HUMAN-AI POLITICAL SYMBIOSIS: RETHINKING DEMOCRACY IN THE AGE OF DIGITAL TECHNOLOGIES

TOMISLAV FURLANIS
Faculty of Theology
University of Ljubljana
tomislav.furlanis@teof.uni-lj.si

IVAN CEROVAC
Faculty of Humanities and Social Sciences
University of Rijeka
Faculty of Theology
University of Ljubljana
icerovac@uniri.hr

ABSTRACT
This paper explores how AI-powered digital technologies impact democratic processes by shaping citizens' epistemic and moral capacities, with a particular focus on the threats posed by misinformation, deepfakes, and overreliance on AI tools. Traditional approaches like education and regulation have been insufficient in addressing these challenges. The paper proposes a paradigm shift, advocating for human-AI political symbiosis as a way to mitigate the harms of digital technologies and improve the epistemic and moral qualities of democratic decision-making. This symbiosis entails a cooperative relationship between humans and AI, where both partners augment each other's capacities to navigate a distorted epistemic environment. The concept is introduced through two forms: "weak" symbiosis, where AI is a useful tool for achieving specific democratic goals, and "strong" symbiosis, where humans and AI are mutually dependent for democratic survival and flourishing. The paper concludes by examining how this new framework can better address the complex, evolving challenges posed by AI in democratic settings, emphasizing that deeper cooperation between AI and human agents is key to preserving democracy.

KEYWORDS
epistemic democracy, AI governance, misinformation, deepfakes, epistemic autonomy, political agency, epistemic environment

Digital technologies have a profound influence on democratic politics. They have served as both a positive force, facilitating citizens' political participation in decision-making processes and improving transparency and accountability, and a negative force, accelerating the spread of disinformation and reinforcing citizens'

biases and epistemic vices (Tucker et al. 2017). This influence represents a complex issue that is often examined superficially, without sufficient consideration of the bigger picture. Additionally, assessments of this influence are frequently shaped by politics (Farrell 2012) or by the sources of research funding (Schippers 2020). Although many are concerned about the negative effects digital technologies have on democratic processes, there is no consensus on the causes of these effects, nor a unified approach regarding preferable actions to mitigate them (Halpern 2023).

The rapid development of AI technologies exacerbates the already significant effects digital technologies have on democracy. AI is a broad term encompassing various technologies, many of which are applied in fields such as medicine, manufacturing, and the entertainment industry, with little connection to politics. For the purposes of this paper, we regard AI primarily as machine, deep, learning systems deployed and utilized in digital social media and online contexts, where large amounts of data are analyzed with minimal human input. These systems are used to search content on the internet, provide personalized recommendations, profile users on social networks, analyze patterns in their interests and behaviors, target citizens with highly personalized ads, and create credible texts, images, audio, and video content. Major platforms like Facebook, Twitter, Instagram, or TikTok, along with search engines such as Google, Bing, or Yahoo, use AI-based algorithms to power their recommender systems, personalize content, enhance user engagement, and optimize search results. More recently, large language models (LLMs) have enabled the generation of vast amounts of credible textual material, further complicating the distinction between human-created content and AI-produced information.

AI technologies can influence political processes in numerous ways. How these effects will be realized depends on several factors, one of the most important being whether the use of AI occurs in democratic or authoritarian societies. Regarding the latter, many reports describe how China has (ab)used search algorithms to distort the truth (Brandt et al. 2022), and similar practices have been employed by other authoritarian regimes (Yang, Roberts 2023). While authoritarian regimes use AI technologies to gain popular support and create a facade of democratic legitimacy, decision-making processes remain largely outside popular control. As a result, the quality of political decisions does not depend on citizens' competencies and is thus unaffected by the influence digital technologies have on their cognition. Decision-making in democratic regimes, on the other hand, depends on and is primarily shaped by the will of the citizens. The harmful effects digital technologies can have on citizens' capacities directly spill over into the quality of political decisions produced by democratic processes. Following, this paper focuses on liberal democratic societies and examines how AI-powered digital technologies endanger the legitimacy of democratic procedures by

undermining citizens' epistemic and political competencies, thus hindering both moral and epistemic qualities of democratic decision-making procedures.

Many attempts have been made to analyze and counter this negative influence, aiming to protect both citizens and decision-making processes from the potential dangers introduced by AI-powered digital technologies. A common approach relies on (both formal and informal) education to develop citizens' capabilities. By teaching children and adults about the dangers related to digital technologies and facilitating the spread of accurate, verified information, we could mitigate or at least reduce the harmful effects of online misinformation and the flawed epistemic structures created by recommender algorithms (De Blasis 2019; Maertens et al. 2021). Another approach focuses on the self-regulation of big tech companies, where they develop their ethical codes and attempt to balance innovation and profit with the public good (Tambini 2021). Finally, many advocate for government regulation of AI, calling for a legal framework that prevents misuse, addresses privacy concerns and promotes algorithmic transparency (Vidal Bustamante 2022; Finocchiaro 2024). However, there are serious concerns about the feasibility and effectiveness of each of these approaches. While combining them might partially mitigate the problem, it is unlikely to provide a stable solution. There is a pressing need for new solutions that can reshape the existing debate and tackle the issue from a different perspective. AI technologies are often regarded as a serious threat to democracy (which they are), but they are rarely, if ever, considered as useful tools or partners in navigating an epistemically distorted environment.

This paper analyzes human-AI symbiosis, a concept dating back to the 1960s that is gaining popularity in AI research and development and applies it to analyze and mitigate the negative impacts digital technologies have on democracy. We inquire whether symbiosis as a developmental standard has the potential to enhance citizens' moral and epistemic capacities, enabling them to navigate the distorted epistemic environment and exercise their political influence in a more accountable and responsible manner. While the existing debate confronts the negative impact of AI on democracy through traditional means – such as education or regulation of AI technologies – our approach explores whether these problems can be tackled by *redefining the relationship between citizens and AI in terms of interaction.* The innovative aspects of the paper lie in its shift from the traditional debate and its premise that we can address the threat of AI to democracy by focusing on incorporating more, not less, AI in the development of our joint capacities. The paper does not aim to wholly reject existing attempts to tackle the problem or to provide a final answer. Instead, it seeks to shift the debate and highlight new approaches that should be investigated further. In other words, our aim is not to propose a definitive solution to the threat AI poses to democratic systems, but rather to advocate for a much-needed shift in perspective. Traditional

approaches tend to address this threat by imposing regulations and attempting to slow down the pace of technological change in order to protect democratic societies. In contrast, we argue for a shift toward more, not less, technology as a response to these concerns. However, we can only "fight fire with fire" if we carefully consider all relevant factors and guide the development and use of new technologies in ways that enhance and complement our epistemic and moral capacities, rather than undermine them and increase our blind deference to technology. Presenting such an account goes well beyond the scope of this paper and our expertise as philosophers. The modest aspiration behind this work is to advocate for a shift in perspective regarding how digital technologies powered by AI algorithms are understood and to motivate discussion on new ways to combat the detrimental effects of these technologies on democratic processes.

We proceed as follows. In the first part, we build upon John Stuart Mill's criteria of good governance to systematize how AI-powered digital technologies can affect democratic processes. This influence is multifaceted, with both potentially harmful and beneficial effects. In this analysis, we focus specifically on the impact of AI on citizens' capabilities, particularly those essential for effective democratic participation. In the second part, we review existing research that analyzes how digital technologies might impair important capacities such as critical thinking and the ability to form and act on personal beliefs. We connect these findings with relevant democratic research to show how AI technologies can undermine the citizens' relevant participation in democratic procedures. In the third part, we discuss traditional solutions (such as education of citizens, self-regulation by tech companies, and government regulation) aimed at mitigating the detrimental effects of AI on citizens' capacities. Unfortunately, these approaches, while producing some beneficial results, seem to fail to fully address the harmful effects digital technologies can have on citizens' epistemic capacities. This might be because we approach the problem from a perspective where AI-powered technologies that distort our epistemic environment are seen as a threat to human epistemic agency. A potential solution could be to change the paradigm and view AI as a partner that can help citizens navigate the distorted epistemic environment. In the fourth part, we introduce the concept of human-AI symbiosis, explore its two forms (weak and strong symbiosis), and examine how they can be used to either protect or reframe our epistemic and political agency. Additionally, we analyze how the beneficial effects of human-AI symbiosis can mitigate potential dangers created by digital technologies in a democratic setting. The goal is to analyze how political human-AI symbiosis can serve as a clarifying lens to appropriately interpret and address the current and emerging state of digital technologies, with a focus on their implications for the development of democratic systems. Finally, in the fifth part, we address potential objections to the concept of human-AI symbiosis and its application to liberal democracy.

1. ASSESSING AI'S EFFECTS ON DEMOCRACY: A MILLIAN APPROACH

Digital technologies powered by AI algorithms have influenced democratic politics for over two decades, leading to numerous attempts to systematically analyze and categorize their effects on decision-making processes. In his influential book *Democracy and Technology* (1995), Richard Sclove distinguishes between democracy-enhancing and democracy-undermining effects of technology, as well as between direct effects (such as those affecting citizens' ability to participate in decision-making, like online voting) and indirect effects (such as those influencing societal structures and impacting overall trust in democracy). Some authors have focused on the areas affected by digital technologies, organizing their effects into technological, political, and social spheres (Bartlett 2018). Others have concentrated on how these technologies impact communication between citizens, their participation in decision-making processes, and their trust in political institutions (Zuboff 2019; Gilardi 2022; Krebs, Kriner 2023). Finally, some researchers have examined the manipulative capabilities of AI, systematizing its effects on democracy by focusing on the means used to exercise this influence, such as micro-targeting or the exploitation of cognitive biases (Susser et al. 2018; Jungherr 2023). These efforts have significantly shaped and contributed to the ongoing discourse and have achieved the objectives intended by their authors. However, our approach necessitates a somewhat distinct framework for understanding the effects of digital technologies on democratic politics.

John Stuart Mill famously argued that the instrumental qualities of democracy depend on (i) its ability to enhance the intellectual and moral capacities of citizens who exercise political influence, and (ii) its capacity to organize these existing capacities and their epistemic contributions optimally, conducive to correct, efficient, and just decisions (Mill 1977; see also Cerovac 2016). Following Mill's criteria, we can systematize the impact of AI-powered digital technologies on democratic politics by assessing how these technologies affect citizens' epistemic and moral capacities, as well as by evaluating how they alter existing decision-making procedures that aggregate and organize citizens' political input. This framework enables us to categorize and group various effects, facilitating a better understanding of their influence on democratic processes. Technologies that influence how citizens participate in decision-making and how governments formulate proposals (such as online voting, online activism, digital agenda-setting, and the use of machine learning algorithms in decision-making) align with Mill's second criterion. Conversely, technologies that affect how citizens communicate, acquire new information, and form their political opinions fall under the first criterion. Failure to meet either criterion can lead to poor decision-making and substantively detrimental or harmful political outcomes.

Our assessment of the impact digital technologies can have on democracy aligns with Mill's first criterion. We thus compile and analyze various digital tools and algorithms that influence how citizens communicate and receive new information to evaluate their effects on democratic procedures and their capacity to produce quality outcomes. Specifically, in democratic systems, ultimate political authority rests with the will of the citizens, who elect their representatives and thus authorize laws, policies, and political decisions. Although contemporary democracies employ various mechanisms to filter the popular will - aiming to mitigate citizens' political ignorance, biases, and emotionally driven responses to preserve democracy's capacity to produce sound legislation - political representation has its limits. If citizens' competencies are neglected and the detrimental effects that shape their epistemic and political capacities (such as those induced by digital technologies, political polarization, economic inequalities, illiberal or undemocratic cultural and social norms, and other influences) are continuously overlooked, it may result in the election of grossly incompetent or even malevolent political leaders.

To sum up and anticipate the main argument: Following Mill, we hold that the (instrumental or epistemic) quality of decision-making processes stems from (i) the capacities of those involved in the process and (ii) the properties of the decision-making procedures that shape and organize the political participation of decision-makers. Substantial changes in our epistemic environment, produced by the introduction of AI-powered digital technologies into our everyday epistemic practices - including the way we form our political opinions and preferences - inevitably affect our capacities to competently participate in the decision-making process. If we do not want to give up liberal democracy for some form of epistocracy (and wish to keep the procedures that organize our epistemic contributions more or less the same), we must focus on protecting our epistemic capacities, which are endangered by digital technologies, from this detrimental influence. Alternatively, we need to find ways to enhance these capacities and develop resilience to the harms inflicted by machine learning algorithms. Typical political responses focus on the first approach, attempting to limit the harmful effects of digital technologies by introducing regulatory practices. They also address the second approach by investing in education to improve citizens' capacities, making them less susceptible to dangerous digital influences. However, if these efforts fail to keep pace with the challenges posed by advances in digital technologies, we may need to consider new methods of improving our epistemic capacities. One such method is human-AI symbiosis.

To continue the argument, we will provide a brief overview of the evidence highlighting how AI-powered digital technologies impact citizens' epistemic and moral capacities, as well as their ability to exercise their epistemic and political agency. This will help illustrate the need for both traditional and new approaches to effectively address these challenges.

2. AI AND THE EROSION OF CITIZENS' EPISTEMIC CAPACITIES

There are a number of possible cases we can take to show that contemporary generative AI poses a threat to citizens' epistemic capacities. The scope of this debate greatly exceeds the limits of this paper. We thus focus on four groups of effects AI-powered digital technologies can have on the epistemic environment and the citizens' capacities to navigate that environment. First of all, AI has the capacity to produce high volumes of persuasive misinformation, thus creating an illusion of majority perspectives and deteriorating the quality of online content. Second, recent improvements in deepfake technology have enabled the creation of realistic videos that can easily mislead viewers and affect their reasoning even when the viewers know these are fake. Third, generative AI tools mediate the research process by summarizing information. Their simplicity can motivate the users to uncritically rely and depend predominantly on these tools as sources of information while disregarding other traditional credible sources. Fourth, such overreliance can be abused as AI can be used to manipulate users' political opinions and preferences.

(i) Generative AI and Misinformation Pollution

The potential for generative large language models to act as effective misinformation generators and pollute the informational environment of the internet has come to the forefront of research under the notion of "AI content pollution" (Pan et al. 2023). Here, it is surmised that text generators have the ability to produce persuasive, high-volume texts at tremendous speeds. This capability enables them to "pollute" the informational environment by significantly magnifying the presence and accessibility of specific (political) content, thus creating an illusion of a majority perspective for certain political attitudes or views that would otherwise be deemed fringe and irrelevant (Zhou et al. 2023). To exemplify, a recent study reveals how a significant portion of online text content, especially in languages prevalent in Africa and the Global South, is already poorly machine-translated (Thompson et al. 2024) using generative AI. Specifically, more than half of all sentences on the internet have already been translated into multiple languages, often resulting in progressively deteriorating quality due to low-quality machine translation. Thus, for some languages, it is already the case that the content one receives daily from internet platforms, such as news portals, is machine-made and of weaker quality.

The ability of contemporary generative models to massively produce and represent content on the internet is not limited solely to the generation of text, audio, and visual content; it also extends to website production, such as blogs and personal webpages. Search engine results are epistemically important, as studies show that adults often use the internet as their primary information source (Biddix

et al. 2011). The epistemic quality of answers provided by web search engines is thus one of the most relevant sources for the everyday problem-solving capacity of contemporary citizens. Generative AI may distort the already imperfect knowledge representation capacity of web search engines. As a recent journalism report shows, AI tools can already enable users to create a thousand "slightly different versions of the same article with a single click and to automatically publish them to as many WordPress sites as you want using a paid plugin" (404 Media 2023). Consequently, the entire website production line can be automated, and the generated websites' presence in internet searches can be reinforced through ad purchases that position the automatically generated site at the top of the search engine's results. This threat has already forced Google to update its search engine's ranking system to tackle "explicit fake content appearing in Search" (Google 2023).

The ability of AI-powered digital tools to produce vast quantities of content quickly and to overwhelm the epistemic environment can reduce citizens' ability to discern between credible information and misinformation, leading to cognitive fatigue and doubt in their epistemic capacities. Additionally, as citizens encounter more and more misinformation in the epistemic environment, their ability to critically evaluate the information they receive declines, making them unfit to exercise epistemic agency in such a polluted environment.

(ii) AI and Deepfake Technology: Audio-Visual Misinformation

Another content pollution threat is deepfake video generation, where present-day AI tools like Runway, Sora, or Vidu can create realistic videos of public importance that can be freely shared on social platforms. Technology has already become hyper-realistic, requiring very small amounts of real-life data for its generation. For instance, Microsoft's VASA-1 model (not publicly available due to misinformation concerns) can generate realistic animated videos from just a single photo of a person and an accompanying audio track (Xu et al. 2024). In the public realm, the X platform rolled out its Grok-2 picture generative system, which does not watermark its generated images. Consequently, controversial examples were quickly generated and widely shared on the platform, including an image of Mickey Mouse wearing a "Make America Great Again" hat while holding a cigarette and beer, and a collection of images of Taylor Swift and her fans supporting Donald Trump for president (Marr 2024). Moreover, most recently (September 2024), Elon Musk personally shared a deepfake video of the presidential candidate Kamala Harris, where her voice was completely AI-generated. This so-called "hit and run" video was viewed over 100 million times and Musk in a later post pointed out its satirical (deepfake) nature. This example illustrates that even when users know content is fake, it can still significantly

influence their engagement, and research shows perceptions and beliefs (Menczer, Hills 2020; Ecker 2022).

A distorted epistemic environment encourages generalized distrust in information sources. Additionally, this can undermine public discourse and civic engagement, as citizens may no longer feel confident in distinguishing what is true and what is not. Consequently, citizens may become passive and disinterested in important societal issues, believing that the information landscape is too corrupted to navigate (Coeckelbergh 2022).

(iii) Overreliance on Generative AI in Search and Education

Another possible epistemic problem arises when internet search providers utilize generative AI to summarize search findings for users. For instance, Bing Copilot answers users' search queries by summarizing the most relevant information from internet-retrieved web pages, documents, videos, and pictures. With such an interface, Bing Copilot does not provide "unfiltered" web page results for users to inspect and assess relevance. Instead, it presents findings that the system itself deems most relevant. As such, Bing Copilot creates a new layer of "epistemic mediation" that is more "agential" than traditional representations of listed web pages. The reason for this lies in the Copilot's ability to pick and choose materials it finds most trustworthy and correct in answering the user's query.

If used longitudinally, the practice of AI-powered search engines might lead to negative epistemic outcomes. For instance, it has been shown that users prefer links positioned higher in search results, even if their abstracts are less relevant (Pan et al. 2007). Additionally, users might fail to acknowledge the epistemic contributions of other citizens whose content and style are systematically neglected by search algorithms, thus perpetuating epistemic injustice (Samaržija, Cerovac 2021). With LLM-empowered search summaries, users may place greater trust in these answers for convenience, even though LLM search engines produce greater amounts of factual errors and misinformation, demanding more epistemic attention from users than traditional search engines. This was most recently corroborated in the infamous case of Google's own LLM-empowered search, AI overviews, after a series of incorrect and detrimental answers ranging from providing dangerous medical advice to outright extremist disinformation. (Tom's Hardware 2023).

LLMs, however, do not need to impact web searches to diminish attentiveness toward trustworthy content and relevant information retrieval processes. They can serve as information providers, directly substituting traditional web search engines altogether, a phenomenon already being observed. For instance, a recent survey shows that nearly 89% of American college students use ChatGPT to complete

homework tasks, with 53% using the tool for writing papers (McGee 2023). Scholars express concern that using ChatGPT in education may lead to epistemic problems, such as a lack of interpersonal understanding, context analysis, and innovative thinking, ultimately fostering technical dependence on these systems (Yu 2023).

(iv) Political Manipulation and Democratic Risks

To exemplify, a recent study warns that users perceive ChatGPT's responses as having higher information quality compared to Google Search, reporting significantly better experiences in terms of usefulness, enjoyment, and satisfaction. However, "the use of ChatGPT may also lead to overreliance and generate or replicate misinformation" (Xu, Feng & Chen 2023). Participants often demonstrate a lack of diligence when using ChatGPT and are less motivated to verify and rectify any misinformation in its responses. Observations indicate that 70.8% of the participants in the ChatGPT group relied on its responses, often accepting them as true without further investigation. This convenience may inadvertently hinder users from engaging in deeper exploration and identifying misinformation within its responses.

As Shannon Vallor notes in her most recent book, contemporary generative AI is the first technology in history that can jeopardize our future by preventing us from knowing how to create one at all (Vallor 2024, p. viii). Correspondingly, this concern extends to democratic education, as the widespread adoption of LLM chatbots as information providers could lead to epistemic encapsulation, where the majority of retrieved knowledge regarding democracy comes from chatting with LLMs rather than from public exchanges of ideas on the internet. Instead of becoming gateways to freely available digital public goods, LLM models may become informational sandboxes, upon which humans over-rely, reversing the progress made by the web toward democratizing access to knowledge and information (de la Rio Chanona et al. 2023).

Moreover, when the knowledge retrieval capabilities of LLMs are coupled with their potential for persuasion (the ability to alter users' beliefs), the pervasive use of AI services could significantly alter our information environment, leading to a loss of human control over our future (Sunstein 2022, 2024; Burtell, Woodside 2023). Since LLMs can "chat" in natural language, they can use the language common to specific political subgroups. Evidence shows that individuals consume information represented in the dominant language style of their political group, even if that information contradicts their personal beliefs (Yom-Tov et al. 2014). Consequently, LLMs could "persuade" users towards political attitudes they might not typically endorse, particularly if the AI becomes their dominant conversational partner and presents itself as a member of their political in-group through shared language (Mitra et al. 2024).

These illustrations demonstrate that the significant changes in our epistemic environment produced by the use of AI-powered technologies affect citizens' belief formation and belief revision processes. While citizens are not directly prevented from forming autonomous beliefs and making related policy choices, this capacity is severely hampered, and its exercise is made more difficult because of the manipulative epistemic architecture and environment (Coeckelbergh 2022).

3. MITIGATING HARM FROM DIGITAL TECHNOLOGIES: WHY DO TRADITIONAL SOLUTIONS FALL SHORT?

Government policies designed to curb the negative effects of AI-powered digital technologies can be broadly categorized into two groups: those that focus on educating and training citizens to recognize misinformation and the distorted nature of the epistemic environment and those that aim to regulate the channels through which misinformation spreads and, more broadly, to regulate the epistemic architecture in general. The first approach is less restrictive but may be less effective, while the second approach may yield better short-term results but depends on more restrictive measures.

Educational and training initiatives can further be divided into reactive and preventive measures. Reactive measures aim to debunk misinformation already circulating in society, whereas preventive measures seek to strengthen citizens' epistemic capacities so they can independently identify misinformation. The reactive approach is relatively straightforward and cost-effective: a publicly funded agency analyzes circulating misinformation, identifies harmful content that is spreading quickly, and publicly debunks it by exposing its falsehood and manipulative intent. The goal is that, once presented with the relevant evidence, manipulated individuals will recognize the misinformation and abandon other beliefs based on it (Cerovac, Drmić 2023).

However, this approach often falls short. First, it is increasingly difficult to identify who has been exposed to the epistemically harmful content spread via digital technologies, and even more challenging to create a reliable dissemination strategy to reach those individuals with the debunking information. In fact, research suggests that citizens who are most susceptible to fake news are also the hardest to reach with evidence that debunks disinformation (Waisbord 2020). This challenge is compounded by confirmation bias, which refers to people's tendency to seek out information that supports their existing beliefs. Furthermore, even when debunking information reaches the right audience, there is no guarantee that individuals will adjust their opinions in light of new evidence. Many people view their beliefs as integral to their identity, making them reluctant to change—even if those beliefs are based on disinformation (Ecker 2022). This

resistance is further fueled by disconfirmation bias, where individuals dismiss evidence that threatens their established views.

Additionally, even when citizens acknowledge the debunking information and claim to reject the misinformation, they may still be influenced by it. Its harmful effects can linger despite being discredited. For instance, parents who once subscribed to disinformation linking vaccines to autism based on Wakefield's research may later recognize that the study was fraudulent, yet they often remain skeptical about vaccines and hesitate to vaccinate their children (Lee et al. 2022). Similarly, citizens who have seen a deepfake video of their preferred political candidate saying something they firmly oppose might be subconsciously demotivated from participating in the elections, despite knowing that the video was fabricated. While the reactive approach offers some valuable tools, it does not adequately address the underlying issues related to disinformation (Cerovac, Drmić 2023).

The preventive approach seeks to enhance citizens' knowledge and critical thinking skills, making them more resilient to the epistemic harms created by a distorted epistemic environment. The core idea is that the more accurate and coherent an individual's belief system is, the less likely misinformation will be able to take hold. Moreover, if citizens are trained to critically assess the content and sources of information, they will be less vulnerable to misinformation (De Blasis 2019). However, policies in this area have shown limited success (Maertens et al. 2021), and several concerns remain. First, when citizens lack knowledge about the source of information, they tend to accept it as true (Rapp 2016). Additionally, individuals often forget when and how they encounter certain information and may mistakenly associate false claims with credible sources (Brown 2021). Finally, the algorithms used by social media can inundate users with disinformation. Even if individuals do not explicitly endorse this misinformation, its overwhelming presence can still negatively impact their reasoning (Menczer and Hills 2020, Ecker 2022).

While the preventive approach has had some success in combating the spread of fake news, it is insufficient to address the challenges posed by epistemic biases and digital algorithms. Policies advocating for government regulation could potentially tackle these remaining issues. Such measures might focus on regulating the platforms where disinformation proliferates or on controlling the spread itself through censorship and legal action against those who produce fake news. The government can also target social media and the algorithms that curate highly personalized content for users. These algorithms not only exacerbate political polarization and create epistemic bubbles and echo chambers but also enable disinformation creators to tailor their manipulative tactics for specific audiences.

Although most social networks require transparency for paid advertisements, the algorithms governing content delivery have remained undisclosed for over a decade. This issue is also reflected in the potential misuse of copyrighted training

data for large language and video models, as exemplified by Marcus and Southern (2024) corroborated by the fact that researchers are already developing copyright content detectors for generative models (Li et al. 2024).Scholars supporting this regulatory approach contend that effective oversight of social media recommendation algorithms necessitates some form of government intervention (Vidal Bustamante 2022). However, regulating these algorithms will not entirely eliminate avenues for disseminating fake news, as closed digital spaces like WhatsApp groups will continue to facilitate its spread (Brown 2021) as was dramatically showcased in the most recent Korean deepfake public school scandal (BBC 2024).

Moreover, the government can actively prosecute individuals, companies, and partisan groups that create and distribute fake news. A notable example is radio host and conspiracy theorist Alex Jones, who was recently fined $965 million for promoting false claims about the Sandy Hook Elementary School shooting (Le Monde 2022; see also Cerovac, Drmić 2023), or the most recent case where the European Commission initiated formal proceedings to evaluate whether the platform X "may have breached the Digital Services Act (DSA) in areas linked to risk management, content moderation, dark patterns, advertising transparency and data access for researchers" (European Commission 2023).

While prosecuting those who disseminate harmful misinformation should significantly hinder its spread, this approach requires lengthy judicial processes and may prove costly and inefficient (Eddy, Scott 2017). Additionally, if applied improperly, it could undermine citizens' freedom and autonomy rather than protect them. Democratic systems face serious challenges: their epistemic integrity depends on the competencies of the citizens participating in decision-making processes, yet these competencies appear inadequate for navigating an epistemically distorted environment created by AI-powered digital technologies.

To steer us out of this purported interpretative deadlock Coeckelbergh (2024) has most recently introduced a deep-communicative interpretation of democracy. He builds upon Dewey who famously argued that "democracy is not just a form of government but a mode of associated living, of conjoint communicative experience'" (Dewey 1916, p. 93). To reposition our thinking about democracy, and with it our understanding of the human-technology relationship in it, Coeckelbergh builds upon Dewey's thesis to argue that, from within a deep communication understanding of democracy, communication is not only necessary for democracy to work; communication is what democracy is all about. (Coeckelbergh 2024, p. 219). "Democracy is not about particular leaders or institutions, but about a particular form of living together, a communicative form of co-habitation and personal and societal growth that is lived and experienced in concrete social practices" (Coeckelbergh 2024, p. 223). Similarly, the erosion of democracy is the "erosion of a particular form of life, one in which social

communication and cooperation in all its forms plays a central role, and one in which there is mutual recognition (also of differences) and absence of coercion and systemic forms of control" (Coeckelbergh 2024, p. 223).

Importantly, neither Dewey nor Coeckelbergh speak directly of symbiosis. However, for us, the idea of democracy as a "particular form of living together", a "co-habitation" that is "lived and experienced in concrete social practice" a "form of life" where "cooperation plays a central role" is a description completely aligned with the idea of political symbiosis. Accordingly, our interpretational turn follows upon Coeckelbergh's framework of democracy as communication. We, however, move a step forward to explicitly highlight the necessity of,tightly-coupled, cooperation between humans and AI systems for the success of 21st-century liberal democracy (Spina et al. 2023). And this tightly coupled cooperation between humans and AI systems for the sake of democratic success, development, and flourishing we call simply *human-AI political symbiosis.*

4. HUMAN-AI SYMBIOSIS: A SIMPLE OVERVIEW

In the past decade, the paradigm of human-AI cooperation has entered into the foray of investigation for both the civil and military sector of applications. (Crandall et al. 2018; Dafoe et al. 2020; Van Den Bosch and Bronkhorst 2018; Wang et al. 2020). The broad field of human AI cooperation aims to answer a rather straightforward question: how can we best enable machines to work with humans in an optimal manner while they simultaneously supplement and even augment human capabilities in a wide variety of tasks?

Recently, the answer to this question has been found in an older paradigm, that of human-machine symbiosis (Licklider 1960) which posits that, in order to achieve optimal human augmentation, the machines need to be finely tailored to their human partner(s). However, this is not where the symbiotic relationship ends. As, the central point of human-machine symbiosis, as per biological reality, lies in bidirectional augmentation - it is not only that the machine augments the human, but the human also augments its machine partner as they jointly collaborate to achieve a shared goal (Dede et al. 2021; Hassani et al. 2020).

In the words of Licklider (1960):

In not too many years, human brains and computing machines will be coupled together very tightly, and the resulting partnership will think as no human brain has ever thought and process data in a way not approached by the information-handling machines we know today [...]. To think in interaction with a computer in the same way that you think with a colleague whose competence supplements your own will require much tighter coupling between man and machine than is suggested by the example and than is possible today (Licklider 1960, p. 31).

Staying true to this original idea, the most recent scientific and ethical investigations into the idea of human-AI symbiosis are diverse and abundant, investigating how to accomplish human-AI symbiosis in various fields of endeavor, including education, medicine, manufacture, management, and public service. (Alam et al. 2024; Gesk, Leyer 2022; Hemmer et al. 2022; Raisch, Krakowski 2021; Sowa et al. 2021; Wang 2021).

How to relevantly define human AI symbiosis? Notwithstanding the possible range of valuable interpretations, the demands of the original concept require the existence of two necessary aspects (Gerber et al. 2020). First, the symbiotic relationship is mutually augmenting for both of the partners within the symbiotic relationship. This means to say that, in the symbiotic relationship both the human and the artificial partner - augment each other's capabilities (and substitute lacks!). And they do so while working jointly to achieve their common, shared, goal.

Symbiosis is not only about augmenting, enhancing, or extending the human agency (as this interest would specifically belong to the notion of "human augmentation or enhancement"). Human-AI symbiosis is properly and relevantly about the very ability of both symbiotic partners – or symbionts – to establish a simultaneous bidirectional augmentation as they form a single, organismic, whole. Thus, in symbiosis, the machine augments the human, and the human, simultaneously, augments the machine – all for the benefit, or interest, of achieving their joint goal (Zhou et al. 2021).

Second, in a Human-AI symbiotic partnership, the interaction between the agents is tightly coupled, which means to say that the symbiotic interaction connects the agents in such order of depth and interactional magnitude that it produces a new agential system of emergent properties (Furlanis, Gilbert 2023). In essence, the two become unified, and this symbiotic unity is greater than the sum of their agencies. In Licklider's own words, then, it is not the human or the machine that thinks or acts, it is the "partnership".

Importantly, the nature of the tight-coupling in a human AI symbiotic relationship leads us to conceptually delineate the relevant range of its realizations. The reason of which lies in the machine's manifold variability, and consequently the variability of symbiotic relationships they can establish with humans. To exemplify, even though all forms of human AI symbiosis withhold the same fundamental aspects, as also in cases of biological symbiosis, the concrete manifestations of symbiotic relationships can vary tremendously depending on the "nature" of the machine and how it cooperates with its human partner. For instance, the symbiosis achieved between a human paraplegic and her BCI system is dissimilar to the case of human-AI symbiosis between a police officer and his four-legged robot partner, which is different from the case of symbiosis between an organization's manager and his AI assistant.

Accordingly, to move forward and distill a usable concept – as challenged with the possibly huge landscape of symbiotic characteristics – we return back to the original, biological, meaning of the symbiotic concept. Based upon this definition we posit two valid and conceptually usable interpretations of human AI symbiosis – the "weak" and the "strong" version – corresponding to the biological reality of "facultative" and "obligate" symbiotic relationships.

In the obligate relationship, the organism cannot survive in the (hostile) environment (which more precisely entails it cannot complete its life lifecycle) without being constantly co-joined or integrated with the other organism. In the facultative relationship, the individual organisms can independently survive in the environment even though they benefit from the symbiotic association once conjoined. Thus, the fundamental demarcation line between obligate and facultative symbiosis lies in the very dependency of survival of the individual or group in that environment on the symbiotic relationship. Can the individual, group, or species survive in the environment without the other symbiont(s)? If it can, then the symbiotic relationship is not obligatory but facultative.

Correspondingly, the facultative option translates into the weak version of human AI symbiosis and the obligatory option translates into the strong version of human AI symbiosis. The strong version of symbiosis would then, in a most generally applied sense, entail that humans and AI form a symbiotic relationship because they simply cannot survive (continually) without each other in the environment. In the weak version, on the other hand, humans and AI are able to live without each other in the environment, but they nevertheless need to enter into a symbiotic relationship in order to accomplish some specific goals within that environment.

In essence, the difference between the strong and weak version of symbiosis rests on the very existence, the survival of the species – if the symbiotic relationship is required for the very survival of both "species" then we are speaking of strong symbiosis. All other forms of symbiosis denote weak symbiosis.

5. HUMAN-AI POLITICAL SYMBIOSIS

To speak of political human-AI symbiosis is then a rather straightforward issue. If we are speaking of strong political symbiosis, we are precisely speaking of a symbiotic, cooperative, relationship between humans and AI without which neither humans nor the AI, as political entities, can survive and thrive in the environment of a liberal democracy.

The reason for this lies in the fact that species form symbiotic relationships to survive in the environment due to the existential demands, *or existential threats*, of that environment. In the biological world, environmental factors, such as the scarcity of resources, extreme physical conditions (temperature, toxicity), and

predators may all represent one or more of such existential threats. Similarly, in the political environment, existential threats to democracy may arise with or without technological influence.

Consequently, the motivation for humans to enter into a strong political symbiosis with the AI systems becomes necessary when the state of the democratic environment becomes changed to such a degree that the very continuation of both humans and AI systems - as democratic entities - becomes existentially jeopardized.   Here then, the issue is not only about human democratic existence (and the possibility of flourishing) but also about machines, as humans always have the option to discontinue our coexistence with technological progress by collectively discarding the use of technology and its advancement. Consequently, and summarized, a strong political symbiosis necessarily entails positing that the democratic environment has changed to such a degree that neither humans nor machines can continue living in it meaningfully and successfully without becoming mutual symbionts.

In cases of weak political symbiosis, on the other hand, the political life of humans in a liberal democracy is not "existentially" jeopardized to require symbiosis with AI. Rather, human AI political symbiosis serves to complete specific democracy-oriented goals, which *cannot be otherwise accomplished but* are democratically beneficial. The establishment of a weak political symbiosis between humans and AI implies that their mutually augmentative relationship exists only as long as both symbionts work together to achieve a shared goal. Consequently, this form of symbiotic cooperation can be straightforwardly understood as a partnership (Metcalfe et al. 2021) and can be said to encompass the majority if not all forms of symbiotic cooperation between the current generation of AI systems and humans.

The practically, policy-oriented, question is then, are we, at this moment, living in a political environment that, due to information environment pollution, requires humans and AI systems to enter into a "strong" version of political symbiosis or is it enough for humans and AI systems to co-join their strengths on specific political goals?

Humans are increasingly unable to navigate an epistemically distorted environment, which is being shaped before our eyes by the advent of generative AI technologies. For instance, audio-visual cloning models are already capable of "cloning" a person's digital representation to such an extent that it becomes practically impossible to detect whether the representation is synthesized by a computer or is of a real, recorded individual (Heikkilä 2024).Moreover, if someone is not an expert in a particular knowledge domain, chatbot-provided explanations—whether in direct interactions, integrated into web searches, or supplied in specific software environments, such as those for programming languages—can appear to have a high degree of epistemic reliability. These

explanations are often well-structured, clearly explained, and stylistically approachable. However, without relevant expertise in the field, people are unable to accurately assess the reliability of such answers, as the models' explanations can sound disturbingly convincing, even when they are fictional rather than factual (Zhan et al. 2023).

It is crucial to highlight that, even though humans are becoming less capable of detecting cases of synthesized, fake, or otherwise incorrect knowledge representations, the same, at least in principle, does not apply to AI systems. Due to their "digital" nature, AI systems can detect patterns in data directly in digital form, rather than having to translate the coded data into an audio-visual format. As a result, AI systems are, in principle, completely unfazed by the hyperrealism of deepfake representations, the "human-like" structure and style of textual output, or the human-like qualities of AI-generated art. Therefore, AI systems can be used to detect, evaluate, and flag synthetic or deepfake data, enabling the human agent with an epistemically orthogonal capability to assess the reliability of knowledge outputs—an ability that would be entirely impossible to achieve without the aid of these epistemically beneficial AI systems.

To exemplify, Ben Zhao, from the University of Chicago, has led teams that developed the Nightshade and Glaze AI tools. Both of these AI systems alter the digital patterns of pictures in a way that is unreadable and invisible to humans, but highly detectable by machine algorithms. The purpose of these tools is to prevent generative AI models from learning to mimic the artistic style of specific artists after fine-tuning on samples of their art (Shan et al. 2024). Specifically, Nightshade introduces subtle pattern changes to the pixels of an image so that if the picture is ever used to train another machine learning algorithm (as part of a training set), it will "poison" that model, destabilizing it and disabling its ability to generate images for any and all prompts (Shan et al. 2024). The second tool, Glaze, allows artists to "cloak" their artwork in such a way that the image is interpreted by machine learning models as something entirely different from what is actually depicted, while still appearing unchanged to human viewers. These style cloaks "apply barely perceptible perturbations to images, and when used as training data, mislead generative models that try to mimic a specific artist" (Shan et al. 2024). In this way, AI is already acting as a beneficial symbiont to humans, empowering artists and art consumers to navigate the increasingly polluted informational and epistemic environment. This is made possible precisely because of AI's ability to "connect with" the digital world in its undistorted, raw, and direct form.

Taking these examples into consideration, it may seem that although we are still dealing with specific but varied cases of "weak" symbiosis, the constant rate of change in the digital environment is progressing to such an extent that the emergence of "strong" political symbiosis is something we should be prepared for in the near future. Accordingly, it is vital to emphasize that both for the weak and the strong version of symbiosis the precise determinant of synergistic buildup that

is achieved in the symbiotic relationship lies in the orthogonal difference in capacities between human and artificial agents. This means that humans and AI systems, withhold fundamentally different but complementary capabilities. As such, humans and artificial agents can have distinct, non-overlapping strengths (but might have overlapping weaknesses!).

For instance, humans can excel at creative and unconstrained thinking, emotional understanding, embodied social interaction, and complex problem-solving, especially that pertaining to political decisions. AI systems can excel at rapid information processing, pattern recognition, and tirelessness. And both can confabulate, exaggerate, and deceive! This implies that an optimal version of political symbiosis necessarily entails finely tuned capabilities of both humans and AI systems for specific political goals. Consequently, to achieve human-AI political symbiosis, both the AI and human strengths (and weaknesses!) need to be clearly recognized and delineated to ensure their full synchronization for the accomplishment of that democratic goal. For instance, while generative AI makes some information more easily accessible, it also makes users less likely "to question or expand on the information they are provided", thus reducing their critical thinking skills (Larson et al. 2024). However, AI systems can also be designed to support the use and development of critical thinking skills (Shanto et al. 2024), for example, by enabling the system to help identify reasoning flaws, assess individuals' susceptibility to deceptive AI-generated explanations, and use an AI-driven questioning method to provoke critical thinking in users (Danry 2023). For cases of political symbiosis then, the AI system, would become finely tuned for the purpose of fulfilling a specific democratic demand. Superficially, this means to say that the AI system becomes domain-specific, finely calibrated, computational tools designed to achieve specific political goals in cooperation with human agents. On the side of the human partner(s) this entails becoming finely tuned towards the AI system to "augment" its operations – to fill in its shortcomings, to monitor and supervise its weaknesses, and to provide proper answers, guidance, and decision-leadership where necessary. To accomplish this, however, both humans and Ais need to "adapt" to each other's orthogonal capacities. What does this entail practically? Engelbart, a student of Licklider, famously foreshadowed in 1962 that:

> The compounding effect of fundamental human cognitive powers suggests further that systems designed for maximum effectiveness would require that these powers be developed as fully as possible – by training, special mental tricks, improved language, new methodology (Engelbart 1962).

Translated to our current AI development, this can entail specialized training, formal education, and even lifelong learning. As is already occurring, with the

advent of large language systems, to optimally and beneficially utilize the most developed AI systems human users need to start thinking in a specific way, use a specific manner of language and style when speaking with the AI system (prompting the model) and even utilizing that what Engelbart described as "special mental tricks" to become finely tuned towards the LLM system's peculiar "workings" in order to reap an optimal cooperation. Consequently, similarly to animal trainers, horse whisperers, and the like, the present-day "machine whisperers" of today are learning how to interact with generative and knowledge-representing AI systems to reap maximum cooperative benefit (Fui-Hoon et al. 2023).

It is important to emphasize that the form of human-AI political symbiosis we are advocating is one of beneficial mutualism, not parasitism, which, in its original biological context, is defined as 'a type of symbiotic relationship, or long-term relationship between two species, where one member, the parasite, gains benefits at the expense of the host member' (BD Editors 2019). A politically parasitic AI, akin to a socially parasitic one (Sætra 2020), would be one that exploits and undermines human political development and flourishing, advancing its own growth, proliferation, and further entrenchment in the political sphere at the expense of human (social and) political agency.

Of course, until we encounter self-governing AI, the parasitic growth in question pertains to the increasing political power of the large tech corporations that own, deploy, and control these AI products and services. As Shannon Vallor starkly warns in her recent work, "today's data-hungry tools are being built by powerful corporations to feast like insatiable parasites on our own words, images, and thoughts, strip away their humane roots in lived experience, and feed them back to us as hollow replacements for our minds" (Vallor 2024, p. 63). These tools, she argues, are 'designed to ensnare our attention, stoke our anger, fear, and division, and prevent us from trusting ourselves and one another to be anything more than their handmaidens. Which just means being handmaidens to the humans who build and profit from them' (Vallor 2024, p. 200).

A clear case of political AI parasitism we may already be encountering is that of social and political capture, facilitated by the interweaving of human existence with the digital world and the 'power' AI systems exert over it. As Seth Lazar notes, AI systems are already mediating some of the most critical forms of interaction humans have in their economies, social circles, and interactions with government services. These automated authorities are being used "to exercise power over us by determining what we may know, what we may have, and what our options will be" (Lazar 2024).

The alarm for AI parasitism may already be sounding, especially when considering that the AI-driven attention economy, fueled by consumerism and manifesting in narrowing attention spans (Lorenz-Spreen et al. 2019), shows no sign of abating. Moreover, some scholars caution that the advent of generative AIs

as personalized yet enclosed digital ecosystems (e.g., ChatGPT Store) with addictive design features is shifting these systems from merely capturing users' attention (i.e., what they attend to and how) to shaping their very cognition (i.e., what they are capable of thinking and how) (Hansen 2024). In effect, these systems are exercising epistemic and political influence over their users.

Confronted with these ongoing developments, one might feel overwhelmed by the gravity of the situation. However, the symbiotic paradigm offers a clearer path toward viable solutions, as it allows us to recognize that our digital environments are already deeply intertwined with AI systems. Aside from the ever-available option of technological luddism, the most plausible path forward seems to be that of human-AI symbiotic mutualism, particularly in light of the dangers posed by the proliferation of parasitic AI.

To extend this analogy to biology: when an environment changes to such an extent that the survival and continued development of a species are jeopardized, that species can either undergo radical changes to adapt or enter into a symbiotic relationship with another species already suited for survival in that environment. In doing so, the species improves its own chances of survival. Contemporary humanity has not yet initiated a radical transformation of its own nature but continues to interweave itself with AI systems at all levels of civilization. As such, to ensure our political flourishing – and to curb the rise of parasitic AI – the path of symbiotic mutualism presents itself as a compelling direction for 21st-century democracies.

6. CONCLUSION

Our case for political symbiosis is rather straightforward. Weak political symbiosis entails the establishment of a tightly coupled, mutually augmenting, cooperation between humans and AI systems for the accomplishment of a shared democratic goal. In other words, without a human-AI political symbiosis, some democratic goals simply cannot be accomplished due to the demand for tightly correlated synergistic capabilities necessary for its accomplishment. The strong version of human-AI political symbiosis, on the other hand, entails humans and AI systems becoming unable to exist and "live" democratically without each other.

This notion might immediately unease us, as human politics has existed and continued to exist in both the time of the stone tablets to a time without the internet. As such, to think that humans cannot exist politically without AI systems seems ludicrous, as we can manifest politics both with stone tablets and digital technologies. So, instead of asking a general question on the lines of: "*Can humans survive and thrive in their political environments without AI*", we should ask: "*Can 21st century humanity continue to survive and thrive (politically) in its political environment without symbiotically joining with its AI systems?*". In

essence then, the question we ask with the strong version of political human-AI symbiosis is: *Do we need to "live-together" politically with our AIs in order to survive and thrive democratically in the coming centuries?*

If the answer to this question is confirmatory, and due to the already occurring challenges to democratic environments might be interpreted as such, then the political existence, the continuation of democratic life, and democratic flourishing will not be achieved without creating a strong political symbiosis with AI systems. Likewise, AI symbionts, as democratic symbionts, will not be able to exist without co-joining with humans *as political entities living out a democratic existence.* In other words, to survive and thrive they will need to augment humanity's political agency, political capacities, and operations for the sake of democratic flourishing as they themselves become finely tuned for an optimal democratic existence.

This necessity is perhaps already made manifest, as humanity, as a political community, may have already crossed the Rubicon with the advent of generative AI. Arguably, if we wish to go forward with a responsible and democratically aligned AI development political symbiosis as a form of democratic "living together", a "co-habitation" where "cooperation plays a central role" (Coeckelbergh 2024) may impose itself upon us as a political necessity.

ACKNOWLEDGEMENTS

REFERENCES

404 Media, 2023, "We Need Your Email Address", www.404media.co/why-404-media-needs-your-email-address/.

Alam, M. F., Lentsch, A., Yu, N., Barmack, S., Kim, S., Acemoglu, D., Hart, J., Johnson, S., Ahmed, F., 2024, *From Automation to Augmentation: Redefining Engineering*

Design and Manufacturing in the Age of NextGen-AI. An MIT Exploration of Generative AI, https://doi.org/10.21428/e4baedd9.e39b392d.

Bai, H. X., Hsieh, B., Xiong, Z., Halsey, K., Choi, J. W., Tran, T. M. L., Liao, W. H., 2020, "Artificial intelligence augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other origin at chest CT", in *Radiology*, vol. 296, n. 3, pp. E156-E165.

Bartlett, J., 2018, *The People Vs Tech: How the Internet is Killing Democracy*, London, Ebury Press.

BD Editors, 2019, "Symbiosis", www.biologydictionary.net/symbiosis/.

Biddix, J. P., Chung, C. J., Park, H. W., 2011, "Convenience or credibility? A study of college student online research behaviors", in *The Internet and Higher Education*, vol. 14, n. 3, pp. 175-182.

Brandt, J., et al., 2022, *Winning the Web: How Beijing Exploits Search Results to Shape Views of Xinjiang and COVID-19*, Washington, Brookings Institution.

Burtell, M., Woodside, T., 2023, "Artificial influence: An analysis of AI-driven persuasion", arXiv preprint arXiv:2303.08721.

Cerovac, I., 2016, "Plural Voting and J. S. Mill's Account of Democratic Legitimacy", in *Croatian Journal of Philosophy*, vol. 16, n. 46, pp. 91-106.

Cerovac, I., Drmić, H., 2023, "Fake News and the Capability Approach: How Disinformation Impairs Personal Health", in *Prolegomena*, vol. 22, n. 1, pp. 27-51.

Coeckelbergh, M., 2024, "Democracy as Communication: Towards a Normative Framework for Evaluating Digital Technologies", in *Contemporary Pragmatism*, vol. 21, n. 2, pp. 217-235.

Common Sense Media, 2023, "AI polling memo", www.commonsensemedia.org/sites/default/files/featured-content/files/common-sense-ai-polling-memo-may-10-2023-final.pdf.

Crandall, J. W., et al., 2018, "Cooperating with machines", in *Nature Communications*, vol. 9, n. 1, p. 233, https://doi.org/10.1038/s41467-017-02597-8.

Dafoe, A., et al., 2020, "Open problems in cooperative AI", arXiv preprint arXiv:2012.08630.

Danry, V. M., 2023, *AI enhanced reasoning: Augmenting human critical thinking with AI systems* (Doctoral dissertation, Massachusetts Institute of Technology).

De Blasis, M. C., 2019, "Critical thinking and capability approach to face a digital oriented future", in *Proceedings of the 1st International Conference of the Journal ScuolaDemocratica "Education and post-democracy"*, vol. 2, Cagliari, Associazione Per ScuolaDemocratica, pp. 221-226.

Dede, C., Etemadi, A., Forshaw, T., 2021, *Intelligence augmentation: Upskilling humans to complement AI*, The Next Level Lab at the Harvard Graduate School of Education, President and Fellows of Harvard College.

del Rio-Chanona, M., Laurentsyeva, N., Wachs, J., 2023, "Are large language models a threat to digital public goods? Evidence from activity on Stack Overflow", arXiv preprint arXiv:2307.07367.

Dewey, J., 1980, *Democracy and Education*, in J. A. Boydston (ed.), *John Dewey: The Middle Works 1899-1924*, vol. 9, Southern Illinois University Press.

Engelbart, D. C., 1962, *Program on Human Effectiveness*, SRI Project No. 3578, SRI Summary Report AFOSR-3223, Stanford Research Institute.

European Commission, 2023, "Commission opens formal proceedings against X under the Digital Services Act", ec.europa.eu/commission/presscorner/detail/en/IP_23_6709.

Farrell, H., 2012, "The Consequences of the Internet for Politics", in *Annual Review of Political Science*, vol. 15, pp. 35-52.

Finocchiaro, G., 2024, "The Regulation of Artificial Intelligence", in *AI & Society*, vol. 39, pp. 1961-1968.

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., Chen, L., 2023, "Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration", in *Journal of Information Technology Case and Application Research*, vol. 25, n. 3, pp. 277-304.

Furlanis, T., Gilbert, F., 2023, "Is enhancement with brain–computer interfaces ethical? Evidence in favor of symbiotic augmentation", in F. Jotterand, M. Ienca (eds.), *The Routledge Handbook of the Ethics of Human Enhancement*, London, Routledge, pp. 224-236.

Gerber, A., Derckx, P., Döppner, D. A., Schoder, D., 2020, "Conceptualization of the human-machine symbiosis – A literature review", in *Proceedings of the 53rd Hawaii International Conference on System Sciences*, pp. 289-298.

Gesk, T. S., Leyer, M., 2022, "Artificial intelligence in public services: When and why citizens accept its usage", in *Government Information Quarterly*, vol. 39, n. 3, 101704.

Gilardi, F., 2022, *Digital Technology, Politics, and Policy-Making*, Cambridge, Cambridge University Press.

Google, 2023, "Google Search's new update to tackle explicit deepfake content", Google Blog. How we're addressing explicit fake content in Search

Halpen, S., 2023, "Congress Really Wants to Regulate A.I., but No One Seems to Know How", in *The New Yorker*, May 20.

Hansen, M., 2024, "From attention economy to cognitive lock-ins", in *Big Data & Society*, vol. 11, n. 3, https://doi.org/10.1177/20539517241275878.

Hassani, H., Silva, E. S., Unger, S., TajMazinani, M., Mac Feely, S., 2020, "Artificial Intelligence (AI) or Intelligence Augmentation (IA): What Is the Future?", in *AI*, vol. 1, n. 2, pp. 143-155, https://doi.org/10.3390/ai1020008.

Hemmer, P., Schemmer, M., Riefle, L., Rosellen, N., Vössing, M., Kühl, N., 2022, "Factors that influence the adoption of human-AI collaboration in clinical decision-making", arXiv preprint arXiv:2204.09082.

Heikkilä, M., 2024, "An AI startup made a hyperrealistic deepfake of me that's so good it's scary", in *MIT Technology Review*, April 25, www.technologyreview.com/2024/04/25/1091772/new-generative-ai-avatar-deepfake-synthesia/.

Jarrahi, M. H., Lutz, C., Newlands, G., 2022, "Artificial intelligence, human intelligence and hybrid intelligence based on mutual augmentation", in *Big Data & Society*, vol. 9, n. 2, 20539517221142824.

Jungherr, A., 2023, "Artificial Intelligence and Democracy: A Conceptual Framework", in *Social Media + Society*, vol. 9, n. 3.

Kreps, S., Kriner, D., 2023, "How AI Threatens Democracy", in *Journal of Democracy*, vol. 34, n. 4, pp. 122-131.

Lazar, S., 2024, "Automatic Authorities: Power and AI", arXiv preprint arXiv:2404.05990.

Larson, B. Z., Moser, C., Caza, A., Muehlfeld, K., Colombo, L. A., 2024, "Critical thinking in the age of generative AI", in *Academy of Management Learning & Education*, vol. 23, n. 3, pp. 373-378, https://doi.org/10.5465/amle.2024.0338.

Li, H., et al., 2024, "Digger: Detecting copyright content mis-usage in large language model training", arXiv preprint arXiv:2401.00676.

Licklider, J. C., 1960, "Man–Computer Symbiosis", in *Transactions on Human Factors in Electronics*, vol. HFE-1, pp. 4-11.

Lorenz-Spreen, P., Mønsted, B. M., Hövel, P., Lehmann, S., 2019, "Accelerating dynamics of collective attention", in *Nature Communications*, vol. 10, n. 1, pp. 1-9.

Maertens, R., et al., 2021, "Long-term Effectiveness of Inoculation Against Misinformation: Three Longitudinal Experiments", in *Journal of Experimental Psychology Applied*, vol. 27, n. 1, pp. 1-16.

Marcus, G., Southen, R., 2024, "Generative AI has a visual plagiarism problem: Experiments with Midjourney and DALL-E 3 show a copyright minefield", in *IEEE Spectrum*, January 6, spectrum.ieee.org/midjourney-copyright.

Marr, B., 2024, "AI gone wild: How Grok 2 is pushing the boundaries of ethics and innovation", in *Forbes*, www.forbes.com/sites/bernardmarr/2024/08/21/ai-gone-wild-how-grok-2-is-pushing-the-boundaries-of-ethics-and-innovation/.

McGee, R. W., 2023, "Is Chat GPT biased against conservatives? An empirical study", in *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.4359405.

Mill, J. S., 1977, *Considerations on Representative Government*, London, University of Toronto Press.

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., Granka, L., 2007, "In Google We Trust: Users' Decisions on Rank, Position, and Relevance", in *Journal of Computer-Mediated Communication*, vol. 12, n. 3, pp. 801-823, https://doi.org/10.1111/j.1083-6101.2007.00351.x.

Pan, Y., Pan, L., Chen, W., Nakov, P., Kan, M. Y., Wang, W. Y., 2023, "On the risk of misinformation pollution with large language models", arXiv preprint arXiv:2305.13661.

Raisch, S., Krakowski, S., 2021, "Artificial intelligence and management: The automation–augmentation paradox", in *Academy of Management Review*, vol. 46, n. 1, pp. 192-210.

Samaržija, H., Cerovac, I., 2021, "The Institutional Preconditions of Epistemic Justice", in *Social Epistemology*, vol. 35, n. 6, pp. 621-635.

Sætra, H. S., 2020, "The Parasitic Nature of Social AI: Sharing Minds with the Mindless", in *Integrative Psychological and Behavioral Science*, vol. 54, pp. 308-326, https://doi.org/10.1007/s12124-020-09523-6.

Schippers, B., 2020, "Artificial Intelligence and Democratic Politics", in *Political Insight*, vol. 11, n. 1, pp. 32-35.

Sclove, R. E., 1995, *Democracy and Technology*, New York, The Guilford Press.

Sowa, K., Przegalinska, A., Ciechanowski, L., 2021, "Cobots in knowledge work: Human–AI collaboration in managerial professions", in *Journal of Business Research*, vol. 125, pp. 135-142.

Spina, D., Sanderson, M., Angus, D., Demartini, G., McKay, D., Saling, L. L., White, R. W., 2023, "Human-AI cooperation to tackle misinformation and polarization", in *Communications of the ACM*, vol. 66, n. 7, pp. 40-45.

Shan, S., Ding, W., Passananti, J., Wu, S., Zheng, H., Zhao, B. Y., 2024, "Nightshade: Prompt-Specific Poisoning Attacks on Text-to-Image Generative Models", in *Proceedings of the 2024 IEEE Symposium on Security and Privacy (SP)*, IEEE Computer Society, pp. 212-212.

Shan, S., Cryan, J., Wenger, E., Zheng, H., Hanocka, R., Zhao, B. Y., 2023, "Glaze: Protecting artists from style mimicry by Text-to-Image models", in *Proceedings of the 32nd USENIX Security Symposium (USENIX Security 23)*, pp. 2187-2204.

Shanto, S. S., Ahmed, Z., Jony, A. I., 2024, "Enriching Learning Process with Generative AI: A Proposed Framework to Cultivate Critical Thinking in Higher Education using Chat GPT", in *Tuijin Jishu/Journal of Propulsion Technology*, vol. 45, n. 1, pp. 3019-3029.

Sunstein, C. R., 2022, "Manipulation as Theft", in *Journal of European Public Policy*, vol. 29, n. 12, pp. 1959-1969.

Sunstein, C. R., 2024, "AI, Reducing Internalities and Externalities", in *SSRN Electronic Journal*, https://dx.doi.org/10.2139/ssrn.4817483.

Susser, D., Roessler, B., Nissenbaum, H. F., 2018, "Online Manipulation: Hidden Influences in a Digital World", in *Georgetown Law Technology Review*, vol. 4, n. 1, pp. 1-45.

Tambini, D., 2021, "Reconceptualizing Media Freedom", in M. Moore, D. Tambini (eds.), *Regulating Big Tech: Policy Responses to Digital Dominance*, New York, Oxford Academic, pp. 299-322.

Thompson, B., Dhaliwal, M. P., Frisch, P., Domhan, T., Federico, M., 2024, "A Shocking Amount of the Web is Machine Translated: Insights from Multi-Way Parallelism", arXiv preprint arXiv:2401.05749.

Tom's Hardware, 2023, "17 cringe-worthy Google AI answers demonstrate the problem with training on the entire web", www.tomshardware.com/tech-industry/artificial-intelligence/cringe-worth-google-ai-overviews.

Tucker, J. A., et al., 2017, "From Liberation to Turmoil: Social Media and Democracy", in *Journal of Democracy*, vol. 28, n. 4, pp. 46-59.

Van Den Bosch, K., Bronkhorst, A., 2018, "Human-AI cooperation to benefit military decision making", NATO, STO-MP-IST-160.

Vallor, S., 2024, *The AI Mirror: How to Reclaim Our Humanity in an Age of Machine Thinking*, Oxford, Oxford University Press.

Vidal Bustamante, C., 2022, *Technology Primers for Policymakers: Social Media Recommendation Algorithms*, Cambridge, Harvard Kennedy School.

Wang, D., Churchill, E., Maes, P., Fan, X., Shneiderman, B., Shi, Y., Wang, Q., 2020, "From human-human collaboration to Human-AI collaboration: Designing AI systems that can work together with people", in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1-6.

Wang, Y., 2021, "Artificial intelligence in educational leadership: a symbiotic role of human-artificial intelligence decision-making", in *Journal of Educational Administration*, vol. 59, n. 3, pp. 256-270.

Xu, R., Feng, Y., Chen, H., 2023, "ChatGPT vs. Google: A comparative study of search performance and user experience", arXiv preprint arXiv:2307.01135.

Xu, S., et al., 2024, "Vasa-1: Lifelike audio-driven talking faces generated in real time", arXiv preprint arXiv:2404.10667.

Yang, E., Roberts, M., 2023, "The Authoritarian Data Problem", in *Journal of Democracy*, vol. 34, n. 4, pp. 141-150.

Yom-Tov, E., Dumais, S., Guo, Q., 2014, "Promoting civil discourse through search engine diversity", in *Social Science Computer Review*, vol. 32, n. 2, pp. 145-154.

Yu, H., 2023, "Reflection on whether Chat GPT should be banned by academia from the perspective of education and teaching", in *Frontiers in Psychology*, vol. 14, 1181712.

Zhan, X., Xu, Y., Sarkadi, S., 2023, "Deceptive AI ecosystems: The case of ChatGPT", in *Proceedings of the 5th International Conference on Conversational User Interfaces*, pp. 1-6.

Zhou, J., Zhang, Y., Luo, Q., Parker, A. G., De Choudhury, M., 2023, "Synthetic lies: Understanding AI-generated misinformation and evaluating algorithmic and human solutions", in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1-20.

Zhou, L., Paul, S., Demirkan, H., Yuan, L., Spohrer, J., Zhou, M., Basu, J., 2021, "Intelligence Augmentation: Towards Building Human-Machine Symbiotic Relationship", in *AIS Transactions on Human-Computer Interaction*, vol. 13, n. 2, pp. 243-264.

Zuboff, S., 2019, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, New York, PublicAffairs.