

Volume 6 \ Issue 2 \ 09-11-2026

DOI 10.34669/WI.WJDS/6.2.6 \ ISSN 2748-5625

Licensed under Creative Commons Attribution 4.0 (CC BY-SA 4.0)

VOICES FOR THE NETWORKED SOCIETY

AI-Enhanced Journalism

How to Strengthen Its Role as an Epistemic and Evaluative
Intermediary in Democracy

Christoph Neuberger

Weizenbaum Institute, Berlin, Germany

christoph.neuberger@weizenbaum-institut.de

ABSTRACT

Journalism has two main functions in democracy: news coverage and the formation of public opinion. This essay develops a theoretical framework for the role of journalism as an epistemic and evaluative intermediary. This framework can be used to discuss the two-part question: how can AI improve the epistemic process of news production, and how can it strengthen the evaluative process of public opinion formation? Veristic social epistemology emphasizes the role of organized skepticism in the process of truth-seeking. One crucial weakness of AI is that it does not meet the necessary condition for collective scrutiny, namely, transparency regarding the process through which it generates truth claims. Consequently, AI can only support the human search for truth. Only a few studies have measured the epistemic quality of AI-generated content and its effects directly and systematically. With regard to public opinion formation, there is a theoretical gap that can be addressed using normative theories of democracy. Values can be derived from these theories, which can serve as yardsticks for democratic opinion formation. Empirical studies also paint a mixed picture. The risks include radicalization and confirmation bias. However, generative AI can also improve the quality of deliberation and enhance the communication skills of disadvantaged groups. The rapid development of AI models requires continuous monitoring and an experimental approach to understand the potential uses and limitations of AI in journalism and the democratic public sphere.

KEYWORDS

artificial intelligence \ journalism \ public opinion formation \ news coverage

1 News and Commentary as Journalistic Functions

Journalism has two main functions in democracy: news and commentary, that is, the description of current affairs and its evaluation, respectively. As *The Guardian* editor C. P. Scott's (1921/2017) famous statement goes, "Comment is free, but facts are sacred." In other words, news should be objective, and there can be only one truth. Opinions can be diverse, but they must be well reasoned. Both are essential for liberal democracy to function (Christians et al., 2009).

Consequently, the question of how AI is changing journalism must be answered from two perspectives: how it is changing the *epistemic process* of news production and how it is changing the *evaluative process* of public opinion formation. In the theory of communicative action, *assertions* are constative speech acts that claim "epistemic truth," while *evaluations* are regulative speech acts that claim "normative rightness" (Habermas, 1981, pp. 435–436). There are established procedures and practices for the genesis of these speech acts and the critical examination of their validity claims in public discourse. To be reviewable, they must be justified by arguments. Accordingly, knowledge is classically defined as justified true beliefs, whereas opinions must be supported by arguments.

As an *epistemic* and *evaluative intermediary* (Bartsch et al., 2025; Neuberger, 2022a; 2025), journalism not only produces news and commentaries but also scrutinizes and disseminates knowledge and opinions from external sources, including actors from different societal sub-systems such as politics, economy, or science. It also considers everyday knowledge and common sense. Therefore, as an intermediary, journalism translates between different forms of knowledge as "knowledge broker" (Reich & Lahav, 2021) and conveys the pluralism of values in modern society as moderator of the discourse. In doing so, it contributes to social cohesion (Deitelhoff & Schmelzle, 2023). Since the breakthrough of generative AI, this raises a crucial question: Despite its obvious weaknesses, can AI be used to strengthen journalism's role as an epistemic and evaluative intermediary? Addressing this question necessitates not only empirical research but also theoretical clarifications that are frequently overlooked in the discussion.

2 Neither Intelligent nor Artificial

First, however, the meaning of "artificial intelligence" must be clarified. Sociologist Elena Esposito (2022) proposes a fundamental shift in perspective. According to her, "artificial intelligence" is the wrong term for machine-learning algorithms. These algorithms simulate *communication*, not intelligence. Furthermore, their results are not purely artificially generated but based on natural (human) communication: "Machine learning algorithms that use big data ... are artificially reproducing not intelligence but communication skills, and they do so by parasitically exploiting the participation of users on the web" (Esposito, 2022, pp. 2–3; see also Gunkel, 2012, pp. 2–10).

For communication to be informative, it must be new. According to Niklas Luhmann (1996, p. 46), the special function of mass media is the generation and processing of irritation (see also Schoenbach, 2007). However, this element of surprise must be relevant and appropriate to human observers who understand the synthetic outputs as communicative contributions and make sense of them. This was the case with Joseph Weizenbaum's chatbot ELIZA, whose responses as a virtual therapist were rated as relevant and helpful by its human interaction partners (Esposito, 2024, pp. 33–35). Therefore, “What algorithms are reproducing is not the intelligence of people but the informativity of communication” (Esposito, 2022, p. 18). Unlike search engine algorithms, which only provide access to human-generated content, large language models (LLMs) can generate communication content (Esposito, 2026, pp. 87–89). They learn to communicate appropriately based on their training data, which consists of large volumes of human examples.

The opacity of AI's internal processes, its “black box” nature, is no different from that of human communication. Humans usually lack transparency about the cognitive processes of their communication partners (and even their own), simply because we cannot read each other's minds (Koskinen, 2024, p. 460). However, this is not considered a disadvantage: “On the contrary, this lack of transparency allows for otherwise unthinkable degrees of freedom and abstraction” (Esposito, 2022, p. XI). Nonetheless, a significant difference remains in the transparency of internal cognitive and computational processes. The results of cognitive processes are evaluated as communicated claims in a collective process that takes place publicly and follows shared practices to uncover errors and biases. In such discourses, individual perspectives are presented and supported by arguments that are either confirmed or rejected (Koskinen, 2024, pp. 460–461, 471–472).

Conversely, AI does not provide such disclosure or justify its results. Instead, it produces a “computational epistemology,” where “truth becomes a probabilistic function rather than a deliberative outcome. The result is a new epistemic logic—one that prioritizes predictive performance over interpretive depth, and algorithmic consistency over reflective justification” (Shin, 2025, p. 1554). In the case of generative AI, the public sphere, as a space in which claims are validated through a deliberative process of rational and critical clarification (Habermas, 2006, p. 413), faces a fundamental challenge because the generation of AI output is largely opaque and, therefore, cannot be justified through argumentation.

Viewing AI as communication fundamentally shifts the focus of analysis. The question of whether intelligence, consciousness, and the human brain are replicable (Somers, 2025) is no longer of high relevance. Instead, the question is whether AI can communicate appropriately. There are many different types of communication (Neuberger, 2009, pp. 22–26): one-way and two-way, single-channel and multi-channel (e.g., text, image, audio), one-to-one (i.e., individual communication), one-to-many (i.e., mass communication), and many-to-many communication (communities). Further distinctions can be drawn based on modes of interaction (e.g., conflict, competition, cooperation, diffusion and scandal; Neuberger, 2022b), degrees of accessibility (i.e., private or public), and sub-systemic contexts (e.g., politics, economy, art, and sports). Communication participants can be categorized based on their roles (source, audience, intermediary) and goals (e.g., to inform, persuade, or entertain).

Analyzing general-purpose AI like ChatGPT (OpenAI) and Gemini (Google) requires a theory of the whole society to capture the full range of their potential uses across different contexts and constellations of communication. Therefore, AI should be viewed as a groundbreaking sociotechnological institution expected to transform society in the same way as earlier institutions of information-processing and decision-making did such as the printing press, democracy, bureaucracy, and market (Farrell et al., 2025; Harari, 2024).

Journalism provides a particularly revealing case because it mediates public communication across society as a whole (Neuberger, 2022a; 2025). With the digital transformation, journalism is no longer limited to one-way, one-to-many communication, as is the case with traditional mass media; it now also involves interactive (two-way) communication with varying numbers of participants. The remainder of this essay investigates the following question: can AI improve news coverage and the formation of public opinion in journalism?

3 Can AI Improve News Coverage?

The topic of objectivity in journalism is widely debated (Neuberger, 2017): is journalism able to report the truth about current events? In epistemology, a veristic position takes a middle ground between extreme skepticism (relativism, constructivism), which denies the possibility of access to reality, and naïve (everyday) realism, which is certain of the attainability of truth. Veristic social epistemology does not assume that absolute certainty about knowledge is attainable and instead emphasizes the role of organized skepticism and collective scrutiny. This weaker understanding of truth as true belief (Goldman, 2019, p. 5) requires a justification with “good” reasons and accepts evidence, at least provisionally (Boghossian, 2006, pp. 10–16; Lynch, 2012, pp. 1–6). Consequently, knowledge claims must be justified and are always open to scrutiny. Justification relies on the proper handling of the epistemic process based on shared practices.

In journalism, epistemic practices (e.g., Kovach & Rosenstiel, 2007) can be systematized according to the phases of the news process: investigation, generation, verification, dissemination, adaptation, discussion, and the correction of mistakes (Neuberger, 2017, pp. 418–419; Neuberger et al., 2023, p. 186). Investigation should use reliable sources and a diversity of perspectives. Factual claims in the news should be verified before publication through cross-checking with a second independent source (Godler & Reich, 2017). News outlets should be transparent about their sources and the methods they apply to conduct investigations and verifications. Journalists should inform their audience about what they know for certain, what they do not yet know, and what they are unsure about. However, the news process does not end with publication. Journalists should also facilitate the critical review of their claims by their audience and correct them retroactively if necessary. Journalists are also increasingly expected to fact-check what others have published and debunk misinformation.

Does generative AI meet these requirements of a transparent and critical epistemic discourse? AI's epistemic weaknesses are many, from unintended (hallucinations and logical fallacies; Cau et al., 2025) to intended distortions of truth (deepfakes, use for disinformation, and fake-bots; von Sikorski & Hameleers, 2025; Simon et al., 2023; on the possible misuses of AI, see Marchal et al., 2024). According to Coeckelbergh (2025, pp. 4–10), these could have far-reaching implications for society: AI-generated misinformation creates a climate of epistemic uncertainty, fosters polarization and epistemic bubbles, and enhances the prominence of ignorance and relativism as epistemic attitudes. In addition, in the long run, it leads to epistemic incest, “a situation in which over time the LLM only interacts with its own input data and output texts” (Coeckelbergh, 2025, p. 7). Using synthetic data rather than authentic data for machine learning, a process known as “AI autophagy” (Xing et al., 2025), can result in model collapse (see also Crawford, 2025, p. 3; Gillman et al., 2024; Wachter et al., 2024, pp. 12–13).

In modern cultures referring to an open future, instead, this circularity results in feedback loops and a serious inability to learn. Algorithms see the reality that results from their intervention and do not learn from what they cannot see because it has been canceled by the consequences of their work. The use of algorithms produces a second-order blindness. (Esposito, 2022, p. 100)

This conservatism and loss of touch with reality impedes democratic learning in public discourse, which requires updated knowledge and fresh ideas (Coeckelbergh, 2025, p. 9), especially in the case of journalism, which reports on current events. In addition to misleading content, Simon (2026) identifies three other types of deception associated with generative AI: deception regarding the ontological status of the communication partner (i.e., human or machine), the capacities of AI (attributing human characteristics such as intelligence and empathy to AI), and the function of AI (claiming that it retrieves existing texts instead of generating new ones).

Despite these epistemological pitfalls, generative AI takes the appearance of the perfect truth-telling machine: “AI falsely gives us the illusion of epistemic certainty and completeness, leaving out the ambiguities and knowledge gaps that are inherent in the political life” (Coeckelbergh, 2025, p. 10). Synthetic messages are “presented with an aura of objectivity, yet these outputs are deeply shaped by design decisions, training datasets, annotation protocols, and embedded assumptions” (Shin, 2025, p. 1557).

The main weakness of generative AI is that it does not meet the necessary condition for its truthfulness to be examined discursively, namely, transparency regarding the process by which it generates truth claims. In other words, it is opaque and lacks explainability and assessability (Jungheer & Schroeder, 2023, p. 4). As a result, AI can only serve as a support for human decision-making. It can develop alternative hypotheses as “epistemic nudges” (Hirmiz, 2024, p. 12), for example, in medical diagnosis and treatment, but these require final human examination. Accordingly, codes of ethics in journalism emphasize human oversight of AI use and transparency towards the audience (Becker et al., 2025, pp. 1588–1589). Due to the opacity of AI, journalism is primarily viewed within the professional metadiscourse as an “AI-assisted intermediary” (Perreault & Ohme, 2025, p. 1916). The results of newsroom studies confirm the cautious adop-

tion of AI in journalism (e.g., Ashuri et al., 2026). Xu et al. (2026) are going a step further and suggest a paradigm shift of AI models toward “scaffolded cognitive friction” to preserve human epistemic sovereignty and critical thinking. In the same vein, Harari (2024, p. 300) calls “to train computers to be aware of their own fallibility” and to “learn to doubt themselves”.

The crucial question now is how generative AI can support epistemic work and “contribute to the strengthening of professional news and information curation” (Jungherr, 2023, p. 6). Empirical analysis illustrates the benefits and pitfalls of using AI in news production processes (see reviews by Chalikiopoulou et al., 2025; Opdahl et al., 2023). Many empirical studies are based on journalists’ reports of their experiences with AI (e.g., Cools & Diakopoulos, 2024) or on tool testing in collaboration with researchers (e.g., Fridman et al., 2025; Gutiérrez-Caneda et al., 2023). These studies demonstrate the variety of epistemically significant applications of generative AI in journalism, for research, fact-checking (for an overview on fake news detection techniques, see Emil & Remus, 2025), data analysis, and news writing. Other studies assess the quality of AI-generated news from the user’s perspective (e.g., Graefe & Bohlken, 2020; Nah et al., 2024) or examine audience trust and other attitudes (e.g., Ross Arguedas, 2024).

Nevertheless, only a few studies have measured the epistemic quality of AI or its effects directly and systematically. Kuznetsova et al. (2025) showed that generative AI (ChatGPT, Bing Chat) can correctly verify political information on five topics in most cases. ChatGPT was able to identify true, false, and borderline statements in 72% of cases, while Bing Chat achieved a 67% rate. The transparency of sources is essential for verifying truth claims. In this regard, Petroni et al. (2023) developed and evaluated a neural network-based system for Wikipedia that suggests more suitable sources of evidence for Wikipedia articles. Furthermore, studies exploring knowledge effects of generative AI are rare. A meta-analysis of experimental studies shows that AI-generated texts are increasingly perceived as credible and accurate, whereas visual representations (deepfakes) tend to be viewed with greater skepticism over time (Herasimenka et al., 2026). Costello et al. (2025) conducted experiments to test the (de)bunking effect of LLMs on conspiracy beliefs and achieved mixed results: LLMs can both increase and decrease conspiracy beliefs. The bunking condition was even more informative, collaborative, and trustworthy than the debunking condition (Costello et al., 2025, p. 6). However, one result gives hope:

A simple intervention (telling the AI to only use accurate and truthful information) substantially diminished the AI’s persuasive abilities for bunking, largely through increasing non-compliance where the AI debunked even though it was asked to bunk. Conversely, the truth constraint did not diminish the efficacy of debunking. (Costello et al., 2025, p. 12)

To sum up, based on the current state of research, it is still difficult to say how generative AI can effectively improve news coverage and epistemic discourses on contentious topics like climate change (Schäfer et al., 2026). The rapid pace at which new AI models are developed makes optimizing their use in journalism and predicting their consequences even more difficult. The long-term structural effects of AI must be taken into account, such as its impact on the established knowledge order and epistemic authority (Bartsch et al., 2025; Neuberger et al., 2023), and the possible loss of collective memory (Wachter et al., 2024, p. 11).

4 Can AI Improve Public Opinion Formation?

Journalism plays a dual role in public opinion formation. On the one hand, it acts as a moderator of public discourse, organizing and guiding interactions between discussants. On the other hand, it participates in this discourse as a commentator. *Moderating* public discourse requires neutrality towards individual opinions. Thus, the media should present the diversity of opinions on a given issue in an unbiased and balanced manner. Additionally, journalists should moderate the public discourse based on deliberative criteria such as inclusiveness, responsiveness, justification, and civility (Wessler, 2018, pp. 86–88). At the same time, journalists are expected to express their own opinions. In doing so, journalistic *commentators* should carefully consider arguments from all sides and explain why one opinion is supported by stronger arguments than others.

Does generative AI meet these democratic requirements for public opinion formation? This is an important question given the increasing use of AI in election campaigns (e.g., Jungherr et al., 2026). Campolo and Schwerzmann (2023) diagnose an “artificial naturalism” in machine learning because norms are not formalized and prescribed as rules but emerge implicitly from examples in training data:

In processing these examples, machine learning models produce representations, which express regularities found in the data and seem to naturally emerge from it. These representations then become normative in a more traditional sense when they influence our behavior. The “ought” of examples turns into the “is” of the representations, and, in a further twist, these representations become normative (“ought”) again when the algorithmic outputs are used to make and legitimate predictions and classifications that regulate our conduct. (Campolo & Schwerzmann, 2023, p. 2)

This implicit, individualized, and dynamic normation of human action by AI undermines the “liberal, rule-based form of political subjectivity” and “the common ground necessary for collective agency and struggle” (Campolo & Schwerzmann, 2023, p. 11). Similarly, Miragoli (2025, p. 527) criticizes the conformism of machine learning “as instantiating a tendency to value or endorse attitudes and behaviours that are commonly accepted simply because they are commonly accepted”. In doing so, they “lean toward averages” (Jungherr 2023, p. 5), limiting the diversity of opinions and ignoring minority viewpoints. As they rely on historical data, these systems are also “inherently conservative” and risk “perpetuating past injustices” (Jungherr & Schroeder 2023, p. 5).

Besides such technical distortion, several practices allow human actors to manipulate the process of public opinion formation, including “impersonating public figures, using synthetic digital personas to simulate grassroots support for or against a cause (‘astroturfing’) and creating falsified media” (Marchal et al., 2024, p. 12). These practices are often employed by so-called “entrepreneurs of polarization” (Mau et al., 2026, pp. 216–219) in the context of politically divisive issues. Additionally, the capacity of generative AI to promote a particular worldview has been debated. For example, it has been suggested that Grok (SpaceX AI) is less “woke” than other AIs. However, this was not confirmed in a comparative test of five models (Rayner, 2025).

“Bias” is the umbrella term used to describe all forms of distorted and unjustified decisions made by AI systems that are subject to criticism. It is broadly defined as “the inclination or prejudice of a decision made by an AI system which is for or against one person or group, especially in a way considered to be unfair” (Ntoutsis et al., 2019, p. 3). According to a computational literature review, “multiple meanings and treatments of bias coexist without an overarching framework and understanding dominating” (Jarrahi et al., 2026, p. 14). Others confirm that theoretical reflection is lacking. In particular, the discussion on AI and democracy “largely takes place at a very general and atheoretical level” (Karppinen et al., 2025, p. 44; see also Helberger et al., 2022, p. 1609).

This theoretical gap can be addressed using normative theories of democracy to clarify expectations placed on the public sphere (Martinsen, 2009) and journalism (Christians et al., 2009). These theories partly overlap but also set different priorities and may even contradict each other on specific points. According to *liberal* theory (Christians et al., 2009, pp. 96–101; Ferree et al., 2002, pp. 290–295; Helberger, 2019, pp. 998–1001), not everyone needs to be able to publicly express their opinion; it is sufficient if representatives of groups articulate the diversity of opinions proportionally. Further, the validation of positions in public discourse is not required. In cases of irreconcilable conflicts of interest, a compromise is a sufficient outcome. Liberal theory thus trusts the self-correcting nature of the “free market of ideas” and places individual (negative) freedom at the center. Accordingly, the state should not intervene to regulate the media, and criticism and control of the state (by “watchdogs”) are necessary to prevent abuses of power. The privacy of individuals must be protected against state interference, but democracy itself should also be defended against its enemies. Meanwhile, *deliberative* theory (Christians et al., 2009, pp. 101–103; Ferree et al., 2002, pp. 300–306; Helberger, 2019, pp. 1004–1007) makes greater demands on the discourse process than other theories. Based on inclusive, responsive, reasoned, and civil exchanges of arguments and counterarguments, consensus must be achieved through coercion-free persuasion (Wessler, 2018, pp. 86–88). Along these lines, *participatory* theory (Christians et al., 2009, pp. 103–105; Ferree et al., 2002, pp. 295–299; Helberger, 2019, pp. 1002–1004) argues that citizens should participate in political discourse and be empowered regardless of the quality of their contributions. All citizens should be able to express themselves authentically, narratively, and emotionally.

Taken together, nine democratic values can be derived from these normative theories of democracy: freedom, equality, diversity, integration, knowledge quality (see above), discursive quality, the distribution of opinion power, critique and control, and security (Neuberger, 2022c; 2024). Tensions (trade-offs) exists between these values as concerns their realization. Furthermore, there are differing views on their meaning and application, which must be clarified within the democratic discourse itself. Nevertheless, these values can serve as yardsticks for both the validity claim of “normative rightness” and the conditions of democratic discourse itself, which demands “value-sensitive designs” (Helbing et al., 2023, p. 3; see also Coeckelbergh, 2024, pp. 70–119), for instance, in the case of news recommenders (Helberger, 2019, pp. 1007–1009).

Within this normative framework, which AI applications are detrimental or conducive to public opinion formation? Empirical evidence is still insufficient to answer this question definitively. The risks associated with generative AI and opinion formation include radicalization and confirmation bias, known as “sycophancy”. Dukić et al. (2026) show that conversations with AI chatbots can lead to cumulative radicalization. Whether this occurs depends mainly on the model and is less influenced by user prompts. Only Claude (Anthropic) consistently de-escalated the conversation. Nogueira et al. (2026) tested 13 large language models in five-turn conversations through asking for its opinion (direct probing) or engaging them in an argumentative debate (indirect probing) on different topics. A sycophantic drift, i.e. the trajectory of models moving in the direction of user’s opinion, could be observed in 44% of conversations. Agreement and disagreement of AI depended on the distinct model and experimental design. In another experiment comparing 11 models (Cheng et al., 2026), AI affirmed users’ actions 49% more often than humans, on average, even in cases of deception, illegality, and other forms of harm. AI confirmation increased users’ conviction that they were right.

However, generative AI can also improve the quality of deliberation (as a research overview, see Friess et al., 2025) and enhance the communication skills of disadvantaged groups. Argyle et al. (2023) demonstrated in an experiment that an AI chat assistant can help a user formulate discussion contributions through politeness, restatement (i.e., repetition to signal understanding), and validation (as confirmation that the other position is legitimate). Participants in the study stated that the quality of the discussion on gun regulation in the United States had improved as a result of using AI. Other research suggests that LLM-powered chatbots can enhance news diversity (Heitz et al., 2022), adapt responses to different audiences on issues such as climate change and Black Lives Matter (Chen et al., 2022), and present compelling arguments on political matters (Palmer & Spirling, 2023). Furthermore, AI can help moderate uncivil user comments (Stockinger et al., 2023). A full agenda on the democratic quality of AI should encompass all values derived from normative theories of democracy.

5 Conclusion

This essay discussed the question of whether AI can enhance journalism from two angles, corresponding to the two democratic functions of journalism: news coverage and public opinion formation. To answer this question, a theoretical framework was developed that provides quality benchmarks for AI applications and its empirical assessment. To this end, values were derived from normative theories of democracy. These define the appropriate conditions for news production and public discourse. To date, only few empirical studies have demonstrated how the use of AI can strengthen the democratic functions of journalism. Instead, research has primarily focused on potential applications and subjective assessments of the advantages and disadvantages of AI from the perspectives of journalists and audiences. Meanwhile, hard empirical evidence of the epistemic and evaluative quality of AI output, its effect on audiences’ political knowledge and attitudes on the micro level, and its impact on public discourse at the macro level is still lacking. However, such studies are essential to

the adaptation of AI in journalism. Their findings could provide editorial teams with valuable feedback. The rapid development of AI models requires continuous monitoring and an experimental approach to understand the potential uses and limitations of AI in journalism and the democratic public sphere. Innovation labs (García-Avilés, 2024) and pioneering communities (Hepp & Loosen, 2021), such as the “AI for Media Network” (<https://aiformedia.network/en/>), are testing grounds for AI-enhanced journalism.

References

- Argyle, L. P., Bail, C. A., Busby, E. C., Gubler, J. R., Howe, T., Rytting, C., ... & Wingate, D. (2023). Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale. *Proceedings of the National Academy of Sciences*, 120(41), Article e2311627120.
- Ashuri, T., Zalmanson, L., & Goldstein, D. (2026). Journalistic epistemic authority in the Age of AI: The incorporation of generative AI in newsrooms. *Digital Journalism*, 14(5), 830–849. <https://doi.org/10.1080/21670811.2026.2640421>
- Bartsch, A., Neuberger, C., Stark, B., Karnowski, V., Maurer, M., Pentzold, C., Quandt, T., Quiring, O., & Schemer, C. (2025). Epistemic authority in the digital public sphere. An integrative conceptual framework and research agenda. *Communication Theory*, 35(1), 37–50. <https://doi.org/10.1093/ct/qtae020>
- Becker, K. B., Simon, F. M., & Crum, C. (2025). Policies in parallel? A comparative study of journalistic AI policies in 52 global news organisations. *Digital Journalism*, 0(0), 1–21. <https://doi.org/10.1080/21670811.2024.2431519>
- Boghossian, P. (2006). *Fear of knowledge: Against relativism and constructivism*. Oxford University Press.
- Campolo, A., & Schwerzmann, K. (2023). From rules to examples: Machine learning’s type of authority. *Big Data & Society*, 10(2). <https://doi.org/10.1177/20539517231188725>
- Cau, E., Pansanella, V., Pedreschi, D., & Rossetti, G. (2025). Selective agreement, not sycophancy: investigating opinion dynamics in LLM interactions. *EPJ Data Science*, 14, Article 59. <https://doi.org/10.1140/epjds/s13688-025-00579-1>
- Chalikiopoulou, E., Saridou, T., & Veglis, A. (2025). The role of generative AI in enhancing audience participation in journalism: A scoping review. *Societies*, 15, Article 358. <https://doi.org/10.3390/soc15120358>
- Chen, K., Shao, A., Burapacheep, J., & Li, Y. (2022). How GPT-3 responds to different publics on climate change and Black Lives Matter: A critical appraisal of equity in conversational AI. *arXiv preprint*, arXiv:2209.13627v2. <https://doi.org/10.48550/arXiv.2209.13627>

- Cheng, M., Lee, C., Khadpe, P., Yu, S., Han, D., & Jurafsky, D. (2026). Sycophantic AI decreases prosocial intentions and promotes dependence. *Science*, 391(6792), Article eaec8352. <https://doi.org/10.1126/science.aec8352>
- Christians, C. G., Glasser, T. L., McQuail, D., Nordenstreng, K., & White, R. A. (2009). *Normative theories of the media: Journalism in democratic societies*. University of Illinois Press.
- Coeckelbergh, M. (2025). LLMs, truth, and democracy: An overview of risks. *Science and Engineering Ethics*, 31, Article 4. <https://doi.org/10.1007/s11948-025-00529-0>
- Coeckelbergh, M. (2024). *Why AI undermines democracy and what to do about it*. Polity.
- Cools, H., & Diakopoulos, N. (2024). Uses of generative AI in the newsroom: Mapping journalists' perceptions of perils and possibilities. *Journalism Practice*, 0(0), 1–19. <https://doi.org/10.1080/17512786.2024.2394558>
- Costello, T. H., Pelrine, K., Kowal, M., Timm, J., Arechar, A. A., Godbout, J.-F., Gleave, A., Pennycook, G., & Rand, D. (2026). AI can effectively promote conspiracies unless it is truth constrained. *arXiv preprint*, arXiv:2601.05050v3. <https://doi.org/10.48550/arXiv.2601.05050>
- Crawford, K. (2025, September). Eating the future: The metabolic logic of AI slop. *e-flux Architecture*. <https://www.e-flux.com/architecture/intensification/6782975/eating-the-future-the-metabolic-logic-of-ai-slop>
- Deitelhoff, N., & Schmelzle, C. (2023). Social integration through conflict: Mechanisms and challenges in pluralist democracies. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 75 (Suppl 1), 69–93. <https://doi.org/10.1007/s11577-023-00886-3>
- Dukić, S., Vela, N., & Venkataramakrishnan, S. (2026). *Radicalisation in closed loops: Risks and intervention opportunities with AI chatbots and companions*. Institute for Strategic Dialogue, AI Security Institute. <https://www.isdglobal.org/publication/radicalisation-in-closed-loops-risks-and-intervention-opportunities-with-ai-chatbots-and-companions/>
- Emil, R. Ş., & Remus, B. (2025). A review of automatic fake news detection: From traditional methods to large language models. *Future Internet*, 17(10), Article 435. <https://doi.org/10.3390/fi17100435>
- Esposito, E. (2022). *Artificial communication: How algorithms produce social intelligence*. MIT Press.
- Esposito, E. (2024). *Kommunikation mit unverständlichen Maschinen*. Residenz Verlag.
- Esposito, E. (2026). Answer engines and other communication partners. *Communication Theory*, 36(2), 86–94. <https://doi.org/10.1093/ct/qtaf036>
- Farrell, H., Gopnik, A., Shalizi, C., & Evans, J. (2025). Large AI models are cultural and social technologies. *Science*, 387(6739), 1153–1156. <https://doi.org/10.1126/science.adt9819>
- Ferree, M. M., Gamson, W. A., Gerhards, J., & Rucht, D. (2002). Four models of the public sphere in modern democracies. *Theory and Society*, 31(3), 289–324. <https://doi.org/10.1023/A:1016284431021>

- Fridman, M., Krøvel, R., & Palumbo, F. (2025). How (not to) run an AI project in investigative journalism. *Journalism Practice*, 19(6), 1362–1379. <https://doi.org/10.1080/17512786.2023.2253797>
- Friess, D., Weinmann, C., & Warné, M. (2025). AI and deliberation: Normative ideals in the light of current AI research in the context of online discussions – A literature review. *Journal of Deliberative Democracy*, 21(1), 1–11. <https://doi.org/10.16997/jdd.1805>
- García-Avilés, J.-A. (2024). Media labs: agents of innovation. In K. Meier, J.-A. García-Avilés, A. Kaltenbrunner, C. Porlezza, V. Wyss, R. Lugschitz, & K. Klinghardt (Eds.), *Innovations in journalism: Comparative research in five European countries* (pp. 227–233). Routledge.
- Gillman, N., Freeman, M., Aggarwal, D., Hsu, C.-H., Luo, C., Tian, Y., & Sun, C. (2024). Self-correcting self-consuming loops for generative model training. *arXiv preprint*, arXiv: 2402.07087v3. <https://doi.org/10.48550/arXiv.2402.07087>
- Godler, Y., & Reich, Z. (2017). Journalistic evidence: Cross-verification as a constituent of mediated knowledge. *Journalism*, 18(5), 666–681. <https://doi.org/10.1177/1464884915620268>
- Goldman, A. I. (1999). *Knowledge in a social world*. Clarendon Press.
- Graefe, A., & Bohlken, N. (2020). Automated journalism: A meta-analysis of readers' perceptions of human-written in comparison to automated news. *Media and Communication*, 8(3), 50–59. <https://doi.org/10.17645/mac.v8i3.3019>
- Gunkel, D. J. (2012). Communication and artificial intelligence: Opportunities and challenges for the 21st century", *communication +1*, 1, Article 1. <https://doi.org/10.7275/R5QJ7F7R>
- Gutiérrez-Caneda, B., Vázquez-Herrero, J., & López-García, X. (2023). AI application in journalism: ChatGPT and the uses and risks of an emergent technology. *Profesional de la información*, 32(5). <https://doi.org/10.3145/epi.2023.sep.14>
- Habermas, J. (1981). *Theorie des kommunikativen Handelns. Band 1: Handlungsrationalität und gesellschaftliche Rationalisierung*. Suhrkamp.
- Habermas, J. (2006). Political communication in media society: Does democracy still enjoy an epistemic dimension? The impact of normative theory on empirical research. *Communication Theory*, 16(4), 411–426. <https://doi.org/10.1111/j.1468-2885.2006.00280.x>
- Harari, Y. N. (2024). *Nexus. A brief history of information networks from the Stone Age to AI*. Fern Press.
- Heitz, L., Lischka, J. A., Birrer, A., Paudel, B., Tolmeijer, S., Laugwitz, L., & Bernstein, A. (2022). Benefits of diverse news recommendations for democracy: A user study. *Digital Journalism*, 10(10), 1710–1730. <https://doi.org/10.1080/21670811.2021.2021804>
- Helberger, N. (2019). On the democratic role of news recommenders. *Digital Journalism*, 7(8), 993–1012. <https://doi.org/10.1080/21670811.2019.1623700>
- Helberger, N., van Drunen, M., Moeller, J., Vrijenhoek, S., & Eskens, S. (2022). Towards a normative perspective on journalistic AI: Embracing the messy reality of normative ideals. *Digital Journalism*, 10(10), 1605–1626. <https://doi.org/10.1080/21670811.2022.2152195>

- Helbing, D., Mahajan, S., Hänggli Fricker, R., Musso, A., Hausladen, C. I., Carissimo, C., Carpentras, D., Stockinger, E., Sanchez-Vaquerizo, J. A., Yang, J. C., Ballandies, M. C., Korecki, M., Dubey, R. K., & Pournaras, E. (2023). Democracy by design: Perspectives for digitally assisted, participatory upgrades of society. *Journal of Computational Science*, 71, Article 102061. <https://doi.org/10.1016/j.jocs.2023.102061>
- Hepp, A., & Loosen, W. (2021). Pioneer journalism: Conceptualizing the role of pioneer journalists and pioneer communities in the organizational re-figuration of journalism. *Journalism*, 22(3), 577–595. <https://doi.org/10.1177/1464884919829277>
- Herasimenka, A., Valenzuela, S., Boulianne, S., Esser, F., Given, L. M., Lewandowsky, S., Navarro-López, E. M., & Howard, P. N. (Eds.). (2026). *Confronting misinformation produced with generative AI: A meta-analysis of experimental scientific evidence*. International Panel on the Information Environment. SR2026.2. <https://doi.org/10.61452/UGTR3022>
- Hirmiz, R. (2024). The epistemic role of AI decision support systems: Neither superiors, nor inferiors, nor peers. *Philosophy & Technology*, 37, Article 127. <https://doi.org/10.1007/s13347-024-00819-8>
- Jarrahi, M. H., Karami, A., Conway, P., Memariani, A., & Lutz, C. (2026). Navigating the muddy waters of bias in artificial intelligence research: Understanding divergent meanings and conceptions. *Technology in Society*, 84, Article 103127. <https://doi.org/10.1016/j.tech-soc.2025.103127>
- Jungherr, A. (2023). Artificial intelligence and democracy: A conceptual framework. *Social Media + Society*, 9(3). <https://doi.org/10.1177/20563051231186353>
- Jungherr, A., & Schroeder, R. (2023). Artificial intelligence and the public arena. *Communication Theory*, 33(2–3), 164–173. <https://doi.org/10.1093/ct/qtad006>
- Karppinen, K., Moe, H., & Svensson, J. (2025). What democracy are we talking about? Reviewing the debate on the implications of generative AI for democracy. *Journal of Information Policy*, 15, 31–56. <https://doi.org/10.5325/jinfopoli.15.2025.0002>
- Koskinen, I. (2024). We have no satisfactory social epistemology of AI-based science. *Social Epistemology*, 38(4), 458–475. <https://doi.org/10.1080/02691728.2023.2286253>
- Kovach, B., & Rosenstiel, T. (2007). *The elements of journalism. What newspeople should know and the public should expect*. Three Rivers Press.
- Kuznetsova, E., Makhortykh, M., Vziatysheva, V., Stolze, M., Baghumyan, A., & Urman, A. (2025). In generative AI we trust: Can chatbots effectively verify political information? *Journal of Computational Social Science*, 8, Article 15. <https://doi.org/10.1007/s42001-024-00338-8>
- Luhmann, N. (1996). *Die Realität der Massenmedien* (2nd ed.). Westdeutscher Verlag.
- Lynch, M. P. (2012). *In praise of reason: Why rationality matters for democracy*. The MIT Press.
- Marchal, N., Xu, R., Elasmr, R., Gabriel, I., Goldberg, B., & Isaac, W. (2024). Generative AI misuse: A taxonomy of tactics and insights from real-world data. *arXiv preprint*, arXiv:2406.13843v2. <https://doi.org/10.48550/arXiv.2406.13843>

- Martinsen, R. (2009). Öffentlichkeit als „Mediendemokratie“ aus der Perspektive konkurrierender Demokratietheorien. In F. Marcinkowski & B. Pfetsch (Eds.), *Politik in der Mediendemokratie* (pp. 38–69). VS Verlag für Sozialwissenschaften.
- Mau, S., Lux, T., & Westheuser, L. (2026). *Trigger points: Inequality and political polarization in contemporary society*. Bristol University Press.
- Miragoli, M. (2025). Conformism, ignorance & injustice: AI as a tool of epistemic oppression. *Episteme*, 22(2), 522–540. <https://doi.org/10.1017/epi.2024.11>
- Nah, S., Luo, J., Kim, S., Chen, M., Mitson, R., & Joo, J. (2024). Algorithmic bias or algorithmic reconstruction? A comparative analysis between AI news and human news. *International Journal of Communication*, 18, 700–729.
- Neuberger, C. (2009). Internet, Journalismus und Öffentlichkeit. Analyse des Medienumbruchs. In C. Neuberger, C. Nuernbergk & M. Rischke (Eds.), *Journalismus im Internet: Profession – Partizipation – Technisierung* (pp. 19–105). VS Verlag für Sozialwissenschaften.
- Neuberger, C. (2017). Journalistische Objektivität. Vorschlag für einen pragmatischen Theorierahmen. *Medien & Kommunikationswissenschaft*, 65(2), 406–431. <https://doi.org/10.5771/1615-634X-2017-2-406>
- Neuberger, C. (2022a). Journalismus und Plattformen als vermittelnde Dritte in der digitalen Öffentlichkeit. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 74(62), 159–181. <https://doi.org/10.1007/s11577-022-00832-9>
- Neuberger, C. (2022b). How to capture the relations and dynamics within the networked public sphere? Modes of interaction as a new concept. In B. Krämer & P. Müller (Eds.), *Questions of communicative change and continuity. In memory of Wolfram Peiser* (pp. 67–95). Nomos.
- Neuberger, C. (2022c). Digitale Öffentlichkeit und liberale Demokratie. In Bundeszentrale für politische Bildung (Ed.), *Repräsentation – Identität – Beteiligung. Zum Zustand und Wandel der Demokratie* (pp. 381–394). Bundeszentrale für politische Bildung.
- Neuberger, C. (2024). Werte als Maßstab der liberal-demokratischen Öffentlichkeit. In M. Prinzing, J. Seethaler, M. Eisenegger, & P. Ettinger (Eds.), *Regulierung, Governance und Medienethik in der digitalen Gesellschaft. Mediensymposium* (pp. 23–43). Springer VS. https://doi.org/10.1007/978-3-658-42478-7_2
- Neuberger, C. (2025). Journalismus in der Gesellschaft. In T. Hanitzsch, W. Loosen, & A. Sehl (Eds.), *Journalismusforschung* (pp. 35–58). Nomos. <https://doi.org/10.5771/9783748932291-35>
- Neuberger, C., Bartsch, A., Fröhlich, R., Hanitzsch, T., Reinemann, C., & Schindler, J. (2023). The digital transformation of knowledge order: a model for the analysis of the epistemic crisis. *Annals of the International Communication Association*, 47(2), 180–201. <https://doi.org/10.1080/23808985.2023.2169950>
- Nogueira, R., Bonás, G. K., Almeida, T. S., Roque, A., Pires, R., Abonizio, H., Laitz, T., Larcher, C., Malaquias Junior, R., & Piau, M. (2026). Measuring opinion bias and sycophancy via LLM-based persuasion. *arXiv preprint*, arXiv:2604.21564v2. <https://doi.org/10.48550/arXiv.2604.21564>

- Ntoutsis, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdil, W., Vidal, M.-E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., Broelemann, K., Kasneci, G., Tiropanis, T., & Staab, S. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *WIREs Data Mining and Knowledge Discovery*, 10, Article e1356. <https://doi.org/10.1002/widm.1356>
- Opdahl, A. L., Tessem, B., Dang-Nguyen, D., Setty, V., Throndsen, E., Tverberg, A., & Trattenner, C. (2023). Trustworthy journalism through AI. *Data & Knowledge Engineering*, 146. <https://doi.org/10.1016/j.datak.2023.102182>
- Palmer, A., & Spirling, A. (2023). Large language models can argue in convincing ways about politics, but humans dislike AI authors: Implications for governance. *Political Science*, 75(3), 281–291. <https://doi.org/10.1080/00323187.2024.2335471>
- Perreault, G., & Ohme, J. (2025). Chatbots as artificial intermediaries? Adaptation to artificial intelligence in newsrooms. *Journalism Studies*, 26(15), 1914–1935. <https://doi.org/10.1080/1461670X.2025.2567894>
- Petroni, F., Broscheit, S., Piktus, A., Lewis, P., Izacard, G., Hosseini, L., ... & Riedel, S. (2023). Improving Wikipedia verifiability with AI. *Nature Machine Intelligence*, 5(10), 1142–1148. <https://doi.org/10.1038/s42256-023-00726-1>
- Rayner, M. (2025). *How woke is Grok? Empirical evidence that xAI's Grok aligns closely with other frontier models*. <https://doi.org/10.13140/RG.2.2.35804.45449>
- Reich, Z., & Lahav, H. (2021). What on earth do journalists know? A new model of knowledge brokers' expertise. *Communication Theory*, 31(1), 62–81. <https://doi.org/10.1093/ct/qtaa013>
- Ross Arguedas, A. (2024). Public attitudes towards the use of AI in journalism. In N. Newman, R. Fletcher, C. T. Robertson, A. Ross Arguedas, & R. K. Nielsen (Eds.). *Reuters Institute Digital News Report 2026* (pp. 39–43). *Reuters Institute for the Study of Journalism*. <https://doi.org/10.60625/risj-vy6n-4v57>
- Schäfer, M. S., Chen, K., Mahl, D., Painter, J., & Volk, S. C. (2026). Climate change communication in the Age of Artificial Intelligence. *Wiley Interdisciplinary Reviews: Climate Change*, 17(3), Article e70073. <https://doi.org/10.1002/wcc.70073>
- Schoenbach, K. (2007). 'The own in the foreign': reliable surprise – an important function of the media? *Media, Culture & Society*, 29(2), 344–353. <https://doi.org/10.1177/0163443707074269>
- Scott, C. P. (1921/2017, October 23). A hundred years. *The Guardian*. <https://www.theguardian.com/sustainability/cp-scott-centenary-essay>
- Shin, D. (2025). Automating epistemology: how AI reconfigures truth, authority, and verification. *AI & Society*, 41, 1553–1559. <https://doi.org/10.1007/s00146-025-02560-y>
- Simon, F. M., Altay, S., & Mercier, H. (2023, October 18). Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. *Harvard Kennedy School Misinformation Review*, 4(5). <https://doi.org/10.37016/mr-2020-127>

- Simon, J. (2026). Generative AI, quadruple deception & trust. *Social Epistemology*, 40(1), 101–115. <https://doi.org/10.1080/02691728.2025.2491087>
- Somers, J. (2025, November 3). The case that A.I. is thinking. ChatGPT does not have an inner life. Yet it seems to know what it's talking about. *The New Yorker*. <https://www.newyorker.com/magazine/2025/11/10/the-case-that-ai-is-thinking>
- Stockinger, A., Schäfer, S., & Lecheler, S. (2023). Navigating the gray areas of content moderation: Professional moderators' perspectives on uncivil user comments and the role of (AI-based) technological tools. *New Media & Society*, 0(0). <https://doi.org/10.1177/14614448231190901>
- von Sikorski, C., & Hameleers, M. (2025). Disinformation in the Age of Artificial Intelligence (AI): Implications for journalism and mass communication. *Journalism & Mass Communication Quarterly*, 102(4), 941–957. <https://doi.org/10.1177/10776990251375097>
- Wachter, S., Mittelstadt, B., & Russell, C. (2024). Do large language models have a legal duty to tell the truth? *Royal Society Open Science*, 11(8), Article 240197. <https://doi.org/10.1098/rsos.240197>
- Wessler, H. (2018). *Habermas and the media*. Polity.
- Xing, X., Shi, F., Huang, J., Wu, Y., Nan, Y., Zhang, S., Fang, Y., Roberts, M., Schönlieb, C.-B., Del Ser, J., & Yang, G. (2025). On the caveats of AI autophagy. *Nature Machine Intelligence*, 7, 172–180. <https://doi.org/10.1038/s42256-025-00984-1>
- Xu, K., Shen, Y., Yan, L., & Ren, Y. (2026). Cognitive agency surrender: Defending epistemic sovereignty via scaffolded AI friction. *arXiv preprint*, arXiv:2603.21735v2. <https://doi.org/10.48550/arXiv.2603.21735>

Date accepted: 17 August 2026