DEMOCRACY IN THE AGE OF GENERATIVE AI:

NAVIGATING RISKS AND HARNESSING OPPORTUNITIES

IRI INTERNATIONAL REPUBLICAN INSTITUTE
Advancing Democracy Worldwide

Democracy in the Age of Generative AI:
Navigating Risks and Harnessing Opportunities

ABOUT US

THE INTERNATIONAL REPUBLICAN INSTITUTE (IRI)

IRI is one of the world's leading international democracy development organizations. The nonpartisan, nongovernmental institute has supported civil society organizations, journalists, democratic governments, and other democratic actors in more than 100 countries since 1983—in Africa, Asia, Eurasia, Europe, Latin America and the Caribbean, and the Middle East and North Africa—with a presence in over 80 others worldwide. IRI's Technology and Democracy Practice works in every region of the world to help grassroots actors use technology to advance democratic principles, solve public problems, and support stronger governance.

IRI'S TECHNOLOGY AND DEMOCRACY PRACTICE

IRI's Technology & Democracy programming aims to make the digital revolution work for democracy rather than against it, specifically:

1. To ensure that the digital revolution advances democracy and freedom and outcompetes digital authoritarians;
2. To build a 21st-century democracy by galvanizing the benefits of technology to make democracy stronger; and
3. To increase the resiliency of democratic actors to digital repression.

IRI's Technology & Democracy Practice has conducted programming on digital democracy, civic and government technology, internet freedom, emerging technologies, and information integrity in every region of the world. IRI has experience supporting civil society organizations (CSOs) and other democracy partners such as governments, political parties, activists, and others in mainstreaming internet freedom tools, and monitoring, countering and building resilience to information integrity threats in repressive environments as well as advancing technology for good governance programming. These programs have included support to actors to deploy digital technologies in ways that advance transparency, inclusion, accountability and responsiveness, while adjusting to a changing threat landscape, including emerging technologies.

ACKNOWLEDGEMENTS

The development of this white paper was made possible through the writing and research of Hui Hui Ooi, Aaron Spitler, and Amanda Zink, with extensive support by Justin Sherman and Devan Desai at Global Cyber Strategies. Thanks to Jared Ford for his review, editing support and strategic guidance throughout the development and finalization of the report. We also appreciate the myriad of IRI staff who provided invaluable feedback throughout the drafting process.

Finally, IRI would like to thank the members of the Generative AI and Democracy Working Group, some of whom are listed below, who shared their insights throughout the working group series, as well as feedback that helped inform the direction of this white paper.

- Nicklas Badum, Democracy X
- Laure de Roucy-Rochegonde, Institut français des relations internationales (IFRI)
- Renee DiResta, Independent
- Konur Metehan Durmaz, Social Media Exchange (SMEX)
- Adam Fivenson, National Endowment for Democracy (NED)
- Allie Funk, Freedom House
- Arzak Khan, Innovation Center for Artificial Intelligence and Human Rights
- Agita Pasaribu, Bullyid Indonesia
- Cristina Martinez Pinto, PIT Policy Lab
- Lisa Reppell, International Foundation for Electoral Systems (IFES)
- Max Rizzuto, Atlantic Council's Digital Forensic Research Lab (DFRLab)
- Priscilla Ruiz, Article 19 Mexico and Central America Regional Office
- Frei Sangil, Layertech Lab
- Maurice Sayinzoga, National Democratic Institute (NDI)

LIST OF ACRONYMS

AFP: Agence France-Presse

CCP: Chinese Community Party

CSO: Civil society organization

C2PA: Steering Committee of the Coalition for Content Provenance and Authenticity

DC: District of Columbia

DFRLab: Digital Forensic Research Lab at the Atlantic Council

ECI: Election Commission of India

EU: European Union

GPAI: Global Partnership on Artificial Intelligence

GRAIL: Governance and Responsible AI Lab at Purdue University

IFES: International Foundation for Electoral Systems

IFRI: Institut français des relations internationales

INGO: International nongovernmental organization

IRP: Incident response plan

IRI: International Republican Institute

IT: Information technology

LGBTQI+: Lesbian, gay, bisexual, transgender, queer, intersex, and more

LLM: Large language model

MIT: Massachusetts Institute of Technology

ML: Machine learning

NED: National Endowment for Democracy

NLP: Natural language processing

NGO: Nongovernmental organization

OECD: Organisation for Economic Co-operation and Development

PRC: People's Republic of China

RAIN: Responsible AI Network (Africa)

SMEX: Social Media Exchange

TFGBV: Technology-facilitate gender-based violence

UNDP: United Nations Development Programme

UNESCO: United Nations Educational, Scientific and Cultural Organization

UNGA: United Nations General Assembly

US: United States

USAID: United States Agency for International Development

TABLE OF CONTENTS

INTRODUCTION: WHY WE WROTE THIS

Democratic societies are no stranger to change. Throughout history, democracies have evolved in response to a diverse set of social, political, and economic factors time and again. Still, there are moments in history when democratic societies face a significant rate of, and pressure for, change. The rise of generative artificial intelligence (GenAI) in late 2022 has seemingly introduced one of those moments, promising a transformation for how society and government operate that many have compared to the transformations brought about by the advent of the internet.[1] As with many technological advancements in the past, governments and civil society can struggle to understand, anticipate, and minimize the harms and maximize the benefits of digital innovations on democratic societies. To avoid a repeat of past mistakes, the Technology and Democracy Practice at the International Republican Institute (IRI) launched a program in the summer of 2023 to respond to partners' demands about preparing for current and potential impacts caused by GenAI.[2]

Many of IRI's partners, particularly governments and civil society actors, struggle to keep pace with technological innovations, often failing to leverage new tools for the benefit of democracy and struggling to proactively advocate or regulate against their harms. This is especially concerning as these two stakeholder groups are key to the effective functioning of democracies. IRI's partners—most often in this case civil society and decision-makers, including elected officials and policymakers—have increasingly recognized the need to upskill themselves on this topic to ensure democracies remain nimble, modern, and effective in the digital age.

This intentional upskilling is especially acute to make sure actors in the Global South are not left behind. This white paper intends to help fill knowledge gaps and support IRI partners in better understanding this innovation, providing analysis of the risks, challenges, and opportunities for GenAI and democracy for lower-capacity government and civil society stakeholders in the Global North and Global South. The white paper provides an overview written to introduce and define GenAI, including an analysis of key threats and risks that GenAI poses to democracy, and a summary of opportunities for civil society and policymakers to leverage it for good.

This white paper intends to help fill knowledge gaps and support IRI partners in better understanding generative AI, providing analysis of the risks, challenges and opportunities for GenAI and democracy for lower-capacity government and civil society stakeholders around the world.

The white paper also includes a chapter on recommendations for civil society, policymakers, and industry to ensure GenAI works for, rather than against, democracy. The final chapter of the white paper focuses on the near-term impacts of GenAI, offering readers a chance to consider possible future impacts. It is important to keep in mind that findings in the white paper are generalized to be broadly applicable to diverse global contexts, and that the threats, opportunities, and recommendations may vary depending on these factors.

This white paper is the culmination of an extensive landscape analysis and interview series, followed by a six-month-long working group that regularly convened over 20 experts - including representatives from civil society, academia, industry, and policy - to discuss existing and potential impacts GenAI will have on democracies, with a particular focus on trends in the Global South.[3] The white paper is a synthesis of the discussions and findings of the working group, intended to provide guidance to civil society and policymakers on how to maximize the benefits and minimize the harms that GenAI poses to democracy.

[1] Bajarin, Tim. "The practical impact of AI for the Masses." Forbes, 28 Nov. 2023, www.forbes.com/sites/timbajarin/2023/11/28/the-practical-impact-of-ai-for-the-masses/?sh=5217fc8f675b.

[2] Hijjawi, Faisal. "AI's Regulatory Challenge: Avoiding the Pitfalls of Past Mistakes." Science and Technology Law Review, 18 Apr. 2023, journals.library.columbia.edu/index.php/stlr/blog/view/583.

[3] "Generative Artificial Intelligence and Democracy Working Group." International Republican Institute, www.iri.org/what-we-do/programs/generative-artificial-intelligence-and-democracy-working-group/.

WHO THIS WAS WRITTEN FOR

The risks, opportunities, and recommendations included in this white paper are relevant for any actor concerned about the future of democracy. For some readers, this may be an obvious area of interest, but for others, it may seem far from their current work or remote from the challenges they face.

Industry, civil society organizations (CSOs), and governments—all key stakeholders in democratic societies—have begun to use these tools, making it necessary for them to at least be familiar with the basics. Below is a sample of the types of groups that might find this white paper beneficial; even if your organization or government does not currently engage with or use GenAI, these tools will eventually begin to appear in your areas of work.

- Advocacy CSOs
- Citizen engagement CSOs
- Policymakers
- Elected officials
- Awareness raising/education CSOs
- Fact-checking and/or media groups
- Tech companies

Consider the following example: you work at a citizen engagement-focused CSO that has not decided whether to intentionally use or interact with GenAI tools. However, in the coming months or even weeks, your government will launch a virtual chatbot that leverages GenAI to address citizens' concerns, complaints, and suggestions. Or perhaps a partner CSO will begin using GenAI tools to quickly and efficiently translate guides explaining how individuals can access government services into multiple languages. In either instance, it will be important—at a minimum—to understand these tools, the risks they present, and the data practices that will be employed to educate citizens and other beneficiaries. These situations are likely to increasingly emerge, so it is important to understand this topic even if your organization does not plan to incorporate these tools.

KEY FINDINGS

Following desk research, expert consultations, and working group discussions on GenAI and its impacts, particularly throughout the Global South, IRI distilled key risks, opportunities, and recommendations applicable to diverse democratic actors. These findings should be considered to help inform the work of civil society, decision-makers, and industry representatives striving to protect democracy and democratic principles globally.

Key Risks

- Increased speed and lowered cost of producing disinformation
- Exacerbation of hate speech, harassment, and technology-facilitated gender-based violence (TFGBV)
- Overall lowered trust in the information environment that may lead to deep skepticism and lowered engagement in democratic processes and institutions
- Widening digital divides (especially impacting the Global South)
- The lack of accountability in GenAI development and use may reduce trust in industry and government
- Increased privacy concerns—e.g., what data is used to develop GenAI (and how it was accessed), use of GenAI for surveillance and manipulation
- Overfocusing on technical solutions (e.g., content provenance) to threats introduced by GenAI may divert attention from equally important approaches addressing the political, regulatory and social context in which GenAI-related threats exist (e.g., public education campaigns)

Key Opportunities

- Strengthen strategies and efforts to protect information environments through improved identification, labeling, and debunking of mis/disinformation
- Power responsive and efficient decision-making by synthesizing complex documents, digesting vast quantities of data to inform policymaking, and more
- Enhance organizational capacity for governments and civil society facing capacity constraints
- Improve institutional transparency by increasing accessibility of government data, service delivery, powering accessible chatbots, and more
- Leverage GenAI to support the scaling of civic education efforts to more directly provide individuals with key governance and civic education information
- Create adaptable cyber defenses to enhance defensive countermeasures and streamline the work needed to fend off cyberattacks
- More accurately predict policy outcomes to spot emerging trends, patterns, and potential scenarios to proactively address policy issues

RECOMMENDATIONS

Below is a set of recommendations distilled from working group conversations for three key stakeholder groups that are central to ensuring GenAI is leveraged democratically: civil society, government/decision-makers, and industry. Each stakeholder group must take action to ensure GenAI is a benefit—or at minimum, not a harm—to democracy. Perhaps even more importantly, these stakeholder groups must collaborate with one another to achieve real impact. The recommendations found below are segmented to speak specifically to each stakeholder group, with many of them referencing the multistakeholder collaboration required to effectively tackle potential harms that GenAI poses.

Civil Society

- Refer to ethical GenAI use frameworks to develop internal GenAI use policies
- Join global coalitions and advocate in those forums for a greater focus on GenAI and AI regulations and governance
- Develop an incident response plan (IRP) that incorporates GenAI-related issues
- Flag government and private sector uses of GenAI that violate human rights
- Support societal education efforts, like civic education campaigns, as well as upskilling internal staff
- Review your country's laws and regulations on the levels of access around GenAI-related data

Government/Decision-Makers

- Create a public resource website that clearly, understandably, and comprehensively describes how GenAI is used in government agencies
- Ensure procurement teams have multidisciplinary staff and an established approach to procure GenAI
- Identify opportunities to integrate GenAI to improve service delivery and accessibility of information
- Assess how GenAI aligns with existing laws and regulations
- Refer to existing guidelines for insights on how to address gaps in current laws and regulations
- Involve citizen feedback in all stages of GenAI integration
- Engage external experts and build in-house capacity to the extent possible

Industry

- Build capacity among industry professionals on human rights, legal, and societal implications of GenAI
- Engage policymakers in national AI lawmaking processes
- Publish more information about GenAI use practices and ethical standards on company websites
- Build products and services according to human rights principles while building multidisciplinary and diverse teams
- Provide sustainable, long-term funding for civil society and academic partners to ensure diverse actors can contribute to solutions to minimze the harms of GenAI
- Create sustained knowledge-sharing opportunities with civil society and academia to ensure diverse voices inform tool development, usage policies, and creation of guardrails
- Tailor harm mitigation efforts for various environments
- Commit resources to develop methods to detect GenAI misuse

INFORMATIONAL PRIMER: UNDERSTANDING GENERATIVE AI

Understanding the Terminology

Understanding GenAI, which is a subset of artificial intelligence (AI), requires first understanding the general definition of AI—often called traditional or narrow AI—which has been a part of technological systems for decades. The Organisation for Economic Co-operation and Development (OECD), which established the OECD Principles on Artificial Intelligence in 2019, defines AI as: "...a machine-based system that can, for a given set of human defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments."[4] AI refers to systems designed to respond to a particular set of inputs, powered by algorithms,[5] and can be used to analyze data and make simple predictions. These systems have been part of everyday life for years, such as AI algorithms that power recommendation engines on Amazon, Netflix, and Google. These AI systems process data inputs, such as information on users' previously selected shows on Netflix's streaming service or the words users input into Google's search engine, and then provide suggested results in response to those inputs. These AI systems have been programmed to follow specific rules to respond to requests, and they do not generate new content.

Much of the current excitement around AI is focused specifically on GenAI; the launch of ChatGPT in late 2022 brought GenAI to the forefront and made it accessible to everyday citizens without GenAI expertise, often for free. GenAI, while built on a similar machine-based system, can also create new content in response to prompts, including images, text, and audio-based content. However, rather than simply following a specific set of rules, GenAI models are trained on a set of data and learn the underlying patterns through a process called machine learning (ML) to generate new data mirroring the training set. This can also be described as calculating the probability of the next token in a sequence, such as what the next word might be in a sentence. GenAI is built on large language models (LLMs) in addition to other types of machine learning models. Recent innovations in GenAI include interactive chatbots and search portals such as OpenAI's ChatGPT, Microsoft's Bing, and Google's Gemini, all of which can generate outputs in response to inputs of various types. This functionality includes, for example, generating an outline for an academic paper with limited instruction or creating a picture of a whale upon request. The more data a GenAI system receives, the stronger its outputs become.

There are many terms related to GenAI that have only been addressed briefly in this primer. A good place to start is familiarizing oneself with some of the terms that frequently come up in conversations about AI, such as algorithms, machine learning, and LLMs. We suggest reviewing CNET's Glossary of AI Terms for a user-friendly overview of key definitions related to understanding AI and GenAI.[6] More detailed resources can help understand how these terms interconnect, such as Medium's Generative AI for Beginners blog series.[7]

[4] Artificial Intelligence and Responsible Business Conduct. OECD, 2021, mneguidelines.oecd.org/RBC-and-artificial-intelligence.pdf.

[5] Algorithms are step-by-step instructions or rules that guide AI to solve a specific problem or perform a particular task. For GenAI, that can include instructing GenAI to learn from data, recognize patterns, generate new content, and adapt over time.

[6] Khan, Imad. "CHATGPT Glossary: 44 AI Terms That Everyone Should Know," CNET, 14 May 2024, www.cnet.com/tech/services-and-software/chatgpt-glossary-44-ai-terms-that-everyone-should-know/.

[7] Gupta, Raja. "Generative AI for Beginners: Part 1 – Introduction to AI," Medium, 8 Feb 2024, https://medium.com/@raja.gupta20/generative-ai-for-beginners-part-1-introduction-to-ai-eadb5a71f07d. Archived Version, https://web.archive.org/web/20240719150037/https://medium.com/@raja.gupta20/generative-ai-for-beginners-part-1-introduction-to-ai-eadb5a71f07d.

Variables Shaping Current Use

Before jumping into the risks, threats, and opportunities that GenAI presents for democracy, it is important to note how GenAI has been used to date and where current conversations on GenAI stand. It is especially important to consider how GenAI's impacts may vary depending on the political context within a country. In the short time since OpenAI launched ChatGPT, its usage has exploded across academia, among ordinary citizens, industry, civil society, the public sector, and others—examples of some of these use cases will be explored later in the white paper. What is especially important to note, including for readers outside of the Global North, is that the way GenAI's use and impact have played out across countries varies depending on regulatory environments, internet access and infrastructure, ethical norms and digital literacy levels, among other factors.

Conversation and analysis on GenAI's impacts often have an imbalanced focus on the Global North, where infrastructure, data privacy, and internet access are typically more reliable, and where the norms and cultures embedded into the development of these tools originate. Less discussed is how GenAI is or is not being used in the Global South, where weaker infrastructure, slower internet, and a lack of protective regulation can result in outsized harm, including the exacerbation of the digital divide. The realities of digital subdivides, such as disparities between urban and rural access, gender inequalities, and limited access for marginalized communities, are rarely mentioned. Frameworks such as USAID's Digital Ecosystem Country Assessment and resources such as Civicspace.tech can help to analyze digital ecosystems and consider civic spaces to inform more nuanced discussions on GenAI's impacts, including how GenAI may exacerbate or resolve existing challenges.

It is also important to note the rapid emergence of AI regulations and frameworks that look quite different depending on the government, entity, or country. For example, at the time of writing this white paper, the European Union (EU) released the EU AI Act,[8] which relies on a risk-based approach to govern the advancement of AI, while the United States' (US') White House Blueprint for an AI Bill of Rights[9] emphasizes nonbinding, guiding principles rather than risk-based, specific rules. Both emphasize a respect for human rights, transparency, and risk management. The People's Republic of China (PRC), meanwhile, has developed the Interim Measures for the Management of Generative AI Services, a rules-based approach that is rigid and focused on specific applications of the technology.[10] Additional frameworks have been developed by leading international nongovernmental organizations (INGOs), such as UNESCO's Recommendations on the Ethics of AI,[11] UNGA's Resolution on AI,[12] and the OECD AI Principles,[13] to name a few. As one considers the implications of AI and GenAI in a particular context, it is important to remember that existing frameworks or regulations may vary widely—or are completely absent—in said country, thereby altering the impact that GenAI may have generally. The AI Watch: Global Regulatory Tracker[14] may be a useful tool to help track evolving policies, as well as the Global Center of AI Governance's Global Index on Responsible AI.[15]

[8] "The EU Artificial Intelligence Act." European Union, artificialintelligenceact.eu/. Accessed 26 June 2024.

[9] "Blueprint for an AI Bill of Rights," The White House, Oct. 2022, www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

[10] Regulatory and Legislation: China's Interim Measures for the Management of Generative Artificial Intelligence Services Officially Implemented. PwC, Aug. 2023, www.pwccn.com/en/tmt/interim-measures-for-generative-ai-services-implemented-aug2023.pdf.

[11] Recommendation on the Ethics of Artificial Intelligence, UNESCO, 2022, unesdoc.unesco.org/ark:/48223/pf0000381137.

[12] General Assembly Adopts Landmark Resolution on Artificial Intelligence," UN News, 21 Mar. 2024, news.un.org/en/story/2024/03/1147831.

[13] "Artificial Intelligence," OECD.

[14] "AI Watch: Global Regulatory Tracker." White & Case Insights, 13 May 2024, www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker.

[15] "Main Findings." Global Index on Responsible AI, 13 June 2024, https://global-index.ai/Takeaways.

GENERATIVE AI'S RISKS TO DEMOCRACY

Since the release of ChatGPT at the end of 2022, followed by tools such as Claude[16], Gemini[17], and Meta AI[18], there has been increased attention and public use of GenAI tools that has escalated the risks posed to democracy, including for countries in the Global South. GenAI-related risks have presented themselves as real-world harms that threaten to get worse. The spread of deepfakes, referring to audio-visual content that has been convincingly manipulated to misrepresent someone's actions, authoritarian misuse, and data privacy risks are exacerbating trends that may erode the core pillars of a healthy democracy; specifically, the (mis)use of GenAI has plagued online information ecosystems, interfered with elections, silenced voices necessary for democratic pluralism, and reduced public trust in institutions. Below are additional details of GenAI's potential threats to democracy.

Increased Speed and Lower Cost of Producing Disinformation

Authoritarian actors in countries such as China and Russia, as well as illiberal actors within democratic countries, can use GenAI to speed up the production of disinformation at a lower cost. The believability of this content varies. Those with a more discerning eye may analyze mouth movements in video deepfakes, for instance, to determine the authenticity of the information; but not everyone will have such a keen sense of detection. Often marginalized groups, such as women and ethnic minorities, have lower levels of digital literacy. Therefore, there is a risk that they could be disproportionately affected or swayed by such content. In some cases, entrepreneurial nonstate groups or individuals seeking financial gain in the disinformation market leverage this technology, which is powered by the accessible, low-cost use of GenAI. This risk is especially harmful in elections as the rapid proliferation of GenAI tools will mean that everyday voters will encounter easily made disinformation that may be increasingly convincing as these tools become more sophisticated. This may impact the ability of voters to authenticate information and make informed choices, potentially resulting in increased political apathy, and may undermine vulnerable candidates, including women. While impacts from GenAI-created content on electoral outcomes remain to be seen and are complex to determine, these factors are concerning for democracy actors around the world. Countries across the globe, including in the Global South, that already struggle with and lack proper resources to counter information integrity threats online will face more challenges as deepfakes increasingly plague the online information space.

Case Study

SLOVAKIA | AI-GENERATED ELECTION DISINFORMATION

In the lead-up to the 2023 Slovak parliamentary election, an audio clip circulated on Facebook purportedly capturing a conversation between Michal Šimečka, leader of the liberal Progressive Slovakia party, and Monika Tódová, a journalist from the center-liberal Denník N newspaper. The recording allegedly featured Šimečka explaining to Tódová how he intended to steal the contest from his rivals, primarily by bribing voters from the country's marginalized Roma community. Both the politician and journalist responded that the audio was completely fabricated, and a report from the fact-checking department of Agence France-Presse (AFP) later concluded that the clip was most likely altered using AI tools.[19] However, the recording was shared during a critical 48-hour blackout period prior to the elections, during which candidates are prohibited from campaigning. Given the timing, watchdog groups found it difficult to properly debunk the disinformation, as citizens had already seen it before they went to the polls. While it is difficult to isolate the impact this event had on the outcome[20] of the election, it is important for democracy actors to continue tracking and mapping uses like this.

[16] "Home." Claude, claude.ai. Accessed 26 June 2024.

[17] "Home." Gemini, gemini.google.com. Accessed 26 June 2024.

[18] "Meet Your New Assistant: Meta AI, Built With Llama 3." Meta, 18 Apr. 2024, about.fb.com/news/2024/04/meta-ai-assistant-built-with-llama-3.

[19] Barca, Robert. "Údajná Nahrávka Telefonátu Predsedu PS a Novinárky Denníka N Vykazuje Podľa Expertov Početné Známky Manipulácie." AFP Fact Check, 28 Sept. 2023, fakty.afp.com/doc.afp.com.33WY9LF.

[20] "Poll of Polls: Slovakian Polls, Trends and Election News for Slovakia." POLITICO, www.politico.eu/europe-poll-of-polls/slovakia. Accessed 26 June 2024.

Case Study

GLOBAL | EXPLOITATION OF AI FOR MANIPULATION AT SCALE

In March 2024, the Insikt Group of Recorded Future — one of the world's largest threat intelligence companies—identified a malign influence network leveraging inauthentic media outlets in the United States, the United Kingdom, and France. The CopyCop network is suspected to be Russian-operated, and Insikt Group reported that "the network extensively used large language models (LLMs) to plagiarize, translate, and modify content from legitimate media sources to tailor political messages with specific biases, which included content critical of Western policies and supportive of Russian perspectives on international issues like the Ukraine conflict..."[21] As this content gained traction, CopyCop began to also post human-produced content along with its AI-generated content as a strategy to generate and distribute content at scale to evoke targeted political emotions. The Economist reported that within two months, CopyCop uploaded more than nineteen thousand deceptive posts.[22] The emergence of GenAI-empowered networks like this creates further chaos ahead of critical elections, making the task of safeguarding elections even more challenging.

Exacerbation of Hate Speech and Harassment[23]

GenAI could accelerate existing trends that have plagued online ecosystems, especially the spread of hate speech, online harassment, and technology-facilitated violence (including gender-based forms). These attacks disproportionately target vulnerable and marginalized communities, such as women and gender nonconforming individuals, youth, LGBTQI+[24] individuals, persons with disabilities, and other religious and racial minorities. GenAI can lead to the creation of sustained and automated attacks, increased reach of hate speech and harassment beyond national borders, and the cheap and easy generation of content such as deepfakes, nonconsensual sexual imagery, and impersonation, all of which increases risks and threats, online and offline, to the communities listed above. As referenced in the UNESCO report on technology-facilitated gender-based violence in the era of GenAI,[25] there has been evidence that open and closed AI models can be used to propagate existing technology-facilitated gender-based violence (TFGBV) harms, such as impersonation, hacking and stalking, and cyber-harassment, among others. Combined, these use cases perpetuate discriminatory attitudes and biases and impose barriers to entry into public and political dialogue for vulnerable and marginalized groups, which negatively impacts political pluralism.

Case Study

GLOBAL | GENERATING DEFAMATORY DEEPFAKES

DeepTrace, a Dutch company that tracks synthetic media, published a report[26] in September 2019 highlighting the state of GenAI-created deepfakes. Alarmingly, their team estimated approximately 96 percent of all deepfake videos found online were nonconsensual pornography.

[21] Insikt Group. "Russia-Linked CopyCop Uses LLMs to Weaponize Influence Content at Scale." Recorded Future, 9 May 2024, www.recordedfuture.com/russia-linked-copycop-uses-llms-to-weaponize-influence-content-at-scale.

[22] "A Russia-linked Network Uses AI to Rewrite Real News Stories." The Economist, 10 May 2024, www.economist.com/science-and-technology/2024/05/10/a-russia-linked-network-uses-ai-to-rewrite-real-news-stories.

[23] The United Nations defines hate speech as any kind of communication in speech, writing or behavior, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, color, descent, gender or other identity factor. See: United Nations. "What Is Hate Speech?" United Nations, www.un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech. Accessed 26 June 2024.

[24] Lesbian, gay, bisexual, transgender, queer, intersex, and more.

[25] "Your opinion doesn't matter, anyway": exposing technology-facilitated gender-based violence in an era of generative AI." UNESCO, 2023, https://unesdoc.unesco.org/ark:/48223/pf0000387483.locale=en.

[26] Ajder, Henry, et al. "The State of Deepfakes: Landscape, Threats, and Impact." Deeptrace, Sept. 2019, regmedia.co.uk/2019/10/08/deepfake_report.pdf.

Their report, which featured data from the five most popular deepfake pornography websites at the time of publication, also noted that all content featured women whose images were used without their permission. While these explicit videos often featured regular people, public figures were targeted as well in what was usually an attempt to defame their professional reputation. Women in a variety of positions, from veteran journalists[27] to government officials,[28] have regularly been at the center of GenAI-enabled campaigns created to push them out of public life. Concerns as to how GenAI is and can be used to amplify the harassment that women face, in particular, have continued to grow since 2019, and this issue remains a concern for other marginalized communities that similarly deal with disparaging attacks.

Overall Distrust in the Information Environment

Malicious uses of GenAI, such as the inexpensive production of written disinformation or deepfakes could further diminish public trust in the reliability of information environments. Simultaneously, the more benign, general-purpose use of GenAI to create large amounts of content is likely to crowd the information space, lowering the quality of information online. As public awareness of GenAI grows, doubt and distrust in online information could manifest broadly. Increased citizen understanding of and awareness about deepfakes may result in deep skepticism overall as to the reliability of information about political actors, campaigns, election results, and more. This skepticisim can lead to a decrease in democratic engagement, as citizens become increasingly disillusioned with democratic actors and, subsequently, democratic processes. According to an OpenAI report released in January 2023,[29] the authors stated that LLMs can create propaganda that is more persuasive and impactful than techniques already in use and at a fraction of a cost; the authors acknowledged that GenAI could significantly enhance influence operations that further sow confusion and skepticism. Experts have speculated that it is likely the governments of the People's Republic of China and Russia—as well as domestic actors, companies, and even influencers—will use GenAI to create convincing content that overwhelms online platforms in targeted countries, challenging moderation efforts and further reducing information quality.

Case Study

CHINA | EXPLOITING PUBLICLY AVAILABLE AI TOOLS

In February 2023, social media analytics firm Graphika[30] released a report[31] on a pro-PRC influence operation that used AI-generated news anchors to spread state-approved propaganda. The campaign, through a fictitious outlet called Wolf News, produced a pair of videos that circulated across social media platforms. The first video included a presenter accusing the United States of failing to take action on gun violence in the country, while the second featured a commentator emphasizing the need for cooperation between Beijing and Washington. Notably, these clips were generated using technology created by Synthesia, a United Kingdom-based company that creates personas for companies to use as a way of engaging their clients' target audience. In their brief, Graphika notes that publicly available tools like the ones provided by Synthesia have been exploited by malign actors. Outside of the PRC, state media in Venezuela has also used Synthesia technology, most notably in a 2023 when they created a pair of fabricated videos featuring English-speaking commentators pushing pro-government messaging. This incident, which was explored in Freedom House's Freedom on the Net 2023 report,[32] shows how authoritarian actors have leveraged GenAI to create propaganda.

[27] Ayyub, Rana. "I Was the Victim of a Deepfake Porn Plot Intended to Silence Me." HuffPost UK, 21 Nov. 2018, www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316.

[28] Jankowicz, Nina. "I Shouldn't Have to Accept Being in Deepfake Porn." The Atlantic, 26 June 2023, www.theatlantic.com/ideas/archive/2023/06/deepfake-porn-ai-misinformation/674475.

[29] "Forecasting Potential Misuses of Language Models for Disinformation Campaigns and How to Reduce Risk." OpenAI, 11 Jan. 2023, openai.com/index/forecasting-misuse.

[30] "Home." Graphika, graphika.com. Accessed 26 June 2024.

[31] "Deepfake It Till You Make It." Graphika, Feb. 2023, public-assets.graphika.com/reports/graphika-report-deepfake-it-till-you-make-it.pdf.

[32] "Freedom of the Net 2023: The Repressive Power of Artificial Intelligence." Freedom House, 2023, freedomhouse.org/report/freedom-net/2023/

Case Study

GLOBAL | AN EXAMPLE OF AI-POWERED DISINFORMATION GENERATORS

In September 2023, an engineer using the alias Nea Paw published a YouTube video detailing their CounterCloud project, an experiment where they trained an LLM to produce false and misleading content.[33] According to The Debrief, the model was designed to scour the internet for content, and once a piece was selected, it prepared counter-articles attributed to fake journalists invented by the system.[34] This synthetic media was then posted to CounterCloud's WordPress-based website, where the AI model then created fake personas to comment on specific pieces to simulate user engagement. Once articles had been created, the AI would then prepare to share links on X (formerly Twitter) and create a series of oftentimes controversial posts to drum up interest. While this experiment was conducted in a closed environment, meaning that articles, personas, and posts never became publicly available, Nea Paw's work drew attention to the ease with which GenAI can be used to quickly produce, disseminate, and draw interest to convincing disinformation. Although evidence of these tactics is still limited, CounterCloud stands as a cheap and scalable model that autocracies may try to emulate.

Widening of Digital Divides

According to the International Telecommunications Union's (ITU) Measuring Digital Development: Facts and Figures Report released in December 2023, 2.6 billion people, or one third of the global population, remain offline.[35] In low-income countries, fewer people are online, and those who use far less data due to limited and unstable access and cost constraints are not able to achieve the full potential of connectivity, as Internet Society states.[36] With the continued rapid evolvement of GenAI risks and capabilities, those without the internet access, skills, resources, and tools to access, use, and leverage GenAI positively will fall further behind. Deployed with intentionality, GenAI solutions can improve access and inclusion and level the playing field for participation in the digital and political space, but the default tends to lead to exacerbation of existing inequalities. This negative trend can lead to political campaigns, civil society actors, and individuals around the world continuing to face accessibility gaps to using GenAI models. The gender digital divide can also be amplified, spanning from limited access to GenAI tools for women to women being disproportionately targeted by these tools. Additional factors resulting in widening digital divides—which may disproportionately affect some countries in the Global South, even though many of these challenges exist in Global North countries as well—include disparities in internet infrastructure, lack of adequate policy tools to regulate GenAI, and gaps in GenAI use due to limited literacy combined with GenAI's reliance on text.

Low Accountability in GenAI Development and Use Resulting in Reduced Trust

Policymakers and election administrators everywhere are struggling to keep pace with GenAI's development and use around elections and information environments. This is especially the case in some Global South countries that may face difficult infrastructure challenges, limited financial resources, a larger digital divide and a diversity of language barriers, foreign dependence on external technologies, or a lack of existing accountability frameworks and regulations, among other factors.

repressive-power-artificial-intelligence#generative-ai-supercharges-disinformation.

[33] "CounterCloud - AI Powered Disinformation Experiment." YouTube, uploaded by neapaw713,. 5 Sept. 2023, www.youtube.com/watch?v=cwGdkrc9i2Y.

[34] Banias, Mj. "Inside CounterCloud: A Fully Autonomous AI Disinformation System." The Debrief, 25 Nov. 2023, thedebrief.org/countercloud-ai-disinformation.

[35] "Measuring Digital Development – Facts and Figures 2023." ITU, 2023, https://www.itu.int/itu-d/reports/statistics/facts-figures-2023. Accessed 19 July 2024.

[36] Aguiar, João. "One-Third of the Global Population Remains Offline." Internet Society Pulse, 6 Dec. 2023, pulse.internetsociety.org/blog/one-third-of-the-global-population-remains-offline.

For instance, as seen in both the Global North and Global South, if a country already struggles to hold local stakeholders accountable for disinformation broadcast off and online, it may struggle greatly with creating accountability mechanisms for political campaigns that misuse GenAI in elections. Even though some efforts are being made to govern AI as referenced in the primer, basic guardrails are not yet in place, leaving policymakers without a roadmap or guidance on developing effective GenAI regulations and governance. Even more concerning is the fact that absent frameworks that encourage rights-based, rights-respecting policies, a lack of frameworks may result in governments passing legislation that does more harm than good. This gap and the lack of accountability around GenAI use and misuse will create opportunities to further incentivize malign political actors, including foreign and domestic governments, to take advantage of the cheap and easy access to GenAI tools to generate manipulated content to distort the information ecosystem and sow distrust.

> "The roadmap [for addressing global digital divides worsened by GenAI] requires taking both economic and social impact together. It starts from the grassroots level and goes through all the different realms. Everyone must have a say in this conversation."
>
> Lavina Ramkissoon (RAIN-Africa)

Increased Privacy Concerns

Companies building GenAI systems rely on large amounts of data about and from individuals, including copyrighted content and data that contains personal information. The rapid advancement and use of these systems without guiding policy regulations increases privacy risks to individuals and may cause the public to distrust government, industry, and civil society uses of GenAI, which often overlap with AI-powered surveillance that some governments use. According to the Electronic Privacy Information Center, a public interest research center based in Washington, DC, marginalized communities are disproportionately harmed by data collection practices and privacy abuses from both the government and private sector, and exploited through surveillance, policing, and algorithmic bias when the use of AI and machine learning is involved.[37] Data collection by authoritarian-leaning governments on their own citizens could also be used for malign purposes to create authoritarian LLMs[38] and GenAI tools to manipulate their citizens as a method of control. Data privacy is seldom a topic raised with the public and its importance not made known, and there are many countries where privacy is often considered a luxury rather than a right.

Overfocusing on Technical Solutions to GenAI Risks

The solutions most often touted to address some of the threats described above are technical solutions that focus on adapting or tweaking GenAI tools. For example, the often-discussed idea of relying on content provenance,[39] which helps people tell what content is authentic and what has been AI-generated, is often considered to be a technical fix for a technical problem that does not account for the sociopolitical nature of disinformation. The tendency to focus on technical solutions to GenAI issues—such as through model training to reduce bias,[40] implementing safeguards, or developing tools to detect GenAI disinformation —de-emphasizes the numerous political, regulatory, and social solutions that are also on the table.

---

[37] "Privacy and Racial Justice." Electronic Privacy Information Center, epic.org/issues/democracy-free-speech/privacy-and-racial-justice. Accessed 26 June 2024.

[38] The Decoder reported that in December 2023, the Chinese government released a dataset to train language models that reflect their political views and in line with its polices. See: Bastian, Matthias. "CCP Releases Politically Approved LLM Dataset With 50 Billion Tokens." The Decoder, 29 Dec. 2023, the-decoder.com/ccp-releases-politically-approved-llm-dataset-with-50-billion-tokens.

[39] Content provenance refers to the basic, trustworthy facts about the origins of a piece of digital content (including images, videos, audio recordings, and documents). It may include information such as who created the content and how, when, and where it was created or edited. The content author always has control over whether provenance data is included as well as what data is included. Included information can be removed in later edits. Provenance also allows for anonymous content. See: "FAQ." Coalition for Content Provenance and Authenticity, c2pa.org/faq. Accessed 26 June 2024.

[40] Training a large language model involves teaching it to understand and generate text by using a huge amount of written material, adjusting it to perform specific tasks or account for bias, and then testing and using it to do things like answer questions or write content. Model training to reduce bias can include using a diverse and representative dataset, removing sensitive variables, using bias mitigation techniques, and using human oversight.

Some of these technical solutions, such as establishing flagging systems for concerning content, can additionally put the onus on targeted individuals to seek recourse or these mechanisms can even be manipulated for use against victims. Many nontechnical solutions may ultimately be more effective, including public education campaigns, fact-checking networks, independent media's role in building norms around truthful and accessible information, and regulations for GenAI companies. If one were to only focus on technical solutions to challenges that GenAI creates, they may risk failing to address threats holistically.

"This strategy [of investing in media literacy efforts] will have more staying power in the long term, rather than hanging onto AI-detection tools that may not be effective. Supporting people to do their due diligence when dealing with disinformation may be the best way to inoculate the public at large."

Industry Representative from the United States

OTHER MAJOR FACTORS THAT LEAD TO INCREASED GENERATIVE AI RISKS TO DEMOCRACY

To better understand the risks outlined above, IRI has compiled a short list of factors that create enabling environments for those risks to emerge. The goal is to help readers anticipate and minimize enabling factors and be better prepared in environments that are likely to allow some of the risks to present themselves.

Lack of Risk Awareness

In many countries, policymakers, politicians, election officials, and other civil society actors may be unaware of existing research on GenAI and its potential impacts. This lack of awareness contributes to a weak understanding of risks generally or knowledge of existing risks and harms. GenAI models are also intrinsically opaque and difficult to understand, from the data sources of the LLMs used in their development to understanding how and why a certain output is generated. Challenges for government to procure GenAI systems, outsourcing GenAI system development and integration (such as a government using private contractors for GenAI-powered public service delivery), and leaving decisions to ill-informed public officials can be contributors to or outgrowths from this same problem. As decision-makers continue to struggle to develop AI literacy across the globe, especially in countries struggling to overcome skills gaps, the gap in understanding is widening, magnifying an existing digital divide. Concerns from partners in Global South countries around existing digital divides have contributed to concerns about those in the Global South being left behind.

Opacity in Design and Use

It is difficult for civil society groups, governments, and citizens to understand how GenAI operates, as many GenAI models are considered to be "black boxes."[41] This is further compounded by the fact that some platforms do not clearly articulate their GenAI content policies and enforcement actions, what training data is used, and how models are designed or tested. It is also difficult for civil society actors and citizens to pursue accountability measures for GenAI used by governments and political campaigns when said use is often secretive or outsourced to private companies. Further complicating this situation is the absence of a clear definition of what transparency means in GenAI, and deployment. It could mean everything from disclosing how a model is built or publishing what training data it relies on, to giving the public more information about its design, development, deployment, and maintenance. The lack of established understanding of what equates to transparency leaves every private company to define transparency for themselves.

[41] The lack of transparency and understanding of how AI systems make their decisions by the developers and users alike results in GenAI systems often being referred to colloquially as a "black box." As explained in a Vice article, "the people who develop AI are increasingly having problems explaining how it works and determining why it has the outputs it has." See: Xiang, Chloe. "Scientists Increasingly Can't Explain How AI Works." Vice, 1 Nov. 2022, https://www.vice.com/en/article/y3pezm/scientists-increasingly-cant-explain-how-ai-works.

Value Gaps

There are major gaps and variations between AI developers' guiding values and the values of the billions of users around the world, so it is not possible for global GenAI tools to reflect such a multitude of cultures and values. Many tools are developed without diverse inputs, making it difficult or impractical for the many different potential end users—who represent various cultures, languages, and histories— to widely use a GenAI tool without localizing it first. There is extensive variation among countries like Brazil, India, Mexico, and Pakistan, to give just a few examples, that is not captured in nonlocalized, "global" GenAI tools. These gaps impact how GenAI is built, used, and regulated in important, context-relevant ways, and these gaps will remain until tools are localized by the many different people who hope to use them.

> "Governments have an extra responsibility for explainability and transparency, and they should keep these priorities in mind before launching anything."
>
> Nicklas Badum (Democracy X)

Companies' Varied Policies and Moderation Challenges

Most technology companies either have nascent, largely untested policies on the use of GenAI content in political contexts or lack such policies altogether. Refer to the below pop-out box for examples of the more popular GenAI tools and their relevant policies. The companies' policies are wide-ranging and are updated on an ad hoc basis, making it challenging for actors to keep track and monitor. Sometimes it is also unclear if these companies' policies apply globally or only in select countries. Additionally, the companies' elections related queries and outputs policies (see below pop-out) including sources to which they cite, mostly pertain to Global North elections and largely ignore the rest of the world. Historically, companies have also underinvested in content moderation in languages other than English. This underinvestment, combined with mass layoffs of the trust and staff staff, mean that GenAI is especially impactful and risky in a year with many elections taking place in countries where English is not the predominant language. Finally, it is important to note that it is extremely difficult to monitor the accuracy of outputs produced by GenAI tools given the variety of and inconsistency in responses that tools often provide. For example, an AI Democracy Projects workshop probing how leading AI models respond to queries that voters might ask in January 2024 revealed that answers from five leading AI models at the time (Anthropic's Claude, Google's Gemini, OpenAI's GPT-4, Meta's Llama 2, and Mistral's Mixtral) were often inaccurate, misleading, and even downright harmful.[42] This will likely be worse with non-English text inputs or generating non-English text outputs because lower availability of data for languages other than English can result in lower reliability of non-English GenAI models.

Overview of Current Content Policies for Widely Used GenAI Tools

Updated as of June 2024

Below are a few of the GenAI content generators that are most widely used globally, along with a list of their policies and trusted resources related to elections, political use, and labelling of GenAI content and other key efforts. This list is not comprehensive and focuses on tools that have at least some form of publicly available policies; the list does not capture all actions platforms are taking to address this problem. Note that there are numerous other GenAI tools created in other regions and countries that are also widely used but may only be accessible to people living in those localities. Brookings published an informative 2023 article on GenAI tools and markets within the Global South.[43]

[42] Angwin, Julia. "Seeking Reliable Election Information? Don't Trust AI." Proof, 27 Feb. 2024, www.proofnews.org/seeking-election-information-dont-trust-ai/?ref=platformer.news.

[43] Okolo, Chinasa T. "AI In the Global South: Opportunities and Challenges Towards More Inclusive Governance." Brookings, 1 Nov. 2023, www.brookings.edu/articles/ai-in-the-global-south-opportunities-and-challenges-towards-more-inclusive-governance.

Please note these listed policies and trusted resources were updated at the time of writing, but they may shift as GenAI and political and regulatory landscapes continue to evolve. We recommend regularly checking the tool's policy webpages. In reading through the policies and resources listed below, there is great variation of how politically sensitive issues and civic processes are (or are not) addressed. Finally, keep in mind that as government regulations continue to emerge, they may impact company policies and require changes, updates, and potentially greater standardization.

OpenAI ChatGPT: 1.8 billion monthly visits[44]

In early 2024, OpenAI said ChatGPT will refer users to CanIVote.org for any queries related to the US elections. The company's policy also forbids impersonation of candidates, prohibits using its tools for campaigning, lobbying, discouraging voting, or misrepresenting the voting process.[45]

OpenAI also announced[46] its referral of users to the European Parliament's official source of voting information for the 2024 European Parliament elections.

OpenAI announced it joined the Steering Committee of the Coalition for Content Provenance and Authenticity (C2PA)[47] in May 2024 and began adding C2PA metadata in ChatGPT to denote AI-generated images with a CR symbol.

Google Gemini (formerly Bard): 414.4 million monthly visits

In December 2023, Google announced that it would avoid US election-related queries, referring users to Google Search instead (this applies to all queries and outputs, not just text).[48]

In March 2024, Google announced[49] that it will also restrict the types of election-related queries for which Gemini will return responses regarding the Indian elections and shared that the company is working with the Election Commission of India (ECI) to surface quality information.[50]

In May 2024, Google announced the expansion of DeepMind's SynthID[51] capabilities to check not only images but also Gemini (text) and Veo (video).

[44] Monthly visit information, referring to the average number of visits within a four-week period, were provided by Similarweb, a data aggregation company whose specialty is in website traffic and performance. Figures account for global users and reflect data from April 2024.

[45] "How OpenAI Is Approaching 2024 Worldwide Elections." OpenAI, 15 Jan. 2024, openai.com/index/how-openai-is-approaching-2024-worldwide-elections.

[46] "How OpenAI Is Approaching 2024 Worldwide Elections."

[47] C2PA is a widely used standard for digital content certification, developed and adopted by a wide range of actors including software companies, camera manufacturers, and online platforms. See: "FAQ." Coalition for Content Provenance and Authenticity, c2pa.org/faq. Accessed 26 June 2024.

[48] Jasper, Susan. "How We're Approaching the 2024 U.S. Elections." Google The Keyword, 19 Dec. 2023, blog.google/outreach-initiatives/civics/how-were-approaching-the-2024-us-elections.

[49] "Watermarking AI-generated Text and Video With SynthID." Google DeepMind, 14 May 2024, deepmind.google/discover/blog/watermarking-ai-generated-text-and-video-with-synthid.

[50] Google India Team. "Supporting the 2024 Indian General Election." Google India Blog, 12 Mar. 2024, blog.google/intl/en-in/company-news/outreach-initiatives/supporting-the-2024-indian-general-election.

[51] "Watermarking AI-generated Text and Video With SynthID." Google DeepMind, 14 May 2024, deepmind.google/discover/blog/watermarking-ai-generated-text-and-video-with-synthid.

Perplexity AI: 70.54 million monthly visits

AI search engine Perplexity AI announced that its algorithms prioritize reliable and reputable sources, such as news outlets, and always provides links so users can verify its output.[52]

Perplexity AI's terms of service state that it is not allowed to be used for unlawful, illegal, fraudulent, or harmful activities, which includes political manipulation.[53] No additional details are provided on elections.

Anthropic Claude: 65.97 million monthly visits

Anthropic's acceptable use policy[54] prohibits[55] the use of its tools in political campaigning, lobbying, inciting/facilitating the disruption of electoral and civic processes, discourage voting, etc. Candidates will not be allowed to use Claude to build chatbots that can pretend to be them or for targeted political campaigns.

Claude is prevented from answering political questions from US users, instead directing users to TurboVote, a nonprofit platform encouraging Americans vote by providing reliable, useful voting guidance.

Microsoft Copilot (formerly Bing): 34.78 million monthly visits

In February 2024, Microsoft will use media provenance capabilities based on cryptographic methods to mark and sign AI-generated content with metadata about its source and history, through the use of C2PA's digital credentials to denote AI-generated images with a special symbol.[56]

In April 2024, Microsoft announced the expansion of its Content Integrity tools to EU political parties and campaigns and news organizations from around the world.[57] This allows Content Credentials[58] to be added to original content and gives anyone access to a website[59] to check content for the existence of Content Credentials.

Microsoft will also ban users who create and distribute deceptive and abusive content from their services such as LinkedIn, Microsoft's gaming network, and other relevant services.

Midjourney: 17.04 million monthly visits

Midjourney, an AI image-generator, does not allow its tool to be used to generate images for political campaigns or to influence the outcome of an election.[60]

[52] Davis, Wes. "How AI Companies Are Reckoning With Elections." The Verge, 19 Mar. 2024, www.theverge.com/2024/3/19/24098381/ai-chatbots-election-misinformation-chatgpt-gemini-copilot-bing-claude?ref=platformer.news.

[53] "Terms of Service." Perplexity, 4 June 2024, www.perplexity.ai/hub/legal/terms-of-service.

[54] "Usage Policy." Anthropic, 6 June 2024, www.anthropic.com/legal/aup.

[55] "Preparing for Global Elections in 2024." Anthropic, 16 Feb. 2024, www.anthropic.com/news/preparing-for-global-elections-in-2024.

[56] Smith, Brad. "Combating Abusive AI-generated Content: A Comprehensive Approach." Microsoft On the Issues, 13 Feb. 2024, blogs.microsoft.com/on-the-issues/2024/02/13/generative-ai-content-abuse-online-safety.

[57] Hutson, Teresa. "Expanding Our Content Integrity Tools to Support Global Elections." Microsoft On the Issues, 22 Apr. 2024, blogs.microsoft.com/on-the-issues/2024/04/22/expanding-our-content-integrity-tools-to-support-global-elections.

[58] Content Credentials reveal information about the origin and history of content available for everyone to access. It lets users understand how much editing or manipulation the content went through. See "Home." Content Credentials, https://contentcredentials.org/. Accessed 26 June 2024.

[59] "Content Integrity." Microsoft, contentintegrity.microsoft.com. Accessed 26 June 2024.

In March 2024, reports showed that Midjourney has started blocking users from creating fake images of the two leading candidates of the 2024 US elections. The company has not yet outlined its policy changes as they relate to the upcoming elections.[61]

Meta AI: 6.53 million monthly visits

Meta announced in November 2023 that it would require political advertisers to disclose if they used AI or other digital techniques to create ads published on its platforms.[62]

Meta has also banned the use of its GenAI tools by political campaigns and advertisers.[63] Meta announced in April 2024 that it will begin labeling a wider range of content as "Made with AI" when it detects industry standard AI image indicators or people self-report.[64]

Mistral's Le Chat: 4.78 million monthly visits

Mistral's terms of service do not contain any policies related to elections or political manipulation, but specifically notes the prohibition of the generation of hateful, harassing, or violent outputs.[65]

Stability AI (Stable Diffusion 3): 3.12 million monthly visits

Stability AI, which created text-to-image AI generator Stable Diffusion, updated its acceptable use policy in March 2024 and prohibits the creation or promotion of disinformation, large-scale disinformation campaigns, political advertisements, propaganda, and influence campaigns.[66]

It is also important to note that in February 2024, major technology companies came together to unveil a tech accord to combat deceptive use of AI in 2024 elections.[67] The list of companies that joined this accord include Amazon, Google, TikTok, among others.[68] The agreement is voluntary and includes commitments from technology companies ranging from creating watermarking technology, detecting and labelling AI-generated content, and assessing the models that underlie AI software to identify risks of abuse and to raise public awareness and education about AI[69]. However, the accord lacks an accountability mechanism, and its focus seems to be short term, applying only to elections taking place in 2024.

[60] "Terms of Service." Midjourney, 7 Mar. 2024, docs.midjourney.com/docs/terms-of-service.

[61] O'Brien, Matt. "AI Image-generator Midjourney Blocks Images of Biden, Trump." Associated Press, 13 Mar. 2024, apnews.com/article/midjourney-ai-imagegenerator-biden-trump-deepfakes-bc6c254ddb20e36c5e750b4570889ce1.

[62] Clegg, Nick. "How Meta Is Planning for Elections in 2024." Meta, 28 Nov. 2023, about.fb.com/news/2023/11/how-meta-is-planning-for-elections-in-2024.

[63] Paul, Katie. "Meta Bars Political Advertisers From Using Generative AI Ads Tools." Reuters, 7 Nov. 2023, www.reuters.com/technology/meta-bar-political-advertisers-using-generative-ai-ads-tools-2023-11-06.

[64] Bickert, Monika. "Our Approach to Labeling AI-Generated Content and Manipulated Media." Meta, 5 Apr. 2024, about.fb.com/news/2024/04/metas-approach-to-labeling-ai-generated-content-and-manipulated-media.

[65] "Legal Terms and Conditions." Mistral AI, mistral.ai/terms. Accessed 26 June 2024.

[66] "Acceptable Use Policy." Stability AI, stability.ai/use-policy. Accessed 26 June 2024.

[67] "AI Elections Accord: A Tech Accord to Combat Deceptive Use of AI in 2024 Elections." AI Elections Accord, www.aielectionsaccord.com. Accessed 26 June 2024.

[68] "AI Elections Accord: A Tech Accord to Combat Deceptive Use of AI in 2024 Elections."

[69] Bond, Shannon. "Tech Giants Pledge Action Against Deceptive AI in Elections." NPR, 16 Feb. 2024, www.npr.org/2024/02/16/1232001889/ai-deepfakes-election-tech-accord.

GENERATIVE AI'S OPPORTUNITIES FOR DEMOCRACY

Despite these threats, GenAI, when used appropriately and responsibly, may provide significant benefits to democracies across the globe. Although GenAI has only recently scaled to become widely accessible, democratic actors—particularly decision-makers and civil society representatives—are beginning to take advantage of GenAI in novel and innovative ways. GenAI solutions may make a difference in promoting inclusive decision-making, forecasting policy impacts, and prompting upskilling efforts, to name just a few examples. This section delves deeper into how GenAI can be a force for good within democratic societies, given certain precautions are met.

Strengthening Strategies to Protect Information Environments

Much has been said about GenAI's ability to create disinformation at scale. However, GenAI may also enable democratic actors to address misleading content more efficiently and effectively. Examining the ways GenAI can be used to strengthen information integrity is premised on protecting individual rights to freedom of expression. A healthy and robust space does not mean shutting down speech, but rather using technology to help advance democratic freedoms and principles. Tools powered by GenAI, such as those made by US-based company Reality Defender,[70] are part of this effort by verifying the authenticity of audio, visual, or image-based content, enabling decision-makers and civil society representatives to identify fabricated content more easily. In these early cases, GenAI makes it simpler to label, remove, and ultimately slow the spread of mis/disinformation. Additionally, GenAI could be leveraged to automate this work, allowing democratic actors to quickly address falsehoods shared by malign actors. Fact-checking systems, which can sort through vast quantities of information, may be able to spot mis/disinformation in ways that cannot be replicated by humans. GenAI-powered models can even provide average users with accurate counterpoints to help debunk false information, a potential application highlighted by researchers from Purdue University's Governance and Responsible AI Lab (GRAIL)[71] in a 2023 article published by The Conversation.[72] The examples listed here are just the beginning when it comes to how GenAI could protect information integrity. As the field continues to evolve, these GenAI-powered tools may play a larger role in how democratic actors approach the complex topic of information integrity.

Powering Responsive and Efficient Decision-Making

For many years, digital solutions have been created to improve collaboration between decision-makers, their partners in civil society, and the public at-large to varying degrees of success. Now that GenAI is more accessible, decision-makers and civil society stakeholders are beginning to see how Gen AI-powered tools could strengthen past efforts, increase their efficacy, and scale their impact. Participatory GenAI tools are especially promising when considering the diverse ways they could ensure more stakeholders are heard. GenAI may be used to simplify complex policy documents into text that can be easily interpreted and used by decision-makers. It could digest vast quantities of data about distinct demographics and communities, providing novel insights that can be used to create more responsive policymaking. GenAI could also be integrated into chatbots that can efficiently answer questions from a variety of constituents, including those with limited digital know-how or understanding of governance. It could even power platforms that synthesize feedback from constituents in a way that helps officials make decisions. GenAI has immense possibilities to make the work of governance more open, inclusive, and responsive to all stakeholders if designed and applied democratically. Many in the democracy community have early optimism about how GenAI could be used to build more participatory, responsive and efficient systems if the appropriate guardrails are in place.

[70] "From Dating to Democracy, AI-generated Media Creates Multifaceted Risks." Deloitte Insights, 19 Oct. 2023, www2.deloitte.com/xe/en/insights/focus/tech-trends/2024/reality-defender-helping-companies-with-ai-content-detection.html.

[71] "Governance and Responsible AI Lab (GRAIL)." Purdue University, www.cla.purdue.edu/academic/polsci/research/labs/grail/index.html. Accessed 26 June 2024.

[72] Schiff, Daniel S., and Kaylyn Jackson Schiff. "Generative AI Like ChatGPT Could Help Boost Democracy – if It Overcomes Key Hurdles." The Conversation, 7 Nov. 2023, theconversation.com/generative-ai-like-chatgpt-could-help-boost-democracy-if-it-overcomes-key-hurdles-212664.

Case Study

TAIWAN | FACILITATING COLLABORATIVE POLICYMAKING

In 2015, decision-makers in Taiwan used a GenAI solution to solicit the opinions of citizens while crafting regulations for ride-sharing services. Polis,[73] a dynamic platform that collects user feedback and identifies shared concerns, was deployed to increase citizen engagement in the decision-making process. Individuals were able to share their thoughts on what policies would be effective, which the platform analyzed to identify areas of consensus. After completing this step, Polis then generated a set of recommendations that guided the design of regulations, which are published in a 2019 Centre for Public Impact report.[74] The success of this experiment has led leaders in Taiwan to continue exploring the place of GenAI in policymaking processes. In 2023, the country's Ministry of Digital Affairs joined the Collective Intelligence Project's Alignment Assemblies initiative, the goal of which is to gather citizen feedback on AI-related risks and share these concerns with industry leaders. Looking at these examples, Taiwan's use of Polis underscores how GenAI can help create responsive policies that address citizens' most pressing needs.[75]

Enhancing Organizational Capacity

Decision-makers across the globe often lack the time, resources, and personnel to adequately respond to the needs of their constituents. Similarly, civil society organizations are hamstrung by unreliable funding, capacity constraints, and logistical challenges to navigating time-intensive tasks. Early findings indicate that GenAI may be part of the solution to these pressing problems, allowing democratic actors to do more with less. GenAI could be used by decision-makers and civil society to prepare communications materials, such as talking points or draft emails, that are suitable for both internal and external audiences. It could be leveraged to summarize and synthesize content, pulling out insights that are relevant to the work of decision-makers and civil society. In some cases, GenAI could be used to improve the delivery of services to constituents needing assistance, such as those experiencing food insecurity, thereby automating processes that may have been inefficient.

For instance, Viamo, a voice-based platform designed for low-connectivity areas, has incorporated GenAI[76] to help answer user queries on healthcare services. While GenAI's efficacy is dependent on users having the required skills and knowledge to be able to use it, GenAI offers promise for augmenting the work of trained professionals navigating a web of limitations and restrictions.

> "Generative AI provides an opportunity for conversational government. Chatbots, for example, enable citizens' questions to be answered 24/7. Instead of having to wait in long queues or complete paper-based forms, they can receive immediate help."
>
> Beth Simone Noveck (Burnes Center for Social Change at Northeastern University)

[73] "Home." Polis, pol.is/home. Accessed 26 June 2024.

[74] "Building Consensus and Compromise on Uber in Taiwan." Centre for Public Impact, 18 Sept. 2019, www.centreforpublicimpact.org/case-study/building-consensus-compromise-uber-taiwan.

[75] Cushing Rodriguez, Sebastian. "Consensus Building in Taiwan, the Poster Child of Digital Democracy." Democracy Technologies, 4 Oct. 2023, democracy-technologies.org/participation/consensus-building-in-taiwan.

[76] Viamo. "Ask Viamo Anything." Viamo, viamo.io/ask-viamo-anything-ai. Accessed 26 June 2024.

Case Study

GLOBAL | BROADEN ACCESS TO GOVERNMENT SERVICES

Many decision-makers and civil society groups are considering how GenAI could expand or simplify citizens' ability to access services. One early example of this is the use of GenAI to better help survivors of domestic violence obtain the support they need. In 2022, Swiss nonprofit Spring ACT released their Sophia chatbot,[77] which collects evidence submitted by users, provides them with up-to-date information on their rights, and offers them a menu of options based on their responses. A similar product named SARA[78] was created in 2023 by Infosegura, a strategic alliance between the United Nations Development Programme (UNDP) and the United States Agency for International Development (USAID). Like Sophia, SARA is a tool that provides women and girls who have experienced violence to access information about gender-based violence as well as available institutional and civil society care services. It also offers legal advice, as well as support in crafting emergency plans among other services. Solutions like these serve as early examples for how civil society can use GenAI to help individuals understand what services are available to them in moments of crisis.

Case Study

ARGENTINA | SIMPLIFYING DAILY TASKS

The Prometea project,[79] which resulted from a collaboration between the University of Buenos Aires' Faculty of Law and the Buenos Aires Attorney General's Office, is a GenAI-powered system designed to reduce the strain placed on civil servants in the municipal judiciary system. In an article written by the leader of the university-based team, this novel tool achieves this in several ways.[80] Firstly, Prometea acts as a virtual assistant to civil servants by answering questions about certain laws or statutes. Its developers note that their solution can even respond to specific queries made through voice commands. Additionally, the system can complete, and even generate, legal documents without input from users. At the same time, it can classify large quantities of these resources in a fraction of the time it would normally take a person to do so. Innovations like Prometea can be invaluable to organizations dealing with capacity constraints. With human involvement, as well as proper oversight, legal professionals around the world may see significant efficiency gains.

Improving Institutional Transparency

A common criticism of GenAI solutions is that they are "black boxes," meaning that how and why GenAI tools conduct analysis and produce outputs are impossible to determine. Although GenAI itself is difficult to fully understand, some in the democracy community are tentatively optimistic about its potential to be leveraged as a tool that strengthens transparency efforts. There are a few ways this could be done. It could create simple visualizations of complex data used by decision-makers in clear, understandable ways. GenAI could present this information in a dynamic format, allowing citizens to interpret confusing data more easily while showing them how this information does (or does not) inform governments' resource allocation and decision-making, among other key tasks.

[77] "Sophia Chatbot." Spring ACT, springact.org/sophia-chatbot. Accessed 26 June 2024.

[78] "Home." Sistema De Atención Y Resguardo De Información Automatizada (SARA), chatbotsara.org. Accessed 26 June 2024.

[79] Ast, Federico. "Prometea, Artificial Intelligence in the Judicial System of Argentina." Medium, 14 Dec. 2021, medium.com/astec/prometea-artificial-intelligence-in-the-judicial-system-of-argentina-4dfbde079c40.

[80] Corvalan, Juan Gustavo. "Prometea: Artificial Intelligence to Transform Justice and Public Organizations." International Journal of Digital and Data Law, vol. Vol 6, 2020, core.ac.uk/download/pdf/322501055.pdf.

It could also make information stored in expansive databases used by decision-makers more accessible. For instance, DC Compass, a chatbot created by Washington, DC officials answers users questions about how their data is managed.[81] If applied and designed democratically, there is promise for GenAI to shine a light on the inner workings of governance systems and increase transparency and access to important information, including for citizens with low digital and data literacy in the Global North and South.

Case Study

INDIA | INCREASING GOVERNMENT OPENNESS

Launched in 2023, the Digital Sansad app[82] was released by decision-makers in India as a way to increase transparency within parliament. One notable feature of this GenAI-powered solution is its ability to automatically transcribe government proceedings into 22 languages via automated speech recognition techniques.[83] The team behind Digital Sansad enables users to open the app and read a transcript of deliberations within a matter of minutes, making the information and process of governance much more accessible to individuals across the country. Additionally, developers of the tool plan to integrate both chat- and voice-based bots that will answer constituent questions. Users could easily learn more about their decision-makers work at any time, even if their level of digital literacy may be relatively low. For countries where digital divides create barriers to accessibility and engagement for many citizens, solutions like Digital Sansad could make a significant difference in increasing citizens' understanding of governmental processes.

Scaling Civic Education Efforts

Decision-makers and civil society groups are experimenting with how GenAI can keep the public informed about how governance works. For example, there is promise for democratic actors to leverage GenAI solutions to create learning modules that can be easily tailored and responsive to citizen inputs. Content in these programs can be altered to meet the individual's knowledge gaps, such as what companies like Teachflow[84] are considering to enhance political science and civic education in school settings. Civic education modules could inform users about governance processes, as well as how they could meaningfully participate.

Additionally, GenAI could be used to address individuals' policy-related questions. For example, chatbots[85] powered by GenAI have been used to answer questions about certain departments or regulations, provide users with up-to-date information, and direct them to relevant resources. Through these applications, GenAI provides individuals with the knowledge they need to meaningfully participate in governance processes. While guardrails should be in place to prevent sharing disinformation, these use cases show how GenAI could help ensure the larger public is civically engaged and accurately informed.

[81] Edinger, Julia. "New Washington, D.C., Tool Uses Generative AI to Make Data Accessible." GovTech, 26 Mar. 2024, www.govtech.com/artificial-intelligence/new-washington-d-c-tool-uses-generative-ai-to-make-data-accessible.

[82] "Home." Digital Sansad, sansad.in/phistory. Accessed 26 June 2024.

[83] Mathur, Swati. "'Digital Sansad' to Use AI to Transcribe House Proceedings Real-time." The Times of India, 26 Apr. 2023, timesofindia.indiatimes.com/india/digital-sansad-to-use-ai-to-transcribe-house-proceedings-real-time/articleshow/99770031.cms.

[84] "The Role of AI in School Political Science and Civics." Teachflow, 25 Jan. 2023, teachflow.ai/the-role-of-ai-in-school-political-science-and-civics.

[85] "KI-Bürgerassistenz Lumi Soll Nutzerinnen Und Nutzer Kennenlernen." Heidlelberg, 24 Apr. 2022, www.heidelberg.de/HD/Presse/24_10_2022+ki-buergerassistenz+lumi+soll+nutzerinnen+und+nutzer+kennenlernen.html.

Case Study

EUROPEAN UNION | IMPROVING CITIZEN/DECISION-MAKER DYNAMICS

Launched in September 2022, the TITAN project[86] is a multi-stakeholder initiative launched to assist citizens as they encounter information on the internet that may be false or misleading. TITAN is a GenAI-powered chatbot designed to guide users in gauging the accuracy and reliability of information found on social media. One goal of this project is to address low levels of trust in institutions. The chatbot will provide users with fact-checked information on how decision-makers work for them. The TITAN team hopes that average citizens might gain trust in their decision-makers, become more knowledgeable about key institutions, and feel more inclined to participate in civic life. Results from this project are not yet available, and it has yet to be determined if TITAN can truly strengthen trust. As with the integration of any GenAI tools, there may be unexpected risks or harms that have yet to be seen. Continued observation of this effort will help demonstrate the opportunities—and limits—to using these tools for civic education and trust-building efforts.

Creating Adaptable Cyber Defenses

Advocates for democracy, whether they are decision-makers or civil society stakeholders, have long contended with cyberattacks designed to derail their work. While GenAI has already begun to play a role in campaigns launched by authoritarian and authoritarian-leaning actors, it is important to note that GenAI could also help enhance defensive countermeasures. Early research has shown that GenAI could be used to track potential threats, identify exploitable vulnerabilities, and even conduct forensic analyses among other critical tasks needed for robust cyber defense.[87] It could also offer critical support to low-capacity stakeholders facing regular attacks. For instance, it could power an informative chatbot that provides threat intelligence to help democratic actors better manage their limited resources when addressing pressing cyber threats. Outside of emergencies, it could also streamline the work needed to fend off cyberattacks, reducing the burden placed on decision-makers and civil society organizations. As another example, GenAI could provide guidance on how to ward off cyberattacks through simulations, preparing exercises that mimic malign actors' tactics and strategies to help prepare decision-makers and civil society leaders. Though guidelines are needed on GenAI's use for cybersecurity before it is leveraged for any of these potential use cases, its ability to proactively deal with digital threats is promising.

Case Study

UKRAINE | TRACKING POTENTIAL CYBERATTACKS

Entrepreneurial start-ups in the tech space have been paying attention to how GenAI solutions could be used to combat the online activities of authoritarian or authoritarian-leaning actors. One such example of this can be seen from the work of Mantis Analytics, a Ukraine-based group that has created a platform that monitors information shared online, with the goal of exposing falsehoods circulating across social media networks. According to the company, their solution can identify coordinated influence operations found online, all while preparing content that corrects and clarifies falsehoods. It also can detect disinformation, as the system is trained to recognize common manipulation techniques.

[86] "Home." Titan, www.titanthinking.eu. Accessed 26 June 2024.

[87] Ferrag, Mohamed Amine, et al. "Generative AI and Large Language Models for Cyber Security: All Insights You Need." arXiv, 21 May 2024, https://doi.org/10.48550/arXiv.2405.12750.

The platform also provides real-time analysis of information trends on an interactive dashboard, condensing online noise to make it more digestible for cybersecurity professionals. Mantis Analytics has assisted the National Security and Defense Council of Ukraine[88] on cyber defense matters and has since received substantial funding from the US-based Alchemist Accelerator.[89] Such examples indicate how GenAI, if deployed responsibly, can help democracies fend off disruptive attacks in the digital domain.

Predicting Potential Policy Outcomes

Decision-makers, such as those in the State of California, have considered how GenAI solutions could be deployed to assess the potential impacts of policy solutions before they are fully implemented.[90] For instance, decision-makers and civil society can use GenAI to map out different scenarios if a policy were to go into effect. In running predictive models to forecast what outcomes may arise, decision-makers could gain insight into adjustments that could be made to mitigate negative or unintended consequences.[91] Decision-makers could also leverage GenAI to spot emerging trends and patterns within larger datasets that may shape an intervention's overall impacts. Civil society could take advantage of GenAI to spot potential shortcomings in decision-makers' policy analyses. These applications demonstrate how GenAI could be key for proactively addressing policy issues before they arise. It would not only give decision-makers the opportunity to course correct before their constituents are impacted, but also provide a unique opening for civil society to offer a more nuanced critique of policy interventions.

Case Study

NIGERIA | MEETING COMMUNITY NEEDS

Sentiment analysis, which refers to a process where text-based content from many different sources is systematically assessed to determine general tone, offers promise for decision-makers interested in using GenAI to better understand their constituents' needs. The development of tools enabling sentiment analysis in the Global South, however, is often hamstrung by the lack of language resources for GenAI developers based in non-English-speaking regions. To address this gap, Masakhane, a grassroots community for natural language processing (NLP) experts throughout Sub-Saharan Africa, has sought to create a language corpus and lexicon for Hausa, Yoruba, and Igbo,[92] three of the most widely spoken languages in Nigeria. With this project, Masakhane seeks to generate the assets needed to enable the creation of tools to conduct sentiment analysis in their own communities. Applying GenAI for sentiment analysis could result in reshaping policies based on constituents' concerns before they are put into effect. Though developing language assets remains a challenge, Masakhane offers a strong example for how community-driven efforts can work to make innovative applications of GenAI for good possible.

[88] Sobachynskyi, Rostyslav. "Ukrainian-made AI Platform Mantis Analytics Detects Russian Fakes in the News." AIN.Capital, 15 Sept. 2023, ain.capital/2023/09/15/ukrainian-mantis-analytics-detects-russian-fakes.

[89] Laoun, Joy. "Ukrainian Mantis Analytics Secures $30k Investment From the Alchemist Accelerator." Vestbee, 11 Apr. 2024, www.vestbee.com/blog/articles/ukrainian-mantis-analytics-secures-a-30k-investment-from-the-alchemist-accelerator.

[90] State of California. State of California Report: Benefits and Risks of Generative Artificial Intelligence. Nov. 2023, www.govops.ca.gov/wp-content/uploads/sites/11/2023/11/GenAI-EO-1-Report_FINAL.pdf.

[91] Digital Twins Digest. "Simulations 2.0: The Role of Generative AI in Creating Accurate and Reliable Models." Medium, 26 June 2023, Archived Version, https://web.archive.org/web/20240604131848/https://medium.com/@thedigitaltwindigest/simulations-2-0-the-role-of-generative-ai-in-creating-accurate-and-reliable-models-b6e45d91d0ba.

[92] "NaijaNLP: Sentiment Lexicon and Hate Speech." Masakhane, www.masakhane.io/ongoing-projects/naijanlp-sentiment-lexicon-hate-speech. Accessed 26 June 2024.

RECOMMENDATIONS

Civil society organizations (CSOs), governments, and private sector companies—from democracy nongovernmental organizations (NGOs) and accountability watchdogs to election administrations to social media platforms—have many opportunities to maximize GenAI's potential to benefit democracy while mitigating its risks. This subsection lays out the multiple concrete recommendations that emerged from IRI's working group to align GenAI development, deployment, and use with democratic values. Recommendations are split across three key sectors: civil society, government, and industry. Note that while the recommendations are organized by sector, many of the recommendations speak to the cross-sectoral collaboration that is needed to effectively minimize the harms and maximize the opportunities that GenAI presents. Diverse voices and perspectives should be included in these efforts by default.

Across all areas, working group members were clear that CSOs, governments, and companies should make sure to conduct inclusive outreach and public consultations to seek out, listen to, and emphasize the perspectives of individuals, communities, and regions that face marginalization and inequities in society broadly. Even more importantly, it is critical that these perspectives are effectively incorporated into actions as a result of these inputs. This includes considering how factors including race, gender, age, socioeconomic status, ability, first language, country, and global region (e.g., Global South countries and variation within the Global South) impact GenAI's opportunities for and risks to those people.

RECOMMENDATIONS FOR CIVIL SOCIETY

Democracy NGOs, fact-checking organizations, accountability watchdogs (across media, politics, technology, and other areas), and other CSOs can take the following steps to help ensure GenAI is used democratically within their own organizations and by their own country's governments, companies, and citizens:

Before taking any other steps: upskill yourselves and your team on the basics of generative AI. It is important to understand what exactly GenAI is and the risks and threats it might present to see where, if at all, it might be useful to one's work. Remember that leveraging GenAI might not be applicable to all organizations, and that many solutions to governance or societal challenges are analog rather than digital. Refer to the primer section earlier in the report, which lists educational resources and explanatory documents to help build up a basic understanding of GenAI-enhanced tools. After learning the basics, consider how to **integrate GenAI into the work of your organization or educating and advocating about the risks and opportunities** GenAI presents. The below recommendations speak to and are segmented by those two needs.

Integrating and Leveraging GenAI in Your Organization

Refer to rights-based GenAI use frameworks to develop internal GenAI use policies. Leaning on frameworks like the US[93] and New Zealand[94] guidance for government agency uses, European Research Area Forum[95] guidelines for research uses, and United Nations[96] AI ethics principles, can inform the creation of internal GenAI use policies that should make clear to CSO staff and partners how they should or should not use GenAI in their day-to-day jobs.

[93] "Responsible Use of Generative Artificial Intelligence for the Federal Workforce." U.S. Office of Personnel Management, www.opm.gov/data/resources/ai-guidance. Accessed 26 June 2024.

[94] "System Leaders' Guidance for Use of GenAI Across the New Zealand Public Service." New Zealand Government, www.digital.govt.nz/assets/Standards-guidance/Technology-and-architecture/Generative-AI/Joint-System-Leads-tactical-guidance-on-public-service-use-of-GenAI-summary-September-2023.pdf. Accessed 26 June 2024.

[95] Directorate-General for Research and Innovation. "Guidelines on the Responsible Use of Generative AI in Research." European Union, 20 Mar. 2024, research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/guidelines-responsible-use-generative-ai-research-developed-european-research-area-forum-2024-03-20_en.

[96] "Principles for the Ethical Use of Artificial Intelligence in the United Nations System." UN System Chief Executives Board for Coordination, 27 Oct. 2022, unsceb.org/sites/default/files/2023-03/CEB_2022_2_Add.1%20%28AI%20ethics%20principles%29.pdf.

To provide an example, a policy could permit staff to use GenAI for language translation, so long as there is subsequent human verification, and also prohibit staff and partners from entering any personal data into GenAI models. Civil society groups could publish these policies on their websites for transparency and for others' reference as well. The goal of adopting these frameworks should be identifying and enabling positive, effective GenAI uses internally, upskilling and educating staff, contributing to GenAI use standards within civil society, and boosting accountability for civil society practices. For instance, the Responsible AI Institute, nonprofit organization that works with key stakeholders using AI tools including civil society, has developed a set of best practices[97] to help guide organizations as they seek to effectively and appropriately integrate GenAI.

Join global technology coalitions (if not already involved), and advocate in those forums for a greater focus on GenAI and AI regulations and governance. Consider joining coalitions such as the nonprofit Partnership on AI, the Aspen Institute's Tech Accountability Coalition, and the Global Partnership on Artificial Intelligence (GPAI), supported by the OECD. If CSOs encounter GenAI use in their day-to-day work, they should share those examples and case studies with other CSOs—and if possible, publish them in some form (e.g., write case study blogs, mention the use cases in webinars and public advocacy) to help educate policymakers working on GenAI and democracy. And when CSOs go in front of policymakers to talk about AI, GenAI, and related issues, they should tap into these global technology coalitions to source new research, news articles, best practices, and other materials. For example, civil society interested in GenAI transparency and auditing could benefit from referencing the "Foundation Model Transparency Index"[98] from researchers at Stanford University, MIT, and Princeton University. Also of interest is Open Loop's case study report[99] on AI transparency efforts among companies in Mexico. It is critical that international nongovernmental organizations and the organizers of these coalitions themselves recognize barriers to participation and identify ways to incorporate civil society voices. High-profile civil society groups should advocate for the intentional lowering of barriers for participation such as language, status, and brand recognition expectations.

Develop an incident response plan that incorporates GenAI use. An incident response plan is typically a written document that includes guidance on how your organization can prepare before, and how to act during and after a security incident. Doing so ensures your organization is ready in case of cyber incidents or risks tied to GenAI, such as a GenAI product leaking sensitive organizational data,[100] so that staff have a roadmap for how to respond. Because GenAI does not exist in a vacuum, organizations that already have a cybersecurity incident response plan or cybersecurity measures in general (e.g., phishing filtering, firewalls) can evaluate how potential data leaks via GenAI products, malicious prompts[101] that circumvent GenAI model rules, and other risks fit into their incident response plan. For organizations without a cyber incident response plan—or with a less-developed one—the US National Institute of Standards and Technology provides a starting point with the three steps in its Incident Response Cycle:[102] detect, respond, and recover. The US Cybersecurity Infrastructure and Security Agency also clearly spells out how organizations should create a written incident response plan[103] document and implement it. And the International Association of Privacy Professionals also published an article[104] on AI incident response plans with useful steps to follow when AI systems have caused harm.

[97] Responsible Artificial Intelligence Institute. "Best Practices in Generative AI Guide." Responsible AI, 2024, www.responsible.ai/best-practices-in-generative-ai-guide.

[98] Bommasani, Rishi, et al. "The Foundation Model Transparency Index." arXiv, 19 Oct. 2023. https://doi.org/10.48550/arXiv.2310.12941.

[99] Del Pozo, Claudia May, et al. "AI Transparency and Explainability (Mexico): Public Policy Prototype on the Transparency of Artificial Intelligence Systems." Open Loop, Aug. 2023, openloop.org/programs/ai-transparency-explainability-mexico.

[100] Muncaster, Phil. "Fifth of CISOs Admit Staff Leaked Data via GenAI." Infosecurity Magazine, 24 Apr. 2024, www.infosecurity-magazine.com/news/fifth-cisos-staff-leaked-data-genai.

[101] Zhu, Banghua, et al. "Generative AI Security: Challenges and Countermeasures." arXiv, 20 Feb. 2024, arxiv.org/abs/2402.12617.

[102] "Incident Response." NIST Computer Security Resource Center, csrc.nist.gov/Projects/incident-response. Accessed 26 June 2024.

[103] "Incident Response Plan (IRP) Basics." Cybersecurity and Infrastructure Security Agency, 31 Jan. 2024, www.cisa.gov/resources-tools/resources/incident-response-plan-irp-basics.

[104] Leong, Brenda, and Daniel Atherton. "AI Incident Response Plans: Not Just for Security Anymore." IAPP, 20 Sept. 2023, iapp.org/news/a/ai-incident-response-plans-not-just-for-security-anymore.

Awareness Raising, Advocacy and Educating on GenAI

Flag government and private sector uses of GenAI that that violate or otherwise undermine or potentially compromise human rights[105] such as privacy and freedom of expression. This could be an ongoing effort by a broad array of CSOs—including those with a specific investment in or focus on AI governance programming and advocacy. This is also relevant for those invested in advocating for vulnerable groups —and in other cases will happen on an ad hoc basis as organizations encounter opaque, risky, or harmful uses of GenAI in their day-to-day work. For example, Access Now, a nonprofit organization dedicated to advancing digital rights, has extensive experience on advocating against the use of AI tools that undermine individual liberties[106] in recent years. Freedom House has also drawn attention to how AI may curtail internet freedoms and impact democracies in general.[107] Look for existing guidelines or agreements to help inform or underpin your advocacy, including looking to regional declarations that may be closer to your organizations' priorities, such as the Montevideo Declaration on Artificial Intelligence and its Impact in Latin America among others.[108]

Support societal education efforts, such as civic education campaigns, across multiple fronts, including upskilling internal staff. Identify other civil society groups working on upskilling programs[109] to share case studies, firsthand experiences with GenAI, and other resources and information that could inform their training and education efforts with populations around the world. Develop educational materials based on your organizations' own experiences with GenAI. For example, a human rights NGO could write explanatory blogs on GenAI and human rights impacts (as Access Now did[110]), or a tech accountability nonprofit could speak up about AI-created content it is seeing on social media (as the Software Freedom Law Center in India did).[111]

> "These [GenAI-powered] technologies are benefiting individuals at the local level. Creating campaign codes of conduct, as well as working with watchdog organizations, are examples of a few solutions we could consider to deal with this [GenAI misuse] problem."
>
> Julia Brothers (National Democratic Institute)

The University of Wollongong in Australia has developed resources summarizing and explaining[112] the impact of GenAI and AI in education, and the Berlin-based think tank Democracy Reporting International has built and released a guide on how to better recognize[113] AI-generated content on the internet. Drawing on these resources, CSOs should additionally use their networks and positioning outside industry and government to amplify the perspectives and needs of different ages, skills, and geographic groups during upskilling, as UNESCO did[114] with its report on AI's impacts on indigenous communities in Latin America and the Caribbean.

---

[105] United Nations General Assembly. The Universal Declaration of Human Rights (UDHR). New York: United Nations General Assembly, 1948, www.un.org/en/about-us/universal-declaration-of-human-rights.

[106] "Artificial Intelligence and Human Rights." Access Now, https://www.accessnow.org/issue/artificial-intelligence. Accessed 19 July 2024.

[107] "Freedom of the Net 2023: The Repressive Power of Artificial Intelligence."

[108] Various Authors. Montevideo Declaration on Artificial Intelligence and its Impact in Latin America. KHIPU, Mar. 2023, https://zenodo.org/records/8208793.

[109] "InnovateUS Workshop Series Artificial Intelligence for the Public Sector." InnovateUS, https://innovate-us.org/workshop-series/artificial-intelligence-for-the-public-sector. Accessed 26 June 2024.

[110] Leufer, Daniel, and Méabh Maguire. "What You Need to Know About Generative AI and Human Rights." Access Now, 24 May 2023, www.accessnow.org/what-you-need-to-know-about-generative-ai-and-human-rights.

[111] "Joint Letters to ECI and Platform Companies on Generative AI and Manipulated Media Content." Software Freedom Law Center (India), 29 Mar. 2024, sflc.in/joint-letters-to-eci-and-platform-companies-on-generative-ai-and-manipulated-media-content.

[112] "AI in Education: Publicly Available genAI Tools." University of Wollongong, 19 Mar. 2024, ltc.uow.edu.au/hub/page/ai-resources-public-generative-ai-tools.

[113] "Synthetic Media Exposed: DRI's Comprehensive Guide to AI Disinformation Detection." Democracy Reporting International, 10 Apr. 2023, democracy-reporting.org/en/office/global/publications/synthetic-media-exposed-DRI-comprehensive-guide-to-ai-disinformation-detection.

[114] "Inteligencia Artificial Centrada En Los Pueblos Indígenas: Perspectivas Desde América Latina Y El Caribe." UNESCO, 2023, unesdoc.unesco.org/ark:/48223/pf0000387814.

In addition to investing in external education efforts, CSOs should provide regular training to staff on the promises and pitfalls of GenAI use and invest in educating other CSOs (e.g., through joint training sessions) on ways to better utilize the technology and talk to constituencies about GenAI. It is even possible for CSOs to consider fine-tuning large language models (LLMs) to meet civil society's specific needs, as has been done[115] with Hugging Face's Bloom LLM.

Look up laws and regulations in the countries in which your organization operates to understand what GenAI-related data you have access to. Some emerging regulations increasingly provide civil society with access to technology companies' data for research purposes. For example, the EU's newly implemented Digital Services Act legally requires companies[116] deemed "very large online platforms" and "very large search engines" to provide select researchers with access to data that contributes to detecting, identifying, and understanding systemic risks in EU countries. CSOs seeking to better understand how GenAI impacts disinformation, hate speech, public service delivery, and more could use these research access rights to study non-public data and publish unique insights into the problem space.

RECOMMENDATIONS FOR GOVERNMENT

Governments can take the following steps to help ensure GenAI is used democratically internally as well as in the service of the public and democratic values. These recommendations could be implemented by different elements of governments, such as public-facing benefits agencies that might use a chatbot to enhance service delivery to citizens, and legislators who might be writing laws and shaping regulations around GenAI:

Before taking any other steps, upskill government officials and staff on the basics of generative AI. It is especially important for government to take time to ensure there is a strong, baseline understanding of what GenAI is and what risks and threats it might present in order to understand where, if at all, it might be useful to a specific process or team. Remember that leveraging GenAI might not be applicable to various governmental processes, and that many solutions to governance or societal challenges are analog rather than digital. Refer to the primer section earlier in the report, which lists different educational resources and explanatory documents to help build up basic knowledge of this tool.

After learning the basics, government officials may identify ways to either integrate GenAI into their own work, revise or develop policies and regulations, or launch education and awareness efforts to ensure citizens are informed on the risks and opportunities that GenAI presents. The below recommendations speak to these different needs.

Invest in clearly, understandably, and comprehensively describing and notifying users of how and when GenAI is used in government services or outputs. This can be through the form of adding labels or notifications when GenAI is used, or through the creation of a public resource website. When building such a website, governments should focus on publishing as much as possible regarding GenAI uses that directly impact citizens—for example, if a research agency uses citizens' data to fine-tune a GenAI model, if a law enforcement organization uses GenAI in its policing, or if a social safety net program uses GenAI chatbots to summarize citizen complaints. Public resource websites should cover, among other areas, where and when GenAI is used, any uses of government-held or -procured personal data, fairness and privacy criteria for procuring commercially developed GenAI products, checks and balances to identify and mitigate GenAI errors, and internal processes such as privacy impact assessments and recordkeeping.

[115] Shixiang Research. "Deep Dive on Hugging Face." Medium, 14 Nov. 2022, https://medium.com/@shixiangtech2019/deep-dive-on-hugging-face-87cfd122c1a4. Archived Version, https://web.archive.org/web/20221114151347/https://medium.com/@shixiangtech2019/deep-dive-on-hugging-face-87cfd122c1a4.

[116] "FAQs: DSA Data Access for Researchers." European Centre for Algorithmic Transparency, 13 Dec. 2023, algorithmic-transparency.ec.europa.eu/news/faqs-dsa-data-access-researchers-2023-12-13_en.

While they are early-stage websites and do not currently include many of these components, the US Office of Personnel Management webpage on GenAI[117] and the State of New Jersey website on state GenAI use[118] are two approaches to describing a collection of GenAI opportunities, risks, and uses understandably. Governments should advertise their educational websites to citizens through public communications campaigns, such as on social media platforms. Note that primarily motivated CSOs will access these websites rather than average citizens, so adding in labels and other notifications on GenAI use is paramount.

Ensure that procurement teams purchasing or licensing GenAI solutions have a multidisciplinary staff and approach to ensure the fair, rights-based, and legal use and application of GenAI. This should include building transparency into procurement processes and decisions. For example, the City of Seattle has published a policy[119] that requires all government agencies to go through Seattle's established information technology (IT) acquisition process before using any GenAI software, software-as-a-service, web-based service, browser plug-in, or smartphone app and follow requisite guidelines on privacy, bias, harm, intellectual property, and accountability—a clear policy that is publicly transparent.

Identify opportunities to integrate GenAI tools to increase the accessibility of government resources and improvement of service delivery. For example, consider where a GenAI-powered chatbot can be integrated to answer constituent questions,[120] improve clarity of language[121] on government websites and in government resources, translate documents into more languages,[122] provide screen-reader access for content, include more subtitles on government videos and public service announcements, and simplify written materials[123] for individuals with low literacy. GenAI tools can also be incorporated to make it easier to access services through similar mechanisms, such as chatbots that can answer questions on qualifying factors to receive services, and to direct users to relevant webpages and guidance on how to apply for them. You can also revisit the case studies on pages 19 and 20 for how GenAI tools can produce efficiency gains. And, to ensure citizen needs and rights remain central to any integration of GenAI, ensure there are remedy and redress opportunities for citizens to use when integration of these tools results in harm or unintended consequences.

Assess how GenAI fits into laws and regulations already in place. This includes consumer protection laws or data privacy regulations; consider pursuing investigations or publishing regulatory guidance for specific GenAI products and use cases. For example, see the Colombian Superintendency of Industry and Commerce's investigation[124] into ChatGPT's compliance with the country's data protection legislation and the US[125] for businesses on AI and US.

If any significant gaps are identified in current laws and regulations, look to existing guidelines to begin a legislative or regulatory process to ensure GenAI systems are used in line with democratic and human rights principles. Useful guidelines to refer to include the United Nations' aforementioned AI ethics principles, resources such as the OECD's AI Policy Observatory,[126] and regulations such as the EU's AI Act.[127]

---

[117] "Responsible Use of Generative Artificial Intelligence for the Federal Workforce."

[118] "Using Generative AI in New Jersey State Government." State of New Jersey Office of Innovation, innovation.nj.gov/skills/ai. Accessed 26 June 2024.

[119] Balgetty, Ben. "City of Seattle Releases Generative Artificial Intelligence Policy Defining Responsible Use for City Employees." Office of Mayor Bruce Harrell, 3 Nov. 2023, harrell.seattle.gov/2023/11/03/city-of-seattle-releases-generative-artificial-intelligence-policy-defining-responsible-use-for-city-employees.

[120] Averyheart, Kathryn. "State and Local Agencies Can Improve the Citizen Experience With AI Customer Service." Technology Solutions That Drive Government, 15 Apr. 2024, statetechmagazine.com/article/2024/04/state-and-local-agencies-can-improve-citizen-experience-ai-customer-service.

[121] Novek, Beth. "Generative AI and Policymaking for the New Frontier." GovTech, 1 Dec. 2023, www.govtech.com/opinion/generative-ai-and-policymaking-for-the-new-frontier.

[122] Douglas, Theo. "State GenAI Opportunities Target Inspections, Language." Industry Insider, 1 Feb. 2024, insider.govtech.com/california/news/state-genai-opportunities-target-inspections-language.

[123] "Generative AI and Policymaking for the New Frontier."

[124] Gonzalez, Geusseppe, and Rodrigo Serrallonga Mejía. "Colombian Regulator Launches Inquiry Into ChatGPT Over Personal Data Concerns." Access Partnership, 16 May 2023, accesspartnership.com/access-alert-colombian-regulator-launches-inquiry-into-chatgpt-over-personal-data-concerns.

[125] Levine, Samuel. "Clear Eye on AI." FTC Business Blog, 10 Apr. 2024, www.ftc.gov/business-guidance/blog/2024/04/business-blogs-clear-eye-ai.

[126] "Policies, Data and Analysis for Trustworthy Artificial Intelligence." OECD Artificial Intelligence Policy Observatory, oecd.ai/en. Accessed 19 July 2024.

[127] "EU AI Act: First Regulation on Artificial Intelligence." European Parliament, 6 June 2023, www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

Governments could consider possible use cases, opportunities, and risks associated with political campaigns using GenAI to create campaign videos (e.g., is it permitted and are campaigns required to label the content accordingly?[128]), citizens reporting errors about them in chatbots outputs (e.g., are companies able, and compelled, to correct inaccuracies?[129]), or companies building GenAI models with personal data (e.g., what does consumer "consent" mean?[130]), among others. Laws and regulations should include principles of fairness, transparency, and accountability to the public.

Evaluate how the government is regulating or auditing industry tools and actions related to GenAI. Consider what actions the government can democratically take to provide oversight and protect the rights of citizens. This could include regulatory agencies evaluating laws and regulations already in place to determine whether and how they might apply to GenAI development, deployment, and use—and issuing either general guidance or formal advisory opinions to provide clarity to businesses and civil society on use cases. For example, the US Federal Trade Commission's public guidance on the competition concerns of GenAI[131] and how US "deceptiveness" law applies to AI models' terms of service[132] are two examples of how to assess existing laws and regulations against GenAI and explain the implications clearly to businesses and the public. India provides another example, issuing an advisory[133] in December 2023 reminding intermediaries and platforms of existing information technology rules that impact GenAI and issuing another advisory[134] in March 2024 further describing how those same companies must use GenAI and large language models under existing law. Other options for governments include establishing third-party auditing systems to evaluate GenAI tools; conducting first-party audits of the corporate governance mechanisms for GenAI systems, such as by leveraging existing regulatory authorities to subpoena access to companies' controls to prevent GenAI model misuse; establish tiered fines, similar to the risk-based regulatory approach in the EU's Artificial Intelligence Act,[135] to hold companies to account for harmful behavior; and issuing government agency requests for information that invite businesses, civil society, and members of the public to respond to GenAI-related questions that could then inform an agency's future regulatory rulemaking.

Involve citizens to ensure their feedback is included and considered for all stages of GenAI integration. This includes creating opportunities for citizens to weigh in before, during, and after a GenAI tool is integrated. This can be done through mechanisms like town halls, virtual feedback forms, public surveying, and even user testing once a government is moving forward with tool integration. Invest in increasing the government's understanding of public sentiment around GenAI to predict both acceptable and controversial GenAI uses. Organizations such as Democracy X in Denmark have done good work in the past to gather input from citizens[136] regarding how decision-makers seek to leverage AI solutions through listening sessions and participatory dialogues. Be intentional about engaging more vulnerable or marginalized groups whose voices are often excluded, such as women and minority rights groups.

Engage external experts to assist with objectives like training and upskilling while not becoming reliant on external experts and building up, over time, the necessary internal capacity, expertise, and staffing. Governments should do so by setting aside funds to create and post a job notice (or notices) when feasible for an AI policy expert (or experts)—or tasking a current government employee, such as a technology policy expert, with developing a plan to increase AI expertise in government.

---

[128] "Our Approach to Labeling AI-Generated Content and Manipulated Media."

[129] "ChatGPT Provides False Information About People, and OpenAI Can't Correct It." Noyb, 29 Apr. 2024, noyb.eu/en/chatgpt-provides-false-information-about-people-and-openai-cant-correct-it.

[130] Voica, Alexandru. "GenAI Apps Are Cloning Your Likeness Without Consent—and Might Make You Famous for All the Wrong Reasons." Fortune, 11 Mar. 2024, fortune.com/2024/03/11/gen-ai-apps-cloning-likeness-consent-tech-artificial-intelligence-risk-law.

[131] Staff in the Bureau of Competition & Office of Technology. "Generative AI Raises Competition Concerns." FTC Office of Technology Blog, 29 June 2023, www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns.

[132] By Staff in the Office of Technology and the Division of Privacy and Identity Protection. "AI (and Other) Companies: Quietly Changing Your Terms of Service Could Be Unfair or Deceptive." FTC Office of Technology Blog, 13 Feb. 2024, www.ftc.gov/policy/advocacy-research/tech-at-ftc/2024/02/ai-other-companies-quietly-changing-your-terms-service-could-be-unfair-or-deceptive.

[133] Ministry of Electronics and IT. MeitY Issues Advisory to All Intermediaries to Comply With Existing IT Rules. 26 Dec. 2023, pib.gov.in/PressReleasePage.aspx?PRID=1990542.

[134] "MeitY Revises AI Advisory, Does Away with Government Permission Requirement – Update." AZB & Partners, 21 Mar. 2024, https://www.azbpartners.com/bank/meity-liberalizes-ai-advisory-dated-march-1-2024-following-industry-concerns-and-issues-revised-advisory-on-march-15-2024/.

[135] "Article 99: Penalties." EU Artificial Intelligence Act, artificialintelligenceact.eu/article/99. Accessed 26 June 2024.

[136] "Participation." The Danish Board of Technology, tekno.dk/service/participation/?lang=en. Accessed 26 June 2024.

At the same time, governments should look to CSOs locally, regionally, and globally that have technology, media, democracy, and AI expertise to explain GenAI topics to civil servants and policymakers, and assist with building a plan for increasing government AI expertise over the next several years.

RECOMMENDATIONS FOR INDUSTRY

Private-sector companies, including social media platforms and GenAI developers, can take the following steps to help ensure that GenAI is used democratically within their own organizations and in organizations that buy or procure their technology—as well as in the service of the public and democratic values. These opportunities for action span many phases and aspects of GenAI creation and use, including training data collection and refinement, model design and development, GenAI deployment, contractual and technical controls around GenAI use, and online platform rules about the handling of GenAI content:

Build capacity among industry professionals on the ethical, legal, societal, and human rights implications of GenAI. Consider doing this through increased development of educational materials and widespread diffusion of existing company resources, GenAI design practices, and GenAI deployment policies. For example, AI developers such as IBM[137] and Google,[138] consulting firms such as Deloitte[139] and PwC,[140] and academic researchers[141] have all published design principles and guidelines for AI and GenAI ethics. A positive cycle can be created when companies intentionally invest in publishing information about their own practices, educate developers about the implications of AI, and provide resources to companies and universities looking to train developers on AI ethics, ultimately creating more incentives for other companies to develop and publish a set of ethical design principles. These steps also boost accountability, because CSOs can use company guidance to identify gaps in industry practices and hold specific firms to the standards. It is also important to invest in engaging with a diversity of experts who can assist in understanding broad societal impacts, including digital rights experts, women's rights scholars, and marginalized group advocates.

Provide sustainable, long-term funding for civil society and academic partners in the larger AI space. Industry has a role to play in ensuring that all stakeholders, especially those outside of government, can meaningfully participate in conversations surrounding the democratic use of GenAI. Offering resources to groups like civil society and academia, and especially to organizations throughout the Global South, can go a long way to ensure a healthy GenAI ecosystem. Funding could go toward supporting research, advocacy, and training activities that are ultimately likely to help encourage industry to minimize the harms that their own tools may pose. Through these contributions, leaders in industry could have a significant role in creating a responsible AI ecosystem and could help to reverse recent funding cuts that have impacted the digital rights ecosystem critical to informing responsible GenAI development.

Engage policymakers in their national AI lawmaking processes to ensure industry views are formed and then communicated to government and civil society. These efforts can occur through direct engagement with policymakers as well as other multi-stakeholder efforts. For example, while not industry-led, the nonprofit Partnership on AI model for this kind of consensus building and information sharing, bringing together nonprofit organizations like the American Association for the Advancement of Science and Code for Africa, academic institutions such as the AI Center in Tokyo and the Australian National University, and companies like Google and Meta to discuss and promote best practices on algorithmic fairness, multi-stakeholder participation, AI and elections, and more. These models serve as a great example to consider emulating in your own context.

[137] "AI Design Ethics Overview." IBM, www.ibm.com/design/ai/ethics. Accessed 26 June 2024.

[138] "Our Principles." Google AI, ai.google/responsibility/principles. Accessed 26 June 2024.

[139] Guszcza, Josh, et al. "Human Values in the Loop." Deloitte Insights, 28 Jan. 2020, www2.deloitte.com/us/en/insights/focus/cognitive-technologies/design-principles-ethical-artificial-intelligence.html.

[140] Axente, Maria Luciana, and Ilana Golbin. "Ten Principles for Ethical AI." PwC Australia Digital Pulse, 12 Apr. 2022, www.pwc.com.au/digitalpulse/ten-principles-ethical-ai.html.

[141] Brey, Philip, and Brandt Dainow. "Ethics by Design for Artificial Intelligence." AI And Ethics, Sept. 2023, https://doi.org/10.1007/s43681-023-00330-4.

Publish more information about GenAI use and development practices, and share ethical and human rights standards on websites and directly to users/customers, journalists, and civil society organizations. Social media platforms, search engines, and other companies should have a single webpage or set of webpages where GenAI and AI uses are clearly, succinctly, and understandably listed and explained—along with relevant links to opt-outs or data correction forms. This should include describing whether the company gathers any data directly from users/customers to train AI systems and whether the company purchases or acquires[142] data about users/customers from third parties (e.g., buying from data brokers, scraping from public websites) to train AI systems. Companies should also engage with civil society organizations about their GenAI ethical standards and use policies through presentations at conferences such as NeurIPS144, one of the largest annual AI gatherings; discussions in multi-stakeholder forums such as the Partnership on AI; and by potentially standing up civil society advisory networks, akin to Twitter's former Trust and Safety Advisory Council[143], where they can continuously consult with and hear from civil society.

Build products and services according to human rights principles, leveraging multidisciplinary teams to do so. This includes creating AI policy teams with lawyers, product engineers, and external civil society expert advisers. As a reference point, the United Nations lays out 10 principles for the ethical use of AI:[144]

- **do no harm;**
- **defined purpose, necessity, and proportionality;**
- **safety and security;**
- **fairness and nondiscrimination;**
- **sustainability;**
- **right to privacy, data protection, and data governance;**
- **human autonomy and oversight;**
- **transparency and explainability;**
- **responsibility and accountability; and**
- **inclusion and participation.**

To implement these principles in GenAI development and use, companies should start by conducting a human rights impact assessment—the Danish Institute for Human Rights explains how to get started[145]—and then regularly conduct updated assessments to evaluate gaps in company policies and engineering practices. They should also engage civil society groups and external experts to conduct external audits, stand up and staff independent review boards, or supervise nonregulatory safeguards for GenAI ethical design, deployment, and use, as well to ensure products and services align with human rights principles.

Intentionally create sustained space and recurrent opportunities for knowledge sharing. It is critical for tool developers to maintain sustained opportunities and efforts to connect with civil society to ensure the perspectives, concerns, and learnings from these groups are regularly integrated into and inform product development. These spaces should continuously be made available beyond the conclusion of an audit or after a product is launched; that way, tools are more likely to holistically reflect the concerns and values of civil society groups. Further, the exchange of this knowledge is a two-way street, so it is important to share technical knowledge from technical teams to assist democratic actors in the development of applications, training, and policies around GenAI, such as sharing technical expertise with civil society on best practices for GenAI use in translation.

---

[142] Burgess, Matt, and Reece Rogers. "How to Stop Your Data From Being Used to Train AI." WIRED, 10 Apr. 2024, www.wired.com/story/how-to-stop-your-data-from-being-used-to-train-ai.

[143] Zakrzewski, Cat, et al. "Twitter Dissolves Trust and Safety Council." Washington Post, 12 Dec. 2022, www.washingtonpost.com/technology/2022/12/12/musk-twitter-harass-yoel-roth.

[144] "Principles for the Ethical Use of Artificial Intelligence in the United Nations System." Accessed 26 June 2024.

[145] "Introduction to Human Rights Impact Assessment." The Danish Institute for Human Rights, www.humanrights.dk/tools/human-rights-impact-assessment-guidance-toolbox/introduction-human-rights-impact-assessment. Accessed 26 June 2024.

Finally, these knowledge-sharing spaces should include opportunities to share insights among other companies developing or using GenAI, especially sharing challenges and learnings related to political use cases and information integrity concerns to inform a more holistic ecosystem approach to these challenges.

Tailor harm mitigation efforts for various environments and identity group needs—for instance, GenAI accessibility and usability gaps between different country contexts and challenges, or disparities in how hate speech and disinformation impact women. Combatting GenAI-created disinformation and hate speech in English in Western Europe, for example, will require different solutions and approaches than targeting GenAI-created disinformation and hate speech in another language (in which many major GenAI chatbots are not trained.)[146]

Commit resources to methods for detecting GenAI misuse, such as identifying GenAI-created disinformation[147] and other content[148] or identifying when individuals manipulate GenAI chatbots with prompts[149] designed to return underlying user data. Engage in collaboration with other companies and governments on those detection methods, and ensure those detection techniques are listed and explained to users/customers on the aforementioned webpages to boost transparency.

EMBRACE CROSS-SECTORAL COLLABORATION

One way to approach maximizing GenAI's opportunities and mitigating GenAI's risks is to break recommendations down by civil society, government, and industry. Another important consideration, working group members noted, is the types of individuals at an organization who implement a recommendation—such as lawyers, engineers, policymakers, or public opinion professionals. Transparency is a great example to illustrate this point; for lawyers, transparency could mean pairing company practices against a checklist based on regulations, with each checklist item open to some interpretation. For an engineer, transparency could mean implementing specific technical measures. For a social scientist, transparency might mean conducting a qualitative and quantitative study of how users interact with GenAI models and content, and whether they view the models or content as reliable. Maximizing the opportunities of GenAI for democracy while mitigating its many risks to democracy will require this kind of broad involvement across product engineers, data scientists, content moderators, policymakers, media watchdogs, and other actors.

[146] North, Madeleine. "Generative AI Is Trained on Just a Few of the World's 7,000 Languages. Here's Why That's a Problem – and What's Being Done About It." WEF Foreign Agenda, 17 May 2024, www.weforum.org/agenda/2024/05/generative-ai-languages-llm.

[147] Loth, Alexander, et al. "Blessing or Curse? A Survey on the Impact of Generative AI on Fake News." arXiv, 3 Apr. 2024, doi.org/10.48550/arXiv.2404.03021.

[148] Perkins, Mike, et al. "GenAI Detection Tools, Adversarial Techniques and Implications for Inclusivity in Higher Education." arXiv, 28 Mar. 2024, doi.org/10.48550/arXiv.2403.19148.

[149] Iyer, Prithvi. "New Study Suggests ChatGPT Vulnerability With Potential Privacy Implications." Tech Policy Press, 29 Nov. 2023, www.techpolicy.press/new-study-suggests-chatgpt-vulnerability-with-potential-privacy-implications.

LOOKING AHEAD: WHAT TO ANTICIPATE

Given the risks and opportunities we laid out, how can we anticipate GenAI's impacts evolving? While many of the white paper recommendations are intended to better equip democratic actors to start taking steps now to address GenAI's impacts, this section takes a forward-looking perspective on how GenAI's impacts may evolve in coming years. Working group members highlighted a number of uncertain potential realities that may come to fruition depending on how the GenAI space evolves. The likelihood of the below scenarios unfolding is dependent on factors that include the extent of GenAI regulation; the effectiveness of governments' regulation of GenAI; the evolution of GenAI tools and the ways in which they become increasingly powerful; and the extent to which industry establishes usage guidelines.

While the below realities remain to be seen, it is important for civil society, government, industry, and academia to consider how GenAI may be applied in the future, and for the democracy community to continue assessing GenAI's impacts.

DATA GATHERING

In the absence of robust data privacy laws and regulations, companies, governments, and research institutions could collect and use personal data—such as copyrighted media, facial images, voice recordings, and videos—in violation of citizens' privacy. As a result, citizens could have their data continually used without their consent to develop GenAI applications, which may further violate their privacy—such as through the creation of increasingly accurate nonconsensual deepfake imagery.

SURVEILLANCE

In the absence of protective legislation, citizens may be subject to government surveillance to collect data to further AI development, including but not limited to the collection of biometric data and surveillance of online behavior. Governments may leverage GenAI to enhance their ability to monitor content, censor speech, and track dissidents' and citizens' activities.

LANGUAGE DIVERSITY OPPORTUNITIES

If GenAI developers invest in non-English language content and training data, opportunities could emerge for democracy nongovernmental organizations (NGOs), accountability watchdogs, political campaigns, and government agencies around the world to expand the accessibility of their content to include different language speakers. These multilingual tools could enhance information integrity efforts across platforms in a wider diversity of languages, including for indigenous or ethnic communities' languages that have often been overlooked.

STRENGTHENED OR FRACTURED INTERNATIONAL NORMS

GenAI is just one component of the broader conversation around information environments, democracies, elections, and norms as they relate to technology. If the rapid development and deployment of GenAI models surpasses even the fastest-developed, nonregulatory GenAI guardrails, the international norms landscape may become even more fragmented or drift further from democratic values including, but not limited to, privacy, transparency, freedom of expression, accountability, and human rights.

LACK OF EXPLAINABILITY OF GENAI OUTPUTS LEADS TO LOWERED ACCOUNTABILITY

Due to persistent and serious explainability problems around how GenAI works, it will remain difficult to explain how and why certain outputs are created. If governments are quick to integrate GenAI into public policy and public benefits processes, the lack of transparency as to how and why GenAI outputs are informing government decisions and processes will reduce the public's ability to hold those processes accountable. Governments and developers may not be able to explain why GenAI systems reached specific decisions. Civil society watchdog groups, media, and citizens could ask, petition, or sue for access to more information about GenAI's usage and development in those situations, but they may not be successful.

DIVERSIFICATION OF THE AI DEVELOPMENT SPACE

The GenAI development space is presently driven heavily by companies based in the United States and China, yet it is likely that more countries will implement GenAI, creating a more diversified base for AI development. This shift may result in more localized tools, resulting in potentially more widely accessible tools but potentially further fragmenting the GenAI space. However, the diversification of AI tools will depend on whether countries embrace open data policies, and on the overall quality and availability of data.

GROWING IMPORTANCE OF NO-CODE TOOLS

No-code GenAI tools, such as Hugging Face among many others, will increase the ability of individuals with little to no coding experience to develop GenAI tools on open-source platforms. This may democratize access and development while also saturating the GenAI space with high- and low-quality tools, potentially confusing users and resulting in a plethora of unregulated services and products with little to no oversight.

FRACTURED TRUST IN THE INFORMATION SPACE

As GenAI continues to advance in capacities, and access to GenAI expands, GenAI-developed mis/disinformation and deepfakes—and awareness and suspicion of digital content as a result—will continue to grow, fracturing trust in online information as a whole.

APPENDIX: ADDITIONAL RESOURCES

There are many resources available to aid in better understanding GenAI, spanning from education resources on terminology to articles tracking its use and impacts around the globe. Please reference this spreadsheet, which will be continuously updated, for a growing list of resources:

Generative AI and Democracy Resources List

INTERNATIONAL REPUBLICAN INSTITUTE

IRI.org | @IRIglobal | P: (202) 408-9450 | E: info@iri.org