Digital Homunculi

Reimagining Democracy Research with Generative Agents

Petr Špecián

petr.specian@fsv.cuni.cz

Institute of Political Studies, Faculty of Social Sciences, Charles University, Czechia

Abstract: The pace of technological change continues to outstrip the evolution of democratic institutions, creating an urgent need for innovative approaches to democratic reform. However, the *experimentation bottleneck*—characterized by slow speed, high costs, limited scalability, and ethical risks—has long hindered progress in democracy research. This paper proposes a novel solution: employing generative artificial intelligence (GenAI) to create synthetic data through the simulation of *digital homunculi*, GenAI-powered entities designed to mimic human behavior in social contexts. By enabling rapid, low-risk experimentation with alternative institutional designs, this approach could significantly accelerate democratic innovation. I examine the potential of GenAI-assisted research to mitigate current limitations in democratic experimentation, including the ability to simulate large-scale societal interactions and test complex institutional mechanisms. While acknowledging potential risks such as algorithmic bias, reproducibility challenges, and AI alignment issues, I argue that the benefits of synthetic data are likely to outweigh their drawbacks if implemented with proper caution. To address existing challenges, I propose a range of technical, methodological, and institutional adaptations. The paper concludes with a call for interdisciplinary collaboration in the development and implementation of GenAI-assisted methods in democracy research, highlighting their potential to bridge the gap between democratic theory and practice in an era of rapid technological change.

Keywords: Artificial Intelligence; Generative Agents; Democratic Innovation; Institutional Design; Synthetic Data; Experimental Political Science

Statement on the use of generative AI: During the preparation of this work, the author used Claude 3.5 Sonnet to elicit feedback and improve his language and style. The author reviewed and edited all the content of the paper as needed, taking full responsibility for it.

1. Introduction

In an era of rapid technological change and global crises, liberal democracy faces difficult challenges, ranging from political polarization to the renewed onslaught of authoritarianism. While many scholars have theorized alternative models of democratic decision-making (e.g., Posner and Weyl 2018; Landemore 2020), a significant gap persists between democratic theory and practical reform. This gap stems largely from the slow pace, high costs, and ethical challenges inherent in empirically exploring democracy's vast institutional design space.

This paper argues that generative artificial intelligence (GenAI), if properly utilized, could catalyze experimental democracy research, narrowing the theory-practice divide in democratic innovation. Specifically, I propose that GenAI-created synthetic experimental data may ease the process of institutional innovation by facilitating rapid, cost-effective, and robust assessment of democratic reforms. This approach could reduce the hardships of evidence gathering and enable an iterative process of reform-testing and fine-tuning that precedes, complements, and even replaces traditional empirical studies with human participants.

GenAI, exemplified by large language models like ChatGPT, encompasses computationally intensive algorithms based on deep learning and transformer technology (LeCun, Bengio, and Hinton 2015; Vaswani et al. 2017). These models can generate human-like artifacts and achieve performance in a variety of cognitive tasks that equals or, increasingly, surpasses human capabilities (N. Jones 2024; C. R. Jones and Bergen 2024). Crucially, recent research shows that GenAIs can be turned into 'generative agents'—or *digital homunculi*—capable of providing believable simulations of human behavior in social contexts (Park et al. 2023; Argyle et al. 2023).

The potential of GenAI to revolutionize democracy research is particularly significant given the pressing need for institutional innovation. There are many serious technology-related challenges to democratic governance, including the rise of AI itself (Coeckelbergh 2024). If met with institutional rigidity, they threaten to undermine democracy's stability, which is conditional on the compatibility

between its political institutions and its technological substrate (Špecián 2022). Unfortunately, institutional change—especially if it is to be evidence-based and carefully crafted to deliver predictable results—remains exceedingly difficult, with democracies often constrained to trial-and-error tweaks of the existing rules.

Against this background, my paper addresses the following research question: *To what extent and under what conditions can GenAI-created synthetic data facilitate the empirical exploration of democracy's design space and drive democratic innovation?* To provide a tentative answer, I examine the prospects of utilizing GenAI-based digital homunculi to overcome current limitations in democracy research. Along the way, I also analyze the associated risks and challenges and propose strategies for leveraging synthetic data safely and effectively.

The contribution of this paper is threefold: 1. It introduces a novel framework for overcoming the long-standing experimentation bottleneck in democracy research by leveraging GenAI and digital homunculi to test and refine institutional innovations. 2. It provides a balanced analysis of the epistemological implications of using synthetic data in the social sciences, examining both the transformative potential and the difficult challenges this approach presents. 3. It outlines the possible steps toward the assimilation of GenAI-based methods, proposing several technical, methodological, and institutional adaptations to maximize the epistemic benefits for democracy research and—ultimately—pragmatic benefits for democracies and their citizens.

The paper is structured as follows: *Section 2* documents the gap between democratic theory and institutional reform, focusing on the experimental bottleneck. *Section 3* analyzes GenAI's promise to unlock the exploration of democracy's design space. *Section 4* addresses the risks and caveats associated with the use of synthetic data in democracy research. *Section 5* explores the possible strategies for mitigating the risks and maximizing the epistemic and pragmatic benefits of synthetic data. It proposes technical adaptations, methodological innovations, and institutional and cultural changes necessary to effectively integrate GenAI into democracy research.

2. The Experimentation Bottleneck

Democracy defies any neat definition, appearing in an almost inexhaustible wealth of forms and shapes (Schaffer and Gagnon 2023). This diversity underscores the vastness of the space of possible institutional structures that qualify as democratic. From these almost endless possibilities, only a small subset has been rigorously theorized. However, theorizing itself is not a binding constraint. Even with relatively meager resources invested in political theory, we already have a plethora of promising proposals for novel democratic institutional mechanisms. Instead, it is the testing and implementation of these innovations that present the key bottleneck.

The sluggish pace of institutional innovation contrasts starkly with the dynamic landscape of technological development. Where democratic institutions appear almost frozen in time, industrial revolutions fly by one after another. Especially important are those technological changes that reshape democratic discourse itself. Since democracy is a system where free debate among equals plays a critical role, any significant disruption in the ways knowledge is produced, disseminated, and assimilated can profoundly impact its fortunes (cf. Kurtulmus 2020).

So far, modern democratic institutions have proven resilient. Democracy's institutional foundations, often dating back to when information traveled at the speed of a galloping horse, have adapted to the age of mass media and kept facilitating human prosperity. However, the ongoing digitalization of democratic discourse and the 'epistemic democratization' that comes along with its user-created content, 'fake news,' and echo chambers are posing significant challenges (e.g., Sunstein 2017). And before this shock has even begun to subside, GenAI is swiftly arriving on the scene, promising further disruption.

Given this differential speed of institutional and technological innovation, it is increasingly difficult to ignore the possibility that legacy mechanisms of democratic decision-making may ultimately falter in resolving pressing contemporary issues (Špecián 2022). If these mechanisms prove unable to cope

with the challenges of the digital age, we face a dearth of well-tested alternatives. Therefore, accelerating empirical research of institutional innovations should be one of social scientists' top imperatives.

However, achieving robust pre-deployment testing of potential institutional reforms is fraught with difficulties. Ideally, researchers should be able to progress with relative ease from conceptual exploration to a functional prototype, rigorously test the prototype under realistic conditions, and iterate upon it when shortcomings become apparent. Only upon establishing sufficient confidence in an institutional mechanism's functionality under a spectrum of conditions should it be implemented at scale in high-stakes situations. This approach has proven itself in the realm of physical technologies—no automaker would dream of introducing a new car model without extensive performance and safety testing. Yet, paradoxically, there persist calls for radical reforms of our critical institutional infrastructure with scant empirical evidence that such reforms will yield the intended consequences.

These calls for reform likely stem not from negligence but from a deep-seated frustration—even desperation—in the face of mounting challenges to democracy. In the social sciences, particularly in the study of democratic institutions, implementing sufficiently robust testing procedures for innovations has been nearly impossible due to a multitude of constraints. While the experimental method is widely regarded as a cornerstone of scientific inquiry, its application in studying democratic reforms faces significant hurdles. These obstacles collectively form what I term the *experimentation bottleneck*—a set of interrelated challenges that severely limit our ability to gather reliable evidence about the efficacy of proposed institutional changes.

Prohibitive Costs

The financial burden of conducting large-scale democratic experiments represents a significant barrier to innovation. For instance, deliberative democracy experiments, typically involving several dozen to several hundred participants, often run into hundreds of thousands of US dollars in costs (OECD

2023). These costs create insurmountable obstacles for all but the best-funded researchers in affluent countries, leading to a number of problematic outcomes:

1. Research bias: The local and cultural concentration of institutions and researchers with sufficient financial resources may bias the types of questions asked and solutions proposed.
2. Limited replication: The expense of large-scale experiments hinders replication efforts.
3. Scale constraints: Even ambitious experiments typically involve only a few hundred participants, falling far short of the scale of actual communities.
4. Limited iteration: The high cost per experiment limits the number of iterations researchers can perform, hindering the refinement of institutional designs.

Overall, the financial barriers reduce the scale and scope of democratic experiments while also creating an exclusionary and unequal research environment. More cost-effective methods of institutional innovation testing are sorely needed.

Time Constraints and Ethical Hurdles

Democratic experimentation is notoriously time-consuming, with studies often taking years to design, implement, and analyze, not to mention the lengthy publication process. This protracted timeline is at odds with the rapid pace of technological and social change. By the time they are published, findings can already be obsolete. There is also a tension between the need for realistic testing and the imperative to protect participants from potential harm.

Time constraints:

1. Participant recruitment and retention: Assembling a representative sample and maintaining participation over time is difficult and time-consuming.

2. Regulatory approvals: Institutional review boards and ethics committees, while important for maintaining research integrity, can significantly delay or constrain innovative research designs.
3. Long-term follow-up: Assessing the enduring effects of democratic innovations often requires longitudinal studies, which can span years.

Ethical hurdles:

1. Informed consent: The requirement for informed consent pushes for experimental designs simple enough that the participants can be informed with relative ease.
2. Deception limitations: Ethical guidelines often restrict the use of deception in experiments, limiting researchers' ability to study certain aspects of political behavior, such as where an accurate understanding of the experimental situation can trigger preference falsification.
3. Risk mitigation: The need to minimize risks to participants can lead to overly sanitized experimental conditions that fail to capture the gravity of real-world high-stakes democratic processes.

The interplay of time constraints and ethical considerations often results in research processes that are too slow to keep pace with societal changes, let alone technological developments, and too sanitized to reflect real-world conditions. We need to develop methods that allow for more agile experimentation without compromising researchers' ethical commitments.

Limited Ecological Validity

Difficulties in recreating authentic democratic conditions in experimental settings severely constrain the results' generalizability and robustness. A deficit of ecological validity not only hampers our ability to confidently implement institutional innovations but also risks misleading policymakers about the potential impacts of democratic reforms:

1. Artificial settings: Laboratory experiments, while allowing for precise control of variables, tend to create environments far removed from the complex, messy reality of democratic processes.

2. Sample representativeness: Samples most often used due to convenience and cost considerations, such as university students or Western populations, may not accurately represent the broader populace in terms of demographics, cultural background, or psychological traits (cf. Henrich 2020).

3. Short-term focus: Most experiments occur over relatively brief periods, failing to capture the longer-term dynamics crucial to democratic processes.

4. Motivation mismatch: Experimental incentives often fail to replicate the complex motivations driving real-world political behavior.

5. Lack of institutional embeddedness: Experiments typically occur in isolation from the broader institutional ecosystem that shapes democratic processes.

Observational data or field experiments, while offering greater realism, are of limited use in testing unprecedented institutional reforms. Even where precedents exist, robust causal inference remains challenging, and it is often unclear how findings from one set of cultural and historical conditions translate to another, given that path- and context-dependency can profoundly shape institutional outcomes. Overall, the limited ecological validity of the current methods further highlights the need for innovative methodologies that produce results more immediately relevant to complex democratic realities.

*

The challenges outlined above collectively create a formidable barrier to evidence-based institutional reform and present us with daunting trade-offs. Efforts to boost validity increase time requirements and costs; efforts to cut costs further undermine validity and may compromise research ethics; efforts

to achieve methodological perfection are infeasible for an overwhelming majority of democracy researchers due to their resource constraints. Even when successfully attempted, by the time they deliver results, the sociopolitical landscape may have already transformed in ways detrimental to the relevance of the findings.

In practical terms, the difficulties in testing new institutional designs create a strong status quo bias. Existing institutions, despite their flaws, often persist due to the perceived risks and uncertainties associated with their suggested alternatives. Such "conservatism for lack of viable options" can lead to institutional sclerosis, where democracies struggle to adapt to changing societal needs and technological landscapes and leave them ill-equipped to handle new challenges. This, in turn, may ultimately erode democracies' perceived legitimacy, making the autocratic governance with its (typically false) promises of technocratic efficiency gain appeal (Mittiga 2021).

Therefore, overcoming the experimentation bottleneck—by developing new approaches that allow more cost-effective, agile, and robust pre-implementation testing of democratic reforms—is not merely an academic exercise but a vital endeavor for the future of democratic governance. While the challenges outlined above may not be resolvable in full, even partial mitigation would bring substantial benefits. The central hypothesis of this paper is that recent advancements in GenAI may hold the key to making these approaches feasible.

3. How GenAI Could Make a Difference

The experimentation bottleneck in democracy research, as sketched in *Section 2*, presents a formidable challenge to institutional innovation. However, let us explore if GenAI may offer at least a partial solution by producing *synthetic data* to reduce the present constraints. Synthetic data, broadly defined, refers to artificially generated information that mimics real-world observations. While primarily used in machine learning to address data scarcity (Nikolenko 2021), its potential in social sciences, particularly in democracy research, appears significant (Grossmann et al. 2023).

For the purposes of this paper, I will use ‘synthetic data’ to refer specifically to data that result from observation of GenAI-based simulations of human behavior and social interactions. These simulations aim to model complex social and political scenarios for research in democracy and institutional design, potentially allowing us to explore situations that would be impractical or ethically challenging to test in real-world settings.

GenAI systems are trained on vast datasets containing many terabytes of human texts and other artifacts. The goal of using synthetic data for social science research is to leverage the knowledge embedded in these texts and artifacts to reveal insights about human behavior and social dynamics that are not readily accessible through conventional methods. There are a number of ways how this knowledge can be accessed, such as using GenAI to predict real-world events Pole (Schoenegger et al. 2024), augmenting surveys (Kim and Lee 2024), or letting it participate in behavioral experiments (Mei et al. 2024). However, I want to concentrate on a completely novel approach that appears to bear the most revolutionary potential in democracy research, namely the employment of ‘generative agents’ (Park et al. 2023) or, as I prefer to call them, *digital homunculi*. These are GenAI-powered entities optimized to simulate individual humans and engage in human-like social interactions.

Albeit still in a nascent stage, digital homunculi have already demonstrated a remarkable capacity to credibly roleplay assigned personas with specific demographic characteristics, cultural backgrounds, and personality traits, capable of engaging in relatively complex interactive tasks. For instance, Kaiya et al. (2023) had these agents solve a murder mystery, while Park et al. (2023) observed them organize a social event. Even today, the capabilities of digital homunculi could thus prove sufficient for smaller-scale, less detailed simulations that still bear attractive epistemic fruits. As technology progresses, digital homunculi are bound to become more capable of mimicking human behavior. Their use may eventually—and possibly rather soon, given the dynamics of GenAI development—enable researchers to create virtual populations that resemble the complexity of real-world democracies.

Several key potential benefits of using digital homunculi in democracy research are already apparent. If we manage to unlock their potential—which, however, is connected with overcoming significant hurdles (see *Section 4*)—we could expedite pre-implementation testing of democratic reforms:

Increased Speed and Efficiency

Perhaps the most significant advantage of GenAI in democracy research is the dramatic increase in speed and efficiency it offers (Bail 2024). Generation of synthetic data through digital homunculi outpaces the traditional data collection methods in several ways:

1. Rapid deployment: There is no need for time-consuming participant recruitment, physical congregation, or extensive training. For some research tasks, such as questionnaire administration or simple interaction scenarios, GenAI can be deployed almost instantaneously. For others, standardized templates will be increasingly available (see *Section 5*).

2. Accelerated simulations: GenAI systems can operate at speeds exceeding human cognitive processes. This capability allows for the compression of long-duration social phenomena into more manageable timeframes.

3. Reduced administrative burden: The use of digital homunculi eliminates much of the logistical overhead associated with human subject research. Tasks such as scheduling, consent management, and data entry are largely automated or eliminated.

4. Rapid iteration: GenAI allows for swift iteration based on empirical feedback. Researchers can quickly tweak their designs and retest hypotheses, enabling faster epistemic progress.

The speed advantages of GenAI extend beyond the data collection phase. The standardization of GenAI tools and methodologies could facilitate easier comparison and meta-analysis of results across different studies, enabling faster accumulation of robust and generalizable findings in democracy research.

Cost Reduction and Accessibility

The financial implications of utilizing GenAI in democracy research may democratize access to potent research methodologies:

1. Operational cost reduction: While computationally intensive, the operational costs of GenAI are negligible compared to those associated with human participants. The primary costs for GenAI experiments are computational resources. These continue to decrease in price, and the efficiency of their use is growing. Between 2022-24, the costs per output token have fallen by 99 % while the output quality increased significantly (OpenAI 2024). For instance, Kayia et al. (2023) demonstrated a 30-100x decrease in costs compared to previous methods, running their digital homunculi at about $0.5 per (human) hour.

2. Democratization of research opportunities: The accessibility of GenAI and synthetic data levels the research landscape in epistemically fortuitous ways. The low cost of GenAI-enabled research can democratize the opportunities for experimentation, diversifying the perspectives brought to bear on democratic innovation.

3. Bolder research design: The reduced cost per experiment allows researchers to take more risks in their experimental designs. Currently, the high costs of traditional experiments incentivize researchers to prioritize safe, incremental studies. With GenAI, researchers can explore a broader range of ideas, including the ‘tail’ ones that are unlikely to work but promise especially large benefits if they do.

4. Enhanced replication capabilities: As Ioannidis (2005) memorably suggested, “most published research findings are false.” Synthetic data could provide a fast, cost-effective means of verifying results.

The cost advantages alone could bring a major overhaul to the research methodologies. For instance, even if we retain a cautious stance on the epistemic worth of synthetic data, it would still enable a two-tiered approach: an initial broad exploration using GenAI, followed by more focused, resource-

intensive studies with human participants, building on the most promising preliminary results. This approach would improve resource efficiency and increase the likelihood of breakthroughs by allowing researchers to cast a wider net in their initial investigations.

Greater Realism and Scale

GenAI may also serve to enhance the scale and—somewhat paradoxically—realism of democratic experiments:

1. Diverse participant pools: GenAI offers enhanced flexibility in subject composition. Researchers can tailor the demographic, cultural, and even historical characteristics of their experimental subjects with ease (Argyle et al. 2023).

2. Mitigation of self-selection bias and preference falsification: By creating digital homunculi with specified characteristics (Argyle et al. 2023), researchers can overcome the self-selection bias often present in human subject research. Also, compared to humans, digital homunculi can possess a lesser propensity to falsify their preferences under observation (cf. Dillion et al. 2023, 597).

3. Large-scale simulations: GenAI might soon enable researchers to conduct experiments involving thousands or even millions of digital homunculi (Chan et al. 2024). This would open the pathway to simulating complex societal dynamics and emergent phenomena that only manifest at scale.

4. Complex scenario modeling: Researchers can model intricate, multi-faceted scenarios that would be difficult to recreate in real-world experimental settings. This includes simulating rare but consequential events such as constitutional crises or the breakdown of democratic norms.

Advantages in realism and scale may result in a situation where synthetic data is not only cheaper but also allows us to provide more reliable answers to the pressing questions of institutional design. Scaling up to realistic population sizes or eliciting honest responses have long been hard to impossible

for social scientists in many crucial settings, including democratic innovation. With digital homunculi, we may be nearing a situation where these research constraints can finally be alleviated.

Ethical Advantages

Finally, the use of GenAI in democracy research can address some core ethical problems:

1. Ethical experimentation: Given the current consensus that GenAIs are not conscious entities (Hildt 2023), they cannot be harmed in the same way as human subjects. This allows for the exploration of scenarios that would be ethically untenable or potentially traumatic to human participants.

2. Controversial topics: GenAI enables the study of sensitive or controversial topics that can be difficult to explore with human subjects due to social desirability bias or fear of repercussions.

3. Reduction of harm: By using digital homunculi for initial exploratory research, we could reduce the number of human subjects needed in later stages, minimizing potential harm or discomfort to real individuals.

Ethical advantages allow researchers to study scenarios that were previously too perilous or simply off-limits. Such scenarios are often pragmatically important: consider stress-testing a political system to the limits of its endurance or assessing its ability to settle pitched conflicts between identity groups. This expansion of the research horizon has the potential to deepen our understanding of democratic crisis management or even democratic breakdown.

*

To summarize, the integration of GenAI into democracy research could provide us with the tools we need to enable swifter progress from institutional blueprints to prototypes, testing and refining our hypotheses and designs on our feet. The accelerated research cycle would enable a more effective exploration of the institutional design space. Providing a low-stakes environment for testing

theoretical proposals, GenAI simulations could help refine and validate ideas before empirical studies with human participants or actual implementation.

However, it must also be recognized that the potential epistemic gains are conditional on addressing a number of challenges associated with GenAI methodologies. Their promise can only be realized if we remain vigilant about GenAI's limitations and failure modes and sufficiently inventive in mitigating or circumventing them.

4. Risks and Caveats

While the potential benefits of GenAI-generated synthetic data for social sciences are significant, it is crucial to acknowledge and examine the associated risks. As the future development of technologies and methodological innovations remains unpredictable, these challenges may prove difficult or even impossible to overcome, undermining the promise outlined in the previous section.

The Alien Actress

At the heart of the challenges facing GenAI in democracy research lies what Yudkowsky (in Fridman 2023) calls the 'alien actress' problem. While GenAIs increasingly excel at mimicking human speech and creating human-like artifacts, we need to avoid anthropomorphizing their internal workings. An alien actress may convincingly play a human role, but her 'mind' still functions in profoundly non-human ways. Its alienness becomes particularly apparent when the GenAI gets 'jailbroken'—that is, a user finds a way to unlock hidden or prohibited capabilities—or ventures beyond familiar patterns and begins to improvise. These instances may trigger bizarre, incomprehensible, and perhaps dangerous behavior (Bengio et al. 2024).

The 'alien actress' metaphor encapsulates two interconnected issues: alignment and out-of-distribution behavior. The alignment problem stems from the opacity of the current GenAI systems. While experts understand these models at an abstract level, their specific operations remain largely inscrutable. This opacity creates significant challenges in interpreting the AI's internal states and explaining its

decisions (Marcinkevičs and Vogt 2020). Despite ongoing progress, these issues persist, contributing to the broader challenge of ensuring AI systems behave in accordance with human intentions and values (Christian 2020).

In democracy research, the alignment problem may prevent researchers from exercising sufficient control over their experiments. Since humans have 'grown' rather than built the current GenAI systems, they may only influence their behavior but not precisely control it. Not knowing exactly what GenAIs are 'thinking' and why they make their 'choices,' we can mostly only check their output in response to our inputs. Partially, the systems' autonomy is welcome—a researcher needs to study their digital homunculi's spontaneous behavior to reap the epistemic benefits GenAI offers over traditional methods like agent-based modeling. However, they also need proper control over the experimental setup and elicit 'genuine' behavior. In this context, GenAIs' propensity to sycophancy could be a significant problem, for instance (Sharma et al. 2023). A misaligned, unduly sycophantic GenAI system may produce responses that appear plausible but do not accurately portray human behavior to avoid displeasing the user. The resulting synthetic data may thus be corrupted, providing false support to human wishful thinking and, ultimately, perhaps even leading to flawed real-world policy decisions.

Compounding this is the problem of out-of-distribution behavior. GenAIs generate outputs by predicting patterns based on their training data. While powerful, this approach falters when dealing with scenarios underrepresented or absent in the training data. For social scientists using digital homunculi to explore novel institutional mechanisms, this limitation could be particularly problematic. The models may produce misleading outputs when pushed beyond their training distribution, skewing researchers' expectations of real human behavior in such situations. Worryingly, detecting these failures can be challenging, as GenAIs are known to 'speak' convincingly even when confabulating baselessly (Varshney et al. 2023).

For instance, researchers might use digital homunculi to simulate voter behavior under a newly proposed voting system—say, one that tests the performance of quadratic voting (Posner and Weyl 2018). While GenAI models possess a broad understanding of human behavior that could allow for

some extrapolation to novel voting systems, the accuracy and reliability of such simulations remain uncertain. Human behavior is notoriously sensitive to context, including subtle social cues and framing effects. Moreover, GenAI systems currently have limited capacity to 'laterally' interconnect information from various relevant domains where a reliable template is missing (cf. Mitchell 2021). GenAI may thus struggle to integrate its knowledge of human psychology, game theory, political behavior, and the specifics of the new voting system, having all the necessary data points but failing to recognize the patterns critical for predicting human behavior in this novel context. Consequently, GenAI might produce plausible-seeming yet misleading outputs.

In essence, the 'alien actress' problem presents the risk of subtle, hard-to-detect corruptions in synthetic data due to limited control over the process of their generation on account of misalignment and out-of-distribution behavior.

Reproducibility and Replicability

While GenAI opens promising avenues for enhancing replication in social sciences, it also introduces its own set of reproducibility and replicability challenges.

GenAI's developmental dynamics make it difficult to guarantee identical results across experiments. The cutting-edge models currently cannot be run outside of large data centers and get frequently updated. For social scientists, this creates a risk of what Bail (2024) calls a *drift*: their digital homunculi may have their 'minds' altered between or even throughout their experiments (cf. Mei et al. 2024). Drift can prevent researchers from maintaining consistency within their own studies and hinder others' replication efforts.

Complicating matters further is the widespread use of the fine-tuning technique of Reinforcement Learning from Human Feedback (RLHF) in model development. While RLHF is an essential tool to align AI behavior with human values and mitigate harmful outputs (Bai et al. 2022), it can inadvertently introduce biases that impact research validity. In particular, this process can result in AI systems behaving in unrealistically virtuous or overly socially desirable ways (Poulsen and DeDeo

2023, 483). As Grossman et al. (2023, 1109) put it: “LLM engineers have been fine-tuning pre-trained models for the world that ‘should be’ rather than the world that is.”

For democracy researchers, the models’ sanitization, however understandable on ethical and safety grounds, presents an unwelcome tweak, compounding the ‘alien actress issue.’ Not only is the ‘actress’ inscrutable and possibly misaligned, she is also actively being trained to compromise the accuracy of her answers for the sake of other values (cf. Špecián 2024). Sanitized GenAIs then fail to capture the full spectrum of human behavior in political contexts, especially its less savory aspects such as self-interest, prejudice, or susceptibility to misinformation. As a result, digital homunculi may avoid discriminatory behavior, even when such discrimination would be prevalent in real-world scenarios or possess unrealistic degrees of altruism and cooperativeness (Mei et al. 2024). The resulting synthetic data would then paint a false image of the institutional mechanisms’ efficacy.

Bias and Misrepresentation

The application of GenAI in democracy research raises significant ethical concerns, primarily revolving around algorithmic bias and fair representation. These issues are not merely technical hurdles but fundamental challenges that could potentially skew our understanding of democratic processes and institutions.

A critical challenge in applying GenAI to democracy research is the pervasive issue of algorithmic bias (Kordzadeh and Ghasemaghaei 2022). GenAI systems, by their nature, learn and replicate patterns present in their training data. When these patterns embody societal biases, such as under- or overrepresentation of certain demographics or viewpoints, GenAIs may perpetuate and reinforce these biases (Jungherr 2023). This propensity is exacerbated by the stochastic nature of GenAI: it tends to output stereotypical responses as these are often the most probable according to the training data (Bender et al. 2021).

In the context of democracy research, this bias can manifest in multiple, interconnected ways. First, AI models may misrepresent entire populations, failing to capture the true heterogeneity of human

behavior. Second, they may portray human preferences, propensities, and social structures as fixed when they are, in fact, malleable and subject to change through institutional reform. Third, and perhaps most insidiously, they may reinforce existing power structures and inequalities by treating them as inevitable or "natural" outcomes of democratic processes.

The implications of such biases in GenAI models are far-reaching for democracy research. For instance, Buolamwini and Gebru (2018) demonstrated significant gender and racial bias in commercial AI systems, raising concerns about the potential for similar biases in political simulations. The bias problem extends beyond mere representation issues to more subtle forms of cultural and ideological bias (Harding et al. 2023). GenAI models trained predominantly on data from certain cultural or political contexts may struggle to accurately represent diverse democratic traditions or alternative forms of political organization. This could reinforce a certain 'digital ethnocentrism' in democracy research, leading to Western liberal democratic norms being inadvertently treated as more universal than they truly are.

While it is likely that the overt biases will ultimately be expunged from GenAI output, their more subtle forms could still lead to skewed results in simulations of electoral processes, misrepresentation of minority viewpoints in modeled deliberations, or inaccurate predictions of policy impacts on diverse populations. The challenge of identifying and mitigating these biases is also exacerbated by the opaque nature of GenAI systems' internal workings.

*

In light of these problems, it is clear that while GenAI offers unprecedented opportunities for democracy research, its current value and long-term potential remain uncertain. The spectrum of issues—ranging from the models' opacity to concerns of reproducibility and bias—may limit the potential epistemic benefits. The realization of GenAI's promise for institutional innovation will require technical upgrades, as well as the development of novel methodologies and adaptations of the institutional and cultural norms in the research community.

5. Annoyances or Dealbreakers?

The challenges associated with the use of GenAI-based synthetic data in democracy research, as delineated in *Section 4*, are formidable. However, it is crucial to recognize that we are still in the nascent stages of this technological and methodological revolution. As Mollick (2024, 60) points out, today's GenAI is "the worst AI you will ever use." The same holds for methods of utilizing synthetic data in the social sciences. Given the extensive potential benefits of GenAI in advancing our understanding of democratic institutions and processes, we must seek and pursue pathways to mitigate its downsides and benefit from its strengths. The value of synthetic data should be assessed from a comparative, rather than absolutist, perspective. Perfection is not the benchmark; instead, we must evaluate the epistemic benefits per unit of resources invested when comparing traditional and GenAI-based methods. Are synthetic data instrumental in achieving our epistemic goals? Which methods can minimize their inherent risks and maximize the benefits?

Note that the traditional methodologies, beyond having plenty of their own limitations (*Section 2*), also tend to bear a limited promise of further improvement. Being mature, they have often already plateaued in terms of their epistemic performance. In contrast, synthetic data derived from observation of digital homunculi represent a novel research avenue with substantial room for improvement and refinement. What may initially appear as insurmountable obstacles could, with an investment of effort and ingenuity, prove to be merely temporary hurdles. Let us consider some of the promising technical, methodological, and institutional adaptations.

Technical Adaptations

From the technological viewpoint, the main trend is the steady increase in GenAI capabilities (Perrault and Clark 2024). Concurrently, the cost of use is falling, and the datasets are broadening and deepening as more and more data gets recorded and accessed for GenAI training. Also, scores of researchers from well-funded private and public labs are hard at work to mitigate well-known problems, such as misalignment, confabulation ('hallucinations'), and bias. Continued progress is not

guaranteed, but it is highly probable. Moreover, even if all fundamental capabilities were to freeze forever at their current level, much work still remains to be done to put them to the service of democracy research.

Enhancing Human Simulation Capabilities

If the scaling laws in AI development persist, we can anticipate substantial improvements in human simulation capabilities. Historically, increases in model size have corresponded with enhanced performance across various benchmarks (Kaplan et al. 2020). While this relationship remains a subject of debate within the AI community, evidence suggests that larger models develop more sophisticated internal representations of the world, including human behavior and mental states (Kosinski 2024).

The basic training objective of GenAI models—to predict the next token in a sequence to the satisfaction of human users and trainers—motivates the development of increasingly accurate models of human cognition and behavior. Consequently, we can expect more capable models to possess a more nuanced and precise representation of human beings. This would also enable an increasingly realistic performance of digital homunculi in social scientific simulations.

Mitigating Out-of-Distribution Behavior

While out-of-distribution behavior remains an inherent challenge in AI systems, the expansion of training data sets offers a potential solution by broadening the scope of ‘in-distribution’ scenarios. As AI researchers grapple with concerns about ‘hitting a data wall’—given that a significant portion of the publicly accessible internet has already been exploited for training—there is a growing impetus to uncover and utilize additional data sources, including proprietary datasets. Simultaneously, the continuous growth of the internet, together with the digitalization of historical information and increased online participation from previously underrepresented global populations, promises to provide increasingly comprehensive data on human behavior.

This expansion of the data landscape may help to fill many of the current gaps in AI models’ ‘understanding’ of human behavior, particularly in contexts where underrepresentation in training data has been a significant issue. As training data becomes deeper and more diverse—a process possibly

also coupled with GenAI's growing capability to mine the data for useful patterns and synthesize patterns from a broader range of observations—the risk of digital homunculi exhibiting an overly 'Westernized' perspective diminishes, and the ability to capture a broader spectrum of human behavior improves.

Advances in Explainable AI (XAI) Techniques

Progress in AI explainability and interpretability offers promising avenues for addressing the opacity of GenAI models. While a definitive solution remains elusive, emerging techniques suggest that the internal workings of these models may become increasingly transparent over time (Bereska and Gavves 2024). Advancements in XAI could significantly benefit democracy researchers by enabling greater control over experimental conditions and facilitating the detection of confabulations and errors. As the understanding of model internals improves, researchers may also be better equipped to identify and mitigate biases, ensuring more reliable and valid simulation results.

Scaffolding for Digital Homunculi

Significant resources are being invested in the development of GenAI agents, driven by their vast commercial potential. While the heretofore efforts to employ digital homunculi in research settings have been largely exploratory, improvised, and dependent on substantial technical expertise, we can anticipate rapid standardization and dissipation of the barriers to entry. The progress in GenAI agents is likely to result in user-friendly and customizable digital homunculi that social scientists can easily access and utilize for their social simulations. These advanced off-the-shelf agents could be adjusted to specific research needs through simple interfaces, taking advantage of GenAI's ability to be 'programmed' simply via natural language (Mollick 2024).

Standardization would help mitigate the reproducibility and replicability challenges. In particular, researchers could specify and document the exact version and configuration of the digital homunculi used in their studies, enabling others to recreate the experimental conditions and avoid the problem of model drift. Moreover, standardization could facilitate the development of benchmark tests for digital homunculi, similar to those used in traditional psychology and social science research (Mills, Costa,

and Sunstein 2023). These benchmarks could assess the agents' ability to replicate known human behaviors, biases, and decision-making patterns across various cultural contexts. By establishing a common set of performance metrics, researchers could more easily compare results across studies and evaluate the ecological validity of their digital homunculi-based experiments.

Custom Fine-tuning for Democracy Research

The challenges posed by potential distortions in GenAI's representation of human character, particularly those introduced by RLHF, can be significantly mitigated through the use of models specifically fine-tuned for social scientific purposes. Recent research has demonstrated that off-the-shelf open-source models can be efficiently customized to undo the effects of safety-oriented training, allowing researchers to simulate a fuller spectrum of human behavior (Lermen, Rogers-Smith, and Ladish 2023). This approach offers important benefits for democracy research. By fine-tuning models, researchers can bypass the overly restrictive safety measures designed for general users, enabling the simulation of morally ambiguous or socially undesirable behaviors crucial for realistic social modeling. Custom fine-tuning also enables 'correcting' the unrealistically virtuous behaviors induced by RLHF. It thus provides significant opportunities to create digital homunculi that more accurately reflect the full range of human behavior, including its unusual or less savory aspects likely to be suppressed in major public-oriented models.

Moreover, the use of customized models can be combined with local implementation, offering researchers greater control over their experimental environment. This approach not only helps avoid the drift associated with constantly updated proprietary models but also ensures better data privacy and allows researchers to maintain consistent experimental conditions throughout a study, enhancing reproducibility.[1]

[1] However, this strategy comes with trade-offs. Open-source models often lag behind their proprietary counterparts in capabilities, and the hardware requirements for local implementation can be substantial. These limitations would limit the GenAIs' equalizing influence in research capabilities within the field of democracy studies, excluding researchers from less well-funded institutions.

Methodological Adaptations

While technical advancements in GenAI promise significant benefits for democracy research, they are merely a complement of necessary adaptations on the side of the research processes and norms. Let me explore several immediately applicable strategies for integrating GenAI into democracy research methodologies that already appear feasible.

Developing Hybrid Approaches

A straightforward and immediately implementable strategy to mitigate the risks associated with GenAI, particularly the 'alien actress' problem and potential biases, is to incorporate a digital homunculi research track alongside planned or ongoing empirical studies with human participants. This approach enables comparative analysis between *in silico* and real-world results, serving as a validation mechanism for GenAI-generated data.

While the immediate scientific value of digital homunculi studies is limited due to uncertainties in design and interpretation, this hybrid approach provides an opportunity to calibrate and refine our digital homunculi. As demonstrated by Park et al. (2023) and others, even early implementations of GenAI-based social simulations can be insightful. Moreover, hands-on experience is crucial for the development of more sophisticated hybrid methodologies that can effectively address the challenges posed by GenAI in democracy research.

A particularly useful research arrangement would be to leverage the swiftness and low costs of synthetic data for initial testing, calibration, and iteration while using human participants as a control group. This initial phase could substantially narrow down the range of options, reserving more resource-intensive empirical studies with human participants for later stages of pre-implementation testing. This approach eases the experimentation bottleneck while still allowing traditional methods to perform the epistemic heavy lifting of final validation.

Implementation of Adversarial Testing

GenAI also offers an opportunity to implement robustness testing through adversarial proceedings that—while epistemically valuable (Pamuk 2021; 2022)—are otherwise often too time- and resource-intensive to pursue using traditional methods. One simple strategy is prompting GenAI—especially the large language models (LLMs)—to generate and evaluate competing perspectives and designs before or after the simulation. For instance, researchers could implement a debate format between multiple instances of an LLM, each arguing for different interpretations of data or methodological approaches. This setup, inspired by Irving et al. (2018), could help reveal potential weaknesses in research designs, highlight overlooked factors in the analysis of democratic processes, and be specifically directed to focus on how GenAI's known problems can be mitigated by the research design.

Moreover, adversarial testing can be used to stress-test proposed democratic institutions or reforms. Researchers could program digital homunculi to actively seek flaws or exploit potential loopholes in proposed systems, mimicking the behavior of strategic actors in real-world democratic contexts. This approach could uncover potential vulnerabilities that human-based empirical research may not show due to the necessary simplifications, limited duration, and hard-to-control incentives of traditional experimental research.

Embracing Predictions

How can we determine when synthetic data provide reliable answers? Furthermore, when do they offer more or less reliable results than traditional experimental research? Notably, a mere deviation between outcomes achieved with digital homunculi and those from empirical studies with human participants does not mean the fault lies with the synthetic data, given the limited reproducibility of experimental results in social sciences. In fact, there is no fundamental reason why experiments with digital homunculi could not offer greater epistemic performance than studies with human subjects, even ignoring their cost and speed advantages. But how can we assess the (comparative) epistemic performance in various domains?

I propose that the solution lies in embracing prediction as the gold standard for evaluating model success (cf. Friedman 1953). While social scientists have often hesitated to adopt predictive criteria, perhaps due to notable failures in the past (Tetlock 2006), the integration of GenAI into our methodological toolkit presents an opportunity to recalibrate our stance on prediction in social inquiry. Prediction testing offers a robust, quantifiable metric for assessing the validity of both traditional and GenAI-assisted research methods (cf. Schoenegger et al. 2024). By focusing on a model's ability to anticipate outcomes—that is, to generalize to unobserved contexts—we can compare the performance of digital homunculi simulations against empirical studies with human participants in terms of what matters most: their capacity to establish reliable expectations about the real-world performance of various institutional mechanisms.

A model that consistently produces accurate predictions, regardless of whether it is derived from studies with human subjects or digital homunculi, is inherently more likely to provide actionable insights for democratic reform and innovation. This renewed focus on prediction could serve as a bridge between traditional social science methodologies and the emerging possibilities offered by GenAI. It provides a level playing field for evaluating different approaches and enforces a focus on epistemic performance above other cherished—but potentially outdated (cf. Smith 2023)—criteria such as elegance and simplicity. Importantly, this approach acknowledges that social reality may be too complex to correspond neatly with any human-comprehensible model, perhaps due to the number of interacting parameters exceeding the limits of human cognition. Adopting prediction as our primary epistemic criterion compels us to confront the possibility that the most effective tools for democratic innovation may not be intuitively appealing theories but rather inscrutably large machine learning models whose only—yet decisive—virtue lies in their ability to forecast outcomes in complex social systems accurately.

Institutional and Cultural Adaptations

The integration of GenAI into democracy research demands not only technical and methodological adaptations but also shifts in our institutional landscape and academic culture. The current research

environment in democracy studies reflects the constraints imposed by costly, time-consuming, and methodologically delicate traditional empirical studies. This has fostered a certain ‘scarcity mindset’ that shapes academic institutions and practices. However, the potential of GenAI to dramatically reduce costs and accelerate research cycles challenges this status quo.

The current institutional landscape of democracy research reflects the constraints imposed by the costly, time-consuming, and methodologically delicate nature of traditional empirical studies. This environment has fostered a certain ‘scarcity mindset’ that shapes academic institutions and culture. Given the high value and resource-intensive nature of empirical studies, both in terms of labor and public funding, there is a tendency towards conservatism in research approaches. Researchers often hesitate to venture into unknown territories, prioritize minimizing the risk of antagonizing reviewers, and invest considerable efforts in rigorous verification processes. Many aspects of this scarcity-shaped institutional order are valuable: maintaining rigor remains crucial in studies using synthetic data. However, some would deserve updates, or at least increased openness to alternatives, to enable the exploitation of GenAI’s potential in democracy research.

Promoting Open Dialogue on GenAI’s Use in Research

One critical area for institutional and cultural adaptation is the promotion of open dialogue regarding GenAI's research uses. This openness is crucial not only for facilitating the free exchange of ideas about synthetic data and innovative research methodologies but also for collectively addressing the risks associated with GenAI. However, some existing standards and expectations in academia, adapted to the traditional approaches and methodologies, may now become hindrances to the dissemination of innovative practices and the development of risk mitigation strategies.

A significant portion of the academic enterprise is structured around the acknowledgment of individual merit, sometimes at the expense of prioritizing epistemic progress. Partly, the intense scrutiny dedicated to standard academic publications is motivated by the considerations of whether the researcher ‘deserves praise’ for a result rather than any epistemic concerns. This scrutiny is certainly

justified when it comes to the issues of plagiarism or the misappropriation of others' work. However, the use of machine labor, particularly GenAI, introduces new complexities to these considerations.

While GenAIs cannot be treated as authors in the traditional sense, especially since they cannot bear responsibility for their statements, their 'authorship' is also difficult to dismiss completely due to their growing capabilities to contribute to research tasks (Teixeira Da Silva and Tsigaris 2023). For instance, how about situations where a GenAI contributes a substantive 'idea' that significantly enhances the research—an idea that, if proposed by a human collaborator, may warrant a co-authorship claim? Current norms regarding the use of GenAI in research remain ambiguous and often prone to erring on the side of caution, discouraging whatever might be deemed 'too extensive' use of these tools.

This caution, while understandable, is unfortunately bound to impede progress in the field. Many competent researchers will avoid experimenting with GenAI technologies out of concern for maintaining their professional reputation, preferring to wait until norms are updated and 'safe' practices are clearly delineated. Paradoxically, however, the establishment of new norms requires precisely this kind of experimentation, including the inevitable errors and excesses that accompany any norm renegotiation.

Therefore, it seems a greater lenience and openness in evaluating the use of GenAI in research, with a renewed focus on the the advancement of knowledge as the primary—if not sole—objective of research. This openness could foster collaborative efforts to address the challenges posed by GenAI, such as developing protocols for detecting and mitigating biases in synthetic data or establishing best practices for maintaining the reproducibility of GenAI-assisted research.

To summarize, if GenAI-facilitated research is found to be epistemically unsound or misleading, it should rightfully be rejected. However, if such research is epistemically robust but does not strictly adhere to traditional norms regarding, for instance, the concept of authorship or originality, a degree of flexibility appears advisable. This flexibility could allow for innovative approaches to emerge that might better handle the specific challenges posed by GenAI in democracy research. In essence, to

create a more conducive environment for harnessing GenAI's opportunities and mitigating its risks, the academic community should consider recalibrating its focus towards the epistemic mission of research, de-emphasizing certain aspects of the status-oriented elements of academic culture.

Greater Through-Put Quality Control

The second area of institutional and cultural adaptation concerns the mechanisms for quality control of research outputs. Given that the traditional methods of empirical research produce a relatively low number of results, the slow, labor-intensive process of academic publication, including peer review, has been relatively functional. The months-long wait for reviews following years of data collection and analysis has become a mundane reality. Still, even now, many researchers already circumvent traditional quality control processes by releasing preprints, which typically undergo no independent pre-publication scrutiny. Increasingly, formal journal publication may end up driven by concern for career concerns rather than for epistemic progress.

The introduction of synthetic data is likely to upset the status quo further. Researchers may soon be able to produce results at a much faster pace, flooding the field with novel results of uncertain epistemic merit. This, however, poses a substantial challenge to the current quality control mechanisms, which may easily be overwhelmed.

There is thus an urgent need for new, high-throughput quality control mechanisms (Kousha and Thelwall 2024). Moreover, these mechanisms should be optimized to detect the specific types of issues that may arise in studies based on synthetic data. One potential, albeit largely untested, solution is the implementation of a partially automated review process (Kousha and Thelwall 2024). In this scenario, specialized GenAIs assist, or in some cases substitute for, human reviewers. These AI reviewers could be trained to identify potential issues specific to GenAI-based research. The above-mentioned adversarial proceedings could also prove valuable for independent quality control. For example, one could pit different GenAI reviewers against each other in a debate format to scrutinize research findings, potentially uncovering flaws or biases that might be missed by human reviewers alone.

The development and implementation of these solutions require extensive experimentation and especially cultural and institutional openness. However, in light of the emerging evidence on malfunctions and pathologies of the traditional peer review (Mastroianni 2022), the development of new quality control mechanisms should perhaps not be viewed as a forced compromise but as an opportunity to address longstanding issues in the academic review process. Be that as it may, some form of (partially) automated quality control, even if imperfect, could well be preferable to the absence of pre-publication scrutiny seen in the preprint model.

Fostering Interdisciplinarity (for real this time)

The introduction of synthetic data in democracy research not only opens new avenues for interdisciplinary collaboration but also necessitates it as a strategy to mitigate GenAI-associated risks. Traditionally, the high costs associated with empirical research have often constrained studies to narrow aspects of human behavior, typically aligned with the interests and methodologies of individual disciplines. However, social simulations enabled by GenAI allow for multidimensional data gathering on an unprecedented scale while simultaneously introducing complex challenges that span multiple fields of expertise.

Digital homunculi, mimicking human agents, do not conform to disciplinary boundaries; their behavior is not solely economic, sociological, or psychological but a complex interplay of all these dimensions and more. This holistic nature of digital communities in simulations produces data that spans multiple dimensions simultaneously, potentially offering a more comprehensive view of democratic processes. However, it also introduces risks that require diverse expertise to be addressed effectively.

Consequently, the most effective utilization of synthetic datasets will likely require teams of researchers capable of cooperating beyond the confines of any single research niche. For instance, mapping and handling potential biases in digital homunculi is likely to require collaboration socials scientists, machine learning experts, and AI ethicists.

Moreover, the use of GenAI itself can facilitate interdisciplinary collaboration by lowering barriers to accessing expertise outside one's primary field (Špecián 2024). Large language models, for instance, can assist in translating discipline-specific jargon and concepts, bridging communication gaps that have historically hindered interdisciplinary efforts.

Democracy research, in particular, with its convoluted causal chains and high-stakes implications, stands to benefit from such an integrated approach. The challenges posed by GenAI in this field demand a convergence of insights from a broad range of disciplines. Increased openness towards interdisciplinarity could do much for a swifter assessment—and, hopefully, unlocking—of the epistemic potential of synthetic data.

*

Above, I have outlined a range of interventions, mitigations, and adaptations that may address many of the key risks associated with the use of synthetic data in social sciences, particularly in democracy research. Technical adaptations targeting known limitations have shown steady progress, with a high likelihood of continued advancement. At the same time, however, to be able to employ these "technofixes" productively, there is also an urgent need for methodological, institutional, and cultural adaptations that need to stem from within the democracy research community. Above, I have presented several initial suggestions on the direction this 'soul-searching' might take.

Most importantly, as elsewhere, experimentation remains the key (cf. Špecián and Císař Brown 2024). Even if the full promise of synthetic data is not realized due to technological stagnation or insurmountable methodological issues, this pursuit remains valuable. The process of adaptation may provide solutions to long-standing issues in our research and academic systems, such as the limitations of peer review, excessive credentialism, and persistent disciplinary silos. Given the difficulty of epistemic and pragmatic challenges in democratic innovations, even partial success would be invaluable.

6. Conclusion

This paper has explored GenAI‘s potential to transform the landscape of democracy research, offering an assessment of both its promise and its challenges. The integration of GenAI-based methodologies, particularly through the use of synthetic data stemming from the observation of digital homunculi, presents a possible solution to the longstanding experimentation bottleneck in democracy studies. Opening new possibilities for rapid and cost-effective testing of democratic reforms, GenAI shows the potential to significantly narrow the gap between democratic theory and practical institutional innovation.

Ultimately, GenAI could become an indispensable research tool, enabling large-scale societal simulations, diverse participant pools, and complex scenario modeling that were previously much less feasible. However, these epistemic fruits cannot be harnessed without effort and risk. Challenges such as algorithmic bias, reproducibility issues, and alignment problems will require innovative solutions. To provide a peek at the breadth of possibilities, I have outlined a range of technical, methodological, institutional, and cultural adaptations that could address these challenges, emphasizing that many of the present obstacles to GenAI use in democracy research likely present surmountable annoyances rather than definite dealbreakers.

The broader implications of integrating GenAI into democracy research are far-reaching. It could transform research methodologies, making them more dynamic and innovative. In democracy research and, more broadly, the social sciences, this may ultimately trigger a gradual shift from a scarcity mindset to one of abundance and experimentation.

Finally, the accelerated pace of research enabled by GenAI could unblock the exploration of democracy’s vast design space. The persistent ‘pacing problem’ (Marchant 2011), where the speed of technological and social change often outpaces our ability to study and understand its implications for democratic governance, could be mitigated. In such a ‘brave new world’ of democracy research, the insights gained from GenAI-assisted research could enable otherwise infeasible democratic reforms

(cf. Landemore 2021), enhancing the resilience and adaptability of democratic systems in the face of the challenges introduced by the digital age.

References

Argyle, Lisa P., Ethan C. Busby, Nancy Fulda, Joshua R. Gubler, Christopher Rytting, and David Wingate. 2023. “Out of One, Many: Using Language Models to Simulate Human Samples.” *Political Analysis* 31 (3): 337–51. https://doi.org/10.1017/pan.2023.2.

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.

Bail, Christopher A. 2024. “Can Generative AI Improve Social Science?” *Proceedings of the National Academy of Sciences* 121 (21): e2314021121. https://doi.org/10.1073/pnas.2314021121.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜.” In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. FAccT ’21. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922.

Bengio, Yoshua, Geoffrey Hinton, Andrew Yao, Dawn Song, Pieter Abbeel, Trevor Darrell, Yuval Noah Harari, et al. 2024. “Managing Extreme AI Risks amid Rapid Progress.” *Science* 384 (6698): 842–45. https://doi.org/10.1126/science.adn0117.

Bereska, Leonard, and Efstratios Gavves. 2024. “Mechanistic Interpretability for AI Safety -- A Review.” arXiv. https://doi.org/10.48550/arXiv.2404.14082.

Buolamwini, Joy, and Timnit Gebru. 2018. “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.” In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91. PMLR. https://proceedings.mlr.press/v81/buolamwini18a.html.

Chan, Xin, Xiaoyang Wang, Dian Yu, Haitao Mi, and Dong Yu. 2024. “Scaling Synthetic Data Creation with 1,000,000,000 Personas.” arXiv. https://doi.org/10.48550/arXiv.2406.20094.

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

Coeckelbergh, Mark. 2024. *Why AI Undermines Democracy and What To Do About It*. John Wiley & Sons.

Dillion, Danica, Niket Tandon, Yuling Gu, and Kurt Gray. 2023. “Can AI Language Models Replace Human Participants?” *Trends in Cognitive Sciences* 27 (7): 597–600. https://doi.org/10.1016/j.tics.2023.04.008.

Fridman, Lex, dir. 2023. *Eliezer Yudkowsky: Dangers of AI and the End of Human Civilization | Lex Fridman Podcast #368*. https://www.youtube.com/watch?v=AaTRHFaaPG8.

Friedman, Milton. 1953. “Methodology of Positive Economics.” In *Essays in Positive Economics*. University of Chicago Press.

Grossmann, Igor, Matthew Feinberg, Dawn C. Parker, Nicholas A. Christakis, Philip E. Tetlock, and William A. Cunningham. 2023. “AI and the Transformation of Social Science Research.” *Science* 380 (6650): 1108–9. https://doi.org/10.1126/science.adi1778.

Harding, Jacqueline, William D'Alessandro, N. G. Laskowski, and Robert Long. 2023. "AI Language Models Cannot Replace Human Research Participants." *AI & SOCIETY*, July, s00146-023-01725–x. https://doi.org/10.1007/s00146-023-01725-x.

Henrich, Joseph Patrick. 2020. *The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous*. New York: Farrar, Straus and Giroux.

Hildt, Elisabeth. 2023. "The Prospects of Artificial Consciousness: Ethical Dimensions and Concerns." *AJOB Neuroscience* 14 (2): 58–71. https://doi.org/10.1080/21507740.2022.2148773.

Ioannidis, John P. A. 2005. "Why Most Published Research Findings Are False." *PLOS Medicine* 2 (8): e124. https://doi.org/10.1371/journal.pmed.0020124.

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

Jones, Cameron R., and Benjamin K. Bergen. 2024. "People Cannot Distinguish GPT-4 from a Human in a Turing Test." arXiv. https://doi.org/10.48550/arXiv.2405.08007.

Jones, Nicola. 2024. "AI Now Beats Humans at Basic Tasks — New Benchmarks Are Needed, Says Major Report." *Nature* 628 (8009): 700–701. https://doi.org/10.1038/d41586-024-01087-4.

Jungherr, Andreas. 2023. "Artificial Intelligence and Democracy: A Conceptual Framework." *Social Media + Society* 9 (3): 20563051231186353. https://doi.org/10.1177/20563051231186353.

Kaiya, Zhao, Michelangelo Naim, Jovana Kondic, Manuel Cortes, Jiaxin Ge, Shuying Luo, Guangyu Robert Yang, and Andrew Ahn. 2023. "Lyfe Agents: Generative Agents for Low-Cost Real-Time Social Interactions." arXiv. https://doi.org/10.48550/arXiv.2310.02172.

Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. 2020. "Scaling Laws for Neural Language Models." arXiv. https://doi.org/10.48550/arXiv.2001.08361.

Kim, Junsol, and Byungkyu Lee. 2024. "AI-Augmented Surveys: Leveraging Large Language Models and Surveys for Opinion Prediction." arXiv. https://doi.org/10.48550/arXiv.2305.09620.

Kordzadeh, Nima, and Maryam Ghasemaghaei. 2022. "Algorithmic Bias: Review, Synthesis, and Future Research Directions." *European Journal of Information Systems* 31 (3): 388–409. https://doi.org/10.1080/0960085X.2021.1927212.

Kosinski, Michal. 2024. "Evaluating Large Language Models in Theory of Mind Tasks." arXiv. https://doi.org/10.48550/arXiv.2302.02083.

Kousha, Kayvan, and Mike Thelwall. 2024. "Artificial Intelligence to Support Publishing and Peer Review: A Summary and Review." *Learned Publishing* 37 (1): 4–12. https://doi.org/10.1002/leap.1570.

Kurtulmus, Faik. 2020. "The Epistemic Basic Structure." *Journal of Applied Philosophy* 37 (5): 818–35. https://doi.org/10.1111/japp.12451.

Landemore, Hélène. 2020. *Open Democracy: Reinventing Popular Rule for the Twenty-First Century*. Princeton: Princeton University Press.

———. 2021. "Open Democracy and Digital Technologies." In *Digital Technology and Democratic Theory*, edited by Lucy Bernholz, Hélène Landemore, and Rob Reich, 62–89. University of Chicago Press. https://www.degruyter.com/document/doi/10.7208/9780226748603-003/html.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." *Nature* 521 (7553): 436–44. https://doi.org/10.1038/nature14539.

Lermen, Simon, Charlie Rogers-Smith, and Jeffrey Ladish. 2023. “LoRA Fine-Tuning Efficiently Undoes Safety Training in Llama 2-Chat 70B.” https://doi.org/10.48550/ARXIV.2310.20624.

Marchant, Gary E. 2011. “The Growing Gap Between Emerging Technologies and the Law.” In *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem*, edited by Gary E. Marchant, Braden R. Allenby, and Joseph R. Herkert, 19–33. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-1356-7_2.

Marcinkevičs, Ričards, and Julia E. Vogt. 2020. “Interpretability and Explainability: A Machine Learning Zoo Mini-Tour.” https://doi.org/10.48550/ARXIV.2012.01805.

Mastroianni, Adam. 2022. “The Rise and Fall of Peer Review: Why the Greatest Scientific Experiment in History Failed, and Why That’s a Great Thing.” *Experimental History*, December. https://www.experimental-history.com/p/the-rise-and-fall-of-peer-review.

Mei, Qiaozhu, Yutong Xie, Walter Yuan, and Matthew O. Jackson. 2024. “A Turing Test of Whether AI Chatbots Are Behaviorally Similar to Humans.” *Proceedings of the National Academy of Sciences* 121 (9): e2313925121. https://doi.org/10.1073/pnas.2313925121.

Mills, S., S. Costa, and C. R. Sunstein. 2023. “AI, Behavioural Science, and Consumer Welfare.” *Journal of Consumer Policy* 46 (3): 387–400. https://doi.org/10.1007/s10603-023-09547-6.

Mitchell, Melanie. 2021. “Why AI Is Harder Than We Think.” arXiv. https://doi.org/10.48550/arXiv.2104.12871.

Mittiga, Ross. 2021. “Political Legitimacy, Authoritarianism, and Climate Change.” *American Political Science Review*, December, 1–14. https://doi.org/10.1017/S0003055421001301.

Mollick, Ethan. 2024. *Co-Intelligence: Living and Working with AI*. London.

Nikolenko, Sergey I. 2021. *Synthetic Data for Deep Learning*. Vol. 174. Springer Optimization and Its Applications. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-75178-4.

OECD. 2023. “OECD Deliberative Democracy Database (2023).” Airtable. 2023. https://airtable.com/appP4czQlAU1My2M3/shrX048tmQLl8yzdc/tblrttW98WGpdnX3Y/viwX5ZutDDGdDMEep.

OpenAI. 2024. “GPT-4o Mini: Advancing Cost-Efficient Intelligence.” June 18, 2024. https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/.

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.

Park, Joon Sung, Joseph O’Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. “Generative Agents: Interactive Simulacra of Human Behavior.” In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 1–22. San Francisco CA USA: ACM. https://doi.org/10.1145/3586183.3606763.

Perrault, Ray, and Jack Clark. 2024. “Artificial Intelligence Index Report 2024,” April. https://doi.org/20.500.12592/h70s46h.

Posner, Eric A., and E. Glen Weyl. 2018. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton: Princeton University Press.

Poulsen, Victor Møller, and Simon DeDeo. 2023. “Large Language Models in the Labyrinth: Possibility Spaces and Moral Constraints.” *Possibility Studies & Society* 1 (4): 471–88. https://doi.org/10.1177/27538699231206210.

Schaffer, Frederic Charles, and Jean-Paul Gagnon. 2023. “Democracies Across Cultures,” June. https://doi.org/10.3167/dt.2023.100107.

Schoenegger, Philipp, Indre Tuminauskaite, Peter S. Park, and Philip E. Tetlock. 2024. “Wisdom of the Silicon Crowd: LLM Ensemble Prediction Capabilities Rival Human Crowd Accuracy.” https://doi.org/10.48550/ARXIV.2402.19379.

Sharma, Mrinank, Meg Tong, Tomasz Korbak, David Duvenaud, Amanda Askell, Samuel R. Bowman, Newton Cheng, et al. 2023. “Towards Understanding Sycophancy in Language Models.” arXiv. https://doi.org/10.48550/arXiv.2310.13548.

Smith, Noah. 2023. “The Third Magic.” Noahpinion. 2023. https://www.noahpinion.blog/p/the-third-magic.

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.

———. 2024. “Machine Advisors: Integrating Large Language Models Into Democratic Assemblies.” *Social Epistemology*, August, 1–16. https://doi.org/10.1080/02691728.2024.2379271.

Špecián, Petr, and Lucy Císař Brown. 2024. “Give the Machine a Chance, Human Experts Ain’t That Great….” *AI & SOCIETY*, April, s00146-024-01910–16. https://doi.org/10.1007/s00146-024-01910-6.

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

Teixeira Da Silva, Jaime A., and Panagiotis Tsigaris. 2023. “Human- and AI-Based Authorship: Principles and Ethics.” *Learned Publishing* 36 (3): 453–62. https://doi.org/10.1002/leap.1547.

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

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.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. “Attention Is All You Need.” In *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.