ARTIFICIAL INTELLIGENCE AND INFORMATION INTEGRITY: LATIN AMERICAN EXPERIENCES

Policy Paper No. 34, 2025

ARTIFICIAL INTELLIGENCE AND INFORMATION INTEGRITY: LATIN AMERICAN EXPERIENCES

Policy Paper No. 34, 2025

Sebastian Becker Castellaro, Mariana Carvalho, Alberto Fernandez Gibaja, Amaro Grassi, Cecilia Hammar, Juliane Muller, Laura Pereira, Victor Piaia, Marco Aurélio Ruediger

Acknowledgements

Abbreviations

AI   Artificial intelligence

DaaS   Disinformation-as-a-service

LGBTQIA+ Lesbian, gay, bisexual, transgender, queer, intersex and asexual

LLMs   Large language models

Contents

EXECUTIVE SUMMARY

Artificial intelligence (AI) could reshape Latin America's electoral landscape by further exacerbating existing problems with how information is created, curated and disseminated. Advances in natural language processing, transformer-based architectures and widespread access to large data sets have led to powerful AI applications. While AI holds promise for fostering civic engagement, driving economic development and potentially improving democratic processes, it also introduces profound risks to information integrity, especially for elections.

This policy paper examines the growing role of AI in Latin American elections, placing a special focus on its implications for elections and marginalized groups. It analyses AI uses across four layers—content creation, curation and moderation as well as political advertising—and documents its uneven but expanding presence in the 2024 elections in Brazil and Mexico. Case studies of the elections in both countries reveal how generative AI was used both to mislead and to engage, including via deepfakes, manipulated audio content and AI-powered chatbots. Particularly concerning is the use of AI to amplify political gender-based violence, with female and LGBTQIA+ candidates disproportionately targeted through non-consensual and demeaning content.

The paper also evaluates regional and global AI policy frameworks. It finds that, while many of these frameworks recognize the need for human oversight and transparency, most neglect key areas—AI-driven electoral manipulation, algorithmic amplification of harmful content, gendered harms and the monetization of disinformation. Latin American AI strategies are often vague or silent particularly when it comes to the electoral consequences of AI and its impact on marginalized communities.

To mitigate the risks posed by the use of AI in elections, the paper recommends that platforms increase transparency by disclosing the data sets used to train algorithms, explaining why certain content is shown to specific users, and by incentivizing democratic values, such as media plurality

and diverse opinions, while curbing exploitative data practices. Clarifying editorial standards for AI-generated material and allowing independent audits of platform data can help prevent further unintended or damaging algorithmic amplification. Renewing public trust in democratic institutions and strengthening the rule of law are essential to shield electoral processes from any harmful consequences of AI for information integrity and guarantee that all voices—especially those of marginalized communities—are heard and represented.

AI's current impact on elections in Latin America remains limited, but its disruptive potential is real and growing. Strengthening democratic institutions, improving regulatory frameworks and centring equity in digital governance are essential to protecting electoral integrity in the age of AI.

INTRODUCTION: THE CHALLENGE OF ARTIFICIAL INTELLIGENCE IN LATIN AMERICA

The integrity of the information environment has been crucial to elections in Latin America—and globally—in recent years. Political actors have sought, with varying degrees of success, to distort the information ecosystem during elections to gain or retain power. These actors have weaponized societal divisions, exploiting the new channels and affordances created by digital platforms. They have flooded the digital public sphere with information and narratives, often using inauthentic or illicit techniques, to advance their interests, eroding trust in the electoral process along the way. In doing so, they have diminished the ability of elections to accurately capture citizens' concerns and opinions and to transform them into policies—a long-standing complaint across Latin America—which is fuelling growing mistrust and disenfranchisement with democracy. From Argentina to Mexico and from Brazil to El Salvador, the contamination of the information environment represents a defining challenge for democracy across the continent.

The issue goes beyond the spread of false or inaccurate information: it involves the manipulation of the information ecosystem to amplify certain political narratives, steer the political agenda and debate, and normalize or legitimize ideas and policies that were, until recently, on the fringes of political discourse. This manipulation is creating bespoke realities for certain groups of citizens—realities that are based not on facts or a peaceful exchange of ideas but on manipulated information and narratives that benefit certain political camps. In Latin America, the subversion of the information ecosystem has relied upon a lack of trust in the media—partly a consequence of the concentration of media ownership—and the increasing hostility that journalists face (Mont'Alverne et al. 2021). These phenomena are fracturing societies by undermining a fundamental requirement for democratic functioning— the existence of a shared understanding of reality based on trustworthy information.

Elections cannot fulfil their role without a shared reality based on facts. Political candidates and campaigns have strategically exploited this erosion of

Political actors have sought, with varying degrees of success, to distort the information ecosystem during elections to gain or retain power.

a shared fact-based reality to question the legitimacy of electoral results and the trustworthiness of electoral management bodies. Across the continent, many political actors have sought to create an alternative reality for their followers, where they represent the only legitimate political force, while any alternative is labelled undemocratic—ironically, often through attempts to delegitimize the electoral process itself. By undermining electoral integrity, these actors are challenging the very legitimacy of government and further fracturing societies. In addition, the political debate on how to address the looming challenges facing the region, from climate change to inequality or corruption, remains muddled and unable to advance.

The consequences have been felt across the region. In Brazil, years of coordinated attacks on the electoral system culminated in a failed coup d'état following the 2022 general election in Brazil (Souza 2024). In Peru, baseless questions about results of the 2021 general elections spurred days of protests and deepened the crisis of trust in the system. Throughout the region, electoral results are frequently challenged, following a global trend (International IDEA 2024). Authoritarian leaders have risen to or cemented power in countries such as El Salvador, Nicaragua and Venezuela.

Although still in its early days, AI has the potential to become a powerful tool for those aiming to distort the information landscape.

In this bleak context, the rapid rise of AI has introduced new, complex threats to the already eroded integrity of information during elections. Although still in its early days, AI—particularly generative AI—has the potential to become a powerful tool for those aiming to distort the information landscape. As seen in the case studies of Brazil and Mexico below, as generative AI improves, it is becoming increasingly capable of producing misleading text, images, audio and video, further undermining a shared reality among citizens. AI can not only convince citizens of an alternative reality but also deliver it directly to their phone screens. Alarmingly, there is still no reliable method to detect AI-generated content with the speed and accuracy needed to mitigate its most harmful effects.

During times of heightened tension, like elections, the danger of this scenario is evident. As we have seen in the region, social tensions during the decisive days of elections can potentially be inflamed by manipulative information, and AI could potentially escalate this tension further. Disputed results or political violence could be fuelled by AI-generated content. An elected government could be portrayed as illegitimate, or, in the worst case, the entire democratic system could be destabilized. The attempted coup in Brazil is a stark reminder of the vulnerability of the system when information is weaponized to undermine electoral results.

Yet, in its current state and based on the experiences from 2024, AI seems to be only marginally enhancing the capacity of political actors to distort the information environment around elections. Its use appears to be focused mostly on increasing confusion and reinforcing existing narratives in an effort to erode trust in the system. Nonetheless, it is not the driving force behind the creation of alternative realities. However, its potential impact is significant enough to warrant analysis, and it is imperative to establish safeguards within the system to prevent its most damaging consequences.

As exemplified by the case studies of Brazil and Mexico, the 2024 elections in the region and globally have shown only isolated, limited uses of the technology, while substantial potential remains for AI to have a broader impact across the electoral cycle.

Two main reasons explain this limited impact so far. First, influencing and altering citizens' capacity to interpret reality is a long-term effort, and widely accessible generative AI has been available only for a short time. Second, disrupting the information environment is not only a matter of content. In fact, content itself might be the least important factor. More important are the capacity to spread disinformation, mechanisms for amplification, the willingness of political actors to construct an alternative reality that is not based on facts and, perhaps most importantly, the nurturing of communities who are willing to embrace this alternative reality. As will be further detailed, the wider distribution of Grok (through Elon Musk's social media platform X) in Brazil has already started to populate Brazilian digital politics with artificial images.

We must also accept that it is very unlikely that we will be able to clearly isolate the effect of AI on information integrity. The way information is created, shared, spread and consumed is a complex process. AI might influence many different phases of this process, altering them slightly and thus altering the final outcome of the elections. Yet, in most cases, it would be nearly impossible to establish clear causal links between the use of AI and the resulting changes.

Notwithstanding this, AI will only improve over time, and so should our defence of democracy and elections. Even if the impact of AI appears to be marginal so far, it is only a matter of time before it becomes more significant. Building a shared, fact-based reality must be central to any strategy for safeguarding elections. Democratic forces must understand that certain principles are non-negotiable—accepting electoral results, ensuring the peaceful transfer of power, respecting and protecting democratic institutions, and upholding the rule of law. Strengthening these foundations is how democracy is defended.

This policy paper will delve into the potential impact AI might have on information integrity during elections by focusing on three different aspects. First, it will examine the way AI affects the creation and distribution of information. Chapter 1 will analyse how generative AI might influence information integrity, exploring how AI is already determining the content we have access to through moderation, curation and distribution, and the potential consequences for marginalized groups and communities. Chapter 2 delves into the cases of Mexico and Brazil with a detailed analysis of the uses of AI in the elections in these countries in 2024. Chapter 3 will explore the different global and regional legal frameworks, evaluating to what extent they consider and include elections and democracy while identifying the main gaps. The final chapter will present a set of policy recommendations to address the challenges identified.

It is unlikely that we will be able to clearly isolate the effect of AI on information integrity. The way information is created, shared, spread and consumed is a complex process.

Chapter 1
AI AND INFORMATION INTEGRITY IN ELECTIONS

The impact of AI on information integrity has been widely discussed in recent months, especially as a number of important elections coincided with a seemingly exponential increase in the capacity of generative AI tools in 2024. Yet, most fears of a deluge of untrustworthy information flooding the information environment and compromising the quality of elections seem to have been largely unfounded, with minimal observable impact on electoral outcomes.

The apparent lack of impact could be explained by a combination of factors, such as the novelty of AI, the fact that it is still under development, or the coordinated efforts of civil society and some tech companies to prevent its misuse. A good example of such a coordinated effort is the Tech Accord to Combat Deceptive Use of AI in 2024 Elections, an initiative of some of the main US-based tech companies to limit the impact of AI in elections.

The capacity of generative AI to create manipulative content was expected to have a major influence on elections, in part by flooding the information ecosystem with manipulative content that would drown out reliable information.

However, the key aspect of manipulative content has never been one of supply but rather of demand (Simon, Altay and Mercier 2023). There are two reasons why having the capacity to create a massive amount of manipulative content does not directly translate into increasing the capacity to manipulate voters. First, influence is very hard to achieve, and research points to a minimal capacity of online information to change voter behaviour (Coppock, Hill and Vavreck 2020). Online content created with the aim of manipulating collective sensemaking can, in the long term, alter people's priorities, but its influence depends less on the quantity of content than on its reach (Rid 2020). Second, demand for manipulative content is highly correlated with low trust in institutions and strong partisanship (Simon, Altay and Mercier 2023).

This implies that manipulative content is used to confirm existing biases and partisan views rather than to shift voters' perceptions.

Although AI's current impact on elections may appear limited, this perception requires nuance and a long-term view. First, AI is already a fundamental part of the information ecosystem, and, as such, it influences how information is distributed and amplified. AI tools are widely used to curate and rank content, as well as to moderate it on all social media platforms. Our feeds are AI-driven, so the information we have access to is determined to a large degree by AI. Second, the exponential improvement in the technology in recent years indicates that our capacity to detect it, which is already low, will continue to decrease. With better, faster and more accurate generative AI applications, the risks to information integrity and the quality of elections will increase. Third, AI is already having an impact on traditionally marginalized communities, and especially on female politicians. This is true at the global level and also in Latin America. Fourth, the empirical analysis of early generative AI usage in elections around the world shows a diversity of applications of the technology beyond direct political manipulation, such as the production of political content that appeals to humour and emotions.

The following sections will delve deeper into these four elements, presenting a framework for understanding the different ways AI might potentially affect information integrity during elections at the global level and in Latin America in the near future.

Although AI's current impact on elections may appear limited, this perception requires nuance and a long-term view.

1.1. **AI AND THE CREATION AND DISTRIBUTION OF INFORMATION IN LATIN AMERICA**

AI plays a fundamental role in the distribution of electoral content online. Although generative AI tools such as ChatGPT, Gemini, Microsoft Copilot and others are popular and well known, tech companies—notably social media platforms—have been using AI for a long time. AI is an important tool for distributing and disseminating content on social media platforms through automated curation and moderation, including in Latin America. Political parties, state actors, politicians and civil society organizations are using social media platforms—and thus AI systems—to create and disseminate election-related political content online.

The processes of designing, developing and deploying content on social media are not neutral. On the contrary, they are driven by political and economic incentives that are shaping the online public sphere, the information environment and, ultimately, democratic processes. On top of that, the institutional legal framework and the geopolitical and the sociopolitical context also shape the electoral information environment. Consequently, the policy approach to ensuring electoral information integrity should be comprehensive, factoring in political, cultural, sociological and geopolitical considerations rather than focusing solely on either media ecosystems or social media

platforms. This approach could help spark a debate to enhance democratic values in online electoral informational contexts.

Considering this, it is important to highlight the impact of AI systems on information integrity within the context of electoral processes in Latin America. Therefore, the following section will explain how AI systems are impacting information integrity at four levels: (a) content creation; (b) content curation; (c) content moderation; and (d) online political advertisement.

1.1.1. Content creation

Generative AI tools can produce sophisticated content on a very large scale; this content is often indistinguishable from human-generated content and has an unprecedented potential to deceive (Juneja 2024). Notably, AI can be used to produce content tailored to shape the views and opinions of specific subsets of the population (Juneja 2024). Content such as texts, visuals or audio have already been used in electoral campaigns in Latin America—in Argentina and Brazil, for example.

The ways in which generative AI, especially large language models (LLMs), can increase the reach, impact and influence of information operations are plenty. It is important to note, however, that we have not yet witnessed the full potential negative effect of the technology on information integrity during elections. Generative AI and LLMs in particular are likely to have mostly a multiplying and amplifying effect. This effect has four aspects: (a) lowering entry barriers; (b) increasing the quantity of deceptive content exponentially; (c) making content more adaptable and targeted; and (d) making information operations more difficult to detect (Rid 2020; Goldstein et al. 2023; Simon, Altay and Mercier 2023; Juneja 2024).

Generative AI lowers the cost and effort required for conducting information operations, reducing barriers to entry and making such operations more accessible.

Generative AI significantly lowers the cost and effort required for conducting information operations, reducing barriers to entry and making such operations more accessible. AI-powered tools can greatly enhance the efficiency of political campaigns. Moreover, they can facilitate the expansion of disinformation-as-a-service (DaaS) providers, enabling a larger number of actors to enter the DaaS market (Soliman and Rinta-Kahila 2024). However, having the capability to employ such tactics does not necessarily translate into the intent to do so. Leveraging manipulative material to sway public opinion through deceit and falsehoods is a strategy adopted by only certain parties and candidates (Törnberg and Chueri 2025). For many political groups, employing deceptive practices is not an appealing approach. Most parties and candidates who adopt these tactics, either directly or through DaaS operators, exploit the growing distrust in the political system and official information. Their goal is to leverage this declining confidence in democratic institutions for political advantage (Törnberg and Chueri 2025).

Generative AI enables political parties and candidates willing to employ manipulative tactics to produce deceptive material on a much larger scale than ever before. For example, LLMs can facilitate the creation of false content for fabricated online media outlets, a strategy commonly used in information

operations, including by foreign actors (Rid 2024). An increase in deceptive content can detach segments of the population from reality in respect of critical issues for the political future of Latin America, such as climate change, political corruption and migration (DiResta 2024). Additionally, it can make it increasingly challenging for individuals to access fact-based information needed to understand the world accurately.

However, the digital information landscape is already inundated with deceptive content, which suggests that only an increase in demand, rather than supply, will turn the enhanced capacity to generate such material into a genuine threat to democracy (Simon, Altay and Mercier 2023). Moreover, overwhelming the information space does not necessarily equate to influence. Disinformation primarily impacts individuals already inclined to believe it, as it reinforces their pre-existing perceptions of reality (DiResta 2024). Factors such as distrust in democratic institutions and strong partisan identification—especially with radical right-wing parties—make people more susceptible to embracing false narratives (Grinberg et al. 2019; Törnberg and Chueri 2025).

A third potential impact of generative AI on information integrity is its ability to enable the extreme personalization of deceptive content. Such content could be used either to harm individuals, particularly through the creation of non-consensual images, or to manipulate their perceptions. Non-consensual pornographic content represents a significant threat posed by generative AI (for an in-depth discussion, see 1.2: AI and marginalized communities). This type of content disproportionately affects female political figures, serving as a form of tech-facilitated gender-based violence that further limits the political space available to women.

Generative AI also enhances the ability to adapt and target content to specific social groups, in theory increasing the persuasive power of deceptive content. Methods for adapting and targeting content include using culturally appropriate language, employing specific references, framing issues in a way that resonates with the target group or crafting content that manipulates reality to appeal to specific demographics. While many foreign-led information operations can be detected due to their unnatural use of language or lack of cultural awareness, generative AI can avoid these pitfalls (Goldstein et al. 2023).

It is even technically conceivable that information operations could become far more granular in the future, targeting individuals with personalized deceptive content. Such operations could leverage individualized data to create tailored manipulative materials designed to influence or harm specific people. As of now, however, generative AI models do not have access to personalized data that can make this type of content possible.

In the case of Latin America, the technology in its current form might struggle to target individuals at such a granular level. AI models are often trained on content originally created in English and then machine-translated into Spanish. Consequently, on the one hand, studies have suggested that these models

are perpetuating racial prejudices, which might impact negatively on the content's persuasion power (Hofmann et al. 2024). Since these models are designed in English and work better in English-language contexts, they create an experience gap between native English speakers and users elsewhere, including in Latin America (Nicholas and Bhatia 2023), where the lack of local context and data in other languages poses a systemic problem in how these technologies are being used (Oversight Board 2024).

Finally, generative AI could enhance the ability of information operations, particularly those conducted by foreign actors, to evade detection (Goldstein et al. 2023). Influence operations conducted by foreign powers often fail because of a lack of precision in their messaging and use of language. In addition, the capacity of social media companies and some state agencies to identify the coordinated inauthentic behaviour typical of such operations often leads to their detection before they can have any real effect (Rid 2024). Generative AI can make it more difficult to detect information operations by creating networks of fake online personas that are distributed in a way that conceals their origin. These fake personas can be used to help spread narratives and shape alternative realities, as they can better mimic local dialects and language usage or even fake accents. They can also overwhelm fact-checking organizations and make their work much more difficult (Kahn 2024).

> **Generative AI can make it more difficult to detect information operations by creating networks of fake online personas that are distributed in a way that conceals their origin.**

1.1.2. Content curation

AI is being used not only to create content on social media platforms but also to rank, promote or demote content in newsfeeds, usually based not only on users' preferences and profiles but also on the use of content designed to keep users as engaged as possible, without considering the quality or accuracy of it (ARTICLE 19 2021; OSCE 2021; Narayanan 2023). The main purpose of AI-powered tools—often referred to as ranking algorithms—is to help people find relevant content based on the commercial and, in some cases, political considerations of the platforms they are using. Given the volume, scale and speed at which content spreads on social media, AI systems play a crucial role in prioritizing and demoting content in users' feeds based on their profiles and the platforms' business models.

Such developments may affect the accuracy, diversity and public interest of the information ecosystem in online spaces, which is especially important during elections (OSCE 2021). The United Nations (2018) has also said that users may have little or no exposure to certain types of critical social or political content on their platforms. This lack of exposure might undermine not only citizens' freedom to access diverse information but also their right to form opinions independently, free of violence, inducement or manipulative interference of any kind (OHCHR 1996). Such limitations on access to diverse content could entail a breach of other fundamental rights, including the right to participate in public affairs.

Consequently, AI-powered tools that recommend content are not neutral. Algorithmic curation is shaped by the values and goals of the algorithm's creator, and by socio-technical dynamics, industry self-regulation and state

oversight (OSCE 2021). These algorithms are fed by a huge volume of personal and behavioural data that is combined with third-party data to create sophisticated profiles that increase user engagement and retention by tailoring the user experience.

The personalization of content is also being used for online political advertising and political campaigns. It has been pointed out that the use of content personalization may entail manipulation, polarization and fragmentation of the political debate (Gorton 2016). Additionally, several studies have shown that content curation or ranking algorithms contribute to the amplification of disinformation, hate speech, harmful content and discrimination on social media platforms (Panoptykon Fundacja 2023). The 'content with highest predicted engagement scores low in terms of quality and trustworthiness' (Panoptykon Fundacja 2023: 3). The Mozilla Foundation (2021) also reported in a study that 71 per cent of regrettable experiences in YouTube raised by volunteers were recommended to viewers by the platform. As people are more likely to click on sensationalist or incendiary content, ranking algorithms privilege harmful content (Amnesty International 2019) instead of accurate information, which undermines electoral integrity in both online and offline contexts.

In Brazil, during the Bolsonaro regime, notably up to the 2022 elections, disinformation campaigns and conspiracy theories were disseminated on Twitter and WhatsApp to discredit electronic voting and the country's democratic process (Souza 2024). The radicalization process that started in 2018 with the first Jair Bolsonaro presidential campaign and culminated in an attempted coup d'état against the elected president of 2022, Luiz Inácio Lula da Silva, shows how impactful disinformation can be.

On another note, AI-powered content curation tools are trained on data that fail to reflect broader contextual, historical and economic realities. Hence, potential biases might appear in how content is distributed, distorting, downplaying or amplifying certain messages in order to engage citizens on social media platforms.

Due to the volume of data processed by AI-powered tools, ranking algorithms should be scrutinized on two levels. First, the data that is collected to train them should be clearly disclosed and categorized. Second, ranking algorithms should explain the reasoning behind their output—why certain content is shown to a particular user—and what data points led to the decision to show certain content.

Designing incentives that promote democratic values, such as media plurality and diversity of opinion, and restricting the data used to train algorithms could mitigate the detrimental effects of content curation. Providing access to the data and enabling third-party audits of algorithms could help reduce the unexpected and problematic outcomes of these AI-powered tools.

Designing incentives that promote democratic values and restricting the data used to train algorithms could mitigate the detrimental effects of content curation.

1.1.3. **Content moderation**

Content moderation often involves large-scale outsourcing combined with automated machine-learning detection. AI systems assist social media companies in identifying, matching, predicting or classifying user-generated content based on its features to determine what should remain online and what should be removed. This process typically occurs after users publish content on a social media platform. In other words, content moderation tools act as gatekeepers, guided by the platform's policies and business model, as well as the national legal framework in which the platform operates.

Content moderation tools may perform the following functions:

*   detecting content that violates the law (e.g. hate speech or the dissemination of intimate images without consent) or the platform's internal policies;

*   identifying violations by assessing whether detected content breaches internal policies; and

*   enforcing policies by flagging or deleting online content created by users based on the platform's guidelines (OSCE 2021).

Platforms often employ practices like de-prioritization and de-ranking to minimize the visibility of harmful or illegal content. Account suspensions and removals are also integral to these content moderation processes.

AI systems facilitate these processes by reviewing, making decisions and restricting information or disinformation at a massive scale, often without human intervention. The automation of these processes can enhance consistency and predictability. Thanks to natural language processing, content moderation techniques can identify positive and negative emotions, classifying them as either 'offensive' or 'not offensive'.

However, the use of AI in content moderation also carries risks. Mislabelling can result both in false positives, where legitimate content is wrongly labelled as illegal or harmful, and in false negatives, where illegal or harmful content goes undetected.

The aforementioned challenges are compounded by biases in the training data used for AI systems, which often rely on content written primarily in English. This reliance on English-language content creates difficulties in detecting disinformation or hate speech in non-English contexts. For instance, in the case of Quechua—an Indigenous language spoken in Latin America—it has been reported that the 'biggest roadblock in addressing online harms is the "lack of high-quality digital-data"' (Shahid 2024). Furthermore, the 'Facebook Papers' revealed that 87 per cent of Facebook's global spending on classifying misinformation was directed at English-language content, even though only 9 per cent of its users are native English speakers (Elswah 2024).

1.1.4. **Online political advertising**

Online platforms rely on online advertising, including online political advertising, to monetize their services. Advertising is at the core of social media platforms' business models because it enables them to monetize user attention and retention. In this context, AI systems are used to deploy and distribute political advertising on social media platforms.

On the one hand, microtargeting techniques (ad targeting) enable political parties and campaigners to identify and target specific population groups with tailored political ads with an incredible level of granularity. It is possible, for example, to present a particular political ad to young women living in Santiago de Chile, aged 18 to 24, who like K-pop music, lean towards the left politically, and are interested in nature and sports. A different ad with a different message altogether could be shown to a completely different demographic. These targeting techniques rely on different sources of data, including personal and behavioural data, data generated by social media platforms themselves and third-party data. Altogether, this data is used to create highly personalized profiles that determine which content gets amplified on a particular user's timeline and which ads are shown to the user (Ali et al. 2019).

On the other hand, social media platforms use amplification techniques to deliver political advertisements to the most relevant recipients (European Partnership for Democracy n.d.). This process, known as ad delivery, involves the platform displaying ads to users based on machine-learning algorithms. These algorithms process behavioural data such as likes, shares, connections, clicks and video views to build comprehensive online user profiles. Using this data, AI systems can predict user behaviours, revealing sensitive characteristics such as political inclinations, religious beliefs, health conditions and more. However, because some users are deemed more valuable than others, this process can result in discriminatory consequences in the distribution of political advertisements (Ali et al. 2019).

Moreover, these techniques help fine-tune ranking algorithms to keep users engaged with content, which may reinforce biases and fuel radical opinions. By disseminating disinformation and misinformation, they contribute to the creation of echo chambers, which can fragment the online public sphere. This fragmentation facilitates manipulation, polarization and the radicalization of citizens.

1.2. **AI AND MARGINALIZED COMMUNITIES**

As with most challenges, marginalized groups within society feel the impacts most acutely. In Latin America, these groups include working-class communities, women, LGBTQIA+ individuals, people with disabilities and Indigenous groups. Marginalized groups in the region already contend with heightened risks, including in digital spaces, which AI systems can exacerbate

when they are designed, deployed or regulated without sufficient consideration of human rights and democratic values.

One of the clearest examples of the potentially negative consequences of AI is its ability to perpetuate and amplify gender-based discrimination and harassment in online environments. Generative AI, for instance, has been weaponized around the world to create deepfake content—including non-consensual explicit material—that disproportionately targets women and sexual minorities (Kira 2024; Burgess 2021). Numerous websites offer deepfake services, posing a tangible threat to women, who are often targeted by AI-generated pornographic content. Applications allow users to upload photos and generate realistic nude images of individuals without their consent. In Mexico, for example, a university student altered 166,000 photos of classmates into pornographic material and sold them on Telegram (Rodríguez 2023).

Such practices are not limited to personal exploitation; they are also misused to undermine female political candidates. In Brazil's recent 2024 municipal elections, for instance, deepfake technology was employed to target female political candidates, depicting women in compromising or explicit scenarios, aiming to damage their reputations and electoral prospects (Farrugia 2024). While male politicians are also targeted by deepfakes—often for their political statements or actions—these attacks usually allow them to preserve a sense of personal integrity. In contrast, deepfake attacks against female politicians tend to revolve around fabricated personal misconduct, which fuels objectification, reinforces harmful stereotypes and undermines women's credibility (Cerdán Martínez and Padilla Castillo 2019). Such weaponization of deepfake technology not only compromises women's dignity but can also deter them from running for office or fully engaging in the electoral process (Souza and Varon 2019; Fundación Multitudes 2023; Ríos Tobar 2024), with repercussions for their mental well-being, personal safety and overall political representation.

Several Latin American countries have begun adopting gender-focused approaches to cybersecurity policies, addressing issues like cyberbullying and non-consensual sharing of intimate images (González-Véliz and Cuzcano-Chavez 2024). Brazil's municipal elections in October 2024 were held under strict new electoral regulations that banned the use of unlabelled AI-generated content by campaigns in an effort to curb AI-fuelled disinformation. Although these measures likely reduced some risks, regulation alone is not enough to fully prevent the spread of harmful, violent deepfakes. Similarly, Mexico's 2020 Olimpia Law criminalizes the non-consensual dissemination of sexual content, although its scope remains too narrow to fully address offences specific to deepfake technology (Piña 2023).

Furthermore, social media platforms that rely on AI for content moderation often fall short in protecting marginalized groups. For instance, research shows that AI-driven content moderation on social media platforms frequently struggles to effectively identify and remove hate speech targeting marginalized groups, often due to the nuanced and context-specific nature of such content

> One of the clearest examples of the potentially negative consequences of AI is its ability to perpetuate and amplify gender-based discrimination and harassment in online environments.

(Udupa, Maronikolakis and Wisiorek 2023). Conversely, AI systems sometimes flag legitimate activism or discussions on gender and LGBTQIA+ issues as violations of platform policies, silencing critical voices and limiting their reach.

Indigenous communities are often particularly vulnerable to the manipulation of electoral information environments. They face an array of challenges, including disinformation campaigns, foreign information manipulation and interference, systemic exclusion from digital spaces and the amplification of hate speech through emerging technologies. Such campaigns frequently exploit cultural stereotypes or misrepresent Indigenous traditions and political demands, undermining the credibility of Indigenous leaders, activists and movements (Aaberg, et al. 2024). See also an example from the Australian election: Kaye (2023).

During electoral periods in Bolivia, for example, disinformation, hate speech and racially charged narratives have been deliberately circulated to delegitimize Indigenous participation in politics (Wood 2019). Such tactics not only distort public discourse but also incite hatred, racism and prejudice against Indigenous peoples, effectively eroding their significance and contributions both domestically and globally. Reports have documented the use of misleading information to foster intolerance, thereby deepening historical grievances and systemic injustices against Indigenous communities (The Carter Center 2021; Funk, Shahbaz and Vesteinsson 2023).

AI technologies have exacerbated the vulnerabilities of Indigenous communities. Automated content generation tools, large-scale bot networks and sophisticated algorithmic targeting can greatly increase the spread, personalization and persistence of disinformation and hate speech (Funk, Shahbaz and Vesteinsson 2023). As a result, Indigenous leaders and movements may find themselves at an even greater disadvantage, struggling to counter narratives that are widely and persistently disseminated via digital platforms.

Algorithmic biases intensify existing forms of exclusion, particularly when AI systems are trained on data sets that do not adequately represent Indigenous communities. As a result, these technologies reinforce harmful stereotypes and fail to meet the distinct needs of Indigenous populations. Methodological shortcomings, such as overgeneralization or the failure to adapt models to local contexts, can distort reality. In doing so, biased algorithms not only misrepresent the truth but also amplify existing inequalities, thereby widening the digital divide that Indigenous communities face (Naciones Unidas en México 2023).

Before considering the use of AI in public processes such as elections, the fundamental issue of access must be addressed. Across Latin America and the Caribbean, Indigenous communities face a pronounced ethnic digital divide, as infrastructure and connectivity often cater primarily to urban areas. This imbalance leaves many Indigenous peoples in rural regions with limited opportunities to access digital tools. Compounding this challenge, widespread

gaps in digital literacy further curtail their ability to fully participate in the digital ecosystem. According to the UN Economic Commission for Latin America and the Caribbean, while 40 per cent of the region's population have basic computer skills, fewer than 30 per cent can use spreadsheets effectively, under 25 per cent can install new devices or software, and only about 7 per cent have programming experience (Naciones Unidas en México 2023).

Building ethical and inclusive AI demands that we close connectivity and skill gaps, correct algorithmic biases, uphold Indigenous data sovereignty and integrate diverse worldviews.

These systemic inequalities perpetuate the exclusion of Indigenous perspectives from AI development. Limited connectivity, insufficient policies and inadequate technical competencies reinforce a cycle of marginalization that diminishes Indigenous knowledge and experiences. Building ethical and inclusive AI demands that we close connectivity and skill gaps, correct algorithmic biases, uphold Indigenous data sovereignty and integrate diverse worldviews. A crucial first step is to recognize Indigenous peoples not only as participants but also as creators and innovators within technological frameworks.

These are just some of the challenges that illustrate the intricate overlap between information integrity, AI and electoral integrity, particularly their disproportionate impact on marginalized groups within society. This policy paper seeks to explore the role of AI in these electoral processes and to develop targeted recommendations for social media companies, policymakers and authorities to effectively regulate AI systems. Such regulations should address the creation, curation and moderation of online content while prioritizing human rights and democratic principles.

Chapter 2
IMPACT OF AI ON ELECTIONS IN MEXICO AND BRAZIL

To understand how AI has impacted electoral processes in Latin America, this chapter will investigate how AI-powered tools were used in the 2024 elections in Mexico and Brazil. For this analysis, we relied on reports produced by FGV Comunicação Rio (2024) on the use of generative AI in elections around the world in 2024, articles published on Brazilian and Mexican news websites, and the continuous monitoring of public debate conducted by DappLab (FGV DAPP 2017), the social media monitoring laboratory of FGV Comunicação Rio.

The elections that took place in Mexico on 2 June 2024 and were among the largest in the country's history, with 92 million eligible voters and a final turnout of over 60 per cent (INE 2024a). The elections were conducted through direct and mandatory voting, using paper ballots, and involved seven political parties and more than 20,000 elected positions, including deputies, senators, the president, state governors, local authorities and a new mayor for Mexico City, the country's capital. Operating under a presidential republican system, the national parliament is bicameral, consisting of a chamber with 500 deputies and another with 128 senators.

The presidential race was mainly contested by three candidates—Claudia Sheinbaum, representing the left-wing governing coalition Sigamos Haciendo Historia, who won with 59.3 per cent of valid votes; Xóchitl Gálvez, from the opposition alliance Fuerza y Corazón, who secured 27.9 per cent of valid votes; and Jorge Álvarez Máynez, from the Citizens' Movement, who obtained 10.4 per cent of valid votes (INE 2024a).

The 2024 elections marked a milestone for women's representation in Mexican politics, as four female governors and the country's first female president were elected (Martínez Holguín 2024)

As the successor to President Andrés Manuel López Obrador, Sheinbaum consistently led in the polls, facing similar criticisms to those directed at López Obrador, particularly regarding the alleged leniency of Mexico's

left-wing government towards drug cartels. In this context, violence and crime dominated public concerns, emerging as central themes in political discussions, particularly regarding the country's high homicide rate, gender-based political violence, cartel disputes and incidents of political violence. Despite these issues, topics related to the economy and health also gained prominence, while immigration policies were discussed more marginally (Stevenson and Verza 2024).

At the same time, disinformation remained a persistent issue, predating the elections. Rumours circulated throughout the electoral period, alleging electoral fraud, a proposed new constitution, the abolition of private property and even claims that Sheinbaum intended to close the Basilica of Our Lady of Guadalupe. Local reports also highlight the spread of false information, including fabricated links between López Obrador, Sheinbaum and drug trafficking, as well as allegations that López Obrador's wife had connections to Nazism.

In this context, the Mexican elections also served as a backdrop for the use of generative AI in political campaigns and information, not only for spreading deepfakes to discredit candidates or disseminate political misinformation but also for financial scams exploiting candidates' images.

Brazil held municipal elections in October 2024, electing mayors, deputy mayors and legislative representatives in 5,569 municipalities. While municipal elections in the country are traditionally less polarized and receive fewer resources than national ones, the 2024 elections took place in a highly charged political atmosphere. Races in several key cities, including São Paulo and Fortaleza, were particularly intense, with mayoral candidates closely aligned with national political figures such as President Luiz Inácio Lula da Silva and former President Jair Bolsonaro, who, despite being ineligible to hold public office, remains a central figure in the opposition. These local elections were widely seen as a strategic prelude to the 2026 general elections, shaping alliances and narratives for the upcoming presidential race.

The elections took place amid growing concerns over the role of digital platforms in Brazilian democracy. Given the anticipation surrounding AI's potential impact on political campaigns, Brazil's main electoral authority, the Superior Electoral Court, issued new regulations before the elections to curb misinformation and the misuse of AI-generated content (TSE 2024). These measures included mandatory labelling of synthetic content in political advertisements and stricter oversight of digital campaigning.

Contrary to expectations, AI was not used as extensively as initially feared.

Contrary to expectations, AI was not used as extensively as initially feared. As in Mexico, much of the observed usage of AI during the electoral period was overshadowed by a series of events that, while linked to the use of the Internet in politics, were not directly related to AI. These events and the documented cases of AI use reveal a complex relationship between innovation and long-standing challenges in Brazilian political communication. Even at this early stage, the characteristics of these AI applications are noteworthy.

2.1. **USE OF GENERATIVE AI IN NATIONAL AND LOCAL ELECTIONS**

In Mexico, the use of AI is still considered a peripheral issue compared with more pressing problems, such as political violence and drug trafficking. Although there have not been many cases that have had significant repercussions for digital platforms, the spread of deepfakes and other AI applications are linked to national issues, particularly in the electoral context, where these technologies have been used both constructively and as tools for disinformation and manipulation (Del Pozo and Arroyo 2024).

In the election that resulted in the country's first female president, disinformation reflected long-standing political issues in Mexico. The potentially harmful uses of AI, especially deepfakes, also followed this trend, attracting greater attention within the Mexican electoral landscape. Notable examples include manipulated videos of President López Obrador, allegedly promoting fraudulent investment schemes, such as Pemex stocks or financial applications falsely attributed to Elon Musk. This material was even verified by Reuters after being shared more than 2,000 times on Facebook (Sosa Santiago 2024).

It is interesting to note that, even in the electoral context, much attention has been given to falsifications that targeted the economic and financial sectors rather than being directly related to the elections. In other words, the fraudulent content used images of political figures but did not mention the elections specifically.

Before being elected president, Sheinbaum was also the target of manipulated content linking her to fraudulent investment schemes, while regional deepfakes gained prominence in local disputes (see Figure 2.1). In Tijuana, audio recordings attributed to political figures were used to influence campaigns, while in the Mexican state of San Luis Potosí, Enrique Galindo Ceballos, a mayoral candidate, denounced the use of AI to launch attacks against his candidacy, leading to a formal complaint to the National Electoral Institute (Flores Saviaga and Savage 2024).

At the national level, deepfakes targeted the main presidential candidates. Sheinbaum, for example, was depicted in manipulated videos with communist symbols in the background while speaking in Russian, whereas Gálvez appeared holding a national flag with an inverted emblem (Flores Saviaga and Savage 2024).

The use of AI was also observed in Mexico's local elections, with the Friedrich Naumann Foundation and the German Council on Foreign Relations (Del Pozo and Arroyo 2024) identifying 44 such cases. One prominent example was an audio clip supposedly of a conversation between Montserrat Caballero, mayor of Tijuana, and Maricarmen Flores, then a candidate for the National Action Party, discussing campaign issues (Del Pozo and Arroyo 2024).

Figure 2.1. **Screenshot from a deepfake of Mexican presidential candidate Claudia Sheinbaum promoting a financial platform**

Source: Mexico News Daily, 'Video circulating of Claudia Sheinbaum is apparently a "deepfake"', 26 January 2024, <https://mexiconewsdaily.com/politics/video-circulating-of-claudia-sheinbaum-is-apparently-a-deepfake>, accessed 24 June 2025.

Other uses of AI reveal how political figures can leverage the technology to shield themselves from accusations, claiming that they were victims of fabricated content—even when such claims remained unproven.

Other uses of AI reveal how political figures can leverage the technology to shield themselves from accusations, claiming that they were victims of fabricated content—even when such claims remained unproven. In the Mexican elections, Martí Batres, a candidate for the presidency, claimed that an audio recording circulating on social media in which he discussed the Morena party's internal candidate selection process had been generated by AI. Despite his defence, experts told Mexican newspapers at the time that it was impossible to determine whether the recording was legitimate or AI-generated (Murphy 2024).

Similar cases of audio recordings with disputed authenticity were also observed in the Brazilian elections, with candidates themselves contesting their validity. However, many of these cases were not confirmed or were dismissed by the country's Electoral Court due to a lack of evidence or the technical tools needed to verify the authenticity of the recordings (Murphy 2024).

The most notable case occurred in the second round of the Brazilian elections, when Fortaleza mayoral candidate André Fernandes accused his opponent Evandro Leitão and his team of creating an allegedly fake audio recording in which Fernandes appeared to endorse bribery and vote buying. Fernandes made the accusation on Instagram, called a press conference and submitted the evidence to the Electoral Court. Similar to the Batres case, fact-checking news outlets consulted experts to verify the authenticity of the claim, but they could not determine whether the audio was truly fake.

Another use of generative AI tools observed in the analysed elections was the dissemination of cheapfakes—simple, low-cost manipulations of images and

videos, including editing cuts, speed changes, or recontextualization. Unlike deepfakes, which use advanced AI to create realistic forgeries, cheapfakes rely on basic edits that the general public could make.

One of the most emblematic examples of this technique in the Brazilian elections occurred in April 2024, before the official campaign period for the municipal elections. At the time, Tabata Amaral, a pre-candidate in the São Paulo mayoral race, mocked the mayor at the time, Ricardo Nunes, after he had referred to her as the 'Barbie of politics'. In response, Amaral posted a video from the *Barbie* movie with Nunes's face superimposed on the character of Ken. Nunes's party, the Brazilian Democratic Movement, filed a complaint with the Electoral Court, claiming that the use of AI was prohibited and accusing Amaral of premature campaigning. In response, the congresswoman deleted the video and published a new version, switching from a deepfake-style edit to a more amateur collage aesthetic (see Figure 2.2) (Resk 2024).

Figure 2.2. **Posts by São Paulo mayoral pre-candidate Tabata Amaral on Instagram**

Source: Resk, F., 'Tabata usa deep fake para expor Nunes como Ken de Barbie e refaz post' [Tabata uses deep fake to expose Nunes as Ken from Barbie and remakes post], Metropoles, 5 April 2024, <https://www.metropoles.com/sao-paulo/tabata-deep-fake-nunes-ken-barbie>, accessed 17 June 2025.

Similar occurrences were observed during the elections, in videos shared by both candidates themselves and their supporters. In these cases, candidates and other prominent political figures were placed in unusual and embarrassing situations. These examples suggest a distinctive form of political participation in Brazil, where humour and criticism are popular expressions of political engagement among citizens.

These cases sparked a debate on the challenges of uniformly enforcing policy. Although Resolution No. 23.732, issued by Brazil's Superior Electoral Court in 2024, sets clear rules for the use of AI in electoral propaganda (TSE 2024), a study by the Laboratory for Governance and Regulation of Artificial Intelligence (Laboratório de Governança e Regulação de Inteligência Artificial), affiliated with the Brazilian Institute for Education, Development and Research (Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa), found that local court rulings across the country lacked consistency in classifying deepfakes despite the clarity of the resolution. For instance, the Brazilian Democratic Movement's complaint regarding the Ken video (mentioned above) was ultimately dismissed by the São Paulo Electoral Court, which ruled that it lacked illicit intent and fell within the scope of freedom of speech. However, other courts across the country adopted a stricter approach, prohibiting any use of deepfakes—even when they were labelled or clearly identifiable (Caballero et al. 2024).

2.2. AI AS AN AMPLIFIER OF POLITICAL GENDER-BASED VIOLENCE

Although Mexico elected a female president for the first time, representing progress in women's political representation, the 2024 elections also saw a significant increase in gender-based political violence. Data from the National Electoral Institute indicate that 205 complaints related to this type of violence were registered during the electoral period, with 42 per cent of them involving attacks on social media. The violence increased in frequency and severity as election day approached, escalating from digital forms to physical threats, assaults and even murders (Rangel 2024).

One of the most prominent cases involved the Chihuahua state senator Andrea Chávez, who denounced on social media the distribution of an AI-manipulated image showing her face on another person's semi-nude body. A report by the Friedrich Naumann Foundation and the German Council on Foreign Relations (Del Pozo and Arroyo 2024) also mentions cases in which AI was used to 'beautify' female candidates. While such image manipulation does not necessarily constitute political gender-based violence, it reinforces stereotypes of female objectification in positions of power and also challenges enforcement of Mexico's national laws regarding technology and gender-based digital violence. As mentioned earlier in this paper, while the so-called Olimpia Law was designed primarily to address the non-consensual distribution of intimate images, recent debates about the law have considered its potential application to AI-generated content, such as deepfake nudes (Borchardt 2024).

Beyond exposure, ridicule and the shame associated with such attacks, these cases also cause material damages to the targets, who must dedicate time and effort to reporting the incidents, holding those responsible accountable or having the content removed. During election campaigns, the time and effort spent combating deepfakes can also harm a candidate's electoral chances. While male politicians are also victims of AI-generated attacks and ridicule,

a clear gendered pattern emerges when examining cases involving women—typically aimed at shaming and silencing them by mobilizing stereotypes, particularly related to sexuality.

2.3. TOOLS FOR THE PROMOTION OF CITIZENSHIP

Despite the challenges posed by AIin electoral contexts, the elections in Brazil and Mexico also demonstrated how this technology can be applied to strengthen democracy and citizen participation. A clear example of this was the use of chatbots and other AI-generated resources to foster closer interaction between candidates and voters.

Although not exclusive to the 2024 elections in Latin America, Brazil's municipal elections provided a positive example of how chatbots can be leveraged to enhance political communication and voter engagement. Their deployment was particularly notable among mayoral candidates in major Brazilian capitals, who used AI-powered tools to interact with voters, present proposals and mobilize supporters for campaign events. The use of chatbots in Brazil's 2024 elections was regulated by the Superior Electoral Court, which mandated that candidates could not simulate real people and had to clearly indicate that interactions were AI-driven (Tribunal Superior Eleitoral 2024).

Even within the limits imposed by electoral regulations, these tools have significant potential to add value to a candidate's image. By automating interactions and providing information to voters, these virtual assistants not only expanded campaign reach but also mobilized cultural and symbolic elements, influencing the public perceptions of candidates.

In Mexico, similar cases demonstrated the positive potential of AI. During the pre-campaign, Xóchitl Gálvez used generative AI to create a digital version of herself, called iXóchitl, to communicate with voters in campaign videos on social media. Gálvez also stated that she used AI tools to prepare for election debates by simulating questions from her opponent (Ormerod 2023).

The National Institute for Elections (INE) also deployed AI tools during the Mexican elections to help combat disinformation, launching the chatbot Inés as a tool for users to report potentially misleading content for verification by specialized fact-checking companies (INE 2024b).

AI-powered chatbots were also adopted by government agencies responsible for overseeing Brazil's electoral process. In the state of Bahia, the Electoral Court used the chatbot Maia, which provided support via WhatsApp and Telegram for campaign finance reporting and facilitated citizens' access to election-related information, promoting greater transparency and accessibility. In Pernambuco, another Brazilian state, the virtual assistant Júlia enabled voters to access election-related information via Telegram, including checking their voter registration status, obtaining their voter ID number, finding their

The elections in Brazil and Mexico demonstrated how AI technology can be applied to strengthen democracy and citizen participation.

The use of AI in the 2024 elections in Brazil and Mexico was primarily characterized by experimentation related to content creation, and it probably had little impact on the final results.

polling location, and receiving guidance on electoral security and campaign regulations (Tribunal Regional Eleitoral–BA 2024).

The use of AI in the 2024 elections in Brazil and Mexico was primarily characterized by experimentation related to content creation, and it probably had little impact on the final results. In this regard, it is noteworthy that deception was not the main reason for the use of AI in several cases. Among many imagined risks associated with AI, the possibility of generating content that is impossible to verify stands out as the most alarming. However, the impact of AI on politics and daily life can manifest in ways that go beyond deception, increasing the potential to construct narratives that ridicule and discredit.

In cases involving deception, low-tech content (such as manipulated images) and the predominance of audio formats stand out, with the latter having the greatest potential to produce content that is nearly impossible to verify.

In terms of content production, it is also important to consider how the use of AI is connected with pre-existing dynamics, as seen in videos where political candidates endorse financial scam platforms. In other words, AI enhances the impact and reach of an existing scam, leveraging the visibility and appeal of candidates during the electoral period to attract new victims.

The Latin American cases analysed reveal that, while various candidates were targeted by AI-generated audiovisual content, female candidates were specifically subjected to attacks involving gender-based violence. In this context, the analysis indicates that the most vulnerable groups tend to experience the negative impacts of AI misuse more intensely, exacerbating pre-existing lines of discrimination and prejudice. The absence of specific regulations addressing the production of deepfakes for harassing women and LGBTQIA+ individuals in the legislative proposals of the Latin American countries analysed further exacerbates this issue, given the recurrence of such practices.

Although AI applications have largely been used to undermine the integrity of the electoral information system, they have also been employed as a tool for enhancing citizenship—such as in the case of chatbots created by regional electoral courts in Brazil—and as an instrument that deepens political reflection and knowledge in the electoral context, exemplified by Mexican candidate Xóchitl Gálvez's experiments with a digital avatar and AI-assisted training for televised debates.

Finally, it is important to highlight the experimental and innovative nature of the popularization of AI tools in everyday life. Recently, in Brazil, former President Bolsonaro leveraged public trust in generative AI tools by calling on citizens to use AI to question the conduct of a Supreme Court judge in a case accusing Bolsonaro of plotting a coup d'état (Poder360 2025). This event took place outside the electoral period and is beyond the scope of this study, but it suggests a previously unidentified way that generative AI tools can be utilized, as a supposedly neutral mediator in social conflicts and disputes.

Chapter 3

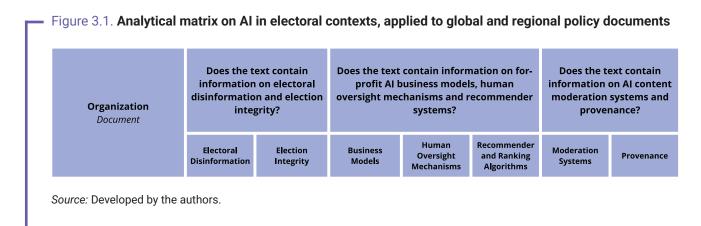# AI POLICIES AND ELECTIONS IN LATIN AMERICA

This chapter provides an overview of the current AI policy landscape both globally and in Latin America, highlighting discussions on the role of AI during elections and its impact on gender and inclusivity in policy frameworks and strategies. The chapter is divided into three subsections—the first explaining the analytical framework, the second covering documents published by international organizations and the third covering governments and regional coalitions in Latin America.

3.1. **ANALYTICAL FRAMEWORK**

The analytical framework was developed to discern the degree to which state actors and international organizations are currently addressing issues relating both to AI and elections and to AI and gender. As these issues are complex and multifaceted, the analytical framework aims to create a gradient with which to assess whether different facets of the issues are being addressed. It does so by dividing the two issues—AI and elections and AI and gender—into respective subsets of aspects relevant to each issue. For example, AI and gender is examined by looking into how actors address technologically facilitated gender-based violence by way of deepfakes, the AI divide, the negative effects of digital technologies on political participation and data set–induced biases. The selected facets are operationalized to corresponding close-ended questions that are directed at international and national policy frameworks and strategies, such as the following: 'Does the text acknowledge the gendered dimension of the AI or digital divide?' There are three possible answers: Yes, Partially or No. The level to which each criterion is satisfied in turn indicates whether, and on what fronts, the larger issue is currently being addressed.

The first matrix (Figure 3.1) examines three aspects of the use of AI in the distribution of online content during elections: (a) the use of AI to spread political disinformation and pollute the information environment, endangering the integrity of electoral information; (b) AI system design and incentives that

have an impact on the information environment; and (c) tools to mitigate the disruptive effect of AI on elections.

Figure 3.1. **Analytical matrix on AI in electoral contexts, applied to global and regional policy documents**

Organization *Document*	Does the text contain information on electoral disinformation and election integrity?		Does the text contain information on for-profit AI business models, human oversight mechanisms and recommender systems?			Does the text contain information on AI content moderation systems and provenance?	
	Electoral Disinformation	Election Integrity	Business Models	Human Oversight Mechanisms	Recommender and Ranking Algorithms	Moderation Systems	Provenance

Source: Developed by the authors.

The second matrix (Figure 3.2) covers AI's impact on women, girls and LGBTQIA+ communities by identifying the presence of four facets of digital marginalization: (a) the significant and disproportionate impact of generative AI on women, girls and LGBTQIA+ individuals; (b) the negative impact of AI technologies on political participation among women, girls and LGBTQIA+ individuals; (c) the transposition of the gendered digital divide into an AI divide, where women, girls and LGBTQIA+ individuals face disparities when it comes to knowledge, proficiency and trust in AI systems; and (d) whether or not AI strategies address discriminatory biases against women, girls and LGBTQIA+ individuals.

Figure 3.2. **Analytical matrix on AI, gender and inclusion, applied to global and regional policy documents**

Organization *Document*	Does the text address use of deepfakes to harass women and LGBTQIA+ individuals?	Does the text raise the potential negative impact of AI on women and LGBTIQA+ individuals' political participation?	Does the text acknowledge the gendered dimension of the AI or digital divide?	Does the text address the issue of data set induced discriminatory biases against women and LGBTQIA+ individuals?

Source: Developed by the authors.

3.2. **GLOBAL ANALYSIS**

All of the frameworks analysed discuss at least some of the selected operational indicators for AI in electoral contexts, with a cumulative explicit coverage rate of 55.1 per cent. However, no single question is addressed more often or left out altogether. The point addressed most often was the need to involve human oversight mechanisms in the development and deployment

Figure 3.3. **Overview of the discussion on AI in electoral contexts in global policy documents**

Organization *Document*	Does the text contain information on electoral disinformation and election integrity?		Does the text contain information on for-profit AI business models, human oversight mechanisms and recommender systems?			Does the text contain information on AI content moderation systems and provenance?	
	Electoral Disinformation	Election Integrity	Business Models	Human Oversight Mechanisms	Recommender and Ranking Algorithms	Moderation Systems	Provenance
UNESCO Recommendation on the Ethics of Artificial Intelligence (2022)	✖	✖	➖	➖	✔	✔	➖
UNESCO Artificial Intelligence and Democracy Report (2024)	✔	✔	✔	✔	✔	✔	✖
OECD Recommendation of the Council on Artificial Intelligence (2019)	➖	✖	✖	✔	✔	➖	✔
Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (2024)	✖	✔	✖	✔	✖	✔	✔
UN Global Digital Compact (2024)	✔	✔	✖	✔	✔	✔	✔
UN Governing AI for Humanity Report (2024)	✔	✔	✔	✔	✖	✔	✖
G20 Artificial Intelligence Guidelines (2019)	✖	✖	✔	✖	✖	✖	✖

Sources: United Nations Educational, Scientific and Cultural Organization (UNESCO), 'Recommendations on the Ethics of Artificial Intelligence', 16 May 2023; UNESCO, 'Artificial Intelligence and Democracy', 16 May 2024; Organisation for Economic Co-operation and Development (OECD), Recommendation of the Council on Artificial Intelligence, adopted on 22 May 2019, amended on 3 May 2024, OECD Legal Instruments; Council of Europe, 'Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law', 5 September 2024; United Nations, Global Digital Compact, 22 September 2024, A/79/L.2; United Nations, *Governing AI for Humanity: Final Report* (New York: United Nations, 2024); and G20, Ministerial Statement on Trade and Digital Economy (AI Principles), 9 June 2019.

Figure 3.4. **Overview of the discussion on AI, gender and inclusion in global policy documents**
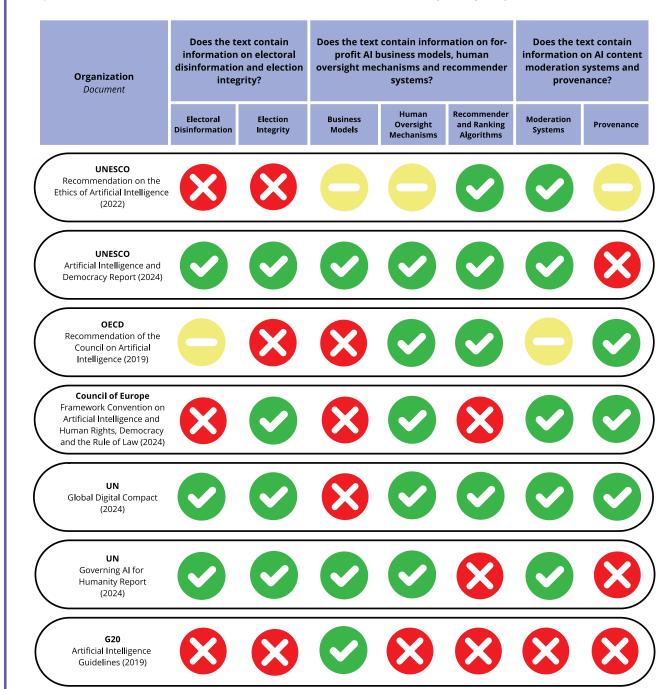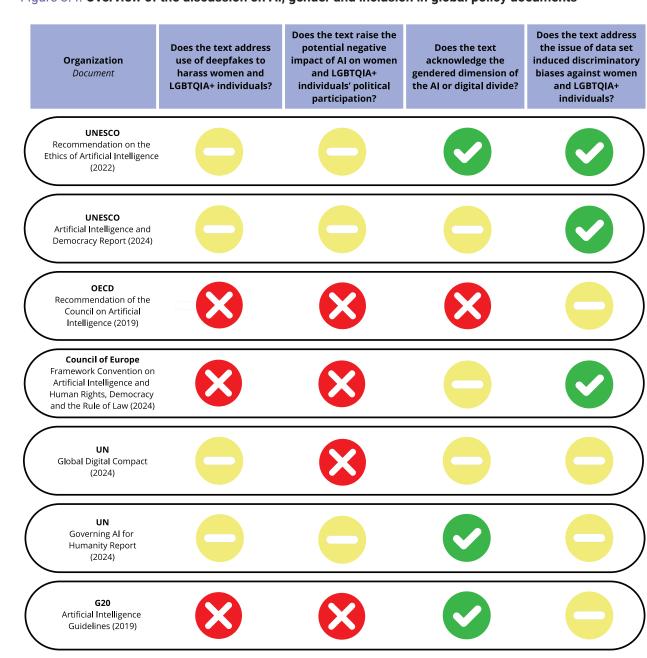
Organization *Document*	Does the text address use of deepfakes to harass women and LGBTQIA+ individuals?	Does the text raise the potential negative impact of AI on women and LGBTQIA+ individuals' political participation?	Does the text acknowledge the gendered dimension of the AI or digital divide?	Does the text address the issue of data set induced discriminatory biases against women and LGBTQIA+ individuals?
UNESCO Recommendation on the Ethics of Artificial Intelligence (2022)	➖	➖	✅	✅
UNESCO Artificial Intelligence and Democracy Report (2024)	➖	➖	➖	✅
OECD Recommendation of the Council on Artificial Intelligence (2019)	❌	❌	❌	➖
Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (2024)	❌	❌	➖	✅
UN Global Digital Compact (2024)	➖	❌	➖	➖
UN Governing AI for Humanity Report (2024)	➖	➖	✅	➖
G20 Artificial Intelligence Guidelines (2019)	❌	❌	✅	➖

Sources: United Nations Educational, Scientific and Cultural Organization (UNESCO), 'Recommendations on the Ethics of Artificial Intelligence', 16 May 2023; UNESCO, 'Artificial Intelligence and Democracy', 16 May 2024; Organisation for Economic Co-operation and Development (OECD), Recommendation of the Council on Artificial Intelligence, adopted on 22 May 2019, amended on 3 May 2024, OECD Legal Instruments; Council of Europe, 'Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law', 5 September 2024; United Nations, Global Digital Compact, 22 September 2024, A/79/L.2; United Nations, *Governing AI for Humanity: Final Report* (New York: United Nations, 2024); and G20, Ministerial Statement on Trade and Digital Economy (AI Principles), 9 June 2019.

of AI, which 71.4 per cent of the frameworks address explicitly and 14.3 per cent address partially. The least commonly addressed topic in the documents is the for-profit business models of AI companies and their associated incentives, which are addressed explicitly in 42.8 per cent of the texts and partially in 14.3 per cent. Notably, the document containing the least amount of discussion of the indicators by a wide margin is the G20's 'Artificial Intelligence Principles', which focuses solely on the for-profit business models of AI companies.

While all the examined policies, recommendations and guidelines address the interplay of AI and the electoral information environment to some extent, the facets and complexity of the discussions vary greatly. This variation can, in part, be attributed to the scope of the documents, as the more comprehensive texts dedicate space to problematizing issues at the intersection of AI and its impact on society. This discrepancy can be exemplified in the extensive discussions on electoral information integrity found in UNESCO's *Artificial Intelligence and Democracy* (2024) report and the UN's *Governing AI for Humanity Report* (2024), compared with the limited examples provided in the G20's *AI Principles* (2019). In addition, instruments with a comprehensive ambition, such as the UN's Global Digital Compact, lack discussions on underlying mechanisms and incentives that are crucial for attaining a complete understanding of why and how AI gives rise to certain threats to the information environment. The lack of cohesion among frameworks risks undermining the multifaceted impact of AI, understating the complexity and diversity of its risks.

Overall, the texts include significantly less extensive coverage of gender-related issues, with a cumulative explicit coverage rate of 21.4 per cent (a partial coverage rate of 50 per cent). The most frequently addressed issue is the existence of data set–induced biases against women and LGBTQIA+ individuals, which is covered explicitly in 42.9 per cent of the texts and partially in 57.1 per cent. The potential negative impact of AI on women and LGBTQIA+ individuals' political participation is the least addressed aspect covered by the analytical framework, with only 42.9 per cent of the frameworks raising the issue. While barely any of the texts explicitly covered more than one of the indicators, the Recommendation of the Organisation for Economic Co-operation and Development's Council on Artificial Intelligence addressed the fewest, with only one point being partially met (OECD 2024).

In terms of gender issues, the texts exhibit similar levels of inconsistency concerning what issues are addressed and who is included in the impacted groups. In general, women and girls are frequently identified among the groups who are disproportionately impacted by AI. It is much less common for members of the LGBTQIA+ community to be explicitly identified as particularly vulnerable. The incomplete representation of groups vulnerable to technologically facilitated gender-based violence risks overlooking the specific harm to and needs of LGBTQIA+ individuals, thereby perpetuating existing structures of oppression in online spaces.

The lack of cohesion among frameworks risks undermining the multifaceted impact of AI, understating the complexity and diversity of its risks.

3.3. REGIONAL ANALYSIS: LATIN AMERICA

Figure 3.5. **Overview of the discussion on AI in electoral contexts in Latin American policy documents**

Organization *Document*	Does the text contain information on electoral disinformation and election integrity?		Does the text contain information on for-profit AI business models, human oversight mechanisms and recommender systems?			Does the text contain information on AI content moderation systems and provenance?	
	Electoral Disinformation	Election Integrity	Business Models	Human Oversight Mechanisms	Recommender and Ranking Algorithms	Moderation Systems	Provenance
Argentina Recommendations for Reliable Artificial Intelligence (2023)	❌	❌	❌	✅	❌	➖	❌
Brazil Artificial Intelligence Strategy (2021)	❌	❌	✅	✅	✅	✅	❌
Chile National Artificial Intelligence Policy (2021)	❌	❌	❌	✅	✅	❌	❌
Costa Rica National Artificial Intelligence Strategy 2024-2027 (2024)	✅	✅	❌	✅	✅	❌	❌
Peru National Artificial Intelligence Strategy (2021)	❌	❌	❌	❌	❌	❌	❌
Mexico National Agenda on Artificial Intelligence (2018)	❌	❌	❌	✅	❌	✅	❌
Montevideo Declaration on Artificial Intelligence and its Impact in Latin America (2023)	❌	❌	❌	❌	❌	❌	❌

Sources: Argentina, Subsecretaría de tecnologías de la información, Recomendaciones para una inteligencia artificial fiable [Recommendations for reliable artificial intelligence], 1 June 2023; Brazil, Ministério da Ciência, Tecnologia e Inovações, Estratégia Brasileira de Inteligência Artificial [Artificial Intelligence Strategy of Brazil], 13 July 2021; Chile, Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, Política nacional de inteligencia artificial [National Artificial Intelligence Policy], 27 October 2021; Costa Rica, Ministry of Science, Innovation, Technology, and Telecommunications (MICITT), National Artificial Intelligence Strategy of Costa Rica 2024–2027 (San José: MICITT, 2024); Peru, Secretaría de Gobierno y Transformación Digital, Estrategia Nacional de Inteligencia Artificial [National Artificial Intelligence Strategy], 2021; IA2030Mx Coalition, Agenda Nacional Mexicana de Ineligencia Artificial [Mexican National Agenda on Artificial Intelligence], September 2020; and the Montevideo Declaration on Artificial Intelligence and its Impact in Latin America, 10 March 2023.

Figure 3.6. Overview of the discussion on AI, gender and inclusion in Latin American policy documents

Organization *Document*	Does the text address use of deepfakes to harass women and LGBTQIA+ individuals?	Does the text raise the potential negative impact of AI on women and LGBTIQA+ individuals' political participation?	Does the text acknowledge the gendered dimension of the AI or digital divide?	Does the text address the issue of data set induced discriminatory biases against women and LGBTQIA+ individuals?
Argentina Recommendations for Reliable Artificial Intelligence (2023)	❌	❌	✅	➖
Brazil Artificial Intelligence Strategy (2021)	❌	❌	✅	✅
Chile National Artificial Intelligence Policy (2021)	❌	❌	✅	✅
Costa Rica National Artificial Intelligence Strategy 2024-2027 (2024)	❌	❌	✅	✅
Peru National Artificial Intelligence Strategy (2021)	❌	❌	❌	➖
Mexico National Agenda on Artificial Intelligence (2018)	❌	❌	✅	➖
Montevideo Declaration on Artificial Intelligence and its Impact in Latin America (2023)	❌	❌	❌	❌

Sources: Argentina, Subsecretaría de tecnologías de la información, Recomendaciones para una inteligencia artificial fiable [Recommendations for reliable artificial intelligence], 1 June 2023; Brazil, Ministério da Ciência, Tecnologia e Inovações, Estratégia Brasileira de Inteligência Artificial [Artificial Intelligence Strategy of Brazil], 13 July 2021; Chile, Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, Política nacional de inteligencia artificial [National Artificial Intelligence Policy], 27 October 2021; Costa Rica, Ministry of Science, Innovation, Technology, and Telecommunications (MICITT), *National Artificial Intelligence Strategy of Costa Rica 2024–2027* (San José: MICITT, 2024); Peru, Secretaría de Gobierno y Transformación Digital, Estrategia Nacional de Inteligencia Artificial [National Artificial Intelligence Strategy], 2021; IA2030Mx Coalition, Agenda Nacional Mexicana de Ineligencia Artificial [Mexican National Agenda on Artificial Intelligence], September 2020; and the Montevideo Declaration on Artificial Intelligence and its Impact in Latin America, 10 March 2023.

The Latin American policy documents analysed generally failed to address most of the indicators, with a cumulative affirmative response rate of 26.5 per cent—28.6 percentage points lower than for the global documents. Two documents (Peru's National Artificial Intelligence Strategy and the Montevideo Declaration on Artificial Intelligence and Its Impact in Latin America) did not address a single indicator. By a wide margin, the indicator discussed most often was the need for human oversight mechanisms, which was raised in 71.4 per cent of the analysed texts. The points discussed least often were provenance systems, which were not mentioned in any of the texts, followed by electoral disinformation, election integrity and AI business models, each covered in only a single document.

Compared with the surveyed global policies, the regional strategies in Latin America infrequently and often inadequately address the connection between AI and the distribution and management of online content. Aside from scarce mentions of the need to employ human oversight and online content moderation, the documents do not address AI's potential influence in electoral information environments, nor do they advocate for specific mechanisms to steward AI, such as provenance systems or transparency in recommender and ranking algorithms.

Several of the surveyed countries have officially negotiated and endorsed global digital policy agreements, such as Peru's and Mexico's active participation in the development of the Council of Europe's Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. As AI policy development is an ongoing global process, inconsistencies between or gaps in national and global strategies may in part be explained by the overlap of many processes taking place in parallel and supposedly informing each other. This explanation is not wholly representative, however, as more recently published national strategies, such as Costa Rica's (2024), still fail to acknowledge key issues of AI and information integrity.

While the analysed documents also lacked discussions on topics related to AI and gender, there are clearer patterns in what areas are addressed. The total explicit coverage rate stands at 28.6 per cent; all the points covered pertain to the gendered AI or digital divide and the existence of structural, discriminatory gendered biases. None of the documents discuss how deepfakes are used as a form of technologically facilitated gender-based violence against women or the potential negative impact of AI on women and gender-marginalized individuals' political participation. In contrast, a majority of the texts explicitly or partially cover the gendered dimension of the AI or digital divide (71.4 per cent explicit mentions) and the issue of data set–induced gendered biases (71.4 per cent explicit mentions, 42.8 per cent partial mentions).

Like their global equivalents, Latin American AI strategies present a patchwork of statements on how AI impacts groups marginalized on the basis of gender. While systematic discriminatory bias is often acknowledged, how gendered bias impacts LGBTQIA+ individuals is not consistently addressed, such as in the cases of Argentina, Costa Rica and Mexico. The texts particularly

lack discussions of how weaponized generative AI disproportionally targets women, girls and LGBTQIA+ individuals and how it could disrupt or discourage them from political participation. Almost all the analysed strategies address how the digital divide limits access for marginalized groups, particularly women, to AI technologies. In most cases, statements on the digital divide segue into recommendations for measures to increase women's participation in AI development and, more broadly, in science, technology, engineering and mathematics (STEM) fields. The potential impact that the AI divide could have on civic participation is absent, however, leaving out how marginalized communities not only lack access to the potential benefits of AI but are also active targets when it is used for harm.

In the Brazilian case, recent policy developments have expanded the country's regulatory and strategic approach to AI. In 2024, Brazil relaunched its national AI strategy, incorporating additional public investments and emphasizing initiatives such as the development of a Portuguese LLM and the application of AI in public services, with investments planned until 2028 (MCTI 2024). Additionally, in December 2024, the Brazilian Senate approved the Brazilian Framework for Artificial Intelligence (Marco Legal da Inteligência Artificial), establishing governance structures for AI development and commercialization (Zanatta and Rielli 2024; Digital Policy Alert n.d.). The framework also introduces provisions to safeguard copyright for creators whose work is used in AI training. However, the legislation largely sidesteps issues related to social media platforms, content moderation or elections—topics that were evaluated as highly contentious among political actors, particularly right-wing parties.

Latin American AI strategies present a patchwork of statements on how AI impacts marginalized groups. While systematic discriminatory bias is often acknowledged, how gendered bias impacts LGBTQIA+ individuals is not consistently addressed.

Chapter 4
POLICY RECOMMENDATIONS

1. *Address high-risk uses of AI in electoral processes.* Global and regional AI policy frameworks must proactively tackle controversial and harmful applications rather than avoiding discussions of malicious or politically disruptive uses. Many AI-driven challenges in Mexico's elections were not adequately covered by national, regional or global regulations, particularly concerning electoral manipulation and gender-based digital violence. Despite challenges in cohesive enforcement, Brazil's pre-emptive judicial regulation by the Superior Electoral Court offers concrete pathways for managing direct AI applications, such as prohibiting unlabelled AI-generated content and permitting regulated AI-driven automated interactions.

2. *Mandate transparency in AI systems used for content curation and political advertising.* Governments should require platforms to disclose the data sets used to train AI algorithms, explain the logic behind content ranking and personalization, and provide clear explanations for why users are shown specific political ads or content.

3. *Develop provenance and labelling mechanisms for AI-generated content.* Countries and platforms should implement robust provenance systems and mandate the labelling of AI-generated material—particularly during electoral periods—to ensure that users can verify the origin and authenticity of online content.

4. *Expand policy focus beyond political manipulation.* AI's role in politics extends beyond traditional disinformation tactics, influencing campaign mobilization, humour-driven engagement and economic exploitation. AI was used in many ways during the 2024 elections in Brazil and Mexico that were not explicitly deceptive; rather, they helped shape political narratives through humour, pictorial representations or monetized content strategies. Future policies must recognize AI's evolving role in electoral dynamics,

ensuring accountability not only for deliberate manipulation but also for its broader effects on political engagement and digital ecosystems.

5. *Address AI-facilitated gender-based violence in legal and regulatory frameworks.* National AI strategies must explicitly recognize and address deepfakes and other forms of technology-facilitated abuse that disproportionately affect women, girls and LGBTQIA+ individuals, especially in political and electoral contexts.

6. *Introduce algorithmic audits and ensure data access for independent oversight.* Public policy should ensure that independent bodies, including electoral commissions and civil society organizations, have access to platform data and algorithms for auditing purposes—especially during elections—to evaluate bias, discrimination and democratic risks.

7. *Ensure that AI strategies promote both the protection and participation of marginalized groups.* Governments must go beyond acknowledging the digital divide and actively promote the inclusion of Indigenous communities, women and LGBTQIA+ individuals in the design, development and governance of AI, including mechanisms for Indigenous data sovereignty and linguistic equity.

8. *Encourage political parties, political campaigns and candidates to sign a code of conduct.* Although codes of conduct are a form of soft law, they can foster cooperation and encourage positive digital campaigns that avoid the negative uses of digital tools to manipulate information ecosystems. Through codes of conduct, political parties, campaigns and candidates can agree to limit the possible impact of digital technologies, especially AI, embrace good practices and provide safer spaces for *healthy political campaigns.*

9. *Stop the monetization of information manipulation and the lack of political finance controls for digital campaigns.* Manipulating the information environment is often profitable, and it may become even more so with the help of AI. Governments, digital platforms and political parties must address the monetization of information manipulation to reduce the financial incentives for engaging in manipulative practices. Similarly, the lack of political finance oversight for digital campaigns often allows parties and candidates to bypass financial regulation while campaigning online.

References

Aaberg, R., Gurung, P., Kandie, C., Méndez, L. V. and Xón, M. V., 'Engaging Indigenous Peoples in Elections: Identifying International Good Practices through Case Studies in Guatemala, Kenya, and Nepal', International Foundation for Electoral Systems, February 2024, 2024), <https://www.ifes.org/sites/default/files/2024-02/IFES%20 Engaging%20Indigenous%20Peoples%20in%20Elections.pdf>, accessed 18 June 2025

Ali, M., Mislove, A. Sapiezynski, P., Rieke, A. and Korolova, A., 'Ad Delivery Algorithms: The Hidden Arbiters of Political Messaging', Proceedings of the 14th ACM International Conference on Web Search and Data Mining, 17 December 2019, <https://www.semanticscholar.org/reader/822a9eed86d0375719ce71f0c2ef3b 09feb458f5>, accessed 18 June 2025

Amnesty International, 'Surveillance giants: How the business model of Google and Facebook threatens human rights', 21 November 2019, <https://www.amnesty.org/ en/documents/pol30/1404/2019/en>, accessed 2 June 2025

ARTICLE 19, 'Taming Big Tech: Protecting Freedom of Expression through the Unbundling of Services, Open Markets, Competition, and Users' Empowerment', 2021, (London: ARTICLE 19, 2021, <https://www.article19.org/wp-content/uploads/ 2021/12/Taming-big-tech_FINAL_8-Dec-1.pdf>, accessed 17 June 2025

Borchardt, C., '¿Qué dice la Ley Olimpia sobre el uso de la IA?' [What does the Olimpia Law say about the use of AI?], 10 October 2024, El Imparcial, 10 October 2024, <https://www.elimparcial.com/mexico/2024/10/11/que-dice-la-ley-olimpia-sobre-el -uso-de-la-ia>, accessed 18 June 2025

Burgess, M., 'The biggest deepfake abuse site is growing in disturbing ways', WIRED, 15 December 2021, <https://www.wired.com/story/deepfake-nude-abuse>, accessed 18 June 2025

Caballero, B. I., Ammon, B., Gomes, C., da Silva, C. C., Gomes, E., Roquete, F., Dias, F., Maia, G., Lopes, J., Costa, J., Mendes, L. S., Teixeira Ferreira, L. M., de Paula Silveira, M., de Oliveira, M., Vogel, S., Junquilho, T. A. and Marcílio, T., *Construindo consensos:* Deep fakes *nas eleições de 2024* [Building Consensus: Deepfakes in the 2024 Elections] (Brasília: Instituto Brasiliense de Direito Público, 2024), <https:// www.idp.edu.br/arquivos/cedis/IDP%20-%20LIA%2C%20CEDIS%20e%20ETHICS4AI %20-%20Nota%20T%C3%A9cnica%20-%20Construindo%20Consensos%20-%20 Deep%20Fakes%20nas%20Elei%C3%A7%C3%B5es%20de%202024.pdf>, accessed 2 June 2025

Carter Center, The, 'Analyzing Bolivia's 2020 General Elections: Final Report', June 2021, <https://www.cartercenter.org/resources/pdfs/news/peace_publications/election _reports/bolivia-2020-final-report.pdf>, accessed 2 June 2025

Cerdán Martínez, V. and Padilla Castillo, G., 'Historia del "fake" audiovisual: "deepfake" y la mujer en un imaginario falsificado y perverso' [History of audiovisual 'fakes': 'Deepfakes' and women in a falsified and perverse imaginary], *Historia y Comunicación Social*, 24/2 (2019), pp. 505–20, <https://doi.org/10.5209/hics .66293>

Coppock, A., Hill, S. J. and Vavreck, L., 'The small effects of political advertising are small regardless of context, message, sender, or receiver: Evidence from 59 real-time randomized experiments', *Science Advances*, 6/26 (2020), <https://doi.org/10.1126/sciadv.abc4046>

Del Pozo, C. M. and Arroyo, D. R., *Generative IA and Its Influence on Mexico's 2024 Elections* (Potsdam-Babelsberg: Friedrich Naumann Foundation and German Council on Foreign Relations, 2024), <https://dgap.org/system/files/article_pdfs/Policy%20Paper%20AI%20on%20Mexico%27s%202024%20Electons%20PDF.pdf>, accessed 2 June 2025

Digital Policy Alert, 'Brazil: Passed Bill No. 2338 of 2023 regulating the use of Artificial Intelligence including data protection measures', [n.d.], <https://digitalpolicyalert.org/event/25132-passed-bill-no-2338-of-2023-regulating-the-use-of-artificial-intelligence-including-data-protection-measures>, accessed 17 June 2025

DiResta, R., *Invisible Rulers: The People Who Turn Lies into Reality* (New York, NY: Public Affairs, 2024)

Elswah, M., 'Investigating content moderation systems in the Global South', Center for Democracy & Technology, 30 January 2024, <https://cdt.org/insights/investigating-content-moderation-systems-in-the-global-south>, accessed 17 June 2025

European Partnership for Democracy, 'Targeting and Amplification in Online Political Advertising', [n.d.], <https://epd.eu/content/uploads/2023/08/Targeting-and-amplification-in-online-political-advertising.pdf>, accessed 17 June 2025

Farrugia, B., 'Brazil's electoral deepfake law tested as AI-generated content targeted local elections', DFRLab, 26 November 2024, <https://dfrlab.org/2024/11/26/brazil-election-ai-deepfakes>, accessed 2 June 2025

FGV (Fundação Getulio Vargas) Comunicação Rio, 'Artificial Intelligence in the World: Mapping Its Uses During the 2024 Elections', 8 November 2024, <https://midiademocracia.fgv.br/en/estudos/inteligencia-artificial-no-mundo-mapeamento-dos-usos-durante-eleicoes-de-2024>, accessed 17 June 2025

FGV (Fundação Getulio Vargas) DAPP, *Nem Tão Simples: O Impacto Da Inteligência Artificial No Ecossistema Informacional* [It's Not Simple: The Impact of Artificial Intelligence on the Informational Ecosystem] (Rio de Janeiro: FGV, 2017), <https://repositorio.fgv.br/server/api/core/bitstreams/4eb748af-90ff-4f93-a3dd-6ff58382580b/content>, accessed 2 June 2025

Flores Saviaga, C. and Savage, S., 'Generative AI misinformation in the 2024 Mexican elections', Series on Generative Artificial Intelligence and Elections, Center for Media Engagement, October 2024, <https://mediaengagement.org/research/generative-artificial-intelligence-and-elections>, accessed 16 June 2025

Fundación Multitudes, *Análisis Comparativo de los Marcos Legales Actuales en América Latina para Combatir la Violencia de Género en Línea y Apoyar el Liderazgo de las Mujeres en la Vida Pública* [Comparative Analysis of Current Legal Frameworks in Latin America to Combat Online Gender-Based Violence and Support Women's Leadership in Public Life] (Fundación Multitudes, 2023), <https://www.boell.de/sites/default/files/2023-11/estudio-comparativo-fm-es.pdf>, accessed 17 June 2025

Funk, A., Shahbaz, A. and Vesteinsson, K., 'The repressive power of artificial intelligence', in A. Funk, A. Shahbaz, K. Vesteinsson, J. Brody, G. Baker, C. Grothe, M. Barak, M. Masinin, R. Modi and E. Sutterlin (eds), *Freedom on the Net 2023* (Washington, DC: Freedom House, 2023), <https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence>, accessed 2 June 2025

G20, Ministerial Statement on Trade and Digital Economy (AI Principles), 9 June 2019, <https://wp.oecd.ai/app/uploads/2021/06/G20-AI-Principles.pdf>, accessed 19 June 2025

Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M. and Sedova, K., 'Generative language models and automated influence operations: Emerging threats and potential mitigations', Georgetown University's Center for Security and Emerging Technology, OpenAI and Stanford Internet Observatory, January 2023, <https://doi.org/10.48550/arXiv.2301.04246>

González-Véliz, C. and Cuzcano-Chavez, X., 'Desafíos y dimensiones de la desinformación en ALAC deepfakes y la urgencia de proteger los derechos de las mujeres' [Challenges and dimensions of disinformation in Latin America and the Caribbean: Deepfakes and the urgent need to protect women's rights], *Revista Espacio I+D Innovación más Desarrollo*, 13 (2024), pp. 162–75, <https://www.researchgate.net/publication/378849055_Desafios_y_dimensiones_de_la_desinformacion_en_ALAC_deepfakes_y_la_urgencia_de_proteger_los_derechos_de_las_mujeres>, accessed 2 June 2025

Gorton, W. A., 'Manipulating citizens: How political campaigns' use of behavioral social science harms democracy', *New Political Science*, 38/1 (2016), pp. 61–80, <https://doi.org/10.1080/07393148.2015.1125119>

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B. and Lazer, D., 'Fake news on Twitter during the 2016 U.S. presidential election', *Science*, 363/6425 (2019), pp. 374–78, <https://www.science.org/doi/10.1126/science.aau2706>

Hofmann, V., Kalluri, P. R., Jurafsky, D. and King, S., 'Dialect prejudice predicts AI decisions about people's character, employability, and criminality', 1 March 2024, <https://doi.org/10.48550/arXiv.2403.00742>

Instituto Nacional Electoral (INE), 'Programa de resultados electorales preliminares 2024' [2024 Preliminary election results programme], 2024a, <https://prep2024.ine.mx/publicacion/nacional/presidencia/nacional/candidatura>, accessed 2 June 2025

—, 'Te Presentamos a "Inés", la asistente virtual del INE para combatir la desinformación sobre las elecciones 2024' [Meet 'Inés', the INE's virtual assistant to combat misinformation about the 2024 elections], 26 April 2024b, <https://centralelectoral.ine.mx/2024/04/26/te-presentamos-a-ines-la-asistente-virtual-del-ine-para-combatir-la-desinfomacion-sobre-las-elecciones-2024>, accessed 17 June 2025

International IDEA, *Global State of Democracy 2024: Strengthening the Legitimacy of Elections in a Time of Radical Uncertainty* (Stockholm: International IDEA, 2024), <https://doi.org/10.31752/idea.2024.55>

Juneja, P., *Artificial Intelligence for Electoral Management* (Stockholm: International IDEA, 2024), <https://doi.org/10.31752/idea.2024.31>

Kahn, G., 'Generative AI is already helping fact-checkers. But it's proving less useful in small languages and outside the West', Reuters Institute for the Study of Journalism, 29 April 2024, <https://reutersinstitute.politics.ox.ac.uk/news/generative-ai-already-helping-fact-checkers-its-proving-less-useful-small-languages-and>, accessed 2 June 2025

Kaye, B, 'Australia's Indigenous Voice referendum faces misinformation linked to COVID influencers', Reuters, 7 September 2023, <https://www.reuters.com/world/asia-pacific/australias-covid-agitators-inject-misinformation-into-indigenous-voice-vote-2023-09-07>, accessed 18 June 2025

Kira, B., 'Deepfakes, the weaponisation of AI against women and possible solutions', Verfassungsblog, 3 June 2024, <https://verfassungsblog.de/deepfakes-ncid-ai-regulation>, accessed 2 June 2025

Martínez Holguín, P., 'Mexico now has 13 female governors after 2024 elections', San Diego Red, 4 June 2024, <https://www.sandiegored.com/en/news/258798/Mexico-now-has-13-female-governors-after-2024-elections>, accessed 18 June 2025

Ministério da Ciência, Tecnologia e Inovação (MCTI), Brazil, 'IA para o Bem de Todos: Proposta de Plano Brasileiro de Inteligência Artificial 2024-2028' [AI for the good of all: Proposal for a Brazilian artificial intelligence plan 2024-2028], 29 July 2024, <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/noticias/2024/07/plano-brasileiro-de-ia-tera-supercomputador-e-investimento-de-r-23-bilhoes-em-quatro-anos/ia_para_o_bem_de_todos.pdf/view>, accessed 17 June 2025

Mont'Alverne, C., Arguedas, A. R., Toff, B. and Badrinathan, S., 'Trust in the media in Latin America: The same old story?', LSE Latin America and Caribbean Blog, 2 February 2021, <https://blogs.lse.ac.uk/latamcaribbean/2021/02/02/trust-in-the-media-in-latin-america-the-same-old-story>, accessed 2 June 2025

Montevideo Declaration on Artificial Intelligence and its Impact in Latin America, 10 March 2023, <https://doi.org/10.5281/zenodo.8208793>

Mozilla Foundation, 'YouTube Regrets: A Crowdsourced Investigation into YouTube's Recommendation Algorithm', July 2021, <https://foundation.mozilla.org/en/youtube/findings>, accessed 2 June 2025

Murphy, H., 'The trouble with deepfakes — liar's dividend', Financial Times Tech Tonic Podcast, 22 August 2024, <https://www.ft.com/content/7f22ce59-1c6c-4d84-bca8-dc539992e286>, accessed 17 June 2025

Naciones Unidas en México, 'Inteligencia artificial centrada en los pueblos indígenas: Perspectivas desde América Latina' [Artificial intelligence focused on Indigenous peoples: Perspectives from Latin America], 14 December 2023, <https://mexico.un.org/es/255872-inteligencia-artificial-centrada-en-los-pueblos-ind%C3%ADgenas-perspectivas-desde-am%C3%A9rica-latina>, accessed 2 June 2025

Narayanan, A., 'Understanding social media recommendation algorithms', Knight First Amendment Institute, Columbia University, 9 March 2023, <https://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms>, accessed 2 June 2025

Nicholas, G. and Bhatia, A., 'Lost in translation: Large language models in non-English content analysis', 12 June 2023, <https://doi.org/10.48550/arXiv.2306.07377>

Office of the United Nations High Commissioner for Human Rights (OHCHR), 'General Comment No. 25: The right to participate in public affairs, voting rights and the right of equal access to public service (Art. 25):. 12/07/96. CCPR/C/21/Rev.1/Add.7, General Comment No. 25', 12 July 1996, <https://www.equalrightstrust.org/ertdocumentbank/general%20comment%2025.pdf>, accessed 2 June 2025

Organisation for Economic Co-operation and Development (OECD), Recommendation of the Council on Artificial Intelligence, adopted on 22 May 2019, amended on 3 May 2024, OECD Legal Instruments, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>, accessed 2 June 2025

Organization for Security and Co-operation in Europe (OSCE), Office of the Representative on Freedom of the Media, *Spotlight on Artificial Intelligence and Freedom of Expression: A Policy Manual* (Vienna: OSCE, 2021), <https://www.osce.org/files/f/documents/8/f/510332_1.pdf>, accessed 2 June 2025

Ormerod, A. G., 'Latin Americans' social feeds are being flooded with AI-generated ads', Rest of the World, 19 July 2023, <https://restofworld.org/2023/newsletter-latin-america-generative-ai-ads>, accessed 17 June 2025

Oversight Board, 'Content Moderation in a New Era for AI and Automation', September 2024, <https://www.oversightboard.com/news/content-moderation-in-a-new-era-for-ai-and-automation>, accessed 2 June 2025

Panoptykon Fundacja, 'Fixing recommender systems. From identification of risk factors to meaningful transparency and mitigation', Briefing Note, 25 August 2023, <https://en.panoptykon.org/fixing-recommender-systems-identification-risk-factors-meaningful-transparency-and-mitigation>, accessed 17 June 2025

Peru, Republic of, Secretaría de Gobierno y Transformación Digital, Estrategia Nacional de Inteligencia Artificial [National Artificial Intelligence Strategy], 2021, <https://cayetano.edu.pe/wp-content/uploads/2023/05/Estrategia-Nacional-de-Inteligencia-Artificial.pdf>, accessed 2 June 2025

Piña, G., 'Cinco casos de "DeepFake porn" en México. La ley es insuficiente: Olimpia Coral' [Five cases of 'DeepFake porn' in Mexico. The law is insufficient: Olimpia Coral], EMEEQUIS, 29 June 2023, <https://m-x.com.mx/entrevistas/cinco-casos-de-deepfake-porn-en-mexico-la-ley-es-insuficiente-olimpia-coral>, accessed 2 June 2025

Poder360, 'Bolsonaro sugere uso de IA para questionar Alexandre de Moraes' [Bolsonaro suggests using AI to question Alexandre de Moraes], 22 February 2025, <https://www.poder360.com.br/poder-justica/bolsonaro-sugere-uso-de-ia-para-questionar-alexandre-de-moraes>, accessed 2 June 2025

Rangel, L., 'INE registró 205 quejas de violencia política de género en este proceso electoral' [INE registered 205 complaints of political gender violence in this electoral process], Animal Político, 29 June 2024, <https://animalpolitico.com/verificacion-de-hechos/te-explico/ine-violencia-politica-genero-elecciones>, accessed 17 June 2025

Resk, F., 'Tabata usa deep fake para expor Nunes como Ken de Barbie e refaz post' [Tabata uses deep fake to expose Nunes as Ken from Barbie and remakes post], Metropoles, 5 April 2024, <https://www.metropoles.com/sao-paulo/tabata-deep-fake-nunes-ken-barbie>, accessed 17 June 2025

Rid, T., Active Measures: The Secret History of Disinformation and Political Warfare (New York: Farrar, Straus and Giroux, 2020)

—, 'The lies Russia tells itself: The country's propagandists target the West—but mislead the Kremlin, too', Foreign Affairs, 30 September 2024, <https://www.foreignaffairs.com/united-states/lies-russia-tells-itself>, accessed 2 June 2025

Ríos Tobar, M., Violencia política de género en la esfera digital en América Latina (Stockholm: International IDEA, 2024), <https://doi.org/10.31752/idea.2024.93>

Rodríguez, A., 'Acusado un universitario de alterar con inteligencia artificial miles de imágenes de alumnas para venderlas como pornografía' [A university student has been accused of using artificial intelligence to alter thousands of images of female students to sell them as pornography], El País, 13 October 2023, <https://elpais.com/mexico/2023-10-14/acusado-un-universitario-de-alterar-con-inteligencia-artificial-miles-de-imagenes-de-alumnas-para-venderlas-como-pornografia.html>, accessed 2 June 2025

Shahid, F., 'Colonialism in content moderation research: The struggles of scholars in Majority World', Center for Democracy & Technology, 26 August 2024, <https://cdt .org/insights/colonialism-in-content-moderation-research-the-struggles-of-scholars -in-the-majority-world>, accessed 18 June 2025

Simon, F. M., Altay, S. and Mercier, H., 'Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown', *HKS Misinformation Review*, 18 October 2023, <https://doi.org/10.37016/mr-2020-127>

Soliman, W. and Rinta-Kahila, T., 'Unethical but not illegal! A critical look at two-sided disinformation platforms: Justifications, critique, and a way forward', *Journal of Information Technology*, 39/3 (2024), pp. 441–76, <https://doi.org/10.1177/ 02683962231181145>

Sosa Santiago, M. F., 'AMLO alerta sobre estafa con IA utilizando su imagen para invertir en Pemex: ¿Cuál es la inversión mínima que piden los estafadores?' [AMLO warns about AI scam using his image to invest in Pemex: What is the minimum investment the scammers are asking for?], El Economista, 26 April 2024, <https:// www.eleconomista.com.mx/politica/AMLO-alerta-sobre-estafa-con-IA-utilizando -su-imagen-para-invertir-en-Pemex-Cual-es-la-inversion-minima-que-piden-los -estafadores-20240426-0049.html>, accessed 18 June 2025

Souza, C. A., 'A Blueprint for Digital Propaganda in the Majority World: Online Coordinated Attacks, Narratives About Content Moderation and Challenges to Electoral Integrity in Brazil', Yale Law School, Information Society Project, Majority World Initiative, July 2024, <https://law.yale.edu/sites/default/files/area/center/isp/ documents/mwi-carlos-affonso-souza_2024-08-01_re-fin.pdf>, accessed 2 June 2025

Souza, L. and Varon, J., 'Violencia Política de Género en Internet' [Online gender-based political violence], Policy Paper, Al Sur, 2019, <https://www.alsur.lat/sites/default/ files/2021-07/Violencia Pol%C3%ADtica de G%C3%A9nero en Internet ES.pdf>, accessed 2 June 2025

Stevenson, M. and Verza, M., 'A violent, polarized Mexico goes to the polls to choose between 2 women presidential candidates', Associated Press 31 May 2024, <https://apnews.com/article/mexico-election-polarized-divided-heat-violence-4d 5f620f0f8f9b7ef6efa8b3083561a8>, accessed 18 June 2025

Törnberg, P. and Chueri, J., 'When do parties lie? Misinformation and radical-right populism across 26 countries', *The International Journal of Press/Politics*, 0/0 (2025), <https://doi.org/10.1177/19401612241311886>

Tribunal Regional Eleitoral–BA, 'Chatbot Maia passa a oferecer informações sobre Registro de Candidaturas' [Chatbot Maia now offers information on candidate registration], 23 October 2024, <https://www.tre-ba.jus.br/comunicacao/noticias/ 2023/Outubro/chatbot-maia-passa-a-oferecer-informacoes-sobre-registro-de -candidaturas>, accessed 18 June 2025

Tribunal Superior Eleitoral (TSE), Resolução nº 23.732, de 27 de fevereiro de 2024 [Resolution No. 23.732], Brasília, 27 February 2024, <https://www.tse.jus.br/ legislacao/compilada/res/2024/resolucao-no-23-732-de-27-de-fevereiro-de-2024>, accessed 2 June 2025

Udupa, S., Maronikolakis, A. and Wisiorek. A., 'Ethical scaling of content moderation: Extreme speech and the (in)significance of artificial intelligence', *Big Data & Society*, 10/1 (2023), <https://doi.org/10.1177/20539517231172424>

United Nations, Promotion and Protection of the Right to Freedom of Opinion and Expression, A/73/348, 29 August 2018, <https://documents-dds-ny.un.org/doc/UNDOC/GEN/N18/270/42/PDF/N1827042.pdf>, accessed 2 June 2025

—, Global Digital Compact, 22 September 2024a, A/79/L.2, <https://www.un.org/digital-emerging-technologies/global-digital-compact>, accessed 2 June 2025

—, *Governing AI for Humanity: Final Report* (New York: UN, 2024b), <https://digitallibrary.un.org/record/4062495?v=pdf>, accessed 19 June 2025

United Nations Educational, Scientific and Cultural Organization (UNESCO), *Artificial Intelligence and Democracy* (Paris: UNESCO, 2024), <https://www.unesco.org/en/articles/artificial-intelligence-and-democracy>, accessed 2 June 2025

Wood, B., 'Could Bolivia's current politics be fueling Indigenous discrimination?', PBS NewsHour, 25 November 2019, <https://www.pbs.org/newshour/world/could-bolivias-current-politics-be-fueling-indigenous-discrimination>, accessed 2 June 2025

Zanatta, R. A. F. and Rielli, M., 'The artificial intelligence legislation in Brazil: Technical analysis of the text to be voted on in the federal senate plenary', DataPrivacy Brazil Research, 12 December 2024, <https://www.dataprivacybr.org/en/the-artificial-intelligence-legislation-in-brazil-technical-analysis-of-the-text-to-be-voted-on-in-the-federal-senate-plenary>, accessed 19 June 2025

About the authors

Sebastian Becker Castellaro is an Associate Programme Officer of the Digitalization and Democracy Programme at International IDEA. His work focuses on the impact of technology in democracies worldwide, providing research, policy analysis, and project support. He holds an LLM in International Law from the Université Libre de Bruxelles and an LLM in Public Law from the University of Chile.

Mariana Carvalho is a researcher and professor at FGV Comunicação. She holds a PhD in Communication from the State University of Rio de Janeiro (UERJ) and a degree in Journalism from the Federal University of Espírito Santo (UFES). Her research focuses on narrative disputes and the role of emotions in political debate on digital platforms.

Alberto Fernandez Gibaja is the Head of the Digitalization and Democracy Programme at International IDEA. He oversees the work of the institute on the aspects at the intersection of technology and democracy. He holds a master's degree in International Studies from the University of Uppsala and a bachelor's degree from the Complutense University of Madrid.

Amaro Grassi is Vice-Director at the School of Communication, Media and Information at FGV Comunicação. He holds a PhD in Political Science and an MSc in Sociology from the Institute of Social and Political Studies at the University of the State of Rio de Janeiro (IESP-UERJ).

Cecilia Hammar is a Programme Assistant of the Digitalization and Democracy Programme at International IDEA. She conducts research and project coordination support, primarily relating to the societal implications of AI. She is currently pursuing an MSc in Political Science at the University of Uppsala.

Juliane Müller is an Associate Programme Officer in International IDEA's Digitalization and Democracy Programme. She specializes in the democratic implications of emerging technologies—particularly AI—with a focus on human rights and electoral integrity. Her work focuses on the Programme's global AI capacity-building initiative for electoral management bodies and includes research and policy analysis. She holds an LLM in International Law from the University of Edinburgh and an LLB from the University of Mannheim.

Laura Pereira holds a master's degree in social sciences (São Paulo State University) and works as project lead at FGV Comunicação. Her work focuses on the intersection of the Internet, communication and politics, with experience in research and policy analysis related to technology regulation and the political impacts of contemporary media consumption.

Victor Piaia holds a PhD in Sociology at the Institute of Social and Political Studies (IESP-UERJ). He is a professor at FGV Comunicação and a leader of the Study Group in Political Sociology and Digital Transformations. He studies the political effects of transformation in daily communication, focusing on social media platforms and messaging apps.

Marco Aurélio Ruediger holds a PhD in Sociology and an MSc in Policy Analysis and Management. He is currently the Director of the School of Communication, Media and Information at FGV Comunicação.

About International IDEA

The International Institute for Democracy and Electoral Assistance (International IDEA) is an intergovernmental organization with 35 Member States founded in 1995, with a mandate to support sustainable democracy worldwide.

WHAT WE DO

We develop policy-friendly research related to elections, parliaments, constitutions, digitalization, climate change, inclusion and political representation, all under the umbrella of the UN Sustainable Development Goals. We assess the performance of democracies around the world through our unique Global State of Democracy Indices and Democracy Tracker.

We provide capacity development and expert advice to democratic actors including governments, parliaments, election officials and civil society. We develop tools and publish databases, books and primers in several languages on topics ranging from voter turnout to gender quotas.

We bring states and non-state actors together for dialogues and lesson sharing. We stand up and speak out to promote and protect democracy worldwide.

WHERE WE WORK

Our headquarters is in Stockholm, and we have regional and country offices in Africa, Asia and the Pacific, Europe, and Latin America and the Caribbean. International IDEA is a Permanent Observer to the United Nations and is accredited to European Union institutions.

OUR PUBLICATIONS AND DATABASES

We have a catalogue with more than 1,000 publications and over 25 databases on our website. Most of our publications can be downloaded free of charge.

<https://www.idea.int>

This policy paper examines the impact of artificial intelligence (AI) on information integrity in electoral contexts across Latin America. Drawing on case studies from Brazil's and Mexico's 2024 elections, it analyses how AI tools are used to create, curate and moderate political content online, with particular attention to their effects on marginalized groups such as women, Indigenous communities and LGBTQIA+ individuals. While the use of generative AI in elections remains limited, its potential to undermine democratic discourse and amplify harmful content demands urgent attention.

To protect democratic processes, this policy paper calls for a robust regulatory framework that includes transparency obligations for platforms, safeguards against AI-facilitated gender-based violence and inclusive governance strategies. It recommends stronger oversight of AI use in elections, provenance systems for synthetic content and independent auditing of content curation algorithms. As AI evolves, so too must our commitment to electoral integrity and the protection of all voices in the democratic sphere.