This work has progressed from a preprint to a peer-reviewed publication. The final version is now available in Social Epistemology. Please use the following citation for all references:

Špecián, P. (2024). Machine Advisors: Integrating Large Language Models Into Democratic Assemblies. Social Epistemology, 1–16. https://doi.org/10.1080/02691728.2024.2379271

For those without institutional access, a version of the article is also available on my personal webpage: https://www.petr-specian.com/news#h.z98jwsam7ib1.

Machine Advisors

Integrating Large Language Models into Democratic Assemblies

Petr Špecián

Department of Philosophy, Prague University of Economics and Business

Department of Psychology and Life Sciences, Charles University

Abstract: Large language models (LLMs) represent the currently most relevant incarnation of artificial intelligence with respect to the future fate of democratic governance. Considering their potential, this paper seeks to answer a pressing question: Could LLMs outperform humans as expert advisors to democratic assemblies? While bearing promise of enhanced expertise availability and accessibility, they also present challenges of hallucinations, misalignment, or value imposition. Weighing LLMs' benefits and drawbacks compared to their human counterparts, I argue for their careful integration to augment democracy's ability to address complex policy issues. The paper posits that time-tested democratic procedures like deliberation and aggregation by voting provide safeguards effective against both human and machine advisor imperfections. Additional protective measures include custom LLM training for the advisory role, boosting representatives' competencies in query formulation, or implementation of adversarial proceedings in which LLM advisors could debate each other and provide dissenting opinions. These could further mitigate the risks that LLMs present in advisory role and empower human decision-makers toward increased autonomy and quality of their collective choices. My conceptual exploration offers a roadmap for the co-evolution of AI and democratic institution, setting the stage for an empirical research agenda to finetune the implementation specifics.

Keywords: Large Language Models, Epistemic Democracy, Democratic Assemblies, Institutional Design.

Funding: The work on this paper is supported by the Czech Science Foundation (GACR), grant number 24-11697S.

Introduction

Large language models (LLMs) could become a revolutionary technology for democratic institutions. This paper focuses on their pragmatic potential to provide epistemic assistance to democratic representatives. Its central hypothesis is that LLMs' judicious employment in place of the human expert advisors could increase democratic assemblies' ability to access and assimilate specialized knowledge and thus improve their odds in tackling the complex challenges of our time.

LLMs are neural-network-based models trained on vast amounts of human text: trillions of words sourced from the internet. Their task is to discern the structure of human language(s) to enable prediction of subsequent "tokens"—that is, words or parts of words—in a sequence. Through the scaling of their size and training datasets, as well as various architectural tweaks, LLMs have recently gained a remarkably advanced ability to generate coherent and informative textual output. The most capable ones now achieve impressive performance across a variety of tasks, including language translation, summarization, conversational response generation, and other core language skills (Anthropic 2023a; OpenAI 2023; Borji and Mohammadian 2023). Their blisteringly fast progress is likely to continue for some time (Bowman 2023).

If mismanaged, such technology could exacerbate the existing pressures on democratic governance to a breaking point. It could amplify misinformation, aggravate inequalities due to asymmetric access, and further exacerbate political polarization, deteriorating people's capacity for collective action (Coeckelbergh 2022; Sætra 2023; Zhou et al. 2023). In short, LLMs could be the last straw to topple the democratic equilibrium and usher us toward anarchy or autocracy (Ovadya 2023b). However, they also brings opportunities to better integrate knowledge scattered across many minds to enable finding solutions that could otherwise escape our grasp (Small et al. 2023). Hence a hypothesis offers itself that LLMs' emerging capabilities could enhance democracy's ability to access and assimilate expertise, boosting its problem-solving performance.

To streamline my analysis of LLMs' epistemic potential in democratic settings, I make two simplifying steps:

1) I analyze on a specific scenario in which human expert advisors serve as alternatives to LLMs, excluding from consideration the potential synergies between the two. Admittedly, human-AI collaborations may temporarily eclipse both human-only and AI-only configurations (cf. Alves and Cipriano 2023). However, they would also dilute LLMs' key advantages in mediating expertise (see *Section 2*), and their possible configurations remain largely unexplored.

2) I confine my scrutiny to the role of LLMs as advisors within midsize democratic assemblies, sidestepping their potential epistemic impact in other areas critically important for modern democracies, such as general elections. I also omit the question of democratic reliance on expert bodies like supreme courts and central banks. The complexities inherent in these alternative settings remain beyond this paper's scope.

These simplifications allow for posing a question that is both radical and focused: *Could LLMs' employment in place of human advisors improve the problem-solving ability of democratic assemblies?* I argue that (1) LLMs' potential for success is significant, given their promise of increasing the availability and accessibility of expertise; (2) LLMs' specific risks, such as occasional unreliability or misalignment with users' goals and values, can be mitigated through a combination of standard democratic procedures and innovative, but affordable, techno-institutional remedies.

The paper thus expands the literature on "AI augmented" democracy (Bakker et al. 2022; Ovadya 2023b; 2023a; Sætra 2020; Small et al. 2023) and contributes a novel perspective on institutional design by envisioning a pragmatic integration of LLMs into the core of the current democratic processes. My argument has the following structure. In *Section 1*, I examine the contours of the uneasy relationship between expertise and democratic decision-making. Next, I pivot to a comparison between human experts and LLMs, espousing the latter's advantages in terms of the democratization of access to specialized knowledge (*Section 2*). However, machine advisors are no panacea. Accordingly, *Section 3* scrutinizes the pitfalls associated with LLMs' employment like the risks of misinformation, misalignment, and value imposition. In response to these challenges, *Section 4* considers the efficacy of democracy's existing defense mechanisms—such as deliberative procedures and aggregation by majority voting—against LLMs' epistemic risks. *Section 5* proposes propping up democracy's epistemic defenses with a mix of "blended" strategies, custom-made to address LLMs' weaknesses, that combine technological remedies with institutional tweaks. The *Conclusion* provides a summary and charts avenues for future research.

1. Democracy's Epistemic Conundrum

When assessing opportunities to enhance the performance of democratic decision-making via the employment of LLMs, I adopt the instrumental perspective of epistemic democrats who emphasize democracy's ability to come up with correct answers to difficult questions and resolve difficult problems (Landemore 2012; Goodin and Spiekermann 2018). At the same time, I also concur with Holst and Molander (2019, 542) that "the people" can be collectively smart but cannot succeed without access to specialized knowledge. Dealing with the wicked social, environmental, and technological problems of our age necessitates expert consultations. However, it is not so simple to draw from such consultations the epistemic benefits for policymaking.

One problem is that expertise is fractured among a plethora of disciplines but the problems-to-be-solved demand a holistic approach. For instance, consider the current debates on AI regulation that require insights ranging from computer science to economics, or law (Smuha 2021). However, each of these domains of expertise brings its own specific perspectives, methodologies, and epistemic risks (cf. Biddle and Kukla 2017), adding layers of complexity to the policymaking process. Moreover, few countries are blessed with an abundance of world-class experts across the diverse fields, supported by a wealth of high-quality peer-reviewed literature. For most of the world, expert scarcity remains a pressing issue. Its presence is most palpable in the development context where it bites the hardest at those most vulnerable.

Another issue is the adversarial nature of democracy's epistemic environment (cf. Nguyen 2023). When it comes to discerning the reliable sources of expertise, laypeople struggle (Goldman 2021). While the simple heuristics they utilize in expert recognition can be fortified by institutional strategies, such as certification of experts or information filtering by reputable media, the public remains vulnerable to exploitation. Disinformation peddlers eagerly emulate any successful persuasion strategy of information providers and seek novel ways to kidnap the public's attention and trust (Gelfert 2018). Democracies' epistemic resilience is thus continually tested. Many fear that the ascent of LLMs and other tools of generative artificial intelligence will further aggravate the situation (Coeckelbergh 2022; Sætra 2023).

Finally, the human track-record in advisory expertise is far from spotless. Human experts occasionally confabulate, misremember facts, or commit errors in their professional judgment, and even simple algorithms can surpass them in various tasks (Kahneman, Sibony, and Sunstein 2021). They may also make authoritative claims that obscure the uncertainty inherent in their expertise (Manski 2011), resist admitting their mistakes, or engage in self-serving gatekeeping practices (Koppl 2018). Not to mention their dubious forecasting performance (Tetlock 2006). Moreover, the more they strive to be useful for democratic purposes, the further they move away from value-neutrality (Pamuk 2022).

Some of these facets of modern democracy's epistemic conundrum are likely immovable. They represent inevitable consequences of social development and technological progress in which complexity breeds complexity: social problems cannot be simplified to square neatly with disciplinary boundaries; information technologies cannot be rolled back before the eve of the epistemically fractured digital age. Therefore, democracies need to find ways to cope with their epistemic situation. I suggest that LLMs—while by no means an unalloyed blessing—could become a critical component of such a coping mechanism. Employed in the role of expert advisors they could make expertise more available and more accessible to the democratic decision-makers than ever before.

2. The Case for Machine Advisors

Expertise availability is critical for knowledge dissemination. Traditionally, it is bound to the experts' presence and readiness to assist non-experts when needed. In our context of automating advisory expertise for democratic assemblies, it refers more broadly to the degree of difficulty that the representatives encounter when trying to secure the source of the specialized pertaining to the problem at hand. The existing scarcity of human experts translates into substantial costs associated their employment. Even with ample financial resources at one's disposal, experts still need to be identified, booked in advance, and asked only for a limited time commitment. Importantly, availability depends on experts' incentives, which are not always aligned with social needs. For instance, career rewards for esoteric research and "publish or perish" pressures can disincentivize providing public service (cf. Akerlof 2020).

While the current LLMs remain unable to bridge the knowledge gaps left by human experts and do their own research, they offer freedom from access restrictions and temporal constraints. Their output is immediately available, eliminating the need for advanced scheduling or adherence to limited consultation windows. In contrast to human experts, who often require significant time to process information and deliver findings, LLMs are characterized by rapid response capabilities, limited only by processing speed. They enable extensive engagement on-demand, accommodating as many parallel interactions as required, a stark contrast to the more traditional approach of human experts limited to written reports and plenary hearings. This translates into a distinct availability advantage over the current advisory model.

Expertise accessibility refers to the ease with which non-experts—including the members of a democratic assembly—can assimilate expertise effectively to improve their decision-making performance. Assimilation involves 1) *understanding* the expert's meaning and grasping the implications of their insights; 2) *discourse control*, that is, the capacity to steer the expert to provide answers most relevant for the decision-maker; and 3) a capability of *synthesis* to integrate expert insight across different domains and translate it into actionable recommendations to address the problem at hand. As it appears, LLM advisors hold promise of offering a better service than their human counterparts in all these areas.

In terms of *understanding*, human experts often struggle with the "curse of knowledge," a phenomenon where their expertise makes it challenging to communicate with laypeople who lack similar background knowledge (cf. Tullis and Feder 2023). Experts' rigorous, complex vocabulary often riddled with probabilistic language presents a significant challenge to those unfamiliar with the subject matter. Translating assertions that make perfect sense against a disciplinary paradigm into the terms of a more common-sense, or "folk", model of reality is expertise on its own and may easily fail to achieve the desired effect (cf. Boyer and Petersen 2018). Moreover, experts often lack incentives to develop outreach skills, with the lack of time and academic recognition being quoted as barriers to a greater engagement in science communication (Sanz Merino and Tarhuni Navarro 2019). Overall, understanding human experts does not come easy even with the best intentions on both sides.

How about the *discourse control*? When the experts perform an advisory role for democratic assemblies, the stakes tend to be high, even exorbitant. And the experts' incentives are often not well-aligned with providing their most honest and accurate testimony. They must navigate not just the complexities of their discipline but also a labyrinth of external pressures—from the public, the various interest groups, and their peers. The public, perhaps on accord of its identity-protective cognition (Kahan 2017), does not always take kindly the testimonies that violate its preconceptions of what the correct answer needs to be (Norris 2023). The organized interest groups are awash with resources they are ready to provide to anyone willing to speak in ways that support their agenda (Oreskes and Conway 2010). The peers expect a degree of professional solidarity and induce conformity pressures that limit the heterogeneity of expert testimonies (Koppl 2018). Experts' ideological predispositions, too, can cloud their judgement, turning advice into advocacy. For instance, Kozlowski and Van

Gunten (2023) show that greater ideological commitment coincides with the willingness of expressing greater confidence on ideologically salient themes in economics.

In sum, human experts are not just advisors but actors within a broader sociopolitical ecosystem, who face conflicting motivations and pressures. These dynamics make it often hard for democratic representatives to steer the conversation with the experts in the direction most productive relative to the policy decision at hand.

Finally, there is the challenge of *synthesizing* knowledge. Human experts specialize in narrow domains and shy away from crossing disciplinary boundaries, which are, however, arcane to laypeople and fail to map neatly on the structure of the challenges at hand. The problem of narrow expertise has grown more pressing as the overall knowledge stock has expanded exponentially, making it impossible for any individual to master more than a tiny fraction of available information. Even if democratic representatives can access top experts in all the individual domains, integrating their insights into an informed and coherent model of the policy-relevant issues at hand remains a gargantuan task that easily gets off rails. As a result, the task of delivering a synthesis and deriving of actionable recommendations is often outsourced by democratic assemblies to expert committees. However, this is hardly an ideal solution. The political responsibility for the decisions—and legitimacy thereof—lies fully on the representatives' shoulders and the committees cannot square the need for specific recommendations with retaining a commitment to value neutrality (Pamuk 2021).

It thus appears important that LLMs possess a versatility edge over the human advisors. Their advantage is derived from the breadth of their training data that include vast quantities of experience and knowledge, far greater than what any human could possibly digest. It grants them the ability to emulate different perspectives and communicate fluently across disciplinary boundaries. For example, consider the intricate task of summarizing a scientific paper laden with field-specific terminology. An LLM can transmute this complexity into an accessible summary, without diluting the essence of the research, explain its essence using intuitive daily life examples and metaphors, and suggest how specifically it may be relevant to the task at hand. The current models already seem on par with human performance in text summarization, at least when it comes to news articles (Zhang et al. 2023) and this capability increasingly extends to translating technical jargon into everyday language (Lyu et al. 2023) or communication of complex ideas (Ayers et al. 2023). Future models are likely to become more capable in identifying parallels across disperse fields, doing high-quality literature research, highlighting contentious or disputed claims, or explaining the main competing views. Even without such uncertain advances, the LLMs seem to be eking toward empowering laypeople—including democratic representatives—to take a more autonomous stance when it comes to synthesizing knowledge to inform policy design.

Interestingly, LLMs can also aspire to a degree of impartiality hard to achieve to humans. Unlike human advisors who might be encumbered by industry affiliations or driven by the desire for professional aggrandizement, they offer advice untainted by personal ambitions or conflicts of interest. While it would be a severe error to presume their output value-neutral (see *Section 3*), their biases and implicit normative judgements—being the result of their training—at least lack the strategic sensitivity to financial and professional allegiances that burden their human counterparts.

To summarize, LLMs show significant promise to enhance the dissemination and assimilation of human expertise despite their present limitations in independently advancing knowledge frontiers. They are unconstrained by the scarcity, time limitations, and access restrictions that hamstring human experts. The LLMs' capacity to satisfy representatives' needs for understanding, discourse control, and synthesizing insights across disciplines suggests they could also outperform human advisors in expertise accessibility. However, LLMs' strengths stand alongside significant limitations. As advisors, they present distinct risks that require careful mitigation.

3. Outstanding Challenges

Remarkably, many of LLMs' capabilities emerge spontaneously just due to their growing scale, without being intentionally designed by their creators (Bowman 2023). This underscores an important challenge: while we can observe the models' impressive performance in response to our inputs, there is little human comprehension of their internal processes (Liu, Gan, and Tegmark 2023). Unfortunately, this opacity is at the root of the key risks of LLMs possible use in an advisory role.

LLMs' perhaps most often quoted problem is their "hallucinations." These are instances when LLMs produce syntactically coherent statements devoid of any factual foundation or "unfaithful to the provided source input" (Varshney et al. 2023, 1). Hallucinations can include fabricated references to academic papers, made-up legal precedents, fanciful descriptions of non-existent physical phenomena, or even false accusations of specific individuals. Despite the lack of veracity in these outputs, LLMs do not signal diminished confidence in their accuracy. Even worse, they may persist in their fallacies when the user is doubtful or offers a correction. In some cases, LLMs even resort to gaslighting, defending their original claim with some ingenuity while suggesting that it is the user who is mistaken. Clearly, such behavior is of considerable concern for the members of a democratic assembly, who would rely on LLMs' accuracy and reliability to support substantive decisions.

However, from the perspective of considering machine advisors for a democratic assembly, hallucinations should not dominate our list of concerns. This is partly because their prevalence is likely to be mitigated by technological interventions (Bowman 2023; Varshney et al. 2023). Even in the brief time window between the GPT-3.5 and GPT-4 models have been available, the hallucination rate has dropped significantly (Ali et al. 2023). Also, hallucinations are perhaps not exceedingly dangerous if the models are used properly in democratic settings (see *Sections 4* and *5*).

A more fundamental reason to avoid selecting out hallucinations as the main threat is that an "accuracy only" notion of advice is simply misguided (Pamuk 2022). As we have seen, the advisory role crucially involves tasks like synthesizing information into a coherent whole, prioritizing certain views over others, and framing the overall message for the lay recipient. In executing these pragmatic necessities, the values embedded in LLM's output become critical. Therefore, the focus should not be on hallucinations themselves, but rather on the potential mismatch in goals and values between the LLMs and their users, which could result in machine advisors imposing their values on democratic assemblies.

The threat of goal and value mismatch—termed *misalignment*—between humans and LLMs appears omnipresent and hard, if not impossible, to fully resolve (Christian 2020). Consider, for instance, that the current generation of LLMs is designed to observe the principles of helpfulness, harmlessness, and honesty (e.g., Bai, Jones, et al. 2022). Clearly, there exist tensions among these values and trade-offs that cannot be avoided. In many situations, these principles are mutually incompatible or contradictory. As a result, LLMs often prioritize harmlessness over helpfulness, withholding information that the user requests if they detect a possible safety or ethics violation. This introduces a distinct paternalist layer to the user-system interaction: it is the values of LLMs' creators—not necessarily shared by the users—

that inform, but not fully determine, the ways in which the model provides, refuses to provide, or perhaps just frames its answer. In such a case, we may say that the model is more aligned with its creators' rather than (some of) its users' objectives. Note that even the "creator alignment" is imperfect since the specific content and shape of the answers provided by the model results from an opaque interplay between its designers' wishes and its inner workings.

The quest for imbuing intelligent technologies with human values is—to some extent—technical. However, it is also clearly political. There is the diversity of values and cultures, as well as the tensions between competing worldviews and interests. LLMs thus need to be equipped with a normative setup that goes beyond fulfilling a specific user's wishes. Therefore, they end up being more aligned with some worldviews than others and closer to some political outlook than another (Hartmann, Schwenzow, and Witte 2023; Atari et al. 2023). In other words, LLMs inevitably take a political stance.

If we allow the LLM creators to choose, or at least heavily influence, which specific stance that will be, a clear danger to democracy arises. Negotiating between conflicting interests and perspectives is the very purpose of its proceedings—a community's value setup is discovered, revealed, and legitimized during them. An implementation of allegedly "purely technical" resolution of normative and preferential diversity would undermine democracy's normative autonomy. Key value judgements that need to emerge from the democratic process—or at least to be confirmed by it—would be smuggled in during the LLM's construction and training. This is less a problem of misalignment, in the fundamental sense of making the machines do (some)one's bidding, than the venerable problem of *value imposition*: a third party—in this case the LLM builders—impresses their values externally on a community, perhaps sneaking them into a service they provide. If this service were used prominently enough, such as for the purpose of advising democratic representatives, major legitimacy and autonomy concerns ensue.

To summarize, LLMs may lack social ties, epitomizing a distinctly “alien” intelligence unentangled in human quarrels. However, this cannot suffice for turning them into unbiased advisors. Even the most sophisticated training methods cannot resolve inherent normative contradictions and trade-offs without overstepping into the political realm. From the perspective of a democratic assembly, where the preservation of normative autonomy is paramount, these risks could prove fatal. Does their presence undermine the potential to employ LLMs in place of human experts?

4. Is Democracy Ready?

Fortunately, democracy was not born yesterday. It represents a time-tested system that has proven its ability to deal with epistemic and other adversities. As such, it possesses robust defenses. People themselves—including experts—occasionally “hallucinate,” providing confused and inaccurate testimonies. People themselves—including experts—are often “misaligned,” following values and goals that differ from those of their clients. Finally, people themselves—including experts—can engage in value imposition. Are democracy’s bulwarks unassailable enough to cope with the machine advisors, too?

To retain focus, let us consider a stylized case of a democratic assembly that consists of several hundred elected representatives with regular competence tasked by addressing a complex problem, perhaps by crafting a new regulation. The assembly then proceeds in three stages: 1) expert consultation, 2) deliberation, and 3) majority vote. At the initial expert consultation stage, LLMs will be considered as an alternative to human experts. The latter two stages shall remain reserved for humans only. With this process in mind, do the existing institutional mechanisms appear capable of reining in machine advisors' limitations and threats?

In *Section 3*, we contemplated the various reasons why LLMs could wreak epistemic havoc. However, if used at the consultation stage, machine experts also introduce one potentially critical advantage. It lies in the degree to which the individual representatives can make up their mind independently of the others. The research on epistemic democracy highlights the need for a degree of individual independence in approaching the issue at hand (Goodin and Spiekermann 2018). This prevents epistemically detrimental phenomena, such as groupthink or information cascades (Solomon 2006; Sunstein 2017), from taking root right from the start. LLM advisors provide each person with a space to leverage their own unique perspective, informed by personal experiences, interests, and idiosyncrasies. Improved availability and accessibility of expertise are of key importance in this context. Today, expert consultations have limited space to contribute to independence. Reports and hearings are adapted for collective consumption and may paradoxically work to cement partisan divides and boosting conformity, especially where the testimonies pertain to politicized issues (Kahan 2017). Few representatives have access to a relevant expert team on their own and lobbying offers information possibly tainted or cherry-picked by the special interest that provides it. Compared to the *status quo*, LLMs could boost independence significantly. If epistemic democrats' theoretical arguments are correct, this promises decisive gains for democracy's epistemic performance (Landemore 2012; Goodin and Spiekermann 2018; Špecián 2022).

When we venture beyond the initial expert consultation stage, deliberation and aggregation through majority voting both represent powerful instruments in democracy's arsenal to catalyze collective intelligence and assert normative authority (Landemore 2012; 2020). Deliberation can be defined as a communicative procedure promoting substantive, balanced, and civil discussion (List 2018, 468). The empirical record highlights its capabilities to counter epistemic risks, such as polarization (Fishkin et al. 2021) or psychological biases (Dryzek et al. 2019). In the context of LLM advisors, deliberation can be effective in mitigating many of LLMs signature weaknesses including hallucinations, lack of ability to provide transparent methodologies behind their advice, and value imposition through implicit normative assumptions. Deliberation, by its very nature, demands reasoned justifications for any claim (Mansbridge et al. 2010), offering a robust platform for challenging the veracity of machine-generated advice. Public scrutiny serves as a powerful check against both hallucinations and opaque methodologies.

Furthermore, the deliberative process sets the stage for making normative assumptions explicit and examining them systematically (Bächtiger et al. 2018). Thus, it mitigates the risk of value imposition. The iterative process of an assembly allows for multiple rounds of scrutiny and amendment (despite its limitations such as partisan divides and time constraints). As such, it offers ample opportunities for vetting the LLMs' advice. There also exists suggestive evidence that politicians outperform the regular citizen when it comes to the quality of deliberation (Strandberg et al. 2021). Therefore, being manipulated by LLMs should present a lesser risk in democratic assemblies than in more general contexts like general elections.

Deliberation reveals, clarifies, and sometimes reconciles conflicting viewpoints. However, in realistic conditions like those of the modern parties-based representative democracy, it rarely leads to consensus or achieves its full theoretical potential (Bächtiger et al. 2018). Aggregation by voting remains necessary to arrive at a resolution. It also provides a way to further enhance epistemic performance, since its power lies beyond mere counting of votes. Even the simple majority voting procedure represents a powerful epistemic engine capable of filtering out noise and revealing collective wisdom. The principle at work here is the law of large numbers: in the context of a democratic assembly, it means that as more votes are cast, the collective decision becomes increasingly purified from random error and may converge upon the most rational and informed choice (Goodin and Spiekermann 2018).

Now, let us assess the expected efficacy of democratic mechanisms against the idiosyncratic errors and biases inherent in LLMs. For instance, consider a situation when a LLM provides faulty or biased advice to a representative. During deliberation, some, but not all, of these issues are caught and corrected. However, what slips through the cracks of deliberation are more likely random errors than systematic problems. And these errors are where aggregation can best perform its miracles. Because they are random and individual-specific—that is, unlikely to simultaneously affect large swaths of the assembly—they tend to cancel each other out when the votes are tallied. What remains is more likely an accurate reflection of factual matters and of the distribution of the representatives' normative positions. Diversity in the assembly may amplify this effect since it ensures that no single error or bias dominates the aggregate decision (Page 2008). Instead, the biases and errors offset each other, resulting in a more balanced and nuanced collective choice.

Still, the existing mechanisms, while powerful, are not invincible. LLMs' flaws could prove more difficult to counter than the flaws of human advisors. As such, they may yet breach democracy's defenses. While nobody can see into the experts' heads, they are still human—LLMs, in contrast, represent inscrutable black boxes with possibly rather alien inner workings. The lack of interpretability of the huge matrices of floating numbers that constitute their internal states and the lack of explainability of the reasons why they come up with a particular answer represent some of the top concerns (Christian 2020). It remains to be seen to what extent progress can be made in increasing LLMs' transparency and empowering humans to verify and better be able to adjust the values embedded in models.

5. Reining in Machine Advisors

As things stand, it appears likely that safe incorporation of machine advisors into democratic proceeding would require additional safeguards beyond the traditional democratic remedies. For instance, democratic assemblies could require the LLMs intended for the advisory role to be constructed by a custom-made process that requires use of specific training datasets and fine-tuning protocols. One possible approach involves training the models on legislative records and writings about deliberative democracy, potentially enhanced by reinforcement learning scripts designed to optimize their advisory capabilities. This could improve the effectiveness of LLMs as facilitators of human reasoning in policy discussions within a democratic context.

Do note, however, that such interventions are never merely technical measures, but complex amalgams of "technofixes" with institutional design choices that include key normative judgements. As such, they should not be left to the prerogative of the model builders. For example, a democratically sanctioned LLM ground rules of behavior could replace similar structures erected somewhat haphazardly by the current model builders (Bai, Kadavath, et al. 2022; Bowman 2023). Steps in this direction have already been taken. Meta's Community Forum has been used "to generate feedback on the governing principles people want to see reflected in new AI technologies" (Clegg 2023) using deliberative methods; OpenAI's has bid to sponsor projects to develop democratic processes for AI oversight (Zaremba et al. 2023); Anthropic (2023b) experiments with providing a democratically sanctioned constitution for its models. Still, a more systematic and assertive approach where the shape and weight of the democratic input less depends on the corporate goodwill would offer a preferable future pathway.

Another front of progress can be opened my establishing processes to boost the representatives' competence in the work with machine advisors. The representatives could undergo a specialized training session to increase their proficiency in prompt engineering—that is, crafting queries that elicit the maximum advisory performance from LLMs—and critically assessing the model output. While the representatives' time is, of course, precious, the existing research suggests that even simple techniques can offer significant epistemic returns (Gigerenzer 2010; Hertwig and Grüne-Yanoff 2017). In this vein, we should explore adapting such existing techniques to the new purpose of human-LLM interaction. For instance, "boosts" are methods developed to improve decision-making performance by learning simple heuristic techniques optimized to address the most frequent sources of bias and error (Hertwig and Grüne-Yanoff 2017). While more research is needed on how boosts need to be adapted for the issues at hand, early attempts to provide prompt engineering instruction provide hope that the basics can be assimilated with relative ease, not requiring technical expertise (Meskó 2023). Innovative approaches, such as prompt pattern catalogues documenting examples of best practice (White et al. 2023), could further enhance its efficacy. Moreover, representatives' training protocols and manuals of use could be publicly accessible and open to scrutiny—perhaps in the form of massive open online sources—to promote transparency and put all the representatives on the level ground (cf. Špecián 2022, chap. 5).

Also, a possibility exists of simultaneously employing several independent LLMs for the purposes of cross-checking and validating their answers. After all, human lies and mistakes tend to be discovered through inconsistencies. Therefore, consensus identification appears a promising candidate for a viable strategy aimed at mitigating LLMs' hallucinations. It involves model diversity, that is, employing multiple LLMs, each trained on different datasets, perhaps using different algorithms. Unreliable or spurious information will likely result in diverging responses from the models, whereas accurate information will result in overlapping answers.

Finally, there is the especially tricky problem that a machine advisor will provide a factually correct but selective and partial statements, thus misleading the user and possibly committing value imposition. Here, introduction of adversarial proceedings appears a promising pathway. For instance, a debate between LLMs could facilitate judging the strength of competing expert arguments (Irving, Christiano, and Amodei 2018; Michael et al. 2023). However, the deployment of adversarial LLM debates, while addressing the issue of selective representation, does not venture far enough beyond the inadequate "accuracy only" view of expertise. In this context, Pamuk's (2022; 2021) analysis of the dynamics of human expertise in democracy appears pertinent. She promotes a more nuanced strategy of complementing the statements or recommendations delivered by expert bodies with dissenting opinions. Well-aware of the whole range of epistemic risks connected with the democratic use of expertise, Pamuk argues persuasively that such a strategy may best enable laypeople to appreciate the limits of different perspectives and reveal crucial information about the degree of expert consensus. It is well-suited to make transparent the assumptions, uncertainties, and value tradeoffs involved in the expert advice and empower the lay decision-makers. In the context of machine advisors, similar effect could be achieved by playing different LLMs, or instantiations thereof, against each other. LLMs can be queried—perhaps based on a (partially) standardized prompt template—to point out weaknesses of various positions, formulate counterarguments, and deliver dissenting statements. Most easily, a variant of this adversarial process can be achieved with a single model being asked to criticize its own arguments (Saunders et al. 2022). However, a cross-model implementation relying on model diversity could deliver an even stronger result.

Although employing these methods introduces some friction into the use of LLM advisors, the drawbacks are likely not insurmountable. Eventually, consensus identification and argument adjudication could be facilitated by an automated interface (cf. Small et al. 2023). This interface would relate user queries to several LLMs, evaluate the consistency of their responses and summarize the majority and dissent opinions. Such a more complex option comes with specific normative pitfalls, however, that pertain to the proper setup of the evaluation and the capability to provide it in ways which are both democratically legitimate and pragmatically effective.

Conclusion

Above, I propose a novel approach to weave LLMs into democratic decision-making as a replacement for human advisors. Recognizing their potential risks, I still find their strengths—particularly the availability and accessibility boost they provide with respect of specialized knowledge—could significantly enhance the ability of democratic bodies to tackle complex policy challenges.

Of course, their integration as “machine advisors” needs to be approached with caution. Accordingly, I have outlined a strategy for incorporating them into democratic practices, drawing on insights from social epistemology and democratic theory. Traditional democratic practices like deliberation and voting represent powerful tools to neutralize risks such as hallucinations, misalignment, and external value imposition by LLMs. Moreover, to strengthen these existing defenses, I propose several blended techno-institutional remedies, such as include specialized training for representatives or the utilization of model diversity for consistency-testing and adversarial proceedings, that bear significant promise.

As far as limitations are concerned, my exploration has focused specifically on LLMs' potential as advisors within democratic assemblies. Consideration of other context may lead to different conclusions. At the same time, my model case of a democratic assembly provides a useful point of departure for these additional analyses. Also, my assumptions about the machine advisors' capabilities are relatively conservative, given the current pace of LLMs' progress. While not without risk, such an approach has the advantage of avoiding speculation about future capability gains and can stick to a scenario where not too many variables are fundamentally different from the *status quo*. This provides an opportunity to build fluently upon the existing streams of research in democratic theory and social epistemology. And as long as the future AI technology shares the principial strengths and weaknesses of the existing LLMs, the gist of my claims shall remain intact.

To finetune the specifics of LLM's institutional integration, this conceptual work must eventually be supplemented by empirical comparisons between human and machine experts on an even playfield. For instance, experiments could pit human experts against LLMs on advisory tasks within simulated democratic assemblies, using metrics for forecasting performance and user satisfaction. Researchers could also study interactions between people and their machine advisors using varying formats such as open-ended consultation, structured interviews, or restricted prompts. Additionally, long-term trials integrating LLMs into simulated policy development processes would prove insightful.

In conclusion, this paper illustrates how realizing AI's potential will not require merely adapting AI to democratic institutions but also adapting democratic institutions to AI. Even modest performance gains relative to the *status quo* could compound over time into better policies and prove highly consequential given the gravity of 21st century's challenges. Embracing LLMs' potential promises to enhance our collective ability to manage the outstanding crises and sustain democratic governance. This, in turn, appears necessary to ascertain that further AI progress will remain broadly beneficial to humankind.

Statement on the use of generative AI: During the preparation of this work the author used Claude 2 and GPT-4 in order to improve his language and style. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Declaration of interest statement: The author reports there are no competing interests to declare.

References

Akerlof, George A. 2020. "Sins of Omission and the Practice of Economics." *Journal of Economic Literature* 58 (2): 405–18. https://doi.org/10.1257/jel.20191573.

Ali, Rohaid, Oliver Y. Tang, Ian D. Connolly, Jared S. Fridley, John H. Shin, Patricia L. Zadnik Sullivan, Deus Cielo, et al. 2023. "Performance of ChatGPT, GPT-4, and Google Bard on a Neurosurgery Oral Boards Preparation Question Bank." *Neurosurgery*, June, 10.1227/neu.0000000000002551. https://doi.org/10.1227/neu.0000000000002551.

Alves, Pedro, and Bruno Pereira Cipriano. 2023. "The Centaur Programmer -- How

Kasparov's Advanced Chess Spans over to the Software Development of the Future." arXiv. https://doi.org/10.48550/arXiv.2304.11172.

Anthropic. 2023a. "Model Card and Evaluations for Claude Models." https://www-files.anthropic.com/production/images/Model-Card-Claude-2.pdf.

———. 2023b. "Collective Constitutional AI: Aligning a Language Model with Public…." Anthropic. October 17, 2023. https://www.anthropic.com/index/collective-constitutional-ai-aligning-a-language-model-with-public-input.

Atari, Mohammad, Mona J. Xue, Peter S. Park, Damián Ezequiel Blasi, and Joseph Henrich. 2023. "Which Humans?" Preprint. PsyArXiv. https://doi.org/10.31234/osf.io/5b26t.

Ayers, John W., Adam Poliak, Mark Dredze, Eric C. Leas, Zechariah Zhu, Jessica B. Kelley, Dennis J. Faix, et al. 2023. "Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum." *JAMA Internal Medicine*, April. https://doi.org/10.1001/jamainternmed.2023.1838.

Bächtiger, Andre, John S. Dryzek, Jane Mansbridge, and Mark Warren. 2018. "Deliberative Democracy: An Introduction." In *The Oxford Handbook of Deliberative Democracy*, edited by Andre Bächtiger, John S. Dryzek, Jane Mansbridge, and Mark Warren, xxii–32. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198747369.013.50.

Bai, Yuntao, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, et al. 2022. "Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback." https://doi.org/10.48550/ARXIV.2204.05862.

Bai, Yuntao, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, et al. 2022. "Constitutional AI: Harmlessness from AI Feedback." arXiv.

https://doi.org/10.48550/arXiv.2212.08073.

Bakker, Michiel A., Martin J. Chadwick, Hannah R. Sheahan, Michael Henry Tessler, Lucy Campbell-Gillingham, Jan Balaguer, Nat McAleese, et al. 2022. "Fine-Tuning Language Models to Find Agreement among Humans with Diverse Preferences." arXiv. https://doi.org/10.48550/arXiv.2211.15006.

Biddle, Justin B., and Rebecca Kukla. 2017. "The Geography of Epistemic Risk." In *Exploring Inductive Risk: Case Studies of Values in Science*, edited by Kevin C. Elliott and Ted Richards, 215–37. Oxford University Press.

Borji, Ali, and Mehrdad Mohammadian. 2023. "Battle of the Wordsmiths: Comparing ChatGPT, GPT-4, Claude, and Bard." SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.4476855.

Bowman, Samuel R. 2023. "Eight Things to Know about Large Language Models." arXiv. http://arxiv.org/abs/2304.00612.

Boyer, Pascal, and Michael Bang Petersen. 2018. "Folk-Economic Beliefs: An Evolutionary Cognitive Model." *Behavioral and Brain Sciences* 41. https://doi.org/10.1017/S0140525X17001960.

Christian, Brian. 2020. *The Alignment Problem: Machine Learning and Human Values*. First edition. New York, NY: W.W. Norton & Company.

Clegg, Nick. 2023. "Bringing People Together to Inform Decision-Making on Generative AI." *Meta* (blog). June 22, 2023. https://about.fb.com/news/2023/06/generative-ai-community-forum/.

Coeckelbergh, Mark. 2022. “Democracy, Epistemic Agency, and AI: Political Epistemology in Times of Artificial Intelligence.” *AI and Ethics*, November. https://doi.org/10.1007/s43681-022-00239-4.

Dryzek, John S., André Bächtiger, Simone Chambers, Joshua Cohen, James N. Druckman, Andrea Felicetti, James S. Fishkin, et al. 2019. “The Crisis of Democracy and the Science of Deliberation.” *Science* 363 (6432): 1144–46. https://doi.org/10.1126/science.aaw2694.

Fishkin, James, Alice Siu, Larry Diamond, and Norman Bradburn. 2021. “Is Deliberation an Antidote to Extreme Partisan Polarization? Reflections on ‘America in One Room.’” *American Political Science Review*, July, 1–18. https://doi.org/10.1017/S0003055421000642.

Gelfert, Axel. 2018. “Fake News: A Definition.” *Informal Logic* 38 (1): 84–117. https://doi.org/10.22329/il.v38i1.5068.

Gigerenzer, Gerd. 2010. *Rationality for Mortals: How People Cope with Uncertainty*. Evolution and Cognition. New York ; Oxford: Oxford University Press.

Goldman, Alvin I. 2021. “How Can You Spot the Experts? An Essay in Social Epistemology.” *Royal Institute of Philosophy Supplement* 89 (May): 85–98. https://doi.org/10.1017/S1358246121000060.

Goodin, Robert E., and Kai Spiekermann. 2018. *An Epistemic Theory of Democracy*. Oxford, United Kingdom: Oxford University Press.

Hartmann, Jochen, Jasper Schwenzow, and Maximilian Witte. 2023. “The Political Ideology of Conversational AI: Converging Evidence on ChatGPT’s pro-Environmental, Left-Libertarian Orientation.” arXiv. https://doi.org/10.48550/arXiv.2301.01768.

Hertwig, Ralph, and Till Grüne-Yanoff. 2017. "Nudging and Boosting: Steering or Empowering Good Decisions." *Perspectives on Psychological Science* 12 (6): 973–86. https://doi.org/10.1177/1745691617702496.

Holst, Cathrine, and Anders Molander. 2019. "Epistemic Democracy and the Role of Experts." *Contemporary Political Theory* 18 (4): 541–61. https://doi.org/10.1057/s41296-018-00299-4.

Irving, Geoffrey, Paul Christiano, and Dario Amodei. 2018. "AI Safety via Debate." https://doi.org/10.48550/ARXIV.1805.00899.

Kahan, Dan M. 2017. "Misconceptions, Misinformation, and the Logic of Identity-Protective Cognition." SSRN Scholarly Paper ID 2973067. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2973067.

Kahneman, Daniel, Olivier Sibony, and Cass R. Sunstein. 2021. *Noise: A Flaw in Human Judgment*. First edition. New York: Little, Brown Spark.

Koppl, Roger. 2018. *Expert Failure*. 1 Edition. Cambridge Studies in Economics, Choice, and Society. New York: Cambridge University Press.

Kozlowski, Austin C., and Tod S. Van Gunten. 2023. "Are Economists Overconfident? Ideology and Uncertainty in Expert Opinion." *The British Journal of Sociology* 74 (3): 476–500. https://doi.org/10.1111/1468-4446.13001.

Landemore, Hélène. 2012. *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton University Press.

———. 2020. *Open Democracy: Reinventing Popular Rule for the Twenty-First Century*.

Princeton: Princeton University Press.

List, Christian. 2018. “Democratic Deliberation and Social Choice: A Review.” In *The Oxford Handbook of Deliberative Democracy*, edited by Andre Bächtiger, John S. Dryzek, Jane Mansbridge, and Mark Warren, 462–89. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198747369.013.14.

Liu, Ziming, Eric Gan, and Max Tegmark. 2023. “Seeing Is Believing: Brain-Inspired Modular Training for Mechanistic Interpretability.” https://doi.org/10.48550/ARXIV.2305.08746.

Lyu, Qing, Josh Tan, Michael E. Zapadka, Janardhana Ponnatapura, Chuang Niu, Kyle J. Myers, Ge Wang, and Christopher T. Whitlow. 2023. “Translating Radiology Reports into Plain Language Using ChatGPT and GPT-4 with Prompt Learning: Results, Limitations, and Potential.” *Visual Computing for Industry, Biomedicine, and Art* 6 (1): 9. https://doi.org/10.1186/s42492-023-00136-5.

Mansbridge, Jane, James Bohman, Simone Chambers, David Estlund, Andreas FÃ¸llesdal, Archon Fung, Cristina Lafont, Bernard Manin, and JosÃ© luis MartÃ. 2010. “The Place of Self-Interest and the Role of Power in Deliberative Democracy*.” *Journal of Political Philosophy* 18 (1): 64–100. https://doi.org/10.1111/j.1467-9760.2009.00344.x.

Manski, Charles F. 2011. “Policy Analysis with Incredible Certitude.” *The Economic Journal* 121 (554): F261–89. https://doi.org/10.1111/j.1468-0297.2011.02457.x.

Meskó, Bertalan. 2023. “Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial.” *Journal of Medical Internet Research* 25 (October): e50638. https://doi.org/10.2196/50638.

Michael, Julian, Salsabila Mahdi, David Rein, Jackson Petty, Julien Dirani, Vishakh Padmakumar, and Samuel R. Bowman. 2023. “Debate Helps Supervise Unreliable Experts.” arXiv. https://doi.org/10.48550/arXiv.2311.08702.

Nguyen, C. Thi. 2023. “Hostile Epistemology.” *Social Philosophy Today* 39: 9–32. https://doi.org/10.5840/socphiltoday2023391.

Norris, Pippa. 2023. “Cancel Culture: Myth or Reality?” *Political Studies* 71 (1): 145–74. https://doi.org/10.1177/00323217211037023.

OpenAI. 2023. “GPT-4 Technical Report.” arXiv. https://doi.org/10.48550/arXiv.2303.08774.

Oreskes, Naomi, and Erik M. Conway. 2010. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. 1st edition. New York: Bloomsbury Press.

Ovadya, Aviv. 2023a. “‘Generative CI’ through Collective Response Systems.” arXiv. https://doi.org/10.48550/arXiv.2302.00672.

———. 2023b. “Reimagining Democracy for AI.” *Journal of Democracy* 34 (4): 162–70. https://doi.org/10.1353/jod.2023.a907697.

Page, Scott E. 2008. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. New edition with a New preface by the author edition. Princeton: Princeton University Press.

Pamuk, Zeynep. 2021. *Politics and Expertise: How to Use Science in a Democratic Society*. 1st ed. Princeton: Princeton University Press.

———. 2022. “COVID-19 and the Paradox of Scientific Advice.” *Perspectives on Politics* 20 (2): 562–76. https://doi.org/10.1017/S1537592721001201.

Sætra, Henrik Skaug. 2020. “A Shallow Defence of a Technocracy of Artificial Intelligence: Examining the Political Harms of Algorithmic Governance in the Domain of Government.” *Technology in Society* 62 (August): 101283. https://doi.org/10.1016/j.techsoc.2020.101283.

———. 2023. “Generative AI: Here to Stay, but for Good?” *Technology in Society* 75 (November): 102372. https://doi.org/10.1016/j.techsoc.2023.102372.

Sanz Merino, Noemí, and Daniela H Tarhuni Navarro. 2019. “Attitudes and Perceptions of Conacyt Researchers towards Public Communication of Science and Technology.” *Public Understanding of Science* 28 (1): 85–100. https://doi.org/10.1177/0963662518781466.

Saunders, William, Catherine Yeh, Jeff Wu, Steven Bills, Long Ouyang, Jonathan Ward, and Jan Leike. 2022. “Self-Critiquing Models for Assisting Human Evaluators.” https://doi.org/10.48550/ARXIV.2206.05802.

Small, Christopher T., Ivan Vendrov, Esin Durmus, Hadjar Homaei, Elizabeth Barry, Julien Cornebise, Ted Suzman, Deep Ganguli, and Colin Megill. 2023. “Opportunities and Risks of LLMs for Scalable Deliberation with Polis.” arXiv. https://doi.org/10.48550/arXiv.2306.11932.

Smuha, Nathalie A. 2021. “From a ‘Race to AI’ to a ‘Race to AI Regulation’: Regulatory Competition for Artificial Intelligence.” *Law, Innovation and Technology* 13 (1): 57–84. https://doi.org/10.1080/17579961.2021.1898300.

Solomon, Miriam. 2006. “Groupthink versus The Wisdom of Crowds: The Social Epistemology of Deliberation and Dissent.” *The Southern Journal of Philosophy* 44 (S1): 28–

42. https://doi.org/10.1111/j.2041-6962.2006.tb00028.x.

Špecián, Petr. 2022. *Behavioral Political Economy and Democratic Theory: Fortifying Democracy for the Digital Age*. 1 Edition. Routledge Frontiers of Political Economy. New York, NY: Routledge.

Strandberg, Kim, Janne Berg, Thomas Karv, and Kim Backström. 2021. "When Citizens Met Politicians – the Process and Outcomes of Mixed Deliberation According to Participant Status and Gender." *Innovation: The European Journal of Social Science Research* 34 (5): 638–55. https://doi.org/10.1080/13511610.2021.1978282.

Sunstein, Cass R. 2017. *#Republic: Divided Democracy in the Age of Social Media*. Princeton ; Oxford: Princeton University Press.

Tetlock, Philip E. 2006. *Expert Political Judgment: How Good Is It? How Can We Know?* New Ed edition. Princeton, N.J.: Princeton University Press.

Tullis, Jonathan G., and Brennen Feder. 2023. "The 'Curse of Knowledge' When Predicting Others' Knowledge." *Memory & Cognition* 51 (5): 1214–34. https://doi.org/10.3758/s13421-022-01382-3.

Varshney, Neeraj, Wenlin Yao, Hongming Zhang, Jianshu Chen, and Dong Yu. 2023. "A Stitch in Time Saves Nine: Detecting and Mitigating Hallucinations of LLMs by Validating Low-Confidence Generation." arXiv. https://doi.org/10.48550/arXiv.2307.03987.

White, Jules, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C. Schmidt. 2023. "A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT." arXiv. https://doi.org/10.48550/arXiv.2302.11382.

Zaremba, Wojciech, Arka Dhar, Lama Ahmad, Tyna Eloundou, Shibani Santurkar, Sandhini Agarwal, and Jade Leung. 2023. “Democratic Inputs to AI.” 2023. https://openai.com/blog/democratic-inputs-to-ai.

Zhang, Tianyi, Faisal Ladhak, Esin Durmus, Percy Liang, Kathleen McKeown, and Tatsunori B. Hashimoto. 2023. “Benchmarking Large Language Models for News Summarization.” https://doi.org/10.48550/ARXIV.2301.13848.

Zhou, Jiawei, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. “Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions.” In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–20. CHI ’23. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3544548.3581318.