Online Media Glob. Commun. 2023; 2(1): 122–143 DE GRUYTER MOUTON

Featured Translated Research Outside the Anglosphere

Valerie Hase*, Daniela Mahl and Mike S. Schäfer

The "computational turn": an "interdisciplinary turn"? A systematic review of text as data approaches in journalism studies

https://doi.org/10.1515/omgc-2023-0003
Received February 23, 2023; accepted February 25, 2023

Abstract: Possibilities of applying automated content analysis in journalism studies include, for example, machine learning to identify topics in journalistic coverage or measuring news diffusion via automated approaches. But how have computational methods been applied thus far? And what are consequences of the "computational turn" in communication science, especially concerning interdisciplinarity? Based on a systematic literature review, this article summarizes the use of automated content analysis in journalism studies. Results illustrate an increasing use of the method by communication scientists, as yet another indicator of methodological interdisciplinarity in communication science. However, there is little evidence of an increase in theoretical interdisciplinarity: Studies relying on computational methods do not increasingly refer to theories from other disciplines. With respect to practical interdisciplinarity, for instance collaborations, our discipline is by no means becoming more interdisciplinary. Instead, we find a shift in favor of technical disciplines. At least up to now, the "computational turn" in

Article Note: This article was originally published in German: Valerie Hase, Daniela Mahl and Mike S. Schäfer (2022). "Der „Computational Turn": ein „interdisziplinärerTurn"? Ein systematischer Überblick zur Nutzung der automatisierten Inhaltsanalyse in der Journalismusforschung" *Medien & Kommunikationswissenschaft*, 70(1–2): 60–78. Permission to translate by *Medien & Kommunikationswissenschaft*. Translators: Sitian Chen and Yinglei Zang, Shanghai International Studies University. Copy Editor: Dane Claussen.

***Corresponding author: Valerie Hase**, Department of Media and Communication, LMU Munich, Munich, Germany, E-mail: valerie.hase@ifkw.lmu.de
Daniela Mahl and Mike S. Schäfer, Department of Communication and Media Research, University of Zurich, Zurich, Switzerland, E-mail: d.mahl@ikmz.uzh.ch (D. Mahl), m.schaefer@ikmz.uzh.ch (M.S. Schäfer)

communication science should thus not be equated with an “interdisciplinary turn.”

Keywords: automated content analysis; computational communication science; computational methods; computational social science; interdisciplinarity; journalism studies

1 Introduction

“Computational methods” such as simulation models, network analyses or automated content analyses, have gained importance in communication science. This is also evident in the establishment of fields such as “computational social science” (CSS) or “computational communication science” (CCS) (Strippel et al. 2018). Because computational methods often stem from other disciplines, such as computer or information science, Hepp et al. (2021) argue that they transform communication science beyond a simple “computational turn” of its methodological repertoire, for instance concerning interdisciplinarity within the field (Theocharis and Jungherr 2021; Windsor 2021).

This study analyzes the application of computational methods, more specifically the use of automated content analysis in journalism studies, and shifts in interdisciplinarity within this research field: to what extent do communication scholars use computational methods? (*methodological interdisciplinarity*). To what extent do they work with theories from outside of communication science? (*theoretical interdisciplinarity*). And how does the “computational turn” affect interdisciplinary collaborations? (*practical interdisciplinarity*).

Focusing on *one* computational method (automated content analysis) and *one* research area closely related to communication science (journalism studies), we shed light on these questions. On the one hand, manual content analysis is one of the few genuine research methods in our discipline (Loosen and Scholl 2012); however, in its automated form, the method is used across disciplines (DiMaggio 2015) – also by communication scientists (Baden et al. 2022). Accordingly, questions revolving around interdisciplinarity play an important role here (Laugwitz 2020). On the other hand, manual content analysis is part of our standard methodological toolkit, especially in the field of journalism studies (Hanitzsch and Engesser 2014; Löffelholz and Rothenberger 2011). However, automated content analysis increasingly gains in importance as well, which has fostered discussions about the use of automated methods in journalism studies (Boczek and Hase 2020; de Grove et al. 2020).

2 Interdisciplinarity in communication science

Interdisciplinarity describes the integration of theories, data, or methods from different disciplines (Wagner et al. 2011). Following and expanding on the definitions of Klein (2017) and von Nordheim et al. (2021), we distinguish between methodological, theoretical, and practical interdisciplinarity:

- *Methodological interdisciplinarity* refers to the use of methods developed by other disciplines, e.g., the use of automated content analysis as a method partly originating from computer or information science in communication science.
- *Theoretical interdisciplinarity* refers to the use of theories developed by other disciplines, for example the use of complexity theory in communication science.
- *Practical interdisciplinarity* includes the dissolution of disciplinary boundaries in scholarly work, e.g., communication scholars publishing with researchers from other disciplines or in journals from other disciplines.

Communication science has always been considered as being comparatively interdisciplinary due to its interlinkages with sociology, political science, or psychology (Walter et al. 2018; Zhu and Fu 2019). Many of the methods commonly employed by our discipline such as experiments were predominantly developed in other disciplines like psychology. Important theories such as framing or structuration theory also have strong roots in other disciplines like psychology and sociology. Interdisciplinary collaborations are also not uncommon in our discipline (Walter et al. 2018). Although discussions around interdisciplinarity are therefore by no means new, computational methods or the "computational turn" in communication science have certainly reignited them (Theocharis and Jungherr 2021; Windsor 2021).

However, as Jacobs and Frickel (2009) criticize, there are hardly any empirical analyses of how interdisciplinary specific disciplines are. Instead, discourses around interdisciplinarity are often shaped by personal experiences—a problem that, according to Zhu and Fu (2019), also affects communication science. In addition, implicit assumptions dominate, such as interdisciplinarity necessarily facilitating solutions for scientific problems, meaning that interdisciplinarity should be understood as an advantage or even a norm (Jacobs and Frickel 2009; Zhu and Fu 2019). However, interdisciplinarity brings about both opportunities and risks, especially with regard to computational methods. Such risks include a lack of methodological standards for applying these methods, a lack of theoretical embeddedness in much of existing computational work, and uncertainties for researchers pursuing interdisciplinary careers within CSS (Theocharis and Jungherr 2021; Windsor 2021).

To illustrate the extent to which the "computational turn" in communication science can be understood as an "interdisciplinary turn" and to underline both opportunities and risks associated with it, we here focus on a specific method and a specific research area: the use of automated content analysis in journalism studies.

3 Automated content analysis in journalism studies

First, we turn to definitions of key terms: What do we mean by journalism studies as a research field and automated content analysis as a method?

3.1 Key terms

We here understand journalism studies as research on the *public use of words, images and sounds by journalists* (Zelizer 2017). This perspective excludes communication by recipients, which is often also studied as a form of audience participation (Loosen 2016). Journalism studies is closely related to communication science but constitutes an independent, interdisciplinary subfield (Hanitzsch and Engesser 2014; Löffelholz and Rothenberger 2011; Steensen and Ahva 2015). Accordingly, both communication scholars but also researchers from other disciplines engage in journalism studies.

Since automated content analysis is a method frequently used (Baden et al. 2022) and discussed (Boczek and Hase 2020; de Grove et al. 2020) in journalism studies, we further focus on this particular method employed to (partially) automate content analysis (Benoit 2020; Günther and Quandt 2016). Such methods include rule-based approaches, e.g., the identification of similar content via similarity metrics, and dictionaries, i.e., word lists. For the latter, a distinction is made between "off-the-shelf" dictionaries, which are often used across genres and topics, and organic dictionaries, which are developed specifically for genres and topics. Other methods include supervised machine learning, which uses algorithms to classify content into predetermined categories, or unsupervised machine learning, such as exploratory analysis of texts via topic modeling. For an overview of the method, we refer readers to Benoit (2020), Grimmer and Stewart (2013), and Günther and Quandt (2016) as well as to Williams et al. (2020) for the analysis of visual content. In the German-speaking regions, work by Geise et al. (2016), Günther (2021), Scharkow (2012), and Wettstein (2016) is particularly noteworthy.

3.2 How could journalism studies use automated content analysis?

In order to illustrate the importance of automated content analysis for journalism studies, we now discuss which theories/concepts or variables can be studied using this method. Based on a framework by Boczek and Hase (2020), which we modified by introducing the public sphere model of Capra (1996) and Waldherr (2017), we illustrate relevant theories/concepts and variables.

First, automated content analysis can be used to analyze "elements" of journalistic communication, i.e., the occurrence of formal features, actors/places, events/topics, and semantic/syntactic properties of language. *Formal aspects* include, but are not limited to, metadata of articles – e.g., timestamps, by means of which we can analyze news diffusion (Buhl et al. 2018). The occurrence of *actors/locations* is often analyzed via "named-entity recognition" to automatically identify, for example, people, organizations, or places mentioned in texts. This is often

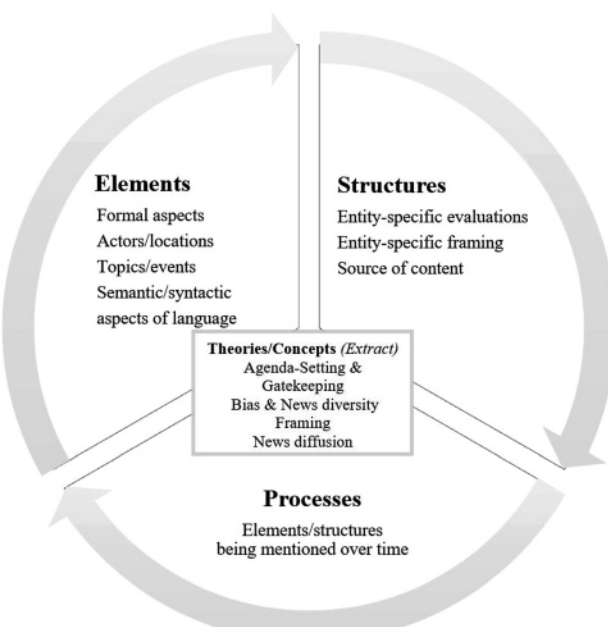

Figure 1: Using automated content analysis in journalism studies. *Note:* Figure based on Capra (1996) and Waldherr (2017).

used to grasp personalization within studies focusing on news values (Burggraaff and Trilling 2020). The automated identification of *topics/events* has become a separate line of research that is also of interest to communication science (Günther 2021; Trilling and van Hoof 2020). Finally, several studies address—often critically—automated measurements of *semantic/syntactic aspects of language*, for instance via sentiment analysis (van Atteveldt et al. 2021).

A second area of application deals with "structures" of journalistic communication, i.e., how content is evaluated, framed, or pushed forward by specific actors. A key point of discussion is the extent to which not only the occurrence of evaluative terms in text can be measured, but also the degree to which *entity-specific evaluations* are used. Such analyses can be employed, for example, to analyze media bias (Hamborg et al. 2019). Similarly, scholars debate whether more complex concepts, e.g., *entity-specific framing*, can be analyzed via automated content analysis (Nicholls and Culpepper 2021). Moreover, the *source of content*, for instance who is quoted in texts, can also be analyzed via automated means, for example to study agenda-setting or gatekeeping.

Third, computational methods help to trace "processes" such as *elements and structures being mentioned in texts over time*—something also referred to as a "temporal turn" (Wells et al. 2019).

3.3 Research questions

To date, there is a lack of empirical analyses on how automated content analysis is used to study specific theories/concepts or variables (except for Baden et al. 2022). Moreover, consequences of the so-called computational turn, especially for interdisciplinarity within the field, are frequently discussed (Theocharis and Jungherr 2021; Windsor 2021) but not empirically examined. Therefore, we focus on two key research questions (RQs) answered through a systematic literature review. Here, readers should note that we examine the use of automated content analysis within the field of journalism studies but discuss potential consequences for the broader context of communication science.

RQ1: How is automated content analysis used in the field of journalism studies?

RQ2: What are consequences of the use of automated content analysis for methodological, theoretical, and practical interdisciplinarity in communication science more broadly?

4 Methodological approach: systematic literature review

Our literature review includes two samples: (1) a CSS sample, i.e., studies that use automated content analysis to analyze journalistic communication or methodologically advance the method for journalism studies, and (2) a benchmark sample, i.e., studies using mostly manual content analysis within journalism studies. An overview of key studies related to specific variables (Table A1), our samples (Table A2), and info on the operationalization of variables (Table A3) can be found in the German-language appendix to this article via https://doi.org/10.17605/OSF.IO/BZQE6.

4.1 CSS sample

We used the Scopus database to identify relevant CSS studies. Compared to the Web of Science and its Social Science Citation Index (SSCI), Scopus includes more journals and especially non-English language journals (Mongeon and Paul-Hus 2016). At the same time, it more strongly focuses on scientific publications than Google Scholar. Since Scopus predominantly covers English-language journals, the online archives of the three most established German-language journals *(Medien & Kommunikationswissenschaft [M&K], Publizistik*, and *Studies in Communication and Media [SCM])* were also searched. Following Song et al. (2020), we included articles in peer-reviewed journals, books, book chapters, and conference publications which included the following terms in their title, abstract, or in their keywords: (computer assisted OR automated OR automatic OR computational) AND (content analysis OR text analysis OR visual analysis) AND (journalis* OR news*).[1]

We took into account all publications up to 2020, although readers should note that the availability of databases varied.[2] This led to a preliminary sample of $N = 435$ publications (see Figure 2). After removing seven duplicates, relevant studies were identified based on the abstract and, where necessary, the full text. We coded studies as relevant if they were (1) empirical, meaning they employed automated content analysis for empirical analyses or if they were methodological, meaning they advanced automated content analysis as a method. We also only coded them as

1 For German-language publications, articles were manually checked via equivalent search terms: (computer-unterstützt OR automatisiert OR automatisch OR digital) AND (Inhaltsanalyse OR Textanalyse OR visuelle Analyse) AND (journalis* OR nachricht*).

2 The availability of publications in Scopus (continuously since 1996) and in the online archives of M&K and Publizistik (both since 2000) differs. For SCM, all studies published since 2011 were retrieved.

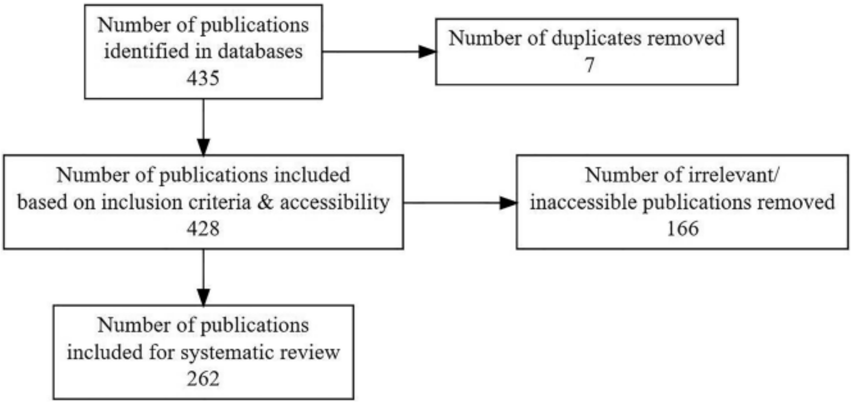

Figure 2: Identification of relevant studies.

relevant if their focus was on (2) journalistic communication as the public use of words, images, and sounds by journalists (Zelizer 2017). The first and the second author conducted all manual coding. After an intercoder test (N = 43, α = 0.9), we coded the relevance of articles in the preliminary sample based on these two inclusion criteria (0 = not relevant, 1 = relevant). After excluding 147 non-relevant and 19 inaccessible publications, N = 262 studies were included for the systematic review (see Section A2 of the Appendix for a list of all publications).

After another intercoder test (N = 27, α_{min} = 0.81), a range of bibliographic-formal, theoretical-conceptual, and methodological-empirical variables were coded (see Section A3 of the Appendix for details).

4.1.1 Bibliographic-formal variables

We coded the *discipline of each author* (α = 0.92) based on their institute affiliation according to the article. In addition, we coded the *discipline of the publication medium.* Journals were classified as “communication science” or “non-communication science” based on the SSCI classification in the Journal Citation Report. Additional information was used for journals, conference publications, and monographs not listed in the SSCI.[3] Furthermore, we coded each *study type* (α = 1), i.e., whether an

3 For non-SSCI-listed journals and conference publications, we referred to their respective websites. Monographs were coded based on their blurb. This variable was not collected as part of manual coding because it was collected at the publication level. The first and second author both coded the disciplinary affiliation of the publication medium. Differences (N = 4) were clarified.

article was empirical or methodological in nature. For methodological studies, coding ended at this point.

4.1.2 Theoretical-conceptual variables

For empirical studies, we coded the extent to which *references to theories/concepts* (α = 0.81) were made. Theories/concepts were derived and extended based on existing handbooks (Löffelholz and Rothenberger 2016) or previous reviews on automated content analysis (Boczek and Hase 2020; de Grove et al. 2020). In addition, we noted the *degree of deductive or inductive orientation* in each article (α = 1), i.e., whether theoretical assumptions were formulated as open research questions or as closed hypotheses.

4.1.3 Methodological-empirical variables

In addition, we coded the *unit of analysis* (α = 1), i.e., whether articles used automated content analysis to study textual, visual, or audio-visual content. We also coded *employed methods* (α_{min} = 0.86 across these five variables), i.e., whether articles used rule-based approaches, organic dictionaries, "off-the-shelf" dictionaries, supervised machine learning, or unsupervised machine learning. Based on our model (see Figure 1), we also analyzed which variables (α_{min} = 0.87 across these seven variables) were measured via automated content analysis: from formal features[4] as an example for elements to entity-specific evaluations as an example of structures to the occurrence of elements over time as an example of processes. We also coded whether studies used a *multi-method design* (α = 0.94) and the extent to which they employed *validation tests* (α = 1). Automated content analyses need to be validated (Grimmer and Stewart 2013), for example by comparing automated and manual codings. Using metrics like precision or recall, validation tests then analyze the extent to which manual and automated coding overlaps. Following Song et al. (2020), we coded whether at least one of these metrics was reported.

4.2 Benchmark sample

We then coded an additional "benchmark" sample of N = 262 studies. This benchmark sample includes studies that examine journalistic communication via (mostly manual) content analysis. It thus represents the breadth of studies employing any

4 This variable was generated inductively after initial construction of the codebook based on open codings.

type of content analysis in journalism studies. This benchmark was used to illustrate whether studies using computational methods, i.e., the CSS sample, differ from research in journalism studies more generally, i.e., the benchmark sample. We did not explicitly exclude the few studies using computational methods in the benchmark sample since the field as a whole was intended to serve as a parameter of comparison. To identify studies for the benchmark sample, we used the same search parameters as before but excluded references to computational methods: (content analysis OR text analysis OR visual analysis) AND (journalis* OR news*).

After an intercoder test ($N = 42$, $\alpha = 0.94$), a total of $N = 262$ relevant articles were identified. For these, after another intercoder test, the *discipline of each author* ($N = 26$, $\alpha = 0.75$) and the *discipline of the publication medium* were, again, coded.

5 Results

In the following, we provide an overview of how automated content analysis is used in journalism studies (RQ1) before discussing consequences for methodological, theoretical, and practical interdisciplinarity in communication science (RQ2).

5.1 On the application of automated content analysis in journalism studies

With regard to our first research question, we find related to the variable *study type* that half of the studies in the CSS sample empirically analyze journalistic communication via automated content analysis (49.6%). The other half is concerned with methodologically advancing automated content analysis (50.4%). In terms of the smaller sample of empirical studies for which we coded additional variables ($N = 130$), analyzing the *deductive or inductive orientation* of studies employing computational methods shows that most are somewhat more inductively oriented. In more than half of all studies in the CSS sample, authors did either not use any hypotheses/research questions but analyzed data without theoretical assumptions or they predominantly used open research questions (56.9%). Studies less often focused more on hypotheses (41.6%) or used research questions and hypotheses equally often (1.5%).

Related to the *unit of analysis*, automated procedures are almost exclusively used to analyze text (98.5%); very few studies considered visual content (1.5%). The spectrum of *employed methods* is more diverse: In addition to rule-based methods (33.8%), studies used dictionaries, with organic dictionaries (46.9%) dominating over

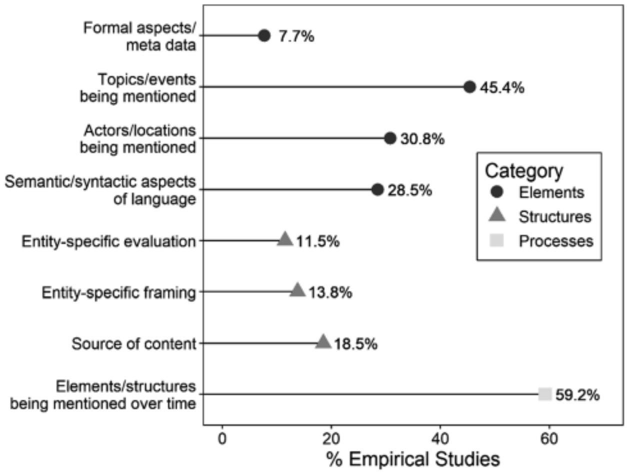

Figure 3: Variables. *Note: N* = 130 (empirical studies, a single study may include several variables).

"off-the-shelf" dictionaries (20%). Unsupervised machine learning (30%), especially topic modeling, and supervised learning (18.5%) were also used.

Related to *variables*, Figure 3 shows that automated content analysis is predominantly used to analyze processes of journalistic communication (59.2%). For example, studies resort to timestamps to trace news diffusion (Buhl et al. 2018). Elements, e.g., which actors/locations (30.8%) are mentioned or the semantic/syntactic aspects of language (28.5%), are also examined. For example, Burggraaff and Trilling (2020) analyze how often news articles mention persons to capture news values such as personalization. Jonkman et al. (2020) combine automated content analysis with a panel survey to analyze how often journalists report on business enterprises to demonstrate agenda-setting effects. More complex structures – e.g., the source of content (18.5%) or entity-specific evaluations (11.5%) – are rarely studied. An exception here is a study by Kroon et al. (2021) on stereotypical representations of minorities. More than one third of all studies use a *multi-method design* (39.2%), i.e., combine automated content analysis for instance with manual content analyses (20.8%) or standardized surveys (6.9%). *Validation tests*, in this case by comparing automated with manual coding, are reported in 40.8% of all studies.

5.2 Interdisciplinarity in communication science

With regard to our second research question, we are interested in how the use of automated content analysis may affect interdisciplinarity in communication science.

5.2.1 Methodological interdisciplinarity

How and how often is automated content analysis as a method originating from computer science used by communication scientists? When only considering studies displaying a clear disciplinary affiliation of authors (*N* = 204), Figure 4 shows that automated content analysis has gained relevance across disciplines. Its methodological development has been advanced since the 2000s; its empirical application has increased only in the last few years. That the method is also being used by our discipline points to a growth of methodological interdisciplinarity in communication science.

However, this increase in methodological interdisciplinarity is primarily driven by empirical applications of computational methods: communication scientists are involved in only a fraction of methodological studies (16.5% with communication science; 83.5% without communication science). Instead, our discipline focuses on empirically applying the method (64.4% with communication science; 35.6% without communication science). Our review also points out differences in *employed methods*: As Table 1 shows, studies involving authors from communication science tend to make greater use of dictionaries. Machine learning, on the other hand, is used less frequently. However, these results should be viewed with caution due to the small number of cases in this subsample and comparably small differences between groups. While there are no clear differences with respect

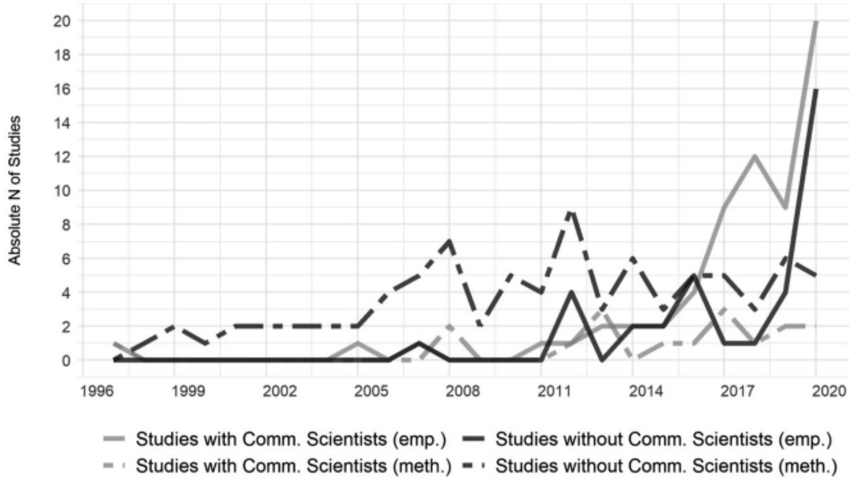

Figure 4: Disciplinary differences: use of automated content analysis. *Note:* *N* = 204 (empirical and methodological studies with disciplinary affiliation); with/without CS describes studies with/without participation of authors from communication science.

Table 1: Disciplinary differences: methods.

Method	**Number of publications (%)**	
	With comm. scientists	**Without comm. scientists**
Rule-based approaches	24 (36.9%)	13 (36.1%)
Organic dictionaries	32 (49.2%)	14 (38.9%)
"Off-the-shelf"-dictionaries	12 (18.5%)	6 (16.7%)
Unsupervised machine learning	16 (24.6%)	12 (33.3%)
Supervised machine learning	12 (18.5%)	8 (22.2%)
N	65	36

Note: N = 101 (empirical studies with clear disciplinary affiliations, publications could be assigned several methods); *with/without comm. sientists* describes whether publications were published by at least one author affiliated with a communication department or not.

to the analysis of specific *variables* or the use of *multi-method designs*, results are more often subjected to *validity tests* when our discipline is involved (47.7% with communication science; 36.1% without communication science).

5.2.2 Theoretical interdisciplinarity

Do communication scientists use theories from outside their own discipline when employing computational methods? Table 2 shows that *references to theories/concepts* often embed studies within classical frameworks, e.g., framing (20% with communication scientists; 27.8% without communication scientists) or agenda-setting (23.1% with communication scientists; 2.8% without communication

Table 2: Disciplinary differences: theories/concepts.

Theories/concepts	**Number of publications (%)**	
	With comm. scientists	**Without comm scientists**
Agenda-setting	15 (23.1%)	1 (2.8%)
Bias etc.	3 (4.6%)	2 (5.6%)
Emotionalization etc.	5 (7.7%)	0 (0%)
Framing	13 (20%)	10 (27.8%)
News diffusion etc.	4 (6.2%)	1 (2.8%)
Other	15 (23.1%)	9 (25%)
Theory/concept missing	10 (15.4%)	13 (36.1%)
N	65	36

Note: N = 101 (empirical studies with clear disciplinary affiliations); *with/without comm. scientists* describes whether publications were published by at least one author affiliated with a communication department or not. *Others* includes theories/concepts mentioned in less than $N < 5$ studies out of all $N = 101$ studies.

scientists). They also rely on concepts like media bias (4.6% with communication scientists; 5.6% without communication scientists). Thus, studies do reference theories/concepts and often those with interdisciplinary origins in psychology or sociology, especially if our discipline is involved. However, these theories/concepts are also part of the theoretical repertoire of communication science outside of its “computational turn” (Steensen and Ahva 2015).

5.2.3 Practical interdisciplinarity

Do communication scholars cooperate in interdisciplinary teams or publish outside their discipline when conducting journalism research via computational methods? Figure 5 visualizes the discipline of authors across studies, both for studies focusing on automated methods (CSS sample) and the field as a whole (benchmark sample). Surprisingly, the use of computational methods may not lead to communication scientists publishing more in interdisciplinary teams: The proportion of studies in which communication scientists cooperate with authors from other disciplines is the same across samples (9.5% CSS sample; 9.5% benchmark sample). However, there is a

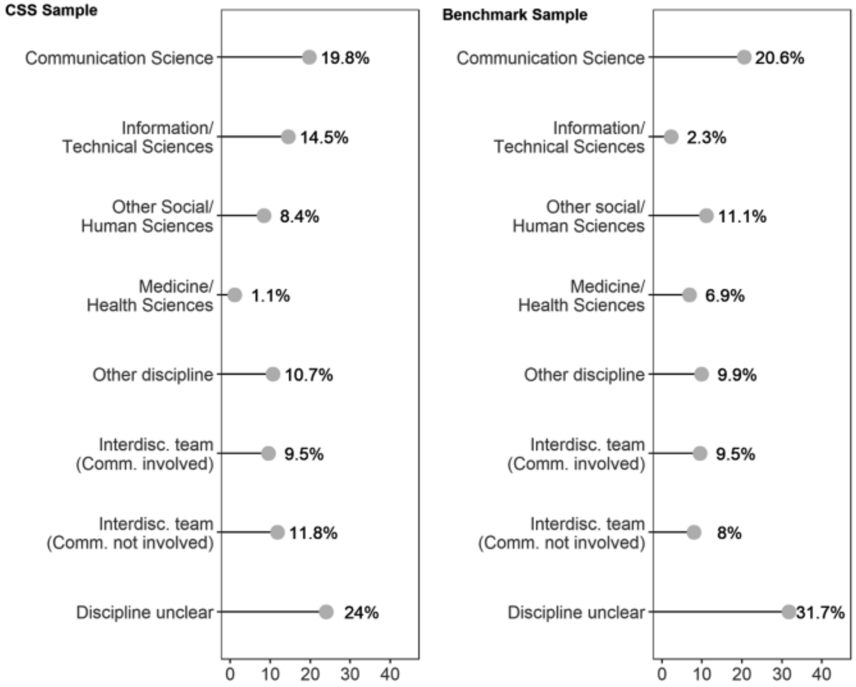

Figure 5: Authors in the field of journalism studies. *Note:* N = 262 studies in both samples.

shift with regard to which other disciplines—whether in cooperation with our discipline or on their own—conduct research in the realm of journalism studies. In the CSS sample, studies are more often conducted by researchers in information science, engineering, or technology without participation of our discipline (14.5% CSS sample; 2.3% benchmark sample), whereas medicine/health sciences or humanities/social sciences, for example, lose influence. If we look at who communication scientists collaborate with, we find that they cooperate more frequently with information scientists, engineers, and technical scientists, especially in interdisciplinary teams, when using automated methods to study journalistic communication.

Lastly, we were also interested in journals in which CSS studies are commonly published. We find that studies using computational methods are less frequently published within our discipline, for example in communication journals (35.5% CSS sample; 42.7% benchmark sample). However, communication scholars are not the driving force behind this development: Communication scientists are, in fact, more frequently publishing studies in communication journals if these studies employ computational methods (81.5% CSS sample; 70.8% benchmark sample). In contrast, researchers from other disciplines are considerably less like to publish their CSS work in communication journals (16.7% CSS sample; 28.3% benchmark sample).

6 The "computational turn": an "interdisciplinary turn"?

When focusing on a single computational method (automated content analysis) and a single research field (journalism studies): Can the "computational turn" be equated with an "interdisciplinary turn"? Our study shows that computational methods may lead to (a) increased methodological interdisciplinarity in communication science, (b) have not had any discernible effects on theoretical interdisciplinarity, and (c) may not increase but change practical interdisciplinarity by opening the field for collaborations with technical disciplines.

6.1 Methodological interdisciplinarity

That communication scientists use automated content analysis signals an increase in methodological interdisciplinarity: methods that are closely affiliated with computer and information science are gaining in importance in our discipline. Which opportunities and risks does this development bring about for communication science?

One opportunity is that computational methods afford the examination of theories/concepts or variables *from a new perspective* or, to some extent, *in an improved way*. Many studies in the CSS sample analyze processes of journalistic communication, as predicted by Wells et al. (2019): Using computational methods to retrieve timestamps, these studies can analyze theories/concepts such as event and issue agendas (Hase et al. 2021) or news diffusion (Buhl et al. 2018) in greater detail and scope than is possible via manual coding. Moreover, as de Grove et al. (2020) argue, computational methods allow to capture complex, hybrid, and nonlinear news flows—and thus extend existing theories/concepts in communication science, thereby allowing us to better understand the increasing complexity of public spheres (Waldherr 2017). In addition, automated analyses enable us to measure variables that have largely been understudied, for example linguistic complexity (Tolochko and Boomgaarden 2019).

However, methodological interdisciplinarity also carries risks. These include a *trivialization of theories/concepts* and a *lack of methodological standards*. Our review illustrates that studies often analyze variables such as themes or sentiment – likely due to the accessibility of comparably “easy” methods to measure these concepts without the need for much technical or statistical knowledge. However, there is little consensus on how to theoretically conceptualize sentiment (van Atteveldt et al. 2021); the same is true for “topics” often measured via topic modeling (Günther 2021). Since studies often focus on variables that lack a clear theoretical framework or a merely adopted from other disciplines without any reflection about their theoretical value, de Grove et al. (2020) fear that the “computational turn” will result in studies that contribute little to theory-building in communication science. In our review, this is exemplified by studies that cite framing theory to then measure frames via automated means—while it remains unclear whether measured variables correspond to theoretical conceptualizations of frames as proposed by framing theory. Constructs automated methods identify empirically often bear little resemblance to what communication science theoretically understands as frames (Nicholls and Culpepper 2021). Already in 2015, Mahrt warned that the trivialization of theoretical constructs will, hopefully, not be defining big data studies in communication science on the long run. Our literature review shows that her concern seems to at least partially be justified.

Second, Theocharis and Jungherr (2021) point out that CSS is currently failing to establish methodological standards across disciplines. While reporting intercoder reliability is for example considered a standard in communication science (Lacy et al. 2015), there continues to be debate about corresponding criteria for assessing automated content analysis (Baden et al. 2022). In a similar vein, our literature review indicates that not all studies employing computational methods validate their results, similar to what previous studies suggested (Song et al. 2020). One reason for this may be that peer-reviewing has not yet established validation tests as a necessary standard for publishing computational work, often because such tests are not well known. Another reason is that different approaches for validating automated analyses co-exist (Grimmer and Stewart 2013). By analyzing whether studies compared automated codings to manually coded "gold standards", we also included only one, somewhat controversial, type of validity tests in our review (DiMaggio 2015). Moreover, standards related to testing the robustness of results or transparency about operationalizations are, to date, similarly debated or simply nonexistent (Baden et al. 2022; Nelson 2019). However, such standards are increasingly being developed by communication scientists (Haim 2021). Overall, the focus of our discipline on the empirical application of computational methods instead of their methodological development at least raises doubts that communication science will play a dominant role in setting methodological standards within the realm of CSS in the near future.

6.2 Theoretical interdisciplinarity

As a potential chance of the computational turn, Waldherr et al. (2021) also argue that the computational turn may open communication science for theoretical approaches developed by other disciplines, such as complexity theory. Our literature review shows that studies using computational methods draw on theories/concepts with interdisciplinary roots – many of which, however, are also closely linked to communication science outside of its "computational turn" (Steensen and Ahva 2015). Thus, there is *little evidence of an increase in theoretical interdisciplinarity*. Rather, our literature review shows that studies using computational methods are often data-driven and exploratory. In addition, many studies rely on comparatively established middle-range theories or do not theoretically embed empirical studies at all. Both these issues are common for communication science as a field (Walter et al. 2018) but more frequently discussed within the context of CSS (Theocharis and Jungherr 2021; Waldherr et al. 2021). Whether the reliance on comparatively traditional theories of medium range in CSS represents an opportunity or a risk remains to be seen. Moreover, the absence of theoretical interdisciplinarity may merely be due to our specific perspective: If we had focused, for example, on simulation models, we may have found stronger indications of theoretical interdisciplinarity.

6.3 Practical interdisciplinarity

Finally, our literature review shows that communication scholars employing computational methods are by no means more often publishing in interdisciplinary teams or in journals outside our discipline. Thus, we find *little evidence of an increase in practical interdisciplinarity*. Rather, we observe a *shift* in terms of which other disciplines are working on research related to communication science and where studies are published: Especially information, engineering, or technical sciences are increasingly analyzing journalistic communication. Moreover, communication scientists are even more likely to publish studies with computational methods in communication journals instead of in journals outside our discipline, in contrast to researchers from other disciplines. That our discipline invites collaborations with technical disciplines can be seen as an opportunity insofar as it could minimize existing risks—such as a lack of methodological standards for using computational methods. One chance emerging out of such collaborations may be the development of research software in interdisciplinary teams (von Nordheim et al. 2021). On the other hand, it should also be noted that, as Boumans and Trilling (2016) have critically pointed out, methodological advances are currently often published without the involvement of and awareness of communication science, something that should be considered a risk.

6.4 Limitations & outlook

Our study comes with several limitations. This includes our analysis being limited by its focus on journalism research and automated content analysis. Our results should not be generalized: For other research fields, e.g., political communication, and other methods, e.g., network analysis or simulation models, computational methods may be used differently and with different consequences. Moreover, journalism research is often defined much more broadly than is the case here, for instance by including content generated by audiences (Loosen 2016). Finally, interdisciplinarity is a process (Wagner et al. 2011) of interdisciplinary convergence (von Nordheim et al. 2021). When considering its "computational turn," communication science is certainly still only at the beginning of such a process.

In this respect, our article is primarily intended to stimulate discussion about implementing change. To date, debates about opportunities, risks and recommendations associated with computational methods and interdisciplinarity are mainly discussed through "bottom-up" initiatives, e.g., the DGPuK working group on "Computational Communication Science (CCS) in teaching" or in self-organized working groups by young scholars, e.g., the "Computational Methods Working Group" in Zurich. Many of these discussions need to be continued, institutionalized, and appropriate measures need to be implemented at the level of universities and institutes. With respect to methodological and theoretical interdisciplinarity, this is especially true for teaching CSS. While computational methods are increasingly being used across institutes, this is not always the case (Strippel et al. 2018); moreover, the development of corresponding syllabi is comparatively costly for lecturers who often have to advance their own methodological and didactical training which requires additional effort (Boczek and Hase 2020).

As Haim (2021) points out, scientists also need stronger incentives to deal with the uncertainties of interdisciplinary career paths (Theocharis and Jungherr 2021; Windsor 2021)—for example concerning the recognition of publications in journals outside of communication science. Such incentives could be provided by better recognizing interdisciplinary career paths and research collaborations (Uth et al. 2020). According to Haim (2021), there is hope in this regard as communication science has started to establish first professorships with a focus on CCS and CSS. Further incentives may include third-party funding (Haim 2021) or support for research infrastructures (Strippel 2021). Only through these incentives, we can turn the "computational turn" into a theoretical, methodological, and practical "interdisciplinary turn", bringing about not only challenges but also opportunities for our discipline.

References

Baden, Christian, Christian Pipal, Martijn Schoonvelde & Mariken A. C. G. van der Velden. 2022. Three gaps in computational text analysis methods for social sciences: A research agenda. *Communication Methods and Measures* 16(1). 1–18.

Benoit, Ken. 2020. Text as data: An overview. In Luigi Curini & Robert J. Franzese (eds.), *The SAGE Handbook of Research Methods in Political Science and International Relations*, 461–497. Thousand Oaks, USA: SAGE Publications Ltd.

Boczek, Karin & Valerie Hase. 2020. Technische Innovation, theoretische Sackgasse? Chancen und Grenzen der automatisierten Inhaltsanalyse in Lehre und Forschung [Technical Innovation, Theoretical Dead End? Opportunities and Limitations of Automated Content Analysis in Teaching and Research]. In Jonas Schützeneder, Klaus Meier & Nina Springer (eds.), *Neujustierung der Journalistik/Journalismusforschung in der digitalen Gesellschaft: Proceedings zur Jahrestagung der Fachgruppe Journalistik/Journalismusforschung der Deutschen Gesellschaft für Publizistik- und Kommunikationswissenschaft 2019*, 117–128. Eichstätt: DEU.

Boumans, Jelle W. & Damian Trilling. 2016. Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars. *Digital Journalism* 4(1). 8–23.

Buhl, Florian, Elisabeth Günther & Thorsten Quandt. 2018. Observing the dynamics of the online news ecosystem: News diffusion processes among German news sites. *Journalism Studies* 19(1). 79–104.

Burggraaff, Christiaan & Damian Trilling. 2020. Through a different gate: An automated content analysis of how online news and print news differ. *Journalism* 21(1). 112–129.

Capra, Fritjof. 1996. *The web of life: A new scientific understanding of living systems*. New York: Doubleday.

De Grove, Frederik, Kristof Boghe & Lieven De Marez. 2020. (What) can journalism studies learn from supervised machine learning. *Journalism Studies* 21(7). 912–927.

DiMaggio, Paul. 2015. Adapting computational text analysis to social science (and vice versa). *Big Data and Society* 2(2). 205395171560290.

Geise, Stephanie, Patrick Rössler & Simon Kruschinski. 2016. Automatisierte Analyse medialer Bildinhalte. Potenziale, Grenzen, methodischtechnischer Status Quo und zukünftige Herausforderungen–eine Bestandsaufnahme [Automated analysis of media image content. Potentials, limits, methodological-technical status quo and future challenges – a review of the status quo]. *Medien & Kommunikationswissenschaft* 64(2). 244–269.

Grimmer, Justin & Brandon M. Stewart. 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis* 21(3). 267–297.

Günther, Elisabeth. 2021. *Topic modeling: algorithmische Themenkonzepte in Gegenstand und Methodik der Kommunikationswissenschaft [Topic modelling: Algorithmic topic concepts in the object and methodology of communication science].* Köln: Von Halem.

Günther, Elisabeth & Thorsten Quandt. 2016. Word counts and topic models. *Digital Journalism* 4(1). 75–88.

Haim, Mario. 2021. Gütekriterien und Handlungsempfehlungen für die Entwicklung von Forschungssoftware in der Kommunikations-und Medienwissenschaft [Quality criteria and recommendations for the development of research software in communication and media science]. *Medien & Kommunikationswissenschaft* 69(1). 65–79.

Hamborg, Felix, Karsten Donnay & Bela Gipp. 2019. Automated identification of media bias in news articles: An interdisciplinary literature review. *International Journal on Digital Libraries* 20(4). 391–415.

Hanitzsch, Thomas & Sven Engesser. 2014. Journalismusforschung als Integrationsdisziplin [Journalism research as an Integration Discipline]. In Matthias Karmasin, Matthias Rath & Barbara Thomaß (eds.), *Kommunikationswissenschaft als integrationsdisziplin*, 137–157. Wiesbaden: Springer Fachmedien Wiesbaden.

Hase, Valerie, Daniela Mahl, Mike S. Schäfer & Tobias R. Keller. 2021. Climate change in news media across the globe: An automated analysis of issue attention and themes in climate change coverage in 10 countries (2006–2018). *Global Environmental Change* 70. 102353.

Hepp, Andreas, Wiebke Loosen & Uwe Hasebrink. 2021. Jenseits des Computational Turn: Methodenentwicklung und Forschungssoftware in der Kommunikations-und Medienwissenschaft– zur Einführung in das Themenheft [Beyond the computational turn: methodological development and research software in communication and media studies – an introduction to the issue]. *Medien & Kommunikationswissenschaft* 69(1). 3–24.

Jacobs, Jerry A. & Scott Frickel. 2009. Interdisciplinarity: A critical assessment. *Annual Review of Sociology* 35. 43–65.

Jonkman, Jeroen G. F., Mark Boukes, Rens Vliegenthart & Piet Verhoeven. 2020. Buffering negative news: Individual-level effects of company visibility, tone, and pre-existing attitudes on corporate reputation. *Mass Communication and Society* 23(2). 272–296.

Klein, Julie Thompson. 2017. Typologies of interdisciplinarity: The boundary work of definition. In Robert Frodeman (ed.), *The Oxford Handbook of Interdisciplinarity*, 21–34. Oxford: Oxford University Press.

Kroon, Anne C., Damian Trilling & Tamara Raats. 2021. Guilty by association: Using word embeddings to measure ethnic stereotypes in news coverage. *Journalism & Mass Communication Quarterly* 98(2). 451–477.

Lacy, Stephen, Brendan R. Watson, Daniel Riffe & Jennette Lovejoy. 2015. Issues and best practices in content analysis. *Journalism & Mass Communication Quarterly* 92(4). 791–811.

Laugwitz, Laura. 2020. *Qualitätskriterien für die automatische Inhaltsanalyse: Zur Integration von Verfahren des maschinellen Lernens in die Kommunikationswissenschaft [Quality criteria for automated content* *analysis: on the integration of machine learning methods in communication science].* Masterarbeit an der FU Berlin. https://doi.org/10.31235/osf.io/gt28f (accessed 30 December 2021).

Löffelholz, Martin & Liane Rothenberger. 2011. Eclectic continuum, distinct discipline or sub-domain of communication studies? Theoretical considerations and empirical findings on the disciplinarity, multidisciplinarity and transdisciplinarity of journalism studies. *Brazilian Journalism Research* 7(1). 7–29.

Löffelholz, Martin & Liane Rothenberger. 2016. *Handbuch Journalismustheorien [Handbook of Journalism Theories].* Wiesbaden: Springer Fachmedien Wiesbaden.

Loosen, Wiebke. 2016. Publikumsbeteiligung im Journalismus [Public Participation in Journalism]. In Klaus Meier & Christoph Neuberger (eds.), *Journalismusforschung: Stand und Perspektiven*, 287–316. Baden-Baden: Nomos.

Loosen, Wiebke & Armin Scholl. 2012. Theorie und Praxis von Mehrmethodendesigns in der Kommunikationswissenschaft [Theory and Practice of Multi-method Designs in Communication Science]. In Wiebke Loosen & Armin Scholl (eds.), *Methodenkombinationen in der Kommunikationswissenschaft: Methodologische Herausforderungen und empirische Praxis*, 9–25. Köln: Halem Verlag.

Mahrt, Merja. 2015. Mit Big Data gegen das Ende der Theorie? [With Big Data against the “End of Theory”?]. In Herausgegeben von Axel Maireder, Julian Ausserhofer, Christina Schumann & Monika Taddicken (eds.), *Digitale Methoden in der Kommunikationswissenschaft*, 23–37. Berlin: Berlin Verlag.

Mongeon, Philippe & Adèle Paul-Hus. 2016. The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics* 106. 213–228.

Nelson, Laura K. 2019. To measure meaning in big data, don’t give me a map, give me transparency and reproducibility. *Sociological Methodology* 49(1). 139–143.

Nicholls, Tom & Pepper D. Culpepper. 2021. Computational identification of media frames: Strengths, weaknesses, and opportunities. *Political Communication* 38(1–2). 159–181.

Scharkow, Michael. 2012. *Automatische Inhaltsanalyse und Maschinelles Lernen [Automated content analysis and machine learning].* Berlin: Dissertation an der Universität der Künste Berlin. epubli.

Song, Hyunjin, Petro Tolochko, Jakob-Moritz Eberl, Olga Eisele, Esther Greussing, Tobias Heidenreich, Fabienne Lind, Sebastian Galyga & Hajo G. Boomgaarden. 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.

Steensen, Steen & Laura Ahva. 2015. Theories of journalism in a digital age: An exploration and introduction. *Digital Journalism* 3(1). 1–18.

Strippel, Christian. 2021. Forschungsinfrastrukturen für die Kommunikations- und Medienforschung im deutschsprachigen Raum: Initiativen, Bedarfe und Perspektiven [Research infrastructures for communication and media research in the German-speaking world: initiatives, needs and perspectives]. *Medien & Kommunikationswissenschaft* 69(1). 136–157.

Strippel, Christian, Annekatrin Bock, Christian Katzenbach, Merja Mahrt, Lisa Merten, Christian Nuernbergk, Christian Pentzold, Cornelius Puschmann & Annie Waldherr. 2018. Die Zukunft der Kommunikationswissenschaft ist schon da, sie ist nur ungleich verteilt [The future of communication science is already here, it is just unevenly distributed]. *Publizistik* 63(1). 11–27.

Theocharis, Yannis & Andreas Jungherr. 2021. Computational social science and the study of political communication. *Political Communication* 38(1–2). 1–22.

Tolochko, Petro & Hajo G. Boomgaarden. 2019. Determining political text complexity: Conceptualizations, measurements& application. *International Journal of Communication* 13. 1784–1804.

Trilling, Damian & Marieke van Hoof. 2020. Between article and topic: News events as level of analysis and their computational identification. *Digital Journalism* 8(10). 1317–1337.

Uth, Bernadette, Bernd Blöbaum, Laura Badura & Katherine M. Engelke. 2020. Institutionalisierte Interdisziplinarität: Chancen für die Neujustierung der Journalismusforschung in einer digitalisierten Welt [Institutionalised interdisciplinarity: opportunities for the reorientation of journalism research in a digitalised world]. In Jonas Schützeneder, Klaus Meier & Nina Springer (eds.), *Neujustierung der Journalistik/Journalismusforschung in der digitalen Gesellschaft: Proceedings zur Jahrestagung der Fachgruppe Journalistik/Journalismusforschung der Deutschen Gesellschaft für Publizistik- und Kommunikationswissenschaft 2019*, 129–139. Eichstätt: DEU.

Van Atteveldt, Wouter, Mariken A. C. G. Van der Velden & Mark Boukes. 2021. The validity of sentiment analysis: Comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms. *Communication Methods and Measures* 15(2). 121–140.

von Nordheim, Gerret, Lars Koppers, Karin Boczek, Jonas Rieger, Carsten Jentsch, Henrik Müller & Jörg Rahnenführer. 2021. Die Entwicklung von Forschungssoftware als praktische Interdisziplinarität [The Development of Research Software as Practical Interdisciplinarity]. *Medien & Kommunikationswissenschaft* 69(1). 80–96.

Wagner, Caroline S., J. David Roessner, Kamau Bobb, Julie Thompson Klein, Kevin W. Boyack, Joann Keyton, Ismael Rafols & Katy Börner. 2011. Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics* 5(1). 14–26.

Waldherr, Annie. 2017. Öffentlichkeit als komplexes System: Theoretischer Entwurf und methodische Konsequenzen [The Public Sphere as a Complex System: Theoretical Outline and Methodological Consequences]. *Medien & Kommunikationswissenschaft* 65(3). 534–549.

Waldherr, Annie, Stephanie Geise, Merja Mahrt, Christian Katzenbach & Christian Nuernbergk. 2021. Toward a stronger theoretical grounding of computational communication science: How macro frameworks shape our research agendas. *Computational Communication Research* 3(2). 152–179.

Walter, Nathan, Michael J. Cody & Sandra J. Ball-Rokeach. 2018. The ebb and flow of communication research: Seven decades of publication trends and research priorities. *Journal of Communication* 68(2). 424–440.

Wells, Chris, Dhavan V. Shah, Jon C. Pevehouse, Jordan Foley, Josephine Lukito, Ayellet Pelled & Junghwan Yang. 2019. The temporal turn in communication research: Time series analyses using computational approaches. *International Journal of Communication* 13. 4021–4043.

Wettstein, Martin. 2016. *Verfahren zur computerunterstützten Inhaltsanalyse in der Kommunikationswissenschaft [Methods for Computer-aided Content Analysis in Communication Science]*. Dissertation, University of Zurich. https://doi.org/10.5167/uzh-127459 (accessed 20 December 2021).

Williams, Nora Webb, Andreu Casas & John D. Wilkerson. 2020. *Images as data for social science research*. Cambridge: Cambridge University Press.

Windsor, Leah Cathryn. 2021. Advancing interdisciplinary work in computational communication science. *Political Communication* 38(1–2). 182–191.

Zelizer, Barbie. 2017. *What journalism could be*. Cambridge: Polity Press.

Zhu, Yuner & King-Wa Fu. 2019. The relationship between interdisciplinarity and journal impact factor in the field of communication during 1997–2016. *Journal of Communication* 69(3). 273–297.