Generative AI and the Future of Democratic Citizenship

PAUL FORMOSA*

Philosophy, Macquarie University, Sydney, Australia, paul.formosa@mq.edu.au

BHANURAJ KASHYAP

Philosophy, Macquarie University, Sydney, Australia, bhanuraj.kashyap@mq.edu.au

SIAVOSH SAHEBI

Philosophy, Macquarie University, Sydney, Australia, siavosh.sahebi@hdr.mq.edu.au

Generative AI technologies have the potential to be socially and politically transformative. In this paper, we focus on exploring the potential impacts that Generative AI could have on the functioning of our democracies and the nature of citizenship. We do so by drawing on accounts of deliberative democracy and the deliberative virtues associated with it, as well as the reciprocal impacts that social media and Generative AI will have on each other and the broader information landscape. Drawing on this background theory, we outline some of the key positive and negative impacts that Generative AI is likely to have on democratic citizenship. The political significance of these impacts suggests the need for further regulation.

CCS CONCEPTS • Applied Computing • Arts and Humanities

Additional Keywords and Phrases: Generative AI, citizenship, social media, Large Language Models (LLM), deliberative democracy

1 Introduction

Generative Artificial Intelligence (AI) refers to a subset of AI technologies that use "complex models" to produce "high-quality, human-like material" [28]. These models can receive and generate multimodal inputs and outputs such as text, images, and sound, and include Large Language Models (LLM) such as ChatGPT [22]. Generative AI can be used to moderate, paraphrase, edit, retrieve information [16], and create new content [28], and these use cases have the potential to be socially and politically transformative, which has profound implications for the ways that democracies function. This includes, for example, Generative AI influencing everything from the way that the executive branch provides services to citizens [9], the way the legislative branch crafts and reviews legislation [55], to the operation of court proceedings [51]. However, beyond the branches of government, the impacts of Generative AI on the citizens of democracies and their cognitive, social, and political virtues are crucial for understanding the broader societal transformation that these technologies could bring about. We shall explore this issue as follows. First, we begin by exploring how Generative AI will impact the information landscape and social media, as this is a key input into democratic citizenship. With this groundwork in place, we first outline some of the potential positive and then negative impacts of Generative AI on democratic citizens. While we consider possible future impacts of Generative AI on democratic citizenship, we base this on extrapolation from existing uses and trends. Our purpose in doing this is to alert different stakeholders, including citizens, regulators, and AI developers, to these potential outcomes so that we can make informed decisions now that might both limit any future harms as well as grasp the opportunities this transformative technology offers.

* Alphabetical order here indicates the equal contributions of the authors.

ACM 2691-199X/2024/05-ART
http://dx.doi.org/10.1145/3674844

2 Social Media and the Information landscape

An important normative assumption that grounds liberal democracies is the expectation that citizens will have access to facts and other relevant information about political affairs and public policies, and they will consider these in developing their political preferences [42]. This assumption animates the relationship between the citizenry and the information landscape they inhabit. By information landscape, we refer not only to traditional media ecosystems, but also to new forms of digital media, political groups, social movements, libraries, and everyday conversations that provide citizens opportunities to access facts and other relevant information about their social and political systems. If this information landscape has systematic flaws [7] or changes unexpectedly [2], this can impact the ability of citizens to participate effectively in democratic politics and culture. New Generative AI technologies, such as ChatGPT, Copilot, and DALL-E, will pose both challenges and create new opportunities within the information landscape by impacting traditional and digital media, and the citizens that consume that content. These changes range from shifts in journalistic practices to the proliferation of personalization tools tailoring outputs to each individual user's interests [3]. As Generative AI transforms our information landscapes, especially our social media platforms, it will also change our democracies. By social media platforms, we refer to digital intermediaries that "host, organize, and circulate users' shared content or social interactions for them" [29], often for commercial purposes.

Over the previous decade or so, social media platforms, such as Facebook and YouTube, have unrecognizably altered the information landscape [29, 2]. The key components of the traditional media, including news outlets, magazines, independent agencies, universities, and the different branches of the state, have all turned to social media to distribute and amplify their content. Social media platforms are, in turn, best considered as constituting and curating the digital public sphere that functions as an important component of the broader public sphere [19]. While theorists often focus on some of the negatives associated with social media [6, 44, 38], it is worth remembering that social media platforms have also contributed to positive outcomes, for example, in diversifying information networks [53] and in spreading social justice movements online [69]. Social media technology has also affected the virtues and vices citizens develop and practice online, which further impacts their participation in deliberative exchanges with one another [71]. It is thus clear that social media companies have had a profound effect on our democratic culture by transforming our information landscape [44].

It may be tempting to theorize how Generative AI, especially in the form of LLM, will change the information landscape independently of social media, such as when citizens use ChatGPT to bring themselves up to speed on a public policy. However, to consider Generative AI in isolation from social media in theorizing the unprecedented changes the information landscape will experience is a mistake, since this fails to consider the transformative potential of these two technologies *in tandem*. We need to consider the effects of LLM on the content social media users produce and consume, and in turn the content that social media companies censor, demote, and amplify, to get an accurate picture of the direction in which our democratic culture may be heading. Social media platforms contain an abundance of information and content, which can feel infinite and limitless, while there are clear limitations to the availability of the attention of users to consume it. Within the attention economy, our

attention is the resource of interest for those looking to control and exploit it for their own ends [18]. The important point to consider here is that the social media attention economy is limited – especially when compared to the vast array of information available to fill it – which means that whatever dominates the attention economy can have significant broader social and political implications [59]. If users overwhelmingly employ Generative AI to create and/or modify the content they share, and in turn consume, on social media, then Generative AI has the potential to be a key force in the attention economy.

It is a mistake to think of the digital world as a closed system, where online users use one digital technology to participate in another digital space, without impacting the offline world at the same time. Instead, we endorse a hybrid approach where the boundaries that separate the offline world from the digital world are impossible to draw and merge into one another at different points [44]. For example, a citizen may use Copilot to learn about a contemporary political event, then use this information to write a post which is edited by ChatGPT before it is shared on Facebook alongside a bespoke image created by DALL-E, which in turn sparks an in-person political conversation with a friend on the same topic who goes on to record a video about this posted on YouTube. The hybrid approach includes not only continuous back-and-forth interactions between LLM, social media platforms, and offline interactions, as the previous example illustrates, but it also includes instances where LLM speed up one task, thereby freeing up time for users to engage with other tasks. We draw on the hybrid approach in articulating below the potential positive and negative impacts of Generative AI on democracy, as this approach most accurately characterizes the contemporary entanglements of digital technologies with the interests of users [44], with many users considering their digital personas as an extension of their selves.

3 Positive impacts of generative Ai on Democratic Citizenship

To understand the potential positive impacts of Generative AI on citizenship, we first need a conception of citizenship. Democratic citizenship includes three elements: a legal status, an identity tied up with being part of a particular political community, and an active role involving citizens as political actors who can play a role in their democratic institutions through, for example, voting and discussing political issues with other citizens [15]. An influential way of conceptualizing this last element is through the idea of deliberative democracy. Deliberative democracy is based on the claim that the legitimacy of political institutions and practices depends on their public justifiability to all citizens as free and equal members of the political community, which in turn requires citizens to partake in deliberative practices that focus on the giving and receiving of reasons [31, 66]. Deliberative democracy therefore requires that citizens develop several distinct virtues. These include: being disposed to recognize each other as equals, to imaginatively engage with the viewpoints of others, and to be honest, sincere, and respectful in deliberations [31]; the virtues of deliberative speech, such as publicity, accountability, and reciprocity [30]; internal deliberative virtues, such as “humility and hope”, that complement the deliberative skills needed to reason and deliberate well [30]; and additional virtues that more privileged speakers may need to cultivate, such as “the virtue of facilitating equitable deliberative exchanges” between unequal parties [66]. As members of a

democracy, citizens should be both able and willing to engage in good faith deliberations about how their community governs itself.

Generative AI has strong implications for democratic citizenship and the development of relevant virtues and vices. We can think about these impacts by drawing on three pathways through which humans and AI interact. These are: one, AI *replacing* human tasks and skills through automation; two, AI *assisting* humans to complete tasks; and three, AI *amplifying* or *augmenting* the skills and capacities of humans [10, 20]. These latter two categories include what Zuboff [1989] calls the "informating" powers of technology which gives humans access to useful information in helpful formats that can guide deliberation and action [36].

Drawing on each of these three pathways, we can identify potential positive impacts that Generative AI might have on democratic citizenship in the near term. In terms of *replacement* through automation, AI promises to relieve us of the "drudgery" of completing mundane and low value tasks [25], such as composing routine emails or simple social media posts, which could free up the time and attentional resources of citizens to focus on higher value politically relevant activities, such as informed deliberations with fellow citizens. Generative AI could also be used to replace in-person and small-scale human facilitated discussions and political meetings (for a non-AI version of such a chatbot, see Kim et al. [2021]; for a general discussion of LLM-based Chatbots, see Kumar et al. [2024]), with online and large-scale AI facilitated interactions, which could break down geographical barriers and allow for greater political engagement. Drawing on our discussion of the hybrid approach to understanding Generative AI impacts outlined above, we can imagine that these exchanges could take place in a broader information landscape in which AI can function as both a participant and moderator in online discussions that merge seamlessly with offline interactions and other political activities, such as voting and opinion formation.

This in turn links to the informating potential of Generative AI across both *assisting* and *augmenting* pathways. Generative AI has the potential to both assist and augment our information accessing and processing powers. LLM, for example, could provide quicker, better targeted, and better curated information that could inform deliberations on a range of topics [24]. As well as better informing citizens, the interactivity of LLM has the potential to cultivate the argumentative and deliberative capacities of citizens. LLM could be used, for example, to provide both sides of an argument, probe for weaknesses in our own arguments, and help us to better understand alternative viewpoints [24, 56]. There is already evidence emerging about the potential effectiveness of non-AI chatbots in this regard. For example, one study showed that facilitation of online deliberative discussion by a chatbot led to better deliberation, suggesting that a "chatbot agent could partially substitute for a human moderator in deliberative discussion" [40]. The greater interactivity of LLM-based Chatbots has the potential to be even more effective in this regard. For example, a more recent study used a Generative AI tool based on the GPT-3 LLM to demonstrate that real-time AI interventions in online deliberations around controversial political topics, such as gun control, could improve the reported conversation quality, as well as promote democratic reciprocity and improve the tone of deliberations [4]. These interventions can help deliberators to hear the other side of a debate, which is an important part of improving deliberations [52], and increase politeness [4], which can help to prevent some of the polarization that can otherwise occur on social media and other online platforms.

Generative AI has the potential to help address two issues with deliberative democracy: argumentative inequality (i.e., citizens are unequal in their capacity to deliberate well) and scalability (i.e., participatory deliberative interactions are hard to scale up). Argumentative inequality is a problem as deliberative democracy assumes that all citizens are free and equal in their capacity to engage in deliberation, but of course in practice the ability of all citizens to reason well is far from equal [8]. By helping those with less advanced communication, reasoning, information retrieval, and argumentation skills to better inform and express themselves politically through the assistance of LLM, Generative AI has the potential to help partly alleviate this problem. This could empower citizens by helping to reduce deliberative inequalities between them. Further, the potential of Generative AI to provide accurate information, promote high-quality deliberation, and engage in dialogue as a critical interlocuter could help to address the scalability problems associated with deliberative democracy. Ackerman and Fishkin [2002] imagine a Deliberation Day, which is a public holiday devoted to all citizens engaging in curated in-person public deliberation with expert information and moderation. But as a practical solution to the scalability problem, this has proven impractical as it is hard to muster the physical and human resources needed to facilitate widespread high-quality deliberative engagements (see the discussion of scalability in Bua [2017]). Deliberative polls and citizen juries [66] are typically limited to small groups as their resource intensity means that rolling them out to a mass audience has proven to be infeasible. The scalability of Generative AI (i.e., its ability to be widely deployed and used by citizens and accessed through existing personal devices, such as mobile phones or laptops) means that such impracticalities could be overcome with technology. Generative AI could be deployed widely to take on the dual role of informing (including by using Retrieval-Augmented Generation (RAG) to access real-time information not in its training dataset – see Chen et al. [2023]) and engaging in reasoned debate with all citizens (including by using "fine-tuning" to create an LLM specialized for this task – see Jeong [2024] for discussion of fine-tuning), as well as skillfully facilitating discussion between citizens by using real-time interventions to promote civility, politeness, and empathy. This could potentially allow for a widespread improvement in the ability of citizens to be both informed and to deliberate well on a range of important political topics.

One of the promises of deliberative democracy is that more deliberative and virtuous citizens will raise the quality, and ultimately the fairness, of political institutions. Insofar as Generative AI could also play a role in improving deliberativeness among citizens, it could help to raise the quality of the broader information landscape that citizens will both consume and produce via social media. But these positive outcomes are certainly not guaranteed, or even highly likely given the tendency of social media to promote polarization and misinformation [59] rather than consensus and informed deliberation, and any benefits that do occur may accrue for only a small subset of citizens given the "technological divide" that exists between different groups in society [54]. This leads to the question of what the potential negative impacts of Generative AI on democratic citizenship are.

4 Negative Impacts of Generative Ai On Democratic Citizenship

When considering the negative impacts of Generative AI within the context of the social media attention economy, we will look at three of the most serious risks associated with it: worse quality or incorrect information due to hallucinations, biased or incorrect training data, and deepfakes and the

weaponization of mass disinformation. We will explore these risks before assessing their influence on democratic citizenship and deliberative democracy.

Generative AI's current propensity to produce low quality, biased, or incorrect information due to hallucinations and biased or incorrect training data is a key concern, especially when considering its use for political deliberation. Biased information outputs are the result of the training data on which the model is developed, while hallucinations are outputs which are "nonsensical or unfaithful to the given source input" [27]. As Stokel-Walker & Van Noorden [2023] note, "ChatGPT and its competitors work by learning the statistical patterns of language in enormous databases of online text — including any untruths, biases or outmoded knowledge". This is not limited to text-based models either, as Cheong et al. [2023] report gender and racial biases in DALL-E. If democratic citizens rely on the outputs of Generative AI models for their political deliberations and decision making, these biases and inaccuracies could be materially consequential. For example, as Kreps & Kriner [2023] argue, the ability of Generative AI to produce outputs at a significant rate and scale could lead to it flooding the information landscape with inaccurate content, which may threaten voters' efforts to deliberate on what elected representatives do. Furthermore, Wach et al. [2023] argue that as Generative AI models develop the capabilities to produce outputs which appear human-like, there is a heightened risk that this content will be relied upon and used for deliberation without critical evaluation or verification of the information.

Deepfakes, which are videos generated by AI depicting people doing or saying something that they did not actually do or say, pose another problem for our information landscape and democracies. Deepfakes are intended to deceive the viewer, and as technology develops further the ability to successfully deceive will only increase. We already have examples of deepfakes of politicians circulating on social media [46]. As Matthews [2022] argues, deepfakes are epistemically harmful due to the undermining of the information being communicated, while also resulting in the erosion of trust in videos and other digital content. The problem for democratic deliberation is twofold here. First, democratic citizens may be deceived by incorrect information that is communicated to them by what they perceive and expect to be a trusted source. Second, they may lose trust in the information they are consuming regardless of its accuracy and develop a mistrust of their fellow citizens who are posting content.

Another negative impact to consider is the use of Generative AI to propagate and weaponize mass disinformation. As Whyte [2020] states, "new abilities to produce even reasonable fidelity fabrications rapidly and at scale combine the multiform shape of the modern digital information environment to make organized influence efforts much more dynamic than has previously been the case". A recent study by Menz et al. [2023] found that by using a publicly available LLM they were able to generate 102 different blog articles containing over 17,000 words of targeted health disinformation regarding vaccines and vaping, including fake patient and clinician testimonials, in just 65 minutes. Those who are well resourced, whether they be competitor states, corporations, or wealthy individuals, could weaponize this ability to flood the information landscape and dominate the attention economy. For example, NewsGuard, a company which tracks and counters online misinformation, fed ChatGPT 100 false narratives from their database and found that it was able to generate false news articles, essays, and TV scripts on politically significant and sensitive topics for 80% of the identified false narratives

[12]. Additionally, Simchon et al.'s [2024] studies highlight the threats posed to democratic decision-making from Generative AI's ability to produce personalized political advertisings at large scales, which can then be used to micro-target voters. We have already seen the use of deepfakes by political parties in India [63, 39], Pakistan [21], and Slovakia [48], to name a few examples. We have also seen Iran-backed hackers use deepfake news to target TV streaming services of countries to reach both domestic and international audiences [50]. This weaponization could be used either to promote false beliefs about controversial political issues or to intensify social division and distrust [47]. This could encourage citizens to be less willing to trust and engage in good faith and informed deliberations with one another, thereby worsening both the democratic culture and the information landscape.

We can further explore the possible negative impacts of Generative AI on democratic citizenship by drawing on the three pathways outlined in the previous section. In terms of replacing human tasks and skills, Generative AI risks atrophying our deliberative skills, thereby making us less informed and less politically engaged. By relying on Generative AI to replace the research, understanding, and critical thinking that we would otherwise need to gain a deeper understanding of a particular issue, we risk a serious de-skilling of the critical deliberation capacities that are vital for a healthy democracy. While the impacts of Generative AI on deliberative skills are yet to be fully realized given that the technology is in its relative infancy, concerns regarding the impact of technology and AI on deskilling critical cognitive and moral capabilities are not new [69, 5, 57, 61]. Highlighting this concern, Pitt [2023] argues that developing the necessary skills to analyze information, form coherent arguments and persuade others of the validity of that argument, cannot be achieved by relying on Generative AI to perform such functions.

Another way in which Generative AI can make citizens worse deliberators is highlighted by Wang et al.'s [2023] study which found that the models they tested (ChatGPT and GPT-4) were prone to being misled by illegitimate arguments by the user, often agreeing with invalid arguments despite independently generating the correct answer in the first instance. They found that instead of providing responses which improved the quality of deliberations, the responses of these models were instead tailored to be preferable to the human user [73]. This signifies a further negative impact that Generative AI may have on citizens by undermining meaningful deliberation and posing risks where false beliefs are reinforced, and misinformation and disinformation are amplified.

The insertion of Generative AI into the information landscape means that it may replace many interpersonal interactions, resulting in citizens directly engaging with each other's viewpoints and perspectives less often. Additionally, Generative AI may lead to a decline in the overall trust between citizens as it becomes increasingly unclear who they are communicating with [35]. Given these potentialities, if citizens are relying on Generative AI to deliberate and assist with making decisions, then there is a real risk of engagement with fellow citizens being severely limited. This could harm citizens' sense of belonging to a common political community. AI powered chatbots are also becoming increasingly influential in online interactions, including in political discussions on social media [60, 34, 32, 33]. As we have discussed earlier, given the scale at which Generative AI can operate, the proliferation of chatbots and their outputs on social media, along with the increased difficulty in differentiating them from humans, means that AI powered chatbots could threaten to replace much of the politically important human interactions conducive to a healthy deliberative

democracy. Furthermore, it may become more difficult for citizens to engage in discourses and organize political actions with each other if it becomes harder for them to get noticed because they need to compete with a flood of Generative AI content that has been optimized to dominate the attention economy.

What the above risks around biased, inaccurate outputs, deepfakes and mass disinformation highlight is that we are facing the prospect of entering an "infopocalypse", which Fallis [2021] defines as a situation "where we cannot tell what is real from what is not". This means that citizens may find themselves unable to differentiate between information which is real or manipulated by Generative AI when they are deliberating on a given issue. In contrast to the potential benefits of replacing mundane and low value tasks by AI, we also risk having to double-check and cross-examine the information that Generative AI provides us, which may ultimately create more work for ourselves as deliberative citizens. This is highlighted by Buchanan et al. [2024] who found that both ChatGPT-3.5 and ChatGPT-4 provided citations which do not exist when asked to write about economic concepts, finding also that the reliability of the model decreased as the prompts became more specific. More importantly, we run the risk of having ill-informed citizens who are no longer engaging with one another and no longer reasoning well together on any meaningful level. The scalability of Generative AI exacerbates this risk, especially on social media, as we face the scenario where the attention economy is dominated and weaponized by deepfakes and mass disinformation campaigns. This could lead to what Rini [2021] refers to as "Weaponized Skepticism", which is the "calculated deployment of landmines across the epistemic commons" with the aim of saturating the "epistemic environment with conflicting accounts so that the truth appears to be only one of the many bickering narratives". The scalability of Generative AI means bad actors can utilize weaponized skepticism to propagate distrust, social division, and ultimately a deep detachment of citizens from both the democratic process and each other.

5 Conclusion

This paper has outlined some of the key positive and negative impacts that Generative AI could have on democratic citizenship. What is clear from this analysis is that there is much at stake for our democracy and citizenship as this technology develops. Given that the development of Generative AI is, at present, largely driven by profit seeking corporations who already have immense political influence [67], ensuring that we maximize the positive and minimize the negative impacts of this technology, while also ensuring accessibility and fairness for all citizens, is vitally important. To address these issues, we need both to ensure that citizens are informed and equipped with the capacities and trusted information needed to navigate this new horizon [64], and to proactively consider the role of regulation in limiting harms to the proper functioning of democracies. While regulation of social media platforms [11] and AI (see Wu & Liu [2023] for a summary of geographic-specific AI regulations) has already begun, it would be beneficial to bring the impacts on democratic citizenship outlined here into these conversations to help ensure that democracies can flourish as the impacts of Generative AI increase.

REFERENCES

[1] Bruce Ackerman and James S. Fishkin. 2002. Deliberation Day. *Journal of Political Philosophy* 10, 2 (2002), 129–152. https://doi.org/10.1111/1467-9760.00146

[2] Mark Andrejevic. 2020. *Automated Media*. Routledge, New York. https://doi.org/10.4324/9780429242595

[3] Amy Ross Arguedas and Felix M Simon. 2023. *Automating Democracy: Generative AI, Journalism, and the Future of Democracy.* Balliol Interdisciplinary Institute, University of Oxford.

[4] Lisa P. Argyle, Christopher A. Bail, Ethan C. Busby, Joshua R. Gubler, Thomas Howe, Christopher Rytting, Taylor Sorensen, and David Wingate. 2023. Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale. *Proc. Natl. Acad. Sci. U.S.A.* 120, 41 (October 2023), e2311627120. https://doi.org/10.1073/pnas.2311627120

[5] Vicky Arnold, Philip A. Collier, Stewart A. Leech, Jacob M. Rose, and Steve G. Sutton. 2023. Can knowledge based systems be designed to counteract deskilling effects? *International Journal of Accounting Information Systems* 50, (September 2023), 100638. https://doi.org/10.1016/j.accinf.2023.100638

[6] Ugur Aytac. 2022. Digital Domination: Social Media and Contestatory Democracy. *Political Studies,* 72, 1 (May 2022), 6-25. https://doi.org/10.1177/00323217221096564

[7] André Bächtiger, John S. Dryzek, Jane Mansbridge, and Mark E. Warren. 2018. *The Oxford Handbook of Deliberative Democracy*. Oxford University Press.

[8] Chris Bail. 2021. Breaking the Social Media Prism: How to Make Our Platforms Less Polarizing. In *Breaking the Social Media Prism*. Princeton University Press. https://doi.org/10.1515/9780691246499

[9] John Bailey. 2023. ChatGOV: Harnessing the Power of AI for Better Government Service Delivery. *American Enterprise Institute - AEI*. Retrieved March 26, 2024 from https://www.aei.org/technology-and-innovation/chatgov-harnessing-the-power-of-ai-for-better-government-service-delivery/

[10] Sarah Bankins and Paul Formosa. 2023. The Ethical Implications of Artificial Intelligence (AI) For Meaningful Work. *Journal of Business Ethics* 185, 4 (July 2023), 725–740. https://doi.org/10.1007/s10551-023-05339-7

[11] Anu Bradford. 2023. *Digital Empires: The Global Battle to Regulate Technology*. Oxford University Press.

[12] Jack Brewster, Lorenzo Arvanitis, and McKenzie Sadeghi. 2023. Could ChatGPT Become A Monster Misinformation Superspreader? *NewsGuard*. Retrieved March 15, 2024 from https://www.newsguardtech.com/misinformation-monitor/jan-2023

[13] Adrian Bua. 2017. Scale and Policy Impact in Participatory-Deliberative Democracy: Lessons from a Multi-Level Process. *Public Administration* 95, 1 (2017), 160–177. https://doi.org/10.1111/padm.12297

[14] Joy Buchanan, Stephen Hill, and Olga Shapoval. 2024. ChatGPT Hallucinates Non-existent Citations: Evidence from Economics. *The American Economist* 69, 1 (March 2024), 80–87. https://doi.org/10.1177/05694345231218454

[15] Joseph H. Carens. 2000. *Culture, Citizenship, and Community: A Contextual Exploration of Justice as Evenhandedness*. Oxford University Press.

[16] Jiawei Chen, Hongyu Lin, Xianpei Han, and Le Sun. 2023. Benchmarking Large Language Models in Retrieval-Augmented Generation. arXiv: 2309.01431. Retrieved from https://doi.org/10.48550/arXiv.2309.01431

[17] Marc Cheong, Ehsan Abedin, Marinus Ferreira, Ritsaart Reimann, Shalom Chalson, Pamela Robinson, Joanne Byrne, Leah Ruppanner, Mark Alfano, and Colin Klein. 2024. Investigating gender and racial biases in DALL-E Mini Images. *ACM Journal of Responsible Computing* (March 2024). https://doi.org/10.1145/3649883

[18] Yves Citton. 2017. *The Ecology of Attention*. John Wiley & Sons.

[19] Joshua Cohen and Archon Fung. 2021. Democracy and the Digital Public Sphere. In *Digital Technology and Democratic Theory*. University of Chicago Press, 23–61. https://doi.org/10.7208/chicago/9780226748603.003.0002

[20] Paul R. Daugherty and H. James Wilson. 2018. *Human + Machine: Reimagining Work in the Age of AI*. Harvard Business Press.

[21] Catalina M. de Abreu. 2024. Artificial Intelligence and deepfakes take over Pakistan's elections. *France 24*. Retrieved March 26, 2024 from https://www.france24.com/en/tv-shows/truth-or-fake/20240208-artificial-intelligence-and-deepfakes-takeover-pakistan-elections

[22] Ziv Epstein, Aaron Hertzmann, the Investigators of Human Creativity, Memo Akten, Hany Farid, Jessica Fjeld, Morgan R. Frank, Matthew Groh, Laura Herman, Neil Leach, Robert Mahari, Alex "Sandy" Pentland, Olga Russakovsky, Hope Schroeder, and Amy Smith. 2023. Art and the science of generative AI. *Science* 380, 6650 (June 2023), 1110–1111. https://doi.org/10.1126/science.adh4451

[23] Don Fallis. 2021. The Epistemic Threat of Deepfakes. *Philosophy & Technology* 34, 4 (December 2021), 623–643. https://doi.org/10.1007/s13347-020-00419-2

[24] Mohammadreza Farrokhnia, Seyyed Kazem Banihashem, Omid Noroozi, and Arjen Wals. 2023. A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International* (March 2023), 1–15. https://doi.org/10.1080/14703297.2023.2195846

[25] Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke, and Effy Vayena. 2018. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds & Machines* 28, 4 (December 2018), 689–707. https://doi.org/10.1007/s11023-018-9482-5

[26] Mirko Franco, Ombretta Gaggi, and Claudio E. Palazzi. 2023. Analyzing the Use of Large Language Models for Content Moderation with ChatGPT Examples. In *3rd International Workshop on Open Challenges in Online Social Networks*, September 04, 2023, Rome Italy. ACM, Rome Italy, 1–8. https://doi.org/10.1145/3599696.3612895

[27] Fiona Fui-Hoon Nah, Ruilin Zheng, Jingyuan Cai, Keng Siau, and Langtao Chen. 2023. Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research* 25, 3 (July 2023), 277–304. https://doi.org/10.1080/15228053.2023.2233814

[28] Francisco García-Peñalvo and Andrea Vázquez-Ingelmo. 2023. What Do We Mean by GenAI? A Systematic Mapping of The Evolution, Trends, and Techniques Involved in Generative AI. *International Journal of Interactive Multimedia and Artificial Intelligence* 8, 4 (2023), 7. https://doi.org/10.9781/ijimai.2023.07.006

[29] Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press.

[30] M. A. Griffin. 2011. Motivating Reflective Citizens: Deliberative Democracy and the Internal Deliberative Virtues. *Journal of Value Inquiry* 45, 2 (May 2011), 175–186. https://doi.org/10.1007/s10790-011-9276-y

[31] Jurgen Habermas. 1998. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. MIT Press.

[32] Rafik Hadfi, Jawad Haqbeen, Sofia Sahab, and Takayuki Ito. 2021. Argumentative Conversational Agents for Online Discussions. *Journal of Systems Science and Systems Engineering* 30, 4 (2021), 450–464. https://doi.org/10.1007/s11518-021-5497-1

[33] Nick Hajli, Usman Saeed, Mina Tajvidi, and Farid Shirazi. 2022. Social Bots and the Spread of Disinformation in Social Media: The Challenges of Artificial Intelligence. *British Journal of Management* 33, 3 (2022), 1238–1253. https://doi.org/10.1111/1467-8551.12554

[34] Martin Hasal, Jana Nowaková, Khalifa Ahmed Saghair, Hussam Abdulla, Václav Snášel, and Lidia Ogiela. 2021. Chatbots: Security, privacy, data protection, and social aspects. *Concurrency and Computation: Practice and Experience* 33, 19 (2021), e6426. https://doi.org/10.1002/cpe.6426

[35] Laura Illia, Elanor Colleoni, and Stelios Zyglidopoulos. 2023. Ethical implications of text generation in the age of artificial intelligence. *Business Ethics, the Environment & Responsibility* 32, 1 (2023), 201–210. https://doi.org/10.1111/beer.12479

[36] Mohammad Hossein Jarrahi. 2019. In the age of the smart artificial intelligence: AI's dual capacities for automating and informating work. *Business Information Review* 36, 4 (December 2019), 178–187. https://doi.org/10.1177/0266382119883999

[37] Cheonsu Jeong. 2024. Fine-tuning and Utilization Methods of Domain-specific LLMs. arXiv: 2401.02981. Retrieved from https://doi.org/10.48550/arXiv.2401.02981

[38] Bhanuraj Kashyap and Paul Formosa. 2023. The Authority to Moderate: Social Media Moderation and its Limits. *Philosophy & Technology* 36, 4 (December 2023), 78. https://doi.org/10.1007/s13347-023-00685-w

[39] Junaid Kathju. 2024. India's politics descends into an AI arms race as deepfakes threaten elections. *South China Morning Post*. Retrieved March 15, 2024 from https://www.scmp.com/week-asia/politics/article/3255388/indias-politics-descends-ai-arms-race-deepfakes-threaten-elections-and-theyre-not-only-ones-risk

[40] Soomin Kim, Jinsu Eun, Joseph Seering, and Joonhwan Lee. 2021. Moderator Chatbot for Deliberative Discussion: Effects of Discussion Structure and Discussant Facilitation. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1 (April 2021), 1–26. https://doi.org/10.1145/3449161

[41] Sarah Kreps and Doug Kriner. 2023. How AI Threatens Democracy. *Journal of Democracy* 34, 4 (2023), 122–131. https://doi.org/10.1353/jod.2023.a907693

[42] James H. Kuklinski, Paul J. Quirk, Jennifer Jerit, David Schwieder, and Robert F. Rich. 2000. Misinformation and the Currency of Democratic Citizenship. *The Journal of Politics* 62, 3 (2000), 790–816. https://doi.org/10.1111/0022-3816.00033

[43] Vimal Kumar, Priyam Srivastava, Ashay Dwivedi, Ishan Budhiraja, Debjani Ghosh, Vikas Goyal, and Ruchika Arora. 2024. Large-Language-Models (LLM)-Based AI Chatbots: Architecture, In-Depth Analysis and Their Performance Evaluation. In *Recent Trends in Image Processing and Pattern Recognition*, 2024, Cham. Springer Nature Switzerland, Cham, 237–249. https://doi.org/10.1007/978-3-031-53085-2_20

[44] Seth Lazar. 2023. Communicative Justice and the Distribution of Attention. Video. (26 January 2023). Retrieved July 28, 2023 from https://hai.stanford.edu/events/tanner-lecture-ai-and-human-values-seth-lazar

[45] Taylor Matthews. 2022. Deepfakes, Intellectual Cynics, and the Cultivation of Digital Sensibility. *Royal Institute of Philosophy Supplement* 92, (2022), 67–85. https://doi.org/10.1017/s1358246122000224

[46] Taylor Matthews and Ian James Kidd. 2023. The Ethics and Epistemology of Deepfakes. In *The Routledge Handbook of Philosophy and Media Ethics*. Routledge. 342-354. https://doi.org/10.4324/9781003134749-33

[47] Spencer McKay and Chris Tenove. 2021. Disinformation as a Threat to Deliberative Democracy. *Political Research Quarterly* 74, 3 (September 2021), 703–717. https://doi.org/10.1177/1065912920938143

[48] Morgan Meaker. 2023. Slovakia's Election Deepfakes Show AI Is a Danger to Democracy. *Wired*. Retrieved March 15, 2024 from https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/

[49] Bradley D. Menz, Natansh D. Modi, Michael J. Sorich, and Ashley M. Hopkins. 2023. Health Disinformation Use Case Highlighting the Urgent Need for Artificial Intelligence Vigilance: Weapons of Mass Disinformation. *JAMA Internal Medicine* (November 2023). https://doi.org/10.1001/jamainternmed.2023.5947

[50] Dan Milmo. 2024. Iran-backed hackers interrupt UAE TV streaming services with deepfake news. *The Guardian*. Retrieved March 15, 2024 from https://www.theguardian.com/technology/2024/feb/08/iran-backed-hackers-interrupt-uae-tv-streaming-services-with-deepfake-news

[51] Trevor Mogg. 2023. Lawyer sorry for fake court citations created by ChatGPT. *Digital Trends.* Retrieved 20 December, 2023 from https://www.digitaltrends.com/computing/lawyer-says-sorry-for-fake-court-citations-created-by-chatgpt/

[52] Diana C. Mutz. 2006. *Hearing the Other Side: Deliberative Versus Participatory Democracy*. Cambridge University Press.

[53] Nic Newman, Richard Fletcher, Antonis Kalogeropoulos, David A. L. Levy, and Rasmus K. Nielsen. 2017. Reuters Institute Digital News Report 2017. *Reuters Institute for the Study of Journalism.* Retrieved 17 December, 2023 from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf

[54] Jessica Nicosia, Andrew J. Aschenbrenner, Sarah L. Adams, Marisol Tahan, Sarah H. Stout, Hannah Wilks, Joyce E. Balls-Berry, John C. Morris, and Jason Hassenstab. 2022. Bridging the Technological Divide: Stigmas and Challenges With Technology in Digital Brain Health Studies of Older Adults. *Frontiers in Digital Health* 4, (April 2022), 880055. https://doi.org/10.3389/fdgth.2022.880055

[55] María L. Paúl. 2023. Porto Alegre city council passes the first AI-drafted law in Brazil. *The Washington Post.* Retrieved 17 December, 2023 from https://www.washingtonpost.com/nation/2023/12/04/ai-written-law-porto-alegre-brazil

[56] Michael A. Peters, Liz Jackson, Marianna Papastephanou, Petar Jandrić, George Lazaroiu, Colin W. Evers, Bill Cope, Mary Kalantzis, Daniel Araya, Marek Tesar, Carl Mika, Lei Chen, Chengbing Wang, Sean Sturm, Sharon Rider, and Steve Fuller. 2023. AI and the future of humanity: ChatGPT-4, philosophy and education – Critical responses. *Educational Philosophy and Theory* (June 2023), 1–35. https://doi.org/10.1080/00131857.2023.2213437

[57] Jeremy Pitt. 2023. ChatSh*t and Other Conversations (That We Should Be Having, But Mostly Are Not). *IEEE Technol. Soc. Mag.* 42, 3 (2023), 7–13. https://doi.org/10.1109/MTS.2023.3299450

[58] Regina Rini. 2021. Weaponized Skepticism: An Analysis of Social Media Deception as Applied Political Epistemology. In *Political Epistemology*. Oxford University Press, 31–48. https://doi.org/10.1093/oso/9780192893338.003.0003

[59] Siavosh Sahebi and Paul Formosa. 2022. Social Media and its Negative Impacts on Autonomy. *Philosophy & Technology* 35, 3 (2022), 70. https://doi.org/10.1007/s13347-022-00567-7

[60] Jared Schroeder. 2018. Toward a discursive marketplace of ideas: Reimaging the marketplace metaphor in the era of social media, fake news, and artificial intelligence. *First Amendment Studies* 52, 1–2 (July 2018), 38–60. https://doi.org/10.1080/21689725.2018.1460215

[61] Nick Schuster and Seth Lazar. 2024. Attention, moral skill, and algorithmic recommendation. *Philosophical Studies* (January 2024). https://doi.org/10.1007/s11098-023-02083-6

[62] Almog Simchon, Matthew Edwards, and Stephan Lewandowsky. 2024. The persuasive effects of political microtargeting in the age of generative artificial intelligence. *PNAS Nexus* 3, 2 (February 2024), pgae035. https://doi.org/10.1093/pnasnexus/pgae035

[63] Yashraj Sharma. 2024. Deepfake democracy: Behind the AI trickery shaping India's 2024 election. *Al Jazeera*. Retrieved March 15, 2024 from https://www.aljazeera.com/news/2024/2/20/deepfake-democracy-behind-the-ai-trickery-shaping-indias-2024-elections

[64] Muhammad Shoaib, Shiyu Jiang, Luo Jin, Donal Fitzpatrick, and Ian Pitt. 2023. An Artificial Intelligence-Based Interactive Learning Platform to Assist Visually Impaired Children in Learning Mathematics. In *HCI International 2023 Posters*, 2023, Cham. Springer Nature Switzerland, Cham, 366–373. https://doi.org/10.1007/978-3-031-35992-7_51

[65] Graham Smith and Corinne Wales. 2000. Citizens' Juries and Deliberative Democracy. *Political Studies* 48, 1 (March 2000), 51–65. https://doi.org/10.1111/1467-9248.00250

[66] Sarah Sorial. 2022. Deliberation and the Problems of Exclusion and Uptake: The Virtues of Actively Facilitating Equitable Deliberation and Testimonial Sensibility. *Ethical Theory Moral Practice* 25, 2 (April 2022), 215–231. https://doi.org/10.1007/s10677-022-10273-0

[67] Nick Srnicek. 2017. *Platform Capitalism*. John Wiley & Sons.

[68] Chris Stokel-Walker and Richard Van Noorden. 2023. What ChatGPT and generative AI mean for science. *Nature* 614, 7947 (February 2023), 214–216. https://doi.org/10.1038/d41586-023-00340-6

[69] Zeynep Tufekci. 2017. *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. Yale University Press.

[70] Shannon Vallor. 2015. Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character. *Philosophy & Technology* 28, 1 (2015), 107–124. https://doi.org/10.1007/s13347-014-0156-9

[71] Shannon Vallor. 2016. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting.* Oxford University Press.

[72] Krzysztof Wach, Cong Doanh Duong, Joanna Ejdys, Rūta Kazlauskaitė, Pawel Korzynski, Grzegorz Mazurek, Joanna Paliszkiewicz, and Ewa Ziemba. 2023. The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review* 11, 2 (2023), 7–30. https://doi.org/10.15678/EBER.2023.110201

[73] Boshi Wang, Xiang Yue, and Huan Sun. 2023. Can ChatGPT Defend its Belief in Truth? Evaluating LLM Reasoning via Debate. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, December 2023, Singapore. Association for Computational Linguistics, Singapore, 11865–11881. https://doi.org/10.18653/v1/2023.findings-emnlp.795

[74] Christopher Whyte. 2020. Deepfake news: AI-enabled disinformation as a multi-level public policy challenge. *Journal of Cyber Policy* 5, 2 (May 2020), 199–217. https://doi.org/10.1080/23738871.2020.1797135

[75] Weiyue Wu and Shaoshan Liu. 2023. A Comprehensive Review and Systematic Analysis of Artificial Intelligence Regulation Policies. arXiv: 2307.12218. Retrieved from https://doi.org/10.48550/arXiv.2307.12218

[76] Shoshana Zuboff. 1989. *In the Age of the Smart Machine.* Basic Books.