Computational Communication Science in a Digital Society

Trilling, D. ; Araujo, T. ; Kroon, A. ; Möller, A.M.; Strycharz, J.; Vermeer, S.

DOI
10.1515/9789048560608-016

Publication date
2024

Document Version
Final published version

Published in
Communication Research into the Digital Society

15. Computational Communication Science in a Digital Society

Damian Trilling, Theo Araujo, Anne Kroon, A. Marthe Möller, Joanna Strycharz, and Susan Vermeer

Abstract
Computational methods have added new approaches to the way many communication scientists do their work. We identify four developments that accelerated the adaption of computational methods: the increasing availability of digital data, the surge of large amounts of user-created data, the need to study new artefacts, and the improved accessibility of computational resources. We describe new data acquisition techniques, new research designs, and new analytical approaches that characterise the field. After discussing contributions to the open source community, to the methodological toolbox, as well as to the testing and development of theories, we sketch in broad strokes a research agenda for the coming years.

Keywords: computational communication science, computational social science, digital society, computational methods

Introduction

For a long time, communication scientists have used computers to aid their work. By the 1970s, if not earlier, some communication scientists had built computer models that searched for keywords in texts to conduct automated content analyses of news (Schönbach, 1982). Yet, only recently, such computational approaches became part of the communication research mainstream. For instance, around 15 years ago, Lazer et al. (2009) popularised the term "computational social science." Analogously, nowadays, "computational communication science" can be considered an established term

Trilling, D., Araujo, T., Kroon, A., Möller, A. M., Strycharz, J., & Vermeer, S. (2024). Computational communication science in a digital society. In T. Araujo & P. Neijens (Eds.), *Communication research into the digital society: Fundamental insights from the Amsterdam School of Communication Research* (pp. 247-263). Amsterdam University Press. https://doi.org/10.5117/9789048560592_ch15

(e.g., Hilbert et al., 2019), with its own division (Computational Methods) at the International Communication Association and a dedicated journal (*Computational Communication Research*) (van Atteveldt et al., 2019).

While it may be too early for a clear consensus within the discipline about the exact characteristics and boundaries of computational communication science, for the purpose of this contribution, we follow Hilbert et al., who "define computational communication science as the endeavor to understand human communication by developing and applying digital tools that often involve a high degree of automation in observational, theoretical, and experimental research" (2019, p. 3915). This definition highlights a number of important aspects: First, neither the objective (understanding human communication) nor the general methodological approaches (theoretical, observational, or experimental work) has changed—we are still doing communication science. Second, to do so, we develop and apply new tools, allowing us to work on a different scale than previously possible.

We observe four intertwined developments that lead to the increasing attention to computational approaches. First, the data that communication scientists are interested in are increasingly available in digital formats. News sites have started to replace printed newspaper editions, an analogue telephone call or chat in a pub may now be observable on social media, and also advertising has often gone digital.

Second, whereas media content typically was created by "institutions" such as newspapers, corporate organisations, or governments, nowadays, media users play an important role in the creation of data, impacting their amount, content, and ownership. Traditional techniques may not scale well enough to collect and analyse such data.

Third, new artefacts such as voice assistants, search engines, chatbots, or recommender systems have become relevant for people's media use. Some of the aspects of such artefacts require new methodological approaches to study them. On a higher level, the increased personalisation of the media environment also falls under this development: There is no longer one newspaper, one advertisement, or one television programme that is largely saying the same thing for (in principle) everyone. Instead, individuals receive different information about different topics at different moments in time. Both the collection and the analysis of such personalised data requires—at least to an extent—automation.

Fourth and finally, advances in and accessibility of storage, computing power, but also powerful and more accessible programming frameworks have enabled things that before would have required immense resources in terms of finances and expertise.

Due to these developments, researchers at our department and beyond have turned to computational methods to study communication with increasing frequency. This chapter provides an overview of how this impacts the work of communication scholars and its implications for the field. As the boundaries of what constitutes computational communication science may sometimes be a bit blurry, our list may not be exhaustive. For instance, our examples tend to emphasise content-analytical and digital trace data, while other types of computational work (such as theory-driven computer simulations) have only slowly been taken up in our department so far. Yet, we believe it covers the main developments we observe at this moment in the work done at ASCoR.

Empirical findings

As we have seen, the boundary between traditional and computational communication science is fluid and, sometimes, it is just about automating and/or scaling up one part of the research pipeline. In this section, we discuss three main parts in which computational techniques are employed: data acquisition, research design, and analytical approaches.

New data acquisition techniques

The collection of digitally available data, such as online news, social media content, or other types of digital traces, requires new methods. Three main paths have been used at ASCoR to collect such data: (1) application programming interfaces (APIs); (2) scraping; and (3) tracking and data donation, that is, partnering with users willing to install trackers or to donate their data to academic research.

Application programming interfaces (*APIs*). Some platforms and services allow scholars to gather data via APIs. These services, originally developed to enable third-party clients to interact with a platform in a standardised format, sometimes also allow searching for and downloading specific data subsets. Over the years, many ASCoR researchers have been using this functionality to collect large amounts of social media data, for example, posts or comments from Facebook, Twitter, or YouTube. While application programming interfaces (APIs) provide researchers with significant benefits, such as structured access to extensive data, it is important to note their limitations. Notably, the availability of API access has been increasingly restricted as platforms impose regulations on data accessibility. For example, Twitter introduced several restrictions that first required academics to apply for full access to the data and now created a prohibitively expensive paid tier to the API that also applies to academics. This makes APIs provided by platforms an unsustainable data collection practice that cannot be trusted to be available and necessitates researchers to turn to different ways of collecting data.

Scraping. In cases when no or very limited access through API is provided by a platform, online content has been commonly scraped. This data collection method involves crawling a website with an automated script that explores the website's structure and collects data using HTTP requests and responses. At ASCoR, scraping has, for instance, routinely been used for conducting content analyses of Dutch online news by scraping news websites at a regular interval. To do so, scholars have collaborated and shared the infrastructure for such content analyses (e.g., Trilling et al., 2018). Scraping has also been used to collect social media content not available via APIs, for example, using open source scrapers to collect visual posts from Instagram or TikTok. While independent from proprietary access given by platforms, it presents a challenge in balancing data collection efforts with currently often unclear regulations that govern scraping different websites.

Tracking and data donation. The usage of APIs or scraping to collect digital trace data (often from social media) enables the creation of large-scale datasets of communication data, yet it often is limited to publicly available data (e.g., Instagram posts that are public), subject to restrictions by platforms, and does not allow for the linkage to individual-level information about the causes (e.g., motivations) or consequences (e.g., perceptions, attitudes) of communication. To address these challenges, ASCoR researchers have extended or developed approaches to tracking and data donation, in which researchers partner with participants willing to share their data. Studies have been conducted with data gathered by apps or trackers that individuals install on their devices, enabling, for example, examining the way news consumers navigate online (Möller et al., 2020; Vermeer et al., 2020) or assessing the differences in self-reported and tracked internet or social media use (Araujo et al., 2017; Verbeij et al., 2021). Another approach is for individuals to request their data from online platforms, and donate it to researchers. ASCoR researchers have not only developed an open source framework for data donation (Araujo et al., 2022), but also several research groups within the institute have created guidelines and methodological studies on the advantages, limitations, and the potential of the method (van Driel et al., 2022). Others have relied on donations of local browsing histories from participants' web browsers (Wojcieszak et al., 2022), analysed participants' WhatsApp data (Vermeer et al., 2021), combined data donations with mobile experience sampling, in which participants get multiple short questions on their smartphone within a short time frame (Otto & Kruikemeier, 2023), or asked participants to donate recordings of the iOS screen time function (Ohme et al., 2020).

New research designs

The increasing digitalisation and datafication of the communication landscape not only enables new forms of collecting communication data for research, but also brings forward a new set of artefacts that, in and by themselves, become the object of communication research. To give an example: Communication scientists can use machine learning to analyse media content—but they can also study how media companies employ machine learning and analyse its consequential impact. Accordingly, as search engines and recommender systems become increasingly central in the media landscape, ASCoR researchers devised new research designs to experimentally study the causes, contents, and consequences of individual interactions with these artefacts.

As content that users are exposed to online is increasingly personalised, ASCoR researchers have turned to designing their own prototypes of so-called recommender systems used for such personalisation to be able to closely observe how individuals interact with them and link these interactions to motivations and consequences. Integrating such prototypes in experimental research allows the increase of ecological validity going beyond scenario studies used in the past. As personalisation is currently present in all domains of communication, such research has been conducted on the impact of recommendation on diversity of news consumption (Loecherbach et al., 2021), possibilities for increasing effectiveness of health information (Nguyen et al., 2017), or achieving a healthy lifestyle through a personalised mobile coaching app (Stuber et al., 2020).

In addition, ASCoR researchers have deployed new research designs to study the emergence of conversational agents (e.g., chatbots and virtual assistants) and social robots in the communication environment. For an overview, see Peter et al. (2024) in this book.

New analytical approaches

New types of data often require new analytical approaches. We saw that already quite early, ASCoR researchers started analysing, for instance, word frequencies. For example, pioneered by Loet Leydesdorff, colleagues examined co-occurrence matrices of words to study topics and frames (e.g., Hellsten et al., 2010; Leydesdorff & Welbers, 2011). Such bottom-up analyses were complemented with top-down dictionary approaches, in which colleagues compiled lists of keywords to automatically search for pre-defined topics, frames, and sentiments.

In contrast, contemporary analyses at ASCoR typically make use of machine learning approaches to, for example, identify frames in news articles, categorise their topic, or to estimate the sentiment of social media data. This means that the top-down dictionary approach (in which researchers come up with a list of keywords that are then automatically counted), in many instances has been replaced by supervised machine learning: We hand-code a subset of the data and estimate a model that can predict the coding based on word frequencies. In other words, and at the danger of oversimplifying matters, rather than pretending that we can come up with a perfect algorithm ourselves, we just let the computer "learn" an algorithm (and associated weights) from a sample of the data. Regarding bottom-up approaches, besides classic techniques like cluster analysis or principal component analysis, unsupervised approaches such as so-called topic models have become popular to identify topics and frames that are not known a priori. In short: supervised machine learning uses hand-coded (labelled) data to learn how to automatically code a larger (unlabelled) dataset, while unsupervised machine learning does not use hand-coded data and recognises patterns in the data instead.

While such classic machine learning approaches (both supervised and unsupervised) are still popular, recent work increasingly focuses on techniques that go beyond this classic approach (in which word order is not taken into account, and in which we cannot account for similarities between words, such as synonyms or antonyms). Researchers now regularly work with word embeddings (vector representations of words, in which more similar words are closer to each other in a vector space), which are either studied in themselves (for instance, to detect biases), or made part of a supervised or unsupervised machine-learning workflow. In particular, researchers work with so-called neural networks. In such a network, features (such as word frequencies) do not directly predict the outcome (such as the topic of a text) anymore but predict some intermediate layers first. Once the number of layers grows, such an approach is also called deep learning. Crucially, deep learning makes it possible to move beyond text: We can as well, for example, use pixels of an image as input features, and the different layers than "learn" how, for instance, the furriness of a cat may predict that it is not a dog.

Very recently, though, so-called transformer-based approaches have revolutionised the field. Even lay people have heard about large language models like ChatGPT, which can generate text based on free-form text prompts, or DALL-E, which can generate images based on text prompts. ASCoR researchers not only study such models (for example, to understand the biases they contain), but also employ them in their research. For example, we learned that neural topic models, which do not "start at zero" but can build on existing language models, provide much more coherent topics than the models that have been used so far. And fine-tuning an existing model by providing it with additional information (e.g., annotations of relevant texts) such that it learns a supervised classification task that we are interested in, often (but not always) outperforms classic approaches and/or requires much fewer training examples (so-called few-shot learning). On the other hand, such approaches also require specific computing resources (graphics processing units, or GPUs), accessibility to which provides a barrier to ASCoR researchers at the moment.

Contributions to the field

So far, we have outlined the developments that have led to innovations within the research done at ASCoR. We now look at the next step and focus not so much on what has been *used* in our group, but about what our group has *provided*. These contributions are not only relevant because they advance the development of practical tools that scholars can use, but also because they can be used to advance our theoretical understanding of communication processes and their consequences, or because they can be used to test theories that were not testable before. Next to the contributions to the computational communication science field, which we outline in the next paragraphs, through intense collaboration with the University of Amsterdam's Institute for Information Law (IViR), lines of research outlined in this chapter also translate into specific policy recommendations.

Contributions to the open source community

Many of the new approaches described above have led to the development of new open source tools that can and have been used for follow-up studies. For example, Araujo (2020) developed a toolkit that offers researchers the possibility of integrating a conversational agent into their experiments; Loecherbach and Trilling (2020) developed a tool to create online experiments with news recommender systems; Votta (e.g., 2021) created multiple R packages to make computational tools more easily accessible; and a large group of colleagues developed OSD2F, an open source data donation framework that allows anybody to set up privacy-respecting data donation studies (Araujo et al., 2022). In sum, these contributions do not only allow gaining access to more and more varied data, but have also increased the possibilities for data creation. Specifically, integrating computational methods into, for example, established research designs (such as social-scientific experiments) has opened a range of new opportunities to conduct studies, often in a more ecologically valid way than previously possible.

Validation and consolidation of the toolbox

New methodological approaches beg the question: How valid and reliable are they—and which ones stand the test of time and should be added to our toolbox? For *new research designs*, multiple efforts have been undertaken, for instance, to understand survey measures of media use vis-à-vis digital trace data. For example, this is done by Valkenburg and her team, who compared digital trace data with survey data gathered among social media users (Verbeij et al., 2021), or by another group focusing on internet use (Araujo et al., 2017). For *new analytical approaches*, others have focused on assessing the quality of such automated content analysis techniques—for instance, the work of Boukes et al. (2020) showing the poor performance of off-the-shelf sentiment analysis dictionaries. The estimates provided by such dictionaries turned out to correlate very weakly with each other and with human coders. And, indeed, also thanks to studies such as theirs, the technique is now considered outdated by many. Finally, much time has been invested (not only in research, but also in teaching) to consolidate and canonise the techniques—for instance, for automated content analysis (Boumans & Trilling, 2016)—that culminated in the publication of a textbook (van Atteveldt et al., 2022).

New avenues for developing and testing theories

Using computational techniques to gather, generate, and analyse data allows scholars to advance their theoretical understanding of communication processes and their effects. The rise of the digital society has led some established paradigms within the field to be questioned and using computational methods in their research equips scholars to understand these phenomena better. Some examples may illustrate such relationships between methodological and theoretical innovations. One example is the acceleration of research on artificial entities in our communication environment—when we see the medium as the communication partner (Peter & Kühne, 2018; Zhao, 2006)—with new methods allowing for chatbot experiments (e.g., Araujo, 2020) helping further develop and test such theoretical approaches.

Scraping in combination with advanced large-scale text comparison approaches and network clustering approaches allow the modification of agenda-setting theory such that it can be applied to a fine-grained news event level, rather than to the comparatively coarse issue level that agenda-setting theory was confined to due to feasibility constraints (Trilling & van Hoof, 2020). Similarly, combinations of scraping, social media APIs, network clustering, and different automated content analysis techniques can help us improve theories about information dissemination and opinion leaders, both regarding brand content (e.g., Himelboim et al., 2023) as well as political content (e.g., Simon et al., 2022). Tracking data (e.g., Merten et al., 2022) are also used for that purpose.

Theories about the political role of interpersonal talk—also hard to test in an offline setting—can now be refined by observing political talk in app data (e.g., Vermeer et al., 2021). The use of large language models (LLMs) now makes it possible to reconcile journalistic genre theories with an analysis of extremely varied and multi-modal news-related online content (Lin et al., 2023). Not only are technological developments changing media content consumption, but they also impact individuals' knowledge and experience with various technologies, leading to new theoretical areas to explore. Focusing on advertisements, in particular, Strycharz and Segijn (2022) describe how the constant collection and usage of consumer data for advertising purposes influence consumer perceptions and behaviour. Such examples may offer a glimpse of the theoretical challenges that the digital society brings with it—but also of the potential that computational methods can have when it comes to developing and testing such theoretical innovations.

Next steps

So far, we outlined multiple promising developments in the field of computational communication science, and it is to be expected that work on all of them will continue. Nevertheless, at the risk of omitting important items, this concluding section provides an attempt to highlight some very recent developments that will most likely shape our research profoundly in the next few years.

Expand research across content domains, languages, and modalities

In recent years, the field of natural language processing (NLP) has been revolutionised by the development of LLMs based on deep learning architectures. LLMs have the potential to significantly augment the efficiency and capabilities of computational analysis applied to media texts in various content domains (including political texts and social media posts) and languages (e.g., Lin et al., 2023). Currently, they are reshaping the benchmarks and standards for automated content analysis. In inductive research, LLMs, powered by tools such as BERTopic, can aid in discovering hidden patterns in text and clustering content around topics, actors, or sentiment (e.g., Simon et al., 2022). For deductive research, pre-trained LLMs can be fine-tuned on small-scale annotated custom datasets to improve the accuracy of concept classification, such as frames, topics, sentiment, or incivility, which are central to communication science. The multilingual and multi-modal analysis capabilities of LLMs are particularly promising, and ASCoR researchers are currently exploring these avenues to enable cross-lingual analyses in different content domains, even including visual and textual content in the same models.

Advance understanding of the impact of algorithmic systems

The advancement of computational methods and the development of LLMs will likely add to the importance of content-based recommendations (over other techniques, such as collaborative filtering)—making such models especially interesting from a research perspective. Consequently, further research on this topic at ASCoR is crucial. Despite the significant performance gains that LLMs bring to the field of computational communication science, there are also concerns about their potential negative influence on research and society at large. A first, salient concern is that LLMs may resonate, reflect, and reinforce existing (historical) biases and inequalities in society. These models are typically trained on large amounts of textual data derived from a broad spectrum of online sources, generated by humans. These data are by no means neutral but reflect human preferences and biases. These biases in the training data may subsequently become embedded in the inner workings of LLMs and produce predictions that are biased in implicit ways. While such biases may be difficult to detect, they can affect society at scale. In addition, there are concerns about the lack of diversity among developers, which may lead to algorithmic design choices that do not adequately address the concerns and input of underrepresented groups. Consequently, biases may become "engineered" in LLMs due to algorithmic design choices and biased training data. This may have consequences for the classification performance of fine-tuned models but may also prompt biases responses in machine-generated communication, such as that created by chatbots. Given the societal importance of these issues, ASCoR researchers have an interest in studying them. This work will build upon previous research on bias in language models conducted by the group's researchers (e.g., Kroon et al., 2021), and focus more in depth on bias in LLMs and consequences for downstream classification. Another significant concern about LLMs relates to their contribution to the so-called authenticity crisis of communication, as it becomes progressively more difficult to distinguish real or authentic communication from fake or unauthentic. There are significant concerns about the ability of LLMs to be (mis)used to generate fake news or ads, spread disinformation, and impersonate real or even fake individuals. This concern has already inspired research by members of ASCoR, for example, on the implications of so-called deepfakes (Dobber et al., 2021). Future research by members of our group will be inspired by these themes.

Computational methods as facilitators of human–machine communication

With the development of methods such as LLMs, computers' ability to "understand" and generate natural language makes them ideal for use in human–machine communication, particularly for conversational agents. As machine-generated communication becomes increasingly difficult to distinguish from human communication, LLMs can inspire and support research at ASCoR into the consequences of conversational agent feature designs and communication styles. These advances can create interesting theoretical questions that draw on state-of-the-art computational techniques. In addition, LLMs are valuable in researching recommender systems, as they can cluster articles that discuss similar topics and/or express similar sentiments towards specific political attitudes, topics, and actors. This clustering enhances recommender systems' ability to personalise content for individuals' prior political attitudes and beliefs, with significant consequences for democratic functioning (e.g., Dobber et al., 2021).

Exposure to and dissemination of information in the digital society

A central theme in communication science is how information spreads in societies. But while in earlier days, for example, agenda-setting studies could suffice with studying relatively coarse issues in a handful of outlets—but in an environment in which not only traditional media, but also social media, alternative media, and fringe platforms outside of the mainstream (such as Telegram) play a role, we need new approaches—both theoretically and methodologically. A lot of work ahead in this area, in particular, devising methodological approaches that combine multiple techniques: APIs, scraping and data donations for data collection, machine learning based on LLMs in combination with network analysis for the identification of topics, themes, and events, and—additionally—simulation-based approaches to model complex non-linear relationships.

These new data acquisition techniques, new research designs, and new analytical approaches open several promising avenues for communication research in the digital society. Advancing our understanding of (the performance of) computational techniques and using this understanding to further develop these techniques are essential steps in providing scholars with more possibilities to analyse the fast quantity of various communication data generated by individuals, groups, and organisations in the digital society. Given the complexity and fast development of these techniques, ASCoR established the Digital Communication Methods Lab (digicomlab), in 2018.[1] It is a place where ASCoR scholars interested in and working with digital methods come together to address these challenges, and to collaborate with the communication science community more broadly.

The road ahead is still long and, as we have shown, not without risk. But ultimately, to understand communication in the digital society, we will need to push ahead and embrace innovative tools that will allow us to study what we need to study. We are working on it.

References

Araujo, T. (2020). Conversational Agent Research Toolkit: An alternative for creating and managing chatbots for experimental research. *Computational Communication Research*, *2*(1), 35–51. https://doi.org/10.5117/CCR2020.1.002.ARAU

Araujo, T., Ausloos, J., van Atteveldt, W., Loecherbach, F., Moeller, J., Ohme, J., Trilling, D., van de Velde, B., de Vreese, C., & Welbers, K. (2022). OSD2F: An open-source data donation framework. *Computational Communication Research*, *4*(2), 372–387. https://doi.org/10.5117/CCR2022.2.001.ARAU

1 https://www.digicomlab.eu.

Araujo, T., Wonneberger, A., Neijens, P., & de Vreese, C. (2017). How much time do you spend online? Understanding and improving the accuracy of self-reported measures of internet use. *Communication Methods and Measures*, *11*(3), 173–190. https://doi.org/10.1080/19312458.2017.1317337

Boukes, M., van de Velde, B., Araujo, T., & Vliegenthart, R. (2020). What's the tone? Easy doesn't do it: Analyzing performance and agreement between off-the-shelf sentiment analysis tools. *Communication Methods and Measures*, *14*(2), 83–104. https://doi.org/10.1080/19312458.2019.1671966

Boumans, J. W., & Trilling, D. (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. https://doi.org/10.1080/21670811.2015.1096598

Dobber, T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. (2021). Do (micro-targeted) deepfakes have real effects on political attitudes? *The International Journal of Press/Politics*, *26*(1), 69–91. https://doi.org/10.1177/1940161220944364

Hellsten, I., Dawson, J., & Leydesdorff, L. (2010). Implicit media frames: Automated analysis of public debate on artificial sweeteners. *Public Understanding of Science*, *19*(5), 590–608. https://doi.org/10.1177/0963662509343136

Hilbert, M., Barnett, G., Blumenstock, J., Contractor, N., Diesner, J., Frey, S., González-Bailón, S., Lamberso, P., Pan, J., Peng, T.-Q., Shen, C., Smaldino, P. E., van Atteveldt, W., Waldherr, A., Zhang, J., & Zhu, J. J. H. (2019). Computational communication science: A methodological catalyzer for a maturing discipline. *International Journal of Communication*, *13*, 3912–3934. https://ijoc.org/index.php/ijoc/article/view/10675

Himelboim, I., Maslowska, E., & Araujo, T. (2023). Integrating network clustering analysis and computational methods to understand communication with and about brands: Opportunities and challenges. *Journal of Advertising*. Advance online publication. https://doi.org/10.1080/00913367.2023.2166629

Kroon, A. C., Trilling, D., & Raats, T. (2021). Guilty by association: Using word embeddings to measure ethnic stereotypes in news coverage. *Journalism & Mass Communication Quarterly*, *98*(2), 451–477. https://doi.org/10.1177/1077699020932304

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & van Alstyne, M. (2009). Computational social science. *Science*, *323*(5915), 721–723. https://doi.org/10.1126/science.1167742

Leydesdorff, L., & Welbers, K. (2011). The semantic mapping of words and co-words in contexts. *Journal of Infometrics*, *5*(3), 469–475. https://doi.org/10.1016/j.joi.2011.01.008

Lin, Z., Welbers, K., Vermeer, S., & Trilling, D. (2023). Beyond discrete genres: Mapping news items onto a multidimensional framework of genre cues. *Proceedings* *of the 17th International AAAI Conference on the Web and Social Media*, *Cyprus*, *17*(1), 542–553. https://doi.org/10.1609/icwsm.v17i1.22167

Loecherbach, F., & Trilling, D. (2020). 3bij3—Developing a framework for researching recommender systems and their effects. *Computational Communication Research*, *2*(1), 53–79. https://doi.org/10.5117/CCR2020.1.003.LOEC

Loecherbach, F., Welbers, K., Moeller, J., Trilling, D., & van Atteveldt, W. (2021). Is this a click towards diversity? Explaining when and why news users make diverse choices. *Proceedings of the 13th ACM Web Science Conference*, 282–290. https://doi.org/10.1145/3447535.3462506

Merten, L., Metoui, N., Makhortykh, M., Trilling, D., & Moeller, J. (2022). News won't find me? Exploring inequalities in social media news use with tracking data. *International Journal of Communication*, *16*, 1127–1147. https://ijoc.org/index.php/ijoc/article/view/17068

Möller, J., van de Velde, R. N., Merten, L., & Puschmann, C. (2020). Explaining online news engagement based on browsing behavior: Creatures of habit? *Social Science Computer Review*, *38*(5), 616–632. https://doi.org/10.1177/0894439319828012

Nguyen, M. H., van Weert, J. C., Bol, N., Loos, E. F., Tytgat, K. M., van de Ven, A. W., & Smets, E. M. (2017). Tailoring the mode of information presentation: Effects on younger and older adults' attention and recall of online information. *Human Communication Research*, *43*(1), 102–126. https://doi.org/10.1111/hcre.12097

Ohme, J., Araujo, T., de Vreese, C. H., & Piotrowski, J. T. (2020). Mobile data donations: Assessing self-report accuracy and sample biases with the iOS Screen Time function. *Mobile Media & Communication*, *9*(2), 293–313. https://doi.org/10.1177/2050157920959106

Otto, L. P., & Kruikemeier, S. (2023). The smartphone as a tool for mobile communication research: Assessing mobile campaign perceptions and effects with experience sampling. *New Media & Society*, *25*(4), 795–815. https://doi.org/10.1177/14614448231158651

Peter, J., Araujo, T., Ischen, C., Shaikh, S. J., van der Goot, M. J., & van Straten, C. L. (2024). Human–machine communication. In T. Araujo & P. Neijens (Eds.), *Communication research into the digital society: Fundamental insights from the Amsterdam School of Communication Research* (pp. 205–220). Amsterdam University Press. https://doi.org/10.5117/9789048560592_ch12

Peter, J., & Kühne, R. (2018). The new frontier in communication research: Why we should study social robots. *Media and Communication*, *6*(3), 73–76. https://doi.org/10.17645/mac.v6i3.1596

Schönbach, K. (1982). "The issues of the seventies": Elektronische inhaltsanalyse und die langfristige beobachtung von agenda-setting-wirkungen der massenmedien. *Publizistik*, *27*(1–2), 129–140.

Simon, M., Welbers, K., Kroon, A. C., & Trilling, D. (2022). Linked in the dark: A network approach to understanding information flows within the Dutch Telegramsphere. *Information, Communication & Society*. Advance online publication. https://doi.org/10.1080/1369118X.2022.2133549

Strycharz, J., & Segijn, C. M. (2022). The future of dataveillance in advertising theory and practice. *Journal of Advertising*, *51*(5), 574–591. https://doi.org/10.1080/00913367.2022.2109781

Stuber, J. M., Mackenbach, J. D., de Boer, F. E., de Bruijn, G-J., Gillebaart, M., Harbers, M. C., Hoenink, J. C., Klein, M. C. A., Middel, C. N. H., van der Schouw, Y. T., Schuitmaker-Warnaar, T. J., Velema, E., Vos, A. L., Waterlander, W. E., Lakerveld, J., & Beulens, J. W. J. (2020). Reducing cardiometabolic risk in adults with a low socioeconomic position: Protocol of the Supreme Nudge parallel cluster-randomised controlled supermarket trial. *Nutrition Journal*, *19*, Article 46. https://doi.org/10.1186/s12937-020-00562-8

Trilling, D., van de Velde, B., Kroon, A. C., Löcherbach, F., Araujo, T., Strycharz, J., Raats, T., de Klerk, L., & Jonkman, J. G. F. (2018). INCA: Infrastructure for content analysis. *Proceedings of the 2018 IEEE 14th International Conference on e-Science* (*e-Science*), *Netherlands*, 329–330. https://doi.org/10.1109/eScience.2018.00078

Trilling, D., & van Hoof, M. (2020). Between article and topic: News events as level of analysis and their computational identification. *Digital Journalism*, *8*(10), 1317–1337. https://doi.org/10.1080/21670811.2020.1839352

van Atteveldt, W., Margolin, D., Shen, C., Trilling, D., & Weber, R. (2019). A roadmap for computational communication research. *Computational Communication Research*, *1*(1), 1–11. https://doi.org/10.5117/CCR2019.1.001.VANA

van Atteveldt, W., Trilling, D., & Arcíla-Calderón, C. (2022). *Computational analysis of communication: A practical introduction to the analysis of texts, networks, and images with code examples in Python and R*. Wiley Blackwell.

van Driel, I. I., Giachanou, A., Pouwels, J. L., Boeschoten, L., Beyens, I., & Valkenburg, P. M. (2022). Promises and pitfalls of social media data donations. *Communication Methods and Measures*, *16*(4), 266–282. https://doi.org/10.1080/19312458.2022.2109608

Verbeij, T., Pouwels, J. L., Beyens, I., & Valkenburg, P. M. (2021). The accuracy and validity of self-reported social media use measures among adolescents. *Computers in Human Behavior Reports*, *3*, Article 100090. https://doi.org/10.1016/j.chbr.2021.100090

Vermeer, S. A. M., Kruikemeier, S., Trilling, D., & de Vreese, C. H. (2021). WhatsApp with politics?! Examining the effects of interpersonal political discussion in instant messaging apps. *The International Journal of Press/Politics*, *26*(2), 410–437. https://doi.org/10.1177/1940161220925020

Vermeer, S., Trilling, D., Kruikemeier, S., & de Vreese, C. (2020). Online news user journeys: The role of social media, news websites, and topics. *Digital Journalism*, *8*(9), 1114–1141. https://doi.org/10.1080/21670811.2020.1767509

Votta, F. (2021, July 13). *Favstats/peRspective* [Original-date: 2019-05-06T21:53:53Z]. GitHub. https://github.com/favstats/peRspective

Wojcieszak, M., von Hohenberg, B. C., Casas, A., Menchen-Trevino, E., de Leeuw, S., Gonçalves, A., & Boon, M. (2022). Null effects of news exposure: A test of the (un)desirable effects of a "news vacation" and "news binging." *Humanities and Social Sciences Communications*, *9*(1), Article 413. https://doi.org/10.1057/s41599-022-01423-x

Zhao, S. (2006). Humanoid social robots as a medium of communication. *New Media & Society*, *8*(3), 401–419. https://doi.org/10.1177/1461444806061951

About the authors

Damian Trilling is a Full Professor and holds the Chair in Journalism Studies at Vrije Universiteit Amsterdam. Previously, he was an Associate Professor and member of the Political Communication & Journalism group at ASCoR. He is interested in the dissemination and consumption of news and societally relevant information in the current digital media ecosystem and in the development of methods to study them. E-mail: D.C.Trilling@vu.nl

Theo Araujo is a Full Professor of Media, Organisations and Society, and Scientific Director of ASCoR as of 2022. His research investigates the dynamic interplay between media and organisations, and what it means for society, with a special focus on trust and technology. His research has also a methodological focus on computational communication science and the implementation of large-scale data collection and analysis for communication research. E-mail: T.B.Araujo@uva.nl

Anne Kroon is an Associate Professor of Corporate Communication at ASCoR. She studies bias in digital media using computational techniques and experiments. Her research examines how social groups are portrayed in digital media and the resulting effects on prejudice and discrimination. E-mail: A.C.Kroon@uva.nl

A. Marthe Möller is an Assistant Professor of Youth & Media Entertainment at ASCoR. In her work, she analyses user-generated information in the form of comments, ratings, and reviews to study social media users' responses to and experiences of entertainment media messages. E-mail: A.M.Moller@uva.nl

Joanna Strycharz is an Assistant Professor of Persuasion & New Media Technologies at ASCoR. Combining traditional and computational methods, she studies algorithmic persuasion, its impact on consumers, their privacy, and consumer empowerment. E-mail: J.Strycharz@uva.nl

Susan Vermeer was an Assistant Professor at ASCoR and currently holds the same title at Wageningen University & Research. By combining traditional research methods with innovative and computational methods, she aims to understand the impact of the digital media environment, such as social media and algorithms, on political behaviour and attitudes. E-mail: Susan1.Vermeer@wur.nl.