessary, or at least highly recommended. Specifically, they advise researchers to obtain informed consent for the collection and content analysis of (1) clearly sensible information (e.g., health-related information, dating preferences) shared in online spaces that are *not* openly accessible, but require a registration, as well as (2) data that represent users as individual subjects or recipients (e.g., interactions or discussions in comment sections), rather than actors or communicators (e.g., producers of YouTube videos).

While these recommendations make sense from an ethical point of view, the question remains how realistic they are – given the obstacles communication researchers might experience when seeking users' informed consent to collect and analyze their social media posts. Sugiura et al. (2017, 189–190) have discussed this issue thoroughly: In their study, they explored interactions and communication about the online purchase of medicine. To do so, they intended to scrape data from six different online forums. Following the recommendations of their university's ethics committee, one of the researchers joined the forums and posted information about their study. These posts included information on how the researchers could be contacted if forum members did not wish their posts to be analyzed. However, this disclosure approach led to various negative reactions, including abusive and suspicious comments from forum members, and the removal of the researcher's posts by the forums' moderators. Therefore, after further discussion with the ethics committee, the researchers chose to collect data exclusively from public forums where moderators (when asked beforehand) did not object to the research. Thus, they proceeded with their research without actively obtaining informed consent from forum members (Sugiura et al., 2017, p. 190).

Overall, their experiences show how difficult it can be to put research ethics for contemporary quantitative content analysis into practice. In consequence, it is not surprising that various researchers call for the development of ethical guidelines that specifically address such issues, to support future studies in navigating ethics when collecting and analyzing social media data (e.g., Hosseini et al., 2022, p. 15; Sugiura et al., 2017, p. 195; Taylor & Pagliari, 2018, p. 2). In addition, it becomes clear that (ethically) justifying the content analysis of social media posts simply by them being publicly available – as it is still done regularly (Roehse, 2022; Stommel & Rijk, 2021, p. 275) – is *not* enough (Salganik, 2018, p. 307). Rather, when they want to collect and analyze social media data, communication researchers are required to actively assess their project's ethical implications. This is necessary whether they decide to seek social media users' informed consent, or not. Specifically, they should reflect on how they can meet their responsibilities to ensure the four principles of ethical research – including the *respect for autonomy* – as far as possible.

Another ethical (and methodological) challenge, particularly when it comes to the principle of *justice*, stems from the *bias* of big data studies: While it may be practical to rely on social media data in order to gather voices and opinions on specific issues and generalize them to larger populations, it is important to keep in mind that users of social network sites are *not representative* of the general population (Hargittai, 2020, pp. 10–11). On the contrary, there are groups of people whose voices and opinions often are systematically excluded from this kind of research (Hargittai, 2020, p. 11), thus 'opposing' the principle of *justice* stating that access, benefits, and burdens of research should be distributed equally.

Notably, the messages derived from a particular social network site are often not even representative of the contents shared on this specific site, as their selection is mostly based on a limited number of hashtags or keywords (Hargittai, 2020, p. 12). And even if that is not the case, it is important to note that there is "no such thing as raw data" (Hosseini et al., 2022, p. 4). When we analyze, for instance, hate speech on social media, the data we refer to is most likely either interpreted by those who generated the data set, or the algorithms of the platform (Gitelman & Jackson, 2013, p. 1; Hosseini et al., 2022, p. 4). Additionally, there is the issue of potential manipulation by trolls or malicious actors, because it is hardly possible to specifically exclude data sourced from fake and bot accounts when generating such big data sets (Hosseini et al., 2022, p. 10). So, overall, in addition to being problematic from a perspective of *justice*, such biases may negatively affect validity of findings, and thus a study's *beneficence* and *nonmaleficence*, as well.

In the following section, I focus on the ethical challenges arising from the data collection in contemporary quantitative content analysis, i.e., the coding. Specifically, I discuss the role of student coders – as members of a vulnerable group (Podschuweit, 2021, p. 310) – and crowdsourcing in contemporary quantitative content analysis.

3.3 Coding

As stated in the beginning, *automatic* and *manual* approaches in content analysis primarily differ when it comes to the coding procedure (Kessler et al., 2023, p. 9): Automatic approaches rely on algorithms, whereas manual content analysis relies on *human coding* (Neuendorf, 2017, p. 40). More specifically, when conducting manual content analysis, we often work with *student* coders (Rössler, 2017, p. 230). Notably, different scientific associations relevant for our discipline offer recommendations on how to deal with *students as research subjects* (e.g., APA, 2017; DGPuK, 2017; DGPS, 2016). However, little information is found when it comes to protecting students from potential risks they may encounter during their work as *research assistants* (Podschuweit, 2021, p. 312). This is surprising, since students often depend on us researchers (as their supervisors in research projects, but also lecturers and examiners) in *multiple* ways, and thus can be considered a *vulnerable group*, for whom we have a special responsibility (Podschuweit, 2021, p. 310). At the same time, however, we frequently ask them to code contents (e.g., hate speech, images and videos of war and terror, or stigmatizing portrayals of health and illness) that can be disturbing and pose potential harms – especially to those who have to continuously face them during the coding process (Schlütz & Möhring, 2016, p. 490).[2] Considering that we often analyze such contents mainly *because* we assume that they have an impact on recipients, this practice is even more problematic. So, overall, working with students as coders is more ethically challenging than content-analytic research has often led on (Rössler, 2017, p. 230). Therefore, in the following, I would like to discuss possible solutions for a responsible ethical conduct, by drawing from the four principles of ethical research.

Firstly, we should enable students to make an informed choice on whether they want to participate in a research project – or not – by offering an extensive briefing regarding the kinds of content they would have to face during the coding process (*respect for autonomy*). In addition, it might be helpful to recruit students as research assistants who have prior knowledge on the research topic; because identifying stigmata or misinformation as such, and dealing with this information will be most likely easier for those familiar with the subject at hand. At the same time, however, we should be aware that familiarity might also come with an increased vulnerability, which leads us back to the necessity of an extensive briefing.

Moreover, it is our responsibility to regularly meet student coders during the data collection process: These meetings allow us to discuss the contents the students face during the coding, the potential harms arising from the contents (e.g., emotional distress), and different ways to deal with such harms.[3] At best, we organize the coding process in a very flexible manner, so that our research assistants can pause regularly from coding disturbing content (*nonmaleficence*). Even though payment is of course an important incentive, we should try to enable student coders (if they are interested) to also benefit professionally from their work in our research project. This includes involving them not only in the coding, but also other steps of our project, such as writing an abstract for a conference (and listing them as co-authors) or similar (*beneficence*). Finally, the workload, as well as the risks and benefits should be distributed equally among the different coders (*justice*).

I am aware that some of these recommendations could negatively affect the validity of the results: After all, to ensure external validity, coders' perception of media content should represent the perception of "average" recipients (Rössler, 2017, p. 157), which is not necessarily the case if all coders show prior knowledge and regularly discuss the contents with others. However, one could also argue that people (1) who are unfamiliar with a topic and then (2) face disturbing content on this issue regularly over a prolonged period (3) without any supervision or support might quickly – contrary to "average" recipients – become numb to these contents. This, in turn, can also affect their coding in a way that leads to invalid findings. So, either way, a critical assessment of the validity of the coding is crucial.

Alternatively, conducting an automated *instead* of a manual content analysis might be the right choice when it comes to disturbing content – presuming this also makes sense from a methodological point of view (Schlütz & Möhring, 2018,

2 Importantly, this also sets them apart from people being exposed to, for example, violent media contents in experimental research for only a short period.

3 While this is not the focus of this paper, I would like to point out that regular meetings with student coders throughout the coding process are also helpful when it comes to ensuring intercoder reliability.

p. 39). However, machine learning and other automated approaches often rely on manually labeled data and human coders as well (Barbosa & Chen, 2019, p. 1). Moreover, they can come with additional ethical challenges – particularly regarding the principle of *justice* – as they are frequently associated with crowdsourcing (Barbosa & Chen, 2019, p. 1; Shmueli et al., 2021, p. 1), i.e., "work practices based on crowd-based innovation and freelancing platforms" (Schlagwein et al., 2019, p. 2), such as, for instance, Amazon Mechanical Turk or TaskRabbit. In the past, crowdsourcing has been criticized for various aspects, including unfair payment and dehumanizing effects (Barbosa & Chen, 2019, p. 1). In these contexts, people who tend to be poor bear the costs of research, while others reap its benefits.

So, overall, when it comes to coding, contemporary quantitative content analyses pose various ethical challenges, particularly when student coders or crowd workers are involved. Therefore, as researchers, it is our responsibility not only to critically assess under which circumstances working with either of these groups is ethically appropriate, but to minimize the burdens and maximize the benefits for the coders. This might entail, for instance, enabling student coders to profit both financially and professionally from their work in our research project, or choosing a more expensive crowd working platform to ensure fair pay.

In empirical research, data collection is commonly followed by reporting, publication, and data management. The next section addresses potential ethical concerns arising during these final steps of a research project and the ways in which they can be met. Specifically, I discuss the importance of openly acknowledging sampling biases, anonymizing data, and following a data protection plan.

3.4 Reporting, publication, and data management

As stated earlier in this paper, one major ethical (but also methodological) challenge in contemporary quantitative content analysis is that our findings are commonly based on biased data. The bias itself (and, thus, its implications regarding the principle of *justice*) may not be preventable. However, simply relying on social media data in order to generalize attitudes and behaviors to larger populations, without critically assessing the limitations of such an approach, *is* preventable (Hargittai, 2020, p. 11). Therefore, future quantitative content analyses should reflect on these issues more actively, openly acknowledge the (potential) bias in their sample, and disclose limitations in an appropriate manner (Camfield, 2019, p. 14; Hargittai, 2020, p. 21).

Another challenge we face is that the data we are interested in (such as, in particular, social media data) often represents actual people whom we might cause harm by analyzing and reporting the information they share (Salganik, 2018, p. 307; Zook et al., 2017, 2) – remember, *all* data is *potentially* sensitive (Salganik, 2018, p. 309). One way to minimize such risks (*nonmaleficence*) is the *anonymization*, i.e., the process of removing obvious personal identifiers (such as names). However, this approach is deeply limited (Salganik, 2018, p. 307; Stommel & Rijk, 2021, p. 290), as even without personal data, identities may be easily deduced from people's postings and affiliations (Taylor & Pagliari, 2018, p. 4). For instance, in a recent study, Mason and Singh (2022, p. 93) have shown that for the majority of tweets quoted in 136 research papers, they were able to identify the original author of a tweet.

In consequence, it is to assume that all data are *potentially* identifiable (Salganik, 2018, p. 309). Thus, before publishing a study, researchers should try to identify potential points of reidentification in their data, and minimize them in their published results as much as possible (Zook et al., 2017, 4). For instance, one way to reduce informational risk in the reporting is to edit contents, such as tweets or comments, to such an extent that they are no longer recognizable, or to not include quotations at all (Mason & Singh, 2022, p. 108). A different approach would be to explicitly seek informed consent for quoting people's social media posts or similar (Mason & Singh, 2022, p. 106). Still, as indicated in chapter 3.2, this does not come without challenges either. Another option to reduce informational risk is following a *data protection plan* (i.e., a set of rules advising researchers how to deal with data)[4] which will decrease firstly the chance of data leaks, and secondly the harm if such a leak occurs nonetheless (Salganik, 2018, p. 312).

Finally, informational risk is particularly salient when we share our data with other researchers (Salganik, 2018, p. 312). Nevertheless, for some projects, data sharing is a crucial step (Zook et al., 2017, 4), as it may significantly increase the societal benefits of research (*beneficence*). A solution to this dilemma – and a way to share data that is ethically sound – might be to only share data with people who both meet certain criteria. These criteria could, for instance, include being informed on how to handle research data in an ethically appropriate manner and agreeing to follow certain rules (such as following a data protection plan themselves) in dealing with the data (Salganik, 2018, p. 313). At the same time, this is somewhat contradictory to the growing demand for *open science* in communication studies (e.g., Dienlin et al., 2021) which explicitly calls for publishing not only material and code, but also data "when appropriate and ethical" (Dienlin et al., 2021, p. 8). Thus, the final decision whether it is ethically appropriate to make data openly accessible (or not) is up to the individual researcher, therefore making a critical assessment of the informational risk essential. If it is deemed low, data sharing should not be problematic; if it is deemed high, it might be advisable to only share material and code. Here, institutional review boards or universities' data protection officers could offer helpful guidance (Salganik, 2018, p. 313).

As a way of compromising, and in order to support ethically sound data sharing among scientists, Lazer et al. (2020, p. 1061) recommend developing (1) large-scale, secure, privacy-preserving, shared infrastructures, which also include metadata describing the collection process as well as (2) structures that connect researchers with shared interests.

Table 2 summarizes the ethical challenges and methodological concerns discussed in this chapter.

4 The specifics of data protection plans vary. However, they generally follow a set of five rules: *safe projects* (limits data use to data that is ethically appropriate), *safe people* (limits data access to people that can be trusted, e.g., those actively involved in the research process or people who have undergone ethical training), *safe data* (data are anonymized and aggregated as far as possible), *safe settings* (data are stored with appropriate protection, e.g., physically in a locked room, password-protected, or similar), and *safe output* (research output is reviewed in a way that minimizes privacy breaches) (Desai et al., 2016, p. 5; Salganik, 2018, p. 312).

Table 2. Ethical challenges and methodological concerns throughout the research process in contemporary quantitative content analysis

	Challenges and concerns	Primary ethical principles in question
Research interest	*Purpose of research* • No clear purpose in many contemporary content analyses • Potential lack of validity	Beneficence and nonmaleficence
	Ethical appropriateness of overall study • Thorough assessment of ethical appropriateness necessary before conducting a study • Availability of data and technological abilities alone do not justify content-analytic research	Beneficence and nonmaleficence
Sampling and data collection	*Sample size* • As big as necessary to ensure statistical power and valid findings • As small as possible to minimize potential risks and burden	Beneficence and nonmaleficence
	Distinction between public and private data • Analysis of private data calls for informed consent; however: seeking informed consent before analyzing online interactions may affect the validity of results	Respect for autonomy Beneficence and nonmaleficence
	Bias of big data studies • Issue of distribution of the burdens of research as well as validity • Systematic exclusion of specific groups, no “raw” data, potential manipulation, e.g., through trolls	Justice
Coding	*Students as coders of potentially disturbing media content* • Members of a vulnerable group who often face potentially disturbing media contents • At the same time: conducting content-analytic research explicitly because we assume that certain contents may (negatively) affect their recipients	Beneficence and nonmaleficence
	Coding disturbing media content automatically • Machine learning or other automated approaches often rely on manually labeled data, and thus human coders, as well • Often associated with crowdsourcing and its additional ethical challenges	Beneficence and nonmaleficence Justice
Reporting, publication, and data management	*Anonymization* • Anonymization as crucial measure to minimize informational risk • Still: all data is potentially identifiable	Beneficence and nonmaleficence
	Data sharing • Data sharing may significantly increase the benefits of research (including open science), while also increasing the informational risk	Beneficence and nonmaleficence

Note. Overview of ethical challenges in contemporary quantitative content analysis, including the ethical principles that are affected primarily by a certain issue. Given beneficence and nonmaleficence require weighing up a study’s benefits and risks (deciding whether they strike the right balance) – and thus are highly related – both principles are presented in combination.

4. Conclusion

It can be concluded that conducting contemporary quantitative content analyses in a way that is both methodologically *and* ethically sound can be quite challenging. Specifically, ethical challenges can be found throughout the research process, from the primary research interest to the publication and data management. Moreover, this paper shows that there are still various uncertainties when it comes to conducting ethically responsible contemporary quantitative content-analytic research (Salganik, 2018, p. 293).

This is particularly true with regards to the challenges arising from the analysis of (big) data mined from social media (e.g., their potential bias, or deciding whether, or not, to obtain users' informed consent to collect and analyze their social media posts). At the same time, ethical challenges in contemporary content-analytic research are *not* limited to the analysis of social media contents. For instance, it is also possible for automatic approaches to more "traditional" media contents, such as war or health coverage in online news media, to lack a clear purpose, and, in consequence, valid findings (thus questioning the study's *beneficence* and *nonmaleficence*). Moreover, potential risks for (student) coders arising from coding, e.g., violent, or stigmatizing media contents, are also not limited to those shared on social media.

However, researchers should not be discouraged. On the contrary: Given that quantitative content analysis as a (supposedly) non-reactive method has been often dismissed in the ethical discussions in communication research (Heise, 2017, p. 769; Neuendorf, 2017, pp. 130–131), acknowledging that it *does* come with ethical challenges, as it has been done for some time now, is a crucial step to meet them. And although the computational turn (Hase et al., 2022, p. 60) in the field raises new ethical issues for contemporary quantitative content analysis, they are not insurmountable (Salganik, 2018, p. 324). Ethical responsible conduct of research is not simply about deciding whether a study is ethical or not, but about critically reflecting under which circumstances it is appropriate (Salganik, 2018, p. 324; Schlütz & Möhring, 2018, p. 34). As I have argued throughout this paper, the four principles of ethical research – *respect for a person's autonomy*, *beneficence*, *nonmaleficence*, and *justice* – offer helpful guidance for this process. Nevertheless, some dilemmas will prevail: As indicated earlier with regards to the risks and benefits of "Infoveillance" in an emerging pandemic, equally meeting all four principles of ethical research will not always be possible. This also applies to balancing methodological and ethical questions, as pointed out, for example, in the context of seeking users' informed consent to collect and analyze social media data. When facing such a dilemma, communication researchers should critically assess how they can ensure the four principles of ethical research as far as possible – without compromising methodological rigor, and, in consequence, validity or further quality criteria (specifically, *reliability*, *reproducibility*, *robustness*, and *replicability*).

Importantly, research ethics are continuous (Salganik, 2018, p. 322). In consequence, given both the rapid development of contemporary quantitative content analysis and the general changes to the field of communication research (including its computational turn as well as the growing demand for open science and thus calls for sharing data, e.g., Dienlin et al., 2021) they require an ongoing discourse (Schlütz & Möhring, 2018, p. 48). Most recently, researchers are also discussing the possibility of using the language model ChatGPT for analyzing content (e.g., Alshami et al., 2023, p. 1). If applied to (big) data gathered from social media, the use of ChatGPT raises further ethical questions: What happens to the data if it is uploaded to ChatGPT? What does the machine itself learn from the data, especially when it consists of disturbing contents, such as hate speech or similar? These and other aspects should be discussed critically in future research.

Against this backdrop, formulating and institutionalizing clear guidelines for approaching ethical questions in contemporary quantitative content analysis can be an important step to (1) reduce the likelihood of ethical issues and (2) do justice to the fact that research ethics and validity are closely linked (Hosseini et al., 2022, p. 15; Schlütz & Möhring, 2018, p. 48). Without such guidelines, "individual researchers are left to their own devices when it comes to acknowledging, facing, and dealing with ethical-methodological dilemmas" (Schlütz & Möhring, 2018, p. 48). These guidelines could be, for instance, informed by surveys or interviews among researchers who regularly conduct contemporary content analyses (e.g., Roehse, 2022) as well as (student) coders or research assistants (e.g., Podschuweit, 2021), but also representatives of institutional review boards. This way, the different experiences of those engaged in the process of conducting content-analytic research, and those regularly debating ethical issues in media and communication studies could be integrated.

Finally, ethical assessments should not only be addressed more thoroughly in academic teaching (Schlütz & Möhring, 2016, p. 490), but also in academic publications (Heise, 2017, p. 773) – for instance, by not only portraying measurements and reliabilities, but also explicitly addressing ethical challenges that have been met throughout the conduct of a quantitative content analysis.

References

Alshami, A., Elsayed, M., Ali, E., Eltoukhy, A. E. E., & Zayed, T. (2023). Harnessing the power of ChatGPT for automating systematic review process: Methodology, case study, limitations, and future directions. *Systems*, *11*(7), 351. https://doi.org/10.3390/systems11070351

APA. (2017). *Ethical principles of psychologists and code of conduct.* American Psychological Association. https://www.apa.org/ethics/code

Baden, C., Pipal, C., Schoonvelde, M., & van der Velden, M. A. C. G. (2022). Three gaps in computational text analysis methods for social sciences: A research agenda. *Communication Methods and Measures*, *16*(1), 1–18. https://doi.org/10.1080/19312458.2021.2015574

Barbosa, N. M., & Chen, M. (2019, May 4). *Rehumanized crowdsourcing: A labeling framework addressing bias and ethics in machine learning*. CHI Conference on Human Factors in Computing Systems Proceedings, Glasgow, Scotland.

Beauchamp, T. L. (2003). Methods and principles in biomedical ethics. *Journal of Medical Ethics*, *29*(5), 269–274. https://doi.org/10.1136/jme.29.5.269

Beauchamp, T. L., & Childress, J. F. (2019). *Principles of biomedical ethics* (8th edition). Oxford University Press.

Bishop, L., & Gray, D. (2018). Ethical challenges of publishing and sharing social media research data. In K. Woodfield (Ed.), *The ethics of online research* (pp. 159–188). Emerald Publishing.

Bond, T. (2012). Ethical imperialism or ethical mindfulness? Rethinking ethical review for social sciences. *Research Ethics*, *8*(2), 97–112. https://doi.org/10.1177/1747016112445419

Camfield, L. (2019). Rigor and ethics in the world of big-team qualitative data: Experiences from research in international development. *American Behavioral Scientist*, *63*(5), 1–18. https://doi.org/10.1177/0002764218784636

Carrig, M. M., & Hoyle, R. H. (2011). Measurement choices: Reliability, validity, and generalizability. In A. T. Panter & S. K. Sterba (Eds.), *Handbook of ethics in quantitative methodology* (pp. 127–158). Routledge.

Casas, A., & Williams, N. W. (2022). Introduction to the special issue on images as data. *Computational Communication Research*, *4*(1), 1–10. https://doi.org/10.5117/CCR2022.1.000.CASA

Chen, J., & Wang, Y. (2021). Social media use for health purposes: Systematic review. *Journal of Medical Internet Research*, *23*(5). https://doi.org/10.2196/17917

Cortina, J. M. (2020). On the whys and hows of quantitative research. *Journal of Business Ethics*, *167*(1), 19–29. https://doi.org/10.1007/s10551-019-04195-8

Crow, G., Wiles, R., Heath, S., & Charles, V. (2006). Research ethics and data quality: The implications of informed consent. *International Journal of Social Research Methodology*, *9*(2), 83–95. https://doi.org/10.1080/13645570600595231

Desai, T., Ritchie, F., & Welpton, R. (2016). *Five safes: Designing data access for research* (Economics Working Paper Series No. 1601). University of the West of England.

DGPS (Ed.). (2016). *Berufsethische Richtlinien des Berufsverbandes Deutscher Psychologinnen und Psychologen e.V. und der Deutschen Gesellschaft für Psychologie e.V.* [Professional ethical guidelines of the Professional Association of German Psychologists and the German Society for Psychology]. https://www.dgps.de/die-dgps/aufgaben-und-ziele/berufsethische-richtlinien/

DGPuK (Ed.). (2017). *Ethik-Kodex der DGPuK* [Code of ethics of the German Communication Association]. https://www.dgpuk.de/sites/default/files/Ethik-Kodex-der-DGPuK-vom-13.-Mai-2015-zuletzt-gea%CC%88ndert-am-31.-Ma%CC%88rz-2017_0.pdf

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., . . . Vreese, C. de (2021). An agenda for open science in communication. *Journal of Communication*, *71*(1), 1–26. https://doi.org/10.1093/joc/jqz052

Döveling, K., Sommer, D., Podschuweit, N., Geise, S., & Roessing, T. (2016). Kommunikationswissenschaftliche Forschungsethik im internationalen und interdisziplinären Vergleich [Communication research ethics in international and interdisciplinary comparison]. In K.-D. Altmeppen, L. Rinsdorf, P. Werner, & T. Pleil (Eds.), *Verantwortung – Gerechtigkeit – Öffentlichkeit: Normative Perspektiven auf Kommunikation* (pp. 1–26). Herbert von Halem.

Gitelman, L., & Jackson, V. (2013). Introduction. In L. Gitelman (Ed.), *Raw data is an oxymoron* (pp. 1–14). MIT Press.

Guttman, N., & Thompson, T. L. (2010). Ethics in health communication. In G. Cheney, S. May, & D. Munshi (Eds.), *The handbook of communication ethics* (pp. 293–308). Routledge.

Ha, Y., Park, K., Kim, S. J., Joo, J., & Cha, M. (2021). Automatically detecting image-text mismatch on Instagram with deep learning. *Journal of Advertising*, *50*(1), 52–62. https://doi.org/10.1080/00913367.2020.1843091

Hargittai, E. (2020). Potential biases in big data: Omitted voices on social media. *Social Science Computer Review*, *38*(1), 10–24. https://doi.org/10.1177/0894439318788322

Hase, V. (2023). Automated content analysis. In F. Oehmer-Pedrazzi, S. H. Kessler, E. Humprecht, K. Sommer, & L. Castro (Eds.), *Standardisierte Inhaltsanalyse in der Kommunikationswissenschaft – Standardized content analysis in communication research: Ein Handbuch – A handbook* (pp. 23–36). Springer.

Hase, V., Mahl, D., & Schäfer, M. S. (2022). Der „Computational Turn“: ein „interdisziplinärer Turn“? Ein systematischer Überblick zur Nutzung der automatisierten Inhaltsanalyse in der Journalismusforschung [The “computational turn”: An “interdisciplinary turn”? A systematic overview of the use of automated content analysis in journalism research]. *Medien & Kommunikationswissenschaft*, *70*(1-2), 60–78. https://doi.org/10.5771/1615-634X-2022-1-2-60

Heise, N. (2017). Warum das Rad neu erfinden? Gedanken zur Diskussion um Forschungsethik in der Kommunikationswissenschaft in Anknüpfung an den Beitrag von Daniela Schlütz und Wiebke Möhring in M&K 4/2016 [Why reinvent the wheel? Thoughts on the discussion of research ethics in communication studies drawing on the contribution by Daniela Schlütz and Wiebke Möhring in M&K 4/2016]. *Medien & Kommunikationswissenschaft*, *65*(4), 766–778. https://doi.org/10.5771/1615-634X-2017-4-766

Heise, N., & Schmidt, J.-H. (2014). Ethik der Onlineforschung [Ethics of online research]. In M. Welker, M. Taddicken, J.-H. Schmidt, & N. Jackob (Eds.), *Neue Schriften zur Online-Forschung: Vol. 12. Handbuch Online-Forschung: Sozialwissenschaftliche Datengewinnung und -auswertung in digitalen Netzen* (pp. 1–23). Herbert von Halem.

Hollingshead, W., Quan-Haase, A., & Chen, W. (2022). Ethics and privacy in computational social science: A call for pedagogy. In U. Engel, A. Quan-Haase, S. X. Liu, & L. Lyberg (Eds.), *Handbook of computational social science, Volume 1: Theory, case studies and ethics* (pp. 171–185). Routledge.

Hopp, F. R., & Weber, R. (2021). Rejoinder: How methodological decisions impact the validity of moral content analyses. *Communication Monographs*, *88*(3), 389–393. https://doi.org/10.1093/joc/jqz052

Hosseini, M., Wieczorek, M., & Gordijn, B. (2022). Ethical issues in social science research employing big data. *Science and Engineering Ethics*, *28*(3), 1–21. https://doi.org/10.1007/s11948-022-00380-7

Hunter, D. (2007). The roles of research ethics committees: Implications for membership. *Research Ethics*, *3*(1), 24–26. https://doi.org/10.1177/174701610700300110

ICA. (2019). *ICA code of ethics*. International Communication Association. https://www.icahdq.org/page/MissionStatement

Janis, I. (2009). The problem of validating content analysis. In K. Krippendorff & M. A. Bock (Eds.), *The content analysis reader* (pp. 358–366). Sage.

Kessler, S. H., Sommer, K., Humprecht, E., & Oehmer-Pedrazzi, F. (2023). Manuelle standardisierte Inhaltsanalyse [Manual standardised content analysis]. In F. Oehmer-Pedrazzi,

S. H. Kessler, E. Humprecht, K. Sommer, & L. Castro (Eds.), *Standardisierte Inhaltsanalyse in der Kommunikationswissenschaft – Standardized content analysis in communication research: Ein Handbuch – A handbook* (pp. 9–22). Springer.

Kozinets, R. V. (2006). Netnography 2.0. In R. Belk & R. W. Belk (Eds.), *Handbook of qualitative research methods in marketing*. Edward Elgar. https://doi.org/10.4337/9781847204127.00018

Krieger, B., Grubmüller, V., & Schäfer, C. (2014). Ethische Herausforderungen bei der sozialwissenschaftlichen Analyse von Social-Media-Inhalten [Ethical challenges in the social scientific analysis of social media content]. *SWS-Rundschau*, *54*(2), 201–216.

Krippendorff, K. (2021). A quadrilogy for (big) data reliabilities. *Communication Methods and Measures*, *15*(3), 165–189. https://doi.org/10.1080/19312458.2020.1861592

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

Lewis, S. C., Zamith, R., & Hermida, A. (2013). Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of Broadcasting & Electronic Media*, *57*(1), 34–52. https://doi.org/10.1080/08838151.2012.761702

Mahl, D., Nordheim, G. von, & Guenther, L. (2022). Noise pollution: A multi-step approach to assessing the consequences of (not) validating search terms on automated content analyses. *Digital Journalism*, 1–23. https://doi.org/10.1080/21670811.2022.2114920

Mark, M. M., & Lenz-Watson, A. L. (2011). Ethics and the conduct of randomized experiments and quasi-experiments in field settings. In A. T. Panter & S. K. Sterba (Eds.), *Handbook of ethics in quantitative methodology* (pp. 185–209). Routledge.

Mason, S., & Singh, L. (2022). Reporting and discoverability of "Tweets" quoted in published scholarship: Current practice and ethical implications. *Research Ethics*, *18*(2), 93–113. https://doi.org/10.1177/17470161221076948

Maxwell, S. E., & Kelley, K. (2011). Ethics and sample size planning. In A. T. Panter & S. K. Sterba (Eds.), *Handbook of ethics in quantitative methodology* (pp. 159–180). Routledge.

McKee, H., & Porter, J. E. (2009). *The ethics of internet research: A rhetorical, case-based process*. Peter Lang.

Nelson, L. K. (2020). Computational grounded theory: A methodological framework. *Sociological Methods & Research*, *49*(1), 3–42. https://doi.org/10.1177/0049124117729703

Neuendorf, K. A. (2017). *The content analysis guidebook*. Sage Publications.

Olteanu, A., Castillo, C., Diaz, F., & Kıcıman, E. (2019). Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data*, *2*, 1–13. https://doi.org/10.3389/fdata.2019.00013

Panter, A. T., & Sterba, S. K. (2011). Ethics in quantitative methodology: An introduction. In A. T. Panter & S. K. Sterba (Eds.), *Handbook of ethics in quantitative methodology* (pp. 1–14). Routledge.

Podschuweit, N. (2021). How ethical challenges of covert observations can be met in practice. *Research Ethics*, *17*(3), 309–327. https://doi.org/10.1177/17470161211008218

Riffe, D., Lacy, S., & Fico, F. G. (1998). *Analyzing media messages: Using quantitative content analysis in research*. Lawrence Erlbaum.

Roehse, E.-M. (2022, October 5). *Social Media Daten als forschungsethische Herausforderung der Inhaltsanalyse: Die Sicht der Forschenden* [Social media data as research ethical challenge of content analysis: The point of view of researchers]. 23rd conference of the methods division of the German Communication Association, Munich, Germany.

Rössler, P. (2017). *Inhaltsanalyse* [Content analysis]. (3rd ed.). UVK.

Salganik, M. J. (2018). *Bit by bit: Social research in the digital age*. Princeton University Press.

Schlagwein, D., Cecez-Kecmanovic, D., & Hanckel, B. (2019). Ethical norms and issues in crowdsourcing practices: A Habermasian analysis. *Information Systems Journal*, *29*(4), 1–27. https://doi.org/10.1111/isj.12227

Schlütz, D., & Möhring, W. (2016). Kommunikationswissenschaftliche Forschungsethik – Sonntagsworte, Selbstzweck, Notwendigkeit? [Communication science research ethics – sunday words, an end in themselves, necessity?]. *Medien & Kommunikationswissenschaft*, *64*(4), 483–496. https://doi.org/10.5771/1615-634X-2016-4-483

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

Shmueli, B., Fell, J., Ray, S., & Ku, L.-W. (2021, June 6-11). *Beyond fair pay: Ethical implications of NLP crowdsourcing*. 2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics.

Signorielli, N. (2008). Research ethics in content analysis. In A. Jordan, D. Kunkel, J. Manganello, & M. Fishbein (Eds.), *Media messages and public health* (pp. 88–96). Routledge.

Song, H., Tolochko, P., Eberl, J.-M., Eisele, O., Greussing, E., Heidenreich, T., Lind, F., Galyga, S., & Boomgaarden, H. G. (2020). In validations we trust? The impact of imperfect human annotations as a gold standard on the quality of validation of automated content analysis. *Political Communication*, *37*(4), 550–572. https://doi.org/10.1080/10584609.2020.1723752

Stommel, W., & Rijk, L. de (2021). Ethical approval: None sought. How discourse analysts report ethical issues around publicly available online data. *Research Ethics*, *17*(3), 275–297. https://doi.org/10.1177/1747016120988767

Sugiura, L., Wiles, R., & Pope, C. (2017). Ethical challenges in online research: Public/private perceptions. *Research Ethics*, *13*(3-4), 184–199. https://doi.org/10.1177/1747016116650720

Taylor, J., & Pagliari, C. (2018). Mining social media data: How are research sponsors and researchers addressing the ethical challenges? *Research Ethics*, *14*(2), 1–39. https://doi.org/10.1177/1747016117738559

Vail, A., Tully, M., Brabin, L., Roberts, S., & McNamee, R. (2009). Methodological considerations in ethical review – 2.: Are the study aims justified and is the design appropriate? *Research Ethics*, *5*(2), 85–88. https://doi.org/10.1177/174701610900500212

van Atteveldt, W., & Peng, T.-Q. (2018). When communication meets computation: Opportunities, challenges, and pitfalls in computational communication science. *Communication Methods and Measures*, *12*, 1–12. https://doi.org/10.1080/19312458.2018.1458084

WAPOR. (2017). *WAPOR code of professional ethics and practices*. World Association of Public Opinion Research. https://wapor.org/about-wapor/code-of-ethics/

Wassenaar, D. R., & Mamotte, N. (2013). Ethical issues and ethics reviews in social science research. In A. Ferrero, Y. Korkut, M. M. Leach, G. Lindsay, & M. J. Stevens (Eds.), *The Oxford handbook of international psychological ethics* (pp. 1–41). Oxford University Press.

Wiles, R. (2012). *What are qualitative research ethics?* Bloomsbury Academic. http://www.doabooks.org/doab?func=fulltext&rid=15451 https://doi.org/10.5040/9781849666558

Williams, M. L., Burnap, P., Sloan, L., Jessop, C., & Lepps, H. (2018). Users' views of ethics in social media research: Informed consent, anonymity, and harm. In K. Woodfield (Ed.), *The ethics of online research* (pp. 27–52). Emerald Publishing.

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), 1-10. https://doi.org/10.1371/journal.pcbi.1005399

Zyphur, M. J., & Pierides, D. C. (2017). Is quantitative research ethical? Tools for ethically practicing, evaluating, and using quantitative research. *Journal of Business Ethics*, *143*(1), 1–16. https://doi.org/10.1007/s10551-017-3549-8