PrEval EXPERTISEN
Themen der internationalen Evaluationsforschung 2 | 2025

Exploring Emerging AI as Subject and Object in Democracy-focused Evaluation

Author **Quito Tsui**

Quito Tsui is a researcher and writer with expertise on technology in the context of the humanitarian and development sectors. She has a keen interest in the development of a technological environment rooted in justice with practical experience of supporting organisations to this end. As a Core Collaborator at the *MERL Tech Initiative*, and *Humanitarian AI + MERL Working Group* lead she has conducted original research on deduplication, and developed guidance on safe and responsible use of AI in social impact efforts, humanitarian applications and research on violence against women. She previously led research on digital identity, biometrics, interoperability and environmental justice at *The Engine Room*. She has also conducted academic research at the *University of Cambridge*, *Stanford University* and the *London School of Economics*. Quito's writing has been featured in the *Forced Migration Review* (*Oxford University*), *Bot Populi* and *The Nation* and her work has been cited by outlets such as *Access Now* and *Geographical*.

Author **Linda Raftree**

Linda Raftree focuses on the ethical use of technology and data in international development, human rights, and humanitarian contexts. She is the founder of the *MERL Tech Initiative* (MTI) which has been exploring the responsible use of digital tools and data since 2014. Through MTI, Linda has played a strong role in bringing together critical thinkers from the monitoring, evaluation, research, and learning (MERL), data science, and digital technology sectors to collaborate, develop capacities and strengthen good practice. Linda convenes the *MERL Tech Natural Language Processing Community of Practice* (NLP-CoP) and the *New York City Technology Salon*. She has advised a range of foundations, bilateral agencies, and non-governmental organizations on responsible data policy and practice, applied research, digital safeguarding, inclusive digital approaches, and safe programming. Linda is a *Certified Information Privacy Professional* (CIPP & CIPM).

Exploring Emerging AI as Subject and Object in Democracy-focused Evaluation

Quito Tsui and Linda Raftree

SUMMARY

Current use of emerging *Artificial Intelligence* (AI) in the evaluation sector remains incipient, with conversations remaining primarily theoretical and lacking in robust evidence. Democracy evaluation – like other sector-specific evaluation practices – requires both a broad and narrow understanding of the implications and possible impact of AI with democracy-focused evaluation. Our research addresses this gap by providing an overview of the use of AI in evaluation with a specific focus on democracy initiatives. It finds that the anti-democratic impact of AI is both widely theorised and increasingly realised in the degradation of information ecosystems and the enhancement of surveillance tools – both of which run counter to healthy democracies. The potential democratic dividends meanwhile are focused on enhanced participation — an intention that is significantly undermined by the anti-democratic ramifications of AI use in online environments especially.

These observations anchor a deeper exploration of AI and *Generative Artificial Intelligence* (GenAI) in democracy evaluation. This paper provides a landscape mapping of current literature identifying how AI is being used for *Monitoring*, *Evaluation*, *Research* and *Learning* (MERL), the key benefits or shortcomings identified through case studies, and the potential uses of AI for MERL being discussed. Further the paper applies these learnings to the context of democracy programming, assessing the extent to which AI-supported MERL can meet assessment needs. This research emphasises the importance of taking a highly specific approach to AI tool selection, analysing discrete applications and their possible utility for democracy evaluations. It pays particular attention to the possible harms both broadly associated with AI in evaluation such as that of biassed data, as well as in the specific deployment of AI for the evaluation of democracy programmes, including the possibility of unintended outcomes. Though there is a wide range of potential uses, actual utility remains underexplored. This area would need to be tackled before evaluation specialists deploy AI tools more widely. To assist with initiating this discussion, the paper considers the ethics of AI in democracy programming evaluation and delineates good practice approaches drawn from AI in MERL and informed by the specific considerations of those evaluating democracy programmes.

Overall, the paper offers a dynamic assessment of AI, democracy and MERL, putting positives and negatives of such use in conversation with each other and considering how limitations in mitigation strategy may constrain potential positive impact or result in exclusionary outcomes.

Contents

1. Introduction and Background

In November 2022, consumer-focused *Artificial Intelligence* (AI) hit the mainstream in the form of *Generative AI* (GenAI) tools built on top of *Large Language Models* (LLMs). These new commercial tools include text-focused applications (*ChatGPT*, *Genesis*, *Perplexity*, *Claude*, etc.) and audio-visual generating applications like *Stable Diffusion*, *Midjourney*, *DALL-E* and *Sora*. This key moment in the evolution of digital technology is expected to usher in a new era that will have huge effects on people, communities, and societies in every sphere.

While the potential of AI is seen as promising for a variety of purposes, the relationship between democracy and AI needs consideration before AI is widely embraced in democracy-focused programming and evaluation. What effects is AI having on privacy and surveillance, democracy, and information integrity? How is AI affecting democracy-focused organisations and how will it alter the contexts in which such organisations work? How are governments adapting to AI – and which contexts and groups are gaining and losing the most from this shift? What assaults on democracy and which anti-democratic practices will be facilitated by AI – and what new democratic processes might it enable? These questions and others that will emerge in the coming years need to be unpacked thoughtfully.

The relationship between democracy and AI needs consideration before AI is widely embraced in democracy-focused programming and evaluation.

As AI begins to take a greater hold on the world, it's critical that those working on governance, policy making and democratic capacity building understand both AI's potential and its downsides. In this paper, we take a two-step approach to exploring democracy, AI and evaluation, first examining the relationship between AI and democracy and then considering its role in the evaluation of this type of programming. The full impact of these rapid techno-social changes is yet to be seen, but they do point to a clear need to gain a deeper understanding of the intersections of these critical areas.

2. Evaluating the Interplay between AI/GenAI and Democracy

The implications of AI on democracy have been interrogated from multiple directions including the angle of mis- and disinformation, AI governance, and the links between unbridled capitalism and economic capture of *Big AI* companies. Indeed, as GenAI technologies have continued to evolve, the discussion of the potential threat or contribution of AI to democracy has increased in tandem. The rapid advancement of GenAI has accelerated fears about how it may expedite the impact of AI tools on democracy and democratic processes (Allen/Weyl 2024). Common in discussion regarding the implications of AI on democracy, is an emphasis on how it can amplify existing anti-democratic patterns. However, the very same amplifying capacity of AI and GenAI is cited as a potential vehicle for making government more efficient and democratic processes more robust and responsive. Understanding the status of discussions regarding AI and democ-

racy is a necessary first step in considering how GenAI can and should be used in evaluating democracy programming.

2.1 Challenges GenAI Brings to Democracy

2.1.1 The Anti-democratic Potential of AI Tools

The *World Economic Forum* deemed AI-driven misinformation and disinformation one of the top global risks of 2024 (*Global Risks* 2024 n.d). The report noted that AI could make the production of synthetic content more accessible, making it easier to generate deepfakes and other false information without the need for specialist skills (Chan 2024). For the democratic context, concern around mis- and disinformation is especially acute (Gül et al. 2020; Leingang 2024; Schmitt et al. 2024; Verma 2023; Whyte 2020). Against the backdrop of the 2024 "super election year", many policy makers and think tanks warned about the ability of AI to "supercharge" misinformation and undermine election integrity (Böhnke 2024).

These concerns were not unreasonable. Manipulation of online media has had real-word impact: in Slovakia a deep-fake audio circulated two days before the 2023 election showed then exit poll-leader and liberal opposition candidate Michal Šimečka colluding with a journalist to buy votes and rig the election (Murphy 2024). Though the recording was falsified, it was shared widely on Social Media — Šimečka ultimately lost the election and the journalist was also a target of hateful attacks (Murphy 2024). As of yet it has remained difficult to assess the impact of the deep- fake video on the election outcome, but experts have pointed to similar instances to indicate the anti-democratic potential of AI-powered misinformation (Murphy 2024).

Meanwhile in Indonesia, an AI-generated deepfake of late Indonesian dictator Suharto (Chen 2024) was used by the *Golkar* party to appeal for votes which one study showed 14.5% of respondents believed to be authentic (Jalli/Wihardja 2024). The reach of such videos often outpaces the ability of journalists and fact checkers to debunk them (Ware 2024) while their proliferation weaponizes information by undermining trust and reducing the ability of liberal democracies to ensure an accessible and accurate information space (Jalli/Wihardja 2024).

The impact of deepfakes and other forms of synthetic media on elections and democratic processes is contested and remains unclear (Maguire 2024). Simon et al. (2023) argue that fears about AI mis- and disinformation are "overblown", with increased quantity or accessibility of misinformation not directly translating into a more gullible public. However, others, including Łabuz and Nehring, suggest that although apocalyptic depictions of AI misinformation have yet to materialise, the more imminent risk of AI mis- and disinformation is the erosion of trust and increased uncertainty within political landscapes (Łabuz/Nehring n.d.; Simon et al. 2023). It is clear that fears around AI are manifold and in themselves could contribute to weakening trust in democratic institutions and uncertainty in where to access trustworthy information sources (Sanchez/Middlemass 2022).

AI Tools Can Contribute to the Degradation of Healthy Information Ecosystems

Beyond elections, others have pointed to the more daily impact AI and GenAI can have on the wider information ecosystem of democracies. Algorithmically constructed online spaces, such as Social Media news feeds, shape the information that individuals engage with in highly specific ways. Micro-targeted political content, a recent UNESCO report argues, contributes to "individualistic and polarised attitude[s]" through the creation of "homogeneous and self-contained virtual communities sharing the same views, thus undermining social cohesion" (Innerarity 2024), with this tailored information pipeline reducing the capacity of individuals to make fully informed decisions (Zhu/Isaacs 2024; Krimmer et al. 2022: 36). AI tools are being employed to refine targeted and personalised content (Gao et al. 2023), utilising the ability of GenAI to quickly synthesise multiple data sources. In doing so AI tools can amplify the risks of individuated online experiences by creating hyper individualised information environments. Content does not need to be political for polarisation to contribute to the fracturing of a shared reality; the segregation of online spaces itself can undermine wider societal cohesion by reinforcing selective information.

Consequences of targeted content can swiftly escalate. In both Myanmar (Zaleznik 2021) and Ethiopia, hate speech and misinformation spread over *Facebook* and amplified by the platform's AI algorithms resulted in widespread violence (Blancho 2024). Misinformation shared on the platform contributed to the ongoing Rohingya Genocide in Rakhine state (Amnesty International 2022; Crystal 2023; Zaleznik 2021) as well as ethnic violence against the Tigrayan community in northern Ethiopia (Ethiopia 2023; Haile 2024; Mackintosh 2021). Though this is not an inevitable outcome of micro-targeted content, it is indicative of the manner in which algorithmically elevated content can, in certain contexts, contribute to devastating outcomes.

AI-enhanced Surveillance Can Enhance Anti-democratic Repression

Research on potential applications of AI have been linked to the expansion of surveillance oriented use of AI. The *Surveillance AI Pipeline* as documented by Kalluri et al. found that the majority of papers related to comput-er vision contribute to "downstream surveillance patents" (Kalluri et al. 2023: 3). Civil society organisations, digital rights advocates and human rights campaigners have already begun to raise concerns about the surveillance capacity and discriminatory potential of AI systems such as smart cities, public facial recognition systems, smart policing and social media surveillance (Bosoer et al. 2022).

AI surveillance has surged in autocratic countries where AI has both given rise to new forms of repression while enhancing the state's abilities to carry out traditional forms of repression (Feldstein 2021). AI tools reduce the resources required for large-scale surveillance, making it cheaper and more accessible for governments globally to monitor citizens. Meanwhile tools such as live facial recognition, web scraping, behaviour monitoring (Hao 2021) and sentiment analysis along with predictive policing can be used to create a nexus of interoperable tools that track and identify individuals in real time. As digital tools have become cheaper and more ubiquitous, it's more likely that governments will surveil without thought to human rights standards (Feldstein 2020).

Illiberal regimes can combine AI surveillance mechanisms with repression tactics to harass and intimidate individuals and stymy dissent. Bots and computer operated algorithms can automate targeting of individuals with disinformation through the generation of false videos and images, as well as harassment, trolling and censorship (*Policy Recommendations* n.d.). Automated repression can also disrupt civil society action and organising through flooding – a method that inundates Social Media with both political and apolitical content to drown out or distract from dissent. Regimes can flood hashtags on Social Media, for instance, to dampen information about protests. This is not limited to within regime borders; these tools have proved highly effective at foreign influence and interference (Cirone/Hobbs 2023). Overseas surveillance and repression of citizens abroad extends the sphere of influence of autocratic regimes beyond national borders and can have a chilling effect (Peterson/Hoffman 2022). These tools enhance anti-democratic forces in both democracies and autocracies and can erode collective democratic progress. The lack of transparency about how and when these tools are used provides covert avenues for quashing dissent and censoring Social Media.

2.2 Uses of AI to Enable Democratic Processes

2.2.1 AI Tools and Enhanced Participation in the Democratic Process

Conversation regarding the possible democratic dividends of AI has largely focused on the role of AI in enhancing participation. This emphasis on participation comes from a wider desire in democracy programming to increase involvement in the democratic process. One key programming effort has been the use of deliberative democracy to engage citizens in decision-making and policy formulation at the local level with the purpose of including the public in the political process (Lambertz 2022) to help combat disillusionment and mistrust (*An Introduction to Deliberative Democracy for Members of Parliament* 2021).

AI tools can be deployed in various ways as part of supporting participation efforts.

AI tools can be deployed in various ways as part of supporting participation efforts. Deliberative democracy methodologies themselves vary widely and can take place both online and in-person – with AI being discussed more predominantly in the online modality. From *Mass Online Deliberation* (Velikanov/Prosser 2017) to mini-publics, these online efforts at participation have seen participants engage in budget allocation, strategic planning and open-ended discussion and opinion gathering regarding high-impact topics such as taxation, urban planning, and low consensus topics such as tackling climate change (Mackisack 2023).

AI Can Help to Facilitate Discourse

Platforms designed for deliberative practice and citizen assemblies include *Polis*, *Citizen Lab* and *Policy Synth* – the latter of which describes itself as attempting to scale crowdsourced policy

making (Bjarnason et al. 2024). These platforms engage participants in different ways including recognisable Social Media mechanisms such as commenting and likes or upvoting. To facilitate discourse these platforms have integrated *Natural Language Processing* (NLP) and machine learning tools to conduct sentiment analysis, group proposals and surface themes. AI-generated graphical representations of themes and sentiments can also be used to help participants better understand the opinion space of the overall discussion (Tsai et al. 2024). These visual representations can be layered with AI-filtered information such as key comments or main reasons given by participants to provide further context. *Consider.it* (*Consider.It – Democracy Technologies Database* n.d.), for instance, provides participants with visual representations of opinions on a spectrum that can be filtered by users based on vectors of interest from demographic to political party. Other ways to use AI to facilitate discourse include the use of chatbots to moderate discussion at scale (*A Moderator ChatBot for Civic Discourse* 2020) as well as to filter — and sometimes remove — speech considered hateful to maintain civility in discussions.

The *Alan Turing Institute* is currently investigating ways to use NLP techniques to enhance citizen policy proposals for local government, with the project suggesting that machine learning and NLP tools can assist with the "information overload" that can impede "collective sense-making" (*Citizen Participation and Machine Learning for a Better Democracy* n.d.). These remarks echo reflections on the challenges of scaling deliberative democracy exercises, which often generate large amounts of input and interaction that can be difficult to sort and moderate. AI assistance can help address some of these limitations and streamline deliberations to make it easier for participants to engage.

Some of the ideas regarding AI and discourse facilitation apply both to specifically designed platforms and wider Social Media environments. In particular counterspeech or speech designed to respond to online hate or correct misinformation is being explored for its potential to keep discourse civil (*A Toolkit on Using Counterspeech to Tackle Online Hate Speech* n.d.; *Counterspeech* n.d.).

AI Can Be Used to Better Inform Citizens

Access to information is a key pillar of engagement within democracies, one that is increasingly critical in the context of political and informational polarisation. In a similar vein to addressing the information overload of deliberative democracy exercises, the breadth and depth of information poses a challenge to citizens seeking to learn more about policies or politicians (Martin et al. 2021). Further micro-targeting and algorithmically mediated news environments can reify existing information silos. Those considering the additive democratic aspect of AI propose that AI can address these informational challenges by helping citizens to educate themselves. Suggestions include making it easier for citizens to learn more about policy issues or key aspects of a politician's stance on issues of interest – all of which help refine existing questions citizens may have and lower the time necessary to access this information. Other suggestions are more ambitious; from political recommender systems integrated with chatbots (Schneier et al. 2023) and AI tracking tools depicting changes in policies or policy discussions to keep citizens updated, to AI tools that suggest the best candidate matches based on an individual's preferences (Kamoen/ Liebrecht 2022).

Access to information is not just about reducing information overload, it is also about ensuring individuals can engage in the ways most comfortable to them and can access the services they need. Translation of government websites, as well as GenAI powered chatbots can help users navigate services better and enable citizens to find answers without the overhead costs of human assistance, or limitations of working hours. In Norway, a chatbot deployed during Covid-19 allowed citizens to get answers to their queries at any time of day. Further, the chatbot reduced the burden on government staff by being able to sufficiently respond 80% of the time – thereby minimising the need for civil servants to engage (OECD 2024: 5). Chatbot adoption by governments has so far focused on "simple advice and information purposes" (Cortés-Cediel et al. 2023) which could point to limitations of AI chatbots as part of more nuanced or complex aspects of citizen engagement.

AI Can Support Sharing Back Public Opinion with Leaders

Democracy is not just about citizens; it is also about their leaders. Recent polling shows how acutely the responsiveness of elected officials shapes citizen experience with democracy (Clancy 2024). It follows therefore that the emphasis on AI in participation has also looked at the feedback loop between citizens and leaders. Similar to suggestions that LLM tools can organise discourse, they have also been proposed as a way of organising public input and opinion for leaders (*AI4Democracy Series 1* 2024). The *World Bank* for instance has pointed to the ability of such tools to delineate location specific variance within public opinion to assist policymakers in identifying concerns specific to areas or communities (*Generative Artificial Intelligence as an Enabler for Citizen Engagement,* n.d.). This can be done with data collected for the express purpose of giving input to policy or elected officials or can be applied to online sources of information through sentiment analysis of Social Media to ascertain what is important to members of the public.

2.3 Evaluating AI's Democratic Dividends

The ability of AI tools to improve democracy remains emergent, both because AI tools are continuing to evolve, and because the way in which they are designed and deployed is also still in flux. Chatbot adoption for instance is still nascent, and research on the impacts of using chatbots are understudied in the literature (Larsen/Følstad 2024; Senadheera et al. 2024). Initial evaluations indicate limited research on the effectiveness of these tools at scale (Cortés-Cediel et al. 2023).

Case studies provide us with some insights into how well AI tools can operate in real-time. *vTaiwan* is an emblematic use case of AI tools in the service of democracy, utilising *Pol.is* to achieve "rough consensus" (*vTaiwan* n.d.) on digital-related policy questions. The use of *vTaiwan* regarding legislation on *Uber* operations sought the views of different key stakeholders which were ultimately reflected in legislative changes in *Uber*'s business model in Taiwan. However, *vTaiwan* itself noted specific challenges including reliance on government funding and willingness to discuss topics, as well as constraints on legislative impact as key shortcomings (Hsiao et al. 2018). These shortcomings are related to political will rather than use of AI tools. But deliberative democracy itself is not a fixed concept and assessments of the value of deliberative democracy fluctuate, in part

because of haphazard evaluation (Tan 2021) but also because micro and macro experiences of democracy are not directly translatable (Palumbo 2024).

More exploratory ideas of how to use AI tools to enhance democracy also pose an obstacle to evaluation. Policy making *sandbox tools* or 'governance of silico' experiments — such as synthetic data simulations, AI agents and digital twins — that blend synthetic data with policy design are only just beginning to be utilised within democracy programming (*Bar Ilan University*, Israel & Kera 2024). French municipality agglomeration Paris-Saclay is currently testing the use of a digital twin of their territory to experiment with different energy management scenarios to assist with long-term planning (OECD 2024: 11). These same kinds of participatory thought experiments with urban planning and policy-making point to a new method of inclusion that allows for more open-ended discussion.

There is, however, a larger discussion absent from many of these examples, and that is a more foundational notion of how democratic value is created. Streamlining information and providing wider discussion spaces are only one aspect of inclusion and access. Indeed, some democracy theorists have raised alarm over how the use of AI and GenAI may substitute rather than augment citizenship (Tsai et al. 2024) by creating pathways of engagement that are disconnected from the levers of change.

3. AI's Contribution to Democracy Evaluation

The possible contributions of AI tools to democracy evaluation specifically remain understudied and minimally understood. In order to fully explain what the use of AI may mean for democracy evaluation, it is critical to have a better understanding of individual applications and approaches. This section outlines these specific uses and provides a matrix to capture applications across the stages of evaluation (cf. also Hense 2025). It is important to note that AI tools have setup costs, requiring time, skills and other resources to ensure they are usable and appropriate. More work must be done to outline these costs and compare the value-add of AI tools relative to other approaches.

3.1 Current State of Democracy Evaluation

Before ascertaining the possibility of using AI to evaluate democracy, it is first necessary to detail current evaluation practices of democracy and democratic programming. Democracy evaluation varies depending on the scale of interest and broadly falls into two main categories: on one end democracy indices focused on country-level assessments across time, and on the other end more intervention focused evaluation of specific projects or interventions.

Large scale data sets of the kind used by *Freedom House* in their *Freedom in the World Index*, or the *V-Dem* Projects' *Varieties of Democracy*, along with *Economist Intelligence Unit*'s *Democracy Index* generate composite measures of democracy to categorise each country in the dataset. They look

at democracy from a macro level, considering facets such as the electoral process, government responsiveness and engagement with citizens, along with civil and political freedoms. Each measure of democracy is calibrated differently, and democracy experts disagree about their validity. These measures can help provide evaluations at the systems level, and help evaluate democratic assistance programs from international donors, elections monitoring by international organisations such as the *Organisation for Economic Co-operation and Development* (OECD).

Methodological debate regarding how best to measure and therefore evaluate shifts in democratic status can result in highly varied conclusions. Little and Meng (2024) contest the degree of democratic backsliding suggested by other scoring frameworks, questioning the subjectivity of expert coders. However, rebuttal from *V-Dem* institute regarding the repeatability of expert coding methods speaks to the opposite (Knutsen et al. 2023). Disagreement on critical matters such as the extent of democracy backsliding is demonstrative of a lack of convergence around what factors indicate democracy. The lack of interchangeability between datasets is indicative of how influential inclusion and exclusion properties can be on measures of democracy (Vaccaro 2021).

[...] The ability to apply AI in democratic evaluation looks different according to the relevant measure in question.

In more intervention-focused programming that looks at strengthening democracy at local levels, the emphasis is on civic and political participation to bolster the democratic process and can span both formal and informal processes. These processes can be evaluated on both quantitative and qualitative dimensions; indeed one study of program evaluation in the democratic space found a wide range of study designs utilised (Link 2024). Increasingly however, we observe programmatic evaluation in the democracy space being influenced by broader qualitative participatory evaluation trends. Evaluation design including semi-structured interviews, surveys, and focus group discussions align with the interest in participatory approaches to policy making and are reflective of the initial democratic intervention. It makes sense therefore that evaluators of democracy programming are increasingly seeking to democratise their own evaluation practices (Link 2024: 17).

This shift has significant implications on the degree to which AI is appropriate for inclusion in the evaluation of democratic programs. These two scales of evaluation are not mutually exclusive – for instance specific community programming or changes to policy making at the local level can be foregrounded against longitudinal data from large scale democracy evaluations. The utility of outlining these potentially complimentary but distinct approaches to measuring change is that the ability to apply AI in democratic evaluation looks different according to the relevant measure in question.

In the next section we will discuss the manner in which AI might be applicable across these two cases, as well as highlight potential drawbacks of using AI to evaluate democratic programming in both large- and small-scale datasets.

3.2 AI Tools for Evaluation of Democracy and Democracy Programs

Though evaluations differ across programming, there is typically a standard shared process across most evaluations; the underlying steps follow a clear pattern: evaluation design, data collection, synthesis and analysis, and reporting. In this section we consider what specific AI tools can be used at each stage of evaluation, provide description and possible applications of each kind of tool, before discussing key challenges that should mediate tool usage.

Discrete AI tools can assist with specific aspects of an evaluation process. There are numerous aspects of the evaluation lifecycle that can be augmented by AI tools. Figure 1 provides a non-exhaustive matrix of potentially relevant AI tools, and the evaluation stages in which they could be used. In drawing up this table we focused on the tools most likely to be used and be useful in the evaluation lifecycle.

Critical to note is that our emphasis is not on replacing but rather complimenting evaluation practices with precise and discrete uses of AI tools. Some general-purpose AI tools are built to respond to an almost unlimited range of prompts and requests and so can span the evaluation lifecycle. However, it is important to consider the appropriateness, possible risk, and utility of each application of general-purpose AI at each step of the evaluative process. To this end the matrix depicts different applications of AI and the stages of evaluation these uses are relevant to.

3.3 Potential Applications of AI with Specific Application or Utility for Democracy Evaluators

Sentiment Analysis/Opinion Mining

Sentiment analysis extracts insights about the emotional inflection of text using a combination of data mining, machine learning, AI and computational linguistics. Typically, sentiment analysis offers an assessment of the degree of positive, negative or neutral sentiment expressed in large quantities of data. This is especially useful for qualitative interviews, deliberations or participatory evaluation.

Democraping

A combination of big data analysis and web scraping, *democraping* refers to the "analysis of digital trace data [...] for political means" (Ulbricht 2020: 427) that extracts and combines insights from a range of feedback mechanisms such as online discussion forums, Social Media information, citizen assemblies and so forth. Currently the focus is on using democraping as part of policymaking, but it can also be used as a benchmark against which incipient policy can be measured as an indicator of the lag between policy making and policy preference.

Fig. 1: **AI Tools matrix for evaluation.**

	Evaluation Design	Data Collection	Synthesis and Analysis	Reporting
Sentiment Analysis/ Opinion Mining		✳		
Demoscraping		✳		
Signal Monitoring		✳		
Codebook and Data Labelling Generation		✳	✳	
Evaluation Synthesis and Summarisation			✳	
Thematic Analysis and Insight Extraction			✳	
Transcription and Translation		✳		✳
Data Labelling			✳	
Data Visualisation			✳	✳
Data Querying		✳		✳
Amalgamating/Process-ing/Comparing Data across Sources		✳	✳	
Digital Trace Analysis		✳	✳	
Location Specific Variance Analysis		✳	✳	
Chatbots		✳		
Text Generation	✳	✳	✳	✳
AI Agents	✳	✳	✳	✳

Signal Monitoring

Signal monitoring uses Social Media scraping, sentiment analysis and emergent topic tracking to indicate growing areas of conflict or potentially antagonistic discussions. In the democracy context these early warning indications can allow practitioners and evaluators to trace more nuanced and subtle shifts within the public arena, and track changes over time. AI tools can be used to track conversations and determine where certain topics or threads of conversation are ema-

nating from and perform network analysis to identify how related information such as posts are being shared. This offers opportunities for longitudinal monitoring and evaluation to understand shifts in public perception over time.

Text Generation

AI generated content allows for more rapid evaluation reporting. If a key aspect of democracy is the feedback loop between citizens and policymakers, the ability of evaluators to create outputs swiftly is critical to supporting this back and forth. By reducing the time taken to create written output evaluators can provide more frequent reports, or more tailored reporting. Reduction of lag time between program termination and program review can ensure that the momentum of successful programs can be upheld.

Transcription and Translation

Scaling participatory approaches to evaluation requires lowering the cost of including larger and more diverse groups. *AI powered transcription and translation* can keep track of larger discussions and enable a wider range of evaluation options elsewhere due to cost savings. This could also help with increasing and improving participation in evaluation of democracy programs and mean more members of linguistically under-represented groups — likely minoritised communities — can also have their voices heard and be better represented within democracy program evaluation.

Data Visualisations

Data visualisation can assist evaluators in two key ways: first it can offer a new way to look at insights by presenting them in a different form, and second graphics, charts, and other visuals can communicate insights to others. AI generated data visuals widen access to data visualisation allowing more evaluators to access this tool and enable wider use of visuals as they can be produced and edited quickly and easily. Data visualisations can also be used in real time as part of feedback or discourse platforms in participatory evaluation processes.

Data Querying

Using AI, evaluators can "converse" with their datasets, to ask questions about the data they have collected. *Retrieval Augmented Generation* (RAG), for example, enhances AI language models by allowing a person to point the AI to relevant information from specific datasets or documents for use when generating responses to questions or prompts. This allows the AI to ground its answers in accurate, specific information rather than relying solely on its training data, which might be quite broad and general. RAG combines language understanding of AI models with accurate, retrievable information from trusted sources, allowing evaluators to better query their data to hone their insights and gain a deeper understanding. *Data querying* can be especially useful for asking questions about demographics that may be obscured in the data.

Digital Trace Analysis

Digital trace analysis refers to the traces left behind by individuals when they use digital technologies. There has been growing interest in the potential insights that can be discerned from this highly granular information, that some argue may offer more nuanced socio-techno insights (Sultan et al. 2023). Digital trace data may be useful for tracking more subtle changes or preferences in relation to democratic programming that are behaviourally rather than verbally signalled.

Location Specific Variance Analysis

Understanding the way outcomes vary across space is vital to discerning less immediately perceptible impacts of democratic programming. AI tools can enhance this level of insight to indicate how program effectiveness is mediated by locale.

Evaluation Synthesis and Summarisation Tools

As AI tools make new forms of data more accessible to evaluators, it is even more pressing that evaluators can process insights from such an array of information. *Text summarisation and synthesis focused NLP tools* can support evaluators in consolidating learnings across dataflows. There are also evaluation-specific methodologies such as synthesising judgements across criteria, sub-criteria, and standards that AI applications may be useful for. Additionally, evidence synthesis of evaluation practices that look at scoping reviews, systematic reviews, umbrella reviews and meta-analyses, could benefit distinctly from the use of AI, helping evaluators to better understand what does and doesn't work. This is critical in the context of democracy programming where vital learnings about what is effective can inform other key interventions to improve the health and robustness of democracy programming.

3.4 Potential Applications of AI with Broad-based Evaluation Utility[1]

Data Cleaning

AI and ML tools can help prepare large datasets for analysis by improving data quality. *Data cleaning tools* remove errors and duplicated information, assist with data validation and data transformation, and identify outliers so that data is ready to use. Machine learning can also help to refine and enhance fuzzy matching between datasets.

Codebook and Data Labelling Generation

Generative text models can be used to *create deductive and inductive codebooks* that make the labelling process more efficient (Katz et al. 2024). Inductive codebooks especially can often be lengthy to put together – an *AI generated codebook* can be checked by human evaluators and tailored as appropriate (thereby reducing initial codebook creation time) or can be used as part of an automated data labelling workflow.

[1] Note this division refers to explicit contribution to democracy evaluation practice in an additive manner. The uses listed here are still important for democracy evaluators but provide value-add for evaluation in a non-specific way.

Data Labelling

Data labelling tools can accelerate data processing by applying a codebook to qualitative research such as interview excerpts and feedback surveys. The use of data labels is known as *supervised learning*. Qualitative research in particular is already drawing on this range of tools that can span suggestion-based tools, to precedent learning tools that utilise few shot learning wherein human evaluators first label several examples to train an AI model. The latter approach could feasibly undertake labelling without an established codebook.

Thematic Analysis and Insight Extraction

NLP can be used to create *thematic analysis* to derive key themes from qualitative evaluation-related input (Zhang et al. 2024). Models usually use either topic modelling or clustering methods to then *extract key themes* (Turobov et al. 2024). These types of procedures are known as unsupervised learning as they don't use data labels. These can be paired with sentiment analysis and fed directly into synthesis or form the basis of a codebook.

Amalgamating/Processing/Comparing Data across Sources

A key aspect of AI tools is the ability to *process information across multiple sources*. Democracy programming, and indeed democracy itself, is an incredibly complex endeavour. Indicators of success therefore may require assessment of several different sources and data types. AI tools, especially multi-model tools that can process different data types, can expand the range or extent of data inputs that evaluators can consider in their evaluations.

Chatbots

Though *chatbots* are not typically thought of as an evaluation tool, AI powered chatbots can be used for conversational surveys that ask open-ended questions (Xiao et al. 2020) to facilitate a more engaged evaluation process. This would allow for individual and individualised engagement where group discussions are not possible or appropriate.

AI Agents

Broadly speaking, an *AI agent* is a piece of autonomous intelligent software that is rooted in a particular information environment and is programmed to recursively carry out multiple tasks according to a particular logic. Like chatbots, AI agents can be seen as a peer or co-worker that can undertake tasks at the prompting of an evaluator, but more so than a standard chatbot, they can be looped to carry out complicated or long processes. Evaluators can interact with AI agents using a Q&A interface like chatbots to ask questions, give instructions and seek support as necessary in the evaluation workflow. There is no predetermined path of interaction, rather AI agents are designed to be multifaceted in their assistance.

3.4.1 AI Tools and Scales of Evaluation in Practice

It is critical to consider use cases when choosing to utilise AI tools and think critically about which tools are useful in practice.

For large scale data sets, the collation and processing of data is critical. Here data querying, demoscraping and location specific variance can enable evaluators to access new sources of information that can help deepen their evaluations and understanding of how and why democratic progress changes. Meanwhile data cleaning can assist with processing and preparing varied data inputs for analysis, and transcription and translation software can allow for more expert evaluators to engage in the evaluation process and provide qualitative background information on their assessments which can be processed using evaluation synthesis and tools designed to bring information from multiple sources and source types together. Together these tools can reduce lag time between data collection and evaluation – which is critical when these datasets are often regarded as policy making tools (Ghioldi 2023).

In the case of intervention-based programming and more participatory approaches to evaluation, lowering the cost and barriers to participation is critical. Chatbots, translation and transcription could allow for greater penetration and participation by allowing evaluators to meet participants where they are. Other tools allow evaluators to access more subtle and implicit information than they may be able to through direct interviews and surveys. Digital trace analysis and sentiment analysis is especially useful to consider here, as they provide evaluators with access to more intangible modalities of evaluation. These methods could also allow evaluators to better understand smaller shifts and changes that may not be perceptible to participants or evaluators themselves and can be used to inform recommendations.

3.4.2 AI and Automated Workflows

Discrete AI tools are just one part of the equation. Larger scale automation of workflows is another possibility that AI introduces. The AI tools outlined above vary in degree of automation but can often be calibrated along the spectrum of suggestions and review based interaction, semi-autonomous and AI-assisted or fully automated. These types of workflows are not clearly delineated and can overlap with one another and one evaluation might employ various of the above tools as part of its methodology.

AI assistants offer a chatbot interface where users can query a large corpus of text, for example program reports or evaluations. In this case, the assistant is trained to search and draw insights from a specific set of documents using an organization's theory of change and other logic models or indicator frameworks to produce synthesised and relevant results. While these uses of AI hold great promise, in practice organisations are still finding challenges in achieving accurate and valid responses (Robert 2024a). AI assistants may be more familiar to evaluators as they are increasingly being integrated within existing proprietary tools and reporting software.

Another approach to achieving extraction of insights from large datasets is a more involved one, which involves creating process-driven AI workflows. Here the idea is that documents, data, and/or methods can be built into a process that can be repeated and executed at scale, utilising AI and other advanced technologies for individual steps in what could be a many-step process. An AI workflow approach aims to train an AI through large and multifaceted requests, split into a series of smaller, single-faceted requests that AI systems can execute with a high degree of reliability. The results of these smaller requests are brought together (like how humans tend to break larger, more-complex jobs into smaller, simpler tasks). While this type of automation has resulted in some level of efficiency, heavy human oversight is still required to review and validate results produced by automated AI workflows for evaluation purposes (Franzen et al. 2022; Robert 2024b).

3.4.3 Participatory Evaluation Using AI

Constructing an evaluation method out of AI parts can be haphazard; the umbrella of participatory evaluation provides a clearer framework to understand how, and to what end, different AI applications can be utilised. Here, many of the AI tools and applications of AI being considered for encouraging engagement and participation in the democratic process can also be utilised for the purposes of making evaluation more participatory (Meylan-Stevenson et al. 2024).

Participatory or deliberative evaluation is not new but has often been burdensome to execute. However, participatory evaluation yields significant democratic returns as participants involved in evaluation are more likely to feel as though they have a stake in the project and its outcomes (Odera 2021). Online tools, along with AI models that can help parse through evaluations in real-time and are financially accessible to evaluators, could usher in new approaches to participatory evaluation (Simon et al. 2017).

Growing use of such platforms in democratic deliberation offers some insights into how participatory evaluation could happen. *ARG-Tech* develops practical AI applications rooted in 'philosophy, linguistics and cognitive science' with an emphasis on argumentation, including argument mining, argument visualisation and analysis, and computation over argument structures (ARG-tech n.d.; Raftree 2024). These tools could be cross-applied for virtual round-table discursive evaluation, or other modes of participatory evaluation where program reviews can be modelled as "arguments" wherein participants are debating the utility and impact of a program.

4. Assessing the Use of AI for Evaluating Democracy Programming

As yet, discussion or case studies of using AI in evaluating democracy programming specifically is minimal. However, there is both speculative and experimental evidence regarding the positive impacts of using AI in general program evaluation.

Most of the upsides noted are related to resources: time (helping evaluators work faster, lower costs); reducing the number of team members required to perform tasks; improving the accuracy of data; and enabling evaluators to gain greater insights into their data. Efficiency gains are a commonly cited positive, especially in regards to knowledge management and corralling large and varied datasets (Adya Makka et al. 2024). The strong pattern recognition ability of AI tools can allow for enhanced analysis for instance through the use of predictive analytics to develop evaluation models for longitudinal and multi-variable programming (York/Bamberger 2020: 2).

More contested positives include the cost saving contribution of AI tools due to the automation of tasks or whole workflows and claims of greater objectivity. Advocates of AI technologies used to predict recidivism in pretrial, parole, and sentencing decisions, for example, have argued that AI-enabled analyses are more accurate and less biassed than those by humans. However, as early as 2016, in an article for *Propublica*, Angwin et al. (2016) showed the opposite. *COMPAS*, an AI tool that predicts future likelihood of criminality, exhibited heavy bias regarding Black people, predicting more often that they would commit future crimes as compared to white people with the same profile. Dressel and Farid (2018) found that *COMPAS* was no more accurate or fair than predictions made by inexperienced individuals given the same task and that a simple linear classifier with only two features achieved the same results as the *COMPAS* algorithm. Non-automated, democratic systems of governance are designed with checks and balances built in with the aim of improving objectivity and reducing bias. Yet in the case of AI, checks and balances are only now being put in place through regulation such as the EU's *2023 Data Act* (European Union 2023).

> *[...] There is both speculative and experimental evidence regarding the positive impacts of using AI in general program evaluation.*

There are several blind spots and downsides to AI approaches to evaluation. These approaches therefore require a dynamic conversation between the potential upsides on one hand and these downsides on the other. There are salient limitations to AI approaches that implicate deployment. Limitations in an evaluation context range from practical limitations such as data quality, hallucinations, and bias, to wider ethical issues such as data labelling labour, supply chain and environmental impacts (Head et al. 2023).

4.1 AI Faces Unreliability and Validity Challenges

Evaluation requires high quality data, and a validated or trustworthy approach to accurately assess a program. Using AI tools within this context means evaluators must be able to discern the extent to which AI tools can meet quality expectations. The performance of AI in establishing trustworthiness or validity varies with the stage and type of data being used, and points to restricted capacity with qualitative trustworthiness (Azzam 2023). Quantitative data proves more straightforward, however there are still important limitations to be mindful of, including the inability of AI

models to observe or collect contextual information to supplement claims of causality (Azzam 2023). This can be mitigated to an extent by reducing how much of this workflow is automated, and supplementing AI facilitated outputs with human expertise, observations, and oversight.

There is a wider unreliability however that has emerged as an increasingly salient limitation of AI tools: "hallucinations" or inaccurate outputs and false information such as false citations or fictional quotes (*When AI Gets It Wrong* n.d.) that can introduce disinformation into reports, data visualisation, and analysis (IBM 2023). To a degree, this shortcoming is built into models which are "designed to always make a prediction, regardless of whether the underlying data is insufficient, inaccurate, or biased, and even when the model does not or cannot know the answer" (Head et al. 2023).

Interest in using LLMs for translation, as well as audio and visual purposes should be scrutinised for the same hallucinatory failures as they rely on conversion of text to output (Rawte et al. 2023). Small changes in input data, bias in data selection, inconsistent or out-of-date information can metastasize through repeated circulation and the inclusion of such outputs into training datasets can result in a contagion effect on outputs. For evaluators of democratic programs considering AI tools for translation, transcription, and analytical purposes it is important to be wary of, as blind spots in foundation models are likely to persist across different use cases (Suresh et al. 2024).

Such limitations are not insurmountable and can be addressed through cross referencing, careful fact checking and human review. However, thorough cross checking is time intensive and may undercut the efficiency claims of AI tool developers and ultimately be insufficient to counter the convergent effect of hallucinations. Other approaches that use external knowledge include interactive question-knowledge alignment, where users can help improve response accuracy by guiding model outputs to better align with factual knowledge (Zhang et al. 2024). This approach may work in settings where bespoke models for evaluation are created but is unlikely to happen with off the shelf solutions. Efforts towards auditability are happening; one method includes using one GPT to query another GPT or engage with a bot to review model validity. Some AI companies are developing ways to address this issue and provide more traceability, and 'cookbooks' have been developed by individuals to guide users of AI chatbots like *Claude* to provide citations and draw out exact quotes (*Anthropic AI* n.d.). However, these methods are still experimental, and the insights such approaches can offer remain uncertain (Mehrotra/Marchman 2024).

4.2 Bias across the AI Lifecycle Can Result in Suboptimal, and Potentially Harmful Outcomes

Well-documented challenges regarding bias in models work against efforts of evaluators to give fair accounting of a program's impacts – both positive and negative. These biases run counter to a growing push to include groups that are often excluded from both democratic processes and evaluation activities. They may also be harder to control as such biases are embedded within tool development — documentation of which is inaccessible to evaluators. In other words, while eval-

uators may make efforts to address bias in tools and outputs or in how they use AI-enabled tools, they have no way to address the fundamental biases that are a core part of how an LLM was built and how it operates. What's more, because algorithms are not transparent or open, evaluators are unable to assess the level of bias inherent in the LLMs on which AI tools are built.

Bias can occur across model development in data collection, data labelling, model training and deployment (*Bias in AI* n.d.). Sampling and selection bias, as well as exclusion bias can shape initial data collection, as datasets are circumscribed by what developers initially choose to include as training data. These decisions are mediated by systematic and cultural biases that privilege certain groups – predominantly white, male, and located in the Global North – resulting in unrepresentative datasets that limit the applicability of systems outside of dominant contexts (Tao et al. 2024). Machine learning models trained on these datasets are imbued with computational biases that reproduce societal prejudices and relational power dynamics, which are then reflected in outputs. This has tangible consequences ranging from uneven distribution of usability and functionality of AI tools (Otis et al. 2023) to downstream semantic bias (Head et al. 2023) — that results in demonstrably stronger alignment between LLMs and western cultural norms and values (Johnson et al. 2022) — to reinforcing racialized and gendered stereotypes that perpetuate real-world harm. A lack of diversity in training sets of facial recognition software has resulted in lower accuracy rates for women and those with darker skin tones, meaning the highest error rate for black women (*Study Finds Gender and Skin-Type Bias in Commercial Artificial-Intelligence Systems* 2018). When training data bias and visualisation bias overlap with societal bias, serious harm can ensue. Johnson and Johnson (2023) describe the consequences of faulty police facial recognition tools that mistakenly identify Black people, finding that use of facial recognition technology disproportionately impacted Black communities and worsened existing inequities in policing.

Proxy measures used when building AI models can contribute to algorithmic bias, with the underlying assumptions of these estimations effectuating further bias. Credit scoring algorithms that look at various aspects of an individual's credit history use data points such as home loan rejection as a proxy for creditworthiness, however individuals from minoritised or low-income backgrounds who are less likely to have as substantive a credit history may receive less accurate credit scores and a higher perceived risk of defaulting (Andrews 2021; O'Shaughnessy 2023).

Across AI models, these different sources of bias compound one another, meaning models do not reflect wider reality and give rise to uneven outcomes. Even more seemingly mundane elements of models can have adverse outcomes, as Guha et al. (2024) found in their work regarding the impact of automated data cleaning on fairness, which notes that different data cleaning methods can reify differences in data quality across demographic groups.

There is also the issue of the technical awareness necessary to determine the particularities of both the AI tool itself and the possible implications of each tool on the intended task. Unintended impacts of AI tools can initially be laborious to identify, and the evolving nature of AI models can place granular details outside the bounds of what evaluators can meaningfully audit. Expertise in both circumstances can circumscribe wider use of AI, and use of off the shelf tools cannot be

reviewed thoroughly or may not be fit for purpose which may then impact the evaluation process or lead evaluators to amend evaluation methods to adhere to what the AI tools can do rather than finding tools that are a good fit for evaluation methods and principles.

4.3 Wider Consequences of Using AI

AI tools do not exist in a vacuum, rather they are products of a wider environment that implicates the appropriateness of using such tools in evaluations of democracy programs.

Foundation models require extensive data collection that poses a risk to digital privacy. Much of the data feeding AI models is collected and used without consent, and the inclusion of personal data within this data harvesting raises privacy concerns. Meanwhile the concentration of the levers of information within *Big Tech* firms is at odds with democratic ownership and access to information and has resulted in *Big Tech* firms having outsized influence over the policy and regulatory space that preferences corporate interests over that of citizens. Reliance on *Big Tech* companies for access to increasingly important foundational models and AI tools further fortifies their power and sway (Kak/Myers West 2023; Khanal et al. 2024). Some would go as far as to say that corporate control of AI systems is fundamentally anti-democratic, considering that *Big Tech* monopolies in the United States have control of data infrastructures that are largely unaccountable to both US citizens and those of other nations. These companies have been described as having "functional sovereignty" due to their size and power (Iazzolino/Stremlau 2024: 10). Algorithmically mediated news undermines journalism and information integrity.

Extractive data practices of model training infringe upon the intellectual property and copyright of citizens, outpacing the ability of regulation to ensure parity between parties and set standards for compensation and data protection for this new use case. Of particular concern for democracy evaluators should be the impact these practices have on journalism and information integrity. Subtle changes to search engines can have adverse effects on how people receive information: AI news summaries that replace search remove critical revenue sources from news outlets and journalists, shrinking their fiscal basis and threatening the viability of journalism (Simon 2024). The bias and unreliability, along with mis- and disinformation of models discussed above, render the sidelining of news outlets a severe threat to the basis of democracy's "third pillar". Financial realities have seen some news outlets fuel these very AI systems by signing licensing deals with large AI firms that allow their data to be used in training models (Brown 2024). The demise of independent news outlets and increased financial precarity for journalists diminishes the imperils of the information ecosystem and reduces the sustainability of trustworthy and unbiased news sources that are central to democracy.

4.3.1 Uneven Regulatory Practices Are Imbued with Geopolitical Power Dynamics

Regulation over how all this data can be used — and how this data usage should be compensated — is still in progress, and legal coverage is haphazard (Sher et al. 2023). As it stands, AI Governance and the ability of governments to properly implement regulation are influenced by larger geopolitical dynamics of economic influence. Research on the current exclusion of the Global Majority in determining AI Governance by the *Carnegie Endowment for International Peace* noted that despite being likely to disproportionately feel the negative effects of AI systems, Global South countries and individuals of the Global Majority were largely sidelined in governance discussions. Global AI Governance is thus heavily skewed towards the experiences and preferences of the Global North despite being intended to apply universally (Anthony et al. 2024). The lack of inclusion and participation in AI Governance decisions is most likely to impact groups that AI development already marginalises, thereby exacerbating exclusionary dimensions of AI systems (Okorie/Marivate 2024). Indeed, Global Majority populations are already bearing the brunt of the human and environmental costs of AI systems even as they struggle to access benefits from systems and tools that are not designed for use in these contexts (Holden/Harsh 2024). Rights-violating working environments for data labellers (Rowe 2023) and environmentally destructive resource extraction are but a few of the acute costs of AI systems. The unequal distribution of benefits and harms between the Global North and Global South, coupled with the relative lack of power and influence of the Global South over AI Governance points to a deeply unjust backdrop to AI systems that is untenable from a democratic perspective.

5. Ethics of Using AI for Democracy Evaluation

The ethics of using AI-enabled evaluation for democracy programming is an emergent consideration, and as yet there is no definitive guidance for this area. We can consider there to be ethical concerns that relate to democracy specifically, and ethical concerns related to AI tools and systems more broadly that have distinct consequences for evaluating democracy programs.

In the context of democracy programming ethical concerns there are direct implications of AI tools and models on democracy and democratic processes. Misalignment of the logic underpinning AI and democratic values of making "decisions in the midst of great ambiguity and contingency" (Bosoer et al. 2022: 7) gives rise to an epistemic difference in how political legitimacy is conferred upon action. AI automation and mediation of democratic processes risks creating distance between elected officials and citizens as AI tools programmed to seek linear pathways flatten difference, or potentially value compromise over representation of outlier opinions (Tsai et al. 2024). Evaluators of democratic programs should be cognisant of how using these tools may reduce representativeness in their analysis. This does not need to be inherent to the AI tool but can also include the way AI tools can rapidly magnify and entrench existing discrimination and oppression. The ease with which AI tools can do so should be part of wider reflection on whether AI tools are appropriate to use for evaluating democracy programs – the same facets that mean such tools

can undermine the resilience of democracy also circumscribes their ability to extend democratic values of inclusion and accessibility.

These shortcomings are unlikely to be addressed in a closed AI development system that locks out the wider populous in model design. The undemocratic development process at the heart of AI systems permeates AI and is visible in the exclusionary dimensions of data collection bias, algorithmic bias and semantic bias to name a few. *Big Tech* companies are actively working to prevent democratic oversight or input into how systems can be developed and used, violating principles of consent and conducting themselves in a covert manner that is in opposition to democratic values of transparency. Holding *Big Tech* companies to account has proven challenging, with numerous *Big Tech* companies using the cover of trade secrets to stymy oversight (Schaake 2020) and courts dependent on *Big Tech* companies themselves to enforce rulings meaning in some instances only *Big Tech* can truly hold itself accountable (Srivastava 2023: 995). Principles of consent, oversight, and environmental rights are routinely violated by *Big Tech* companies who are not subject to democratic oversight. Vesting critical means of democratic engagement and evaluation consolidates the influence such companies have over critical aspects of societal organisation.

A wider risk of cementing a technical approach to evaluation is a creeping technocratisation of how democracy programs are evaluated. Increased use of AI tools privileges a technocratic approach suffused with a worldview that skews towards Global North values that are rooted in histories of colonialism, imperialism and extraction and crucially which asserts a particular approach to knowledge making that often overlooks indigenous and community forms of knowledge. Evaluations of democracy programming form a feedback loop with policy making and other decision-making regarding the shape of democracy. Evaluators must therefore not overlook the more surreptitious ways that use of AI tools could alter how democracy and democracy programming is understood and engaged with.

The ease with which AI tools can reinforce and entrench existing discrimination and oppression should be part of a broader reflection on whether AI tools are appropriate for evaluating democracy programs.

Lastly, a core principle of democracy is that of inclusion and equality. Questions remain over which groups of citizens will be able to utilise more participatory approaches to democracy programming or subsequent evaluations. Involvement is often determined by external factors such as time, resources and awareness – all of which are heavily delineated by gender, race and socio-economic class (Mikhaylovskaya 2024: 8), meaning that the groups currently most marginalised from the democratic process are likely to continue to face barriers to inclusion in AI enabled democratic programming.

5.1 Addressing Ethical Challenges

Evaluators of democracy programs must think critically about the ramification of these ethical challenges for evaluation. Efforts to address some of these issues are underway, foremost of which is the push to democratise AI. Proponents of democratising AI have looked at intervening in different parts of the AI life cycle and introducing means of participation and engagement with different groups of people – especially those likely to be impacted by AI systems. The *Centre for AI Governance* has divided democratisation efforts into four categories: "democratisation of AI use, democratisation of AI development, democratisation of AI benefits, and democratisation of AI governance" (Seger 2023: 1). Participatory AI proffers similar ideas – the OECD's *Participatory AI Framework* encourages co-design and consultation throughout the AI lifecycle (OECD AI n.d.).

Grassroots efforts to address data collection for AI models focus on countering both the extractive nature of current data collection and the lack of representation both within data collection and data collectors. Participatory data stewardship models (Kelly et al. 2023: 1783; Sridharan/Girish 2022: 10–13) such as the *MT* project by the *Masakhane* NLP community showcase one example of community-centred AI. *Masakhane* is a volunteer-led open-source project working to create a machine translation tool which will then translate African papers into African languages. This fills in key gaps for African languages overlooked by larger ML translation services that leave many African countries on the backfoot in terms of access to scientific research and translation tools (*Masakhane – Masakhane MT* n.d.). Meanwhile the *Te Hiku* NLP project seeks to both empower and protect Maori speakers and the Maori language by engaging Maori community members as co-creators in building NLP tools for the Maori language and keep Te Reo Maori linguistic data away from *Big Tech* (Hao 2022; *Papa Reo – a Language Platform for a Multilingual Aotearoa* n.d.). Lastly, *Ubuntu AI* works to undo the extraction of value from African artists by collaboratively creating a licensable portfolio of work to avoid their intellectual property being used without permission or compensation (*Ubuntu AI* 2023). These grassroots efforts are a reclamation of AI systems and the data that feeds them and are representative of a wider push to remake AI systems so that they serve communities first and foremost.

Another approach is to address the transnational scales at which *Big Tech* Companies and the models they build operate. Localism in AI systems seeks to refocus decision making in more proximate structures such as cities and communities, where local decision makers look at how AI is used in their vicinity (Verhulst/Sloance 2020; *The Gov Lab* n.d.). One example of how this could look is *Urban AI*, a French thinktank trying to create a more dynamic exchange between municipalities and AI tool use (Nelson et al. 2023; Urban AI n.d.).

More institutionalised efforts to democratise the governance of AI would point to rather different understanding of what democratisation entails. Key AI companies such as *Meta* (Meta 2022), *Open AI*, *Microsoft* (Microsoft News Center 2016), *Hugging Face* and *Anthropic* have all expressed their interest in democratising AI (Wetherall-Grujić 2023). Democratisation in this sense is animated by altogether separate claims about ease of use or increased access to a skill set – such as illustration – previously the purview of experienced or trained individuals. Several initiatives

utilise AI itself in these efforts; *Anthropic* used *Pol.is* for input on a publicly-sourced constitution for an AI system (*Anthropic* 2023).

Divergence between these approaches belies deeper differences regarding what democratisation or participation is meant to achieve (Birhane et al. 2022). Grassroots visions of democratising AI suggest a more local and proximate approach to AI design, development and deployment that is rooted in community oversight and potentially extends to sovereignty over the data used to build models. Corporate conceptions meanwhile suggest democratisation is about the reach of AI tools, and the ability to deploy tools across a wider range of contexts. What the purpose of democratisation is changes who is spearheading the democratisation process, and by extension who has the power to influence key outcomes regarding AI (Seger et al. 2023).

6. Recommendations for Responsible Adoption of AI in Evaluation of Democracy Programming

The integration of AI into democracy evaluation presents a complex landscape of potential benefits and challenges. While AI tools offer potential enhancements to democratic systems and subsequent evaluation efforts, concerns persist about their development, implementation, and impact. These tools bring a range of capabilities that could add value, yet there are worries about inherent biases and the fundamentally anti-democratic nature of their creation. Issues of inclusiveness and the potential detachment of emerging technologies from democratic principles also loom large. The field's nascent state is evident in the lack of rigorous assessment of AI for evaluation, particularly in the context of democracy programming. This situation is complicated further by the limited AI expertise among evaluators, who may feel pressured to adopt these tools despite reservations. The growing corporate dominance in AI and AI for development raises additional concerns, as evaluators find themselves with little influence over the direction of *Big Tech*'s innovations. Balancing these potential downsides against the possible upsides of AI in democracy programming requires careful consideration and ongoing scrutiny.

A robust, evidence informed, and holistic approach to exploration of these tools and their use in evaluation of democracy and democracy programming is needed. This will require input and investment from various stakeholders, including policy makers, evaluation associations and academic institutes, democracy programming organisations, and evaluators themselves.

6.1 Recommendations for Policymakers

Policymakers play a key role in regulating AI, funding systematic exploration of the potential for AI in evaluation for democracy and democracy programming, and strengthening literacy and capacity in the area of AI at the level of ministries and local governments and within government sections that touch on democracy programming such as education departments, election boards, and monitoring and evaluation departments and staff.

Policymakers are already involved in legislation such as the EU *AI Act* and relevant national legislation. Ensuring that these are filtered into local levels will be an important role of governments. Additionally, as policymakers commission evaluations, they will need to be clear about the risks and benefits of AI for evaluation and in democracy programming. Governments can provide funding or more intensive research on the role of AI in society, including how it affects democracy. Governments should remain vigilant about the potential downsides of AI if they integrate AI into their governing processes. Policymakers can lay the groundwork for national level AI literacy and adoption of responsible uses of AI, and they can fund efforts such as communities of practice and academic institutions to enable shared learning.

6.2 Recommendations for Evaluators

Evaluators assessing democracy programs should approach the use of AI tools with caution and strategic consideration. Given the current lack of robust and standardised evaluations for LLM responsibility, it is crucial to be judicious about when and how AI tools are employed. To address this challenge, evaluators should first develop or identify evaluation standards that align with democratic principles and uphold the highest evaluation practices. They should then map out the evaluation process, identifying time-intensive activities and pain points where AI assistance could be beneficial, prior to any AI implementation discussions.

When integrating AI, evaluators should focus on precise and discrete uses where uncertainty has minimal ripple effects, prioritising simpler and more constrained applications. Transparency is key; methodologies should clearly outline AI usage and openly acknowledge potential limitations or impacts of these tools within evaluation frameworks. It's also important to consider alternative non-AI solutions that may mitigate risks or uncertainties, especially when addressing goals like increased participation.

A 'democracy by design' approach should be adopted, extending to the development of AI models themselves, with consideration given to sector-specific AI development. Throughout this process, evaluators must exercise utmost care, minimising the introduction of uncertainty and reflecting methodologically on any constraints imposed by AI tools. Evaluators will need to navigate the complex landscape of AI in democracy programming, ensuring that technological advancements complement rather than compromise democratic principles and evaluation integrity. A critical point will be keeping a 'human in the loop' and ensuring that humans review and verify outputs generated by AI.

When using AI, evaluators must remain aware of the terms and conditions of use that AI tool makers lay out and the privacy implications of different commercial models, open-source tools, and bespoke applications. Overall, evaluators need to constantly upskill themselves to ensure that they are not putting people at risk if they integrate AI into their practice.

6.3 Recommendations for Evaluation Societies and Communities of Practice

Evaluation societies and communities of practice play a key role in supporting the establishment of shared standards and practices. In the case of emerging technologies such as AI and their use in democracy evaluation, it is critical that these associations create space for conversations, shared learning and agreed on good practices and professional standards. These can both help to set ethical standpoints and to generate shared positions on where, when, why and how emerging AI should be used.

7. Conclusion

The use of AI tools in democracy evaluation begets considerations that should precede the application of such tools. Our research approach takes a holistic account of AI use in democratic processes, along with the implications of AI tools *on* democracy to better map the possibilities and limitations of using AI tools to evaluate democracy programs. While AI tools have received much attention for how they might change evaluation, we find it is first critical to establish which evaluation methodologies and processes can realistically benefit from and be integrated with AI tool use. Further, deploying thoughtful and workable evaluation systems that use AI tools requires prior knowledge and familiarity. Expertise and experience using AI tools is a limiting factor that must be at the front of mind when considering the extent to which evaluators can utilise AI tools in practice.

A key question regarding the use of AI tools is the way they may implicate democracy itself, and therefore the extent to which such tools may be appropriate for assessing democracy programs. One developing area of research is the possibility of using AI tools themselves to democratise the process of AI development. If we look at participation as one paradigm of achieving democratic AI, then we encounter a clear ceiling. At the foundation model layer Suresh et al. 2024 challenge the viability of improved participation — especially of marginalised groups in the development of generalised models — suggesting instead that downstream AI for inclusion discussions are more feasible. This misalignment may in part be because of what Himmelreich describes as an incorrect "response to the injustices that animate the call" (Himmelreich 2023: 8) to democratise AI in the first place. Crawford goes further arguing that "[t]o suggest that we democratise AI to reduce asymmetries of power is a little like arguing for democratising weapons manufacturing in the service of peace" (Crawford 2022: 223).

AI tools should enhance evaluation and the work of evaluators rather than merely being a substitute. For this to happen, evaluation practitioners and democracy programmers must first consider what rigorous, long-term evaluation that is in-built from the beginning of projects looks like (Dodsworth/Cheeseman 2018: 5) and what kind of more diffuse

A key question is the way AI tools may implicate democracy itself, and therefore the extent to which such tools may be appropriate for assessing democracy programs.

evaluation vectors need to be captured for this to happen. A consideration of more fundamental questions about the role of evaluation in contributing to change, as well as the role of evaluation for democratic programs specifically can provide scaffolding for deeper technical discussions about the use of specific AI tools (Griffith Centre for Systems Innovation 2024). This may be a tall order for any individual practitioner; here communities of practice can facilitate learning out loud and reduce the barriers to understanding and properly assessing AI tools in evaluation. The AI space is still nascent and the complexity of using tools indicates the importance of a specific and tempered approach to adoption – one that doesn't stymy uptake but rather ensures informed, ethical and effective deployment of AI tools in the sector.

8. Literature

AI4Democracy Series 1: Using AI to Inform Policymaking 2024: IE University, CGC, 17.6.2024, in: https://www.ie.edu/cgc/news-and-events/news/using-ai-to-inform-policymaking-what-we-can-learn-from-3-use-cases-of-talk-to-the-city/ (last accessed: August 11, 2024).

Allen, Danielle/Weyl, Glen E. 2024: The Real Dangers of Generative AI, in: Journal of Democracy 35: 1, 147–162.

Amnesty International 2022: Myanmar: Facebook's systems promoted violence against Rohingya; Meta owes reparations – new report, 29.9.2022, in: https://www.amnesty.org/en/latest/news/2022/09/myanmar-facebooks-systems-promoted-violence-against-rohingya-meta-owes-reparations-new-report/ (last accessed: August 2, 2024).

Amnesty International 2023: Ethiopia: Meta's Failures Contributed to Abuses against Tigrayan Community during Conflict in Northern Ethiopiain, 31.10.2023, in: https://www.amnesty.org/en/latest/news/2023/10/meta-failure-contributed-to-abuses-against-tigray-ethiopia/ (last accessed: August 2, 2024).

A Moderator ChatBot for Civic Discourse 2020: Stanford University, Human-Centered Artificial Intelligence, 6.2.2020, in: https://hai.stanford.edu/news/moderator-chatbot-civic-discourse (last accessed: August 2, 2024).

Andrews, Edmund L. 2021: How Flawed Data Aggravates Inequality in Credit, in: Stanford University, Human-Centred Artificial Intelligence, 6.8.2021, https://hai.stanford.edu/news/how-flawed-data-aggravates-inequality-credit (last accessed: August 15, 2024).

Angwin, Julia/Larson, Jeff/Mattu, Surya/Kirchner, Lauren 2016: Machine Bias, in: Propublica, 23.5.2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (last accessed: August 15, 2024).

An introduction to deliberative democracy for members of parliament 2021: Westminster Foundation for Democracy, in: https://www.wfd.org/sites/default/files/2021-11/WFD_newDemocracy_An-introduction-to-deliberative-democracy-for-members-of-parliament_2021.pdf (last accessed: August 15, 2024).

Anthony, Aubra/Sharma, Lakshmee/Noor, Elina 2024: Advancing a More Global Agenda for Trustworthy Artificial Intelligence, in: Carnegie Endowment for International Peace, 30.4.2024, https://carnegieendowment.org/research/2024/04/advancing-a-more-global-agenda-for-trustworthy-artificial-intelligence?lang=en (last accessed: August 15, 2024).

Anthropic 2023: Collective Constitutional AI: Aligning a Language Model with Public Input, 17.10.2023, in: https://www.anthropic.com/news/collective-constitutional-ai-aligning-a-language-model-with-public-input (last accessed: August 16, 2024).

ARG-tech n.d.: Research, Centre for Argument Technology, in: https://www.arg.tech/index.php/research/ (last accessed: August 13, 2024).

A Toolkit on Using Counterspeech to Tackle Online Hate Speech n.d.: The Future of Free Speech, in: https://futurefreespeech.org/a-toolkit-on-using-counterspeech-to-tackle-online-hate-speech/ (last accessed: August 2, 2024).

Azzam, Tarek 2023: Artificial Intelligence and Validity, in: New Directions for Evaluation 178–179, 85–95.

Bias in AI n.d.: Artificial Intelligence (AI) Hub, Chapman University, in: https://www.chapman.edu/ai/bias-in-ai.aspx (last accessed: August 13, 2024).

Birhane, Abeba/Isaac, William/Prabhakaran, Vinodkumar/Diaz, Mark/Elish, Madeleine Clare/Gabriel, Iason/Mohamed, Shakir 2022: Power to the People? Opportunities and Challenges for Participatory AI, in: Equity and Access in Algorithms, Mechanisms, and Optimization, EAAMO '22: Equity and Access in Algorithms, Mechanisms, and Optimization, Arlington VA USA, 1–8, https://dl.acm.org/doi/10.1145/3551624.3555290 (last accessed: August 16, 2024).

Bjarnason, Róbert/Gambrell, Dane/Lanthier-Welch, Joshua 2024: Using Artificial Intelligence to Accelerate Collective Intelligence. Policy Synth and Smarter Crowdsourcing, in: http://arxiv.org/abs/2407.13960 (last accessed: August 2, 2024).

Blancho, Martì 2024: Spanish Inmates Not to Be Automatically Monitored in Fear of AI Act, in: AlgorithmWatch,

28.3.2024, https://algorithmwatch.org/en/spanish-inmates-not-monitored-in-fear-of-ai-act/ (last accessed: May 15, 2025).

Böhnke, Olaf 2024: The 2024 Super-Election Year: Navigating the Geopolitical Disinformation Maze, in: Foundation for European Progressive Studies, 4.4.2024, https://feps-europe.eu/the-2024-super-election-year-navigating-the-geopolitical-disinformation-maze/ (last accessed: August 2, 2024).

Bosoer, Lucia/Cantero Gamito, Marta/Galariotis, Ioannis/Innerarity, Daniel 2022: Neither Fish nor Fowl. The Challenges of a Democratic AI, in: School of Transnational Governance, Policy Papers Issue: 2022/30, DOI: 10.2870/81475.

Brown, Pete 2024: Licensing Deals, Litigation Raise Raft of Familiar Questions in Fraught World of Platforms and Publishers, in: Columbia Journalism Review, 22.5.2024, https://www.cjr.org/tow_center/licensing-deals-litigation-raise-raft-of-familiar-questions-in-fraught-world-of-platforms-and-publishers.php (last accessed: August 15, 2024).

Chan, Kelvin 2024: AI-Powered Misinformation Is the World's Biggest Short-Term Threat, Davos Report Says, in: AP News, 10.1.2024, https://apnews.com/article/artificial-intelligence-davos-misinformation-disinformation-climate-change-106a1347ca9f987bf71da1f86a141968 (last accessed: August 15, 2024).

Chen, Heather 2024: AI 'Resurrects' Long Dead Dictator in Murky New Era of Deepfake Electioneering, in: CNN, 11.2.2024, https://www.cnn.com/2024/02/12/asia/suharto-deepfake-ai-scam-indonesia-election-hnk-intl/index.html (last accessed: August 2, 2024).

Cirone, Alexandra/ Hobbs, William 2023: Asymmetric Flooding as a Tool for Foreign Influence on Social Media, in: Political Science Research and Methods 11: 1, 160–171.

Citizen Participation and Machine Learning for a Better Democracy n.d.: The Alan Turing Institute, in: https://www.turing.ac.uk/research/research-projects/citizen-participation-and-machine-learning-better-democracy (last accessed: August 2, 2024).

Consider.It – Democracy Technologies Database n.d.: Democracy Technologies, in: https://democracy-technologies.org/tool/consider-it/ (last accessed: August 2, 2024).

Cortés-Cediel, María E./Segura-Tinoco, Andrés/Cantador, Iván/ Rodríguez Bolívar, Manuel Pedro 2023: Trends and challenges of e-government chatbots: Advances in exploring open government data and citizen participation content, in: Government Information Quarterly 40: 4, 101877.

Counterspeech: A Better Way of Tackling Online Hate? n.d.: The Alan Turing Institute, 1.7.2022, in: https://www.turing.ac.uk/blog/counterspeech-better-way-tackling-online-hate (last accessed: August 2, 2024).

Crawford, Kate 2022: Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence, Yale University Press, in: https://yalebooks.co.uk/9780300264630/atlas-of-ai (last accessed: August 16, 2024).

Crystal, Caroline 2023: Facebook, Telegram, and the Ongoing Struggle Against Online Hate Speech, in: Carnegie Endowment for International Peace, 7.9.2023, https://carnegieendowment.org/research/2023/09/facebook-telegram-and-the-ongoing-struggle-against-online-hate-speech?lang=en (last accessed: August 2, 2024).

Dodsworth, Susan/ Cheeseman, Nic 2018: Ten Challenges in Democracy Support – and How to Overcome Them, in: Global Policy 9: 3, 301–312.

Dressel, Julia/Farid, Hany 2018: The accuracy, fairness, and limits of predicting recidivism, in: Science Advances 4: 1, eaao5580.

Feldstein, Steven 2020: State Surveillance and Implications for Children, in: Unicef, Issue Brief No. 1, https://www.unicef.org/innocenti/media/1136/file/UNICEF-Global-Insight-data-governance-surveillance-issue-brief-2020.pdf (last accessed: April 10, 2025).

Feldstein, Steven 2021: The Rise of Digital Repression. How Technology is Reshaping Power, Politics, and Resistance, Oxford University Press, New York, in: https://global.oup.com/academic/product/the-rise-of-digital-repression-9780190057497?cc=de&lang=en& (last accessed: April 10, 2025).

Franzen Samuel/Quang, Coung/Schweizer, Lukas/Budzier, Alexander/Gold, Jenny/Vellez, Mercedes/Ramirez, Santiago/Raimondo, Estelle 2022: Advanced Content Analysis: Can Artificial Intelligence Accelerate Theory-Driven Complex Program Evaluation?, in: IEG, World Bank Group, 15.2.2022, https://ieg.worldbankgroup.org/methods-resource/advanced-content-analysis-can-artificial-intelligence-accelerate-theory-driven-complex (last accessed: August 15, 2024).

Gao, Biao/Wang, Yiming/Xie, Huiqin/Hu, Yi/Hu, Yi 2023: Artificial Intelligence in Advertising: Advancements, Challenges, and Ethical Considerations in Targeting, Personalization, Content Creation, and Ad Optimization, in: SAGE Open 13:4, 21582440231210759, DOI: 10.1177/21582440231210759.

Generative Artificial Intelligence as an Enabler for Citizen Engagement n.d.: World Bank Blogs, in: https://blogs.worldbank.org/en/governance/generative-artificial-intelligence-enabler-citizen-engagement (last accessed: August 11, 2024).

Ghioldi, Mario 2023: Development Aid and the Democratic Process in the OSCE Region, in: Mihr, Anja/Sorbello, Paolo/Weiffen, Brigitte (Hrsg.): Securitization and Democracy in Eurasia. Transformation and Development in the OSCE Region, Cham, 19–41.

Global Risks 2024: Disinformation Tops Global Risks 2024 as Environmental Threats Intensify 2024: World Economic Forum, 10.1.2024, in: https://www.weforum.org/press/2024/01/global-risks-report-2024-press-release/ (last accessed: August 2, 2024).

Griffith Centre for Systems Innovation 2024: Now We Are All Measuring Impact — but Is Anything Changing?, in: Good Shift, in: https://medium.com/good-shift/now-we-are-all-measuring-impact-but-is-anything-changing-1dc998b95c83 (last accessed: August 16, 2024).

Grøndahl Larsen, Anna/Følstad, Asbjørn 2024: The impact of chatbots on public service provision: A qualitative interview study with citizens and public service providers, in: Government Information Quarterly 41: 2, 101927.

Guha, Shubha/Khan, Falaah Arif/Stoyanovich, Julia Schelter, Sebastian 2023: Automated Data Cleaning Can Hurt Fairness in Machine Learning-based Decision Making, in: 2023 IEEE 39th International Conference on Data Engineering (ICDE), Anaheim, CA, 3747–3754.

Gül, Ayşen Akkor/Ertürk, Yıldız Dilek/Elmer, Paul (Hrsg.) 2020: Digital Transformation in Media & Society, Istanbul, in: https://iupress.istanbul.edu.tr/en/book/digital-transformation-in-media-society/home (last accessed: April 10, 2025).

Haile, Jemal Mohammed 2024: Social media for diffusion of conflict & violence in Ethiopia: Beyond gratifications, in: International Journal of Educational Development 108: 103063.

Hao, Karen 2021: Artificial Intelligence. How Facebook got addicted to spreading misinformation, in: MIT Technology Review, 11.3.2021, https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/ (last accessed: August 16, 2024).

Hao, Karen 2022: Artificial Intelligence. A New Vision of Artificial Intelligence for the People, in: MIT Technology Review, 22.4.2022, https://www.technologyreview.com/2022/04/22/1050394/artificial-intelligence-for-the-people/ (last accessed: August 16, 2024).

Head, Cari Beth/Jasper, Paul/McConnachie, Matthew/Raftree, Linda/Higdon, Grace 2023: Large Language Model Applications for Evaluation: Opportunities and Ethical Implications, in: New Directions for Evaluation 2023: 178–179, 33–46.

Himmelreich, Johannes 2023: Against "Democratizing AI", in: AI & SOCIETY 38: 4, 1333–1346.

Holden, Kerry/Harsh, Matthew 2024: On pipelines, readiness and annotative labour: Political geographies of AI and data infrastructures in Africa, in: Political Geography 113: 103150.

Hsiao, Yu-Tang/Lin, Shu-Yang/Tang, Audrey/Narayanan, Darshana/Sarahe, Claudina 2018: vTaiwan: An Empirical Study of Open Consultation Process in Taiwan, in: https://osf.io/xyhft (last accessed: August 11, 2024).

Iazzolino, Gianluca/Stremlau, Nicole 2024: AI for social good and the corporate capture of global development, in: Information Technology for Development 30: 4, 626–643.

IBM 2023: What Are AI Hallucinations?, in: IBM, 1.9.2023, https://www.ibm.com/topics/ai-hallucinations (last accessed: August 13, 2024).

Innerarity, Daniel 2024: Artificial Intelligence and Democracy (programme and meeting document), in: Unesco, https://unesdoc.unesco.org/ark:/48223/pf0000389736 (last accessed: August 2, 2024).

Jalli, Nuurrianti/Wihardja, Maria Monica 2024: Election Integrity in the Age of Artificial Intelligence: Lessons from Indonesia 2024: in: FULCRUM, 25.6.2024, https://fulcrum.sg/election-integrity-in-the-age-of-artificial-intelligence-lessons-from-indonesia/ (last accessed: August 2, 2024).

Johnson, Rebecca L./Pistilli, Giada/Menédez-González, Natalia/Duran, Leslye Denisse Dias Duran/Panai, Enrico/Kalpokiene, Julija/Bertulfo, Donald Jay 2022: The Ghost in the Machine has an American accent: value conflict in GPT-3, in: http://arxiv.org/abs/2203.07785 (last accessed: August 15, 2024).

Johnson, Thaddeus L./Johnson, Natasha N. 2023: Police Facial Recognition Technology Can't Tell Black People Apart, in: Scientific American, 18.5.2023, https://www.scientificamerican.com/article/police-facial-recognition-technology-cant-tell-black-people-apart/ (last accessed: August 15, 2024).

Kalluri, Pratyusha Ria/Agnew, William/Cheng, Myra/Owens, Kentrell/Soldaini, Luca/Birhane, Abeba 2023: The Surveillance AI Pipeline, in: http://arxiv.org/abs/2309.15084 (last accessed: August 2, 2024).

Kak, Amba/ Myers West, Sarah 2023: 2023 Landscape: Confronting Tech Power, in: AI Now Institute, 11.4.2023, https://ainowinstitute.org/2023-landscape (last accesse: August 15, 2024).

Kamoen, Naomi/Liebrecht, Christine 2022: I Need a CAVAA: How Conversational Agent Voting Advice Applications (CAVAAs) Affect Users' Political Knowledge and Tool Experience, in: Frontiers in Artificial Intelligence 5: 835505.

Katz, Andrew/Gerhardt, Mitchell/Soledad, Michelle 2024: Using Generative Text Models to Create Qualitative Codebooks for Student Evaluations of Teaching, in: http://arxiv.org/abs/2403.11984 (last accessed: August 11, 2024).

Kelly, Sage/Kaye, Sherrie-Anne/White, Katherine M./Oviedo-Trespalacios, Oliver 2023: Clearing the way for participatory data stewardship in artificial intelligence development: a mixed methods approach, in: Ergonomics 66: 11, 1782–1799.

Kera, Denisa Reshef/Navon, Eilat/Wellner, Galit/Kalvas, Frantisek 2024: Governance in Silico: Experimental Sanbox for Policymaking over AI Agents, in: DRS2024, Boston, in: https://dl.designresearchsociety.org/drs-conference-papers/drs2024/researchpapers/11/ (last accessed: August 11, 2024).

Khanal, Shaleen/Zhang, Hongzhou/Taeihagh, Araz 2024: Why and how is the power of Big Tech increasing in the policy process? The case of generative AI, in: Policy and Society: 0: 0, 1–18.

Knutsen, Carl Henrik/Marquardt, Kyle L./Seim, Brigitte/Coppedge, Michael/Edgell, Amanda B./Medzihorsky, Juraj/Pemstein, Daniel/Teorell, Jan/Gerring, John/Lindberg, Staffan I. 2023: Conceptual and Measurement Issues in Assessing Democratic Backsliding, in: PS: Political Science & Politics 57: 2, 162–177.

Krimmer, Robert/Rabitsch, Armin/Kužel, Rast'o/ Achler, Marta/Licht, Nathan 2022: Elections in Digital Times: A Guide for Electoral Practitioners, in: Unesco, https://unesdoc.unesco.org/ark:/48223/pf0000382102/PDF/382102eng.pdf.multi (last accessed: August 2, 2024).

Łabuz, Mateusz/Nehring, Christopher 2024: Information Apocalypse or Overblown Fears—What AI Mis- and Disinformation Is All about? Shifting Away from Technology toward Human Reactions, in: Politics & Policy 52:4, 874–891, https://onlinelibrary.wiley.com/doi/abs/10.1111/polp.12617 (last accessed: August 2, 2024).

Lambertz, Karl-Heinz 2022: Beyond Elections: The Use of Deliberative Methods in European Municipalities and Regions, in: https://rm.coe.int/0900001680a5b00d (last accessed: April 10, 2025).

Leingang, Rachel 2024: 'Disinformation on Steroids': Is the US Prepared for AI's Influence on the Election?, in: The Guardian, 26.2.2024, https://www.theguardian.com/us-news/2024/feb/26/ai-deepfakes-disinformation-election (last accessed: August 2, 2024).

Link, Rouven 2024: Programs Improving Civic and Political Participation: A Rapid Review, in: APO, Analysis & Policy Observatory, 28.3.2024, https://apo.org.au/node/326632 (last accessed: August 11, 2024).

Little, Andrew T./Meng, Anne 2024: Measuring Democratic Backsliding, in: PS: Political Science & Politics 57: 2, 149–161.

Mackintosh, Eliza 2021: Facebook Knew It Was Being Used to Incite Violence in Ethiopia. It Did Little to Stop the Spread, Documents Show, in: CNN, 25.10.2021, https://www.cnn.com/2021/10/25/business/ethiopia-violence-facebook-papers-cmd-intl/index.html (last accessed: August 2, 2024).

Mackisack, Daniel 2023: 6 Participation Tools That Use AI. How Participation Platforms are Integrating Artificial Intelligence, in: Democracy Technologies, 2.5.2023, https://democracy-technologies.org/getting-started/6-participation-tools-using-ai-data/ (last accessed: August 2, 2024).

Makkar, Adya/Rawat, Anmol Singh/Tidiane Toure, Cheick Ahmed/Oza, Priyanka/Sampson, Rob/Ravintula, Sid/Verma, Tanmay/Khan, Zia 2024: How Your Organisation Can Use AI to Improve Efficiency, in: IDinsight, 29.2.2024, https://www.idinsight.org/article/how-your-organisation-can-use-ai-to-improve-efficiency/ (last accessed: August 13, 2024).

Martin, Andreas/Allagha, Mohammad/Misnikov, Yuri 2021: emyConnect: Building a Smart Public Service for Young Mobile Europeans, in: Leitner, Christine/Ganz, Walter/Satterfield, Debra/Bassano, Clara (Hrsg.): Advances in the Human Side of Service Engineering, Cham, 37–44.

Masakhane MT: Decolonise Science n.d.: Masakhane, in: https://www.masakhane.io/ongoing-projects/masakhane-mt-decolonise-science (last accessed: August 16, 2024).

Mehrotra, Dhruv/Marchman, Tim 2024: Perplexity Is a Bullshit Machine, in: Wired, 19.6.2024, https://www.wired.com/story/perplexity-is-a-bullshit-machine/ (last accessed: August 15, 2024).

Meta 2022: Democratizing access to large-scale language models with OPT-175B, 3.5.2022, in: https://ai.meta.com/blog/democratizing-access-to-large-scale-language-models-with-opt-175b/ (last accessed: August 16, 2024).

Meylan-Stevenson, Adam/Hawes, Ben/Ryan, Matt 2024: Democracy Online: Technologies for Democratic Deliberation, in: University of Southampton, WSI White Paper 2024: 2, DOI: 10.5258/SOTON/WSI-WP009, in: https://eprints.soton.ac.uk/491741/ (last accessed: August 13, 2024).

Microsoft News Center 2016: Democratizing AI. For every person and every organization, 26.9.2016, in: https://news.microsoft.com/features/democratizing-ai/ (last accessed: August 16, 2024).

Mikhaylovskaya, Anna 2024: Enhancing Deliberation with Digital Democratic Innovations, in: Philosophy & Technology 37: 1, 3.

Murphy, Hannah 2024: The rising threat to democracy of AI-powered disinformation, in: Financial Times, 11.1.2024, https://www.ft.com/content/16f23c01-fa51-408e-acf5-0d30a5a1ebf2 (last accessed: August 2, 2024).

Nelson, Ruth/Naing, Khant Min/Beroche, Hubert 2023: Generative AI for Urban Governance, in: Urban AI, https://urbanai.fr/wp-content/uploads/2023/10/Generative-AI-Report.pdf (last accessed: April 10, 2025).

Odera, Erica L. 2021: Capturing the Added Value of Participatory Evaluation, in: American Journal of Evaluation 42: 2, 201–220.

OECD 2024: Governing with Artificial Intelligence: Are Governments Ready?, in: OECD Artificial Intelligence Papers: 20, https://www.oecd.org/en/publications/governing-with-artificial-intelligence_26324bc2-en.html (last accessed: August 2, 2024).

OECD AI n.d.: Participatory AI Framework, in: OECD AI Policy Observatory, 27.3.2023, https://oecd.ai/en/catalogue/tools/participatory-ai-framework (last accessed: August 16, 2024).

Okorie, Chijioke/Marivate, Vukosi 2024: How African NLP Experts Are Navigating the Challenges of Copyright, Innovation, and Access, in: Carnegie Endowment for International Peace, 30.4.2024, https://carnegieendowment.org/research/2024/04/how-african-nlp-experts-are-navigating-the-challenges-of-copyright-innovation-and-access?lang=en (last accessed: August 15, 2024).

O'Shaughnessy, Matt 2023: Challenges of Implementing AI With "Democratic Values": Lessons From Algorithmic Transparency, in: Lawfare, 26.4.2024, https://www.lawfaremedia.org/article/challenges-of-implementing-ai-with-democratic-values-lessons-from-algorithmic-transparency (last accessed: August 15, 2024).

Otis, Nicholas G./Clarke, Rowan/Delecourt, Solene/Holtz, David/Koning, Rem 2023: The Uneven Impact of Generative AI on Entrepreneurial Performance, in: https://osf.io/hdjpk (last accessed: August 15, 2024).

Palumbo, Antonino 2024: The Deliberative Turn in Democratic Theory: Models, Methods, Misconceptions (The Theories, Concepts and Practices of Democracy), Cham, in: https://link.springer.com/10.1007/978-3-031-56513-7 (last accessed: August 11, 2024).

Papa Reo – A Language Platform for a Multilingual Aotearoa n.d.: Te Hiku Media, in: https://tehiku.nz/c.C9 (last accessed: August 16, 2024).

Peterso, Dahlia/Hoffman, Samantha 2022: Geopolitical Implications of AI and Digital Surveillance Adoption, in: Brookings, 6.2022, https://www.brookings.edu/articles/geopolitical-implications-of-ai-and-digital-surveillance-adoption (last accessed: August 2, 2024).

Pírková, Eliška/Leufer, Daniel/Maguire, Méabh 2024: Generative AI and Human Rights: What You Should Know, Access Now, in: https://www.accessnow.org/generative-ai-election-disinformation/ (last accessed: August 2, 2024).

Policy Recommendations: Strengthening Democracy n.d.: Freedom House, in: https://freedomhouse.org/policy-recommendations/strengthening-democracy-abroad (last accessed: August 2, 2024).

Raftree, Linda 2024: When Might We Use AI for Evaluation Purposes? A Discussion with New Directions for Evaluation (NDE) Authors, in: The MERL Tech Initiative, 9.4.2024, https://merltech.org/when-might-we-use-ai-for-evaluation-purposes-a-discussion-with-new-directions-for-evaluation-nde-authors/ (last accessed: August 13, 2024).

Rawte, Vipula/Sheth, Amit/Das, Amitava 2023: A Survey of Hallucination in Large Foundation Models, in: http://arxiv.org/abs/2309.05922 (last accessed: August 13, 2024).

Robert, Christopher 2024a: Beyond POCs: Evaluation and QA for production-quality chatbots, in: LinkedIn, 21.3.2024, https://www.linkedin.com/pulse/beyond-pocs-evaluation-qa-production-quality-chatbots-robert-qs7pe/?trackingId=zKgKXule7DxnpLa64EBnJg%3D%3D (last accessed: August 15, 2024).

Robert, Christopher 2024b: Repeatable, reliable, and transparent: Graduating from AI assistance to AI workflows, in: LinkedIn, 1.11. 2024, https://www.linkedin.com/pulse/repeatable-reliable-transparent-graduating-from-ai-workflows-robert-nb4ge/?trackingId=lCmhAUzWQUWevcsKgreMNw%3D%3D (last accessed: August 15, 2024).

Rowe, Niamh 2023: 'It's Destroyed Me Completely': Kenyan Moderators Decry Toll of Training of AI Models, in: The Guardian, 2.8.2023, https://www.theguardian.com/technology/2023/aug/02/ai-chatbot-training-human-toll-content-moderator-meta-openai (last accessed: August 15, 2024).

Sanchez, Gabriel R./Middlemass, Keesha 2022: Misinformation Is Eroding the Public's Confidence in Democracy, in: Brookings, 26.7.2022, https://www.brookings.edu/articles/misinformation-is-eroding-the-publics-confidence-in-democracy/ (last accessed: August 2, 2024).

Schaake, Marietje 2020: Trade Secrets Shouldn't Shield Tech Companies' Algorithms from Oversight, in: Brookings, 4.5.2020, https://www.brookings.edu/articles/trade-secrets-shouldnt-shield-tech-companies-algorithms-from-oversight/ (last accessed: August 16, 2024).

Schmitt, Vera/Tesch, Jakob/Lopez, Eva/Polzehl, Tim/Burchardt, Aljoscha/Neumann, Konstanze/Mohtaj, Salar/Möller, Sebastian 2024: Implications of Regulations on Large Generative AI Models in the Super-Election Year and the Impact on Disinformation, in: Siegert, Ingo/Choukri, Khalid (Hrsg.): Proceedings of the Workshop on Legal and Ethical Issues in Human Language Technologies @ LREC-COLING 2024, Torino, Italia, 28–38, https://aclanthology.org/2024.legal-1.6 (last accessed: August 2, 2024).

Schneier, Bruce/Farrell, Henry/Sanders, Nathan E. 2023: How Artificial Intelligence Can Aid Democracy, in: Slate, 21.4.2023, https://slate.com/technology/2023/04/ai-public-option.html (last accessed: August 2, 2024).

Seger, Elizabeth/Ovadya, Aviv/Siddarth, Divya/Garfinkel, Ben/Dafoe, Allan 2023: Democratising AI: Multiple Meanings, Goals, and Methods, in: Rossi, Francesca/Das, Sanmay/Davis, Jenny/Firth-Butterfield, Kay/John, Alex (Hrsg.): Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society, AIES '23: AAAI/ACM Conference on AI, Ethics, and Society, Montreal, QC Canada, 715–722, https://dl.acm.org/doi/10.1145/3600211.3604693 (last accessed: August 16, 2024).

Seger, Elizabeth 2023: What Do We Mean When We Talk About "AI Democratisation"?, in: Centre for the Governance of AI, 7.2.2023, https://www.governance.ai/post/what-do-we-mean-when-we-talk-about-ai-democratisation (last accessed: August 16, 2024).

Senadheera, Sajani/Yigitcanlar, Tan/Desouza, Kevin C./Mossberger, Karen/Corchado, Juan/Mehmood, Rashid/Man Li, Rita Yi/Hope Cheong, Pauline 2024: Understanding Chatbot Adoption in Local Governments: A Review and Framework, in: Journal of Urban Technology: 1–35.

Sher, Gai/Benchlouch, Ariela 2023: The Privacy Paradox with AI, in: Reuters, 31.10.2023, https://www.reuters.com/legal/legalindustry/privacy-paradox-with-ai-2023-10-31/ (last accessed: August 15, 2024).

Silver, Laura/Fagan, Moira/Huang, Christine/Clancy, Laura 2024: What Can Improve Democracy?, in: Pew Research Center, 13.3.2024, https://www.pewresearch.org/global/2024/03/13/what-can-improve-democracy/ (last accessed: August 11, 2024).

Simon, Julie/Bass, Theo/Boelman, Victoria/Mulgan, Geoff 2017: Digital Democracy. The tools transforming political engagement, in: Nesta, 20.2.2017, https://www.nesta.org.uk/report/digital-democracy-the-tools-transforming-political-engagement/ (last accessed: April 11, 2025).

Simon, Felix M./Altay, Sacha/Mercier, Hugo 2023: Misinformation Reloaded? Fears about the Impact of Generative AI on Misinformation Are Overblown, in: Harvard Kennedy School Misinformation Review, 18.10.2023, https://misinforeview.hks.harvard.edu/article/misinformation-reloaded-fears-about-the-impact-of-generative-ai-on-misinformation-are-overblown/ (last accessed: August 2, 2024).

Simon, Felix M. 2024: AI Search & Implications for Publishers and the Public Sphere, in: Medium, 16.5.2024, https://felixsimon.medium.com/ai-search-implications-for-publishers-and-the-public-sphere-db931b6c8274 (last accessed: August 15, 2024).

Sridharan, Soujanya/Girish, Stefali 2022: Data Stewards in Service of Artificial Intelligence: Reimagining Ai Futures Towards a Participatory Paradigm for Technological Innovation, in: https://papers.ssrn.com/abstract=4299355 (last accessed: August 16, 2024).

Srivastava, Swati 2023: Algorithmic Governance and the International Politics of Big Tech, in: Perspectives on Politics 21: 3, 989–1000.

Study Finds Gender and Skin-Type Bias in Commercial Artificial-Intelligence Systems 2018: MIT News, Massachusetts Institute of Technology, 11.2.2018, in: https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212 (last accessed: August 15, 2024).

Sultan, Mubashir/Scholz, Christin/van den Bos, Wouter 2023: Leaving traces behind: Using social media digital trace data to study adolescent wellbeing, in: Computers in Human Behavior Reports 10: 100281.

Suresh, Harini/Tseng, Emily/Young, Meg/Gray, Mary/Pierson, Emma/Levy, Karen 2024: Participation in the Age of Foundation Models, in: The 2024 ACM Conference on Fairness, Accountability, and Transparency, FAccT '24: The

2024 ACM Conference on Fairness, Accountability, and Transparency, Rio de Janeiro, Brazil, 1609–1621, https://dl.acm.org/doi/10.1145/3630106.3658992 (last accessed: August 13, 2024).

Tan, Michael 2021: Evaluating Deliberative Democratic Designs. Theory of Change and Citizens' Assembly Pilot in Lebanon, in: Institute for Global Prosperity, University College London, https://discovery.ucl.ac.uk/id/eprint/10129242/1/Tan_Democratic%20Designs.pdf (last accessed: April 11, 2025).

Tao, Yan/Viberg, Olga/Baker, Ryan S./Kizilec, Rene F. 2024: Cultural Bias and Cultural Alignment of Large Language Models, in: http://arxiv.org/abs/2311.14096 (last accessed: August 13, 2024).

The Gov Lab n.d.: AI Localism, in: https://ailocalism.org/# (last accessed: August 16, 2024).

Tsai, Lily L./Pentland, Alex/Braley, Alia/Chen, Nuole/Enríquez, José Ramón/Reuel, Anka 2024: Generative AI for Pro-Democracy Platforms, in: An MIT Exploration of Generative AI, 27.3.2024, https://mit-genai.pubpub.org/pub/mn45hexw/release/1 (last accessed: August 2, 2024).

Turobov, Aleksei/Coyle, Diane/Harding, Verity 2024: Using ChatGPT for thematic analysis (Working Paper), in: Bennett Institute for Public Policy, University of Cambridge, 5.2024, https://www.bennettinstitute.cam.ac.uk/wp-content/uploads/2024/05/Using-ChatGPT-for-analytics-WP.pdf (last accessed: April 11, 2025).

Ubuntu-AI 2023: Democratic and Equitable Machine Learning, in: Generativejustice, 29.9.2023, https://generativejustice.org/uai/ (last accessed: August 16, 2024).

Ulbricht, Lena 2020: Scraping the demos. Digitalization, web scraping and the democratic project, in: Democratization 27: 3, 426–442.

Urban AI n.d.: About Us, in: https://urbanai.fr/about-us/ (last accessed: August 16, 2024).

Vaccaro, Andrea 2021: Comparing Measures of Democracy: Statistical Properties, Convergence, and Interchangeability, in: European Political Science 20: 4, 666–684.

Velikanov, Cyril/Prosser, Alexander 2017: Mass Online Deliberation in Participatory Policy-Making—Part I, in: Paulin, Alois A./Anthopoulis, Leonidas G./Reddick, Christopher G. (Hrsg.): Beyond Bureaucracy. Towards Sustainable Governance Informatisation, Cham, 209–234.

Verhulst, Stefaan G./Sloane, Mona 2020: Realizing the Potential of AI Localism, in: Project Syndicate, 7.2.2020, https://www.project-syndicate.org/commentary/local-regulation-of-artificial-intelligence-uses-by-stefaan-g-verhulst-1-and-mona-sloane-2020-02 (last accessed: August 16, 2024).

Verma, Pranshu 2023: The Rise of AI Fake News Is Creating a 'Misinformation Superspreader', in: Washington Post, 17.12.2023, https://www.washingtonpost.com/technology/2023/12/17/ai-fake-news-misinformation/ (last accessed: August 2, 2024).

vTaiwan n.d.: Rethinking Democracy, in: https://info.vtaiwan.tw/ (last accessed: August 11, 2024).

Ware, Gemma 2024: Deepfakes and Disinformation Swirl Ahead of Indonesian Election – Podcast, in: The Conversation, 12.2.2024, http://theconversation.com/deepfakes-and-disinformation-swirl-ahead-of-indonesian-election-podcast-223119 (last accessed: August 2, 2024).

Wetherall-Grujić, Graham 2023: The Race to Democratise AI, in: Democracy Technologies, 25.10.2023, https://democracy-technologies.org/ai-data/the-race-to-democratise-ai-data/ (last accessed: August 16, 2024).

When AI Gets It Wrong: Addressing AI Hallucinations and Bias n.d.: MIT Sloan Teaching & Learning Technologies, in: https://mitsloanedtech.mit.edu/ai/basics/addressing-ai-hallucinations-and-bias/ (last accessed: August 13, 2024).

Whyte, Christopher 2020: Deepfake news: AI-enabled disinformation as a multi-level public policy challenge, in: Journal of Cyber Policy 5: 2, 199–217.

Xiao, Ziang/Zhou, Michelle X./Liao, Q. Vera/Mark, Gloria/Chi, Changyan/Chen, Wenxi/Yang, Huahai 2020: Tell Me About Yourself: Using an AI-Powered Chatbot to Conduct Conversational Surveys with Open-Ended Questions, in: ACM Transactions on Computer-Human Interaction 27: 3, 1–37.

York, Pete/Bamberger, Michael 2020: Measuring Results and Impact in the Age of Big Data: The Nexus of Evaluation, Analytics,and Digital Technology (Measurement and Evaluation), in: The Rockefeller Foundation, 3.2020, https://www.rockefellerfoundation.org/wp-content/uploads/Measuring-results-and-impact-in-the-age-of-big-data-by-York-and-Bamberger-March-2020.pdf (last accessed: April 11, 2025).

Zaleznik, Daniel 2021: Facebook and Genocide: How Facebook Contributed to Genocide in Myanmar and Why It Will Not Be Held Accountable, in: Harvard Law School, Systemic Justice Project, 6.2021, https://systemicjustice.org/article/facebook-and-genocide-how-facebook-contributed-to-genocide-in-myanmar-and-why-it-will-not-be-held-accountable/ (last accessed: August 2, 2024).

Zhang, He/Wu, Chuhau/Xie, Jingyi/Lyu, Yao/Cai, Jie/Carroll, John M. 2024: Redefining Qualitative Analysis in the AI Era: Utilizing ChatGPT for Efficient Thematic Analysis, in: http://arxiv.org/abs/2309.10771 (last accessed: August 13, 2024).

Zhang, Shuo/Pan, Liangming/Zhao, Junzhuo/Yang Wang, William 2024: The Knowledge Alignment Problem: Bridging Human and External Knowledge for Large Language Models, in: http://arxiv.org/abs/2305.13669 (last accessed: August 13, 2024).

Zhu, Junyan/Isaacs, Rachel 2024: Campaign Microtargeting and AI Can Jeopardize Democracy, in: LSE British Politics and Policy blog, 29.5.2024, https://blogs.lse.ac.uk/politicsandpolicy/campaign-microtargeting-and-ai-can-jeopardize-democracy/ (last accessed: August 2, 2024).

PrEval – Zukunftswerkstätten

is a research and transfer project on evaluation in extremism prevention, democracy promotion and civic education in Germany. The project is funded by the German Federal Ministry of the Interior (BMI). This expert report was written as part of the PrEval "Zukunftswerkstatt" on Evaluation Research, which is coordinated by the German Centre for Integration and Migration Research (DeZIM). Further information on the project and the partners involved can be found at https://preval.hsfk.de/en

At a Glance

The accompanying "PrEval Handreichung":
https://preval.hsfk.de/handreichung0225
or directly via the QR-Code.

Proposed citation:
Tsui, Quito/Raftree, Linda: Exploring Emerging AI as Subject and Object in Democracy-Focused Evaluation, PrEval Expertise 2/2025, Frankfurt/M., DOI: 10.48809/PrEvalExp2502.

Peace Research Institute Frankfurt (PRIF)

Baseler Straße 27–31 | 60329 Frankfurt am Main
Phone: +49 69 95 91 04-0 | E-Mail: preval@prif.org
https://www.prif.org

Cover Photo: iStock.com/borzaya

The "PrEval Expertisen" reflect the opinions of the authors

Peace Research Institute Frankfurt
Leibniz-Institut für Friedens- und Konfliktforschung

DeZIM Deutsches Zentrum für Integrations- und Migrationsforschung

With funding from the
Federal Ministry of the Interior
by decision of the German Bundestag