**The Question of AI and Democracy: Four Categories of AI Governance**

Sung Jun Han

forthcoming in *Philosophy & Technology*

**Abstract**

This paper examines the integration of Artificial Intelligence (AI) into democratic governance, focusing on the tension between democracy’s epistemic shortcomings—often manifested as voter ignorance—and AI’s capacity to improve decision-making. Building on the Input-Process-Output (IPO) model, the paper distinguishes AI applications into four categories based on the democratic source of their inputs (i.e., whether they originate from the citizenry) and the binding nature of their outputs (i.e., whether AI decisions carry legal or authoritative weight). Each category—democratic binding AI, undemocratic binding AI, democratic unbinding AI, and undemocratic unbinding AI—is then evaluated against core democratic elements: inclusive and equal participation, quality of decisions, deliberation, and the autonomy

of citizens to set the political agenda. While some undemocratic binding AI risks centralizing power into the hands of a few, certain forms of AI, such as AI advisers, AI delegates with deliberative consent, and AI nudger, can enhance democratic processes by helping citizens overcome epistemic barriers, refine their political views, and participate more effectively in governance. The paper concludes that carefully implemented AI has the potential to enhance democratic governance while preserving its core ideals.

The Question of AI and Democracy: Four Categories of AI Governance

 Recent advancements of Artificial Intelligence (AI) projects its becoming an integral part of our life for its unequaled epistemic abilities in processing vast data sets, detecting patterns, or aggregating dispersed information, raising profound questions about its role in shaping our political judgment. Simultaneously, democracy faces challenges in good governance due to the epistemic limit of citizens in informed monitoring of politicians, often called the problem of voter ignorance. These seemingly disparate phenomena prompt critical questions like “Can AI help alleviate voter ignorance?” and “If so, what should the democratic application of AI look like?”

 This paper argues that democracy can benefit from the integration of AI, particularly in addressing the epistemic challenges posed by voter ignorance. While the deliberative model of democracy is not without criticism, particularly from philosophers like Chantal Mouffe, Jacques Rancière, and Iris Marion Young, the deliberative approach remains uniquely valuable in contexts involving complex epistemic challenges, like voter ignorance, precisely because it fosters informed consensus through structured reasoning. In the deliberative and, more recently, epistemic terms, a key function of democratic governance is not only to uphold the ideal of self-governance among political equals but also to ensure decisions of good quality (Estlund, 2009; Talisse, 2019). However, voter ignorance undermines the ability of democracies to make well-informed decisions (Guerrero, 2024). AI, with its epistemic power in memorizing contents (Panigrahi et al., 2018), predicting accurately (Agrawal et al. 2019), processing and analyzing vast amounts of data consistently over different situations (LeCun et al., 2015), offers a potential answer to this problem as an epistemic tool.

 Yet, AI-driven tools are far from perfect for the issues of transparency (Pasquale, 2015), accountability (Diakopoulos, 2016), manipulation (Susser et al., 2019), publicity (Lazar, 2024) and fairness of algorithmic decisions (Mehrabi et al., 2021). While these discussions are certainly crucial, I present a framework of the Input-Process-Output (IPO) model to distinguish between the questions of AI governance and the issues of algorithms. In brief, algorithmic issues are questions of the process, while AI governance is more oriented toward the democratic nature of input and binding nature of output. In this way, I keep the algorithmic issues aside and discuss representative models of AI governance in democracy, assuming that reasonably acceptable criteria for regulating algorithms and their designers could be established.

 Based on the IPO model, this paper distinguishes four categories of applications of AI in democratic decision-making, focusing on the democratic nature of inputs and the binding aspect of outputs. These categories–democratic binding AI, undemocratic binding AI, democratic unbinding AI, and undemocratic unbinding AI–are evaluated against key democratic questions: (1) *who* participates in decision-making (inclusion and equality), (2) *how* decisions are made (deliberation and quality), and (3) *what* issues are addressed (autonomy). The paper argues that unbinding AI and some carefully limited binding AI can properly answer the questions while effectively mitigating voter ignorance.

In the first section of the paper, I examine these key questions of democracy. These questions call attention to the central elements of democratic governance of empowered and equal participation, deliberative decision-making with good epistemic reasons, and the autonomy of ordinary citizens in setting the political agenda. By situating these tasks in the context of AI-driven decision-making, I highlight a central tension: Can AI enhance the epistemic quality of democratic decisions without sacrificing those core elements of democratic governance?

Then, I turn to the practical problem of categorizing AI’s role in decision-making. Using the IPO model, I identify how the sources of input (democratic or not) and the authority of outputs (binding or unbinding) combine to produce four broad categories of AI applications in democratic decision-making. This categorization captures the basic distinction between different models of employing AI in the context of democracy.

In the final section, I aim to clarify which types of AI can be integrated into democratic decision-making without significantly compromising core elements of democratic governance. Specifically, I argue that a particular model of democratic binding AI, which I call AI Delegate with Deliberative Consent, and unbinding AI models can be introduced in democratic decision-making processes. Note that within each category, there exists a wide range of potential AI applications. In this paper, I will focus on representative cases within each category to illustrate the key issues and arguments, rather than attempting to provide an exhaustive analysis of all possible variations. Additionally, the primary contribution of this paper is intended to be theoretical illustrations of democratically acceptable AI agents. For this reason, though it engages in real-world examples of AI when they exist, it primarily aims to illustrate what each AI agent would look like–even based on fictional or hypothetical cases–setting aside empirical analysis of each of them. Since the goal of the paper is to discuss the theoretical frameworks for categories of AI in terms of democratic decision-making, it may not appear as a limit but as a motivation for further empirical research.

1. Setting the Criteria: The Core Elements of Democratic Governance

One of the defining questions of our time is how to deal with artificial intelligence (AI) in various levels of our society. Governance is not an exception, asking the potential role of AI in decision-making. Yet, people worry that AI would deteriorate rather than enhance democratic governance due to the superior epistemic power in collecting and processing the vast amount of information, making most citizens feel unduly diffident in their epistemic competence (Hendrycks 2023; Zerilli et al. 2019b). This is a fair worry because rule by AI’s decisions contradicts democracy as a political system for self-governance. However, AI differs from the human as an epistemic agent, which requires the incorporation of AI into decision-making not in the form of replacement but supplement. The foremost difference between the human and the machine learning based AI is the mode of reasoning–human thought is largely sequential, but AI operates mainly with parallel processing (LeCun et al., 2015). As Matteo Pasquinelli (2020) argues, this leads to a problem: AI’s understanding—if we can call it like that—is shaped by the massive datasets it is trained on, which reflects not only its technical capacities but also the historical and material contexts of knowledge production. For example, ImageNet—an influential database used in deep learning—not only serves as a repository of images but also shapes research practices by embedding and perpetuating socio-epistemological biases (Denton et al., 2021). This illustrates that while AI’s ability to process large volumes of data is epistemically powerful, simply accumulating more data may not lead to the right kind of judgment. AI’s abilities must, therefore, be understood not only as fundamentally different from human reasoning but also inherently limited.

This limitation of AI can nonetheless support its supplementary role for human agents whose epistemic capacities constrain them to process relatively smaller datasets. In other words, while being cautious about the limit of AI, its different but powerful ability in processing massive data could help alleviate some of the epistemic problems of human decision-makers like voter ignorance by mitigating the cognitive burden of human agents.

Thus, our question should be how to employ AI in democratic decision-making without significantly compromising the elements that a democratic system must deliver for democratic governance. At its core, democracy involves three interrelated inquiries addressing the “who,” “how,” and “what” questions of governance: who participates in decision-making, how decisions are made, and what problems are prioritized and addressed. These projects call attention to inclusion, equality, quality of decisions, deliberation, and the alignment of governance with the concerns of the people.[[1]](#footnote-0)

Democracy begins with inclusivity: it is a political system where decision-making power is shared broadly among citizens rather than concentrated in the hands of a few. This inclusiveness is important because it reflects the idea that all individuals, regardless of their social, economic, ethnic, or cultural background, are equally entitled to participate in decisions that affect their lives. As Anderson (2009) and Young (2000) argue, democracy is built on the recognition of individuals as moral equals who deserve a say in collective decision-making.

Importantly, this inclusivity has an epistemic value. We expect democratic systems to reflect a comprehensive understanding of reality, one that incorporates the diverse perspectives and experiences of the demos (Guerrero 2024; Landemore 2020). In contrast, decision-making restricted to elites, however efficient or competent, risks bias and myopia. Elites may possess superior technical expertise, but they are not immune to self-interest or blind spots that distort their grasp of the broader social reality. For instance, policies crafted exclusively by economic elites often fail to address the needs of marginalized communities, as their lived experiences and perspectives are excluded from the process. This is why inclusion is more than a symbolic gesture; it is essential for capturing the complex and multifaceted nature of societal problems.

Yet, mere inclusion is insufficient. Democracy requires empowered inclusion, granting citizens enforceable rights to influence decisions, such as voting, representation, and veto powers (Warren, 2017). Without such empowerment, inclusion becomes performative. Consider a dictator who consults representatives from every socioeconomic group but retains unilateral decision-making authority. While the process may appear inclusive, the ultimate decisions reflect the dictator’s will, alien to the people, threatening genuine self-governance. Empowered inclusion ensures that citizens are not mere spectators in the political process but active participants with real influence.

Furthermore, participation must be substantially, not just formally, equal.[[2]](#footnote-1) While formal mechanisms like equal voting rights express equality, they often mask underlying power imbalances. Legislators, for instance, frequently shape policies in ways that disproportionately favor certain groups, such as trade unions, wealthy corporations, and other powerful people, over the general population (Baumgartner et al. 2009; Lessig 2011; Mutch 2014). Substantive equality requires mechanisms to counteract these imbalances, ensuring that no group’s voice dominates the democratic process. Without this, the principle of political equality remains superficial, and democracy risks becoming a system of domination rather than self-rule.

Yet, democracy is not only about participation; it must also produce decisions that people regard as proper answers to their problems. The quality of decisions is a central concern because citizens expect governance to benefit them by producing good policies and avoiding harmful ones (Estlund 2009; Brennan 2016). While the criteria for what constitutes a good decision—effectiveness, fairness, or societal well-being—may vary across contexts, the expectation remains consistent: democratic systems are meant to produce outcomes that improve the lives of the governed by finding good answers for their problems.

Nevertheless, quality alone is not enough. A key characteristic of democratic governance is that decisions are perceived as ours. This sense of ownership is what distinguishes democratic decisions from those made under authoritarian regimes or technocracies (Lafont 2020). Even highly effective decisions, if made without fair and broad participation of citizens, may fail to be regarded as democratically legitimate. For instance, the MY World survey, despite garnering over 7 million responses, revealed significant demographic imbalances—overrepresenting voices from low-HDI countries and underrepresenting those from high-HDI regions—which casts doubt on whether the resulting policy priorities truly reflect the diversity of global civil society and thereby on its democratic legitimacy (Gellers, 2016). Thus, democratic decision-making must produce qualified outcomes that people would regard as theirs.[[3]](#footnote-2)

One line of worry concerning the sense of ownership is majoritarianism. However inclusive and equal participation is, when decisions reflect only bare preferences of a majority, it would fail to induce the citizenry to identify the decisions as theirs. This is why democracy requires more than majoritarianism or preference aggregation. While the inclusion of all voices is a necessary condition, decisions must also emerge from deliberation—the exchange of reasons among citizens who recognize one another as equals (Cohen 1996; Habermas 1996).[[4]](#footnote-3) Deliberative processes ensure that decisions reflect reasoned agreement rather than the mere aggregation of preferences. Without deliberation, democratic systems risk devolving into majoritarian rule, where decisions represent the will of the majority but fail to consider the broader public interest or respect the equal standing of minorities. By fostering reasoned dialogue, deliberation transforms decision-making into a collective endeavor for producing decisions by the people, not a simple majority.

Finally, democracy must prioritize and address the problems that genuinely concern ordinary citizens who constitute the majority of the population. This speaks to the scope and focus of governance: whose issues are being addressed, and whose voices shape the political agenda?[[5]](#footnote-4) This has a long tradition of discussion in political theory between elitism and non-elitism. For example, Joseph Schumpeter’s (2008 [1942]) democratic elitism claimed for giving power to elites rather than the electorate–though people still hold the power to elect politicians–because complex political tasks like agenda-setting and policy-making cannot be handled by ordinary people who tend to lack relevant expertise. More recently, this line of thought has developed into what is called *epistocracy* where not all people but only knowledgeable people can be the rulers (Brennan, 2016).

Though elitist rule may have epistemic benefits from their holding expertise in politics, it can easily become an oligarchic rule where the interests of elites and dominant cultural majorities overly influence politics, meaning it fails to serve the broader population. Socioeconomically powerful groups, such as wealthy elites, often possess disproportionate influence over political agendas, steering governance toward issues that benefit them while sidelining the needs of marginalized communities. Similarly, ethnocultural majorities may impose their own priorities at the expense of minority groups. For democracy to be a system of self-governance, it must ensure that public problems are fairly identified and addressed in ways that reflect the diverse needs of the population.

Crucially, this requires that citizens have the power not only to participate in decision-making but also to shape the agenda itself.[[6]](#footnote-5) This capacity for agenda-setting, or autonomy, is essential for authentic self-governance. If citizens merely deliberate on problems set by external forces—be the elites, technocrats, or AI systems—their role is reduced to reactive decision-making rather than proactive self-determination. A CEO, for example, may have the authority to approve company initiatives, but if she cannot influence the projects of the company, her power is superficial and lacks authenticity. Similarly, citizens who cannot shape the political agenda cannot meaningfully govern themselves.[[7]](#footnote-6)

At this point, one might ask whether the introduction of AI into democratic systems is truly necessary. After all, representative democracy has historically managed to address the core elements of democratic governance—inclusive and equal participation in deliberative procedures to address what matters for citizens. The challenge, however, lies in the persistent problem of voter ignorance and other epistemic limits of citizens in obstructing these elements (Achen and Bartels 2016; Caplan 2008; Carpini and Keeter 1997; Somin 2016). Voter ignorance involves at least two key dimensions–rational and irrational aspects. Citizens often lack the information, resources, or expertise necessary to form well-reasoned opinions about complex policy issues, leading to poor judgment like uninformed voting, susceptibility to misinformation, and an inability to hold representatives accountable for their actions. In this case, their poor judgment is primarily due to their lack of relevant political information. In fact, this is not a malfunction of our rationality (Downs 1957).[[8]](#footnote-7) We are suboptimal information processors in the sense that we are not able to consider all relevant information in every decision-making situation (Giubilini and Savulescu 2018).

However, voter ignorance is not solely a matter of lacking information because it also involves irrational reasoning where emotional judgments and cognitive biases significantly shape decision-making processes. Even when voters have access to quality information, these psychological factors like partisan bias and social identity can distort how that information is processed and applied (Taber and Lodge 2006). For instance, deeply ingrained biases may cause voters to interpret neutral or even favorable information about an opposing political stance through a negative lens simply because it conflicts with their pre-existing attitudes. Thus, while aiming to lower the informational cost for voters with AI, it may not, at least in isolation, resolve the complex dynamics of irrational reasoning in voter ignorance.

Then, why do we need to care about voter ignorance? This is because these epistemic limitations undermine the quality of decisions, genuine deliberations, and identifying their true problems. For instance, when voters are uninformed or biased, their preferences may not reflect their genuine values, leading to policies that are misaligned with their true interest. Similarly, voter ignorance can hinder the deliberative processes necessary for achieving reasoned agreement among citizens. Likewise, uninformed citizens may misunderstand what really matters for them to improve their situation but be swayed by manipulation by powerful groups. These require us to think how to alleviate voter ignorance and if AI agents can have a role here.

Yet, AI holds promise primarily for alleviating the information gap rather than the irrational aspects of ignorance.[[9]](#footnote-8) By its capacity to process and synthesize massive datasets, AI can deliver personalized and accessible sets of information that lower the cost of getting informed. This is especially valuable in democratic contexts, where the sheer complexity of policy issues can overwhelm individual cognitive resources. Notably, several attempts have been made to apply AI to political matters in real-world settings. One prominent example is SAM (Semantic Analysis Machine), developed to represent and aggregate the political views of New Zealanders and was hoped to run for office (CNN Wire, 2017). SAM functioned as a chatbot on popular social media platforms like *Facebook* and *Twitter* (now known as *X*), where it engaged users in natural language conversations about policy issues and aimed to capture and generate public opinion.[[10]](#footnote-9)

Yet, while AI may enhance informational access, it does not automatically rectify the distortions caused by emotional and biased reasoning. Thus, AI can be an epistemically useful tool for its powerful epistemic capacities, but it may not treat all epistemic ailments of current democracies. In the sections that follow, I will outline four distinct categories of AI and explore which type can be integrated into democratic decision-making while performing these fundamental democratic elements.

1. Setting the Target: Four Categories of AI for Democratic Governance

One commonly used framework for describing a system is the Input-Process-Output (IPO) model. As its name suggests, the IPO model explains how a system transforms inputs into outputs through a central processing stage (Curry et al., 2006). This model is widely applicable to various decision-making systems. Consider, for example, the decision-making process in a representative democracy: citizens provide inputs (such as votes, petitions, and protests), representatives process these inputs by articulating them as politically relevant reasons in legislative forums (Pitkin 1967), and finally, legislators produce an output in the form of laws or policies that guide citizens’ lives.

A similar pattern emerges with AI decision-making. Take a chatbot as an everyday example: the user asks a question (input), the system’s algorithm examines its database and knowledge sources to identify the most relevant information (process), and then the chatbot delivers an answer (output). While the internal workings of each stage–input, process, and output–may differ significantly between human and AI systems, the IPO framework naturally lends itself to describing how AI makes decisions.

When we apply the IPO model to AI’s decision-making in democracy, it is worth noting that each stage raises distinct questions. First, the input stage invites us to ask who is providing those inputs. In the context of democracy, inputs are considered democratic if they originate from people who would be subject to the output. For example, if citizens collectively convey their preferences or opinions to an AI system that bases its decisions on these inputs, we can consider the input stage democratic. By contrast, if the inputs come from a specific subset of the demos but the output binds the whole demos, the input stage would not be democratic, even if the subsequent process and output stages remain unchanged from the democratic one.

Second, the process stage determines whether a system is considered AI-driven. The key factor here is the use of machine learning algorithms. While the level of complexity and the degree of human intervention in the algorithm may vary, it is the algorithmic nature of the process that typically characterizes an AI decision-making system (Ziewitz, 2016). If a system’s process does not involve a sufficiently complex algorithm—however that threshold is determined—then even if some algorithmic steps are present, we may hesitate to classify it as an AI decision-making system. Since the algorithm is the key factor in the process stage, many of the current debates on algorithmic issues, such as transparency, accountability, publicity, manipulation, and fairness, can be understood as questioning the process, instead of the input or output.

Third, the output stage raises questions about whether AI’s judgments are binding. For instance, when I ask ChatGPT what to wear today, I am not compelled to follow its suggestion. However, if a group collectively agrees that they will abide by whatever AI recommends, then the output becomes binding. At the political level, this might entail AI functioning as a lawmaker, producing decisions that citizens must follow. Conversely, if the AI’s outputs are purely advisory and carry no legal force, the AI remains a consultant rather than an authoritative decision-maker.

| Input | Output | Process |
| --- | --- | --- |
| Democratic | AI-driven | Binding |
| Undemocratic | Non-AI-driven | Unbinding |

Applying the IPO model, we can identify eight possible categories of AI decision-making–democratic or not, algorithmic or not, and binding or not. Each stage—input, process, and output—raises its own philosophically significant questions. In recent discourse, however, much attention has been devoted to the process stage, particularly the complexities of algorithms, including their fairness, publicity, transparency, and accountability (Ananny and Crawford, 2018; Binns, 2018; Hamon et al. 2021; Lazar 2022; Mehrabi et al., 2021; Pasquale, 2015; Zerilli et al., 2019a). These debates undeniably matter, as they discuss the issues of algorithmic processes.

Nevertheless, my argument primarily focuses on the input and output stages of the decision-making process rather than on the algorithmic process itself. Although it is true that these dimensions are interrelated—with the process stage potentially influencing both the inputs and outputs—temporarily setting aside the intricate issues of algorithm design allows us to concentrate on direct political effects of the input and output dimensions. For example, algorithmic decision-making may result in a *predictive lock-in* whereby the system pre-structures decisions toward predetermined outcomes and, in doing so, implicitly shapes or constrains the inputs it receives (Kitchin, 2014). In such cases, citizens’ preferences are filtered through algorithmic expectations, which calls into question the authenticity of the democratic inputs and, consequently, the legitimacy of any binding outputs. In a more extreme scenario, citizens might be reduced to mere data points classified into biased sociological categories. Such a process has the potential to transform political decision-making and communication in profound ways—much like television not only transmitted information but fundamentally reshaped political discourse by emphasizing visual spectacle and media-driven narratives.

While these concerns about process effects must be addressed before fully implementing AI agents in democratic decision-making, setting aside these issues for now creates a distinct analytical space. In this space, we can explore which types of AI agents might be incorporated into democratic decision-making without adopting an overly skeptical or agnostic stance. Consider, for example, a scenario in which an AI agent reduces contextually embedded citizen preferences to mere numerical data.[[11]](#footnote-10) This reduction risks neglecting the underlying socio-cultural motivations behind citizens’ views, treating delicate and meaningful inputs as if they were simple scalar values—reminding of the dictator who listens superficially to the people while making decisions based solely on her own agenda.

However, this potential distortion does not necessarily denounce the entire system. Instead, it highlights a critical need for additional democratic mechanisms—such as avenues for contestation, protest, or petition—to complement and challenge the outputs generated by the AI. Even if the initial data processing became reductionist, the AI system could incorporate iterative feedback loops or serve as a preliminary filter, with its outputs subsequently subject to review and further deliberation by individual users, citizen assemblies, or oversight committees. Though not a panacea, this would allow public oversight of any distortion in receiving and processing inputs while still utilizing AI’s capacity to process vast amounts of data.

Ultimately, my argument is not intended to downplay the significance of process-level issues. Rather, it contends that, through the adoption of established or emerging standards, regulatory frameworks, and technological innovations, it is possible to ensure that the process stage meets acceptable levels of transparency, fairness, publicity, and accountability. With such assurances in place, we can then focus on the input and output stages and pose critical democratic questions: How should we shape the inputs that feed into AI systems? Should AI be empowered to make laws on our behalf, and under what conditions would AI-generated decisions be genuinely democratic? In the following section, I will discuss representative cases of each category of AI in democratic decision-making.

1. Examining Each Category: Democratic Applications of Four Categories of AI
2. Undemocratic Binding

 An AI decision-making system that makes binding political decisions but does not receive inputs from citizens belongs to this category. An illustration of this category of AI can be found in Isaac Asimov’s short story, *Franchise*. Asimov illustrates a fictional future where elections are replaced by the internal processes of a supercomputer, Multivac. In the story, Multivac selects Norman Muller as the most representative American voter and asks questions to him based on which it determines the outcome of an election. Though Muller is the only voter who does not even vote in the traditional way of selecting among candidates, the electoral outcome is considered democratic because the candidate who has been elected by Multivac is the person who would have been elected if everyone had voted. In the last part of the story, Asimov writes “the sovereign citizens of the first and greatest Electronic Democracy had, through Norman Muller (through *him*!) exercised once again its free, untrammeled franchise.” (Asimov, 1955: 15)

 Building on the previously discussed elements of democratic governance—empowered inclusion, political equality, quality, deliberation, and autonomy to set agendas—it becomes clear that a system like Multivac in *Franchise* falls short of democratic governance. Democratic governance requires empowered inclusion, meaning that decision-making should involve the broad demos with legal measures to influence decisions—nearly everyone subject to the laws—rather than an isolated subset. In Asimov’s story, however, the entirety of the electorate’s role is usurped by a single individual, Norman Muller, chosen by Multivac. While the system claims Muller’s inputs are representative, actual widespread participation or control of Multivac’s decisions by the public is absent. The rest of the citizenry is put aside; they do not vote, deliberate, or shape the outcome. Even though the narrative presents this as ‘the will of the sovereign citizens,’ the reality is that citizens are neither actively involved nor empowered.[[12]](#footnote-11) The result is not governance by the people, but a narrow, top-down process that allows an AI and a single individual to speak for everyone, undermining the very idea of rule by the demos.

Moreover, democratic governance is not only about who is included but also about how inclusively and equally their voices are heard in making reasoned decisions. Democracy calls for collective decision-making, where power is not just concentrated in an executive or an elite few, and for deliberation, where citizens collectively reason about the best answers for their problems. While Multivac’s process might exhibit a form of deliberation with Muller—a conversation or interrogation that leads to policy outcomes—this is not collective deliberation. There is no chance of learning the views of others through which individuals would improve their opinions to be more than mere expressions of raw preferences.

Finally, democratic governance requires not only that citizens participate in decisions but also that they have a role in determining which problems deserve collective attention. This agenda-setting capacity is central to autonomy: without the ability to shape the range of topics debated and decided, people cannot truly govern themselves. In *Franchise*, the agenda and criteria for selecting leaders remain opaque and outside citizens’ control. While it might be possible for citizens to influence agenda-setting outside the interaction with Multivac, it remains dubious how much political elites would care of the views of the citizenry when the decision-making power resides with only one citizen. Thus, the people’s autonomy—their ability to define what counts as a pressing issue or a just solution—would be compromised.

These shortcomings highlight that while such a system of AI governance may deliver outcomes that resemble a certain kind of democratic decisions, it fails the empowered and equal inclusion of all citizens, collective deliberation necessary for authentic self-government, and the autonomy to determine the political agenda. Therefore, an undemocratic and binding AI agent like Multivac, as portrayed in *Franchise*, fails on the democratic criteria.

1. Democratic Binding

Perhaps, the fundamental problem with undemocratic binding AI like Multivac is the undemocratic nature of inputs because granting binding authority to a different agent other than the people themselves is not so rare in democratic governance as we see in representatives. The difference between Multivac and democratic representatives might be who gives the inputs. Thus, if the demos are allowed to give inputs to the AI system which then processes those inputs according to reasonably regulated algorithms, it may not necessarily undermine democratic governance. Instead, they can serve as instruments that embody what people really want by overcoming epistemic obstacles like ignorance.

To explore this perspective, I discuss two potential models of democratic binding AI: Artificial Intelligence-Enlightened Preference Voting (AI-EPV) and AI Delegates with Deliberative Consent. The former aims for epistemically informed decisions but falls short of some aspects of democratic governance, while the latter seeks to address the core elements of democracy more effectively by overcoming epistemic limits of citizens.

AI-EPV adapts Jason Brennan’s idea of Enlightened Preference Voting (EPV). In EPV, citizens provide detailed inputs about their preferences, demographic information, and knowledge on election day. For example, citizens would enter their demographic information such as their ethnicity, gender, religion, income, wealth, and place of living, along with their preferred policy or candidate, and they take a knowledge test. If they perform poorly on the test, it means that their preferences are based on the lack of relevant information. Likewise, their good performance on the test—i.e., answering all the questions correctly—means that their preferences are the products of informed reasoning. After collecting all the information including the test results, a central committee uses predetermined methods to estimate what the fully informed public would want, producing a policy outcome that is considered as the people’s decision under ideal epistemic conditions. The reason behind it is straightforward: Uninformed citizens would prefer the same policy or candidate as what their informed counterparts have chosen, were they informed on the issue.

By replacing the human committee with an AI system that processes these inputs, EPV can be an instance of democratic decision-maker AI. In AI-EPV, citizens submit detailed information of their preferences, demographic backgrounds, and knowledge levels. An AI system then estimates what a fully informed electorate would decide, assuming that less informed citizens would adopt the choices of their better-informed counterparts if given equivalent knowledge. By aggregating and processing these inputs algorithmically, AI-EPV aims to produce epistemically justified decisions, ensuring the quality of decisions.[[13]](#footnote-12)

Yet, despite its potential for producing epistemically better decisions, AI-EPV struggles with empowered and equal inclusion and deliberation. While AI-EPV ensures that everyone can provide inputs, it does not guarantee that all inputs influence the final decision.[[14]](#footnote-13) At best, the system disproportionately amplifies the inputs of those who are viewed as informed based on the knowledge test and, at the worst, it simply dismisses the preferences of uninformed citizens assuming that their preferences would change were they informed on the issue. Thus, AI-EPV creates a hierarchy of influence, where the preferences of some citizens carry more weight than others, undermining equality and excluding uninformed citizens in effect.

 Similarly, AI-EPV is not deliberative because there is no meaningful communication between citizens or even with AI agents before decision-making but only unilateral making of inputs to the system. Surely, it is still possible that citizens deliberate before the election day, and AI-EPV is not against deliberation to that extent. Some of the positive effects of deliberation include grasping better understanding of the views of others and, based on that, improving one’s own perspective by listening to others’ arguments and comparing one’s argument with another’s. In AI-EPV, however, this collective deliberation does not occur because individuals' inputs are merely data entries about themselves, not contributions to a shared process of reaching a broadly agreeable decision.

For autonomy, whether AI-EPV addresses the concerns most relevant to the people depends on how the whole democratic system looks like. If there is a citizen body whose primary function is to set the agenda, a democratic system with AI-EPV would not sacrifice autonomy. In the original EPV, Brennan suggests incorporating mini-publics like deliberative polling for such democratic inputs (Brennan and Landemore, 2020).[[15]](#footnote-14) Alternatively, if AI sets the agenda based on democratic inputs–e.g., actively gauging individuals’ current concerns or conducting a poll on popular issues–autonomy could be protected by such a measure as a right to explanation (Vredenburgh, 2021). Autonomy in understanding the decisions and controlling our circumstances lies as the motivations behind the right (Taylor, 2023; Wachter, Mittelstadt, and Russell, 2018). For example, people could ask why adjusting the interest rate should be prioritized over more welfarist policies and continue discussing the issue in response to the answers. Thus, it remains open whether AI-EPV ensures autonomy of citizens.

Since AI-EPV does not satisfactorily address some elements of democratic governance, should we conclude that there is no viable model of democratic-binding AI? While acknowledging the potential risks associated with granting AI decision-making power, it is still possible to envision models of AI that could make binding decisions with reduced risks. Consider an AI model that works in the following way. An individual voluntarily chooses an AI agent that goes through multiple questions and answers with the person. Some of these questions may resemble EPV’s knowledge tests, but most are carefully structured to encourage reflection on political issues by considering diverse perspectives. After the deliberative conversation, they agree with a policy-direction that the person prefers over other directions. Then, the person freely consents to delegate her decision-making power, like voting, until the upcoming election.[[16]](#footnote-15) This consent differs from blindly delegating her voting power to an AI agent to replace human judgment because her decision to delegate after the deliberative procedure is an informed decision.[[17]](#footnote-16)

Not surprisingly, such an AI delegate would help improve the quality of decisions for two reasons. First, citizens can get informed about politics through deliberative communication with AI. Since delegation is not mandatory, citizens can learn about politics without delegating their voting power, improving their epistemic competence. Even when they choose to delegate, they do not delegate their voting power as if giving it up but rather do so in an informed way while enhancing their political knowledge. Second, while individuals might not have the epistemic ability to check all details of political issues and all relevant up-to-date information, the AI agent can do this.[[18]](#footnote-17) Thus, AI would decide based on the mutually agreed policy-direction and vast amount of information relevant to the election, ensuring the epistemic quality of its decision.

Now, we should ask if the AI Delegate with Deliberative Consent fails the democratic elements. While this type of AI makes binding decisions by delegating citizens’ voting rights, it does not exclude people or creates inequality but provides an alternative mechanism for participation.[[19]](#footnote-18) This is because the AI delegate does not fundamentally change the decision-making process, but it only adds a delegatory way of decision-making. Rather than diminishing citizens’ agency, it enhances it by enabling individuals to participate even if they lack the time, knowledge, or resources to engage directly (Viehoff, 2024). Insofar as every individual retains the same opportunity to use an AI delegate, their voice would not be disproportionately amplified or diminished. This way, it is more inclusive without compromising equality.[[20]](#footnote-19)

Furthermore, using an AI delegate would enhance the deliberative aspect of decision-making for two reasons. First, citizens would get better informed about politics over deliberative conversations with the AI agent who serves as a personalized guide. Through these deliberative conversations, individuals can better comprehend complex political topics, learn diverse perspectives, and develop better informed views (Viehoff, 2024).[[21]](#footnote-20) Second, the AI does not merely delegate the person’s raw preference but her informed policy preference, making the political outcome the product of not raw opinions but deliberations. Deliberative conversations allow not only the user but also the AI delegate to refine and transform her raw preferences into an informed policy stance. This ensures that the decision-making process is not shaped by superficial or uninformed opinions but by reasoned judgments rooted in thoughtful deliberation.[[22]](#footnote-21) The result is a more epistemically robust political outcome—one that reflects the reasoned input of an engaged and informed citizenry.

Lastly, the AI delegate, through deliberative communication with its user, can help individuals articulate their concerns more effectively. Many citizens may struggle to connect their lived experiences with specific policy solutions. The AI delegate can bridge this gap by guiding users through the implications of various policy directions, helping them enunciate informed and actionable preferences (Benjamin, 2019). For instance, a citizen deeply worried about climate change may still lack the expertise to identify which policies would effectively address their concerns. The AI delegate can process the individual’s priorities, provide insights into various approaches, and help the user develop a clear and informed stance—whether it be advocating for renewable energy subsidies, supporting carbon pricing mechanisms, or prioritizing climate adaptation measures. This process ensures that each citizen’s voice contributes meaningfully and coherently to collective decision-making by identifying the right policy issues to be addressed, elevating the overall quality of democratic outcomes.

 Thus, while democratic binding AI like AI-EPV is limited in some of the elements of democratic governance, AI Delegates with Deliberative Consent are better positioned to help alleviate the epistemic limits of citizens without significantly harming democratic governance.

1. Democratic Unbinding

 Democratic unbinding AI makes unbinding decisions like recommendation and advice while receiving input from individuals. Using “ChatGPT”, “Siri”, “Gemini”, or other conversational AI before making a decision can be a good example of advisory AI. You can ask about various topics, and the answers can be useful as they are based on processing massive data which you alone cannot manage to understand. For instance, you can ask a question about a policy and even whether a person like you who has such-and-such political and social backgrounds is better to support the policy. Whatever the AI says, you can still decide based on your beliefs. For this reason, these usages of AI are not binding.[[23]](#footnote-22)

 Since advisory AI does not necessarily interfere with an individual's decision-making, it can be a natural part of democratic decision-making. Similar to the AI delegate, advisory AI is inclusive for its wide accessibility, potentially bridging gaps for those who are busy with their daily life lacking time, resources, expertise, and energy to acquire political knowledge (Elliott, 2023). Chatbots already exemplify how AI can provide tailored advice, enabling citizens to engage with political issues in a manner that suits their individual needs.[[24]](#footnote-23) Furthermore, advisory AI does not inherently privilege certain individuals or groups though a critical limitation of this optimistic perspective–i.e. uneven distribution of digital infrastructure and resources due to the underlying socio-economic inequalities–is needed.[[25]](#footnote-24) Its advice is personalized based on user queries, ensuring that each citizen’s interaction with the AI is tailored to their specific concerns or contexts.

 AI advisers have two distinct advantages in terms of deliberation. First, an AI adviser facilitates “deliberation within” (Goodin, 2000). Deliberation-within or internal deliberation is distinct from external deliberation because it occurs within one’s own mind before deliberating with others. In external deliberation, individuals bring their own reasons, beliefs, and perspectives that they have before deliberation. Unlike deliberative democratic theorists who tend to believe that deliberation can change the beliefs and preferences of its participants, those who emphasize internal deliberation argue that such a change occurs prior to deliberation with others. Consider Goodin and Niemeyer’s analysis of an Australian citizens’ jury where people are randomly selected to discuss the policy of building a track that raises the tension between the community’s convenience and environmental concerns. In their description, the stage where many jurors changed their attitudes was not the deliberation phase wherein they could deliberate about the issue with others. It was during the information phase that the juries had chances to see the actual sites at issue, reflect on the issue, and listen to the experts’ talk, and based on these, they changed their policy attitudes (Goodin and Niemeyer, 2008).

 Communicating with AI before making a decision can play a similar role to internal deliberation. Advisory AI can be utilized for learning information as in the Australian citizens’ jury case. AI can provide some empirical materials for them to reflect. Moreover, AI can be an interlocutor who plays the role of political others, but it is not an actual political other. This is like when you imagine the opinions of political others in a monologic way. With AI advisers, instead of imagining political others, you can communicate with an artificial agent who can provide relevant information about political others, such as their beliefs, perspectives, and recent events, by playing the role of political others. The role of imagination in internal deliberation can be replaced by communication with AI. Thus, using AI advisers can effectively realize some benefits of internal deliberation.

 Second, interacting with AI can come without the adversarial and emotional dynamics often present in interpersonal political debates. Citizens acquire political knowledge through traditional mass media such as the newspaper or news or by interacting with other citizens online or in person. Likewise, they encounter political others through news, online platforms such as Twitter, or personal communication with them. As critics of deliberative democracy point out, however, such political interactions tend to strengthen one’s own views rather than bring ideal deliberative effects such as a better understanding of different political views (Brennan, 2016; Taber and Lodge, 2006). Biases such as the confirmation bias explain how simply experiencing different opinions is not enough to achieve positive deliberative effects (Pereira et al., 2023). Nevertheless, knowing different political opinions is important for reasonable judgment. If ignorance includes biased knowledge rather than just a lack of knowledge, understanding different political opinions is important for making reasonable political judgments. However, prejudices and biases prevent giving due credit to different opinions.[[26]](#footnote-25) Even when you are not prejudiced and biased, negative emotions from political engagement can make you be.

Interacting with AI advisers can help to conduct such political interactions without political others and, thus, the involvement of negative emotions. Less involvement in negative emotions means that citizens can have better chances for reason-based interaction with different political perspectives (Talisse, 2021). In many democratic societies, citizens tend to act more like sports fans than rational thinkers and political debate becomes close to a cheering competition. Giving citizens an option for non-interpersonal political engagement with AI in addition to regular interpersonal interaction would mean that they have more chances to think over the non-emotional communications rather than to emotionally react. In this way, non-interpersonal interaction through AI could provide a means for the democratic society to deal with its problem of ignorance with a less inclination of emotional civic conflicts.

Lastly, AI advisers can help citizens identify, contextualize, and articulate their concerns in a way that can influence democratic decision-making processes by mitigating the practical difficulties in judging based on massive relevant information. AI advisers can filter through the overwhelming volume of relevant information to highlight the issues most relevant to an individual’s interests, values, or demographic background. Through personalized analysis of a user’s stated concerns, past voting behavior, or expressed values, an AI adviser can recommend key policy areas or emerging problems that the individual may have overlooked. For example, a citizen who expresses concern about the environment but remains unaware of related labor market implications might receive suggestions to explore the intersection of environmental policy and just transition programs for affected workers. In this way, AI can illuminate connections between seemingly disparate issues, helping citizens better understand the multifaceted nature of politics and encouraging more holistic judgments.

Besides navigating massive information, the AI adviser can also help citizens articulate their own priorities more coherently. Many citizens feel vaguely troubled by certain social conditions like rising inequality, unemployment, or climate change, but they often cannot translate these concerns into clear political objectives. By offering clarifying questions, contextual explanations, and comparative analyses, the AI adviser helps individuals identify their core interests. Over time, as the citizen refines their inquiries, the AI’s responses can sharpen their understanding of which topics genuinely align with their values and which may be secondary.[[27]](#footnote-26)

1. Undemocratic Unbinding

Undemocratic unbinding AI agents are already present in current societies. For instance, YouTube’s recommendation algorithms suggest videos without the user actively seeking them out, and Google’s search algorithms prioritize certain results over others without consent. Similarly, targeted advertisements rely on algorithms to present content the user did not choose. In a different vein, AI can operate to protect privacy by disturbing AI’s profiling of citizens (Auliya et al., 2024). In all these cases, individuals are exposed to information curated by AI algorithms.

While these raise their own set of ethical, social, and political concerns, I want to focus on a more deliberate and reflective type of unbinding AI intervention, which can be called the AI nudger (Hirmiz, 2024).[[28]](#footnote-27) Unlike AI delegates or advisers where the human agent actively seeks for the AI's answers, the AI nudgers highlights tensions or inconsistencies between her preferences and the AI’s autonomous assessment of the situation. By design, the nudger is topic-specific, but the user sets only the general domain or subject matter—such as political candidates, product purchases, or health choices. The nudger makes its judgments independently, drawing from news reports, social media, expert analyses, and other credible data sources. It does not simply reflect the user’s own views, but rather signals when there may be reason to pause, reconsider, and re-examine one’s reasoning.

Importantly, the AI nudger’s role is strictly unbinding. It cannot force the user’s hand, nor does it impose penalties for ignoring its suggestions. Its purpose is to spark reflection: a gentle prompt to the user to double-check her reasoning and perhaps seek additional information before finalizing a decision. To illustrate, imagine a voter who prefers candidate P over candidate Q because of P’s strong environmental platform. As the election nears, the AI nudger might highlight recent evidence that P’s campaign accepted donations from fossil fuel interests, or that independent analyses suggest P’s policies may be less robust than initially portrayed. Presented with this concise summary and supporting links, the user can choose to delve deeper, verify the claim, compare candidates more thoroughly, and potentially adjust their stance. Still, after consideration, the user might still support P, concluding that other aspects of P’s platform outweigh these discrepancies.

In this scenario, the nudger’s primary achievement is not to change the user’s ultimate choice, but to ensure that the choice is more considered and informed. Because the nudger’s interventions are entirely optional and non-coercive, it does not inherently undermine the elements of democratic governance.[[29]](#footnote-28) The key for maintaining democratic integrity here lies in the transparency and fairness of the underlying algorithms, which determine what counts as a relevant discrepancy, whose perspectives are included, and how evidence is weighted. While these questions of algorithmic fairness and openness are set aside here, it is sufficient to note that if designed and governed ethically, an AI nudger could enhance the quality of individual reasoning without compromising essential democratic values.

IV. Conclusion

This paper has sought to navigate the uncharted waters of AI's burgeoning role in democratic governance. By applying a framework that distinguishes AI systems based on the democratic nature of their inputs and the binding force of their outputs, I have illustrated the categories of analyzing AI models in democratic governance. While we already live with unbinding AI in various aspects of our daily lives, it might look repulsive to have AI with binding authority. Though I agree with the worry, I argued that there could be potentially promising and less risky models of AI for epistemically better decision-making, assisting citizens to fight against their epistemic limits. These models, by empowering citizens with information and facilitating more nuanced engagement, offer a path toward addressing the chronic illness of voter ignorance while upholding the fundamental elements of inclusion, equality, deliberation, and agenda-setting autonomy. They can be summarized as follows:

|  | **Inclusive & Equal Participation** | **Deliberation** | **Quality of Decision** | **Autonomy (or Agenda-setting)** |
| --- | --- | --- | --- | --- |
| **Multivac (Undemocratic Binding)** | **Low**(Only one representative is used; the broader demos are excluded)  | **Low**(No collective dialogue; only a single individual’s input is considered) | **Low**(Decisions reflect a top-down process that does not capture the informed preferences of the public) | **Low/Unclear**(Agenda-setting is opaque) |
| **AI-EPV(Democratic Binding)** | **Moderate**(Broad participation is enabled, but the system may overweight informed inputs, potentially marginalizing less knowledgeable citizens | **Low**(Inputs are gathered individually with little collective discussion) | **High**(By aggregating detailed inputs, it aims to simulate the decision of a fully informed elected) | **Low/Unclear**(The filtering of inputs might constrain genuine agenda-setting if diverse voices are downplayed) |
| **AI Delegates with Deliberative Consent(Democratic Binding)** | **High**(Provides an alternative inclusive mechanism for participation, especially for citizens with limited time or expertise) | **High**(Involves deliberative communication between citizens and AI agents) | **High**(Combines personalized deliberation with AI’s data processing to produce well-informed decisions) | **Potentially High**(Preserves citizens autonomy by allowing informed delegation and continuous feedback; citizens retain ultimate control over decision-making) |
| **AI Advisor(Democratic Unbinding)** | **High**(Widely accessible; citizens voluntarily seek advice) | **High**(Facilitates internal deliberation through tailored and accessible information) | **High**(Enhances decision quality by providing personalized information that helps make better choices) | **Potentially High**(While it clarifies political issues and options, it remains non-binding so final decisions stay with the citizen) |
| **AI Nudger(Undemocratic Unbinding)** | **High**(Widely accessible; citizens voluntarily seek AI’s double-check for a pause-and-reflection) | **Moderate**(Stimulates individual reflection by highlighting discrepancies, yet lacks a mechanism for interactive deliberation) | **Moderate**(Encourages reconsideration of choices, potentially improving decision quality without enforcing consensus) | **Potentially High**(Nudges may subtly influence agenda-setting, but citizens still have control over final decisions) |

It is crucial, however, to acknowledge the limitations inherent in the scope of my argument. My discussion of the core elements of democracy is limited to its deliberative and epistemic conception, which is mainly due to the scope of my argument–addressing epistemic deficiencies in contemporary democracies like voter ignorance by AI. Also, I have intentionally bracketed the thorny, yet undeniably vital, issues of algorithmic transparency, manipulation, publicity, fairness, and accountability. These concerns, centered primarily on the *process* stage of AI systems, are not to be dismissed easily. Indeed, they represent a critical frontier for implementing AI models in reality because the integration of AI into democratic decision-making processes will reshape how political discourse is conducted and how citizens are targeted and categorized. Machine learning algorithms can influence which voices are amplified and which issues are prioritized, affecting who is represented and how political preferences are constructed and thereby influencing the outcomes of the democratic decision-making system. Surely, algorithms are not neutral arbiters; they are imbued with the values, biases, and potential for manipulation of their creators. I admit that a truly robust framework for democratic governance with AI must grapple with these challenges, ensuring that the inner workings of these powerful tools are subject to rigorous scrutiny, ethical guidelines, and democratic oversight.

However, the decision to focus on the *input* and *output* stages of AI in this paper was a deliberate one, born not of disregard but of a desire to illuminate a different, yet equally fundamental, set of questions. Before we can fully assess the implications of how AI processes information, we must first grapple with who provides that information andwhat is done with the resulting outputs. By focusing on these often-overlooked aspects, this paper has sought to establish a foundational understanding of the democratic potential—and perils—of AI integration into our political lives.

Funding

The author declares that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests

The author has no relevant financial or non-financial interests to disclose.

Author Contributions

All authors contributed to the research.

Ethics Approval

This is a theoretical research. No ethical approval is required.

Consent to Participate

The research does not involve human subjects.

Consent to Publish

The research does not involve human subjects.

Bibliography

Achen, C. H., & Bartels, L. M. (2016). *Democracy for realists: Why elections do not produce responsive government*. Princeton University Press.

Agrawal, A., Gans, J., & Goldfarb, A. (2019). Exploring the impact of artificial intelligence: Prediction versus judgment. *Information Economics and Policy*, *47*, 1-6.<https://doi.org/10.1016/j.infoecopol.2019.05.001>

Ali, O., Scheidt, N., Gegov, A., Haig, E., Adda, M., & Aziz, B. (2020). Automated detection of racial microaggressions using machine learning. *Proceedings of IEEE SSCI 2020*, 2477–2484.

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

Anderson, E. (2009). Democracy: Instrumental vs. non‐instrumental value. In T. Christiano & J. P. Christman (Eds.), *Contemporary debates in political philosophy* (pp. 213–227). Wiley-Blackwell.

Arendt, Hannah. (1958). *The Human Condition*. Chicago: University of Chicago Press.

Asimov, I. (1955). Franchise. In *If: Worlds of science fiction*. Quinn Publishing Company, Inc. Retrieved from<https://www.astro.sunysb.edu/fwalter/HON301/franchise.pdf>

Auliya, S.F., Kudina, O., Ding, A.Y. *et al.* AI versus AI for democracy: exploring the potential of adversarial machine learning to enhance privacy and deliberative decision-making in elections. *AI Ethics* (2024). https://doi.org/10.1007/s43681-024-00588-2

Barton, A., & Grüne-Yanoff, T. (2015). From libertarian paternalism to nudging—and beyond. *Review of Philosophy and Psychology*, *6*(3), 341–359.<https://doi.org/10.1007/s13164-015-0268-x>

Baumgartner, F. R., Berry, J. M., Hojnacki, M., Kimball, D. C., & Leech, B. L. (2009). *Lobbying and policy change: Who wins, who loses, and why*. University of Chicago Press.

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity.

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *Proceedings of Machine Learning Research*, *81*, 1-11. Retrieved from [https://ssrn.com/abstract=3086546](https://ssrn.com/abstract%3D3086546)

Birhane, A., Isaac, W., Prabhakaran, V., Diaz, M., Elish, M. C., Gabriel, I., & Mohamed, S. (2022). Power to the people? Opportunities and challenges for participatory AI. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (EAAMO '22), Article 6, 1–8.<https://doi.org/10.1145/3551624.3555290>

Breitfeller, L., Ahn, E., Jurgens, D., & Tsvetkov, Y. (2019). Finding microaggressions in the wild: A case for locating elusive phenomena in social media posts. *EMNLP-IJCNLP 2019*, 1664–1674.<https://doi.org/10.18653/v1/D19-1176>

Brennan, J. (2016). *Against democracy*. Princeton University Press.

Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review*, *81*(5), 825-836.<https://doi.org/10.1111/puar.13293>

Caplan, B. (2008). *The myth of the rational voter: Why democracies choose bad policies*. Princeton University Press.

Carpini, M. X. D., & Keeter, S. (1996). *What Americans know about politics and why it matters*. Yale University Press.

Christiano, T. (1996). *The rule of the many: Fundamental issues in democratic theory*. Routledge.

Citron, D. K. (2008). Open code governance. *University of Chicago Legal Forum*, 355-387. Retrieved from [https://ssrn.com/abstract=1081689](https://ssrn.com/abstract%3D1081689)

CNN Wire. (2017). Meet SAM: This virtual politician wants to run for office. <https://whnt.com/news/meet-sam-this-virtual-politician-wants-to-run-for-office/>

Coeckelbergh, M. (2023). Democracy, epistemic agency, and AI: political epistemology in times of artificial intelligence. *AI Ethics* 3, 1341–1350. <https://doi.org/10.1007/s43681-022-00239-4>

Coglianese, C., & Lai, A. (2022). Assessing automated administration. In J. B. Bullock, M. O. Lazar, and A. Seger (Eds.), *The Oxford handbook of AI governance* (online ed.). Oxford Academic. https://doi.org/10.1093/oxfordhb/9780197579329.013.30

Cohen, J. (1996). Procedure and substance in deliberative democracy. In S. Benhabib (Ed.), *Democracy and difference: Contesting the boundaries of the political* (pp. 95-119). Princeton University Press.

Curry, A., Flett, P., & Hollingsworth, I. (2006). *Managing information and systems: The business perspective*. Routledge.

Dahl, R. A. (1989). *Democracy and its critics*. Yale University Press.

Denton, E., Hanna, A., Amironesei, R., Smart, A., & Nicole, H. (2021). On the genealogy of machine learning datasets: A critical history of ImageNet. Big Data & Society, 8(2). https://doi.org/10.1177/20539517211035955

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, *59*(2), 56–62.<https://doi.org/10.1145/2844110>

Downs, A. (1957). *An economic theory of democracy*. Harper.

Estlund, D. (2009). *Democratic authority: A philosophical framework*. Princeton University Press.

Gellers, J.C. Crowdsourcing global governance: sustainable development goals, civil society, and the pursuit of democratic legitimacy. *Int Environ Agreements* 16, 415–432 (2016). https://doi.org/10.1007/s10784-016-9322-0

Giubilini, A., & Savulescu, J. (2018). The artificial moral advisor. The ‘ideal observer’ meets artificial intelligence. *Philosophy & Technology*, *31*(2), 169–188.<https://doi.org/10.1007/s13347-017-0285-z>

Goodin, R. E. (2000). Democratic deliberation within. *Philosophy & Public Affairs*, *29*(1), 81-109.

Goodin, R. E., & Niemeyer, S. J. (2008). When does deliberation begin? In R. E. Goodin (Ed.), *Innovating democracy: Democratic theory and practice after the deliberative turn*. Oxford University Press.

Guerrero, A. A. (2024). *Lottocracy: Democracy without elections*. Oxford University Press.

Habermas, J. (1996). *Between facts and norms: Contributions to a discourse theory of law and democracy* (W. Rehg, Trans.). The MIT Press.

Hamon, R., Junklewitz, H., & Sanchez, I. (2021). Impossible explanations? Beyond explainable AI in the GDPR from a COVID-19 use case scenario. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (FAccT '21) (pp. 549–559).<https://doi.org/10.1145/3442188.3445917>

Hendrycks, D. (2023). Natural selection favors AIs over humans. *arXiv*.<https://doi.org/10.48550/arXiv.2303.16200>

Himmelreich, J. (2022). Against “democratizing AI.” *AI & SOCIETY*, 1–14.

Hirmiz, R. (2024). The epistemic role of AI decision support systems: Neither superiors, nor inferiors, nor peers. *Philosophy & Technology*, *37*(4), 127.<https://doi.org/10.1007/s13347-024-00819-8>

Inioluwa, D. R., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (AIES '19) (pp. 429–435). [https://doi.org/10.1145/3306618.3314244](https://doi.org/10%3C22%3E.1145/3306618.3314244)

Lafont, C. (2020). *Democracy without shortcuts: A participatory conception of deliberative democracy*. Oxford University Press.

Landemore, H. (2012). *Democratic reason: Politics, collective intelligence, and the rule of the many*. Princeton University Press.

Landemore, H. (2020). *Open democracy: Reinventing popular rule for the twenty-first century*. Princeton University Press.

Lara, F., & Deckers, J. (2020). Artificial intelligence as a Socratic assistant for moral enhancement. *Neuroethics*, *13*(3), 275–287.

Lazar, S. (2024). Legitimacy, authority, and democratic duties of explanation. In D. Sobel & S. Wall (Eds.), *Oxford studies in political philosophy* (Vol. 10).<https://doi.org/10.1093/oso/9780198909460.003.0002>

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

Lessig, L. (2011). *Republic, lost: How money corrupts Congress–and a plan to stop it*. Twelve.

Lipset, S. M. (1959). Some social requisites of democracy: Economic development and political legitimacy. *American Political Science Review*, *53*(1), 69-105.<https://doi.org/10.2307/1951731>

Lovett, A., & Zuehl, J. (2022). The possibility of democratic autonomy. *Philosophy & Public Affairs*, *50*(4), 467-498.<https://doi.org/10.1111/papa.12223>

Lupia, Arthur. 2006. “How Elitism Undermines the Study of Voter Competence.” *Critical Review* 18 (1–3): 217–32.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, *54*(6), 1–35.<https://doi.org/10.1145/3457607>

Mollick, E. (2024). *Co-intelligence: Living and working with AI*. Portfolio.

Mouffe, Chantal. (2005). *On the Political*. New York: Routledge.

Mutch, R. E. (2014). *Buying the vote: A history of campaign finance reform*. Oxford University Press.

Noorman, M., & Swierstra, T. (2023). Democratizing AI from a sociotechnical perspective. *Minds and Machines*, *33*(4), 563–586.

Panigrahi, R., Srivastava, P. R., & Sharma, D. (2018). Online learning: Adoption, continuance, and learning outcome—A review of literature. *International Journal of Information Management*, *43*, 1-14.<https://doi.org/10.1016/j.ijinfomgt.2018.05.005>

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Pereira, A., Harris, E., & Van Bavel, J. (2023). Identity concerns drive belief: The impact of partisan identity on the belief and dissemination of true and false news. *Group Processes & Intergroup Relations*, *26*(1), 24–47.

Rancière, Jacques. (1999). *Disagreement: Politics and Philosophy*. Minneapolis: University of Minnesota Press.

Sætra, H. S., & Danaher, J. (2022). To each technology its own ethics: The problem of ethical proliferation. *Philosophy & Technology*, *35*(4), Article 93.

Savulescu, J., & Maslen, H. (2015). Moral enhancement and artificial intelligence: Moral AI?. In J. Romportl, E. Zackova, & J. Kelemen (Eds.), *Beyond artificial intelligence. The disappearing human-machine divide* (pp. 79–95). Springer.<https://doi.org/10.1007/978-3-319-09668-1_6>

Seger, E., Ovadya, A., Garfinkel, B., Siddarth, D., & Dafoe, A. (2023). Democratising AI: Multiple meanings, goals, and methods. *arXiv*.

Schumpeter, Joseph. 2008 [1942]. *Capitalism, Socialism, and Democracy*. New York: Harper Perennial Modern Thought.

Smids, J. (2012). The voluntariness of persuasive technology. In *International conference on persuasive technology, persuasive technology. Design for health and safety* (pp. 123–132). Springer.<https://doi.org/10.1007/978-3-642-31037-9_14>

Somin, I. (2016). *Democracy and political ignorance: Why smaller government is smarter*. Stanford Law Books.

Susser, D., Roessler, B., & Nissenbaum, H. (2019). Technology, autonomy, and manipulation. *Internet Policy Review*, *8*(2).<https://doi.org/10.14763/2019.2.1410>

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*(3), 755-769.<https://doi.org/10.1111/j.1540-5907.2006.00214.x>

Talisse, R. B. (2019). *Overdoing democracy: Why we must put politics in its place*. Oxford University Press.

Talisse, R. B. (2021). *Sustaining democracy*. Oxford University Press.

Taylor, E. (2023). Explanation and the right to explanation. *Journal of the American Philosophical Association*, *10*(3), 467–482.

Veale, M., & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, *4*(2).<https://doi.org/10.1177/2053951717743530>

Viehoff, J. (2024). Beyond justice: Artificial intelligence and the value of community. In J. B. Bullock, M. O. Lazar, and A. Seger (Eds.), *The Oxford handbook of AI governance* (online ed.). Oxford Academic.<https://doi.org/10.1093/oxfordhb/9780197579329.013.71>

Vredenburgh, K. (2022). The right to explanation. *Journal of Political Philosophy*, *30*(2), 209-229.<https://doi.org/10.1111/jopp.12262>

Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law and Technology*, *31*(2), 841–887.

Warren, M. E. (2017). A problem-based approach to democratic theory. *American Political Science Review*, *111*(1), 39–53.

Young, I. M. (2000). *Inclusion and democracy*. Oxford University Press.

Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019a). Transparency in algorithmic and human decision-making: Is there a double standard?. *Philosophy & Technology*, *32*(4), 661–683.<https://doi.org/10.1007/s13347-018-0330-6>

Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019b). Algorithmic decision-making and the control problem. *Minds and Machines*, *29*(4), 555–578.<https://doi.org/10.1007/s11023-019-09513-7>

Ziewitz, M. (2016). Special issue introduction: Governing algorithms: Myth, mess, and methods. *Science, Technology, & Human Values*, *41*(1), 3–16.

1. Before going further, I should admit that my analysis is grounded in an epistemic and deliberative understanding of democracy, which stresses citizens' ability to access and utilize knowledge for informed decision-making. I acknowledge that this perspective is debated in political philosophy. Critics like Chantal Mouffe (2005), Jacques Rancière (2004), and Hannah Arendt (1958) argue that democratic legitimacy stems not only from rational decision-making but also from agonistic pluralism, the self-education of marginalized people, and contestation. While my framework highlights how information and deliberation enhance decision quality, I do not advocate for a purely technocratic democracy. Instead, my argument is part of a larger discussion on strengthening democratic processes with AI, acknowledging the complex interplay of rationality, emotion, and power in democratic life. [↑](#footnote-ref-0)
2. Hélène Landemore emphasizes inclusiveness and equality, understanding them to constitute the core value of democracy, “democraticity.” It seems to me that her idea of democraticity is an attempt to answer the who-question. See Landemore 2020. [↑](#footnote-ref-1)
3. While I do not intend to advocate for a specific version of proceduralism, I maintain that inclusive and equal participation is epistemically valuable in democratic contexts. Many scholars have emphasized the procedural benefits of such inclusiveness, with the Condorcet Jury Theorem serving as one well-known illustration of this principle. More recently, Landemore (2012), drawing on the Diversity Trumps Ability Theorem, argued that cognitively diverse decision-making groups can outperform decisions made by a small number of elites. Given that an in-depth discussion of which procedural approach is superior is beyond the scope of this article, I will limit my argument to the minimal claim that inclusive and equal participation is not only morally valuable—promoting freedom and fairness—but also epistemically beneficial. In this respect, my view aligns with Estlund’s (2007) epistemic proceduralism, which holds that decision-making procedures should allow for the participation of all citizens—thereby ensuring procedural legitimacy—while also requiring that decisions be acceptable to qualified perspectives, thus achieving a high-quality outcome. [↑](#footnote-ref-2)
4. In addition to this moral aspect of deliberation, Landemore (2012) gives a comprehensive argument over the epistemic value of collective deliberation, connecting the epistemic and moral aspects of deliberative decision-making. [↑](#footnote-ref-3)
5. John Dewey viewed democracy as an ongoing, collective process for solving problems that arise from new challenges and the unexpected consequences of current practices (Dewey, 1910). These challenges prompt public reflection and the search for solutions. Hélène Landemore builds on this idea, arguing that democracy's strength lies not just in finding answers but in identifying and defining the problems themselves. This is especially important when facing political uncertainty, where diverse perspectives are crucial for navigating complexity (Brennan and Landemore 2022, p. 203). [↑](#footnote-ref-4)
6. Though not specifically about autonomously determining our own problems, Lovett and Zuehl (2022) argue that joint intention as the crux of collective authorship of decisions. Considering our problems reflect our intention, their argument can be viewed as an attempt to address the *what*-question. [↑](#footnote-ref-5)
7. Sparta, for instance, had a popular assembly, yet it is generally considered undemocratic because its agenda was controlled by the aristocratic *Gerousia* (Hansen 1990, pp. 288-290; 295). This contrasts with Athenian democracy, where the Assembly held considerable power to shape the political agenda. Thus, even in ancient Greece, the ability of citizens to influence which issues were addressed was a key indicator of democracy. [↑](#footnote-ref-6)
8. Anthony Downs (1957) argued that due to the minimal impact of an individual vote, citizens rationally choose not to expend the effort required to obtain comprehensive information. This highlights the cost–benefit calculation underlying the information deficit of voters. [↑](#footnote-ref-7)
9. While my primary focus is on how AI can help mitigate voter ignorance by providing better information, it is also essential to consider a deeper, political-epistemological question: How does AI, as a technological mediator, reshape the very conditions under which human knowledge is produced and exercised? In line with Mark Coeckelbergh’s (2022) argument, technology is not merely a passive instrument but an active mediator of epistemic agency. AI systems, by processing and synthesizing vast datasets, not only supply information but also reshape the cognitive landscape in which voters interpret political reality. This perspective urges to critically examine not only the efficacy of AI in epistemically enhancing decision-making but also the broader implications for human epistemic capacities and democratic participation. While this paper focuses on addressing the informational deficits of voters, I acknowledge that the deployment of AI in democratic contexts carries with it profound implications for how citizens understand and interact with knowledge. [↑](#footnote-ref-8)
10. *Polis* is another well-known example of utilizing AI in politics. On *polis* machine learning algorithms are used to cluster and visualize the diverse opinions of participants to reveal the areas of consensus and divergence. See <https://pol.is/home>. [↑](#footnote-ref-9)
11. AI-EPV, which will be discussed as a case of democratic binding AI, may have such a risk. [↑](#footnote-ref-10)
12. A fundamental problem with an AI-driven system like Multivac is that humans are not “in the loop” of decision-making because it violates the principle of procedural due process. Except for Muller, the rest of the citizens must obey Multivac's decision blindly. For the right to an accountable decision-maker, see Busuioc 2020; Citron 2008; Mollick 2024, Ch.3. [↑](#footnote-ref-11)
13. It should be noted that the currently described AI-EPV may appear to rely primarily on techniques resembling multiple linear regression—where a set of input variables such as various demographic indicators and knowledge test scores are assumed to relate linearly to a predicted outcome. In this simplified view, the system seems to generate scalar responses along the knowledge test axis rather than machine learning results. However, the potential of AI-EPV extends beyond this linear approach. When AI-EPV is expanded to cover multiple election cycles and continuously incorporates dynamic, real-time data, it can leverage more sophisticated machine learning techniques that capture non-linear relationships and complex interactions between variables. For example, advanced models—such as deep neural networks or ensemble methods—could integrate diverse data sources beyond direct voter input. By assimilating information from social media behavior, historical voting patterns, and real-time sentiment analysis, the system can construct a multi-dimensional profile of voter behavior, ultimately leading to better-informed democratic outcomes that grasp voters' underlying interests. [↑](#footnote-ref-12)
14. A more fundamental issue might be found in the design of questions. As the revisionists of voter ignorance sometimes point out, the questions to test voter’s knowledge tend to be elitist (Lupia 2006). For example, it might be practically easy for college graduates to learn basic social scientific knowledge, but not all people go to a college. On this line of view, the issue of voter ignorance might have been exaggerated. [↑](#footnote-ref-13)
15. AI-EPV can also adopt a citizen body for creating or monitoring algorithms to ensure its democratic nature. Consider “Civic AI” which is a collectively constructed algorithm by citizen juries whose participants collaboratively build an algorithm to assist political decision-making. See <https://civicai.cat/en/>. For participatory AI, see also Birhane et al. 2022; Himmelreich 2022; Saetra and Danaher. 2022; Noorman and Swierstra 2023; Segar et al. 2023. [↑](#footnote-ref-14)
16. Note two things. First, citizens can decide not to delegate and also withdraw their decision to delegate at any time before the election. Second, the epistemic benefit of delegation rather than directly voting lies in the fact that AI can decide based on vast amounts of up-to-date political information. The relation between the person and the AI delegate is like that of citizens and elected politicians described in Thomas Christiano (1996). In his explanation, the primary role of citizens is to set the political goal, not deciding detailed policies or laws, because the latter type of task requires political expertise. For this reason, elected politicians, despite their greater political power, do not rule over the citizens insofar as citizens set the political agenda. [↑](#footnote-ref-15)
17. One relevant point is the private nondelegation doctrine which is a principle in U.S. constitutional law that limits Congress's ability to delegate its legislative power to private entities. It stems from the broader nondelegation doctrine, which restricts Congress from excessively delegating its powers to other branches of government, particularly the executive branch. However, they are mainly to avoid corruption and may not fit the context of machine learning. See Coglianese and Lai 2022. [↑](#footnote-ref-16)
18. Though earlier AI models like ChatGPT 3.5 used to be trained on data about two years old, recent models with internet browsing abilities can provide the up-to-date information. [↑](#footnote-ref-17)
19. Recent discussions on algorithmic justice reveal at least two inequality-related problems of algorithms–implicit bias and microaggressions. Due to the unequal nature of data originating in human activities, AI can present implicit biases and microaggressions in its outcome. However, AI does not always strengthen such inequalities but can be used to weaken them by detecting biases and microaggressions of users. For example, current algorithms on digital platforms can identify implicit biases and microaggressions (Ali et al., 2020; Breitfeller et al., 2019). [↑](#footnote-ref-18)
20. One might worry that accessibility to AI can be exclusive and unequal because of the cost of using an AI agent. This is commonly observed in the unequal accessibility to digital platforms. For instance, elderly people or financially devastated people might not have equal access to digital platforms and thereby naturally become excluded. A similar issue might arise with AI. However, this is not an inherent issue with AI delegates because the accessibility can be ensured by external measures like institutional assistance, funding, and service workers. If citizens and the government endeavor to keep the cost of using an AI agent at a reasonable price range while providing necessary assistance, opening the option of delegation to an AI agent should not be inherently exclusive and unequal. [↑](#footnote-ref-19)
21. While the claim that an AI delegate enhances deliberation may initially seem overly optimistic, there are reasons to be cautiously positive. First, because AI agents lack emotional bias, they can diminish the influence of irrational factors—such as partisan attitudes—that often contribute to voter ignorance. This dispassionate approach can foster more balanced and information-based discussions. Second, engaging in structured deliberative interactions with an AI delegate signals a citizen’s commitment to improving her own epistemic competence. For example, although traditional media are frequently criticized for disseminating misinformation, reliable sources like the Pew Research Center provide well-substantiated information—even if processing that information requires considerable cognitive effort. A citizen’s willingness to seek out and engage with such sources indicates a proactive effort to enhance political knowledge. Similarly, voluntary participation in deliberative exchanges with an AI suggests an active attempt to overcome ignorance, which bodes well for the overall quality of democratic engagement over time. [↑](#footnote-ref-20)
22. The epistemic enhancement through Socratic deliberation with an AI agent was discussed as an alternative to an AI agent with an exhaustive decision-making power. This model does not threaten autonomy of human agents because the Socratic conversation empowers the users to overcome their epistemic limitations like ignorance. See Lara and Deckers 2020. [↑](#footnote-ref-21)
23. A similar model of advisor was suggested in neuroethics where AI produces moral judgments based on the criteria set by the human user who in turn decides to follow or reject the advice. See Giubilini and Savulescu 2018; Savulescu and Maslen 2015. [↑](#footnote-ref-22)
24. A potential worry with AI advisors is its working as persuasive technology to make the users naturally act according to the AI designer’s intentions. Williams (2024) explains that persuasive technology has three features: (a) the designer’s intentions to (b) shape human behavior in (c) a noncoercive manner. Considering that most large language models (LLMs) are controlled by big techs due to the cost-related accessibility to massive data, AI advisors could work as persuasive technology. While it may be impossible to entirely eliminate the persuasive nature of AI advisors, discussions of algorithmic justice give potential answers. As Pasquale (2015) emphasizes, transparency of algorithms should be required to disclose the intentions and objectives of the designers, including explicitly indicating the AI’s goals, funding sources, and development motivations. The right to explanation should be protected for the users to have access to clear and accessible explanations of why AI provided specific advice, including the underlying data, cross-check, algorithms, and reasoning processes (Danks, 2024; Vredenburgh, 2022). Similarly, AI advisors should allow users to calibrate the system to align with their own values and priorities rather than being subject to the preset by the designers (Ananny and Crawford, 2018). Additionally, governments and independent organizations must establish systems to audit AI systems for undue influence (Raji and Buolamwini, 2019). Finally, the monopolistic hold of big corporations over AI agents should be diversified by supporting open-source projects to prevent corporate biases dominating user interactions (Veale and Binns, 2017). [↑](#footnote-ref-23)
25. Similar to the AI delegate, unequal accessibility might arise as a problem. Yet, it is important to note that this is not an inherent problem of AI advisors but of society. Thus, when an issue of such an inequality arises, it needs to be addressed as a problem not necessarily of the AI agent itself but rather of the broader socio-economic situation behind it. See also fn. 20. [↑](#footnote-ref-24)
26. My view is not committed to a particular conception of Rawlsian duty of civility, but by reasonable political judgment, I want to keep the idea that democratic citizens bear a sense of obligation to try to base their political judgment on reasons accessible to other citizens. [↑](#footnote-ref-25)
27. A potential worry with AI advisers might be the control problem according to which human agents will be too reliant on machines becoming unduly diffident due to the superior performance of machines. In the context of the AI advisers, citizens could defer to AI’s recommendations all the time, undermining the civic value of interests in politics. As Zerilli et al. (2019) says, however, the control problem can be alleviated by distinguishing between what AI is best suited for and what a human user can do better. Since AI outperforms data-processing, individuals can share the burden of data-collection and data-process with AI, while holding the power of value-judgment, because they may believe that they are still better judges in value-judgment than AI. This way of collaboration between human and AI of tasks that they respectively best suited for could mitigate the chance of the control problem. [↑](#footnote-ref-26)
28. Note that the meaning of nudging in AI nudger differs from Thaler and Sunstein’s (2008). The latter notion of nudge aims to make intended decisions without direct interventions due to the practical limits in rational choice in decisions. In contrast, the AI nudger’s role is not to make particular decisions but to help the human user pause and reflect on a decision. Yet, a certain aspect is common between them. Consider a heuristic-blocking nudge—e.g., cool-off period between purchases to prevent irrational shopping. An AI nudger's intervention can function as such a heuristic-blocking nudge to make the user pause and rethink her decision. See Barton and Grüne-Yanoff 2015; Sunstein 2015. [↑](#footnote-ref-27)
29. A potential concern with AI nudgers is its potential limit of the user’s freedom of action. Smids (2012) argues that a car’s constant beeping to make the driver fasten the seat belt could constrain her capacity for self-control. In a similar vein, frequent interventions by an AI nudger could harm one’s autonomy despite its unbinding nature. However, this kind of problem can be addressed by the user’s setting the frequency and topics of AI interruption to serve her needs. [↑](#footnote-ref-28)