


COMMENTARY

# Disinformation by design: leveraging solutions to combat misinformation in the Philippines' 2025 election

Tetiana Schipper 

German Institute for Global and Area Studies, Hamburg, Germany

Email: [schippertetiana@gmail.com](mailto:schippertetiana@gmail.com)

**Received:** 21 November 2024; **Revised:** 19 January 2025; **Accepted:** 13 April 2025

**Keywords:** AI-driven misinformation; AI ethics; deepfakes; digital autocratisation; digital media literacy; disinformation; electoral integrity; fact-checking; generative AI; historical revisionism; large language models; media regulation; Philippine elections; political campaigns; social media platforms

## Abstract

This commentary examines the dual role of artificial intelligence (AI) in shaping electoral integrity and combating misinformation, with a focus on the 2025 Philippine elections. It investigates how AI has been weaponised to manipulate narratives and suggests strategies to counteract disinformation. Drawing on case studies from the Philippines, Taiwan, and India—regions in the Indo-Pacific with vibrant democracies, high digital engagement, and recent experiences with election-related misinformation—it highlights the risks of AI-driven content and the innovative measures used to address its spread. The commentary advocates for a balanced approach that incorporates technological solutions, regulatory frameworks, and digital literacy to safeguard democratic processes and promote informed public participation. The rise of generative AI tools has significantly amplified the risks of disinformation, such as deepfakes, and algorithmic biases. These technologies have been exploited to influence voter perceptions and undermine democratic systems, creating a pressing need for protective measures. In the Philippines, social media platforms have been used to spread revisionist narratives, while Taiwan employs AI for real-time fact-checking. India's proactive approach, including a public misinformation tipline, showcases effective countermeasures. These examples highlight the complex challenges and opportunities presented by AI in different electoral contexts. The commentary stresses the need for regulatory frameworks designed to address AI's dual-use nature, advocating for transparency, real-time monitoring, and collaboration between governments, civil society, and the private sector. It also explores the criteria for effective AI solutions, including scalability, adaptability, and ethical considerations, to guide future interventions. Ultimately, it underscores the importance of digital literacy and resilient information ecosystems in supporting informed democratic participation.

## Policy Significance Statement

Artificial Intelligence (AI) is transforming electoral integrity worldwide, posing significant risks and offering innovative solutions. This commentary examines their impact, particularly in the context of the Philippines' 2025 midterm elections, where disinformation threatens democratic processes. Policymakers are urged to implement robust regulatory frameworks, prioritising transparency, real-time monitoring, and enhanced digital literacy. Cross-sector collaboration and the development of culturally and linguistically tailored AI tools are critical to building resilient information ecosystems. These measures will help protect electoral integrity, promote informed public participation, and reinforce trust in democratic institutions.

## 1. Introduction

The rapid proliferation of artificial intelligence (AI) technologies, including generative AI (GAI), large language models (LLMs), and natural language processing (NLP), is reshaping global information ecosystems. While these technologies enable transparency and accessibility, they also exacerbate misinformation risks, threatening electoral integrity. This shift is significant alongside developments in the Indo-Pacific, a region marked by diverse political systems, rising geopolitical tensions, and widespread digital engagement. The combination of evolving democratic institutions, varying levels of media literacy, and rapid technological adoption creates both opportunities and heightened challenges for managing misinformation. These factors make the Indo-Pacific especially vulnerable to AI-driven disinformation, with significant implications for electoral integrity and regional stability. As democracies increasingly rely on digital platforms, understanding the dual role of AI is vital for addressing disinformation and fostering trust in democratic processes.

This commentary explores AI's role in electoral misinformation, focusing on the Philippines' 2025 midterm elections, with comparative insights from Taiwan and India. These regions are selected due to their diverse political environments, significant digital engagement, and recent experiences with misinformation in elections, making them valuable case studies for understanding AI's impact in democratic processes. This analysis underscores the importance of regulatory frameworks, collaborative initiatives, and technological innovation to mitigate misinformation's impact. By combining case studies and actionable recommendations, this report contributes to the broader discourse on leveraging AI responsibly to protect democracy. It explores the dual role of AI and proposes scalable solutions and regulatory measures to combat disinformation effectively.

## 2. Section 1: AI as a double-edged sword

This section explores the dual role of AI in modern electoral processes around the world, highlighting its ability to both combat and propagate misinformation. It examines how AI contributes to misinformation through deepfakes (AI-generated manipulations of media), algorithmic biases, and hyper-realistic synthetic media. At the same time, these technologies offer solutions such as automated fact-checking, content moderation, and enhanced accessibility to political information. The section sets the stage for understanding AI's opportunities and risks within the broader context of democratic integrity.

It is essential to differentiate AI-generated material from disinformation, which can also stem from non-synthetic sources and often becomes more persuasive when grounded in partial truths or emotionally resonant narratives. Moreover, distinguishing between “fake news” (deliberately fabricated content) and “distorted news” (subtly manipulated facts) is essential. The latter is often more persuasive and difficult to identify, especially as the boundaries between fact and fiction are shaped by personal beliefs and context. Legal and cognitive challenges arise in defining and regulating misinformation, as these distinctions are not always clear-cut (Neuwirth, 2021).

Misinformation spreads faster and more broadly than the truth across all types of information, with false political news being particularly impactful (Vosoughi et al., 2018). Fake news often mimics legitimate content, making it difficult to differentiate the two, while its rapid spread outpaces fact-checking efforts. For instance, during the 2024 U.S. elections, X's AI chatbot, Grok, spread false information about the process for adding new candidates to the ballot. While X initially resisted corrections, election officials intervened to clarify the facts (Leingang, 2024). Similarly, in elections in Indonesia and Pakistan, AI-generated “softfakes”—manipulated media portraying candidates favourably—raised ethical concerns about voter manipulation and the risks to democratic processes (Chowdhury, 2024). Although sophisticated AI-generated disinformation had minimal impact on recent elections in the UK, France, and the European Parliament, it mostly reinforced existing beliefs (Stockwell, 2024). However, traditional methods, like bots (discussed in 3.1) and influencers, were more effective in reaching a broader audience and spreading disinformation (Heikkilä, 2024). While AI algorithms are helpful in addressing these issues, they have limitations. As deceptive tactics evolve, a multidisciplinary approach becomes essential (Aïmeur et al., 2023).

As AI technologies continue to shape global political discourse, the need for AI-specific policies and accountability measures becomes ever more urgent. The 2024 “super election year” saw AI-driven disinformation influence campaigns in over 60 countries. While safeguards such as policy protections, industry standards, and voter scepticism helped limit the negative effects, three key trends remain: the development of increasingly persuasive AI tools, the growing prevalence of AI-generated content, and public disengagement from political discourse (Carr and Köhler, 2024). Also, trust disparities remain stark. Developing countries report higher levels of trust in institutions compared to G7 nations. Governments face significant distrust, driven by perceptions of incompetence, unethical behaviour, and the belief that leaders intentionally mislead the public. Furthermore, poorly managed innovation and the perception of political interference in science further exacerbate trust issues, particularly in developed nations (Edelman, 2024). Also, Filipinos now demand tangible proof before extending trust, urging institutions to adopt transparency, competence, and ethical conduct as foundational values (EON The Stakeholders Relations Group and Ateneo de Manila University, 2024).

On the other hand, AI-generated content, especially from well-trained models, can positively contribute to democracy by enhancing access to accurate information, supporting fact-checking initiatives, and fostering informed public discourse. For example, AI, including NLP (discussed in 3.3) and machine learning, has been used in peacebuilding efforts by the United Nations (UN), enabling large-scale digital dialogues in conflict zones to identify shared concerns and potential areas of consensus (Alavi et al., 2022). Also, the 2024 Nobel Prizes in Physics and Chemistry recognised groundbreaking contributions to AI, underscoring its immense potential in shaping the future of medicine and science (Li and Gilbert, 2024).

The section underscores that while AI offers transformative opportunities for improving electoral integrity, its misuse poses significant risks. This duality necessitates a nuanced approach to leveraging these technologies responsibly. The next section delves into regional case studies to illustrate how AI-driven misinformation and countermeasures manifest in diverse political contexts.

### 3. Section 2: regional challenges and strategies

Disinformation is a global challenge affecting democracies at all stages of development, not just those with weak regulation or political instability. In the Indo-Pacific, the Philippines, Taiwan, and India provide distinct perspectives on how political, cultural, and technological factors influence the spread and management of disinformation.

The Philippines faces significant risks due to its young democracy, high digital engagement, and history of political manipulation through social media. These risks are heightened by increasing geopolitical pressure from China (Council on Foreign Relations, 2024), similar to Taiwan’s situation. Taiwan has built strong defences against disinformation through proactive regulation, real-time fact-checking, and media literacy—strategies the Philippines could adopt. India’s experience, with its vast, diverse population, offers insights into combating disinformation on a large scale through public reporting mechanisms and digital literacy initiatives, relevant to the Philippines’ own regional diversity across regions like Luzon, Visayas, and Mindanao.

This section provides analysis of how disinformation has influenced electoral processes in the Philippines, Taiwan, and India. It highlights the Philippines’ challenges with historical revisionism and social media exploitation, Taiwan’s strategies to counter geopolitical disinformation, and India’s innovative public reporting mechanisms during its general elections. By comparing these case studies, the section showcases both the commonalities and distinct responses across the region.

#### 3.1. Philippines: the role of social media in shaping the 2025 election

Social media’s influence is especially pronounced in the Philippines, which has one of the highest rates of social media usage globally (Balita, 2023; Telenor Asia, 2023). As the country approaches the 2025 midterm elections, it faces a growing threat from AI-driven disinformation. Building on patterns from previous election cycles discussed below, this threat now includes the added complexity of AI-generated

content, such as deepfakes, which has the potential to intensify disinformation and undermine electoral integrity. Digital literacy remains limited, and entrenched political interests continue to benefit from the spread of disinformation (Enriquez, 2024).

In 2016, the Philippines earned the label “patient zero” in the global disinformation epidemic due to rampant false narratives. Former President Duterte’s campaign effectively used social media to promote aggressive rhetoric, while media literacy efforts lagged. Disinformation networks like Twinmark Media amplified Duterte’s message through platforms such as Trending News Portal (TNP). Although Twinmark was banned from Facebook in 2019 for “coordinated inauthentic behaviour,” it quickly resurfaced with the help of micro-influencers, bypassing platform regulations (Fallorina et al., 2023; Hapal, 2024). During Duterte’s presidency (2016–2022), authoritarian policies, like the anti-drug campaign, gained support through disinformation from state-backed “troll farms,” framing opposition figures as communist sympathisers and silencing critics (Arugay and Mendoza, 2024).

In 2022, President Marcos Jr. constructed a complex media ecosystem blending historical revisionism with influencer narratives, polarising the political landscape and evading regulatory oversight. The Marcos Jr. campaign focused on rehabilitating the Marcos family image and swaying public opinion, particularly among young Filipinos who were digitally active but vulnerable to disinformation due to limited media literacy (Chua and Khan, 2023; Marcelino, 2023). Many young people in the Philippines, unaware of the dictatorship’s history of human rights abuses and corruption, have developed favourable views influenced, among other factors, by economic struggles and the punitive nature of post-Marcos reforms (Tigno et al., 2024). TikTok played a critical role in Marcos Jr.’s digital strategy, with influencers sharing videos portraying the Marcos regime as a time of prosperity and stability (de Guzman, 2022). TikTok’s algorithm amplified these messages, allowing them to go viral rapidly. This environment enabled the spread of revisionist narratives, including the “Marcos gold” myth and conspiracy theories about the EDSA revolution, which were presented as a fabricated power grab (Marcelino, 2023; Arugay and Mendoza, 2024). Marcos Jr. use of social media allowed him to avoid traditional media channels, which often critique the Marcos legacy (de Guzman, 2022; Marcelino, 2023). The blend of historical nostalgia, a desire for continuity, and regional loyalty outweighed secondary factors like age, education, and socioeconomic status in the success of Marcos Jr.’s campaign (Dulay et al., 2023). Disinformation campaigns have targeted both political figures and governmental institutions, spreading false narratives that led to harassment, violence, and stigmatisation (Fallorina et al., 2023).

This drift towards *digital autocratisation* under Duterte and Marcos Jr., fuelled by state-backed disinformation, poses a serious challenge (Arugay and Mendoza, 2024). Efforts to combat disinformation, led by civil society, academia, and media, focus on integrating Media and Information Literacy into school curricula and fostering fact-checking collaborations (Chua and Khan, 2023). Also, the National Library of the Philippines offers virtual reference services to ensure equitable access to information (Romero and Fuellos, 2024).

However, significant challenges remain, including legal, ethical, and privacy concerns, limited AI awareness, and resource constraints, as highlighted in the National AI Roadmap and the Philippine Innovative Startup Act (Marcelino, 2023; Amil, 2024). The country’s weak regulatory framework previously allowed entities like Cambridge Analytica to test online propaganda tactics (Wylie, 2019). While there is no direct evidence that Duterte and Marcos Jr. have used AI-driven tools, the growing use of these technologies in the Philippines highlights a serious threat to electoral integrity. Addressing these issues requires regulatory reforms and a collaborative, multi-stakeholder approach to dismantling entrenched disinformation networks (Enriquez, 2024). To regulate deepfake creation and distribution, the House of Representatives introduced the Deepfake Accountability and Transparency Act (Bill 10,567), requiring clear verbal and written disclosures for AI-generated content (Digital Policy Alert, 2024). Similarly, the Commission on Elections (COMELEC) has issued guidelines for the 2025 election to counter AI-driven disinformation. These include mandating transparency in AI-generated content and banning deepfakes used to spread falsehoods (Enriquez, 2024). However, disinformation campaigns remain highly adaptive, exploiting encrypted platforms like WhatsApp, which are challenging for AI systems to monitor.

To address the risks posed by deepfakes and synthetic media, scholars suggest strengthening existing laws, such as the Data Privacy Act, the Intellectual Property Code, and the Consumer Act, rather than introducing new regulations. They also recommend implementing a charge system to penalise irresponsible AI use and promoting co-regulation, which involves collaboration between government, industry, and civil society. Additionally, integrating AI governance into the National AI Roadmap is advised (Dayrit et al., 2024). Also, partnerships with international organisations, governments, and the private sector are crucial for technology transfer, capacity building, and improving digital literacy. A positive development is President Marcos Jr.'s emphasis on balanced global partnerships to enhance the country's internet infrastructure and cybersecurity (Schipper, 2024). Similarly, the Philippine Department of Information and Communications Technology (DICT) is collaborating with AI providers such as OpenAI and Google to counter the threat of deepfakes ahead of the 2025 midterm elections (Dizon, 2024). The DICT advocates embedding watermarks (discussed in 3.4) in AI-generated content to indicate its investments in tools to monitor and detect fake content online. Inspired by Singapore's approach, the DICT is exploring fact-checking mechanisms that allow disputed posts to remain visible but include government-verified information to provide balanced perspectives. However, improving media literacy among the population is essential for this measure to be effective.

Given the prevalence of historical revisionism in the Philippines, AI-driven tools must prioritise detecting and mitigating narrative manipulation, especially in politically sensitive contexts where cultural identity and national history are critical. Similar to the Philippines, other countries in the region, such as Taiwan, face their own unique challenges with misinformation, demonstrating the global nature of AI's dual use in electoral integrity.

### 3.2. *Taiwan: GAI in democratic engagement and disinformation*

Similar to the Philippines, Taiwan faces unique challenges with misinformation, albeit through different mediums and countermeasures. Taiwan, a stable democracy facing constant geopolitical pressure—primarily from China—implements proactive measures such as real-time fact-checking and media resilience. In Taiwan's 2018 local elections, the University of Queensland (Australia) employed an advanced AI algorithm to detect and explain fake news. This system not only identified false information but also clarified how it reached its conclusions, prioritising transparency and accountability (Sadiq and Demartini, 2024). During the 2020 elections, Taiwan's strategy also focused on swiftly identifying, combating, and punishing disinformation while promoting transparency. Key elements of this strategy include media literacy, rapid debunking, and coordination between government and civil society (Kuo, 2021).

Taiwan amended its laws in 2023, including the Presidential and Vice-Presidential Election and Recall Act and the Civil Servants Election and Recall Act, to impose severe penalties for deepfakes. The government continued collaboration with civil society and independent fact-checkers to combat disinformation. Taiwan AI Labs plays a proactive role, developing solutions like the "Infodemic" platform for real-time monitoring and analysis of disinformation (Council of Asian Liberals and Democrats, 2024).

However, during Taiwan's 2024 presidential election, GAI tools played a dual role as both allies and adversaries in the fight for democratic integrity. Incorporation of social, cultural, and political symbols into TikTok-based anti-disinformation campaigns highlights the complementary role of symbolic communication in fostering engagement and trust (Bhattacharya et al., 2024). Media outlets like Taiwan Television Broadcasting System (TVBS) and Formosa Television (FTV) leveraged GAI to counter disinformation effectively. However, challenges persisted with the rapid spread of AI-generated content on platforms such as YouTube and Douyin (TikTok). The proliferation of deepfakes and other AI-generated content blurred the distinction between factual and fabricated material, exacerbating political divisions and increasing susceptibility to foreign influence (Hung et al., 2024). Taiwan's experience underscores the urgent need for the Philippines to actively cultivate high media literacy and foster strong civil society engagement.



### 3.3. India: battle against AI-generated misinformation in the 2024 general election

India, as the world's largest democracy, faces challenges related to large-scale misinformation across diverse regions and languages, offering valuable insights into managing disinformation at a national level. To address this, the country integrates information literacy into education, promotes digital literacy through initiatives like the Digital India campaign, and improves access to trustworthy information sources (Bhakte, 2024). India's proactive integration of digital literacy into education curricula and community-driven fact-checking can serve as a model for the Philippines to emulate.

During the 2024 Indian General Election, AI-generated misinformation peaked, becoming a significant challenge. In response, the Misinformation Combat Alliance launched the Deepfakes Analysis Unit (DAU), a pioneering initiative that enabled the public to report suspicious audio and video content via a WhatsApp tipline (Nannaware et al., 2025). The DAU categorised content into *deepfake*, *cheapfake*, and *AI-generated*, aiding in the identification of misleading materials.

The tipline received hundreds of submissions—mainly videos—which were analysed using AI detection tools. When manipulation was confirmed, the DAU collaborated with fact-checkers to verify the content and publish public reports, offering guidance on identifying synthetic media. The initiative also highlighted the surge in *cheapfakes*—low-quality AI-generated content—which outnumbered sophisticated *deepfakes* during the election cycle (Raina, 2024).

By partnering with media outlets and detection experts, the DAU raised public awareness about AI-driven misinformation and set a precedent for global collaboration. To further combat AI-related election misinformation, India seeks to strengthen existing laws, such as the Information Technology Act, and encourage self-regulation for high-risk AI applications. Drawing inspiration from the EU's AI Act and the U.S.'s voluntary frameworks (as discussed in Section 4), India plans to develop targeted guidelines through collaborative governance, potentially through the proposed Artificial Intelligence Standards Institute (AISi) (Mohanty and Sahu, 2024).

Combating disinformation requires context-specific strategies, as no universal solution fits all. The Philippines faces widespread distrust in government due to past false narratives, now worsened by AI-driven disinformation, demanding proactive and culturally sensitive responses. Taiwan, despite strong democratic institutions, struggles with disinformation from foreign influence. India's vast diversity and linguistic complexity make misinformation harder to manage. Success in one country may create new challenges elsewhere. Governments must continuously adapt and build public trust to effectively counter disinformation. Building on these insights, the next section examines selected tools and frameworks available to combat misinformation effectively.

## 4. Section 3: leveraging solutions to combat misinformation

This section focuses on the technological tools and regulatory frameworks essential for combating misinformation. It discusses the potential of GAI—which can create new content and simulate human-like creativity—, LLMs—deep learning models that can understand and generate human language—, and advanced NLP methods—such as those used in chatbots and language translation—to detect and counter false narratives. The section also highlights ethical considerations, such as transparency, fairness, and human oversight, while analysing global and regional regulatory approaches to AI governance.

### 4.1. Generative AI (GAI)

GAI can create hyper-realistic synthetic content, such as deepfakes, which poses significant risks in spreading misinformation. For instance, non-consensual deepfake pornography underscores the urgent need for regulatory measures (Roseman, 2024). Similarly, deepfake videos and audio are increasingly weaponised in disinformation campaigns, influencing public opinion in concerning ways (Li and Calligari, 2024). Disinformation bots can misuse GAI to spread false narratives or manipulate information at scale. Bots, in particular, flooded social media, making it hard for users to distinguish real from fake content, undermining trust in the electoral process. Also, tools like DALL-E and ChatGPT, which

contribute to training datasets, risk creating a negative feedback loop, making even less convincing AI-generated content pose risks. This could degrade model quality over time, reinforcing biases and reducing the diversity of future AI systems (Martínez et al., 2023; Angwin et al., 2024; Chafetz et al., 2024).

On the positive side, GAI enhances academic research by streamlining idea development, automating content creation, expediting data analysis, and fostering interdisciplinary collaboration, significantly improving publishing efficiency (Khalifa and Albadawy, 2024). Moreover, GAI can strengthen information ecosystems by automating content moderation and detecting manipulated media. These applications highlight the dual nature of GAI and the necessity for strict oversight and ethical guidelines to harness its potential responsibly.

#### 4.2. Large language models (LLMs)

Disinformation on social media often follows a predictable pattern. AI tools, like LLMs, generate convincing false content, which is amplified by social media algorithms prioritising engagement. Analytics then target specific demographics, boosting disinformation through likes, shares, and comments (Barman et al., 2024).

However, LLMs also play a critical role in addressing disinformation despite the amplification of false content by social media algorithms. Scalable solutions, such as the RoBERTa model, achieve up to 98% accuracy in detecting fake news, offering a promising tool for countering misinformation effectively (Wang et al., 2024). These models integrate seamlessly with systems like Facebook's DeepText and Google's Perspective API. In low-resource settings, few-shot learning frameworks like DetectYSF improve efficiency by reducing the need for large datasets. DetectYSF leverages pre-trained models and advanced techniques to achieve high accuracy with limited data, incorporating social context and misinformation patterns to improve performance, especially in politically sensitive environments (Jin et al., 2024).

Several strategies are being explored to enhance LLMs in combating misinformation. These include expanding training data, using active learning to focus on the most relevant information, and guiding models to provide more accurate responses (Zeng et al., 2024; Manfredi Sánchez & Ufarte Ruiz, 2020). A promising technique, adversarial contrastive learning, helps LLMs identify and separate truthful information from falsehoods more effectively. Meta-learning allows LLMs to adapt quickly to emerging misinformation trends, ensuring their effectiveness in real time (Chen and Shu, 2024). Additional methods, such as knowledge-augmented strategies, integrate external information to improve fact-checking, while multilingual fact-checking ensures accuracy across different languages. LLMs could also flag false information in real-time, preventing it from spreading (Vykolpal et al., 2024). However, many AI models are not optimised for regional languages like Ilocano or Cebuano (Philippines), limiting their effectiveness in rural areas.

Moreover, LLMs carry inherent biases shaped by their design and regional contexts. Western models, for instance, often prioritise individual freedom, while non-Western models emphasise state security and stability (Vecellio Segate, 2022; Buyl et al., 2024). These biases can influence political discourse, particularly during elections, where they may amplify misinformation or favour specific ideologies. The Expanded ASEAN Guide on AI Governance and Ethics showcases regional initiatives like Singapore's Moonshot Project and Vietnam's PhoGPT, which promote collaboration and culturally relevant AI tools. The Moonshot Project evaluates LLMs through benchmarking and automated red-teaming, ensuring safety and alignment with ASEAN contexts, while PhoGPT, tailored for Vietnamese language and culture, fosters innovation and addresses gaps in mainstream models (ASEAN, 2024).

#### 4.3. NLP and LSTM networks

NLP is a field of AI that enables machines to understand, interpret, and generate human language. It involves text analysis, language generation, speech recognition, machine translation, text summarization,

and question answering. NLP techniques were used to categorise TikTok posts and comments based on the presence and type of social, cultural, and political symbols, leveraging advanced models like OpenAI's GPT-4 for detailed analysis and interpretation of large datasets (Bhattacharya et al., 2024). NLP plays a critical role in improving accessibility to political information, helping combat misinformation by making complex data, such as parliamentary speeches, more comprehensible (Alcoforado et al., 2024).

Long Short-Term Memory (LSTM) networks, a type of AI adept at learning from sequential data, offer significant potential for identifying misinformation patterns. By retaining long-term dependencies through memory cells and gates, LSTM can analyse social media for anomalies like rapid content spread or spikes in activity—common indicators of disinformation campaigns. However, fairness is vital to ensure these systems do not disproportionately target specific groups (Han et al., 2024).

#### **4.4. AI detection tools and automated fact-checking**

To tackle AI-generated deepfakes, experts recommend a multi-layered approach that combines detection tools, public awareness, and legal measures. Companies like OpenAI and Microsoft have developed tools to identify synthetic media, while AI detection systems provide extra protection. Digital watermarks, which embed hidden data in AI-generated content, can be detected using advanced detection systems, like Microsoft's Video Authenticator, ensuring traceability without affecting the content's appearance (AI Team, 2024). The EU's AI Act, for example, includes requirements for providers of AI systems to mark their output as AI-generated content. Also, authenticity standards, supported by the Coalition for Content Provenance and Authenticity (C2PA), are vital in distinguishing authentic from manipulated content (Li and Callegari, 2024).

Automated fact-checking is an emerging tool in the fight against disinformation, but fully automated solutions are still being developed. One challenge in creating these systems is detecting complex truth claims, which may require more flexible categories than the rigid true/false dichotomy that current systems typically use (Kavtaradze, 2024). Recently, Meta, influenced by the X platform, ended its third-party fact-checking program, allowing user corrections instead (Isaac and Schleifer, 2025). While seen by some as a win for free speech, this shift risks fuelling misinformation. In the Philippines, with the 2025 elections nearing, it could erode trust in internet voting and amplify disinformation targeting overseas voters, underscoring the need for robust local safeguards (Pangalangan, 2025).

#### **4.5. Transparency and accountability**

To maximise the positive impact of AI, it is crucial to establish systems of transparency and accountability. These systems help ensure AI tools are applied in constructive ways, such as supporting fact-checking efforts and enhancing media literacy, while preventing their misuse in spreading misinformation (Endert, 2024). Promoting digital literacy empowers individuals to critically assess online information, helping to curb the spread of false content. Public education and modernised libraries, for instance, offering reliable information sources are crucial in developing countries. Effective information literacy relies on fostering critical thinking, ensuring access to high-quality information, and enhancing the ability to evaluate source reliability (Haque et al., 2024). In the Philippines, AI adoption is growing rapidly, particularly among knowledge workers who view it as essential for business competitiveness (Microsoft and LinkedIn, 2024). However, frequent use of AI tools negatively affects critical thinking, especially among younger users who heavily rely on AI (Gerlich, 2025). This dependence increases vulnerability to misinformation, particularly in a country facing digital literacy challenges. Addressing this requires improved training to foster critical engagement with AI and reduce cognitive dependence, preventing its political misuse.

Moreover, effective transparency requires comprehensive auditing frameworks. Governance audits ensure adherence to ethical practices, model audits evaluate performance and identify biases, and application audits track real-world usage to prevent the spread of disinformation. This multi-layered approach is vital for safeguarding the integrity of AI systems in the battle against misinformation (Mökander et al., 2024).



A regional sample focusing on youth development could indeed be critical, with California serving as a noteworthy example. The state is particularly proactive in combating misinformation, especially in the context of AI-generated content. Through legislation like AB 2839, SB 942, and AB 2013, California has introduced clear mandates aimed at enhancing transparency and accountability in digital media. These laws not only regulate the manipulation of political content but also require AI developers to provide tools that help users detect synthetic media and mandate transparency in AI training data (Pinto, 2024). The Philippines can learn from California's efforts to balance free speech with anti-disinformation initiatives in AB 2839 (Rabiu, 2024). However, successful implementation would require addressing enforcement challenges, strengthening partnerships with tech companies, and investing in digital infrastructure. Public education on the risks of misinformation is crucial for building support for such regulations.

#### **4.6. Building resilience: enhancing digital security through proactive design**

The *technological singularity* refers to a point at which AI exceeds human intelligence, potentially amplifying risks such as AI-driven cyberattacks and disinformation (Radanliev et al., 2022). While AI strengthens digital security, it also introduces vulnerabilities like data poisoning and AI-driven phishing (Vassilev et al., 2024). As AI is used to develop increasingly sophisticated malware that adapts to evade detection (Gaber et al., 2024), the risk of manipulation rises, particularly in the spread of fake news. Deepfake technology, powered by GAI, enables social engineering attacks, such as impersonating executives in phishing schemes. These activities have become so widespread that the EU has initiated legal action against Meta for failing to adequately prevent malicious actors, including a Russian influence campaign, from exploiting its platform (McMahon, 2024). A resilience-by-design approach—creating systems that can quickly recover from disruptions and ensuring their continued function—coupled with defense-in-depth strategies—implementing multiple security measures at various levels—can help mitigate these risks (Sai, 2024). However, gaps often exist between intentions and implementation due to resource and integration challenges. Addressing these requires better resource allocation, clear policies, and regular assessments (Radanliev, 2024).

The success of discussed technologies in addressing misinformation depends on their ethical application and strong oversight. The next section will explore practical recommendations for the responsible use of AI, particularly in electoral contexts.

### **5. Section 4: regulatory frameworks for safe and ethical AI usage**

This section synthesizes key insights to propose strategies for addressing AI-driven misinformation. It highlights the need for comprehensive regulatory frameworks, enhanced digital literacy, and collaboration among governments, tech companies, and civil society to tackle misinformation effectively.

#### **5.1. The evolving regulatory landscape for AI**

The regulatory landscape for AI is evolving, necessitating clear guidelines to ensure AI systems are safe, reliable, and accountable. Regulations must balance technical safety with broader societal concerns, including the risks of disinformation and AI's impact on governance and security. In the Indo-Pacific, Big Tech's influence exacerbates vulnerabilities due to limited local resources and expertise, hindering innovation and eroding sovereignty. Therefore, equitable regulation is essential to safeguard societal interests (Bak, 2024).

#### **5.2. International cooperation and diverging approaches**

International cooperation is critical to managing AI risks, as definitions of AI safety vary widely across countries. Regulations and approaches from regions such as the US, EU, Singapore, and China often serve as models in the Indo-Pacific (Dayrit et al., 2024; Dizon, 2024; Mohanty and Sahu, 2024). Southeast Asia adopts a flexible, business-friendly approach, guided by voluntary principles like the ASEAN Guide on

AI Governance and Ethics (Haie et al., 2024). Singapore provides a key example of responsible AI use, with governance frameworks and contingency planning prioritizing AI failure responses, offering valuable lessons for Southeast Asia (Soon and Quek, 2024). China's emphasis on national security and information control can shape disinformation dynamics, while stricter regulations may reduce transparency, inadvertently fostering unchecked misinformation (Guest, 2024). The EU's AI Act, while intended to protect consumers, risks stifling innovation if not carefully crafted, similar to concerns raised about the EU's broader regulatory environment that may hinder tech growth (Bradford, 2024; Graf, 2024). Key areas such as liability, privacy, intellectual property, and cybersecurity remain underdeveloped, leaving gaps that could hinder technological advancement (Novelli et al., 2024). The EU's General Data Protection Regulation (GDPR) provides a model for transparency and accountability in data processing. The U.S. focuses more on fostering innovation than on providing regulatory clarity (Guest, 2024). A proposed UN Office for the Coordination of AI could centralize efforts to foster global collaboration and responsible AI development (Fournier-Tombs and Siddiqui, 2024).

### ***5.3. AI in electoral contexts: Necessity for comprehensive regulations***

In electoral contexts, comprehensive AI regulations are essential to ensure transparency, accountability, and fairness. These regulations must address bias, establish international standards for consistency, and incorporate ongoing monitoring to preserve electoral integrity (Juneja, 2024). AI's global impact on elections requires nuanced regulations that balance its benefits with the need to protect electoral integrity (Hasan, 2024). AI tools that provide accurate, responsible information are crucial in elections.

### ***5.4. Ensuring ethical AI Use in fact-checking: the need for human oversight and transparency***

AI can assist in fact-checking by analysing language patterns to identify misleading content, but its effectiveness depends on the quality of training data and algorithmic design. Biases in data or flaws in algorithms can compromise accuracy, highlighting the need for human oversight in fact-checking (Toner-Rodgers, 2024). To mitigate AI-related risks, transparency, human oversight, and certification standards are crucial. Ultimately, human involvement ensures that AI tools are used ethically and effectively. This includes maintaining oversight, human decision-making, and preparing for AI failures through staff training and contingency planning (Cortés et al., 2023).

### ***5.5. Designing effective anti-disinformation regulations***

Additionally, regulations aimed at combating disinformation must be designed carefully to avoid misuse, particularly in politically sensitive contexts like elections. Poorly crafted laws could inadvertently suppress opposition or manipulate the democratic process (Mahapatra et al., 2024). Anti-disinformation measures must, therefore, be clear, transparent, and subject to independent oversight to safeguard credibility and prevent abuse.

By adopting these recommendations, stakeholders can build resilient information ecosystems that safeguard electoral integrity and uphold public trust in democratic processes. As AI technology continues to evolve, these strategies must adapt to protect democracy in the digital age, beyond the 2025 Philippine elections.

## **6. Conclusions and recommendations**

The challenges posed by AI-driven disinformation, particularly during elections, underscore the urgent need for a balanced approach to leveraging AI technologies. While the principles of transparency, human oversight, and robust regulatory frameworks are widely acknowledged, their practical implementation remains a critical issue. The experiences of countries like Taiwan, India, and the Philippines offer valuable insights into addressing these challenges, particularly as the Philippine midterm elections in May 2025 approach.

To safeguard democratic processes and counter the risks associated with AI-driven misinformation, a multifaceted strategy is essential:

### 1. Enhancing Digital Literacy

- Widespread educational initiatives on AI and digital literacy should be prioritised, targeting younger populations who are more susceptible to misinformation. These programmes must foster critical thinking and awareness of AI-generated content.
- **Key Insight:** Taiwan's grassroots digital literacy campaigns, which integrate public participation and rapid fact-checking, provide an effective model for empowering citizens to critically assess online information.

### 2. Developing Comprehensive Regulatory Frameworks

- Governments must collaborate with international bodies and technology companies to establish clear and enforceable AI regulations that address safety, bias, and disinformation. Such frameworks should balance innovation with societal protections.
- **Key Insight:** India's integration of AI governance into existing laws, alongside public reporting mechanisms like the DAU, highlights the importance of regulatory adaptability and inclusiveness in combating misinformation.

### 3. Promoting Cross-Sector Collaboration

- Partnerships among governments, civil society, and the private sector should be strengthened to create scalable, transparent, and accountable solutions. These collaborations must prioritise resource sharing and establish standards for AI systems.
- **Key Insight:** The Philippines' multi-stakeholder approach, including partnerships with international organisations and tech companies, demonstrates the value of collective action in combating disinformation effectively.

### 4. Strengthening Human Oversight in AI Applications

- Human decision-making must remain central to AI systems, particularly in fact-checking and disinformation detection. Training programmes for AI developers and regulators should focus on recognising and mitigating algorithmic biases.
- **Key Insight:** The Philippines faces challenges in addressing disinformation, including limited digital literacy and resource constraints. Efforts by the government, such as the proposed Deep Fake Accountability and Transparency Act, combined with initiatives by civil society to enhance media literacy, highlight the critical role of human oversight in implementing AI-driven countermeasures.

### 5. Implementing Election-Specific Countermeasures

- Measures such as media watermarking and authenticity standards for AI-generated content are critical during electoral processes to ensure transparency and maintain public trust.
- **Key Insight:** Taiwan's real-time fact-checking systems, alongside its legal amendments targeting deepfakes, underscore the need for proactive election-specific measures to protect electoral integrity.

### 6. Ensuring Ethical AI Use in Politically Sensitive Contexts

- Anti-disinformation laws must be designed with clear, transparent mechanisms for independent oversight. These frameworks should strike a balance between preventing misuse and protecting democratic freedoms.
- **Key Insight:** Lessons from China's focus on controlling information illustrate the risks of overregulation, underscoring the necessity of balanced frameworks that promote both transparency and accountability.

## 7. Looking ahead

The evolving nature of AI-driven disinformation demands continuous refinement of strategies. Developing multilingual AI tools, particularly for underrepresented languages, will be crucial in addressing

diverse contexts. Moreover, adaptive regulatory frameworks that evolve alongside technological advancements are essential to ensure resilience against emerging threats.

The Philippines' unique context, combined with lessons from Taiwan and India, highlights the need for urgent action. By fostering collaboration, prioritising ethical AI practices, and empowering citizens through digital literacy, stakeholders can navigate the complexities of AI and misinformation. These measures will help uphold the integrity of electoral processes and sustain public trust in democratic institutions in the digital age.

#### Abbreviations

AB	Assembly Bill
AI	Artificial Intelligence
EU	European Union
GAI	Generative Artificial Intelligence
LLM	Large Language Model
LSTM	Long Short-Term Memory
NLP	Natural Language Processing
RAG	Retrieval-Augmented Generation
ReAct	Reasoning and Acting
US	United States
UK	United Kingdom

**Data availability statement.** No datasets were generated or analysed during the preparation of this commentary, making data sharing inapplicable.

**Author contribution.** Dr. Tetiana Schipper conceptualised the study, conducted the analysis, and authored the entire commentary. She is solely responsible for the work's content and conclusions.

**Funding statement.** This work did not receive funding from any public, commercial, or non-profit entities.

**Competing interests.** The author declares none.

#### References

- AI Team.** (2024) *AI Pulse: Election deepfakes, Disasters, Scams & More*. Trend Micro. Available at [https://www.trendmicro.com/en\\_us/research/24/j/ai-election-deepfakes.html](https://www.trendmicro.com/en_us/research/24/j/ai-election-deepfakes.html) (accessed 8 November 2024).
- Aïmeur E, Amri S and Brassard G** (2023) Fake news, disinformation and misinformation in social media: A review. *Social Network Analysis and Mining* 13, 30. <https://doi.org/10.1007/s13278-023-01028-5> (accessed 12 November 2024).
- Alavi DM, Wählisch M and Konya A** (2022) Using artificial intelligence for peacebuilding. *Peace and Conflict Studies* 17(2), 154–168. <https://doi.org/10.1177/15423166221102757>
- Alcoforado A, Ferraz TP, Bustos E, Oliveira AS, Gerber R, Du Mont Santoro GL, Fama IC, Veloso BM, Siqueira FL, and Realí Costa AH** (2024) Augmented democracy: Artificial intelligence as a tool to fight disinformation. *Estudos Avançados* 38(111), 417–420. <https://doi.org/10.1590/s0103-4014.202438111.021> (accessed 8 November 2024).
- Amil AC** (2024) Integration of artificial intelligence (AI) in Philippine public administration: Legal and regulatory frameworks, challenges, and strategies. *International Journal of Multidisciplinary Research & Reviews* 3(3), 82–88. <https://doi.org/10.56815/IJMRR.V3I3.2024/82-88>
- Angwin J, Nelson A and Palta R** (2024). Seeking reliable election information? Don't trust AI. *Proof News*. Available at <https://www.proofnews.org/seeking-election-information-dont-trust-ai/> (accessed 12 November 2024).
- Arugay AA and Mendoza MEH** (2024) Digital Autocratisation and Electoral Disinformation in the Philippines No. 53, ISSN 2335–6677. ISEAS—Yusof Ishak Institute (accessed 8 November 2024).
- ASEAN** (2024). *Expanded ASEAN Guide on AI Governance and Ethics: Generative AI*. ASEAN. Available at <https://asean.org/book/expanded-asean-guide-on-ai-governance-and-ethics-generative-ai/> (accessed 19 January 2025).
- Bak ML** (2024). *Unmasking Big Tech's AI Policy Playbook: A Warning to Global South Policymakers*. Tech Policy Press. Available at <https://www.techpolicy.press/unmasking-big-techs-ai-policy-playbook-a-warning-to-global-south-policymakers/> (accessed 19 November 2024).
- Balita C** (2023). *Number of Social Media Users in the Philippines from 2017 to 2029*. Statista. Available at <https://www.statista.com/statistics/489180/number-of-social-network-users-in-philippines> (accessed 7 November 2024).
- Barman D, Guo Z and Conlan O** (2024) The dark side of language models: Exploring the potential of LLMs in multimedia disinformation generation and dissemination. *Machine Learning with Applications* 16, 100545. <https://doi.org/10.1016/j.mlwa.2024.100545>

- Bhakte A** (2024) Information literacy in India. *IOSR Journal of Humanities and Social Science (IOSR-JHSS)* 29(11), 09–12. <https://doi.org/10.9790/0837-2911100912>
- Bhattacharya S, Agarwal N and Poudel D** (2024) *Analyzing the Impact of Symbols in Taiwan's Anti-Disinformation Campaign on TikTok During Elections*. University of Arkansas at Little Rock. <https://doi.org/10.21203/rs.3.rs-5182951/v1>
- Bradford A** (2024) The false choice between digital regulation and innovation. *Columbia Law School Scholarship Archive* 119(2). Available at [https://scholarship.law.columbia.edu/cgi/viewcontent.cgi?params=/context/faculty\\_scholarship/article/5567/](https://scholarship.law.columbia.edu/cgi/viewcontent.cgi?params=/context/faculty_scholarship/article/5567/) (accessed 20 January 2025).
- Buyl M, Rogiers A, Noels S, Dominguez-Catena I, Heiter E, Romero R, Johary I, Mara A-C, Lijffijt J and De Bie T** (2024) Large language models reflect the ideology of their creators. <https://doi.org/10.48550/arXiv.2410.18417>
- Carr R and Köhler P** (2024) *AI-pocalypse Now? Disinformation, AI, and the Super Election Year*. Munich Security Conference. <https://doi.org/10.47342/VPRS3682> (accessed 12 November 2024).
- Chafetz H, Saxena S and Verhulst SG** (2024) A fourth wave of open data? Exploring the spectrum of scenarios for open data and generative AI. <https://doi.org/10.48550/arXiv.2405.04333> (accessed 7 November 2024).
- Chen C and Shu K** (2024) Combating misinformation in the age of LLMs: Opportunities and challenges. *AI Magazine* 45(2), 354–368. <https://doi.org/10.1002/aaai.12188>
- Chowdhury R** (2024) AI-fuelled election campaigns are here—where are the rules? *Nature* 628, 237. <https://doi.org/10.1038/d41586-024-00995-9> (accessed 12 November 2024.9)
- Chua YT and Khan RE** (2023) Countering disinformation: Tools and initiatives in the Philippines. In *Background Paper Presented at the IMS Asia Disinformation Learning Forum, 16–17 May*. International Media Support. Available at <https://www.mediasupport.org/wp-content/uploads/2023/06/Countering-disinformation-tools-and-initiatives-in-the-Philippines-May2023.pdf> (accessed 7 November 2024).
- Cortés E, Norden L, Frase H and Hoffmann M** (2023) Safeguards for using artificial intelligence in election administration. *Brennan Center for Justice*. Available at <https://www.brennancenter.org/our-work/research-reports/safeguards-using-artificial-intelligence-election-administration> (accessed 13 November 2024).
- Council of Asian Liberals and Democrats** (2024) *AI in Elections in East and Southeast Asia: Opportunities, Challenges, and Ways Forward for Democrats and Liberals*. Friedrich Naumann Foundation for Freedom. Available at <https://cald.org/wp-content/uploads/2024/12/CALD-AI-Policy-Paper.pdf>
- Council on Foreign Relations** (2024) Territorial disputes in the South China Sea. *Global Conflict Tracker*. Available at <https://www.cfr.org/global-conflict-tracker/conflict/territorial-disputes-south-china-sea>
- Dayrit M, Nalagon GB, Pajo DG, Pineda JG and Rivera JA** (2024) *Regulating Artificial Intelligence in the Philippines: Policy Paper*. University of the Philippines Los Baños, Department of Economics, College of Economics and Management. Available at [https://swisscognitive.ch/wp-content/uploads/2024/02/POLICY-DESIGN-PAPER\\_AI-GROUP.pdf](https://swisscognitive.ch/wp-content/uploads/2024/02/POLICY-DESIGN-PAPER_AI-GROUP.pdf) (accessed 13 April 2025).
- de Guzman C** (2022) Why Bongbong Marcos, a Philippine dictator's son, leads the race for the presidency. *Time*. Available at <https://time.com/6162028/bongbong-marcos-philippines-president-popular/> (accessed 7 November 2024).
- Digital Policy Alert** (2024) *Introduced: Deepfake Accountability and Transparency Act (Bill 10567)*. Digital Policy Alert. Available at <https://digitalpolicyalert.org/event/22040-introduced-deepfake-accountability-and-transparency-act-bill-10567> (accessed 17 January 2025).
- Dizon D** (2024) Philippines working with AI providers to root out deepfakes. *ABS-CBN News*. Available at <https://news.abs-cbn.com/news/2024/10/16/philippines-working-with-ai-providers-to-root-out-deepfakes-1030> (accessed 19 November 2024).
- Dulay DC, Hicken A, Menon A and Holmes R** (2023) Continuity, history, and identity: Why Bongbong Marcos won the 2022 Philippine presidential elections. *Pacific Affairs* 96(1), 85–104. <https://doi.org/10.5509/202396185> (accessed 8 November 2024).
- Edelman** (2024) *Edelman Trust Barometer 2024: A Collision of Trust, Innovation, and Politics*. Edelman. Available at <https://www.edelman.com/trust/2024/trust-barometer>
- Enderit J** (2024) Generative AI is the ultimate disinformation amplifier. *DW Akademie*. Available at <https://akademie.dw.com/en/generative-ai-is-the-ultimate-disinformation-amplifier/a-68593890> (accessed 12 November 2024).
- Enriquez JM** (2024) Tempering the Philippines' AI disinformation storm. *East Asia Forum*. <https://doi.org/10.59425/eabc.1729116000> (accessed 7 November 2024).
- EON The Stakeholders Relations Group and Ateneo de Manila University** (2024) The accountability revolution: Filipinos demand proof before trust. *Philippine Daily Inquirer*. Available at <https://business.inquirer.net/488422/the-accountability-revolution-filipinos-demand-proof-before-trust> (accessed 19 January 2025).
- Fallorina R, Lanuza JMH, Felix JG, Sanchez FH, Ong JC and Curato N** (2023) From disinformation to influence operations: The evolution of disinformation in three electoral cycles. *Internews*. Available at <https://internews.org/resource/from-disinformation-to-influence-operations-the-evolution-of-disinformation-in-three-electoral-cycles/> (accessed 7 November 2024).
- Fournier-Tombs E and Siddiqui M** (2024) Wie kann künstliche Intelligenz global gesteuert werden? *Zeitschrift Vereinte Nationen* 72(5), 195–201. <https://doi.org/10.35998/VN-2024-0021>
- Gaber MG, Ahmed M and Janicke H** (2024) Malware detection with artificial intelligence: A systematic literature review. *ACM Computing Surveys* 56(6), 1–33. <https://doi.org/10.1145/3638552>
- Gerlich M** (2025) AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies* 15(1), 6. <https://doi.org/10.3390/soc15010006>



- Graf J-P** (2024) Two models of regulation: Artificial-intelligence compliance in the United States and the European Union. *Compliance Elliance Journal* 10(2), 3–21. Available at <https://cej-online.com/2024-vol-10-no-02-compliance-in-trade-and-information-technology/#6> (accessed 12 November 2024).
- Guest O** (2024) *Chinese AI Safety Institute Counterparts*. Institute for AI Policy and Strategy. Available at <https://www.iaps.org/chinese-ai-safety-institute-counterparts> (accessed 12 November 2024).
- Haie A-G, Chitrakroth N, Avramidou M and Lamsam R** (2024) Comparing EU, Southeast Asia approaches to AI regulation. *Tilleke & Gibbins*. Available at <https://www.tilleke.com/insights/comparing-eu-southeast-asia-approaches-to-ai-regulation/> (accessed 20 November 2024).
- Han Y, Lam JCK, Li YOK, Newbery D, Guo P and Chan K** (2024) A deep learning approach for fairness-based time of use tariff design. *Energy Policy* 192, 114230. <https://doi.org/10.1016/j.enpol.2024.114230>. Available at <https://www.sciencedirect.com/science/article/pii/S0301421524002507> (accessed 12 November 2024).
- Hapal DK** (2024) The Philippines disinformation machine. *New Internationalist*. Available at <https://ul.qucosa.de/api/qucosa%3A94303/attachment/ATT-0> (accessed 12 November 2024).
- Haque R, Senathirajah ARbin S, Qazi SZ, Afrin N, Ahmed MN and Khalil MI** (2024) Factors of information literacy preventing fake news: A case study of libraries in developing countries. *International Journal of Religion* 5(7), 804–817. <https://doi.org/10.61707/vqbfj15>
- Hasan S** (2024) *The Effect of AI on Elections Around the world and what to do About It*. Brennan Center for Justice. Available at <https://www.brennancenter.org/our-work/analysis-opinion/effect-ai-elections-around-world-and-what-do-about-it> (accessed 19 November 2024).
- Heikkilä M** (2024) *AI-generated Content Doesn't Seem to Have Swayed Recent European Elections*. MIT Technology Review. Available at <https://www.technologyreview.com/2024/09/18/1104178/ai-generated-content-doesnt-seem-to-have-swayed-recent-european-elections/> (accessed 10 May 2025).
- Hung C-L, Fu W-C, Liu C-D and Tsai H-J** (2024) AI disinformation attacks and Taiwan's responses during the 2024 presidential election. *Graduate Institute of Journalism, National Taiwan University*. Available at [https://www.thomsonfoundation.org/media/268943/ai\\_disinformation\\_attacks\\_taiwan.pdf](https://www.thomsonfoundation.org/media/268943/ai_disinformation_attacks_taiwan.pdf) (accessed 8 November 2024).
- Isaac M and Schleifer T** (2025) Meta says it will end its fact-checking program on social media posts. *The New York Times*. Available at <https://www.nytimes.com/live/2025/01/07/business/meta-fact-checking> (accessed 17 January 2025).
- Jin W, Wang N, Tao T, Shi B, Bi H, Zhao B, Wu H, Duan H and Yang G** (2024) A veracity dissemination consistency-based few-shot fake news detection framework by synergizing adversarial and contrastive self-supervised learning. *Scientific Reports* 14(1), 19470. <https://doi.org/10.1038/s41598-024-70039-9>
- Juneja P** (2024) *Artificial intelligence for electoral management*. International IDEA. Available at <https://www.idea.int/sites/default/files/2024-04/artificial-intelligence-for-electoral-management.pdf> (accessed 13 November 2024).
- Kavtaradze L** (2024) Challenges of automating fact-checking: A technographic case study. *Emerging Media* 2(2), 236–258. <https://doi.org/10.1177/27523543241280195>
- Khalifa M and Albadawy M** (2024) Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update* 5, 100145. <https://doi.org/10.1016/j.cmpbup.2024.100145>
- Kuo S-S** (2021) *Taiwan's Response to Disinformation: A model for Coordination to Counter a Complicated Threat*. National Bureau of Asian Research. Available at [https://www.nbr.org/wp-content/uploads/pdfs/publications/sr93\\_taiwan\\_sep2021.pdf](https://www.nbr.org/wp-content/uploads/pdfs/publications/sr93_taiwan_sep2021.pdf) (accessed 15 January 2025).
- Leingang R** (2024) X's AI chatbot spread voter misinformation—and election officials fought back. *The Guardian*. Available at <https://www.theguardian.com/technology/2024/sep/12/x-ai-chatbot-election-misinformation> (accessed 12 November 2024).
- Li C and Callegari A** (2024) *Stopping AI Disinformation: Protecting Truth in the Digital World*. World Economic Forum. Available at <https://www.weforum.org/stories/2024/06/ai-combat-online-misinformation-disinformation/> (accessed 8 November 2024).
- Li B and Gilbert S** (2024) Artificial intelligence awarded two Nobel Prizes for innovations that will shape the future of medicine. *npj Digital Medicine* 7, 336. <https://doi.org/10.1038/s41746-024-01345-9>
- Mahapatra S, Sombatpoonsiri J and Ufen A** (2024) *Repression by Legal Means: Governments' Anti-Fake News Lawfare*. GIGA Focus Global, 1. German Institute for Global and Area Studies (GIGA). <https://doi.org/10.57671/gfgl-24012>
- Manfredi Sánchez JL and Ufarte Ruiz MJ** (2020) Inteligencia artificial y periodismo: Una herramienta contra la desinformación. *Revista CIDOB d'Afers Internacionals* 124, 49–72. <https://doi.org/10.24241/rcai.2020.124.1.49> (accessed 8 November 2024).
- Marcelino KN** (2023) Machine-made deception: Dangers of AI-powered disinformation in Philippine social media. EngageMedia. Available at <https://engagemedia.org/2023/youth-philippines-disinformation/> (accessed 7 November 2024).
- Martínez G, Hernández JA, Watson L, Juárez M, Reviriego P and Sarkar R** (2023) Towards understanding the interplay of generative artificial intelligence and the internet. Available at <https://arxiv.org/pdf/2306.0613> (accessed 12 November 2024).
- McMahon L** (2024) Meta faces EU probe over Russian disinformation. *BBC News*. <https://www.bbc.com/news/articles/c72p1dr0mk8o> (accessed 17 January 2025).
- Microsoft & LinkedIn** (2024) Work Trend Index Annual Report: AI at Work Is Here. Now Comes the Hard Part. Available at [https://assets-c4akfrf5b4d3f4b7.z01.azurefile.net/assets/2024/05/2024\\_Work\\_Trend\\_Index\\_Annual\\_Report\\_6\\_7\\_24\\_666b2e2fafceb.pdf](https://assets-c4akfrf5b4d3f4b7.z01.azurefile.net/assets/2024/05/2024_Work_Trend_Index_Annual_Report_6_7_24_666b2e2fafceb.pdf) (accessed 17 January 2025).

- Mohanty A and Sahu S** (2024) *India's Advance on AI Regulation*. Carnegie Endowment for International Peace. Available at <https://carnegieendowment.org/research/2024/11/indias-advance-on-ai-regulation?lang=en> (accessed 15 January 2025).
- Mökander J, Schuett J, Kirk HR and Floridi L** (2024) Auditing large language models: A three-layered approach. *AI and Ethics* 4(4), 1085–1115. <https://doi.org/10.1007/s43681-023-00289-2>
- Nannaware SC, Pillai R and Kate N** (2025) Deepfakes in action: Exploring use cases across industries. In Gupta G, Bohara S, Kovid RK and Pandla K (eds), *Deepfakes and their Impact on Business*. Hershey, PA: IGI Global, pp. 71–98. <https://doi.org/10.4018/979-8-3693-6890-9.ch004>
- Neuwirth RJ** (2021) The global regulation of “fake news” in the time of oxymora: Facts and fictions about the Covid-19 pandemic as coincidences or predictive programming? *International Journal for the Semiotics of Law—Revue Internationale de Sémiotique Juridique* 35(3), 831–857. <https://doi.org/10.1007/s11196-021-09840-y>
- Novelli C, Casolari F, Hacker P, Spedicato G and Floridi L** (2024) Generative AI in EU law: Liability, privacy, intellectual property, and cybersecurity. *Computer Law & Security Review* 55, 106066. <https://doi.org/10.1016/j.clsr.2024.106066>
- Pangalangan PA** (2025) Meta's move: Risks if Asia is left unchecked. *Philippine Daily Inquirer*. Available at <https://opinion.inquirer.net/180078/metass-move-risks-if-asia-is-left-unchecked> (accessed 20 January 2025).
- Pinto T** (2024) *California's New AI Laws: What They Mean for Developers*. Michalsons. Available at <https://www.michalsons.com/blog/californias-new-ai-laws-what-they-mean-for-developers/76026> (accessed 17 January 2025).
- Rabiu M** (2024) *First Amendment Roadblock? Regulating the Misuse of Generative AI Technologies: Impersonation and Appropriation of Likeness without Permission*. Old Dominion University. Available at <https://digitalcommons.odu.edu/cgi/viewcontent.cgi?article=1110&context=covacci-undergraduateresearch> (accessed 17 January 2025).
- Radanliev P** (2024) Digital security by design. *Security Journal* 37, 1640–1679. <https://doi.org/10.1057/s41284-024-00435-3>
- Radanliev P, De Roure D, Maple C and Ani, U** (2022) Super-forecasting the ‘technological singularity’ risks from artificial intelligence. *Evolving Systems* 13, 747–757. <https://doi.org/10.1007/s12530-022-09431-7>
- Raina P** (2024) *Year of Elections: Lessons from India's Fight Against AI-Generated Misinformation*. World Economic Forum. Available at <https://www.weforum.org/agenda/2024/08/india-fight-ai-generated-misinformation> (accessed 12 November 2024).
- Romero XB and Fuelles GKJ** (2024) *Equity and Accessibility: The National Library of the Philippines' Gateway to Virtual Reference and Information Services*. IFLA. Available at <https://www.ifla.org/news/equity-and-accessibility-the-national-library-of-the-philippines-gateway-to-virtual-reference-and-information-services/> (accessed 15 January 2024).
- Roseman R** (2024) Silicon valley calls it deepfake technology. Let's call it what it is: Algorithmic sexual violence. *Simple Machine*. Available at <https://rachelroseman.substack.com/p/vol-4-when-is-a-woman-just-a-woman> (accessed 12 November 2024).
- Sadiq S and Demartini G** (2024) *How AI is being Used to Fight Fake News*. University of Queensland. Available at <https://sponsored.chronicle.com/how-ai-is-being-used-to-fight-fake-news/index.html> (accessed 12 November 2024).
- Sai K** (2024) *Sustaining Digital Resilience with Secure by Design*. SolarWinds. Available at <https://orangematter.solarwinds.com/2024/08/08/sustaining-digital-resilience-with-secure-by-design>
- Schipper Tetiana** (2024) *Philippine President Embraces Global Alliances to Push Digitalisation*, *GIGA Focus Asia*, Hamburg: German Institute for Global and Area Studies (GIGA). <https://doi.org/10.57671/gfas-24072>
- Soon C and Quek S** (2024) *Safeguarding Elections from Threats Posed by Artificial Intelligence*. Institute of Policy Studies, Lee Kuan Yew School of Public Policy. Available at [https://lkyspp.nus.edu.sg/docs/default-source/ips/ips-working-paper-no-56\\_safeguarding-elections-from-threats-posed-by-artificial-intelligence.pdf](https://lkyspp.nus.edu.sg/docs/default-source/ips/ips-working-paper-no-56_safeguarding-elections-from-threats-posed-by-artificial-intelligence.pdf) (accessed 13 November 2024).
- Stockwell S** (2024) *AI-Enabled Influence Operations: Threat Analysis of the 2024 UK and European Elections*. Centre for Emerging Technology and Security. Available at <https://cetas.turing.ac.uk/publications/ai-enabled-influence-operations-threat-analysis-2024-uk-and-european-elections>
- Telenor Asia** (2023) *Digital Lives Decoded: Play [Report]*. Telenor. Available at <https://www.telenor.com/stories/include/telenor-asia-digital-lives-decoded-play/> (accessed 7 November 2024).
- Tigno JV, Ducanes GM, Rood S and Licudine VJA** (2024) They never left: Drivers of memory of dictatorship and impressions of Ferdinand E. Marcos as president after February 1986. *Journal of Current Southeast Asian Affairs* 43(1), 1–22. <https://doi.org/10.1177/18681034241248763>
- Toner-Rodgers A** (2024) Artificial intelligence, scientific discovery, and product innovation. *General Economics*, arXiv. <https://doi.org/10.48550/arXiv.2412.17866>
- Vassilev A, Oprea A, Fordyce A and Anderson H** (2024) *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations (NIST AI 100-2e2023)*. Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology. Robust Intelligence, Inc. <https://doi.org/10.6028/NIST.AI.100-2e2023>
- Vecellio Segate R** (2022) Horizontalizing insecurity or securitizing privacy? Two narratives of a rule-of-law misalignment between a special administrative region and its state. *The Chinese Journal of Comparative Law* 10(1), 56–89. <https://doi.org/10.1093/cjcl/cxac002>
- Vosoughi S, Roy D and Aral S** (2018) The spread of true and false news online. *Science* 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- Vykopal I, Pikuliak M, Ostermann S and Šimko M** (2024) Generative large language models in automated fact-checking: A survey. <https://doi.org/10.48550/arXiv.2407.02351>

- Wang X, Zhang Y, Liu Z and Zhang C** (2024) Fake news detection with BERT-based models: An effective solution for social media misinformation. *Journal of Artificial Intelligence Research* 45(3), 221–240. <https://doi.org/10.1016/j.jair.2024.04.005>
- Wylie C** (2019) Cambridge Analytica exploited the Philippines’ weak regulations to test online propaganda, whistle-blower reveals. *Rappler*. Available at <https://www.rappler.com/technology/social-media/239606-cambridge-analytica-philippines-online-propaganda-christopher-wylie/> (accessed 15 November 2024).
- Zeng J, Huang R, Malik W, Yin L, Babic B, Shacham D, Yan X, Yang J and He Q** (2024) Large language models for social networks: Applications, challenges, and solutions. <https://doi.org/10.48550/arXiv.2401.02575> (accessed 7 November 2024).