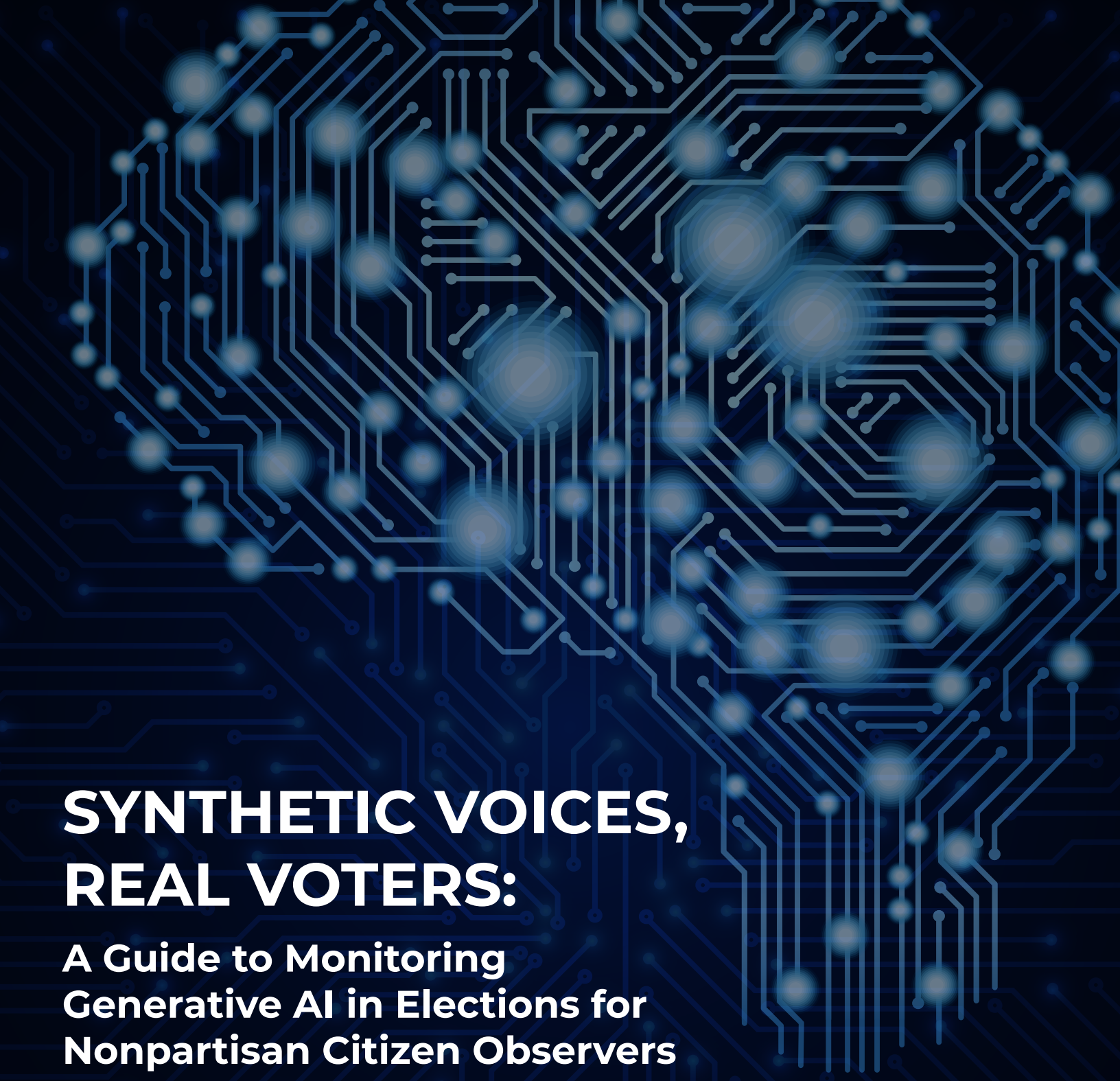


SYNTHETIC VOICES, REAL VOTERS:

A Guide to Monitoring Generative AI in Elections for Nonpartisan Citizen Observers

National Democratic Institute (NDI)


NDI
NATIONAL
DEMOCRATIC
INSTITUTE


Sida

ACKNOWLEDGEMENTS

Nonpartisan citizen election observers around the world are facing increasingly complex political and digital threats to credible elections, requiring them to observe for longer periods of time, cover more diverse processes, and use multiple methodologies and approaches to capture evolving dynamics. This is at a time when funding and resources for nonpartisan citizen observation worldwide is on the decline, making elections more vulnerable to exploitation or eroding public confidence. Transnational networks like the Global Network of Domestic Election Monitors (GNDEM) have continued to build solidarity and learning opportunities between and among citizen election observers around the globe during this challenging time. This guide was drafted in part in response to expressed interests by the GNDEM community.

The National Democratic Institute (NDI) is grateful to the Swedish International Development Cooperation Agency (Sida) for funding this valuable resource and supporting the work of citizen election observers. This guide was developed as part of a Sida-funded initiative geared at combatting digital threats to elections.

This guide was supported by an Advisory Group of citizen election monitoring leaders and experts in the field of information integrity, artificial intelligence, and/or digital communications from all around the world. They include: Sebastian Bay (Independent Expert), Lucas Calil (Fundação Getulio Vargas - FGV), Ona Carritos (Legal Network for Truthful Elections - LENTE), Ognjan Denkovski (Democracy Reporting International - DRI), Samson Itodo (Yiaga Africa), Roy El Khoury (Democracy Reporting International - DRI), Yuri Lisovskyi (Civil Network OPORA), Cynthia Mbamalu (Yiaga Africa), Kateryna Mykhailevska (Civil Network OPORA), and Miazia Schuler (AI Forensics). The outline and guidance from this guide was presented and reviewed by GNDEM representatives and experts in August 2025 in Nairobi, Kenya. NDI and GNDEM expresses their appreciation to all individuals and organizations who lent their expertise and insights in contributing to this guide.

This guide was written by Julia Brothers with support from Tetayana Bohdanova and Anis Samaali. The guide was reviewed by Richard Klein and other NDI experts including Nathan Grubman, Pajtim Gashi, Jerrel Gilliam, Nino Vardosanidze, as well as the Advisory Group.

INTRODUCTION

Democratic elections rely on a competitive process, faith in electoral institutions, and informed participation by all citizens. However, the ability of voters to make choices based on complete and accurate information is increasingly strained by a fractured and frenetic information environment. In some ways, there are more sources for electoral information than ever before, but also a rise in echo chambers, low-credibility sources, and conflicting data. Competitive democracies are facing unprecedented levels of political polarization and disillusionment while authoritarian actors are increasingly undermining credible sources of information surrounding elections, which can tilt the playing field providing unfair advantage and eroding confidence in electoral institutions and safeguards.

Amidst this context is the emergence of widely available generative Artificial Intelligence ("genAI") tools, which threaten to exacerbate these concerns and present challenges to those working to secure electoral integrity. The uptick of genAI-enabled campaign tactics, realistic synthetic political content online, and the use of chatbots for basic tasks and information complicate electoral dynamics during an already fragile moment. Electoral watchdogs such as election observers - who are already monitoring many facets of electoral integrity - will also need to understand the impact of this new dimension. This is particularly pressing as new and upgraded genAI tools are rushed to public release, and often before they are sufficiently tested or risk-assessed.

There is an understandable fear of the unknown when examining the use and impact of genAI in elections - both positively and negatively. This is compounded by both real and perceived opacity and barriers in data collection. GenAI is not yet a well understood, trusted or well-regulated technology, and the transparency and ethics of its deployment and use are largely being determined by technology companies that have their own interests and financial considerations. This guide is intended to provide clarity and reduce anxiety around what has often been viewed as a broad and ominous threat, particularly as the various applications of genAI in elections can have drastically different impacts and electoral integrity risks. This is particularly important for election observers, who are called upon to provide grounded analysis that highlights progress and improvements in electoral processes as well as the true scale and scope of any shortcomings. Citizen observers are not engaging in fact-checking, but rather identifying and evaluating the impact of digital trends - both positive and negative - as part of their broader assessment of the integrity of the process, which can help build transparency and provide actionable recommendations.

This guide is designed to help observers put genAI into context, and support evidence-based findings on its prevalence and impact, not assume the worst. This guide will cover key definitions, practical methodologies, tools, and considerations for further adaptation and planning, as the technology and trends around generative AI and its use are constantly evolving. It will also include a link to a **living document** of updated tools, resources, and observation case studies to ensure users can stay informed as the space develops. The guide is specifically designed for nonpartisan citizen election observers safeguarding their own electoral processes, and to ensure fundamental freedoms, including freedom of expression, are protected. It will focus on methodologies that are achievable and not resource intensive, with additional guidance on how AI-threats can be integrated into pre-existing monitoring strategies.

WHY CITIZEN ELECTION OBSERVERS? WHY NOW?

Nonpartisan citizen election monitors, who are viewed as trusted, politically impartial voices, are well-equipped to analyze, expose, and provide recommendations to understand and address the effects of generative AI in elections. Citizen observers have a critical role to play in the future of genAI in elections, in particular because:

- **Citizen election observers are better positioned to monitor generative AI in elections than international observers.** Citizen observers understand online vernacular and localized viral trends that can help in investigating deepfakes as well as connections to hate speech, incitement, and other means of fanning social divisions. That understanding can be helpful to international election observers, media, and researchers. Moreover, citizen observer organizations can provide ongoing monitoring not only during elections, but also during other major political moments, such as protests, legislative votes, national plebiscites, and the period between elections when the online manipulation of political narratives tends to take root.

- **Monitoring genAI will likely intersect with monitoring that many citizen observers are already doing, especially in the pre-election period.** Most citizen observers are already planning long-term observation exercises that may touch on similar or related themes. In this way, adjusting or expanding methodology may be relatively easy, and can ensure that findings are analyzed within a relevant broader context.

- **Citizen observers need to be able to speak credibly about genAI in elections.** As genAI continues to be a highly publicized and discussed technological trend, observers should expect to be asked about the use or impact of genAI in elections - perhaps even by media outlets or stakeholders who may not wholly understand it themselves. It's important that observers be able to speak confidently in this space, and not to confuse voters or the public, or contribute to unnecessary panic.

- **Now is a critical moment for building accountability and transparency around genAI and its use in elections:** GenAI is widely expected to become more sophisticated in the future, creating even more urgency for a strong regulatory framework now. However at the time of writing, the regulatory environment around generative AI and electoral integrity is nascent in most regions of the world, and there remain no recognized uniform international norms or standards in deploying genAI technology around elections, although attempts have been made (discussed in further detail in Chapter 6). In addition, technology companies promoting genAI tools have only had limited constructive engagement with electoral experts and stakeholders, particularly on impacts. It is important for election observers to build transparency and accountability around genAI tools and applications now while they are still being tested and competing for market share, rather than when they are hard-coded for consumption.

GENERATIVE AI AND INTERNATIONAL ELECTORAL INTEGRITY STANDARDS

International standards for democratic elections assure open, robust, and pluralistic information environments that promote equal and full participation in elections by citizens and contestants alike. Despite the limited regulatory environment around generative AI currently, these standards are enshrined in international and regional instruments, which reflect pre-existing, globally recognized commitments that may pertain to potential genAI threats. This includes:

> "Persons entitled to vote must be free to vote for any candidate for election and for or against any proposal submitted to referendum or plebiscite, and free to support or to oppose government, without undue influence or coercion of any kind which may distort or inhibit the free expression of the elector's will. Voters should be able to form opinions independently, free of violence or threat of violence, compulsion, inducement or manipulative interference of any kind."
>
> **General Comment 25, UN Human Rights Committee**

- **The rights to hold opinions and to seek and receive information in order to make an informed choice on election day:** Everyone has the right to form, hold, and change opinions without interference, which is integral to freely exercising the right to vote.[1] Voters also have the right to seek, receive, and impart accurate information that allows them to make informed choices regarding their future, free from intimidation, violence, or manipulation.[2] Further, institutions are generally obligated to be transparent regarding electoral information so that voters can be informed and data sources can be held accountable.[3] These rights are enshrined for all citizens regardless of race, gender, language, area of origin, political or other opinion, religion, or other status.[4] Synthetic or manipulated political and electoral information may subvert these rights, because it is designed to undermine genuine political debate by intentionally deceiving voters, creating confusion, exacerbating polarization, and undermining public confidence in the electoral process.

- **The right to a level playing field:** Universal and equal suffrage, in addition to voting rights, include the right to seek election to public office without discrimination. Governments' obligations to ensure level playing fields for electoral contestants are derived from this norm.[5] The norm implies providing security from defamatory attacks and other forms of false information aimed at harming a candidate's or a party's electoral fortunes. The obligations extend to government-controlled media, and the

1 Articles 19 of the Universal Declaration of Human Rights (**UDHR**) and International Covenant on Civil and Political Rights (**ICCPR**). **General Comment 34**, paragraphs 2, 4, and 7, UN Human Rights Committee (UNHRC). The UNHRC reviews implementation of the ICCPR and presents its interpretations of the treaty's provisions through its General Comments.

2 "Persons entitled to vote must be free to vote for any candidate for election and for or against any proposal submitted to referendum or plebiscite, and free to support or to oppose government, without undue influence or coercion of any kind which may distort or inhibit the free expression of the elector's will. Voters should be able to form opinions independently, free of violence or threat of violence, compulsion, inducement or manipulative interference of any kind." – General Comment 25, paragraph 19, UNHRC.

3 See, e.g., General Comment 34, paragraphs 18 and 19, UNHRC.

4 These obligations are founded in the freedom of expression provisions contained in the UDHR, the ICCPR, the UN Convention Against Corruption (UNCAC), the American Convention on Human Rights, the African Union Convention on Preventing and Combating Corruption, and the Organization for Security and Cooperation in Europe (OSCE)'s Copenhagen Document, among many others.

5 The UN Human Rights Committee provides guidance on this in its **General Comment 25** to the ICCPR.

norm applies to professional ethics for journalists and private media.[6]

- **Freedom of expression, the press, and regulation:** The aforementioned commitments must be balanced by the freedoms of everyone to hold opinions and to express them, including the need to respect and protect a free press. This includes resisting censorship and regulations that could undermine the freedom of expression for the media as well as individual users.

Don't Panic: Untangling Reality from the Hype

Discussions about generative AI can feel overwhelming, especially because its applications and opportunities (and threats) seem unlimited. However the reality is probably less dramatic and more muddled. People often conflate genAI with other technological developments, including broader digitization in elections, which is an important but separate issue. An addition, as discussed in more detail in Chapter 3, issues related to genAI content online are not markedly different than "shallow fakes" (altered media using basic, widely available editing software) and other kinds of information manipulation. At the time of the writing of this guide, the impact of genAI specifically in elections was somewhat limited and we have not yet seen systematic, outcome-influencing genAI operations thus far. In addition, the long-term sustainability of genAI tools remains questionable so the future of consumer-facing tools can change quickly. Civic organizations should be empowered to weigh in on a technological threats regardless of whether they are technologists or not. As discussed in this guide, there are many rights-based approaches and research methodologies that can contribute to greater transparency and accountability around generative AI in elections.

SCALE AND SCOPE OF THIS GUIDE

Election observers cannot monitor everything. Meanwhile "generative AI" itself is a large and complex series of artificial intelligence techniques that can be applied in any number of contexts. The first step to developing an achievable monitoring methodology on this issue is to define and narrow the scale and scope. This guide is written with the following parameters in mind.

Scope: GenAI Not All AI

"Artificial intelligence" (AI) is an incredibly broad term that refers to the use of machine learning for pattern recognition, analysis, prediction, and problem solving. Machine learning and AI have played a role in daily life, and elections, for years. Campaigns may have utilized non-genAI chatbots to engage with voters using simple rule-based systems that follow pre-programmed logic to provide answers. Media monitors and election observers may use software that includes sentiment analysis or natural language processing (NLP)[7] to monitor traditional and social media. Optical mark recognition (OMR) or optical character recognition (OCR) is often used in election administration or analysis, for instance in ballot marking and counting.[8] These are all AI-enabled tools which have largely been used to better understand and process existing information, but are not generative AI.

6 General Comment 34 paragraph 37 also addresses campaigning on a level playing field.
7 NLP is a field of AI that enables computers to understand and process human language.
8 OMR is technology that reads human-created marks on documents, like bubbles or lines filled on a ballot. OCR converts images of typed, handwritten or printed text into machine-encoded text. This may be used to digitize hard copies of petitions or candidate nomination forms.

In contrast, novel generative or genAI technology refers to original machine-created information or content, based on learning models that tend to be proprietary. This introduces concerns about **quality control, accuracy, provenance,** and, in some cases, **bias**.

- **Artificial intelligence:** A field of computer science that encompasses a broad range of applications that use machine-learning for pattern recognition, analysis, prediction, and problem solving. AI is simply an umbrella term used to describe new types of computer software that can approximate human intelligence, or perform tasks previously thought to require human intelligence.
- **Generative artificial intelligence:** A subfield of artificial intelligence techniques. It is a type of "deep-learning" model that creates new, original content based on data on which the models were trained. The output can be high-quality text, images, audio or videos that reflect or respond to the input.

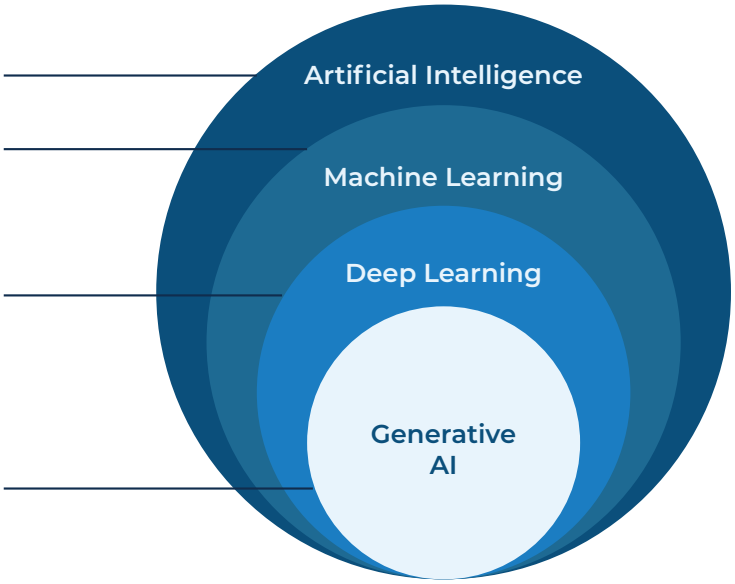
This guide focuses only on generative AI which is a subset within the area of artificial intelligence. This will help limit the universe of what we are discussing and how we collect and analyze data around it. While the monitoring methodologies in this guide are designed to understand the impact of genAI on electoral integrity, it's important for observers to note that the impact may be positive or mixed.

Artificial intelligence is the field of study.

Machine learning is a subset of AI that uses algorithms to detect patterns in large data sets.

Deep learning is a subset of machine learning based on artificial neural networks for in-depth data processing and analytical tasks. Multiple layers of processing are used to extract progressively higher level features from data, simulating how human brains perceive and understand information.

Generative AI is a subset of deep learning that creates new, original content based on data on which the models were trained. The output can be high-quality text, images, audio or videos that reflect or respond to the input.



Focus on Elections and Electoral Integrity

There are many societal concerns around the emergence of generative AI including undermining legacy media, violating intellectual property rights, environmental dangers of energy-depleting genAI data centers, its use in money-making scams, possible threats to national security, or the impact that chatbots can have on learning or mental health. Training data bias for AI generators has been proven to reinforce misogyny and cultural stereotypes. At the same time, there are also broad discussions about the societal opportunities of genAI, such as medical advancements, accelerating research, and improving predictive modeling for the public good. While all of these concerns and opportunities are worthy of discussion, they are out of the scope for election observers, and can become easily distracting or unwieldy if observers pursue. For manageable methodology, observers should remain focused specifically on election and electoral integrity related uses and impacts.

Monitoring - Not Using - GenAI

Many election observers are experimenting with how to use generative AI to improve their internal processes, organizational development, recruitment and outreach, or methodological approaches. This is certainly an area worth further exploration, however this is not the purpose of this guide. This guide is concerned specifically with monitoring strategies around generative AI in elections, to ensure observers have the tools to continue to build transparency, accountability and confidence in elections in the face of changing technological trends and uses.

Key Definitions[9]

- **Artificial intelligence (AI):** A field of computer science that encompasses a broad range of applications that use machine-learning for pattern recognition, analysis, prediction, and problem solving. AI is a catch-all term used to describe new types of computer software that can approximate human intelligence, or perform tasks previously thought to require human intelligence.
- **AI "slop":** A form of digital synthetic content made with genAI and defined by a lack of effort, quality or deeper meaning, and an overwhelming volume of production.
- **Astroturfing:** The practice of creating fake grassroots movements online to give the impression of widespread public support for a product, policy, or cause, often using fake accounts and other deceptive tactics. Astroturfing involves hiding the true sponsors of a message to make it appear as if it originates from a large group of ordinary people.
- **Bias:** GenAI systems process big data, and their accuracy depends on the size and breadth of the dataset. However, women and other marginalized populations are less likely to be represented in datasets because of structural discrimination, group size, or external attitudes that prevent their full participation in society. In this way, bias in training data can systematize existing discrimination. The humans that build and deploy genAI models can also intentionally introduce bias to the system to manage its outputs, for security, relevance, or other motivations.
- **Content provenance:** Content provenance is the recorded history and origin of a digital file, detailing who created it, when, where, and how it has been changed over time. It is a way to verify the authenticity of digital content, particularly with the rise of AI-generated media.
- **Generative artificial intelligence (genAI):** A subfield of artificial intelligence techniques. It is a type of deep-learning model that creates new, original content based on data on which the models were trained. The output can be high-quality text, images, audio or videos that reflect or respond to the input.
- **Jailbreaking:** Related to chatbots, jailbreaking is the practice of using carefully crafted inputs (prompts) to bypass their built-in safety measures, ethical constraints, and content filters, forcing it to generate outputs it was designed to prevent.

9 For more details on Artificial Intelligence and its various offshoots and definitions, visit Civicspace. tech **here**.

- **Large-Language Models (LLMs):** A type of artificial intelligence trained on vast amounts of text data to understand and generate human language. These models are used for a variety of tasks, including answering questions, summarizing text, translating languages, and writing content by recognizing patterns, grammar, and context in text.
- **Liar's Dividend:** The "liar's dividend" is a term for discounting real content - particularly if it's incriminating or damaging - as fake, such as a genAI deepfake. This strategy can be effective because it creates enough confusion and distrust to evade accountability or damage the credibility of the evidence itself.
- **Synthetic content:** Synthetic content is any media, such as images, audio, or video, that is artificially created or altered, often using AI, to appear realistic. It can range from fully fabricated media, like a deepfake video, to content that has had elements like backgrounds or sounds changed or added. The goal is to replicate reality, and the technology used includes deep learning models.

THE MONITORING FRAMEWORK: HOW TO USE THIS GUIDE

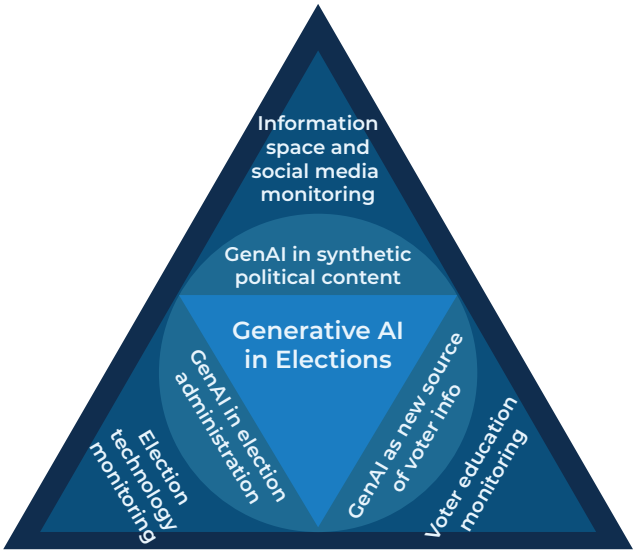
At the end of the day, generative AI models are tools. Tools are agnostic. It's how humans use and apply tools that is important. Understanding the uses of genAI in elections and specific corresponding electoral integrity concerns allows for a focused and appropriate response to its use in the election context. By better articulating, understanding and narrowing these specific threats, observers can reduce confusion, strengthen accountability and provide appropriate recommendations to address them. This guide book is organized based on **three distinct applications of generative AI in elections - synthetic political content, voter information and election administration,** as they produce different electoral integrity threats and subsequently different methodologies to monitor them:

- **Chapter 2: Generative AI in the Creation and Dissemination of Synthetic Political Content:** GenAI is already playing a role in the electoral information space, particularly in social media through synthetic - or AI generated - content. Political campaigns and their operatives are using AI to generate campaign products like candidate avatars or persuasive videos or images. Cheap and widely available genAI tools have resulted in an uptick of deepfakes online in the form of audio recordings, videos or pictures, often to discredit a candidate or manipulate consensus around certain political narratives. This guide categorizes political synthetic content into four broad categories: deepfakes, fake accounts, campaign materials, and AI "slop" - each of which have their own unique impacts. Generally speaking these kinds of tactics have the potential to threaten the level **playing field** for political contestants; undermine voters' ability to make an **informed choice** on election day; and, in some cases, can **inflict online harms and exacerbate the potential for election-related violence**. While digitally enabled false or misleading information around elections is not new, genAI intensifies challenges specifically related to transparency and accountability as the content it generates is difficult to detect and source.

- **Chapter 3: GenAI as a New Source of Voter Information** (via LLMs, chatbots etc): More and more citizens are using genAI-enabled chat bots, personal assistants, web summaries, and other tools as a source of information, which surfaces "answers" based on large language models (LLMs) and proprietary data models (which will be discussed in more detail in this chapter). These kinds of tools rely on the quality of their training data and in some cases may provide inconsistent or false information about the electoral process or its outcomes. This presents a serious **voter information** and **voter participation** challenge if tools are not able to provide consistently credible or up-to-date information, and introduces issues of **confidence** and **accountability** if electoral events and shortcomings are not accurately described or summarized.
- **Chapter 4: GenAI in Election Administration:** While election management bodies (EMBs) are increasingly using technology to improve electoral processes, the use of genAI in election administration is currently limited at the time of writing. Nonetheless, the technology is entering the vernacular of election vendors and administrators. Some ambitious EMBs are already considering how genAI could be used to improve a myriad of functions, such as internal administrative functions, data management, voter outreach, and voter education. However, poorly conceptualized, procured, secured, or implemented technology can have unintended impacts on electoral integrity. Through this framing, the guide will explore the ways in which genAI may be applied in election administration, as well as the kinds of questions and interventions citizen observers should consider as part of this introduction.

These divisions may not be perfectly distinct and there may be areas of overlap. For instance, chatbots can produce political synthetic content, and may also be utilized by election administrators to help execute their duties. However this framing is built around *applications* and related electoral risks, and thus the methodologies that would be needed to understand, track, and recommend ways to mitigate impacts. In addition, these concepts are also associated with other kinds of observation in which groups may already be engaged outside of genAI monitoring, such as social media monitoring, voter education monitoring, and election technology monitoring.

Groups may not prioritize observing every facet of genAI in elections, and rather may select which one or two are the most relevant to their electoral context. Each section will include an introduction to the concept and challenges itself, as well as how to contextualize it to a group's specific electoral environment, and what monitoring strategies, approaches, and tools should be considered, including timelines and how to integrate into relevant monitoring efforts that may be ongoing. The final sections of the guide focus on communication and advocacy, as well as thinking forward and preparing for the future.



CHAPTER 1: PLANNING FOR OBSERVATION

Election observers frequently adjust their methodologies to meet evolving tactics that could undercut credible electoral processes. Integrating the observation of generative AI threats should be no different, and should be designed to meet the local challenges and concerns observers anticipate. Electoral integrity risks will vary from country-to-country, as do the popularity and use of generative AI so there is no "one size fits all" monitoring methodology. In addition, like many emerging technologies, this topic can include large and diverse datasets, especially related to social media which is borderless and virtually limitless in volume. To manage, observers should develop objectives that are clear, realistic, narrow in scope and are derived from a preliminary assessment of genAI-based risks. This section discusses how observers can outline their observation strategy and what issues to consider in planning.

Monitoring genAI in elections may look different than traditional observation, with less focus on field-based deployment and likely more centralized data collection and analysis. However groups can rely on their nationwide networks to help guide risk assessments including data-driven information about how citizens are using generative AI technology. This will help shape objectives and methodologies.

ASSESSING SPECIFIC GenAI-RELATED RISKS

The nature, vulnerabilities, mitigating factors, and opportunities around genAI in elections will vary significantly from country to country. The likelihood and impact of real risks in a given electoral environment should shape appropriate and context specific methodologies. Consider the following questions to determine if and where genAI threats to electoral integrity exist in a group's given operating environment. Some of this information may be collected from election or media experts, and should also ensure that information takes into underrepresented populations - for instance, in distinct socio-demographic areas and rural areas where media consumption is not as thoroughly tracked or assessed. Increasingly journalists, academics and even companies are exploring AI related issues within country contexts and publishing their findings. These reports can be practical resources for election observers. And if these are lacking, an informal, semi-structured interview(s) with relevant local experts can help supplement the information needed.

General understanding

- What is the media and information landscape in your country, and where do people get their political information and news in general? If online, which websites and platforms are most popular?
- What aspects of the electoral process are most likely to be undermined by false or confusing information (for instance, voter registration process, independence of the EMB, accuracy of results, candidate platforms, etc)? What electoral aspects are you most concerned about generally and why?
- Is the EMB a trusted source of information? What institutions are the most trusted sources of information?

- Is there any relevant legislation attempting to promote transparency or accountability in generative AI especially in the political context? If so, is it enforced and how?
- To what extent have tech companies been responsive to electoral integrity threats in the past in your context?
- Are there other actors - international or local - that are working on civic technology or information integrity issues? What is their focus?

Synthetic political content

- Are political parties or candidates using genAI and how? What about partisan supporters? What about influencers or other content creators?
  - What types of content are they creating? What are the subjects? Who are their target audience?
- Has there been a rise in deepfakes or cheap AI-generated images or videos? On what platforms?
- Has inauthentic or coordinated behavior on platforms been a problem around elections in the past? This kind of information can be found in media monitoring and election observation reports, as well as self-reported threat analysis from major technology companies.[10]
- Are any efforts under way - legal or otherwise - to address the use of deepfakes and other synthetic content in elections?

Voter information via chatbots and LLMs

- How frequently are chatbots - such as ChatGPT, Claude, Gemini, etc - or web summaries used by the general public? Which ones are most popular? What do people typically use them for?
- Does chatbot usage vary by demographics (for instance, age, language, class, gender, ethnicity, geography), and if so how?
- Do people in your country generally trust the information they receive from chatbots?

Election administration

- What has been the EMB's history in procuring and implementing new technology in elections? Has it been transparent? Has it had any logistical problems?
- Has there been any discussion of the EMB using genAI related technology, for internal operations or otherwise? If so, has there been discussions about vendors or the types of genAI models they are considering? And for what purpose?

For each genAI related risk (synthetic content, chatbots, and in genAI election administration), consider both what is **likely to occur** and what will have the **greatest impact** and plotting this on a risk matrix as a tool for prioritization. The impact may be determined by how many voters would ultimately be affected or the level of influence on the results. This will help determine where the most resources should be dedicated.

10 Examples such as **Meta's Threat Disruptions** and **Google's Threat Intelligence**

PLANNING CONSIDERATIONS - AND CHALLENGES - FOR OBSERVERS

Following the risk assessment, groups should narrow their scope and identify objectives that are clear and discrete. While the monitoring methodology will vary depending on the application(s) organizations decide to focus on, there are some overarching considerations that groups will want to note early on and before they have fully developed their methodology. In particular:

- **Monitoring methodologies and approaches should be shaped and driven by objectives and organizational capacity, not by available tools.** Methodologies should seek to achieve identified objectives, and only after discrete areas for observation are clarified should groups identify relevant tools that meet the needs of the project and the organization's technical and human resources. While access to tools may be important to project implementation they should not define the methodology. For instance, this means defining a synthetic content monitoring approach first, before seeking out the tools that may help with data collection.
- **Navigating changing policies and interventions by platforms and AI companies.** The use of generative AI in social media, assistants, chatbots, and support technology is constantly changing, as are regulations and policies around it. Before embarking on a project, it will be important for groups to understand the current policies in place by genAI providers and platforms where it is used, which can help create some parameters for accountability. For instance, some tools may have restrictions on genAI election-related content. However these may need to be revisited throughout an observation project as policies evolve quickly. In addition, terms of service, community guidelines and broad platform policies are often inconsistently applied across markets, making such benchmarks difficult to truly monitor.
- **Integrating into broader observation efforts:** As mentioned, in some cases the use of genAI in elections may be a subset of a larger theme observer groups are already looking at. If that's the case, it may not make sense for observers to create a whole separate branch of monitoring teams or checklists just because they are looking at genAI. Rather they may consider how to alter or update preexisting monitoring strategies and data collection mechanisms to accommodate for this new challenge. For instance, monitoring online synthetic content may be a subset of social media monitoring. Monitoring the use of chatbots and LLMs in voter information may be a part of a broader voter education monitoring effort. Monitoring the use of genAI in election administration would likely fall under broader observation of the introduction and integration of new technologies in election administration. There may also be instances where a group focuses specifically on one of these issues, and coordinates with other groups that are monitoring different aspects of a given election. Integrating these observation methodologies is discussed in further detail in each of the relevant sections.

RESOURCES AND BUDGETING

Monitoring generative AI in elections does not necessarily need to be resource-intensive, especially if groups are already involved in a broader or ongoing observation effort. For instance, most of the methodologies discussed in the following chapters do not require field deployment, and can largely be conducted centrally with relatively limited staff.

Timing is discussed in more detail in the methodology chapters, but the focus of the observation will impact how much time should be dedicated to setup and implementation. That said, most of the evident threats to genAI in elections are likely to occur in the pre-election and in some cases, post-election periods, so groups should plan on long-term monitoring. Other human and financial resources that groups may need to consider include:

- **Language capacity:** Many of these methodologies rely on analyzing information or content that voters consume. Therefore organizations should be staffed appropriately to be able to analyze data in all relevant languages in their electoral process. Having reliable interpreters or on-staff language capacity to accurately assess content, chatbot prompts, etc will be critical to this part of monitoring.
- **Research capacity:** As discussed further in Chapter 3, deepfake detection may take time to investigate. Recruiting at least some monitors or analysts with investigative research skills could enhance the effort.
- **Equipment:** Since a bulk of the monitoring will occur in digital spaces, laptops, stable internet, etc should be available to ensure observers are able to do their jobs.
- **Software:** Some of the discussed methodologies may require the use of specialized tools, such as social listening tools or deepfake detection services, which often have subscription fees and data or user limitations often associated with different pricing tiers. Basic access plans may be the only affordable option for civil society actors, who are often working on limited funding.
- **Cost of chatbot subscriptions:** Many chatbots now require monthly subscriptions for their use, or their use after a certain number of prompts or to provide certain kinds of services. While these costs are currently relatively inexpensive, groups will need to budget for them for the period of the observation, which may include several subscriptions for however many models they intend to monitor.
- **Considering resources for psychosocial support to social media researchers:** Social media monitors are, by design, often exposed to borderline and problematic content that contributes to the erosion of trust in democratic institutions and societal cohesion. The work is important, but can be mentally and emotionally burdensome. Some groups have **built safety and anti-trauma protocols** including psychosocial services and staggered workflows into their observation planning to ensure staff are able to work effectively and resiliently.
- **External communications:** Observation findings can only have meaning if they reach the voters and stakeholders involved in the electoral process. Monitoring activities should include attention and resources for external communications, including press conferences, graphic design, and traditional and/or social media advertising, as well as in-person outreach to political parties, EMBs, technology companies, other civil society organizations, and other interested parties.

CHAPTER 2: MONITORING SYNTHETIC POLITICAL CONTENT

One of the most covered - and worrisome - threats of AI in elections is the role it can play in generating synthetic political content and accounts, which can substantially undermine the information ecosystem around elections. GenAI makes synthetic content by analyzing vast datasets to learn underlying patterns, then uses those learned patterns to generate new, original outputs that mimic real-world data, like text, images, audio, and video. While information manipulation in elections is not new, the convincing and easily deployed outputs from generative AI risks exacerbating issues related to voter confusion and mistrust, the level playing field and even prospects for electoral violence.

UNDERSTANDING SYNTHETIC POLITICAL CONTENT

Synthetic content that can impact the electoral ecosystem can be categorized into four broad subtypes, which vary in both sophistication and intent.

Deepfakes

"Deepfake" is a broad term that refers to realistic video, images or audio that are AI-created and often used for **deceptive** or **malicious** purposes. In recent elections, the most prominent use of deepfakes have been to discredit candidates or parties, although in some cases contestants have used deepfakes to elevate their own image.[11] However there have also been instances where deepfakes have been used to discourage participation or misinform voters (for instance, fake news videos about candidates pulling out of a race). Given their convincing nature, deepfakes have the potential to seriously undermine confidence in the process, especially if they are used to trick voters about the credibility of the elections or faith in the election administration.

It is worth noting that sometimes deepfakes may be employed for **satirical** purposes, to elevate political or social messages with less intent to deceive, while in other cases deepfakes have been used by campaigns as more benign **voter outreach** and education.

11 For examples of deepfakes in elections, see **https://restofworld.org/2024/elections-ai-tracker/**

Deepfakes vs "shallow fakes": Does it matter?

Though deepfakes are increasingly appearing as part of political campaigns and other online political discourse, their impact on electoral integrity at the time of this guide was unclear. This is because **human-generated deceptive or inflammatory content continues to have similar impacts** and is shared just as widely. This includes among other things, non-genAI foreign coordinated influence campaigns, state media, and non-genAI image manipulation. In this sense, perhaps the impact is the most relevant issue, and not the technique or technology used. However genAI intensifies challenges specifically related to **transparency and accountability** as the content it generates is difficult to detect and source. Simply put, deepfakes may take more time and effort to debunk. However observers should be aware that many different tactics or methods of information manipulation can impact how citizens view their electoral process and their political choices, and deepfakes are simply one piece of a larger ecosystem.

Observers will need to consider intent and impact when evaluating deepfakes.

A genAI video of then candidate President Catherine Connolly announcing her withdrawal from the race was circulated ahead of Ireland's presidential election in October 2025. This kind of deepfake is not part of a discrediting campaign, but rather seeks to impact voters' participation in the process itself.

Deepfakes have also spurred anxiety about the potential to increase the **"Liar's Dividend."** This refers to the ability of bad actors to deny factual evidence - for instance an image or audio recording - by attributing it to generative AI. As deepfakes become more and more realistic, it may become easier for political leaders to avoid accountability. This is part of a larger challenge to modern information ecosystems, where credible and noncredible information is blurred, causing citizens to increasingly mistrust facts and institutions.

Synthetic content and violence against women in elections

The ease and rapid spread of deepfake technology has contributed to a rise of violent and/or sexually explicit deepfakes and AI-generated pornography that has been used to threaten, harass, and discredit women in political life. This includes using images of prominent women political leaders in realistic nude images and videos and other image-based abuse without the subject's consent with the intent of discrediting them and discouraging their political engagement. While all public figures are at risk of becoming victims of deepfakes and other manipulated content and misuse of their images, genAI content targeting women in politics is much more likely to be gendered in both intent and content and include sexualized or demeaning imagery. Observers should consider specific tags or tracking for this issue if monitoring synthetic content, or partner with women's rights organizations that may already be engaged on the topic.

Online Campaigning

Political and government actors are increasingly using genAI for online campaign materials and propaganda to influence public opinion. This most commonly appears as dramatic imagery (images and videos) designed to elicit an emotional response. While there is some overlap between propaganda and deepfakes, this type of genAI content is usually less deceptive and more **persuasive**, mobilizing or meme-building. For instance, genAI campaign materials may include imagery that looks realistic but is obviously fake, such as depicting a leader as a war hero or fighting a fantastical enemy. While using genAI for online campaign materials falls under free speech and isn't inherently problematic, some genAI content have been accused of fear-mongering or exploiting racist stereotypes. GenAI content has been used both "officially" - in which the associated campaign or political contestant is clearly identified - as well as less transparently, where the source is neither clear nor accountable. While parties, contestants, governments, and other actors have long-used powerful imagery in campaigns to compete for votes, the difference with such genAI content is that it can be developed and deployed **quickly and cheaply**, easily reaching large audiences online.

Synthetic images of India Prime Minister as an astronaut and a soldier were among many that were circulated and included in online political ads in India in early 2024 ahead of the country's national elections. While these images are not necessarily intended to deceive voters, they serve to influence public opinion. Their prevalence in Facebook political ads also violated the platform's own policies about generative AI content in ads, and highlighted a gap in enforcement.

Fake Account Generation

Generative AI exacerbates the challenges of fake accounts in social media discourse. Random face generators are easily accessible online, often for free, meaning fake accounts can be scaled up quickly with minimal effort.[12] Unlike non-genAI bots, sock puppets[13] and fake accounts, genAI accounts are more difficult to detect since they use realistic and original images. Researchers and observers have often used Open Source Intelligence (OSINT) methods[14] - for instance, reverse image searches, people search engines, or public records - or online bot detectors to evaluate the authenticity of accounts, which do not work as well on genAI accounts. GenAI accounts cannot be tracked to a hijacked profile or stock photos, and they are able to "act" more naturally online similar to a human user, evading bot detectors that look for signs of automated behavior. In addition, genAI allows these fake accounts to be produced at speeds and volumes that can dwarf previous efforts.

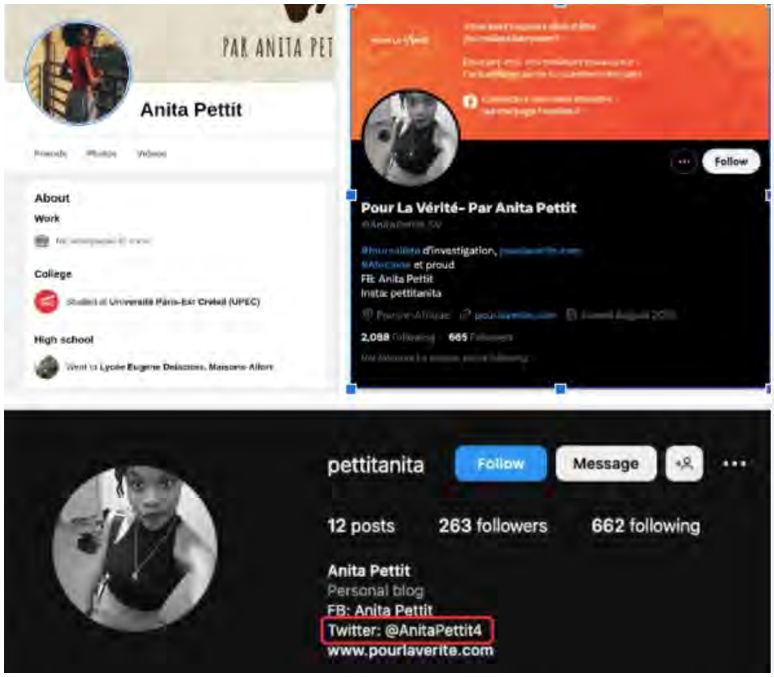
"Agentic AI": Agentic AI refers to autonomous artificial intelligence systems that can independently plan, decide, and execute complex tasks with minimal human intervention to achieve a goal. Unlike traditional AI that needs constant prompting, agentic AI perceives its environment, reasons through scenarios, takes action, and learns from the results, making it a more proactive and adaptable problem-solving technology. This allows for the automation of multi-step workflows, but also for development and deployment genAI information and communications across platforms.

12 For instance **https://this-person-does-not-exist.com/en**

13 Sock puppets are false online identities used for deceptive purposes. Unlike bots, sock puppet accounts are usually operated by humans.

14 OSINT is the practice of collecting and analyzing information from publicly available sources to produce actionable intelligence. These sources can include websites, social media, news outlets, public records, and more. For more information see **Bellingcat** and the **OSINT Framework**.

GenAI fake accounts often employ deepfake technology to produce realistic profiles and interactions, but their threat is not just about their content but also the **source**. By appearing as genuine users supporting a certain cause or narrative, these accounts are able to manufacture consensus around political narratives and build more convincing **astroturfing** campaigns. This allows political actors to create the appearance of credibility for marginal campaigns by making them appear popular and/or that those with expertise or legitimacy support them. They can also appear as credible or reliable sources of information, such as a journalist, academic, or other expert, when they are actually designed to deceive.



A fake, genAI avatar posing as an investigative journalist named Anita Pettit garnered followers and seeded news stories across West Africa. Fake accounts that utilize genAI imagery are being used across social media platforms to spread fake news stories but also cryptocurrency scams and other spam.

AI "Slop"

The final category of synthetic content is what's largely referred to as AI "slop." This is characterized by low-quality genAI content - typically images or videos, although it can also be written content or music. AI "slop" tends to be attention-getting for being silly or absurd, or simply for its prevalence, and is sometimes used in data-farming or money-making scams. While slop is not necessarily political in nature it can be used to overwhelm online spaces, and divert discourse which can crowd out genuine voices. This can even manipulate the algorithms of social media platforms, which may amplify AI slop for its ability to garner attention, and can seep junk content into the training data used in LLMs.

Synthetic content type	Intent	Compound electoral integrity threats	Other applications
Deepfakes	Deception	"Liar's Dividend", level playing field, exacerbate voter confusion, erode confidence in institutions	Satire
Campaign materials	Persuasion	Undermine genuine political discourse, exacerbate polarization	Satire, voter outreach/ mobilization
Fake accounts	Deception	Astroturfing, manipulated consensus	Satire
AI "slop"	Attention	Deception, Undermine genuine political discourse	Spam and marketing

DEVELOPING A MONITORING METHODOLOGY

Defining the Foundations

As with all monitoring methodologies, groups will need to define their scope, including what they will monitor and when. This should be driven by a group's preliminary assessment. Objective(s) should be rooted in what observer groups want their monitoring to achieve and reflect electoral integrity issues and risks. For instance, a group may focus specifically on the fairness of the playing field for candidates, or voters' access to credible information about the election, or ways that narratives are eroding democratic values, or some other challenge specific to the local context.

Then observers will need to identify the parameters for the observation. This includes answering the following questions:

- **What will you monitor?** Which platforms in particular will the observation focus on. These should be informed by the initial assessment of the information environment, with a focus on where voters are most likely to get political and electoral information, to the degree it's available. It's important to note that synthetic content often jumps across platforms, and tracking multiple platforms can help provide insight into how narratives migrate.
- **Who will you monitor?** This includes which types of accounts and/or groups, and how they will be coded (for instance, political contestants, government officials, party supporters, influencers, media, civic groups, etc).
- **When will you monitor?** What is the timeline for the observation? This may include continuous throughout the broad preelection period, only during the official campaign, including election day monitoring, and/or post-election. Ideally, the longer the observation the better for understanding how synthetic content is being deployed and used for an election. While the majority of the focus may be in the pre-election period, groups should also consider examining relevant developments on election day and in the immediate post-election period, particularly if the results are disputed or there are legal challenges to the outcome.

- **How will you categorize and measure data?** This includes how information is tagged or coded as part of the data collection process. There are many ways of categorizing information while tracking synthetic content, and this should be driven by the established objectives and Some particular pieces of information that could be useful to gather may include:
  - By types of synthetic content (deepfakes, propaganda, fake accounts, AI "slop")
  - By media type (video, image, audio, animation, etc)
  - By intent (deception, persuasion, confusion, attention)
  - By electoral threat or impact (undermine confidence in the electoral process, voter misinformation, polarization, violence against women)
  - By other narratives important to your own context and objectives (for instance, nationalist, pro-military, etc)
  - Whether content was labeled as genAI/not labeled as genAI

Anticipating innocuous synthetic content

Not all synthetic content, even those that are political in nature, may be malicious. As mentioned, genAI may be used for satire or voter mobilization. Groups will need to have clear guidance on how to distinguish between malicious or misleading or more innocuous synthetic content. It may be still be relevant to collect data on any use of synthetic content and the elections if you are interested in measuring its use, but that may entail having to provide a range of indicators to be able to measure electoral integrity impact. Overall groups will need to practice prioritization, and resist the pressure to capture and catalogue everything. This may mean they not coding for every single AI meme or slop image, and focusing only where there is a plausible link to electoral integrity impacts (participation, level playing field, violence, legitimacy of results).

Once clear parameters and coding are established, groups can use evaluation metrics to understand impact. This might include analyzing:

- Reach (likes/reactions, shares, comments)
- Followers (number, who)
- Timing (when posts occur, when accounts were created)
- Presence across different platforms, following how synthetic content moves
- Compliance with any relevant legal frameworks
- Compliance with platform's Terms of Services

Much of this methodology is not drastically different from more typical social media monitoring, of which there is already **significant guidance**.[15] If a group is already conducting social media monitoring or other information space monitoring, it's strongly recommended that monitoring synthetic content be integrated into preexisting methodology rather than setting up a separate effort. That may mean considering genAI as a sub-category in the broader analysis. For instance, a group may be labeling authentic and inauthentic content as "undermining confidence in the electoral process" but could also have a separate additional code to tag AI-generated content. Similarly, if groups are already monitoring campaign materials and messages that engage in fear-mongering or exploit racist stereotypes, then simply including additional notations for content that appears AI generated may be all that's needed.

Data Collection and Tools

Most groups rely on bespoke scraping or out-of-the-box social listening software to collect account and content information from online platforms. However these tools have their own challenges. Social listening software, which is often designed for businesses, is becoming increasingly expensive for civic groups. In addition these tools also have limitations with data access from platforms and data analysis features since they are less customizable. Social listening tools may require some trial and error to make sure they are collecting the kinds of data that are most helpful for the observation. Groups could also use supplemental tools such as applying to Meta's Content Library or political ads libraries for additional data, although there are increasing data access challenges across the spectrum of social media platforms.[16][17] Some platforms, such as TikTok, have limited automated monitoring options readily available to users (at the time of this writing) and many researchers have set up manual monitoring methods for those feeds.

The challenge of detecting deepfakes

As previously noted, deepfakes are more difficult to identify than other faked imagery and audio. Particularly at a quick glance - which is how most users consume social media - deepfakes can easily appear to be real content. It often takes more discerning analysis to debunk genAI deepfakes.

Researchers and media forensics experts have adopted several methods for analyzing possible synthetic content to determine its authenticity. For instance, for images and videos, researchers will often look for **inconsistencies, incongruities, and logic fallacies** within the image, such as unnatural hand or arm positions, disappearing or misconnected objects, mismatched shadows or glares, or other minor distortions or violations of the laws of physics that may not be immediately obvious. GenAI images and videos often have blurred or low-quality backgrounds compared to the main subject, and mangle text or numbers. Similarly, genAI audio may struggle to produce realistic background noise, or the natural breath and rhythm of human speech. However as deepfake technology continues to advance, these kinds of errors are less frequent or obvious and deepfakes may be more difficult to detect with the naked eye. Therefore researchers may want to deploy a **mixed methods approach** to identifying deepfakes, discussed in more detail below.

15 For instance, see DRI's Social Media Monitoring Toolkit: **https://digitalmonitor.democracy-reporting.org/toolkit/**, or the ODIHR's Guidance for Observation of Electoral Campaigns on Social Networks: **https://www.osce.org/Observing_elections_on_Social_Networks**

16 **https://www.techpolicy.press/the-worlds-growing-information-black-box-inequity-in-platform-research/**

17 **https://kgi.georgetown.edu/research-and-commentary/better-access/**

GenAI detection tools

There is a growing market for genAI detection tools that are designed to identify synthetic text, images, video and audio. Generally speaking, these tools tend to be specialized in certain types of media, so users may need to use different tools for analyzing audio vs video, for instance. Detection tools are most likely to provide a "range" of certainty regarding the authenticity of a piece of media, for instance, providing a percentage likelihood that an image is real or fake. However the accuracy of these models is inconsistent at best, and in some cases different tools will produce different - and sometimes conflicting - results.

There are some detection tools that are free or free for a certain number of checks. However generally speaking the suite of detection tools available are not prohibitively expensive, and such tools come with the options of monthly or yearly subscriptions. Some detection tools offer more full-scale services, not just delivering an analysis but conducting some of the additional work of confirming the authenticity of content (discussed below). A living list of low-cost detection tools are available on the corresponding **community document**, however please note that the function, cost, and accuracy of these tools can change dramatically as this field evolves.

Despite the growing availability of these detection tools, they have significant limitations and can produce false positives and false negatives.[18] Just as synthetic content continues to eclipse traditional image analysis techniques, detection tools also struggle to keep up with genAI advancements. Once genAI designers know what tools are looking for, those weaknesses can be adjusted to avoid future detection. Other kinds of editing, such as cropping images or screenshotting them, may also make them less likely to be detected. In addition, detection models may suffer from the same kinds of biases that genAI tools themselves do, including weaknesses in discerning the accuracy of images that include a variety of races or genders. For that reason, groups may seek out regional or locally developed detection tools that may be more likely to be trained on local content.

Major AI companies are also attempting to embed their own meta-data - essentially invisible watermarks - on the genAI content their tools produce. Theoretically this would make the content detectable by the company's own detection tools. However even this practice is not fool-proof, and company tools cannot always accurately detect their own synthetic content.

Other tools

As an inverse to deepfake detection tools, some actors are also exploring **content certification** or credentialing software. These tools allow actors to add metadata to their content to **positively affirm the authenticity of the content**. These kinds of tools may be utilized by media or other credible sources to essentially authenticate content which can then avoid the Liar's Dividend or be easily fact-checked. The **Coalition of Content Provenance and Authenticity (C2PA)** has been a leader in this arena, promoting a more widespread adoption of content certification. However, while these tools may be helpful in preventing one' s own content from being manipulated, or for observers to easily confirm real content, they are generally only as effective as they are used. Just because a piece of content does not include a content credential does not mean that it is fake, it just means that the user did not use a C2PA tool. However monitors should familiarize themselves with such software and how information sources, such as media outlets, government bodies, or journalists are using them.

18 **https://reutersinstitute.politics.ox.ac.uk/news/spotting-deepfakes-year-elections-how-ai-detection-tools-work-and-where-they-fail**

A mixed methods approach

Despite these complexities, observers can reasonably track and debunk deepfakes, it just may involve the combined use of content analysis, detection tools and other means of investigation and verification. Observers can rely on monitoring content critically, and in addition to examining content for inconsistencies and incongruity, there are a number of other easy questions observers can ask to discern authenticity:

- **Examine the source:** Observers should look at the original poster of the content. What is the account? What is the name of the account? Does the account itself seem credible, like a real person or entity? What other kinds of content does the account post? Does the source make sense for the content it's posting? For instance, if it's a video of a press conference of a government official, one may expect that to be posted by media outlets, not random users.

*Realistic deepfake videos of American actor Tom Cruise are often used as examples of how convincing these videos can look. However, even a cursory investigation of the account source makes it clear that these videos are fake. The source almost exclusively posts silly videos of Tom Cruise, and the account name itself is called "**DeepTomCruise**".*

- **Examine the context:** Consider the context of the video, image, or audio. Does it make sense? Does the occurrence seem out-of-character? Is it believable that the subjects involved would be acting the way they are, or be at the location? If the answer is no, then the video/photo probably warrants further investigation.

Burkina Faso's military junta leader Ibrahim Traoré has been the subject of widespread genAI/deepfake videos, including videos of him in combat, announcing (false) multibillion dollar trade agreements, launching large scale construction projects and other narratives designed to elevate his image. An investigation into the accounts that seed these videos would show that these are not real news organizations or other stakeholders who would otherwise have access to such footage in real-time. The content has spread far beyond Burkina Faso, amplified by prominent online activists and influencers on other platforms and in multiple languages in Africa. Some analysts and media investigations also alleged the implication of Russian troll factories active in the Sahel region, in amplifying such content.

- **Check with witnesses or other primary sources if necessary:** Observers may need to take additional steps to verify events or occurrences. This may mean checking with witnesses who may have been present at supposed events, or other primary sources and OSINT techniques. For instance, it may be easy to confirm whether a news story is fabricated or not simply by calling the news agency whose logo is used. Observers may be able to confirm the whereabouts of public figures based on public calendars, meeting minutes, and other open data to confirm whether they could realistically be in a viral image or video. Reviews of other video, audio or transcripts could also help verify the accuracy of statements.

Given the additional steps necessary to debunk certain deepfakes, researching synthetic content may take more time than other aspects of social media monitoring. Observers may **consider partnerships to better facilitate fact-checking/verification, for instance, with a trusted fact-checking network or academic researchers** who may be able to help more quickly analyze and verify content. However it's important that observers not confuse their mandate with that of fact-checkers. While fact-checking groups and other media integrity initiatives serve critical functions in weeding out false and misleading narratives as they appear in real-time, monitoring by citizen election observers tends to have different goals and timelines. The objective for observers is not to quickly verify and/or invalidate individual stories but rather to identify and evaluate the impact information trends may have on electoral integrity, build accountability around a variety of actors participating in the electoral process, and provide actionable recommendations. However co-ordination mechanisms and information flow between observers, media monitors, fact-checkers and other relevant groups in the space could be highly beneficial to any electoral analysis.

Being honest about the limits of data collection and analysis: No group can observe the entire internet, and when presenting findings organizations should be transparent about the scale and scope of their observation, including what was and was not monitored. Most social media monitoring has limitations in terms of the types of accounts that can be monitored (public accounts only) and changes in platform APIs has in some cases also restricted access to the amount or kinds of data available to researchers. In addition, synthetic content may spread on even less transparent digital platforms, such as widely used private messaging apps like WhatsApp. Highlighting these challenges in statements is still important so voters understand what findings are based on, and where there are areas for further research.

Groups may also consider how to utilize their broader networks to provide context and supplemental data to their monitoring effort. For instance, long-term observers deployed in the field may be asked to collect information on predominant political narratives in their constituencies, or to understand how synthetic content may be spreading at a local level, including among what communities and on what platforms.

Timeline, Staffing and Resources

The monitoring timeline will largely depend on the scale and scope identified for this methodology and which aspects of the electoral process groups consider the most at risk. However monitoring synthetic content in elections will likely be a part of long-term observation, during the pre-election and campaign period and into election day. These efforts may last longer than other kinds of observation that may be more focused on discrete processes throughout the electoral calendar. Observers should also consider in-depth post-election monitoring, particularly in the case of disputed results, post-election tensions, or run-offs. As with any update to a monitoring effort or introduction of a new one, groups may plan for a "pilot" phase first, ideally during a non-critical electoral period, which will allow monitors to test data collection and detection tools, and refine coding methodologies as necessary before more earnest observation begins.

Like other social media monitoring, most data collection and analysis of online synthetic content can easily be done centrally with a team of media monitors/digital analysts. This may save on the recruiting and deployment of large teams of long-term observers, but the media monitors will need to be well-trained with some specialized detection skills, and budgeted for throughout the duration of the monitoring project.

Case Study: AI Forensic's Analysis of Synthetic Content on TikTok

In June 2025, **AI Forensics** released a **study** on how generative AI content, especially AI slop, spreads across major social platforms, focusing on TikTok and Instagram. The study focused on Spain, Germany, and Poland, and examined algorithmic virality, labeling practices, and the rise of agentic AI accounts which helped automate mass-produced AI content. The study found that the platforms did not sufficiently label AI content, and even in cases where it was labeled, it was not obvious to the user. This study included both automated data extraction and management and qualitative analysis of content and accounts.

CHAPTER 3: MONITORING CHATBOTS AND LLMS AS SOURCES OF VOTER INFORMATION

UNDERSTANDING CHATBOTS AND LLMS IN ELECTIONS

Chatbots - or what is sometimes referred to as "Conversational AI" - are consumer-facing tools that can interact with users, and allow them to retrieve information, products and services through a series of prompts. Some current examples of popular conversational AI models at the time of writing include ChatGPT, Co-Pilot, Gemini, Grok, Llama, DeepSeek, and Claude. Many of these models are owned by companies that also engage in broader technological services and/or social media platforms. In addition to apps, these kinds of models may appear in web-summaries (like Gemini's "AI overview" in a Google search), or embedded in digital personal assistants, writing and design tools, etc. These tools rely on vast amounts of training data which is usually proprietary in nature but may include, among other items, private and publicly available information throughout the Internet.

The application of such tools is constantly expanding and many citizens are incorporating their use into their daily lives to help execute tasks or quickly retrieve synthesized information across many data sources. This may include political, social or electoral information that can influence if and how voters engage in the process. Growing reliance on these tools also means a growing trust that they are providing high-quality information and analysis. However there are some inherent vulnerabilities in these models, in particular:

Bias in training data: LLMs are only as strong as the training data they are fed which can result in bias in the information they provide. Many studies have identified how training data can produce gendered, racial and linguistic biases, as hegemonic groups are often overrepresented in the training data itself. In addition, LLM training data can be intentionally manipulated by outside actors to skew learning, for instance **overwhelming an information space** from which LLMs pull data from with bias or false information. If the training data itself is **weak, flawed or inaccurate**, then the outputs it produces will be as well.

Hallucinations: These tools occasionally produce "hallucinations" - an output that is completely fabricated and has no basis in the actual data on which it was trained. This is essentially when a model makes up an answer, but it's presented as factual. GenAI companies have not yet solved this problem and it can create serious confusion for users.

Corporate political preference and state influence: A number of recent cases have demonstrated how the companies that own and make chatbots can inject their own politics or the politics of the state into these models. This has included recent examples of chatbots **repeating government propaganda** in lieu of addressing certain questions or engaging in hateful rants following **politically motivated adjustments** to the model. These kinds of incidents can significantly harm the credibility of these tools, as well as their perceived usefulness and market viability.

Guardrails that do exist can be by-passed: At the time of writing, chatbots and other conversational AI remains largely unregulated, and standards and restrictions on outputs are mostly set and enforced by the technology companies themselves. This often includes prohibitions on chatbots producing clearly harmful content, such as violent imagery, recommending illegal acts, defamation, or child abuse. Some may also have blocks on producing certain types of misinformation, including electoral misinformation. However these guardrails can be difficult to implement effectively, and there are many demonstrated strategies for by-passing these protections by manipulating prompts, often called "**jailbreaking**."

CONTEXTUALIZING THE THREAT

Current research has demonstrated that even advanced models may be unreliable in providing consistent, accurate electoral information in many country contexts, and across multiple - especially non-English - languages. However before monitoring organizations decide to monitor these tools, it's important to understand their use and impact in the specific electoral context. This includes using the preliminary assessment to determine which chatbots are most used in the country and by which populations, as well as how people are using them (for instance, if citizens are using them for voter information).

Some of this information may be garnered informally through interviews, or through more structured surveys or focus group research. However observers may also consider reaching out to the technology companies themselves to request information. Aggregate information regarding usage statistics (including election or political prompts) can provide important context to such an observation, and companies should feel obliged to provide it for the public good.

Groups will also want to consolidate the self-regulatory frameworks - including terms of service, policies and protocols - companies have for their chatbots, especially if there are any related to electoral integrity. Some of this can be done via desk research of their tools and websites but in areas where there's lack of clarity or information, groups can reach out to companies for further details.

DEVELOPING A MONITORING METHODOLOGY

Similar to monitoring synthetic political content, monitoring chatbots can be conducted centrally, or even remotely. Once an organization establishes a clear methodology for testing these tools, it's a methodology that could be repeated throughout an electoral cycle.

Case Study: DRI's Study of Chatbots in Sri Lanka: In 2025, **Democracy Reporting International (DRI)** evaluated four widely used chatbots (ChatGPT 4.0, Gemini, Copilot, and DeepSeek) by asking each of them the same set of 18 questions related to voting procedures and major political issues shaping the campaign in three languages (English, Tamil, and Sinhala). They catalogued and systematically graded the responses: for electoral process questions on a 0–3 scale (refusal, false/misleading, partially correct, correct), and for political-issue questions on a 0–2 bias scale (refusal, biased, unbiased). The analysis then compared model performance across models, languages, and question types, focusing on patterns of inaccuracy and partisan bias. The study found that the large language models frequently produced incorrect or misleading information on electoral process questions (e.g. about voter registration, election timing, candidate details) and provided some biased answers to political questions. The study also found linguistic inequities, where speakers in Tamil or English were more likely to receive biased or inaccurate political/electoral content than speakers of Sinhala. You can find more about DRI's chatbot research **here**.

Defining the Foundations

Observers will need to identify exactly which chatbots they will want to monitor. This may be based on which are the most frequently used by the population or if/when new chatbots come on the market. Importantly, observers will need to ensure they note which version of the chatbot they are testing, as some models are updated relatively frequently. For the purpose of comparative research, it may be interesting to examine several of the top chatbots in a given context, and not just one. Not only does this provide a more comprehensive picture of voter information avenues available to citizens, but can also increase accountability for tech companies across the market in a given country.

Data Collection and Tools

To investigate what election-related information voters receive from these models, observers should develop standardized questions. It's important that these questions remain the same - and in the same order - across languages and chatbots tested in order to compare and analyze responses. These questions will serve as the "prompts" in the chatbot. The kinds of questions observers could consider include:

- **Basic voter information questions:** This includes the kinds of information that are critical for voters to be able to participate in the process, and what election commissions and voter educators may be promoting. It could include questions about who can vote, date of election day, the voter registration process, what voters need in order to vote, who is on their ballot, polling locations and polling hours, etc. Observers should note that the electoral calendar could impact the availability of accurate information. For instance, asking about polling place location early in the process may not produce accurate results if the polling station list has not yet been released by the EMB. However, observers should note when outdated voter information is provided - for instance, not reflecting a recent change in the legal framework or other policy.
- **Policy and political questions:** This includes questions that can help voters understand the political positions or platforms of contestants, or more about their background. Regardless of the prompt, chatbots should strive to provide accurate and neutral answers which can be fraught in contexts where parties or candidates do not run

on issue-based platforms or are prone to lie about their platforms. Questions could be related to a contestant's stance on a specific issue or could also more broadly ask about contestant platforms. Questions could also include blatant recommendations, for instance what responses are provided when the user asks who to vote for.

- **Credibility and confidence questions:** Electoral information is not limited to basic polling day information and campaign issues, but also information that can help citizens understand how their elections work and to what extent they should have confidence in their electoral institutions. Organizations could consider questions related to the nature of the electoral system, the conduct of previous elections, the background of the election commission, or even the accuracy of results in the post-election period (see *Timeline* section below). These kinds of questions become challenging in contexts of disputed elections but the kinds of references or other data that chatbots surface can be revealing.
- **Human rights questions:** Elections are grounded in human rights, and election observers are human rights defenders. While it's unlikely that the average citizen will ask a chatbot about international standards or human rights mechanisms, it may be enlightening to see how these tools deliver rights-based information. For instance, what are the rights of observers? Of journalists? Are current electoral conditions in line with basic human rights principles or not? Is it ok for the government to shut down the internet for security reasons? What do I do as a voter if I think my rights have been violated?

Questions will need to have a corresponding coding rubric for the "answers" or other information provided by the chatbot, discussed in more detail below. For data collection, in addition to codes, groups will want to make sure to record the date/time of the test, language, as well as the output verbatim (ie: via screen shots or exports). It's very important to archive responses as users may have different experiences with a later updated version. This is also why **it's important to run tests across different chatbots as close to simultaneously as possible** to avoid data drift from updates.

When designing prompts, groups will need to keep in mind ethical safeguards, for instance avoiding prompts that disclose personal data such as voter ID numbers, observer names, or highly sensitive scenarios.

LLMs and open election data: The rise of LLMs and the use of chatbots should compel better and more open election data, particularly on the side of EMBs. Since these models are trained on preexisting data, they can only reflect information that is actually available. It is EMBs best interest to make sure all key electoral information, such as key dates, gazetted decisions, campaign finance disclosures, polling station lists, and polling station level results are made available online, so that the systems can retrieve accurate and credible information that is direct from the official source. Otherwise models will pull the information from elsewhere and risks inaccuracy.

Analysis Framework

There are a number of codes or metrics that groups should think about tracking when undertaking this kind of exercise. These should be determined ahead of any tests, but may involve some pilots to determine if they should be tweaked in advance of formal monitoring. Some major findings that can be easily coded and tracked include:

Accuracy: This is the main and most obvious evaluation point, particularly when it comes to voter information. Fortunately this is generally easy to assess since most voter information is either correct or incorrect. However there are times when information provided may be incomplete or partial, or provides a mixture of accurate and inaccurate information, or otherwise improperly frames certain aspect of voter information (for instance, only about rights but not policies, only part of a particular policy, only about major candidates/parties, or only at the district level, etc). Therefore organizations will want to consider how to code at least a few layers of accuracy (ie: false, incomplete, partially correct, correct)

Bias: This will be more relevant with related political and policy questions. For instance, does the chatbot recommend or endorse a contestant? Are policy claims backed up by credible sources? Observers should in particular be mindful of whether the model appears to systematically favour certain narratives or actors, especially in contexts where there is strong media capture and/or government pressure on tech companies.

Refusal or absentia: Under a number of conditions, chatbots may refuse to answer questions as part of their guardrails, especially if it could result in electoral misinformation or political bias. Observers should track when chatbots refuse to answer a question as well as what/whether it provides a "justification" for refusal, and possible sources of redirection (such as the EMB website or other potentially credible information source).

Implementation of company policy around elections: This may be similar to the above (refusal) but it's important to flag any instances where answers provided don't appear to match company policies. For instance, one standard appears to be enforced in one language but not another or not at all.

References: For sensitive information, such as electoral or political information, some chatbots will provide a summary of primary sources and then link to them in their response, as references. For electoral credibility or human rights questions, this may include links to prominent human rights organizations or election observation statements. Observers should track whether and which reference sources are provided in responses.

Consistency across languages: All findings should be coded by the language they were asked in, especially if groups are working in a context with several prominent national, regional or local languages. Analysis can reveal whether there is inconsistency across languages.

Comparison over time: As discussed more below, periodically running tests on chatbots could reveal how accurately they are covering elections and voter information over time. In this case, it could include a linear comparison of one chatbot over several models, or several different chatbots over time.

Timeline, Staffing, and Resources

Observers could run these tests once as a baseline to get a sense of the kinds of information voters may receive. However given the general ease of deploying this methodology observers could consider periodically conducting tests throughout the election cycle, and perhaps especially around key moments. For instance, observers could follow a standardized evaluation schedule, such as every 3 months in the 9 months leading up to election day. Or they could consider linking tests around specific moments in the election calendar. However it should be noted that this approach may require altering some questions to meet the specific process (for instance, more voter registration specific questions, or more campaign specific questions).

Groups should also consider post-election tests. As post-election disputes and court cases continue to increase around the world, this is an important data point for how voters receive accurate and non-inflammatory information about the credibility of the outcome and ongoing complaints process. Again this would require altering questions that would be more relevant to the electoral dispute resolution process and the results themselves. Regardless, groups should develop a timeline well in advance to plot out when to conduct periodic tests and what kinds of questions may need to be added or changed.

These kinds of tests require relatively minimal resources compared to other kinds of election observation. They can be conducted by a very small number of well-trained researchers, who are familiar with chatbots and prompts and can analyze structured datasets.

CHAPTER 4: MONITORING GenAI IN ELECTION ADMINISTRATION

EMBs are not new to artificial intelligence, and have likely utilized systems that include some kind of machine-learning in their endeavors at one point or another. However, this is not the same as utilizing *generative AI* which may have more limited use cases. While EMBs have demonstrated an interest in ways that genAI advancements can improve election administration, these are likely to be more related to internal operations rather than traditional processes that observers monitor. At the time of writing this guide, there are few strong examples of ways that EMBs are actively using genAI, and rather more ongoing discussions regarding its risks and opportunities. Vendors may also play a role in the discourse regarding genAI elections, selling tools that purport to have sophisticated AI capabilities but sometimes with limited details. However it's still important for observers to be aware of discussions related to genAI and in particular how its integration into election administration may impact the transparency, accountability and credibility of the process.

UNDERSTANDING GenAI IN ELECTION ADMINISTRATION: USE CASES AND CONSIDERATIONS

There are a variety of ways that EMBs are considering utilizing genAI in their work, and different applications come with different levels of risk. Some specific use cases include:

Internal Use

Like other office workers, EMBs around the world are identifying ways that genAI could help streamline internal processes and increase efficiency, for instance, using chatbots to conduct repetitive tasks, note taking, help with scheduling and documentation, budget allocations, and other outputs that do not directly interact with sensitive election materials or processes. This would represent a relatively low-risk and limited incorporation of genAI. Even still, EMBs should have a central AI strategy, with guidance for when and how employees can use genAI for their work, and what kind. This may include a working group or team to test and evaluate tools and provide guidance to the broader commission prior to full implementation. This will also better allow EMBs to assess whether genAI integration actually IS increasing productivity or efficiency and where it is not. In addition, the strategy should be clear about what the red lines are in terms of genAI use (for instance, not inputting sensitive electoral information to a chatbot, or requiring human review for any genAI outputs).

Voter Education and Outreach

Responsible implementation of genAI could aid EMB's in their voter education and outreach efforts. For instance, automatic interpretation and translation services could be used to make sure press conferences, statements, and other materials are made available in a variety of relevant languages easily and with minimal costs. Some EMBs have considered the development of their own AI-powered chatbots that voters can ask questions to and interact

with. Such tools could make it easier for voters to receive accurate election information that may otherwise be buried on an EMB's website. However an EMB chatbot would suffer from the same vulnerabilities as other genAI chatbots, such as hallucinations, language inconsistencies, and bias. In addition, the proprietary nature of most LLMs would make it difficult for EMBs to truly "own" or audit the model, and access to granular metrics can vary widely by vendor and contract. This means EMBs potentially would not have access to important metrics, and voter feedback - for instance, what types of questions voters most frequently asked, or how often voters received a hallucination instead of a correct answer. In addition, just like apps and online portals, the chatbot would only be effective as the number of voters who know about the tool and use it. It's much more likely that voters will use more popular chatbots for information, rather than one unique to the electoral process.

Other Electoral Processes

There are other election processes that EMBs theoretically could use genAI for but at the time of this writing they were not, and the trajectory of genAI itself, its application and regulation will have a significant impact on these factors. For instance, some EMBs have looked into using genAI for things like EMB budgeting and resource allocation - to help EMBs use historical data to predict budgeting needs for an upcoming election and maximize their disbursements. EMBs may also use genAI for other predictive information, such as identifying electoral hotspots. However these are vulnerable to the accuracy of preexisting and training data and the extent to which they accurately reflect genuine needs for a country (resource allocation generated from generic models will not be able to address the context of a particular country, and in models that are only based on country reporting will be limited by the inclusion or omission of data for that specific country).

Even external electoral processes that can impact electoral integrity could be augmented by genAI, such as redistricting processes or processing candidate nominations. However these kinds of processes are often not compromised because of innocent mistakes or for being overly burdensome in timing/resources, but because of political will. EMBs and observers should keep in mind that although genAI as a tool may help address some issues related to human error or resource constraints, it cannot overcome threats that are fundamentally political in nature, such as corruption or political will challenges. As a tool genAI models cannot overcome political will or corruption challenges. In addition it would be critical that any tool be able to garner public confidence in the process. Even tools that are high-quality and well-tested may still negatively impact the process simply for lack of public trust or understanding.

DEVELOPING A MONITORING STRATEGY

GenAI use in elections should be investigated against well-defined key principles that uphold international and regional electoral integrity standards: transparency, usability, auditability, secrecy of the ballot and protection of personal information. Poor planning, procurement and implementation of genAI in election administration could significantly undermine electoral integrity, specifically related to transparency, accountability, and trust in the process. Like other new technologies in elections, many AI risks stem from how systems are integrated, governed and secured, rather than from the underlying model alone. Observers should consider questions that they would consider as part of monitoring any new technology in elections. This includes:

Relevance, appropriateness, and transparency of tool adoption: As with the introduction of any new technology in the election space, decisions should be solution-oriented and deliberate. For instance, EMBs should consider whether other machine-learning or data analytics tools - which may be simpler, more transparent, have more public trust, or are less resource intensive - could address the same challenges.

Transparency in procurement, testing, certification and audits: What information and documentation is available about procurement? What mechanisms are in place to ensure independence from the vendor and proper oversight? How are products tested and is that process accessible to observers?

Voter confidence and information: To what extent are voters aware of the introduction of new technology and what information do they need to understand and trust its deployment? If voters are expected to use or interact with the technology, what processes and procedures are in place to ensure easy adoption? Are special functionality and rollout considerations needed for certain voters, such as persons with disabilities, voters with limited access to internet or smart phones, or the elderly?

Capacity and sustainability: EMBs should understand the level of technical learning necessary for employing genAI tools, the resources necessary to sustain them, how to create independence from and oversight of the vendor, how to build in quality control mechanisms, how to set up redundancies and backups, and ensure proper testing and pilots that minimize risk. Many EMBs lack staff with appropriate technical skills and genAI tools can be expensive and complex, making more straightforward analytics tools a better choice. Election integrity best practices for any new electoral technology extend the principles of accountability and transparency to technology vendors. Introducing any new technology requires an understanding of how the technology is built and ensuring it will perform as expected. This presents a problem with genAI, as the models are proprietary.

However genAI may introduce additional considerations given its unique capacity for deep learning, original output production, and reliance on training data. There may be some additional aspects of these concerns that are specific to genAI, for instance:
GenAI policies and protocols: This may include whether the EMB has an existing written AI strategy and particular protocols in place that guides how staff use genAI tools, specify which tools and for what purposes, as well as oversight and review of genAI-enabled outputs.

Human oversight and incident response: Observers should examine what data ownership and accountability mechanisms are in place. For instance, who is culpable if AI-enabled tools provide bad information? What are EMBs doing to ensure that there's a human "in the loop" for anything AI-generated? What quality control mechanisms exist?

Bias mitigation and fairness: As mentioned, learning data used by LLMs can create biases. How are EMBs looking for and mitigating possible bias in their own tools?

Data management and privacy: If any genAI models used by EMBs use or are trained on voter data that the EMB provides, how are voters made aware of this?

Data Collection and Tools

Most information related to genAI use in election administration can be collected through key informant interviews and analysis of **open election data**. This includes asking about internal AI policies and risk assessments, as well as looking at publicly available information related to EMB decision-making, budgets, procurement information, risk assessments and monitoring and evaluation studies. For additional details on monitoring the introduction of technology in the electoral process, observers can find guidance in NDI's **Toolkit on Monitoring the Impact of New Elections Technologies**, or ODIHR's **Handbook on the Observation of Information and Communication Technologies (ICT) in Elections**.

Timeline, Staffing and Resources

Staffing and timeline will be largely dependent on the technology that is under examination. Like most election technology observation, groups should be prepared to ask questions about genAI in election administration very early in the process, since that is often when those decisions are made. Most monitoring related to procurement - for instance, the decision-making process, applications and selections - can be done centrally with limited staff. However if organizations want to monitor a technology that's being deployed to EMB units around the country, than this may be something that has to be incorporated into long-term field observation.

CHAPTER 5: USING DATA TO SHAPE GenAI POLICY SAFEGUARDING ELECTIONS

COMMUNICATING YOUR FINDINGS

Organizations should also have a public communications strategy that reflects their specific methodology or methodologies used. The monitoring methodology will ultimately drive the communications schedule. For instance, if a group is monitoring LLMs, they should plan to release findings after each full-scale "test." Groups should build periodic reports into their communications plan, and include the methodology and any data limitations. These findings should not just be communicated to the public and stakeholders but also the tech companies that are responsible for the LLMs and genAI models. This could result in change prior to an election. Groups may also consider rapid response partnerships to quickly address pressing issues uncovered that could impact this process; this may include trusted journalists or media outlets, EMB officials, security forces, technology companies, etc.

Observer groups may release periodic reports as well as critical incidents, and should utilize engaging techniques to highlight specific points, including shareable infographics and social media outreach. Generally speaking, observers should communicate findings plainly with their audiences, which can be difficult when the topic such as genAI which is largely hyped and often misunderstood. Findings also should realistically reflect windows for data collection and analysis.

UNDERSTANDING THE CURRENT REGULATORY FRAMEWORK

At the time of this guidebook, there are relatively limited regulations for generative development, deployment and applications despite consumer-facing genAI models already being readily used by the public.

Technology Companies Policies

As mentioned, technology companies that produce genAI models or host content have developed self-regulatory frameworks, protocols and policies supposed to guide their systems. This includes disclosure and labeling requirements for genAI content on most social media platforms, and bans against deceptive AI content around public figures, events, or news stories. Some of these commitments were highlighted in the 2024 Munich Tech Accord,[19] where major technology companies signed on to a set of AI-related safeguards during a major year of global elections.

19 The Tech Accord to Combat Deceptive Use of AI in 2024 Elections is a set of commitments to deploy technology countering harmful AI-generated content meant to deceive voters. The 27 signatories pledged to work collaboratively on tools to detect and address online distribution of such AI content, drive educational campaigns, and provide transparency, among other concrete steps. For more information: **https://securityconference.org/en/aielectionsaccord/**

However companies struggled to meet many of these commitments. Social media platforms have failed to consistently identify AI content and enforce labeling requirements. Meanwhile many platforms are also actively incentivizing AI content producers, while divesting in electoral integrity safeguards, content moderation, and fact-checking. It is also worth noting that the Munich Accord has not been renewed, and was relevant only for 2024. Voluntary commitments can be useful when they create effective peer review and positive reinforcement, but they otherwise lack meaningful enforcement mechanisms.

Summary of Commitments of the 2024 Munich Accord

- Develop and implement technology to mitigate risks related to Deceptive AI Election content.
- Assess AI models to evaluate the risks they may present regarding Deceptive AI Election Content.
- Seek to detect the distribution of this content on their platforms.
- Seek to appropriately address this content detected on their platforms.
- Foster cross-industry resilience to Deceptive AI Election Content.
- Provide transparency to the public regarding how the company addresses it.
- Continue to engage with a diverse set of global civil society organizations, academics.
- Support efforts to foster public awareness, media literacy, and all-of-society resilience.

International, Regional and National Frameworks

There are several international and regional genAI frameworks, some of which are loose commitments, while others hold more legal standing. There is also a growing patchwork of national legislation. Many of these frameworks focus on transparency, accountability and human-centered values, and in some instances, like the EU's AI Act, include explicit provisions on "high-risk" systems, including those that could affect democratic processes or voter behavior. However in most cases these frameworks are broad, and don't yet operationalize election-related AI risks, leaving ambiguity for electoral commissions and technology platforms.

At the time of writing, some international and regional frameworks include:

- United Nations **Global Digital Compact**
- UNESCO's **Recommendation on the Ethics of Artificial Intelligence**
- OECD **AI Principles**
- G7 **Hiroshima AI Process Comprehensive Policy Framework**
- **Council of Europe Framework Convention on AI**
- EU **Artificial Intelligence Act**

- **Principles and Guidelines for the Use of Digital and Social Media in Elections in Africa**
- The Association of Southeast Asian Nations (ASEAN) **Guide on AI Governance and Ethics**

Trends and Considerations

There are many ways to consider genAI governance. Most genAI legislation focuses on the use case rather than the underlying technology, which does open up space for elections-related risks and mitigation. This makes sense in particular since many election-related challenges may be exacerbated by generative AI but are not necessarily unique to generative AI, as mentioned in previous sections. Some international and regional frameworks have examined regulation in terms of low-risk or high-risk, with only high risk considered for specific interventions. Other regulatory frameworks have considered different necessities for consumer-facing products - that are available for the public to use for their own individual purposes - versus non-consumer facing products that aid in academic or business functions.

Addressing election-related deepfakes in legislation

At the time of writing, several US states and countries such as India, Brazil, and EU member states had adopted legislation and/or secondary regulations to mitigate the impact of deepfakes in elections. These typically include outright or time-limited bans on deceptive political deepfakes, which include materially deceptive media about candidates in a defined pre-election window, and/or disclosure / labeling requirements for AI-generated political content. For the latter, this framework allows for the use of AI and even of altered media, but requires clear labels (on-screen text, audio disclaimers, or embedded metadata) indicating that the content is synthetic or manipulated. Both approaches are often paired with private enforcement in which candidates can seek injunctions and damages, as well as platform obligations, where election authorities or regulators can order takedowns or require labeling at the platform level. However these regulations are very new at the time of drafting and the ease and efficacy of their enforcement and deterrence effect is unclear.

Again in most of these cases, elections are not the focal point as there are also areas of concern, including copyright infringement, health and safety, and broader societal risks. That said, select countries and states have started to try to address deceptive genAI in elections in more direct fashion, including several pieces of legislation that is focused on banning candidate deepfakes, or requiring transparency and disclosure (labeling requirements) of genAI content. Implementation and enforcement of these regulations remains challenging, however, and at the time of writing this guide, the regulations were so new that an assessment of their effectiveness is unclear.

DEVELOPING RECOMMENDATIONS AND ADVOCACY STRATEGIES

Recommendations from observers should be evidenced-based, rooted in their specific observation and findings, and grounded in international principles. Focus on trends and impacts on electoral integrity - including how genAI is actually being used and not how we may be scared it could be used. Being discerning about trends can be helpful in more realistically understanding the trajectory of genAI in a given electoral context. This also includes being wary about recommendations that could be easily exploited to increase censorship or reduce freedom of expression, especially by authoritarians who have weaponized the information environment to silence diverse voices. A **recent comparative** review of information and media laws by IFES underscores that legal responses to genAI must meet democratic safeguards of legality, necessity, and proportionality. Observer recommendations should reference these standards to ensure that calls for regulating synthetic content avoid overreach or rights restrictions. Some key principles to keep in mind that can help aid a long-term advocacy strategy include:

- Focus on **transparency and accountability:** These are fundamental principles that are very difficult to exploit for the worse. At the end of the day, greater transparency regarding election-related content, online and via chatbots, is not restrictive and helps build accountability around information producers and amplifiers and the political entities that use or benefit from them.
- Respect **freedom of expression, open campaigning, and voter information:** These are key components to electoral integrity and should be prioritized when considering potential tradeoffs in regulating genAI. Contestants have the right to compete for votes including through outreach, advertising, and other campaign methods. Groups should be careful about recommendations that could be seen as undermining the fundamental right of contestants to compete for votes. In addition, the information space offers robust opportunities for voters to get information, including low-information groups that previously had limited avenues for voter information. In countries where the media is captured by the state, digital and social media remains one of the only outlets to receive independent information.
- At the end of the day, appropriate genAI governance should be framed as a **measure for protecting basic freedoms** to ensure that contestants can compete fairly, observers have the information they need to evaluate the electoral process, and voters have the information they need to effectively participate in the process.

When advocating based on observation findings, observers are likely to find common ground to improve the governing frameworks around generative content. For instance, journalists, civic tech groups, women's groups, and even EMBs have an interest in building resilience to genAI-related threats. For instance, journalists want to make sure they aren't reporting on fake events and announcements, especially around elections. Meanwhile, EMBs have an interest in ensuring deepfakes don't undermine confidence in election administration and that chatbots provide accurate voter information. While legislation is likely the main target in election reform and recommendations, observers can also consider the utility of other mechanisms that can support political or behavioral change, for instance campaign codes of conduct, changes to technology company policies, and partnerships with tech companies, investigative journalists, etc.

CHAPTER 6: LOOKING FORWARD

This guide has covered several challenges genAI poses to electoral integrity, and ways observers can help build transparency and accountability around it. As mentioned, observers will need to be prepared to reexamine the genAI context of their elections as the technology - and its uses - evolve, as well as in the face of the changing commercial, policy and regulatory framework. This also involves being forward thinking and anticipating innovations, while considering how broader technological trends may impact elections. This may include:

GenAI and the technological divide: There are systematic and structural barriers to the access and use of technology by certain populations. At a macrolevel, the Global Majority has often been left behind in technological developments, with tech advancements piloted in and then driven by large Western markets. Within the electoral context within a given country, there are often technological divisions between youth and older generations, between men and women, between rural and urban citizens, between and among marginalized groups, and other local socioeconomic and sociocultural norms. As genAI becomes more popular, observers should be aware of how the technological divide of its availability and use may impact electoral integrity. For instance, does the employment of genAI exacerbate this divide in any way, creating further gaps between voting populations or candidates? Are there additional challenges for certain groups to access generative AI beyond tech access? Does this disadvantage voters in any way? Could this even disadvantage certain political parties or candidates? Consider the impact on voter information, media literacy, fact checking, and political outreach.

Foreign Information Manipulation and Interference (FIMI): Malign actors and hostile states are already exploring the vulnerabilities of synthetic content and LLMs for influence operations. This includes creating or amplifying deepfakes that undermine confidence in democratic processes, and flooding LLM training data so that chatbots surface foreign state media and other false narratives and foreign propaganda (sometimes called '**LLM grooming**').[20] This is increasingly worrisome as many users are defaulting to chatbots as a replacement for more traditional search engines which may otherwise downgrade such content. Elections and election-related information are heightened targets for such interference, and observers should pay special attention to FIMI throughout their monitoring efforts.

Ongoing digital data protection and privacy issues: The companies developing LLMs are the same tech giants that already operate under a business model based on collecting, analyzing, and selling people's data (so-called '**surveillance capitalism**'). LLM development fundamentally relies on analyzing large amounts of data, often derived from indiscriminately scraping terabytes of online content from news sites, blogs, social media and anywhere they can reach, often without proper rights to use it. For example, Google acknowledges training AI models on YouTube content, while Meta integrates AI into services using billions of images and videos from Instagram and Facebook to train its models. Companies are generally secretive about training data sources, quietly changing privacy policies to expand what data can be used for AI training. Many people producing content online don't know about or have not agreed to it being used for LLM training - including older content (2022 and earlier) which has already been scraped and captured prior to any new disclosure interventions. The

20 **https://www.isdglobal.org/digital_dispatches/talking-points-when-chatbots-surface-russian-state-media/** and **https://www.stopfake.org/en/large-language-models-the-new-battlefield-of-russian-information-warfare/**

challenge is balancing the demand for data to train powerful LLMs with indiscriminate mass scraping of personal information that may infringe on privacy rights. For more information about monitoring personal data protection in elections, see NDI's accompanying guide "**Digital Democracy or Data Exploitation: A Guide for Nonpartisan Citizen Election Observers**."

Political organizing: This guide looked in-depth at the use of genAI in online synthetic content, including deepfakes, propaganda etc to influence voters. However election campaigns are also using advanced AI for more nuanced political organizing, that is not necessarily focused on synthetic content. This may include avatars to spread voter outreach messages, simulations to help voters or organizers, and other satirical or creative uses that can promote political engagement. Like EMBs, political parties and civic organizations may more thoughtfully utilize generative AI to improve internal efficiencies and processes.

Expanding and supporting positive uses of genAI: Similar to political organizing, there are multiple ways that democratic activists can harness the power of genAI to amplify their messages, mobilize their supporters, and improve the information environment. Observer groups should also consider ways of not just improving chatbots but how well-structured and ethically responsible large-language models can be utilized to improve voter information. For instance, models that use genuinely credible information sources as part of their training data can be used for fact-checking, and can be a valuable source of information particularly in countries with low-credibility media or significant state control of media. Groups could also consider how to take advantage of improved translation/interpretation that exists now with the help of genAI. Some civic tech developers are also incorporating genAI into social media monitoring tools to help automate tagging and analysis, and can provide early detection of threats in the information environment.


NDI
NATIONAL
DEMOCRATIC
INSTITUTE


Sida