Communication Theory, 2023, **33**, 164–173
https://doi.org/10.1093/ct/qtad006
Advance access publication 19 June 2023
Original Article

Artificial intelligence and the public arena

Andreas Jungherr [1,*], **Ralph Schroeder**[2]

[1]Institute for Political Science, University of Bamberg, Bamberg, Germany
[2]Oxford Internet Institute, University of Oxford, Oxford, UK

*Corresponding author: Andreas Jungherr. Email: andreas.jungherr@uni-bamberg.de

Abstract

The public arena relies on artificial intelligence (AI) to ever greater degrees. Media structures hosting the public arena—such as Facebook, TikTok, Twitter, and YouTube—increasingly rely on AI-enabled applications to shape information environments, autonomously generate content, and communicate with people. These applications affect the public arena's functions: make society visible to itself and provide spaces for the formation of publics and counterpublics. We offer a framework that allows for the conceptualization and empirical examination of AI's structural impact on the public arena. Based on this perspective, we argue that the growing uses of AI will lead to a strengthening of intermediary structures that can exercise a greater degree of control over the public arena. In addition, the data-driven nature of most AI-applications threatens to push challenges to the political status quo out of sight and obstruct the assessability of AI-enabled interventions.

Keywords: public sphere, public arena, artificial intelligence (AI), large language models (LLMs), ChatGPT

Important parts of the contemporary public sphere rely on digital structures that increasingly depend on artificial intelligence (AI). This has given rise to widespread fears regarding whether the associated affordances, ownership, and workings distort the public sphere (Habermas, 2022). Understanding these intermediary structures and their impact on the contemporary public sphere is important for communication research. There have been instructive accounts of AI's impact on the perceptions and behaviors of actors (Esposito, 2022; Guzman & Lewis, 2020; Natale, 2021) and specific institutions (Diakopoulos, 2019; Napoli, 2014; Simon, 2022) within the public sphere. We add to this literature by focusing explicitly on AI's structural impact on the *public arena*. Using this term emphasizes our concern with conditions, changes, and functions on the structural level, providing an alternative to the more expansive *public sphere*, which tends to combine a specific normative ideal of content and forms of discourse with structural considerations, making it difficult to untangle those in practice.

The contemporary public arena relies on AI to ever greater degrees. Media structures—such as Facebook, TikTok, Twitter, and YouTube—increasingly rely on AI-driven applications to shape information environments and user behavior, autonomously generate content, and communicate with people. These applications thus affect the public arena's functions: make society visible to itself and provide spaces for the formation of publics and counterpublics. They also impact the assessability of structures.

This article presents a novel theoretical account focused on AI's structural impact in the public arena. We build on extensive prior empirical work and connect earlier findings to a larger synthetic and theoretical account about AI's role in the contemporary public arena. After defining AI and discussing its workings, we identify the functionings of the public arena and distinguish it from the public sphere. We continue by discussing specific applications of AI within the public arena, focusing on shaping information and behavior, content generation, and AI's role as communicating agent. We close by discussing AI's impact on the assessability of the public arena. We argue that these developments will lead to a strengthening of intermediary structures of the public arena, the data-driven submersion of challenges to the status quo, and an associated strengthening of control by gatekeepers.

Artificial intelligence

Examples abound of the uses of AI in the public arena (Schäfer & Wessler, 2020). Researchers have examined AI in moderating speech (Douek, 2021), shaping information flows in news and everyday digital environments (Bandy & Diakopoulos, 2021; Elahi et al., 2022; Trielli & Diakopoulos, 2019), and journalism and content creation (Diakopoulos, 2019; Napoli, 2014; Simon, 2022).

AI can be defined as "the study and construction of agents that do the right thing" (Russell & Norvig, [1995] 2021, p. 22). Understood narrowly, this could be the ability of machines to pursue specific tasks of varying difficulty in appropriate ways or, more broadly, the ability to set goals autonomously, reason, and adapt to unforeseen circumstances. Current successes are closely connected with advances in machine learning, especially within the field of deep learning (LeCun et al., 2015). The public discussion of AI focuses predominantly on machine systems with human-level cognition and decision-making capabilities—so-called artificial general intelligence (AGI). But this type of AI remains elusive (Larson, 2021; Smith, 2019), and current AI is best characterized as *narrow AI*—AI-enabled systems developed for a specific, singular, or limited task (Mitchell, 2019, p. 45f). This difference is important, considering that most public expectations overestimate the power of AI while ignoring its limitations and the preconditions for its successful application.

Current discussion features the surprising successes of applications using generative models for text and image generation in response to prompts (e.g., ChatGPT or

Received: 21 September 2022. **Revised:** 30 April 2023. **Accepted:** 5 May 2023

Midjourney) (Brown et al., 2020; Vaswani et al., 2017). This has given renewed rise to unfocused enthusiasm and fears about AGI that hide AI's actual workings and effects. Instead, we must also account for other more mundane forms and uses of AI, including some that until recently have been mostly discussed as *algorithmic decision making*.

Algorithms, which are predefined series of steps in pursuit of addressing given problems, are foundational building blocks of computing and form the basis of computer-enabled automation (Cormen et al., 2022). Their steps can be specified for narrow problems, such as treating tweets that receive replies as four times more relevant than those that receive a single retweet in deciding whether to suggest them to other users. Alternatively, the predefined series of steps can focus on a meta-task, defining general steps in determining how best to find solutions to a wide variety of tasks (Kelleher, 2019). For example, instead of hard coding a rule for ranking the relevance of tweets, a machine learning algorithm could be programmed to identify from data the kinds of tweets users treated as relevant. The machine would simply follow a predefined sequence of steps that allow it to identify the best way of solving a given type of problem based on data documenting inputs and outputs—and thus the machine learns how to rank tweets on its own. This process corresponds to the *narrow AI* concept.

These approaches both allow for automation through algorithmic decision making. The rules-based approach lets humans develop rules for specific types of tasks and automatically deploy them through an algorithm; in the data-driven approach, the machine itself follows a predefined set of steps for best identifying rules, based on data, and then deploys them broadly. Both approaches raise fears about the widespread, indiscriminate rollout of rules, with additional concerns regarding the appropriate development and deployment of rules derived from data in the second approach. Our focus in the following discussion is on this second approach, a form of *narrow AI*, and the specific challenges associated with purely data-driven learning and automated decision making for structures hosting the public arena.

For data-driven learning to provide valid inferences about the world, several preconditions must hold. The successful application of AI depends on the availability of machine-readable data that objectively and abundantly document the domain of interest. Further, the connection between available data inputs and outcomes of interest must be stable over time. Also, AI systems face challenges in cases where, normatively speaking, past patterns should not be replicated in the present and future (Jungherr, 2023). These conditions determine whether AI can or, better, should pursue tasks based on prior learnings.

AI-based learnings are inherently conservative, and data-driven predictions generally lean toward the average of a given set of cases. By relying on patterns found in past data, AI will pursue tasks in ways that succeeded in the past but may no longer be appropriate (Vela et al., 2022), either due to unobserved shifts between inputs and outputs (Lazer et al., 2014) or a shift in values and norms that make past learnings obsolete (Bender et al., 2021). This makes AI-supported shapings conservative. Further, data-driven predictions will pull outcomes toward the average. By learning the expected average outcome given a set of inputs and absent some correction, AI will push systems to that average.

Consider large language models (LLMs) such as BERT, GPT-4, and LLaMA, which are often used to translate or edit texts. They do so based on probability distributions of word sequences in response to given inputs. These probabilities have been identified through neural networks trained on extensive text corpora and stored as weights within the neural network (Wolfram, 2023). This works perfectly well in contexts where abundant data allow for robust identification of stable patterns; with scarce training data, the model performs worse. For example, translating a standard news item from English into German should be easy, but a correct, nuanced translation of a highly idiosyncratic opinion piece or the culturally coded voicing of grievances in a public forum will be more difficult. We can expect the model to fall back on general language patterns and edit out the idiosyncrasies or cultural codes that made the original content meaningful. Once a specific context is not well documented or correctly identified, data-driven predictions will tend to the average and edit out specifics or idiosyncrasies.

The public arena

The public arena and the public sphere

The public arena depends on media structures that make society visible to itself. It allows societies to settle crucial issues and control elites and governments, and for groups with shared concerns to emerge. It also provides spaces for political competition. These structures increasingly rely on AI, at least in part, and thus understanding the effects of AI is important for understanding public arena functioning.

In previous work (Jungherr & Schroeder, 2022), we defined the public arena as interconnected communicative spaces hosted by media structures—such as news media, digital platforms, and discursive institutions—that enable and constrain the publication, distribution, reception, and contestation of information that allows people to exercise their rights and duties in pursuit of the public good. These structures mediate the relationship between people and political elites and between civil society and the state, hosting public discourse and political competition while also providing people with information about common concerns. The concept comes with some normative expectations as to these *functions* of the public arena, but not with specific normative expectations about the *form* that this pursuit of common concerns should take.

We follow prior work that emphasizes the importance of structures that host public reflection and debate (Gerhards & Neidhardt, 1991; Habermas, [1962] 1990, 1992; Peters, 2007). More specifically, we build on previous uses of the term *arena* to conceptualize these spaces. But unlike Gerhards & Neidhardt (1991), for example, our idea of the public arena does not focus exclusively on competition between politicians, factions, and interest groups in a space where the public is an observer, as in a gallery.

We also do not commit to *one* normative account of the specific form that political discourse should take. Instead, we recognize the competition between different normative ideals of discourse and present a structural account open to different normative ideals. For example, Habermas' idea of the public sphere features explicit normative conditions about the form of these public exchanges (the need for "rationality") (Habermas, [1962] 1990, 1992) and assigns to these

expectations a specific function of intermediating between political systems, the lifeworld, and specific societal sub-systems (Habermas, 1992, p. 451f.). Instead, we share, with Peters (2007), the recognition that there are different empirical constellations of structures, publics, and forms of exchange. Unlike Peters (2007), though, our conceptualization is not committed to a specific normative ideal of discourse serving as an ideal type by which to measure the qualities or deficiencies of the public arena. Rather, in our view, the public arena remains open to other normatively grounded forms of exchange in what Asenbaum (2022) has recently termed the "kaleidoscope of democratic theory"—which includes participatory, agonistic, and deliberative exchanges. This is in keeping with our focus on the structures hosting the public arena, not the specific shapes, rhythms, or rules of the spectacle within it.

This is not to negate or contest the specifics of either normative conceptualizations of the public sphere or of democratic discourse. Instead, our choice recognizes the plurality of different, competing, and sometimes contradictory normative expectations of what constitutes legitimate discourse, political contestation, and competition. Even insightful discussions of the impact of digital technology in the public sphere committed to either of these normatively demanding concepts often become more of an exercise in defending specific normative choices or lament that contemporary developments fall short of expectations based on ideal-typical considerations. Focusing instead on *structures* of the public arena allows a more focused discussion about the (shifting) conditions under which political discourse, competition, and self-reflection occur. This surfaces persistence in the features and effects of and balance between various forces constituting the public arena, as well as the shifts driven by technology, the economics of media infrastructures, and the changes in norms and their impact on discourse and democracy.

Functions of the public arena

The public arena serves two important structural functions in society (Rauchfleisch & Kovic, 2016): make society visible to itself (Luhmann, [1995] 2017); and provide common and counterpublic spaces for people to pursue the public good and develop shared political identities. To succeed, it is crucial that the public arena allows for the contestation of structures and information that in turn creates demands for their assessability.

With respect to visibility the public arena makes elites visible to people, people visible to elites, and people visible to each other. The mechanics and demands vary. Elites become visible in the public arena by openly competing for support, advertising positions, and publicly documenting activities (Gerhards & Neidhardt, 1991). The public arena allows for public "supervision" of government initiatives (Warner, 1990, p. 41), as well as for public control of elites (Gurevitch & Blumler, 1990). This corresponds with the "watchdog" role of news media, as captured by the term "fourth estate" in liberal democratic theory and journalism studies (Schultz, 1998).

The public arena makes people visible to elites by allowing public support of parties and social movements and the formation of new political associations, which establishes visibility of political factions' relative strength (Gerhards & Neidhardt, 1991; Peters, [1994] 2007, p. 70ff.). It also allows elites to see the outcomes and dynamics of public debates over political events, personnel, or current issues.

People also find each other in the public arena and recognize or construct common identities (Peters, [1994] 2007, p. 81; Rauchfleisch & Kovic, 2016). By providing people with spaces to formulate their concerns, interests, and views, the public arena allows people to find others like them and even develop a new sense of shared identity (Bourdieu, 1990). This visibility of people to others includes those who are part of privileged majority groups with political and cultural power and those in minority groups who are disenfranchised or suffer discrimination. Those in the privileged majority enjoy access to (or may even *own*) major venues—indeed, central structures—of the public arena with broad reach and strong cultural and discursive power, such as major media outlets or institutions (e.g., corporations) with privileged access to mass media. This is where the established balance of power in society tends to manifest itself; this access leads to their strong representation and dominance within the public arena and the potential exclusion of outsiders (Bennett, 1990). Accordingly, the role of mass media and corporate influence in reinforcing the power relations in society within the public arena has been heavily criticized in the normative tradition of public sphere theory (Castells, [2009] 2013; Habermas, 1992; Peters, [1994] 2007).

Outsiders, in turn, can become visible in fringe venues of the public arena—such as dedicated niche publications, informal gathering places, and internet forums—where they can exchange information and coordinate beyond the sometimes stifling gaze of majority attention. Here, counterpublics can form that challenge and contest the legitimacy of dominant publics and public arena structures (Fraser, 1990; Warner, 2002). These include minority groups interested in greater representation and extending rights within the existing political system, but also extremist groups that use segmented spaces within the public arena to agitate and mobilize clandestinely against the state, political system, and social groups. Legitimate and illegitimate challenges to existing power structures in society can emerge (Castells, [2009] 2013; Jungherr, Schroeder, et al., 2019).

Within structures that host the public arena, there is an inherent tension between creating a common space for all or a fragmented space that segments groups from the larger common public and shared political identity, and where these groups develop and stage a challenge to majority opinion, for good or ill (Ferree et al., 2002; Peters, [1994] 2007, p. 81). Both types of spaces are important. The common spaces are crucial for the public sphere to be open and not exclusionary (Habermas, 1992, p. 451f.; Peters, [1994] 2007, pp. 70–82). They contribute to a shared sense of a common endeavor for contributors and audiences in the public arena, and are crucial for the emergence of shared agendas (Peters, [1994] 2007, pp. 82–89) and a shared sense of mediated reality (Luhmann, [1995] 2017). They allow for the emergence or contestation of commonly recognized facts (Arendt, [1967] 1968) and the emergence of shared discourse (Peters, [1994] 2007, pp. 89–97), which in turn provides the basis for collectively binding decision-making in a polity. For the public arena to provide common spaces it is not necessary for people to contribute through the same media structures, as long as their contributions on different media reference each other and are interconnected (Jungherr, Posegga, et al., 2019; Peters, [1994] 2007, p. 81; Taylor, 1995).

The need for segmented spaces for likeminded people to find each other, discover shared interests and concerns, and develop political identities and programs of action has also been well established (Fraser, 1990; Warner, 2002). Thus, it seems best not to demand from the public arena that it provides either a space shared by all and at all times, or a fragmented space for competing publics to form and flourish. Instead, it should provide for both at different times.

To make society visible to itself, the public arena has a minimal set of functional preconditions. It is a purely structural account, but at least one normative precondition remains: *transparency*, or better put, *assessability* (Müller, 2021, p. 139) of its workings, governance, and effects. Transparency emphasizes efforts of structures to document and critically reflect openly on technical workings and governance rules (see Ananny & Crawford 2018 for a critique). More broadly, assessability emphasizes the ability of outsiders, such as the public, elites, professional observers, and regulators, to assess the workings, governance, and effects of the structures that host the public arena—to allow the observing of the observers (Luhmann, [1995] 2017). Normative rules for journalism, such as the division between impartial coverage of events and commentary or opinion sections, are an example (Kovach & Rosenstiel, 2021). This rule, though often broken, gives the public the chance to interrogate news practices and contest the breaching of norms. The advent of new structures in today's public arena is challenging these old institutional norms and practices. This helps explain the contemporary unease about communication environments that host the public arena.

Structures of the public arena

News media have long been a core structure of the public area. News as an institution serves society in observing elites and government and provides civil society a space to become visible and observe itself. The news shares specific norms that are transmitted institutionally, as in journalism schools, and policed institutionally, such as through professional boards. Commitments to neutral political coverage and a clear demarcation between news coverage and opinion or commentary are important institutional norms.

Other media have emerged that follow different norms, including partisan media supporting select political factions, news media dedicated to advocating specific causes, and even those dedicated to muckraking in pursuit of attention they can monetize. Their different norms and motives extend the public arena, providing challengers of the status quo greater opportunities to gain access and find representation within it (Jungherr, Schroeder, et al., 2019) while weakening the power of institutional news media to decide which types of information, actors, and topics can gain access it.

The digital transformation has also added new kinds of structures as hosts of the public arena. These include digital platforms such as Facebook, TikTok, Twitter, and YouTube—algorithmically shaped environments that allow different actors to publish information and find audiences, and people to find information; as well as services that provide people with opportunities to publish information themselves, such as blogs and online discussion forums. This adds to the diversity of information, opinions, and voices. This makes it important to understand these environments' governance processes, usage patterns, and effects on information flows, audience behavior, and attitudes (Jungherr et al., 2020).

The companies that provide these new structures are not primarily in the news business; they have no institutional commitment to the public arena, and even less so to the normative demands of the public sphere. They seek user retention and advertising display. Nevertheless, they are powerful interlocutors between established news media, political elites, and publics (Nielsen & Ganter, 2022) and hold considerable sway within the public arena, without many of the regulatory burdens traditional media companies face.

This raises concerns regarding the legitimacy of power within the normative framework of the public sphere. Here, the power media structures exercise over information flows is seen as legitimate as long as they act distinct from other social subsystems—such as politics or business—and are self-regulating according to specific norms (Habermas, 2006, pp. 418–420). This, of course, is not true for these new digital structures, raising non-trivial concerns and foregrounding the conflict between commercial interests and the public arena's functions—the alignment of which societies need to address. It may not be possible to resolve the associated tensions completely, but they must be surfaced and negotiated in public.

The growing importance of AI has further extended the type of companies that now possess power within the public arena (Simon, 2022). While the theoretical breakthroughs in the current wave of AI began at universities, it is firms that lead in their practical application, further development, and broad rollout. This includes platform companies such as Facebook and TikTok that have added AI development to their portfolios, as well as dedicated AI development houses such as OpenAI and dedicated AI service providers (Ahmed et al., 2023; Metz, 2021). Again, these companies have no commitment to the normative expectations of the public sphere, and are notoriously opaque to outsiders, which makes it difficult to critically interrogate their contributions to and effects on the public arena—thus raising new regulatory and civil society oversight issues. Their growing power over information environments can be challenged along the same normative lines as digital infrastructures in general (Habermas, 2006): they belong to different functional subsystems of society, are governed according to commercial logics, and do not adhere to self-regulation according to the norms of news as an institution.

The adoption of AI by companies providing structures hosting the public arena, and their use of AI-enabled tools, extends from commercial structures, such as digital platforms and commercial news media, to public structures, such as public broadcasters, as well as not-for-profit structures. This is driven by functional reasons. Engaging in digital communication spaces requires completing tasks such as content recommendation and moderation efficiently and at scale, which AI allows. Structures lose audiences or drown in maintenance tasks if they cannot. The logic of AI-enabled workings will therefore permeate different types of structures in the public arena with commitments to different norms and production logics. The technological logic of AI, in turn, will shape the workings of actors following commercial as well as journalism norms and production logics.

Some see these structures of the contemporary public arena as conflicting with the ideals of normatively demanding concepts of the public sphere and its contribution to political

legitimacy through discourse, by catering to peoples' psychological weaknesses, enabling commercial domination and exploitation, and providing opportunities for manipulation (Habermas, 2022). But this reading misconstrues much of the available empirical evidence and obscures the democratically enabling features of digital media, such as empowering politically marginalized groups, surfacing both legitimate and illegitimate grievances, and enabling the contestation of institutions in crisis or conflict (Jungherr & Schroeder, 2021). While it is easy to see current developments predominantly as signs of decline within the contemporary public arena, there are also encouraging signs of revitalization and empowerment.

Artificial intelligence in the public arena

AI features strongly in the public arena. It is deployed in structures to pursue various tasks in three large categories:

- Shaping information and behavior
- Generating content
- Communicating

Each presents specific opportunities and challenges for the public arena.

Shaping

Shaping information and behavior is perhaps the most pervasive use of AI in the public arena, while also the most hidden. AI shapes the information people see and what information they are allowed or incentivized to publish. AI thus clearly impacts the way structures hosting the public arena allow society to become visible to itself and provide spaces for mutual recognition, identity formation, and mobilization.

Many digital media structures—such as Facebook, Google, TikTok, Twitter, and YouTube—rely (at least in part) on data-driven recommender systems that determine which information to display (or suggest) (Narayanan, 2023). Details of these systems' implementation and prominence vary, but they share a similar logic: by examining past data, algorithms determine what content is likely to lead users to the targeted interaction.

By shifting from a predominantly social-graph and subscription model of information exposure to one strongly based on content features and interaction patterns, AI-based recommendations lessen the influence of individual sources and content creators (Narayanan, 2023). The current diversity of sources depends on the opportunity of small digital media sources and actors to establish themselves over time. Without these opportunities, sources will become concentrated: attention will focus on a few established media brands whose coverage will be interspersed in AI-curated information environments with algorithmically identified pieces of content from various sources, with attention and monetization opportunities for small- and medium-size information sources and individual content creators being limited. There will be less diversity of sources and voices within the public arena.

AI-based algorithmic systems are also used to identify and label potentially harmful or illegal content, based on characteristics of other content labeled as such. Such systems can be deployed as upload filters at time of publication or afterwards as moderation (Douek, 2021; Gorwa et al., 2020). This also affects the public arena. By training models on the characteristics of harmful, illegal, or offensive content, tech companies determine what type of information or form of speech gets access to the public arena. Violent imagery documenting institutional brutality, or consciously offensive speech by challengers of the status quo, could be algorithmically censored. Over time, political speech within the dominant digital structures of the public arena would tend to become homogenous and safe. Challenges would have to conform to accepted rules of expression or face being submerged by algorithms. This may seem promising given current fears about online hate speech and radicalization, but it would also threaten the voicing of legitimate societal or political challenges.

AI-enabled shaping also reinforces earlier trends of metric-based governance in news (Christin, 2020). It influences the behavior of information providers in the public arena by creating economic incentives for them to feature or ignore specific information. News organizations can use AI to reverse engineer the type of content likely to receive broad distribution in larger algorithmically curated information environments, or to determine content especially attractive or unattractive for adjoining ad displays (MacKenzie, 2022). Surfacing stories, issues, people, and concerns based on expectations of audience interest and commercial viability turns the inherent biases in AI-based, data-driven visibility and invisibility of specific societal groups (Barocas & Selbst, 2016) into a determining factor for their visibility in the public arena. Visibility of society to itself is, therefore, potentially subject to the same inequalities currently being discussed in the context of algorithmic fairness (Mitchell et al., 2021).

This will also likely impact the behavior of smaller news organizations and content creators. Rather than investing in developing audiences over time, they will focus on reverse engineering the content or interaction signals sought by algorithmic recommender systems and adjust their information offerings accordingly. The AI-shaped future of independent content creation might look much more like TikTok than Instagram or the blogosphere.

AI-based systems introduce a new kind of iron cage, incentivizing journalists, editors, platform engineers, and users to reverse engineer and adjust to algorithmically determined signals of relevance. Recommender systems thereby become a coordinating device, raising the visibility of information in the public arena that conforms to patterns based on content acknowledged in the past and submerging information deviating from acknowledged patterns or that conforms to prior patterns of deviance. Over time, discourse would include fewer central topics, voices, and positions, with non-conforming content pushed out of sight.

Data-driven decision systems are inherently conservative. The uncritical application of AI that looks to the past to find rules to shape the present and future risks perpetuating past injustices (Bender et al., 2021). This raises significant challenges for the functioning of the public arena, especially how society is made visible to itself and the provision of counter-public spaces. If outsiders and challengers to the status quo find themselves unrepresented in the public arena, they may reject it and turn to different means to influence society and gain power, thereby raising the likelihood of conflict. And because many structures use AI-driven tools that work in similar ways, the underlying shifts in how content is visible in the public arena can happen largely unobserved. Further, since no *true* distribution of content is available, it becomes difficult

to identify any shift or bias reliably. If AI-driven shaping and distortion of the public arena can happen unobserved, it eliminates the possibility of deliberation or reflection on the change.

Generating

Recent advances in generative AI have raised public awareness of the capabilities of automated content generation (Brown et al., 2020; Ramesh et al., 2022; Vaswani et al., 2017). Prominent examples include LLMs such as Google's BERT, Open AI's GPT-4, and Facebook's LLaMA, and models that translate text prompts into images, used in applications such as DALL-E, Midjourney, and Stable Diffusion. These advances have made generative models widely accessible and broadly demonstrate how difficult it is to tell AI-generated content apart from human-created content. Autonomously created content can provide legitimate information for the public arena, but AI can also flood the public arena with content that contributes to the deterioration of information quality.

Automatically generated content based on raw information or event data, such as articles based on data from sporting events or real-time data from the stock market (Diakopoulos, 2019), is less concerning than some other types of content. Such automated coverage can be understood as translating one type of information—event or numerical data—into another, more reader-friendly format. Having AI contribute content like this to the public arena seems largely unproblematic. Of greater concern is the expected negative impact of generative AI on information quality control within the public arena. AI-generated content will lead to an increase in false or misleading information in the public arena, either purposeful or accidental. Already, nefarious actors can deliberately generate large amounts of targeted false or misleading information quickly (Goldstein et al., 2023). Widespread uses of AI can also contribute unintentionally to the decline of information quality. Generative AI is not committed to truth in the content it creates, only plausibility; there is no guarantee of factual correctness. In fact, a lot of AI-generated information will be false and misleading, while at the same time seeming perfectly plausible.

It is true that other AI-based systems can contribute to automated fact-checking, removal of poor-quality content, and the like, but it is doubtful that their existence will create a positive view of AI's contribution to information quality within the public arena. Instead, mutually competing and contesting adversarial AI-based systems labeling their respective contributions as false or misleading are more likely to create a sense of information insecurity and epistemic relativism regarding "facts" and a retreat to factionally aligned information intermediaries and AIs.

This situation raises broader structural challenges. Publics and elites rely on the public arena for information about society and the world at large. The legitimacy of structures depends on their ability to provide this information; as this deteriorates, so too do the functions of the public arena. The "flooding the zone" tactic is a good example: interested parties flood information environments with content designed to hide relevant information and hinder publics and elites from converging on a common diagnosis or agenda (Illing, 2020). AI-generated content will strengthen the impact of this tactic and potentially increase public frustration with the public arena's functioning.

There is also the prospect of AI imitating people (Natale, 2021, pp. 87–106) and their contributions in the form of automated opinion pieces or automated participation in comment threads or social media. To be sure, today's fear of bots (Rauchfleisch & Kaiser, 2020) and deepfakes is exaggerated; they are at best a nuisance at present. More advanced AI, though, could make disinformation a larger problem. Contributions from AI-based applications imitating people in digital communication environments—for example, automated harassment of people sharing political opinions or so-called "astroturfing" in digital communication environments to manipulate the "vox digitalis"—threaten a deterioration of the speech environment. If digital structures fail to adjust to this challenge, or are widely seen as failing, then the public arena will lose legitimacy as a space to learn about the world or for political identity building, formation of political programs, and mobilization. People will turn to other structures for these purposes.

These developments could also lead to a greater appreciation of intermediary institutions, such as media brands that feature professional journalism, exercise "gatekeeping" or strong editorial control, and enjoy the trust that comes with being seen by their audiences as authoritative. It is plausible that as the perceived decline of open information spaces increases, there will be greater support for controlled central media and information institutions. This, of course, holds only if these institutions are perceived as not subject to the errors of AI-driven information distribution or generation. So again, broad uses of AI may lead to a strengthening of central control and shared focus in the public arena, in contrast to the contemporary fears of fragmentation.

Communicating

Recent advances in LLMs have also foregrounded the role of AI as agent in digitally mediated communication. People interact with AI-based systems that are clearly identified as artificial agent, as with voice assistants such as Amazon's Alexa or Apple's Siri, or that are masquerading as humans, as with some chatbots in customer support or nefarious chatbots intended to manipulate people. This development, which has led to some rethinking of human-AI communication processes focused predominantly on the actor level (Esposito, 2022; Guzman & Lewis, 2020; Natale, 2021), has affected the structures of the public arena.

AI-enabled systems such as chatbots and voice assistants have become interfaces for people to access the public arena. People query these AI interfaces with questions and issue commands. Where the responses repeat or condense available information (Esposito, 2022), the structural impact of AI can be expected to remain limited—although some researchers have raised concerns regarding AI's anthropomorphized imitation of humans (Natale, 2021, pp. 107–126). The advent of LLM-enabled chatbots such as ChatGPT has significantly extended these capabilities. LLMs can mimic authoritative sources and voices, generating responses based on patterns identified in training data or data found online. But AI has no commitment to the truth of an argument or observation; it is only imitating their likeness. As Smith (2019) has argued, AI today remains committed only to the representation of the world, an object, or an argument available to it, not to the world, object, or argument as such. Using LLM-enabled chatbots thus risks generating false or misleading answers to prompts that nevertheless seem plausible.

Various services use LLM-enabled interfaces to produce responses to user prompts. Examples relevant to the public arena include Microsoft's Bing search and Google's Bard. Whereas querying a search engine returns a list of links to topically relevant sources, LLM-enabled search returns answers in text form, based on the underlying model or information found in sources identified as relevant—meaning people are guided to AI-based synopses and accounts, not directly to sources and actors within the public arena. Beyond the risk of generating factually incorrect answers, this has at least two important structural consequences: AI-based synopses of topics, accounts, and concerns are subject to the mechanisms of a data-driven pull toward the mean, which threatens to weaken idiosyncrasies and specific cultural signals within the public arena; and by providing synopses in lieu of links to media sites, AI-enabled search interfaces monopolize attention rather than distributing it to structures in the public arena that produce and invest in information production. Denying them monetization opportunities will weaken the economic foundation of these structures further. This is further complicated by many different types of actors and web services providing dedicated plugins for GPT-4, marking an attempt by Open AI to establish a platform position for AI-based services akin to what the App Store did for Apple's in the mobile internet. This extends these concerns well beyond search.

The role of AI as communicating agent in and access point to the public arena shifts the balance of power further away from traditional news media toward new mediating structures in the public arena such as search engines. We can also expect AI-enabled filtering of content within the public arena through LLMs to weaken challenges to the status quo and strengthen positions and topics that revert to the societal mean.

Artificial intelligence and the assessability of the public arena

The workings of the public arena as an "intermediary institution" between people and elites depends on its assessability (Müller, 2021, p. 139 f.)—the ability of outsiders to assess the workings and effects of structures hosting the public arena. Philosopher Onora O'Neill identifies assessability as a crucial feature of digital communication: "What originators seek to communicate (. . .) must be assessable (. . .) in ways that support understanding and interpretation, and enable forms of check and challenge" (O'Neill, 2022, p. 3). The use of AI challenges this public assessability.

Transparency—the duty of structures to document their inner workings to outsiders—is an important element for assessability, but only one of many. The workings of structures, their adherence to norms, and their effects on audiences need to be open to interrogation and contestation through specialist observers such as academics as well as by elites and the broader public. For meaningful assessability of digital structures, we need society-wide skillsets that allow regulators, journalists, academics, and the public to ask the right questions. Without such assessability, the legitimacy of the public arena, its structures, and its mediation of discourse will be contested, and a crucial coordinating feature in societies will deteriorate in its functioning.

AI and its uses today, however, contribute to opacity rather than assessability. At a very basic level, there is much unknown about many of the actual uses of AI, its inner workings, and effects in society and for the public arena. This is exacerbated by the plethora of AI types and models. While current debate focuses on LLMs, such as GPT-4 and applications built on it, there will soon be many different AIs developed by different companies and dedicated to specific uses. We need to be able to navigate this arena of unknowns and unknowables. For example, while we know that AI is used to shape the public arena, the exact extent and kinds of uses are largely unknown. More generally, absent a clear understanding of the workings of any given AI, its output can appear to be the result of machine rationality and unassailable. This is especially troubling because AI's failures have received little attention (Raji et al., 2022), making them unknowns, and so public debate is dominated by apparent AI success stories rather than a sober assessment of potential and evaluations of both successes and failures. This will render the functioning of the structures of the public arena uncertain and, over time, help deteriorate their legitimacy.

Somewhat more difficult is the assessability of effects of AI within the public arena. For example, AI-supported shapings of audiences, information flows, and agendas can have positive or negative consequences for the public arena's functioning. At the system level, individual AI-driven shapings do not necessarily skew the availability of content in one direction or another. But the aggregated and coordinated way in which this new technological mediation shifts content imperceptibly over time can negatively affect the functioning of the public arena. It is also possible that AI improves the functioning of the public arena by making certain actors or content more visible. These effects are even more difficult to document because there is no objective list of issues, actors, and audience compositions to compare to those shaped by AI—making biases more challenging to identify. This area requires consistent monitoring, public debate, and, if necessary, contestation to ensure assessability of the public arena.

But the challenge of assessability goes even deeper. In its current state, much AI—especially that which depends on deep learning—comes with inherent opaqueness, and it is often unclear how it achieves its success. We know AI learns the connection between available signals and predefined outputs in available data, but what signals does it use to do so? Potential errors abound. For one, AI can pick up on spurious signals that were correlated with outcomes of interest in training data but not causally connected, leading AI to fail once deployed or fall victim to targeted attacks (Szegedy et al., 2014). Looking from the outside, it remains unclear what AI has learned to predict outcomes, so it also remains unclear whether an AI-pursued outcome actually aligns with the goals of the actors deploying it (Christian, 2020). For example, translating an underlying goal into a metric and having AI optimize for that metric could create unintended consequences, as metrics and actual goals can deviate from each other (Hand, 2016, p. 17). By optimizing for the metric, AI might actually counteract the pursued goal. These sources of errors have given rise to the call for explainable AI (Gunning et al., 2019), but it remains unclear whether and how potential sources of error can be satisfactorily addressed.

Further, the data AI uses can be biased. Many studies have shown that inherent gender, racial, and other biases lead AI to reproduce societal inequalities and injustices (Buolamwini & Gebru, 2018; Caliskan et al., 2017). This emphasizes the need to interrogate the inputs and outputs of AI-based

systems and assess the fairness of the associated results (Mitchell et al., 2021).

Finally, companies rolling out AI often bias automated learnings through sensible safety interventions that try to block potentially harmful uses of AI. This includes filtering potentially harmful content or filtered responses to prompts aimed at creating harmful reactions, such as having a LLM provide instructions for a terrorist attack or trying to get a chatbot to respond with racist speech. These interventions, which are often opaque to the public and even professional observers, have unknown effects on the public arena and structural exclusion of challengers. Though often well-intentioned, these interventions can turn into de-facto control of political speech on the part of AI providers and platform companies by defining the confines of legitimate and illegitimate speech and contestation without democratic oversight or accountability.

Academic research, while urgently needed, cannot alone solve these problems. Actors running structures of the public arena, AI developers, and those who provide AI-based services must enable meaningful assessability of the uses and effects of AI-based applications.

We should, though, not overestimate these challenges. While much about AI's workings and effects may be unknown, they are not necessarily unknowable. Many current challenges might only be engineering problems persisting in AI's early developmental stage. Once AI models, data, applications, and effects are better understood, assessability-enabled critical interrogation will allow for development of engineering fixes and of social processes and practices structuring their use and that of the applications they enable.

Still, today's combination of unknowns and unknowables introduced to the public arena through AI reduces its assessability. This is especially troubling with respect to determining whether AI conforms or conflicts with the functioning of media infrastructures as distinct social subsystems, and their self-regulation along specific norms (Habermas, 2006, pp. 418–420). People, elites, and even actors running or contributing to the public arena's structures become uncertain about its functioning, and may lose their faith in its fairness. If AI imperceptibly reshapes the diversity, inclusiveness, objectivity, and watchdog role of the public arena—in short, its function of making society visible to itself—is the public arena losing some of that proper functioning? Absent assessability, this can degenerate into factions and elites blaming the public arena for being misrepresentative or being kept invisible. Over time, the public arena would lose its legitimacy, and its societal function would deteriorate. The imperative is thus that both those running structures hosting the public arena and its participants must work to ensure the assessability of the AI-enabled public arena and, wherever possible, reduce the unknowns and make, whenever possible, apparent unknowables known. Strengthening, or in many cases establishing, the capabilities and competencies of intermediary institutions such as news media and academia to assess the specific uses, workings, and effects of AI in the public arena is required.

Making artificial intelligence work for the public arena

Even in its current *narrow* manifestation, AI features in the functioning of the media structures hosting the public arena. This makes AI, its workings, and its effects important topics within communication studies. There is growing interest in this question, with many researchers addressing specific questions and applications. But we still need a theoretical synthesis focused on AI's structural impact within the public arena. The many unknowns, AI's implementation in media infrastructures, its workings, and its effects on information flows and user behavior make this no easy task.

We have used the societal functions of the public arena as a frame to sketch central tensions in AI's role and impact. Today's public debate is dominated by AI's perceived risks in society, whether through the mechanical reinforcement of societal biases, algorithmic cages that skew information flows in digital communication environments, or the increased and opaque power of having only a few companies providing AI-based applications and services. Our structural lens introduces a set of additional concerns.

We expect greater use of AI in the public arena to lead to a strengthening of control. This can happen directly by AI applications submerging new, challenging, or offending voices. It can also happen indirectly, through greater demand for intermediary structures providing vetted and gatekept information as a counterweight to open communication spaces flooded with unreliable or deliberately misleading AI-generated content. Demonetizing smaller media brands and sources by offering access to the public arena through communicating agents is another indirect path to the strengthening of central institutions or media brands. Through these developments, an AI-reliant public arena will largely shift away from the current state—a predominantly open, noisy, and sometimes offensive web—toward structures allowing for greater control over safe and vetted spaces. This will further empower different types of gatekeepers, weaken challengers of the status quo, and reinforce the status quo and established power relations. These are dangers that need to be monitored.

Still, we should not forget AI's potential benefits in improving the public arena in its functions for society. This holds both for how AI supports structures of the public arena in making society visible to itself and in the shaping of common and counterpublic spaces that allow people to find themselves and coordinate in the pursuit and contestation of the public good. As Esposito (2022) argues in a different context, perhaps AI's problem lies not in its general potential for bias, but rather that it may not be biased enough toward human goals—or in this case, the functionings of the public arena. Perhaps makers, shapers, and users of AI should assert greater control by deliberately biasing AI in this direction. While this might seem easy enough, it would require that society first agree about beneficial or detrimental biases in the workings of AI. That would make this an inherently social and political challenge, not a primarily technological one.

For that to happen, AI must be assessable. Only by enabling people, participants, and professional observers of the public arena to assess AI and its relationship to content and structures will AI contribute to a broad societal empowerment of the public arena. If society, providers of media infrastructures, and purveyors of AI settle for opaque solutions and applications, conversely, we can expect a deterioration of the public arena in its function for society. For AI to contribute to a more vibrant, inclusive, and empowering public arena, it must be made knowable and known, and so reestablish confidence in its well-functioning contribution.

AI's role and effects in the public arena are not preordained. Perhaps AI itself can be used to improve its own functions. Perhaps unobserved or uncritically deployed AI contributes to a deterioration of the public arena and its subsequent loss of legitimacy. Perhaps AI-based applications turn out to fail in their utility to the structures hosting the public arena and are dropped. Whatever the eventual outcome, academics must engage with AI's uses and effects in the public arena on the structural level by adapting or developing concepts and measurements that allow society to reflect and improve on those uses.

Acknowledgments

We thank Scott Cooper, Pascal Jürgens, Oliver Posegga, Adrian Rauchfleisch, Felix Simon, two anonymous reviewers, and the editors of the special issue for their valuable feedback.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research underlying this article by Andreas Jungherr has been generously supported by the VolkswagenStiftung.

Conflicts of interest: The authors declare no potential conflicts of interest or competing interests with respect to the research, authorship, and/or publication of this article.

References

Ahmed, N., Wahed, M., & Thompson, N. C. (2023). The growing influence of industry in AI research. *Science (New York, N.Y.)*, *379*(6635), 884–886. https://doi.org/10.1126/science.ade24

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Arendt, H. ([1967] 1968). Truth and politics. In *Between past and future: Eight exercises in political thought* (pp. 227–264). Viking Press.

Asenbaum, H. (2022). Rethinking democratic innovations: A look through the kaleidoscope of democratic theory. *Political Studies Review*, *20*(4), 680–690. https://doi.org/10.1177/14789299211052890

Bandy, J., & Diakopoulos, N. (2021). More accounts, fewer links: How algorithmic curation impacts media exposure in Twitter timelines. In *Proceedings of the ACM on human-computer interaction* (Vol. *5*, pp. 1–28). ACM. https://doi.org/10.1145/3449152

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, *104*, 671–732. https://doi.org/10.15779/Z38BG31

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *FAccT '21: Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Bennett, W. L. (1990). Towards a theory of press-state relations in the US. *Journal of Communication*, *40*(2), 103–127. https://doi.org/10.1111/j.1460-2466.1990.tb02265.x

Bourdieu, P. (1990). Social space and symbolic power. In Adamson M., (Ed.), *In other words: Essays towards a reflexive sociology*. (pp. 123–139). Stanford University Press.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., … Amodei, D. (2020). Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems* (Vol. *33*, pp. 1877–1901). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In S. A. Friedler & C. Wilson (Eds.), *Proceedings of the 1st conference on fairness, accountability and transparency* (Vol. *81*, pp. 77–91). Proceedings of Machine Learning Research (PMLR). https://proceedings.mlr.press/v81/buolamwini18a.html

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science (New York, N.Y.)*, *356*(6334), 183–186. https://doi.org/10.1126/science.aal4230

Castells, M. ([2009] 2013). *Communication power* (2nd ed.). Oxford University Press.

Christian, B. (2020). *The alignment problem: Machine learning and human values*. W. W. Norton & Company.

Christin, A. (2020). *Metrics at work: Journalism and the contested meaning of algorithms*. Princeton University Press.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2022). *Introduction to algorithms* (4th ed.). The MIT Press.

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.

Douek, E. (2021). Governing online speech: From "posts-as-trumps" to proportionality and probability. *Columbia Law Review*, *121*(3), 759–834. https://doi.org/10.2139/ssrn.3679607

Elahi, M., Jannach, D., Skjærven, L., Knudsen, E., Sjøvaag, H., Tolonen, K., … Trattner, C. (2022). Towards responsible media recommendation. *AI and Ethics*, *2*(1), 103–114. https://doi.org/10.1007/s43681-021-00107-7

Esposito, E. (2022). *Artificial communication: How algorithms produce social intelligence*. The MIT Press. https://doi.org/10.7551/mitpress/14189.001.0001

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

Fraser, N. (1990). Rethinking the public sphere: A contribution to the critique of actually existing democracy. *Social Text* (25/26), 56–80. https://doi.org/10.2307/466240

Gerhards, J., & Neidhardt, F. (1991). Strukturen und Funktionen moderner Öffentlichkeit: Fragestellungen und Ansätze. In S. Müller-Doohm & K. Neumann-Braun (Eds.), *Öffentlichkeit, Kultur, Massenkommunikation: Beiträge zur Medien- und Kommunikationssoziologie* (pp. 31–90). Bibliotheks- und Informationssystem der Universität Oldenburg.

Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M., & Sedova, K. (2023). Generative language models and automated influence operations: Emerging threats and potential mitigations. *arXiv*. https://doi.org/10.48550/arXiv.2301.04246

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, *7*(1), 1–15. https://doi.org/10.1177/2053951719897945

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI – explainable artificial intelligence. *Science Robotics*, *4*(37), eaay7120. https://doi.org/10.1126/scirobotics.aay712

Gurevitch, M., & Blumler, J. G. (1990). Political communication systems and democratic values. In J. Lichtenberg (Ed.), *Democracy and the mass media: A collection of essays* (pp. 269–289). Cambridge University Press. https://doi.org/10.1017/CBO9781139172271.011

Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A human–machine communication research agenda. *New Media & Society*, *22*(1), 70–86. https://doi.org/10.1177/1461444819858691

Habermas, J. ([1962] 1990). *Strukturwandel der Öffentlichkeit: Untersuchungen zu einer Kategorie der bürgerlichen Gesellschaft*. Suhrkamp.

Habermas, J. (1992). *Faktizität und Geltung: Beiträge zur Diskurstheorie des Rechts und des demokratischen Rechtsstaats*. 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

Habermas, J. (2022). *Ein neuer Strukturwandel der Öffentlichkeit und die deliberative Politik*. Suhrkamp.

Hand, D. J. (2016). *Measurement: A very short introduction*. Oxford University Press.

Illing, S. (2020). "Flood the zone with shit": How misinformation overwhelmed our democracy. *Vox*. https://www.vox.com/policy-and-politics/2020/1/16/20991816/impeachment-trial-trump-bannon-misinformation

Jungherr, A. (2023). Artificial intelligence and democracy: Workings and areas of contact. *Working Paper*.

Jungherr, A., Posegga, O., & An, J. (2019). Discursive power in contemporary media systems: A comparative framework. *The International Journal of Press/Politics*, *24*(4), 404–425. https://doi.org/10.1177/1940161219841543

Jungherr, A., Rivero, G., & Gayo-Avello, D. (2020). *Retooling politics: How digital media are shaping democracy*. Cambridge University Press. https://doi.org/10.1017/9781108297820

Jungherr, A., & Schroeder, R. (2021). Disinformation and the structural transformations of the public arena: Addressing the actual challenges to democracy. *Social Media + Society*, *7*(1), 1–13. https://doi.org/10.1177/2056305121988928

Jungherr, A., & Schroeder, R. (2022). *Digital transformations of the public arena*. Cambridge University Press. https://doi.org/10.1017/9781009064484

Jungherr, A., Schroeder, R., & Stier, S. (2019). Digital media and the surge of political outsiders: Explaining the success of political challengers in the United States, Germany, and China. *Social Media + Society*, *5*(3), 1–12. https://doi.org/10.1177/2056305119875439

Kelleher, J. D. (2019). *Deep learning*. The MIT Press.

Kovach, B., & Rosenstiel, T. (2021). *The elements of journalism: What newspeople should know and the public should expect* (4th ed.). The Crown Publishing Group.

Larson, E. J. (2021). *The myth of artificial intelligence: Why computers can't think the way we do*. The Belknap Press of Harvard University.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science (New York, N.Y.)*, *343*(6176), 1203–1205. https://doi.org/10.1126/science.1248506

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Luhmann, N. ([1995] 2017). *Die Realität der Massenmedien* (5th ed.). Springer VS. https://doi.org/10.1007/978-3-658-17738-6

MacKenzie, D. (2022). Blink, bid, buy. *London Review of Books*, *44*(9). https://www.lrb.co.uk/the-paper/v44/n09/donald-mackenzie/blink-bid-buy

Metz, C. (2021). *Genius makers: The mavericks who brought AI to Google, Facebook, and the world*. Dutton.

Mitchell, M. (2019). *Artificial intelligence: A guide for thinking humans*. Giroux.

Mitchell, S., Potash, E., Barocas, S., D'Amour, A., & Lum, K. (2021). Algorithmic fairness: Choices, assumptions, and definitions. *Annual review of statistics and its application*, *8*(1), 141–163. https://doi.org/10.1146/annurev-statistics-042720-125902

Müller, J.-W. (2021). *Democracy rules*. Allen Lane.

Napoli, P. M. (2014). Automated media: An institutional theory perspective on algorithmic media production and consumption. *Communication Theory*, *24*(3), 340–360. https://doi.org/10.1111/comt.12039

Narayanan, A. (2023). Understanding social media recommendation algorithms. *Knight First Amendment Institute at Columbia University*. https://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms

Natale, S. (2021). *Deceitful media: Artificial intelligence and social life after the Turing Test*. Oxford University Press. https://doi.org/10.1093/oso/9780190080365.001.0001

Nielsen, R. K., & Ganter, S. A. (2022). *The power of platforms: Shaping media and society*. Oxford University Press. https://doi.org/10.1093/oso/9780190908850.001.0001

O'Neill, O. (2022). *A philosopher looks at digital communication*. Cambridge University Press. https://doi.org/10.1017/9781108981583

Peters, B. ([1994] 2007). Der Sinn von Öffentlichkeit. In H. Wessler (Ed.), *Der Sinn von Öffentlichkeit*. (pp. 55–102). Suhrkamp.

Peters, B. (2007). *Der Sinn von Öffentlichkeit*. (H. Wessler, Ed.). Suhrkamp.

Raji, I. D., Kumar, I. E., Horowitz, A., & Selbst, A. (2022). The fallacy of AI functionality. In C. Isbell, S. Lazar, A. Oh, & A. Xiang (Eds.), *FAccT '22: 2022 ACM conference on fairness, accountability, and transparency* (pp. 959–972). ACM. https://doi.org/10.1145/3531146.3533158

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with CLIP latents. *arXiv*. https://doi.org/10.48550/arXiv.2204.06125

Rauchfleisch, A., & Kaiser, J. (2020). The false positive problem of automatic bot detection in social science research. *PloS One*, *15*(10), e0241045. https://doi.org/10.1371/journal.pone.0241045

Rauchfleisch, A., & Kovic, M. (2016). The internet and generalized functions of the public sphere: Transformative potentials from a comparative perspective. *Social Media + Society*, *2*(2), 1–15. https://doi.org/10.1177/2056305116646393

Russell, S., & Norvig, P. ([1995] 2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson Education.

Schäfer, M. S., & Wessler, H. (2020). Öffentliche Kommunikation in Zeiten künstlicher Intelligenz. *Publizistik*, *65*(3), 307–331. https://doi.org/10.1007/s11616-020-00592-6

Schultz, J. (1998). *Reviving the fourth estate: Democracy, accountability and the media*. Cambridge University Press. https://doi.org/10.1017/CBO9780511597138

Simon, F. M. (2022). Uneasy bedfellows: AI in the news, platform companies and the issue of journalistic autonomy. *Digital Journalism*, *10*(10), 1832–1854. https://doi.org/10.1080/21670811.2022.2063150

Smith, B. C. (2019). *The promise of artificial intelligence: Reckoning and judgment*. The MIT Press.

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2014). Intriguing properties of neural networks. *International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.1312.6199

Taylor, C. (1995). Liberal politics and the public sphere. In C. Taylor (Ed.), *Philosophical arguments*. (pp. 257–288). Harvard University Press.

Trielli, D., & Diakopoulos, N. (2019). Search as news curator: The role of Google in shaping attention to news information. In S. Brewster, G. Fitzpatrick, A. Cox, & V. Kostakos (Eds.), *CHI '19: Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–15). ACM. https://doi.org/10.1145/3290605.3300683

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. V. N. Vishwanathan, & R. Garnett (Eds.), *NIPS 2017: 31st conference on neural information processing systems*. (Vol. *30*, pp. 1–11). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Vela, D., Sharp, A., Zhang, R., Nguyen, T., Oleg, S., & Pianykh, A. H. A (2022). Temporal quality degradation in AI models. *Scientific Reports*, *12*(1), 1–12. https://doi.org/10.1038/s41598-022-15245-z

Warner, M. (1990). *Letters of the republic: Publication and the public sphere in eighteenth-century America*. Harvard University Press.

Warner, M. (2002). *Publics and counterpublics*. Zone Books.

Wolfram, S. (2023). What is ChatGPT doing . . . and why does it work? *Stephen Wolfram: Writings*. https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/