Artificial intelligence (AI) in action: A preliminary review of AI use for democracy support

Policy paper

Grahm Tuohy-Gaydos, September 2024

Abstract

This policy paper provides a working definition of AI for Westminster Foundation for Democracy (WFD) and the broader democracy support sector. It then provides a preliminary review of how AI is being used to enhance democratic practices worldwide, focusing on several themes including: accountability and transparency, elections, environmental democracy, inclusion, openness and participation, and women's political leadership. The paper also highlights potential risks and areas of development in the future. Finally, the paper shares five recommendations for WFD and democracy support organisations to consider advancing their 'digital democracy' agenda.

This policy paper also offers additional information regarding AI classification and other resources for identifying good practice and innovative solutions. Its findings may be relevant to WFD staff members, international development practitioners, civil society organisations, and persons interested in using emerging technologies within governmental settings.

Contents

Author and acknowledgements

The publication "Artificial Intelligence (AI) in Action: a Preliminary Review of AI Use for Democracy Support" was written by Grahm Tuohy-Gaydos (WFD Research Intern). It received peer-review comments and edits from Alex Scales (WFD Evidence and Learning Manager), Charlotte Egan (Director of WFD's Democratic Resilience in a Digital World Programme), Szelim Simandi (Head of Systems Improvement) and Alex Adegoke (Data Systems Manager).

Rights and disclaimer

Introduction

The rapid advancement of artificial intelligence (AI), and more specifically generative artificial intelligence (GenAI), has increased interest in machine learning technologies and encouraged conversations about the applications of AI across a variety of industries. This has been accompanied by a growing focus on the interrelated fields of AI safety and ethics, with the UK holding the first-ever AI Safety Summit in November 2023. Whether AI poses a threat to[1] or an opportunity for democracies and democratic actors, or a mix of both, is yet to be fully understood.

As this fast-moving landscape shifts, actionable insights concerning AI use by state institutions and democracy support organisations have remained comparatively limited. In response, this report provides:

1. A working definition of artificial intelligence for Westminster Foundation for Democracy (WFD), which may be of interest to the broader democracy support sector.
2. A preliminary literature review of AI use and early adoption practices within government and democracy support settings, and the opportunities and challenges posed by each.
3. Risks and suggestions of areas for future advancement given the current state of AI deployment.
4. Recommended next steps for WFD and the democracy support sector to consider when integrating emerging technologies like AI and GenAI into democracy support, as well as how to approach AI governance and internal use.

This policy paper forms part of WFD's Democratic Resilience in a Digital World (DRDW) Programme[2], and sits alongside other publications including WFD's Guidelines for AI in Parliaments and A Democratic Approach to Global Artificial Intelligence (AI) Safety. Together, these resources aim to offer insights to support parliaments to integrate AI into the parliamentary workspace, strengthen democratic governance of AI, and support the responsible application of AI within democracy support settings.

Defining artificial intelligence

There is no universally accepted definition of artificial intelligence.[3] This lack of definitional agreement has made both governance and clarificatory efforts a challenge for invested parties and regulators alike. Many of the core characteristics of what would normally be classified as 'AI'—

[1] Kreps, S., & Kriner, D. (2023). How AI Threatens Democracy. Journal of Democracy 34(4), 122-131. https://dx.doi.org/10.1353/jod.2023.a907693.

[2] The DRDW Programme is a pilot programme to explore how best to integrate emerging digital technology (including AI) into several different democracy support programmes in Sri Lanka, Kenya, and Bosnia and Herzegovina.

[3] Concettina Cassa et al., "Strengthening Multistakeholder Approach to Global AI Governance, Protecting the Environment and Human Rights in the Era of Generative AI," (Internet Governance Forum, October 2023), https://www.intgovforum.org/en/filedepot_download/282/26545, p. 1.

neural networks, supervised learning, and reliance on large data sets—are also descriptive of the algorithms which underline platforms like Google and Facebook.[4] This in turn poses a challenge to integratory and regulatory efforts, which as a result have turned towards novel typological methods when drafting governance on AI.

The Organisation for Economic Co-operation and Development (OECD) has led the most notable effort to define AI, establishing a multi-year working group on the subject. The OECD's definition classifies AI as:

> *'A machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.'* [5]

This definition has been widely used, including by organisations like the Internet Governance Forum (IGF)[6], the International Institute for Democracy and Electoral Assistance (International IDEA)[7], and has been referenced by multilateral organisations such as the United Nations Economic and Social Council and the United Nations Educational, Scientific, and Cultural Organisation (UNESCO).[8] It has also been accepted by 47 states which have utilised the above definition when crafting regulatory efforts.[9]

While the OECD definition is widely accepted, it remains relatively broad and captures a wide array of technologies. As such, many organisations have taken the OECD's definition as a framework to build off, considering additional factors when classifying AI.[10] These efforts have resulted in a multitude of frameworks for categorising AI across a variety of axes, which are described in greater

[4] Wolfgang Hoffmann-Riem, "Artificial Intelligence as a Challenge for Law and Regulation," in Regulating Artificial Intelligence, ed. Thomas Wischmeyer and Timo Rademacher (Cham: Springer International Publishing, 2020), 1–29, https://doi.org/10.1007/978-3-030-32361-5_1, p. 2.

[5] "Explanatory Memorandum on the Updated OECD Definition of an AI System," OECD Artificial Intelligence Papers, vol. 8, OECD Artificial Intelligence Papers, December 19, 2023, https://doi.org/10.1787/623da898-en.

[6] Cassa et al., 2023, p. 1.

[7] Prathm Juneja, Artificial Intelligence for Electoral Management (International Institute for Democracy and Electoral Assistance, 2024), https://doi.org/10.31752/idea.2024.31, p. 11.

[8] Committee of Experts on Public Administration, "Artificial Intelligence Governance to Reinforce the 2030 Agenda and Leave No One Behind" (United Nations Economic and Social Council, January 29, 2024).

[9] Organisation for Economic Co-operation and Development, "Governments That Have Committed to the AI Principles," July 2024, https://oecd.ai/en/principles.

[10] See Cassa et al., 2023, p. 1 and Council of the European Union, "Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)," Pub. L. No. ST_7536_2024_INIT (2024).

detail in Annexe A of this report. This catalogues a select number of approaches to AI classification, as well as additional background on AI systems.

A working definition of AI

In line with its broader efforts surrounding AI use, WFD should use the OECD definition of artificial intelligence with specific emphasis towards novel uses of generative and discriminative frontier models. For clarity, frontier models can be classified as, "highly capable general-purpose AI . . . [able to] perform a wide variety of tasks - as well as relevant specific narrow AI that could exhibit capabilities that cause harm - which match or exceed the capabilities present in today's most advanced models."[11] Furthermore, generative AI specifically refers to systems capable of identifying patterns in data, then generating outputs which could theoretically fit within the data. Discriminative AI, by contrast, largely analyses given data to elicit predictions or decision-making.[12] It is important to note that this definition draws heavily from multi-dimensional classificatory efforts which consider technical, social, and ethical factors when analysing AI systems. Further information on multi-modal, matrix-based frameworks can be found in Annexe A, which offers essential insights into these approaches and their contribution to the above definition. It is advised to read both this brief and Annexe A in its entirety.

> ***'WESTMINSTER FOUNDATION FOR DEMOCRACY SHOULD USE THE OECD DEFINITION OF ARTIFICIAL INTELLIGENCE, WITH SPECIFIC FOCUS ON NOVEL GENERATIVE AND DISCRIMINATIVE FRONTIER MODELS.'***

While broad, utilising the OECD definition with an emphasis towards discriminative and generative frontier models serves to sufficiently narrow the scope of analysis and focuses discussions on emerging models rather than pre-existing technologies and practices. Importantly, this should be taken as a *working definition*: this is an emerging field and additional advancements in AI models and technical approaches have the potential to rapidly affect terminological disambiguation. A working definition offers a flexible framework for continual adjustment, aligning WFD's efforts with accepted practices in the AI governance field.

---

[11] Government of United Kingdom, "The Bletchley Declaration by Countries Attending the AI Safety Summit," GOV.UK, November 2023, https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023.

[12] Committee on Digital Economy Policy, "OECD Framework for the Classification of AI Systems," OECD's Artificial Intelligence in Work, Innovation, Productivity and Skills (AI-WIPS) Programme, no. 323 (February 2022), p. 45.

Preliminary AI use cases

A variety of AI systems and technologies are already being used by state actors, democracy support organisations, and stakeholders across a variety of contexts. The following sections outline artificial intelligence use across six key areas of democracy support: accountability and transparency, elections, environmental democracy, inclusion, openness and participation, and women's political leadership[13]. Each section is further subdivided by general application: data classification, monitoring, transcription and translation, chatbot and engagement tools, and content generation. If the given thematic area includes unique applications, they have been identified under 'Additional Use Cases' at the end of the section. If there were no identifiable deployments within an AI sub-class, the section was then excluded from the report.

Accountability and transparency

AI-integrated accountability and transparency tools represent one of the most promising areas of technological deployment. Data classification, monitoring systems, and predictive modelling are core areas of current use, with applications in each setting.

Data classification

Multiple organisations and states have begun to utilise AI tools to classify data and create predictive models. An early pattern, especially within state practice, s to use machine learning models to classify and tag both qualitative and quantitative data for future use. For example, the United States National Archives and Records administration is utilising AI models to meet metadata standards for federal documents, as well as to classify and summarise documents using natural language processing.[14] Similarly, the United Nations Development Programme (UNDP) is applying AI to thematically tag UNDP projects for further organisation, as well as another program to streamline the process of completing development analytics.[15] The United Nations Office of the High Commissioner for Human Rights (UNOHCHR) is also pursuing a digital image verification and classification project. Given the challenges that prior human rights investigations have faced when collecting data, UNOHCHR is producing open-source models that will help to filter the massive influx of digital information the office receives when investigating abuses.[16] The United States Department of State is pursuing a similar program utilising machine learning and satellite imagery

[13] Note: This is a preliminary mapping effort and these categories have been interpreted broadly to permit the identification of additional use cases with possible relevance to the democracy support sector.

[14] National AI Advisory Committee, "Federal AI Use Case Inventory 2023," 2023, p. 43.

[15] International Telecommunication Union, "United Nations Activities on Artificial Intelligence (AI)" (International Telecommunication Union, 2023), p. 132.

[16] International Telecommunication Union, 2023, p. 132.

to identify and document war crimes and abuses in Ukraine, training models to recognise damage to buildings, hospitals, and other critical infrastructure.[17]

Monitoring systems

Monitoring systems have seen rapid uptake, with multiple states deploying both discriminative and generative AI. Tanzania, Brazil, and Colombia have all created platforms to monitor governmental procurement and monetary transfer for corruption, with early findings in Brazil identifying an additional 3,044 agreements that were potentially suspect.[18] Similar programmes are also in development elsewhere, and anti-corruption monitoring seems to be a clear emerging use case across a variety of states and contexts. For example, the United States Bureau of Conflict and Stabilization Operations (USBCSO) is also pursuing a similarly aligned program using the technology service Sealr to verify the delivery of foreign aid to areas in which the Bureau cannot have an on-the-ground presence. USBCSO reported promising early results with the programme.[19] Singapore-based Securade.ai is also using generative AI to perform video analytics on CCTV footage to quickly identify workplace safety violations and issue risk alerts, potentially averting health-related risks while maximising the protection of workers in high-risk fields.[20]

Chatbot and engagement tools

The utilisation of chatbots and engagement tools to support governmental transparency and accountability has been limited globally. In the United States, the National Archives and Records Administration is implementing machine learning to respond to Freedom of Information (FOI) requests, using natural language processing and data classification to identify necessary documentation while additional AI technologies will streamline the redaction process.[21] The program will ease wait times and increase the ability of the organisation to respond to FOI requests. The U.S. State of Georgia, as well as both the Chamber of Deputies and the Senate in Italy, are also utilising AI technologies to scrape legislative data and proceedings to make the information more accessible to the public and easier to understand.[22]

---

[17] National AI Advisory Committee, 2023, p. 43.

[18] International Telecommunication Union, "AI for Good-Innovate for Impact Final Report" (International Telecommunication Union, 2024), p. 34; OECD and CAF Development Bank of Latin America, *The Strategic and Responsible Use of Artificial Intelligence in the Public Sector of Latin America and the Caribbean*, OECD Public Governance Reviews (OECD, 2022), https://doi.org/10.1787/1f334543-en, pp. 48-49.

[19] National AI Advisory Committee, 2023, p. 43.

[20] International Telecommunication Union, 2024, p. 91.

[21] National AI Advisory Committee, 2023, p. 65.

[22] Colin Wood, "State CIOs Share Early Use Cases for Generative AI," *StateScoop* (blog), October 11, 2023, https://statescoop.com/state-government-generative-ai-uses/; Inter-Parliamentary Union, "Index of Parliamentary Use Cases," July 19, 2024, https://docs.google.com/document/d/1tBRg5CPiW9kBUlrsn5RBpFpaqAVuC9Bk-36w7CkDqf0/edit?usp=embed_facebook.

Elections

The impact of artificial intelligence on electoral practice has become a key concern over the past year, especially given the number of high-profile elections in 2024. Recent advancements in AI include data classification and prediction, monitoring systems, chatbot and engagement tools, and microtargeting.

Data classification

AI represents a useful tool for electoral management bodies, capable of completing many otherwise menial or repetitive tasks. International IDEA, for example, notes that AI is "particularly well suited" for voter list management, as well as records matching on behalf of voters.[23] AI systems may also be able to assist in classifying qualitative information on polling place incidents, providing additional assistance to polling watch organisations or electoral management bodies.[24]

Monitoring systems

AI is already used extensively within monitoring systems in the electoral context. Most biometric tools utilise deep learning algorithms, as do signature matching tools.[25] Similarly, AI is also being used to assist in election monitoring. The UNDP is developing iVerify, a set of open-source tools to track mal-information and hate speech on social media which can then be directly countered by local partners.[26] Private companies have also begun to develop similar technologies, with social media monitoring representing an emerging market class for political parties and other stakeholders.[27] The African Union Development Agency reports that AI has also been utilised in Kenya and South Africa to monitor elections for potential abuse.[28]

Chatbot and engagement tools

Despite growing use across the international development sector as a tool to provide populations with service-related information and discussion in popular media of utilising chatbots to provide voters with electoral information, limited use cases were identified in this study. Perhaps the clearest examples come from political parties, which have deployed such technologies in limited capacities. Indeed, research has shown positive outcomes for chatbot architectures for senior

---

[23] Juneja, 2024, p. 15.

[24] Juneja, 2024, p. 27.

[25] Juneja, 2024, p. 16.

[26] International Telecommunication Union, 2023, p. 123.

[27] Juneja, 2024, p. 23

[28] AUDA-NEPAD, "Harnessing Artificial Intelligence (AI) for Transparent Elections: A New Dawn for African Democracy," June 20, 2024, https://www.nepad.org/blog/harnessing-artificial-intelligence-ai-transparent-elections-new-dawn-african-democracy; Patrick Meier, "Artificial Intelligence for Monitoring Elections (AIME)," LinkedIn, April 14, 2015, https://www.linkedin.com/pulse/artificial-intelligence-monitoring-elections-aime-patrick-meier/.

citizens and first-time voters seeking information on electoral information, polling place locations, and beyond.[29] Another recent study tested the use of AI (specifically GPT-4 Turbo) to engage with conspiracy theorists to deliver tailored counterevidence on specific conspiracy theories, such as the belief of widespread fraud in the 2020 US Presidential Election, with researchers reporting "robust evidence" that the intervention reduced conspiracy belief by ~20%.[30]

In this general absence, it is worth noting the emergence of reports that popular general AI chatbots (including Chat GPT 3.5 and 4.0, Copilot, and Gemini) "do not seem fit for purpose to provide accurate information on electoral processes"[31] – a fact supported by other investigations.[32]

Content generation

Again, there is limited available data on using content generation to strengthen democratic systems and support efforts. Whilst there seem to be limited applications of this technology by election management bodies and democracy support actors, there is an emerging trend of harnessing AI technologies for political and campaign mobilisation, especially by political parties and outside groups or persons. However, there are also associated concerns about how this has been used across democracies. For example, GenAI played a prominent role in the Indonesian elections, where Prabowo Subianto utilised an AI-generated cartoon version of himself to soften his reputation and record of human rights abuses.[33]

Similarly, deep fakes—realistic video and audio artificially produced by an AI model—were used heavily in the 2024 Indian Elections. Deep fake usage included audio-realistic translations of candidates, commercialised use on behalf of candidates with fewer resources to quickly produce advertisements, and even the 'resurrection' of past political figures for their endorsement.[34]

Non-state and state-adjacent actors have also begun to utilise GenAI, and especially deep fakes, to impact elections. For example, in Slovakia, an AI-produced audio in which a pro-Europe candidate appeared to suggest rigging the election was released during the media blackout period, wreaking havoc in the absence of resources to disprove the deep fake.[35] Many other uses of GenAI

[29] Juenja, 2024, p. 21.

[30] Costello, T. H., Pennycook, G., & Rand, D. G. (2024, April 3). Durably reducing conspiracy beliefs through dialogues with AI. https://doi.org/10.31234/osf.io/xcwdn. A

[31] Democracy Reporting International, 2024. Are Chatbots Misinforming Us About the European Elections? Yes. https://democracy-reporting.org/en/office/global/publications/chatbot-audit

[32] Reuters Institute, 2024. How AI chatbots responded to questions about the 2024 UK election. https://reutersinstitute.politics.ox.ac.uk/news/how-ai-chatbots-responded-questions-about-2024-uk-election

[33] Kat Duffy, "AI in Context: Indonesian Elections Challenge GenAI Policies," *Council on Foreign Relations* (blog), February 13, 2024, https://www.cfr.org/blog/ai-context-indonesian-elections-challenge-genai-policies.

[34] Nilesh Christopher, "Indian Voters Are Being Bombarded With Millions of Deepfakes. Political Candidates Approve," *Wired*, May 28, 2024, https://www.wired.com/story/indian-elections-ai-deepfakes/.

[35] Daniel Atherton, "Incident Number 573," ed. Daniel Atherton, *AI Incident Database*, 2023, https://incidentdatabase.ai/cite/573.

have been noted in these above strains, and organisations such as *Wired* and the *AI Incident Database* are attempting to agglomerate cases of use and misuse within the electoral context.

Additional use cases

The practice of microtargeting is another arena for potential AI use within elections. Microtargeting, which, "involves deducing psychological attributes that are not readily observable, such as personality traits, from individuals' online behaviour and personal data . . . to craft highly personalized messages tailored to each individual,"[36] is by no means a new development, first gaining prominence in 2016. Simchon et al. have released recent findings that demonstrate that this personalised messaging can be effective, and note that the combination of microtargeting techniques with GenAI may allow campaigns and political parties to produce more specific advertising at a far greater scale than possible in prior elections.[37] The use of AI-powered microtargeting algorithms alongside GenAI also represents a potentially cheaper campaign tool for smaller parties and organisations, and can act as an equaliser given that much of the underlying technology and systems behind microtargeting, as well as certain forms of GenAI like large language models and deep fakes, are otherwise free or open-source.[38] Given its prior use and the potential upside of its uptake alongside generative technologies, microtargeting represents another area of both opportunity and concern within the electoral setting and will require additional consideration in the future.

[36] Almog Simchon, Matthew Edwards, and Stephan Lewandowsky, "The Persuasive Effects of Political Microtargeting in the Age of Generative Artificial Intelligence," *PNAS Nexus* 3, no. 2 (February 1, 2024): pgae035, https://doi.org/10.1093/pnasnexus/pgae035, p. 1.

[37] Simchon, Edwards, and Lewandowsky, 2024, pp. 2-3.

[38] Bernard Siman, "Emerging Hybrid Threats: AI And Microtargeting Disinformation As A Security Threat To The Protection Of International Forces," *Defence Horizon Journal*, October 2023, p. 68; Angela Busacca and Melchiorre Alberto Monaca, "Deepfake: Creation, Purpose, Risks," in *Innovations and Economic and Social Changes Due to Artificial Intelligence: The State of the Art*, ed. Domenico Marino and Melchiorre Alberto Monaca (Cham: Springer Nature Switzerland, 2023), 55–68, https://doi.org/10.1007/978-3-031-33461-0_6, p. 56.

Environmental democracy

Environmental protection has been at the core of many AI deployment efforts, with such tools serving to promote public sector interaction with citizens, stakeholders, and affected groups. While outside the scope of this report, there are impressive efforts underway to utilise AI in scientific settings for monitoring and predictive analytics as well as potential additional uses for AI systems and modalities as part of broader technical solutions to address the climate crisis. Within the scope of environmental democracy efforts, AI has largely been utilised for data classification, monitoring, and citizen engagement efforts, as well as part of unique user-centric environmental protection efforts.

Data classification

Use cases for data classification purposes have been varied. Many state and advocacy group efforts have surrounded the creation of tool repositories and classification models for farming, drought resistance, and other environmental factors to assist impacted communities.[39] The United States Department of Agriculture, meanwhile, is utilising natural language processing algorithms to assist in comment analysis and grouping semantically similar comments together, allowing it to better align its efforts to meet citizens' needs and redistribute costs towards other areas.[40]

Beyond classification efforts, other organisations have begun to utilise machine learning models to identify and predict vulnerabilities and forecast food and water security. These include an in-development pilot programme by USAID which relies on satellite imagery to map informal settlements often left out of urban vulnerability assessments, while the Internet Governance Forum (IGF) has utilised the Measurement Indicators for Resilience Analysis dataset to predict stressors and where vulnerabilities may arise in Southern Malawi, allowing on-the-ground organisations to better direct assistance and meet the needs of those affected.[41]

Monitoring systems

The utilisation of AI-augmented monitoring tools has been notable in environmental democracy efforts. The UN Committee of Experts on Public Administration have noted how AI models can help to, “track pollution levels, enabling local governments to alert the public of dangerous levels.”[42] The United States Environmental Protection Agency and the University of Chicago have also demonstrated how AI can improve the enforcement of environmental regulations, better ensuring

---

[39] International Telecommunication Union, 2024, p. 80.

[40] National AI Advisory Committee, 2023, p. 77.

[41] Concettina Cassa et al., “Strengthening Multistakeholder Approach to Global AI Governance, Protecting the Environment and Human Rights in the Era of Generative AI,” IGF POLICY NETWORK ON ARTIFICIAL INTELLIGENCE (Internet Governance Forum, October 2023), https://www.intgovforum.org/en/filedepot_download/282/26545, pp. 38-39.

[42] Committee of Experts on Public Administration, “Artificial Intelligence Governance to Reinforce the 2030 Agenda and Leave No One Behind” (United Nations Economic and Social Council, January 29, 2024).

that areas of concern are being identified and properly addressed.[43] In Hungary, the Global Green Growth Institute has created its Green Growth Simulation Tool, which helps to quantify the benefits and impact of Hungary's sustainable development goals and co-benefits in other areas, like social inclusion and gender.[44]

Chatbot and engagement tools

Chatbot and engagement tools represent a key nexus of AI-powered environmental democracy efforts. The Development Monitoring and Evaluation Office in India has designed an AI-powered chatbot to assist farmers, helping them, "predict the best plans, policies, and strategies to increase crop productivity and economic growth."[45] In Libya, Leapchat AI has created a chatbot to assist the families of people missing after the Derna Flooding Crisis, allowing family members to notify authorities and NGOs in the area, provide potentially important identificatory information, as well as update the status of missing individuals.[46] There are also additional potentialities surrounding chatbot use, which have been found to help improve access for vulnerable groups to an array of basic services across a variety of contexts.[47]

Additional use cases

In Brazil, the company Umgrauemeio has produced a unique technological solution which merits further discussion. Described as a "holistic, multi-stakeholder approach," to fire prevention utilising predictive modelling, local knowledge, and "community empowering technologies," the company's Pantera software system attempts to mitigate fire risk through a community-centric approach.[48] Using firetowers and computer vision, the company reports it can detect fire outbreaks within 3 minutes and then assist with firefighting operations. The system also allows brigades and other groups to provide data which the system can then analyse, potentially allowing officials to, "Increment prevention in future firefighting operations, increasing overall safety, efficiency, loss, and emissions reductions."[49] The Pantera software is already in use within Brazil, and has seen additional uptake in Portugal and India.[50]

---

[43] National AI Advisory Committee, 2023, p. 46.

[44] International Telecommunication Union, 2024, p. 15.

[45] International Telecommunication Union, 2024, p. 23.

[46] International Telecommunication Union, 2024, pp. 77-78.

[47] Committee of Experts on Public Administration, 2024, p. 6.

[48] International Telecommunication Union, 2024, p. 130.

[49] Oscar Bambini, "Umgrauemeio 1.5ºC," MIT SOLVE, accessed August 6, 2024, https://solve.mit.edu/challenges/resilient-ecosystems/solutions/47562.

[50] International Telecommunication Union, 2024, p. 131.

Inclusion

Whilst human rights concerns about AI systems exist – particularly related to the risks of discrimination, exclusion and a lack of accountability[51]– and should be noted (see risks section below for more), AI also has the potential to increase the inclusiveness and accessibility of political environments. Indeed, multiple researchers, organisations and states are already beginning to produce such systems: to date, current solutions have largely been in the realm of data classification, transcription and translation technologies, and chatbot and engagement tools. There have also been additional use cases which more broadly address specific challenges faced by certain communities.

Data classification

Accessibility, especially of government documents, has remained a challenge. The State of Colorado has reported that it has utilised AI to catalogue its databases and repost PDFs in more accessible formats.[52] Additionally, public sector organisations can utilise AI-empowered systems to identify accessibility errors in their websites and programmes, as well as complex or otherwise confusing text or resources that could be streamlined.[53] These efforts could help make government services more inclusive, and allow greater access for people with disabilities or otherwise not served as a result of these potential errors.

Transcription and translation

Transcription and translation services represent a key area of growth. AI tools are able to provide real-time translation and sign language interpretation, especially where resources have not traditionally been present, including within many government services.[54] The Swindon Borough Council has created a GenAI system to produce Easy Read documents, which, "let people with learning disabilities know what they need to, so they can make key decisions about important areas of their life."[55] The Cambodia Academy of Digital Technology is currently developing an open-source, free tool to provide translation for Khmer Braille, allowing people who are visually and sight

---

[51] Human Rights Watch, 2023. Pandora's Box: Generative AI Companies, ChatGPT, and Human Rights. https://www.hrw.org/news/2023/05/03/pandoras-box-generative-ai-companies-chatgpt-and-human-rights

[52] Colin Wood, "State CIOs Share Early Use Cases for Generative AI," *StateScoop* (blog), October 11, 2023, https://statescoop.com/state-government-generative-ai-uses/.

[53] Emily Warrender, "How AI Can Make Public Sector Services More Inclusive and Accessible," *Open Access Government* (blog), July 22, 2024, https://www.openaccessgovernment.org/how-ai-can-make-public-sector-services-more-inclusive-and-accessible/179942/.

[54] Warrender, 2024; Inclusion Scotland, "Disabled People's Rights in an Artificial Intelligence World An Overview" (Inclusion Scotland, September 2023), https://inclusionscotland.org/wp-content/uploads/2023/11/Overview-Disabled-Peoples-Rights-in-an-Artificial-Intelligence-World.pdf.

[55] Swindon Borough Council, "Council Using AI to Help People with Learning Disabilities," Swindon Borough Council, November 30, 2023, https://www.swindon.gov.uk/news/article/958/council_using_ai_to_help_people_with_learning_disabilities.

impaired to better access services, information, and educational opportunities.[56] In Brazil, Lenovo is currently developing the Libras Project, which utilises AI to translate, "[Brazilian] Sign Language (Libras) into Brazilian text and audio, and vice versa, using digital avatars," allowing seamless communication, "between individuals using sign language and those using spoken language."[57] In Estonia, the government has already produced five AI components for governmental use, including a speech recognition and synthesis tool, a text keyword extractor, and a translation engine, further increasing ease of access for citizens within the country.[58] Similarly, the Chamber of Deputies of Italy and the Chamber of Deputies of Brazil both have utilised automatic subtitling of Assembly session recordings and speech-to-text transcription to increase accessibility.[59]

Chatbot and engagement tools

Chatbots and other engagement tools represent another area of positive development. In the United States, chatbots have been used to answer common questions in an accessible and easy-to-understand manner, and have assisted millions of users.[60] The UN Committee of Experts on Public Administration has also noted how AI models could potentially promote financial inclusion, "Offering affordable and accessible banking services to disadvantaged population groups."[61] In Estonia, the state is endeavouring to create a *Bürokratt*, which will be a virtual assistant able to assist with access to all public services through a single interface. The tool will be highly accessible, offering voice-based interaction as well to ensure ease of use for all citizens.[62]

Additional use cases

In addition to the above, efforts are currently being made to increase the accessibility of AI tools themselves, especially for underrepresented communities. The SeaLLMs project is currently attempting to produce large language model applications trained for the specific nuances of regional languages in Southeast Asia that have often seen performance gaps to English-trained models, which exhibit strong linguistic bias.[63] While just one of many efforts, the SeaLLMs project points to the need for more holistic and inclusively-minded AI systems.

---

[56] International Telecommunication Union, 2024, p. 12.

[57] International Telecommunication Union, 2024, p. 46.

[58] Martin Ebers and Paloma Krõõt Tupay, *Artificial Intelligence and Machine Learning Powered Public Service Delivery in Estonia: Opportunities and Legal Challenges*, vol. 2, Data Science, Machine Intelligence, and Law (Cham: Springer International Publishing, 2023), https://doi.org/10.1007/978-3-031-19667-6, p. 17.

[59] Inter-Parliamentary Union, "Index of Parliamentary Use Cases," July 19, 2024, https://docs.google.com/document/d/1tBRg5CPiW9kBUlrsn5RBpFpaqAVuC9Bk-36w7CkDqf0/edit?usp=embed_facebook.

[60] National AI Advisory Committee, 2023, p. 46.

[61] Committee of Experts on Public Administration, 2024, p. 7.

[62] Ebers and Tupay, 2023, p. 18.

[63] International Telecommunication Union, 2024, p. 49.

Openness and participation

How artificial intelligence could potentially increase the openness of government and citizen participation has been a long-standing discussion within the GovTech field. As such, AI technologies have been utilised across nearly every axis, with a multitude of data classification tools, monitoring systems, transcription and translation models, chatbots, and content generation use cases. There have also been additional efforts which represent unique advancements and areas for continued development.

Data classification

AI has already been utilised for a host of data classification purposes that can serve to increase openness and political participation. The United Nations has piloted an AI-powered polling system in Libya and the Middle East to facilitate dialogue through 'digital focus groups' in which participants answer both multiple-choice and open-ended questions, providing their thoughts on political matters. The system then agglomerates these surveys and identifies common sentiments, which are then assessed again by other participants, producing a set of viewpoints and queries broadly shared within the sample. In Libya, the results of these digital focus groups were then posed on live television to candidates for the Government of National Unity. Masood Alavi et al. note that these digital dialogues and outreach efforts, "appeared to provide some sort of legitimacy [to the Government of National Unity], which it had lacked just a few months earlier."[64]

From a similar perspective, CitizenLab is an AI tool utilising machine-learning algorithms to process and analyse citizen contributions and ideas, providing easy-to-understand dashboards and insights for public servants.[65] Of particular interest was its use in the Youth for Climate Belgium movement, which utilised the tool to collect ideas and thoughts surrounding climate change. The programme reported positive results, producing over 700 ideas and thousands of comments and votes on initiatives, which were then analysed, clustered, and reported to elected officials as 16 core policy recommendations.[66] A similar tool was utilised to track, identify, and provide insights into bills concerning voting rights in the U.S. state of Georgia, where America Votes Georgia (AVG) deployed Plural Policy's AI system to identify relevant bills, tag and keep track of them, and update coalition members on any governmental action.[67] The system was critical to its efforts and helped

---

[64] Daanish Masood Alavi et al., "Using Artificial Intelligence for Peacebuilding," *Journal of Peacebuilding & Development* 17, no. 2 (August 2022): 239–43, https://doi.org/10.1177/15423166221102757, pp. 240-241.

[65] Jamie Berryhill et al., "Hello, World: Artificial Intelligence and Its Use in the Public Sector," OECD Working Papers on Public Governance, vol. 36, OECD Working Papers on Public Governance, November 21, 2019, https://doi.org/10.1787/726fd39d-en, p. 141.

[66] Berryhill et al., 2019, p. 141.

[67] Plural Policy, "How Plural Helped a Coalition of 39 Organizations Stop 50+ Bills Threatening Voting Rights in Georgia," Plural Policy, May 12, 2023, https://pluralpolicy.com/case-studies/how-plural-helped-a-coalition-of-39-organizations-stop-50-bills-threatening-voting-rights-in-georgia/.

AVG block over 70% of the anti-voting rights bills that legislators were attempting to pass in the state.[68]

Monitoring systems

Monitoring systems represent another area of considerable advancement. The Inter-Parliamentary Union notes that the Brazilian Chamber of Deputies has utilised AI to identify and categorise arguments for and against bills based on citizens' comments, polling, and other resources.[69] In Estonia, the Estonian Unemployment Insurance Fund, which is a quasi-governmental organisation, utilises AI to support staff, calculating "the probability of moving into employment for unemployed persons and the probability of becoming unemployed again for people that got a new job."[70] The system allows consultants to better tailor their efforts to underserved communities and prioritise certain clients who the Fund otherwise would have missed or been unable to serve, and has contributed positively to the organisation's work.

Other organisations have also utilised AI to monitor broader state trends and identify areas for concern. The University of Pennsylvania Machine Learning for Peace (MLP) programme uses AI and analytics to identify and forecast civic activity and major political events across the world. The MLP's Civic Space Early Warning System works by "continuously scraping and processing tens of millions of articles published by more than 300 local, regional, and international news sources in nearly 40 languages . . . [providing] up-to-date data on recent and historical trends in civic space and foreign influence . . . that learn from historical patterns to predict how conditions are likely to change in the near future."[71] The MLP Lab has had substantive success in its efforts, providing important insights and helping stakeholders plan around increased civic activity in their respective region or state.

Transcription and translation

Transcription and translation efforts represent another area of continual development. The Italian Chamber of Deputies, the Senate of Italy, and the Brazilian Chamber of Deputies are all utilising AI to help produce audio-to-text transcriptions of parliamentary proceedings for both public sector use and public accessibility, combining these tools with natural language processing search tools to allow for greater openness.[72] The Chamber of Deputies of Italy and Senate of Italy also use AI to produce real-time subtitles for live proceedings.[73] The European Commission's Joint Research

---

[68] Plural Policy, 2023.

[69] Inter-Parliamentary Union, 2024.

[70] Joint Research Centre., *AI Watch: European Landscape on the Use of Artificial Intelligence by the Public Sector. Annex II, Case Studies Description.* (LU: Publications Office, 2022), https://data.europa.eu/doi/10.2760/481674, p. 13.

[71] Machine Learning for Peace Development Lab, "The Machine Learning for Peace Project," 2024, https://web.sas.upenn.edu/mlp-devlab/.

[72] Inter-Parliamentary Union, 2024.

[73] Inter-Parliamentary Union, 2024.

Centre notes that beyond the above use cases, nearly 24% of AI applications in the EU utilise natural language processing techniques to further ensure openness and participation.[74]

Chatbot and engagement tools

Since the early 2020s, governments and other organisations have actively sought to deploy chatbot and engagement tools to assist the public sector. In Italy, Naccari Carlizzi et al. have begun testing the ARCHIMEDE platform, which provides certified editorial content, immediate insights into political subjects, clear event reporting, opportunities for civic participation and the exchange of ideas, as well as petitions and voting opportunities. The platform has been found to provide productive opportunities for communities, utilising technology to maximise democratic outcomes and elicit conversation, debate, and public participation in a safe atmosphere.[75] In a similar vein, the Querido Diário project in Brazil is utilising AI to, "classify, contextualise, and expand the information contained in Brazilian official newspapers," as well as make them more accessible.[76]

Many chatbot and engagement tools seek to make access to government information and services easier and more accessible, thereby promoting greater engagement. The U-Ask programme in the United Arab Emirates utilises AI to solve inefficiencies in access to services, while the State Government of Alagoas, Brazil has deployed Jaque, who assists users and can guide them through all information in the State's 'services guide', which is a catalogue of all the public services the State offers.[77] Jaque has been so successful that the State Government of Alagoas is planning to expand Jaque's presence across all of its websites and even social media to further assist its citizens and promote greater access to its services.[78]

A final use case in this subsection is for mediation. The United Nations Department of Political and Peacebuilding Affairs and Department of Peace Operations (UNDPPA) is currently exploring the use of AI for mediators and stakeholders to receive input from invested parties and affected groups, with the ability "to hold real-time consultations with a large group of individuals in local dialects and language, [and] allowing for analyses and segmentation based on demographic interests.[79] The Project has been piloted in Yemen, as part of the Libyan Political Dialogue Forum, and as part of the UN Assistance Mission for Iraq. The UNDPPA Innovation Cell has also built five

---

[74] European Commission. Joint Research Centre., *AI Watch: European Landscape on the Use of Artificial Intelligence by the Public Sector.* (LU: Publications Office, 2022), https://data.europa.eu/doi/10.2760/39336.

[75] Demetrio Naccari Carlizzi et al., "Decision Making and E-Democracy: Archimede, a Tool to Support the Data-Driven Process," in *Innovations and Economic and Social Changes Due to Artificial Intelligence: The State of the Art*, ed. Domenico Marino and Melchiorre Alberto Monaca (Cham: Springer Nature Switzerland, 2023), 39–53, https://doi.org/10.1007/978-3-031-33461-0_5, p. 51.

[76] OECD and CAF Development Bank of Latin America, *The Strategic and Responsible Use of Artificial Intelligence in the Public Sector of Latin America and the Caribbean*, OECD Public Governance Reviews (OECD, 2022), https://doi.org/10.1787/1f334543-en, p. 40.

[77] International Telecommunication Union, p. 39; OECD and CAF Development Bank of Latin America, 2022, p. 38.

[78] OECD and CAF Development Bank of Latin America, 2022, p. 38.

[79] International Telecommunication Union, 2023, p. 149.

dialect dictionaries for Yemeni, Libyan, Iraq, Palestinian, and Sudanese Arabic, allowing the Department to better understand each and represent said groups within its efforts.[80]

Content generation

While concerns around deep fakes in the electoral context was discussed earlier in this paper, there has been some discussion surrounding a separate category of AI-generated imagery referred to as 'soft fakes' in which the person within the generated video or audio has explicitly chosen to have their image and likeness utilised in the given manner. Busacca and Monaca describe this type of content as a soft fake, "because the representation of the storyteller will be a fake, but the content of storytelling will be perfectly true and match history and literature."[81] Within the context of Indian elections, for example, soft fakes were utilised to translate candidates' words into additional languages or dialects that the candidate does not speak, allowing for greater interaction with traditionally underrepresented groups and citizens.[82] There is some possibility that this practice may follow such candidates into office and become a manner by which to bring a greater portion of the citizenry into politics, increasing their presence in democratic processes.

While not a singular example, AI use is already widespread within many governments, which have begun utilising commercially available AI tools like ChatGPT as part of their efforts. In a survey of AI use in government in the United Kingdom, 32% of those queried utilised AI in their work, a number which is likely to continue to grow.[83] The tasks that AI are already being used for vary widely, but many augment existing efforts and are meant to cut back on repetitive tasks. In the U.S. State of New Hampshire, for example, GenAI is already being used to write job descriptions for positions in government, simplifying the process and allowing staff to focus greater resources on identifying strong candidates.[84]

Additional use cases

Two additional use cases are of interest: The PretorIA system in Colombia and AuroraAI in Finland. The PretorIA system concerns Colombia's Acción de Tutela, an instrument in the country which allows citizens to seek protection, "against any violation of fundamental rights resulting from the act or omission of a public authority or individual."[85] Unfortunately, the Tutela system has overburdened the Constitutional Court, which must select from thousands of Tutelas it receives each day to set legal precedents. The PretorIA system analyses complaints, reviews the Tutelas for predefined criteria, and produces reports and statistics based on information in the document.

---

[80] International Telecommunication Union, 2023, pp. 149-150.

[81] Busacca and Monaca, 2023, p. 61.

[82] Christopher, 2024.

[83] Jonathan Bright et al., "Generative AI Is Already Widespread in the Public Sector" (arXiv, January 2, 2024), http://arxiv.org/abs/2401.01291, p. 4.

[84] Wood, 2023.

[85] OECD and CAF Development Bank of Latin America, 2022, p. 37.

The information is then forwarded to a judge, who can utilise the information that the PretorIA system provides in addition to the original request. The version launched in mid-2020 in response to backlash from civil society groups utilises symbolic programming and topic modelling in place of traditional machine learning modalities, ensuring that the system is fully explainable and limiting the 'black box' concern surrounding neural networks. As noted by its developer, "The Laboratory of Innovation and Artificial Intelligence of the Faculty of Law of the University of Buenos Aires (IALAB) . . . [it is] the first predictive AI system to be used in a high court in the world."[86]

AuroraAI represents a broader approach. The Finnish Government views AuroraAI as a holistic platform which will deliver services based on needs and life-events-based planning. For example, an early experimental application of the system focused on the life event of 'moving to a place of study'. Based on surveys of student populations, the government then was able to cluster students into groups by different levels and types of support, meeting students' needs and identifying further steps that each city respectively could take.[87] The Finnish Government is planning to continue expanding on its efforts through the next three years, and hopes to have a more substantive system in place by 2027 based on its early testing. Singapore is also pursuing a similar model of incorporating AI as part of a platform with its Open Digital Platform, pointing to the possibility of future 'AI as a comprehensive platform'-based approach to digitalisation.[88]

[86] OECD and CAF Development Bank of Latin America, 2022, p. 37.

[87] Maciej Kuziemski and Gianluca Misuraca, "AI Governance in the Public Sector: Three Tales from the Frontiers of Automated Decision-Making in Democratic Settings," *Telecommunications Policy* 44, no. 6 (July 2020): 101976, https://doi.org/10.1016/j.telpol.2020.101976, p. 37; "The AuroraAI: A Human-Centric and Life-Event Based Public Sector Transformation," Observatory of Public Sector Innovation (blog), accessed August 6, 2024, https://oecd-opsi.org/innovations/auroraai/.

[88] "Open Digital Platform (ODP)," Observatory of Public Sector Innovation (blog), accessed August 6, 2024, https://oecd-opsi.org/innovations/open-digital-platform-odp/.

Women's political leadership

Artificial intelligence is already being utilised to support women's political leadership, especially in efforts to counter sexism and misogyny on online platforms. Early efforts have largely centred on data classification, monitoring systems, and chatbot and engagement tools.

Data classification

Data classification represents a key area of development for AI tools supporting women's political leadership. These efforts have largely centred on tracking and tagging sexist text and image content on social media. An example of these efforts is the UN Women's Latin American Countries AI Project, which utilises a trained model which can identify sexist or abusive language online that contributes to or perpetuates harmful stereotypes. By tracking this use, the project hopes to be able to produce substantive interventions that will be able to specifically target these stereotypes and cases of sexism and contribute to the creation of a more productive online space.[89] More broadly, Cassa et al. note that many AI systems can potentially be designed to reduce bias or otherwise support women's empowerment efforts within the democratic setting.[90]

Monitoring systems

There has been considerable discussion about the potential uses of AI monitoring tools to track, block, report, and delete sexist messaging or commentary on social media. MarvelousAI's StoryArc, for example, helps campaigns to, "identify and push back against sexist framing on social media and take control of their own narratives much more readily."[91] Early findings have demonstrated that these types of monitoring systems can be useful in quickly and summarily countering such messages on social media, allowing campaigns to remove comments where possible and identify and challenge negative stereotypes more broadly.[92]

Other monitoring systems have taken a broader view. In Brazil, a team at the newspaper AzMina produced the Political Misogynistic Discourse Monitor, which, "Monitored attacks on social media on women candidates for the municipal elections in Brazil."[93] The UNDP Data Futures Exchange,

---

[89] International Telecommunication Union, 2023, p. 227.

[90] Concettina Cassa et al., "Strengthening Multistakeholder Approach to Global AI Governance, Protecting the Environment and Human Rights in the Era of Generative AI," IGF POLICY NETWORK ON ARTIFICIAL INTELLIGENCE (Internet Governance Forum, October 2023), https://www.intgovforum.org/en/filedepot_download/282/26545, p. 30.

[91] Sarah Oates et al., "Running While Female: Using AI to Track How Twitter Commentary Disadvantages Women in the 2020 U.S. Primaries," SSRN Scholarly Paper (Rochester, NY, August 28, 2019), https://doi.org/10.2139/ssrn.3444200, p. 1.

[92] Oates et al., 2023, p. 27.

[93] Sabrina Argoub, "The Need to Represent: How AI Can Help Counter Gender Disparity in the News," *Polis* (blog), March 16, 2022, https://blogs.lse.ac.uk/polis/2022/03/16/the-need-to-represent-how-ai-can-help-counter-gender-disparity-in-the-news/.

meanwhile, has produced the Gender Social Media Monitoring Tool, which similarly tracks conversations across social media networks to help guide gender-responsive policymaking efforts. The tool supports over 100 languages and has already been utilised in a variety of contexts.[94]

Chatbot and engagement tools

While the use of chatbots to support women's political leadership has been limited, there have been some efforts to utilise engagement tools and trained models to reduce inequality, and especially to promote financial independence. For example, Women's World Banking and Mujer Financiera have created a tool which promotes financial inclusion for women in Latin America, as well as to help support personal finance management.[95] While limited, exploratory efforts like these point towards real possibilities regarding machine learning use for empowerment, and multiple sources recognise the potentiality of AI-augmented tools and approaches for such efforts.

[94] International Telecommunication Union, 2023, p. 135; Data Futures Exchange, "Gender Social Media Monitoring," UNDP, July 29, 2022, https://data.undp.org/insights/gender-social-media-monitoring.

[95] Cassa et al., 2023, p. 30.

Key findings, risks and the future of AI use in democratic contexts

Below, this paper offers an overview of key findings emerging from the review of the AI applications described in the prior section. It also provides insight on the potential risks of these systems, as well as areas for future advancements. Given the fast-moving pace of AI development, it is important to recognise that both risks and use cases are likely to shift quickly and may not follow a direct path. Additionally, emphasis on future AI advancements tends to obfuscate the current use of AI systems and can serve to redirect the conversation from the immediate impacts of artificial intelligence.

Key findings

- **AI use is already widespread.** From publicly available tools like ChatGPT and Microsoft Copilot to bespoke solutions, artificial intelligence is already being developed, piloted, and deployed. While the complexity of these use cases may differ, organisations, companies, and states are continuing to invest in AI-based tools. In broad democratic settings, applications to date have been centred on data classification, monitoring systems, transcription and translation, content generation, and chatbot and engagement tools.
- **AI has the potential to positively contribute to democracy.** While still a developing arena, AI-augmented solutions have shown varying degrees of promise when used to promote democratic practices. The range of use cases identified and the promising results reported reflect a widespread belief in the potential of AI to support positive change.
- **Commercial and private sector companies are key contributors.** While the public sector has spearheaded some efforts, private companies have taken a leading role in the development of AI tools on behalf of governments. There is a burgeoning industry of AI-centric solutions which will likely continue to expand, with inherent risks emerging because of this commercialisation.
- **There remain no clearly defined best practices.** While there are some similarities across each use case, states and organisations have yet to develop a common set of best practices beyond general ethical guidelines. There remains limited information on the feasibility of certain solutions or the specific considerations behind them, especially when AI is applied to solve highly specific or individualised challenges.
- **Information-sharing remains limited.** There remains a critical lack of publicly available information on AI use, algorithmic design, the data used to train models, and the results/impact of deployment efforts. This problem is especially notable among non-governmental organisations, many of which do not clearly identify their use of AI tools. While governments are often more open about their use of artificial intelligence by comparison, information remains restricted, thereby also constraining effective cross-sector learning.

- **There remain considerable risks in AI systems.** While early evidence suggests AI can contribute to democracy support efforts, AI (and particularly GenAI) has the potential to cause considerable harm, even when used with positive intentions. As such, further consideration of AI risks is necessary, especially as part of AI governance and regulatory efforts. Similarly, as knowledge of the best ways to mitigate these risks grows across the democracy support sector, there may need to be new ways to share this learning effectively.

Cross-thematic risks

While there are impacts within each specific thematic setting that require consideration and are of key importance, many risks often identified in AI models are cross-contextual. As such, it is wiser to consider these potential impacts more broadly.

Systemic bias

Much of the current literature on AI centres on the potentially inherent bias present in the data-driven machine-learning models which power a majority of modern AI systems. Studies have found that AI are often significantly biased against people of colour,[96] people with disabilities,[97] and women.[98] Importantly, bias can arise at multiple stages through the development process. Ntoutsi et al. identify how bias can arise in the data generation and collection stage, where the representativeness of the data or the implicit inclusion of institutional biases can affect outcomes; or in the programming stage, where the weighting of variables or training process can similarly produce bias.[99] Ferrara has described seven forms of biases: sampling bias, algorithmic bias, representation bias, confirmation bias, measurement bias, interaction bias, and generative bias.[100] These discriminatory patterns can have outsized effects and pose an existential threat to any application of AI, and as such have been recognised as key concerns when developing artificial intelligence of any kind.

Key industry figures have also noted how such biases can also lead to AI applications producing content that does not align with the sociocultural expectations of specific groups or cultures. For example, the Rt Hon Nick Clegg[101] (President of Global Affairs at Meta) articulated that a key challenge for Meta and other leading AI companies is how to train models to account for

---

[96] Eirini Ntoutsi et al., "Bias in Data-Driven Artificial Intelligence Systems—An Introductory Survey," *WIREs Data Mining and Knowledge Discovery* 10, no. 3 (2020): e1356, https://doi.org/10.1002/widm.1356, p. 1

[97] Inclusion Scotland, 2023, p. 13.

[98] Ayesha Nadeem, Olivera Marjanovic, and Babak Abedin, "Gender Bias in AI-Based Decision-Making Systems: A Systematic Literature Review," *Australasian Journal of Information Systems* 26 (December 21, 2022), https://doi.org/10.3127/ajis.v26i0.3835, p. 1.

[99] Ntoutsi et al., 2020, p. 3.

[100] Emilio Ferrara, "Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies," *Sci* 6, no. 1 (March 2024): 3, https://doi.org/10.3390/sci6010003, pp. 3-4.

[101] For more, see: In conversation with Nick Clegg: Can democracy survive the pace of technology? (chathamhouse.org)

sociocultural complexities spanning diverse expectations in attitudes, behaviours, norms and language. Especially so, given existing models' disproportionate reliance on American content for training, coupled with stringent data protection regulations (particularly in the UK and EU) that pose regulatory barriers to efforts to diversify the content used for model training. In general use cases, this may mean AI systems produce sub-optimal outputs relating to differences in language use and interpretation. However, these risks are amplified if AI is used in potentially sensitive settings like democracy support. For instance, this could lead to culturally inappropriate outputs that undermine trust in any system and its associated actors. Whilst it is beyond the scope of this paper to address the risk of systemic bias further, these risks are important to acknowledge here.

The explainability problem

Many of today's most popular AI tools, like OpenAI's ChatGPT or Google's Gemini, utilise machine learning techniques to produce outputs. Importantly, many of these popular approaches to AI development are 'black boxes'. Rudin and Radin note that since, "These black box models are created directly from data by an algorithm . . . [we] cannot understand how variables are being combined to make predictions.[102] While there are explainable AI solutions which do not utilise approaches which produce black boxes, many systems still rely on these methodologies since they represent the most advanced and technologically powerful solutions. The challenge with these black box models is that they produce an explainability problem: without an understanding of how the system came to a specific result, it is challenging to adjust the system to increase its effectiveness or mitigate other risks. The explainability problem can as a result serve to compound bias and decrease trust in outputs, all of which pose a critical threat. Especially so in democratic support settings, where political trust and perceived fairness are often a prerequisite.

Digital divides and corporate overreach

The emerging technologies literature has noted three levels of digital divides: the first refers to unequal access to the internet and digital technologies, the second to inequalities in skills and use of digital technologies, and the third to inequalities in the ability to transfer these skills into favourable offline outcomes.[103] While the intersection between these digital divides and AI use is still an emerging subject within academic literature, differences in access remain a key concern surrounding AI technologies, which have largely been studied and deployed in the Global North and emerging economies. This relative inequality has the potential to centralise authority over AI development and sideline concerns or considerations about these tools, which could in turn further increase the possibility of divergent outcomes from these systems. Furthermore, since many of the examples outlined throughout this report were developed by private companies, there remains a risk that AI tools and the technologies that underlie them may become overly commercialised or

[102] Cynthia Rudin and Joanna Radin, "Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From an Explainable AI Competition," *Harvard Data Science Review* 1, no. 2 (November 1, 2019), https://doi.org/10.1162/99608f92.5a8a3a3d, p. 3.

[103] Christoph Lutz, "Digital Inequalities in the Age of Artificial Intelligence and Big Data," *Human Behavior and Emerging Technologies* 1, no. 2 (2019): 141–48, https://doi.org/10.1002/hbe2.140, pp. 142-144.

increase corporate presence in democratic processes, which in turn introduces additional risks. As such, considerations surrounding digital divides and corporate overreach must remain a key area for further analysis given the outsized risks each presents.

Accountability and transparency

Risks

While the risks described in the above section are all highly applicable to discussions of accountability and transparency, there remain additional risks which are highly creditable. More than simply ‘tools for good’, AI technologies have seen rapid uptake in authoritarian states. China, for example, has utilised AI to augment its surveillance capacities, as well as part of predictive policing regimes.[104] AI technologies have also been noted as capable of reinforcing repression in other authoritarian regimes.[105] Critically, even within democratic contexts, AI systems pose potential dangers to democratic values, obfuscating government practice and exerting greater control over citizens’ lives. This is especially concerning in the context of broader platform-based approaches to AI development which many states are pursuing and could serve to inadvertently increase the surveillance power of government actors. AI systems can also introduce a responsibility gap in which governments can transfer blame to a model’s failure, thereby decreasing its accountability for adverse effects.[106]

Future use

Future uses of AI technologies for accountability and transparency purposes will likely remain consistent with current developments, largely supporting evidence-gathering and streamlining efforts which can support government transparency. Given the early development of monitoring systems for conflict zones in Ukraine and similar uses of such systems in other thematic areas, there is likely room for additional advancement in monitoring systems both inside government and by democracy support organisations, who can both help states pursue bespoke solutions to transparency or accountability issues or internally utilise AI systems to identify potential areas of concern within state practice.

[104] Jinghan Zeng, “Artificial Intelligence and China’s Authoritarian Governance,” *International Affairs* 96, no. 6 (November 1, 2020): 1441–59, https://doi.org/10.1093/ia/iiaa172, p. 1452.

[105] Grant Baker, “The Repressive Power of Artificial Intelligence,” Freedom on the Net (Freedom House, 2023), https://freedomhouse.org/sites/default/files/2023-11/FOTN2023Final.pdf, p. 4.

[106] Andreas Matthias, “The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata,” *Ethics and Information Technology* 6, no. 3 (2004): 175–83, https://doi.org/10.1007/s10676-004-3422-1, p. 177.

Elections

Risks

In the electoral context, generative AI use is already producing grave threats to electoral practice. States have begun utilising publicly available tools like ChatGPT and open source technologies like deep fakes to disrupt elections in other countries, while candidates themselves have utilised AI in ethically dubious manners.[107] Early studies have reported that political messaging generated by ChatGPT-4 could increase issue stance support by up to 12%,[108] while Juneja and McBride have noted that generative AI has the potential to hyper-focus political messaging and argumentation.[109] The recent 2024 Indonesian Elections offer a clear case study for many of these concerns, where Shidiq et al. have noted that, "false information, secretive computational propaganda campaigns and unchecked digital advertising [have] increasingly undermine[d] election integrity."[110] Multiple campaigns launched AI-powered digital platforms, and artificial intelligence use was widespread across the entire political spectrum.[111] AI has even affected political practice outside of the context of generative content; when a video clip of a candidate in India arguing contentious viewpoints emerged on social media, the campaign quickly moved to counter backlash by claiming the clip was a deep fake even though it was not.[112] This points to the second-order risks surrounding AI use; more than simply increasing mal-information in the electoral sphere, GenAI has corrosive potential. By fostering an environment of distrust and allowing political actors to avoid responsibility for potentially relevant actions or words, AI systems can contribute to the erosion of institutional and social firebreaks *even when GenAI is not in use*. When it is nearly impossible to tell what is real or generated, citizens are left with little insight into what is truthful. This represents a grave risk

---

[107] "STOIC Hits India with 'Zero Zeno': Israeli Firm Tries to Disrupt Lok Sabha Elections; Pushed Anti-BJP, pro-Congress Content," *The Economic Times*, June 1, 2024, https://economictimes.indiatimes.com/news/elections/lok-sabha/india/stoic-hits-india-with-zero-zeno-israeli-firm-tries-to-disrupt-lok-sabha-elections-pushed-anti-bjp-pro-congress-content/articleshow/110611373.cms?from=mdr; Morgan Meaker, "Russia Is Targeting Germany With Fake Information as Europe Votes," *Wired*, accessed July 1, 2024, https://www.wired.com/story/european-union-elections-russia-germany-disinformation-campaigns/; Marianna Spring, "Sadiq Khan Says Fake AI Audio of Him Nearly Led to Serious Disorder," *BBC News*, February 13, 2024, sec. UK, https://www.bbc.com/news/uk-68146053; Morgan Meaker, "Slovakia's Election Deepfakes Show AI Is a Danger to Democracy," *Wired*, accessed July 2, 2024, https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/.

[108] Kobi Hackenburg and Helen Margetts, "Evaluating the Persuasive Influence of Political Microtargeting with Large Language Models," *Proceedings of the National Academy of Sciences* 121, no. 24 (June 11, 2024): e2403116121, https://doi.org/10.1073/pnas.2403116121, p. 1.

[109] Prathm Juneja and Keegan McBride, "How Data and Artificial Intelligence Are Actually Transforming American Elections," *Oxford Internet Institute* (blog), accessed July 17, 2024, https://www.oii.ox.ac.uk/news-events/how-data-and-artificial-intelligence-are-actually-transforming-american-elections.

[110] Rizal Shidiq, Diyi Liu, and Justin Yeung, "Zooming in on the Digital Aspects of the Indonesian Elections 2024," *Oxford Internet Institute* (blog), February 9, 2024, https://www.oii.ox.ac.uk/news-events/zooming-in-on-the-digital-aspects-of-the-indonesian-elections-2024.

[111] Shidiq et al., 2024.

[112] Christopher, 2024.

to core elements of democratic governance, especially if citizens lose faith in their ability to discern accurate information.

Further insight into the potential impacts of AI in electoral contexts is an absolute necessity and is heavily advised given the number of upcoming elections worldwide. Additionally, greater discussion is needed about how to counter these efforts both in the short-term and long-term.

Future use

AI is likely to see continued use by both campaigns and electoral management bodies, for both positive and negative purposes. AI-produced microtargeting campaigns, while largely limited in scale, represent a key area of possible development in the future and one of considerable concern. Similarly, it is likely that states will continue to utilise AI as part of mal-information efforts, whether as 'agents' on social media, to produce content, or to identify and enflame trends that could otherwise impact electoral practices. Regardless, AI use in elections will require continued review and consideration, as well as the pursuit of actionable methodologies by which to counter latent risks that could potentially emerge.

Environmental democracy

Risks

While the risks of artificial intelligence in the context of environmental democracy are a concern, it is important to recognise the global impact of AI systems on the climate. AI requires large amounts of computing power, especially for large models like ChatGPT. Early studies have found that training models can require as much as 300 metric tonnes of CO_2, the equivalent of what 65 cars produce in a year.[113] While there is considerable enthusiasm about the potential benefits of AI as part of climate change mitigation efforts,[114] it is important to recognise the impact that machine learning models can potentially have on the very environment they are attempting to protect.

Future use

As noted above, there has been considerable discussion about the utilisation of AI to help limit current emissions and potentially mitigate further emissions in the future. While discussed to a limited degree within this paper, there are a variety of efforts to utilise AI to help combat food insecurity and contribute to better farming practices which are less impactful on the environment. Furthermore, there is considerable room for advancement in AI-driven monitoring systems for both governments and polluters, many of which could help take the burden off human inspectors and help ensure continual oversight. There has also been considerable discussion of the algorithmic

---

[113] John Naughton, "Why AI Is a Disaster for the Climate," *The Observer*, December 23, 2023, sec. Opinion, https://www.theguardian.com/commentisfree/2023/dec/23/ai-chat-gpt-environmental-impact-energy-carbon-intensive-technology.

[114] Payal Dhar, "The Carbon Impact of Artificial Intelligence," *Nature Machine Intelligence* 2, no. 8 (August 1, 2020): 423–25, https://doi.org/10.1038/s42256-020-0219-9, p. 3.

basis of carbon capture and other engineering-based solutions, many of which already rely on AI or additional digital technologies.

Inclusion

Risks

While noted earlier within this section, it is important to again emphasise the inherent bias which many AI systems can exhibit. Artificial intelligence and machine learning models have continually and consistently been found to display bias against people of colour, people with disabilities, and members of the LGBTQ+ community, and while companies have made efforts to combat these in-built prejudices, many have unfortunately ended in failure. This is in part because the training data which many systems rely on is in itself biased, which in turn causes AI models to perpetuate and sometimes enflame stereotypes and misconceptions. This represents a key concern, especially since many companies have chosen to proceed ahead even with the knowledge that these biases may exist. Continued consideration and analysis of bias in AI systems is therefore an absolute necessity and should likely be considered as part of ethical analyses of AI and machine learning models.

Future use

While the inherent bias of AI systems is a key concern, artificial intelligence does have the potential to positively contribute to efforts to promote inclusivity and accessibility in the democratic sphere. AI tools have the potential to help quickly convert documents and proceedings into easily readable, viewable, or listenable formats, as well as generate content into additional formats. There is likely to be considerable advancement within this arena in the foreseeable future, and multiple organisations are already beginning to pursue new and innovative use cases for artificial intelligence that have transformative potential.

Openness and participation

Risks

Many of the risks described in the prior section concerning accountability and transparency also hold true for considerations of openness and participation. Similarly, bias and digital divides, especially the second (inequalities in skills and use of digital technologies) and third (transferring these skills into favourable offline outcomes) order digital divides, pose additional concerns for openness and participation. Importantly, the openness and availability of information on early pilots of AI tools, as well as the data sets and programming behind them, remains limited. This is especially prevalent where state or non-state actors utilise a commercially-produced program. While the publication of some of these may not be an option, especially if the release of the training data may reveal otherwise private information, the fact remains that many of the tools described in this report remain opaque with limited systems information publicly available. There is considerable irony in the fact that information about systems directly meant to promote openness and participation is not easily accessible, but this remains a continued risk as states and organisations pursue new use cases for AI technologies.

Future use

As the current use cases in this thematic area show, AI technologies have real potential to contribute to the openness of democratic government and promote participation. Additionally, there is reason to believe that AI, and especially GenAI, has the potential to help lower barriers to entry, offering stakeholders and citizens tools by which to quickly produce, distribute, and gain insight into issues of importance. Future developments will likely be highly bespoke, but there is reason to believe that AI systems will see continued use within parliamentary settings, where many use cases have the potential to increase ease of access to information and offer citizens methodologies by which to become involved in political settings. Public-facing surveys and conversational tools represent a key arena for future development, and it is likely that tools like CitizenLab and ARCHIMEDE will continue to see further development and testing in the field.

Women's political leadership

Risks

Like prior sections, much of the risk in this thematic area centres on issues of bias, which have been found to have harmful effects along gendered lines. That said, AI and GenAI has been particularly harmful. The term 'deep fake' originated on Reddit, where the technology was originally utilised to produce illicit and compromising images of women, including female politicians.[115] The use of deep fakes has been noted as, "Potent vectors of tech-facilitated gender-based violence,"[116] while the algorithmic basis of platforms like Instagram and LinkedIn have been noted to be likelier to suppress content featuring women.[117] Furthermore, many hiring algorithms have been found to be biased against women, potentially removing strong candidates and minimising opportunities for selection and promotion.[118] These challenges are further exacerbated by a lack of gender representation in science and technology sectors, which remain male-dominated.[119]

Future use

---

[115] Busacca, 2023, p. 57.

[116] Vandinika Shukla, "Deepfakes and Elections: The Risk to Women's Political Participation | TechPolicy.Press," Tech Policy Press, February 29, 2024, https://techpolicy.press/deepfakes-and-elections-the-risk-to-womens-political-participation.

[117] Sophie Gardner, "Women and the Dark Side of AI," POLITICO, August 2, 2024, https://www.politico.com/newsletters/women-rule/2023/05/19/women-and-the-dark-side-of-ai-00097853.

[118] Gardner, 2024.

[119] Iris-Panagiota Efthymiou et al., "Using AI Changes the Paradigm of Women's Participation in Politics," *HAPSc Policy Briefs Series* 1 (December 29, 2020): 26, https://doi.org/10.12681/hapscpbs.26479, p. 28.

AI technologies have the potential to contribute positively to women's political leadership, even with the above considerations in mind. That said, further efforts are needed to ensure that the inherent risks present in many AI systems are addressed and mitigated given their already outsized impact across a variety of contexts. Given the current array of tools already in use, it is likely that there will be continued development of monitoring tools, both for general trends as well as specific systems which can target, delete, and ban content which may perpetuate gendered assumptions or biases. Additionally, there is ample room for the development of engagement tools which can streamline or otherwise assist women in becoming involved in government or running for office, which when paired with additional support mechanisms, have the potential to contribute positively to this area.

Recommendations: five next steps

The below recommendations have been developed for the consideration of WFD and other democracy support actors.

1. Continue to test AI-driven solutions and invest in research

- Artificial intelligence offers real opportunities to contribute positively to democracy support efforts and the current array of tools being produced by both states and on-the-ground organisations point to a variety of opportunities for advancement through AI-driven methodologies.
- Beyond just a tool, AI and broadly available GenAI systems do present considerable risks and raise additional questions. Democracy support actors must commit to continuing to pursue and consider the impact of AI across its broad applications. To monitor risks effectively, it will be vital that organisations routinely test AI systems for biases by using specific tools, techniques and human expertise to detect discrepancies in outcomes based on sensitive factors, such as race, gender, disability, culture, and socio-political views.
- Serious evidence gaps remain, including on the risks of commercialisation of AI, the impact of AI in elections (especially related to disinformation and microtargeting), and bias and exclusion effects of AI systems on democratic systems. Further research should respond to the questions and risks highlighted in this paper.

2. Consider establishing a set of ethical guidelines and good practices for applying AI

- A clear outline of ethical beliefs and good practices would support efforts to integrate AI technologies into the democracy support sector in the most responsible way, including by helping to mitigate the risks noted earlier in this paper. These guidelines would also support other organisations to better understand how to apply artificial intelligence responsibly.
- There remains an opportunity to play a formative role in supporting classificatory efforts of AI. While technical considerations will continue to exert an outsized role in these discussions, the creation of a democracy support-centric framework for understanding and interpreting AI models may be able to influence this arena and highlight opportunities, risks and trade-offs with democratic values.
- Any use of AI in democracy support must prioritise equitable and inclusive practice, ensuring that AI systems are not unjustly biased. This should also be routinely measured and tested to ensure accountability.

3. Promote the open sharing of data, backend information, and pilot results

- The lack of clear information on many AI applications remains a critical concern. The democracy support community should advocate for greater openness and transparency of the (increasingly commercialised) AI sector, including joining calls for commitments to use open-source and transparent technology in any applications.
- The open sharing of data and the results of early pilot testing or limited deployments remains scarce. WFD and democracy support actors should consider how best to publish results and lessons of their AI use, as well as how knowledge exchange could be systematised in new or existing forums.

4. Recognise the benefits and limits of both general and bespoke approaches to AI

- Publicly available AI tools can offer real benefits to stakeholders and individuals. Notwithstanding the limitations discussed in this paper, GenAI represents a real paradigm shift which can allow smaller organisations and actors to actively engage with the public on a scale previously not possible.
- The most effective AI solutions will remain those that are designed to respond to specific, on-the-ground challenges. Use cases and applications which engage a variety of stakeholders and are designed around specific needs are more likely to address problems in a transformative way.

5. Recognise that AI should only be used when it's the best solution to a problem

- **It is imperative to recognise that AI is not the answer to every challenge.** While AI has the potential to positively impact democracy support efforts, it will only sometimes be appropriate to deploy AI efforts in a given context. It is critical to engage with stakeholders and undergo rigorous problem analysis to ensure that the issue at hand is one that AI is best placed to address. If this is not the case, another solution should be prioritised.
- The introduction of AI can also serve to exacerbate existing digital divides or risks. It is important to recognise that the advancement of other digital democracy efforts may better fit the needs of specific stakeholders.
- The solution to a technological problem may not be technological in nature. While AI has many potential benefits, it is critical to avoid conflating 'emerging technologies' with 'effective solutions'. Certain challenges may be better addressed by other methods.

Annexe A: AI typologies

Artificial intelligence poses a terminological challenge in large part because the term contains little technical meaning, instead referring to an 'artificial' form of the nebulous concept of intelligence. While certain definitional factors have gained prominence in the V
Zture (like machine learning), even these core characteristics remain relatively plastic given the array of considerations at play. This has led researchers and AI ethics advocates to pursue additional typologies by which to otherwise define AI models. A select subset of these typological approaches are outlined below.

Typological Approach	Methodology
Narrow vs. General	• Classifies AI by level of intelligence • Narrow AI systems are capable of completing specific tasks, while general AI is capable of matching human cognitive capabilities • All current systems are currently classified as narrow AI
Vertical vs. Horizontal	• Classifies AI systems by their breadth of function • Vertical AI systems have highly specific use cases, like natural language processing, image identifications, etc. • Horizontal models are generalist architectures able to complete a variety of tasks across many different contexts
Symbolic vs. Statistical	• Symbolic models utilise mathematical or logical constraints; often found in logic programming • Statistical models utilise complex algorithms to identify patterns in data, usually without explicit programming • Hybrid models utilise both symbolic and statistical processes to produce outcomes
Discriminative vs. Generative	• Discriminative AI largely focuses on analytical processes of data sets • Generative AI utilises statistical models to generate information and data which could fit with the given data set • Mixed AI is able to perform both discriminative and generative tasks

The 'Ladder' Approach	• Eschews technically-founded typologizing and instead focuses on AI risk, categorizing systems by threat level • Has become a popular regulatory approach
The 'Matrix' Approach	• Classifies AI across a variety of axes in an attempt to capture technical, social, political, and ethical considerations • There are a variety of different matrices developed for categorizing AI

Narrow vs. general

A common method to analyse AI models is through the lens of their capabilities. All AI systems can be classified as either 'narrow' or 'general'—narrow AI models are only able to complete specific tasks but fail to match human intelligence, while general AI (AGI) matches or exceeds human intelligence.[120] Certain theorists have also proposed superintelligent AI which vastly expands beyond human capabilities, but the distinction has largely been a theoretical one.

While this approach has become popular for understanding AI models, it provides little information of note given that all existing AI systems are narrow AI. Furthermore, there remains considerable debate about whether AGI models are even possible, and the ethical concerns that they would produce.[121] These discussions also remain largely semantic; Martinez notes that, "As technology advances, so too do the tasks computers can accomplish. As machines accomplish more tasks, we tend not to consider them as reaching intelligence, but instead we move the threshold of intelligence farther away and then treat that specific task as unindicative of intelligence."[122] Given these factors, the 'narrow' vs. 'general' distinction remains an effort largely confined to discussions surrounding the threat of AI and does not offer an effective methodology by which to define specific systems.

Vertical vs. horizontal AI

Another approach to understanding advancements in AI is to consider the scope of each models' abilities. Machines and systems designed to complete domain-specific tasks or actions are

---

[120] Institute of Data, "Exploring the Differences Between Narrow AI, General AI, and Superintelligent AI | Institute of Data," October 6, 2023, https://www.institutedata.com/blog/exploring-the-differences-between-narrow-ai-general-ai-and-superintelligent-ai/.

[121] IBM, "What Is Strong AI?," *IBM* (blog), October 13, 2021, https://www.ibm.com/topics/strong-ai.

[122] Rex Martinez, "Artificial Intelligence: Distinguishing Between Types & Definitions," ARTIFICIAL INTELLIGENCE 19 (n.d.).

considered to be 'vertical' AI, while generalist AI models like OpenAI's ChatGPT or Google DeepMind's Gemini are considered to be 'horizontal' AI, capable of completing an array of tasks.[123]

While not as popularly utilised, the vertical vs. horizontal AI dichotomy is a useful tool for understanding the current AI landscape and is popular in efforts to catalogue AI companies. While a useful method to understand the desired focus of specific models and systems, the vertical vs. horizontal dichotomy largely serves as another basic methodology by which to categorize artificial intelligence while lacking the necessary complexity to sufficiently serve governments and other invested organisations.

Symbolic vs. statistical AI

Another approach is to categorize AI by whether it is a symbolic or statistical model. Symbolic systems utilise basic logic programming based on constraints such as, "rules, ontologies and search algorithms and rely on explicit descriptions of variables – agents like humans, entities like factories, objects like machines, variables that can be stock conditions – and descriptions of the interrelations between these variables."[124] Good Old-Fashioned AI (GOFAI), models utilising solely logical programming, usually are symbolic systems, and are still used in a variety of contexts within computing and everyday life. Statistical, or adaptive AI, utilises complex statistics to identify patterns within a given data set and complete tasks without explicit instruction.[125] Many of the largest and oft-discussed advancements in computing—neural networks, for example—utilise statistical methodologies. Newer models are also beginning to integrate both symbolic and statistical methodologies to complete tasks and represent a third class of systems under this framework.

While challenging to correctly identify given that many companies do not publicly share information about the internal structures of their systems, differentiating between symbolic and statistical methodologies represents a useful manner by which to consider AI systems without additional technical knowledge.

Discriminative vs. generative AI

Given the rapid advancement and high levels of interest in programs like ChatGPT, Llama, Sora, and Microsoft Copilot, there has been considerable discussion of the difference between discriminative and generative AI systems, especially given that many technologies, from Google Search to text recognition, utilise statistical methodologies while not fitting within the traditional

---

[123] Shahar Chen, "Horizontal Versus Vertical AI Solutions: Which Is Best?," *RTInsights* (blog), August 31, 2023, https://www.rtinsights.com/unleashing-the-power-of-horizontal-and-vertical-ai-solutions/.

[124] Committee on Digital Economy Policy, "OECD Framework for the Classification of AI Systems," *OECD's Artificial Intelligence in Work, Innovation, Productivity and Skills (AI-WIPS) Programme*, no. 323 (February 2022), p. 44.

[125] Jakob Mökander et al., "The Switch, the Ladder, and the Matrix: Models for Classifying AI Systems," Minds and Machines 33, no. 1 (March 1, 2023): 221–48, https://doi.org/10.1007/s11023-022-09620-y, p. 229; see Committee on Digital Economy Policy, 2022, p. 229.

schema of ‘artificial intelligence’. The above systems and a majority of the recent advancements in AI have been with generative systems, which given a set of data, then uses statistical techniques and pattern recognition to produce an outcome that would otherwise fit within the given data.[126] By contrast, a discriminative model is able to identify between data types and clarify additional patterns.[127]

While still a simplified framework given the complexity of many of today’s most popular AI technologies, the distinction between discriminative and generative artificial intelligence is a useful one and captures the array of technologies and applications which could be popularly classified as ‘AI’ while avoiding many of the potential pitfalls. As such, it differs from other approaches and offers additional benefits. Importantly, the differentiation between generative and discriminative models directly identifies and incorporates analysis of the generative tools which have captured public consciousness while ensuring equal emphasis on discriminative tools which offer meaningful benefits across a broad array of use cases. While insufficiently specific when defining AI systems, the generative vs. discriminative framework is a necessary one and as such is a key component of the proposed working definition.

The ‘ladder’ approach

While the prior approaches largely attempt to establish a dichotomy *between* models, attempting to determine an explicit point at which a certain system becomes ‘AI’, the broad variety of research and advancements in computing have made such approaches a challenge. While a reasonable desire, the complexity of these AI models and their nature as ‘black boxes’ into which stakeholders cannot see has complicated such attempts. Regulators and other institutions have as a result pursued different means by which to categorize such systems. The outcome is what Mökander et al. have described as the ‘ladder’ Approach.[128]

The ladder approach typically eschews analysis of the system itself as a method of classification in favour of an outcome-oriented approach. Drawing from the burgeoning field of AI ethics, the approach emphasises the *potential* of a specific model to cause harm and the *severity* of its impact.[129] Regulators then classify systems according to different levels of risk (the ‘rungs’ of the ladder).[130] This risk-based methodology has become the approach of choice, especially within Europe, where the European Parliament utilised it within its Artificial Intelligence Act.[131] The

---

[126] Committee on Digital Economy Policy, 2022, p. 45.

[127] Google Staff, “Background: What Is a Generative Model?,” Google for Developers, accessed August 4, 2024, https://developers.google.com/machine-learning/gan/generative.

[128] Mökander et al., 2022, p. 235.

[129] Mökander et al., 2022, p. 235.

[130] Mökander et al., 2022, p. 237.

[131] Council of the European Union, “Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act),” Pub. L. No.

ladder's popularity as a regulatory tool is in large part a result of its emphasis on the ethical use of AI technologies and focus on potential risks. While remaining relatively broad, it offers a framework that avoids the rigidity of technical approaches to classification while centring conversation on a core necessity for AI regulation: to prevent damage or abuse. That said, draft proposals utilising the ladder approach have often been unclear in defining exactly how regulations will be applied, and determination of the potential threats of given models may require high levels of technical knowledge and analysis which many state actors and stakeholders may not have.

The 'matrix' approach

The final methodology for classifying AI system avoids establishing an explicit point at which a model becomes 'AI', but also abstains from using a singular measure for classifying AI. Instead, many institutions have pursued a methodology that Mökander et al. call the 'matrix' approach.[132] Matrix-centred approaches utilise multiple dimensions to classify AI systems, allowing for the consideration of multiple factors that may not be captured in a technical or risk-based definition. These may include but are not limited to certain social or ethical implications. As Mökander et al. note, this provides inherent benefits over other approaches by adding nuance to the equation. More than simply determining what AI *is* or *isn't*, multi-dimensional approaches can, "also help organisations identify which precautionary measures are appropriate when designing or implementing a specific AI system."[133] Furthermore, matrix-based methodologies can offer both specificity and broadness, allowing organisations to consider shared risks and concerns across a variety of contexts while preserving specificity in others where considerations may differ.[134] This is not to imply that these frameworks are easy-to-use—the multi-modal nature of the matrix approach does increase the complexity of efforts to classify AI systems. That said, it promotes a deeper understanding of artificial intelligence and allows a far greater level of insight than other approaches, prompting deeper, more thoughtful analysis of AI from multiple perspectives.

There are many matrix-based methodologies which have already seen development. The OECD model classifies AI along five dimensions: people and planet, economic context, data and input, AI model, and task and output. Across each dimension, there are additional criteria by which to analyse and classify a specific system, adding to 37 in total.[135] The definition utilised within this paper, for example, draws heavily from the criteria within the 'AI model' dimension to increase technical specificity while explicitly allowing for the potential incorporation of value-based frameworks in the future. The Centre for Security and Emerging Technology (CSET), meanwhile,

---

ST_7536_2024_INIT (2024), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CONSIL%3AST_7536_2024_INIT&qid=1716971248813.

[132] Mökander et al., 2022, p. 237.

[133] Mökander et al., 2022, p. 241.

[134] OECD, "OECD Framework for the Classification of AI Systems" (Paris: OECD, February 22, 2022), https://doi.org/10.1787/cb6d9eca-en, p. 3.

[135] A chart of both the dimensions and specific criteria can be found on page 18 of the OECD Framework for the classification of AI Systems.

utilised four core dimensions in a comparison effort of dimensional frameworks: context, input, model, and output. The CSET matrix contained nine criteria in total and was found to exhibit a higher rate of consistent classification across all dimensions compared to frameworks with less.[136] Straub et al., by comparison, use another framework with broader dimensions to classify AI systems specifically in government: operational fitness, epistemic completeness, and normative salience. 'Operational fitness' refers to, "the degree to which the composition and functions . . . [of] an AI application . . . aligns with codified standards of an organisation and system construction and functioning [standards]."[137] 'Epistemic completeness' is the degree to which backend information about the data and composition of an application is aligned with knowledge sharing practices and standards, while normative salience refers to whether the behaviour of the AI system aligns with institutional and ethical standards.[138] Internally, Straub et al.'s model utilises a quasi-ladder based methodology, classifying AI systems on one of three levels.[139] This points to the flexibility exhibited by multi-dimensional frameworks; rather than simply utilising a singular approach, matrix-based classification methodologies can incorporate best practice across a variety of contexts and offer as much or as little complexity as needed.

While by no means a perfect solution, the matrix approach offers enormous benefits compared to other classification methods. As such, the AI definition proposed within this report attempts to offer flexibility for future enhancement via a multi-dimension classificatory framework while maintaining enough specificity for current use. As echoed within the recommendations presented at the end of this brief, the pursuit of a matrix-based framework for AI applications within democracy support is highly advisable given key differences from AI use in the public sector or elsewhere, and merits further consideration beyond the confines of this document.

---

136 Center for Security and Emerging Technology and Catherine Aiken, "Classifying AI Systems" (Center for Security and Emerging Technology, November 2021), https://doi.org/10.51593/20200025, p. 26.

137 Vincent Straub et al., *Artificial Intelligence in Government: Concepts, Standards, and a Unified Framework*, 2022, p. 15.

138 Straub et al., 2022, p. 15.

139 Straub et al., 2022, p. 16.