IPS Working Papers

SAFEGUARDING ELECTIONS FROM THREATS POSED BY ARTIFICIAL INTELLIGENCE

CAROL SOON
and
SAMANTHA QUEK

August 2024
IPS Working Papers No. 56

Institute of Policy Studies

About Institute of Policy Studies (IPS)

The Institute of Policy Studies (IPS) was established in 1988 to promote a greater awareness of policy issues and good governance. Today, IPS is a think-tank within the Lee Kuan Yew School of Public Policy (LKYSPP) at the National University of Singapore. It seeks to cultivate clarity of thought, forward thinking and a big-picture perspective on issues of critical national interest through strategic deliberation and research. It adopts a multi-disciplinary approach in its analysis and takes the long-term view. It studies the attitudes and aspirations of Singaporeans which have an impact on policy development and the relevant areas of diplomacy and international affairs. The Institute bridges and engages the diverse stakeholders through its conferences and seminars, closed-door discussions, publications, and surveys on public perceptions of policy.

IPS Working Papers No. 56

SAFEGUARDING ELECTIONS FROM THREATS POSED BY ARTIFICIAL INTELLIGENCE

CAROL SOON

Principal Research Fellow

Head, Society & Culture

Institute of Policy Studies

carol.soon@nus.edu.sg

and

SAMANTHA QUEK

Research Assistant

Society & Culture

Institute of Policy Studies

August 2024

CONTENTS

SAFEGUARDING ELECTIONS FROM THE THREATS POSED BY ARTIFICIAL INTELLIGENCE

Executive Summary

Artificial intelligence (AI) presents a novel challenge to a record-breaking year of elections, with over 20 national elections held since the start of 2024. Much of the public discourse centres on AI's ability to produce and disseminate mis- and disinformation, which has already impacted some elections. For instance, the deepfake of late Indonesian President Suharto endorsing Golkar candidates circulated during the lead-up in the country's 2024 general election.

The use of AI during elections is beneficial for increasing the efficiency of political campaigning processes, improving information parity and detecting AI-generated and AI-manipulated information. However, the technology has its pitfalls too. Its generative functionality has caused it to become a powerful manipulation tool. When used exploitatively to generate and manipulate content, the technology poses a significant threat to the democratic discourse and electoral integrity.

Tech companies are rolling out regulations to safeguard against the pitfalls of AI use in elections. For instance, 20 tech companies signed an accord to combat the deceptive use of AI in elections happening around the world in 2024. Individually, tech companies have also rolled out policies ahead of the upcoming elections.

Governments are responding to the potential threats of generative AI in different ways, and they include South Korea, Brazil and some states in the US. South Korea amended its Public Official Election Act, banning the use of AI-generated deepfakes in political campaigns for 90 days leading up to election day, with exceptions for deepfakes that promote voter participation and intraparty activities (e.g., AI-generated campaign mottos, song lyrics or speeches). Brazil's Superior Electoral Court has also approved 12 resolutions that will guide the upcoming elections. One of the resolutions providing for electoral propaganda prohibit the use of deepfakes, restricts the use of chatbots and avatars to simulate dialogue between a candidate and a real person and

requires disclosure for AI-generated and AI-manipulated content. In the US, several states have implemented bills that safeguard against the deceptive use of AI in elections.

Other efforts to address the threats of AI use in elections include the European Union's Artificial Intelligence Act (EU AIA) and the UK's response paper to its white paper on AI. Under the EU AIA, AI systems that interact with natural persons or generate content posing risks of impersonation or deception in the context of democratic processes, civic discourse and electoral processes, are subject to the same transparency obligations as high-risk AI systems. The UK response paper to the white paper on AI briefly discusses what responsible use of generative AI in elections entails. Other approaches undertaken by governments and intergovernmental institutions include voluntary code of conducts, frameworks and principles.

In Singapore, there have been efforts to promote responsible AI use. They include the continuous refinement of AI guidelines, increasing funding support for AI research, development of an AI evaluation toolkit and governance framework. The Personal Data Protection Commission (PDPC) has issued advisory guidelines that provide guidance to organisations on the use of personal data when developing and deploying AI systems. Singapore also funds AI research initiatives like the National Multimodal Large Language Model Programme (NMLP) for Singapore and the Southeast Asia region, and an LLM evaluation toolkit for companies to evaluate the safety of their generative AI technologies. Finally, Singapore recently introduced two governance frameworks, the Model AI Governance Framework for Generative AI aimed at fostering a trustworthy AI ecosystem, and the AI Governance Playbook for Digital Forum of Small States to guide small states in responsible development and adoption of AI technologies.

Currently, Singapore has laws and advisories that target three key types of threats, namely, mis- and disinformation, political advertising that influences election outcomes, and foreign interference. For Singapore to respond to the evolving tactics of malicious actors, it can consider expanding some of its current laws. Since the Singapore government is amid considering how to regulate deepfake content and the possibility of a temporary ban during election time, this paper recommends that the

government takes into consideration the definition of what constitutes a deepfake and of what is permissible and not permissible (i.e., type of content, purpose and intent). Singapore will also gain from greater collaboration between its government and tech companies, considering the role that tech companies play in elections. At present, Singapore's Centre for Advanced Technologies in Online Safety (CATOS) is working with Adobe to develop content provenance technologies. During the recent Presidential Election 2023, TikTok users were directed to authoritative information from the Elections Department Singapore (ELD). Moving forward, other potential collaborations include the creation of in-app election centres that push out promptly curated election-related information and AI chatbots that direct users to ELD's official website when they pose election-related queries.

With the possibility that Singapore may implement AI tools in the upcoming elections, it is necessary for election officials to evaluate AI tools used in electoral processes, ensuring that these tools promote quality, privacy and security. Additionally, the government should step up dialogues and engagement with members of the public regarding challenges of AI in the context of elections.

Overall, this working paper highlights the rapid advancements in AI and the new set of challenges this brings for governments around the world. With the elections for Singapore to be held no later than 23 November 2025, it is paramount for Singapore to tackle the threats posed by AI to election integrity.

SAFEGUARDING ELECTIONS FROM THREATS POSED BY ARTIFICIAL INTELLIGENCE

1. BACKGROUND

Since the start of 2024, elections have been held in over 20 countries. They include Bangladesh, India, Indonesia and Russia. Voters from the UK and US will be heading to the polls in the coming months and Singapore must hold its next general election no later than 23 November 2025. In total, approximately three billion voters will go to the ballot box in 2024 and 2025 (World Economic Forum, 2024). Instances where generative Artificial Intelligence (AI) were used in elections have attracted much attention and generated concerns over their impact on elections and societies. For instance, a three-minute deepfake of late Indonesian president Suharto that gained over 4.7 million views on X — in which he reminded voters about the importance of their votes for the upcoming Indonesian election — was thought to influence public support for the Golkar party (Chen, 2024). In the lead-up to the Indian election, fake videos of Bollywood actors criticising Prime Minister Narendra Modi, and of his top aides making false claims, fuelled concerns over the impact of disinformation (Channel News Asia, 2024).

This working paper builds on IPS Working Papers No. 52 which reviewed developments in the governance of AI in different jurisdictions and put forth recommendations for Singapore to strike the balance between innovation and risk mitigation (Soon & Tan, 2023). In this current working paper, we focus on the potential impact of AI, in particular, generative AI on elections. We review the potential impact of the technology on elections; what stakeholders like technology companies and regulators in different jurisdictions are doing to counter the threats posed by the technology; and the laws in Singapore that seek to protect election integrity. In so doing, we propose what needs to be done to safeguard elections in Singapore from the threats of AI.

2. THE BENEFITS AND PITFALLS OF AI DURING ELECTIONS

The use of AI during elections holds several benefits, such as increasing the efficiency of the political campaigning process and freeing up labour for more high-touch campaigning activities. For example, ChatGPT has been used to create campaigning speeches, marketing materials, fundraising emails and texts (Curry, 2023). AI tools can also be used to perform operational tasks like scheduling and budgeting (Bagri, 2023). Additionally, AI tools allow election officers to administer elections more efficiently by optimising resource allocation at voting stations (Brand, 2024). Given their lower barriers to adoption and use, AI tools can also potentially level the playing field between well-resourced and less-resourced parties and candidates. Candidates who have fewer financial resources and smaller manpower can use AI tools to perform a greater variety of campaigning tasks, like disseminating large volumes of targeted ads to expand their reach (LaChapelle & Tucker, 2023).

Second, AI tools can potentially increase information parity among voters when they are used to reproduce election-related information in native languages. Voters can use AI tools to translate important information on the election and political campaigns into their native languages (Wirtschafter, 2024). This helps to narrow information gaps between voters from different provinces and states in linguistically diverse countries. One example from a non-election context is Prime Minister Narendra Modi tapping Bhashini AI's speech-to-speech machine translation and automatic speech recognition features to communicate with students from different parts of India (Suraksha P, 2023).

Third, AI tools play a role in detecting AI-generated and AI-manipulated election-related information (e.g., Deep Media, AI Speech Classifier, Deepware scanner and InVID). For example, Intel's FakeCatcher software detects changes in blood flow that is absent in faces created by deepfakes; it also analyses eye movements to verify the authenticity of the person in the video (Clayton, 2023). AI detection tools can also help fact-checkers determine with greater certainty whether a content is a deepfake or not, thereby reducing the chances of journalists reporting misinformation (Adami, 2024).

However, generative AI poses significant threats to elections in several ways. First, while chatbots lower operational costs for candidates, they contribute to the problem

of AI hallucinations through the fabrication of information, thereby exacerbating the problem of misinformation (Alkaissi & McFarlane, 2023; Giansiracusa & Panditharatne, 2023; Vadde, 2024). For instance, when AI Democracy Projects posed queries that voters might ask during an election to five chatbots (i.e., Anthropic's Claude, Google's Gemini, OpenAI's ChatGPT-4, Meta's Llama 2, and Mistral's Mixtral), it found that about half of the answers returned by the chatbots were inaccurate (Angwin et al., 2024). Second, given the low cost and ease of use of generative AI tools, malicious actors can now produce and disseminate disinformation at scale (Mirza, 2024; Robins-Early, 2023). For instance, an AI-generated robocall message imitating Joe Biden reached thousands of voters within two days just before the New Hampshire presidential primary. The robocall reportedly cost Stever Kramer (a veteran political consultant working for a rival candidate) only USD150 and could have discouraged voter turnout (Ramer, 2024; Seitz-Wald, 2024). Hallucinations created by generative AI and the dissemination of disinformation at scale negatively affect people's access to trustworthy information and make it more difficult for them to discern facts and false information.

In addition, the proliferation of deepfakes makes it easier for political candidates to falsely claim genuine content to be manipulated or generated by AI (Goldstein & Lohn, 2024). This enables political candidates to avoid being held accountable for their speech or actions that might cause embarrassment or affect their campaigns, hence benefitting from the "liar's dividend" (Chesney & Citron, 2019; Goldstein & Lohn, 2024) and profiting from a polluted information ecosystem. For example, during the recent Turkish election, a video that showed compromising images of an electoral candidate was said to be a deepfake when it was in fact real (Mirza, 2024; Sparrow & Ünker, 2023). Politicians' lies and evasion of accountability heightens citizens' distrust of the candidates and integrity of the election.

Fourth, generative AI tools may also reinforce certain values and ideologies among users, given that current approaches used to develop AI chatbots imbue them with political biases. For example, several studies have found that ChatGPT leans towards libertarianism (Hartmann et al., 2023; Rozado, 2023; Sullivan-Paul, 2023). One possible source of bias lies with the training data (Baum & Villasenor, 2023; Mowshowitz, 2024). ChatGPT, like most LLMs, is trained on datasets gathered from

the internet that originate from influential institutions in the West, like mainstream news media outlets, prestigious universities and social media platforms (Rozado, 2023). A significant number of individuals working in these influential institutions are politically left-leaning, which might account for ChatGPT's left-leaning bias (Abrams & Khalid, 2020; Cornia et al., 2016; Skelding, 2021; Weaver et al., 2019). Another bias, and perhaps more consequential, lies in the reinforcement learning with human feedback (RLHF) that ChatGPT undergoes. During RLHF, human labellers who are not representative of the general population and have their own political leanings skew chatbots towards certain biases (Baum & Villasenor, 2023). In a podcast, OpenAI's Sam Altman said, "The bias I'm most nervous about is the bias of the human feedback raters. … We're great at the pre-training machinery. We're now trying to figure out how we're going to select those people." (Fridman, 2023). This has implications for elections since individuals turn to AI chatbots to seek information when making voting decisions.

3. TECH COMPANIES APPROACH TO COMBAT DECEPTIVE USE OF AI IN ELECTIONS

As social media platforms provide verdant ground for the dissemination of misinformation and disinformation, tech companies have adopted myriad approaches to mitigate the negative effects of AI on elections. For one, a number of tech companies are collaborating via the Tech Accord to Combat Deceptive Use of AI in 2024 Elections (BBC, 2024). Launched in February 2024, the accord articulates a set of commitments by signatories to counter harmful AI-generated content that might deceive voters (AI Elections Accord, 2024a). It was signed by 20 tech companies including Amazon, Google, IBM, LinkedIn, McAfee, Meta, Microsoft, OpenAI and TikTok[1] (Microsoft, 2024). The accord is a voluntary framework that consists of seven principle goals, namely: prevention, provenance, detection, responsive protection, evaluation, public awareness and resilience. See Table 1 for the definitions of the principle goals.

[1] The 20 tech companies that signed the Tech Accord on 16 February 2024 are Adobe, Amazon, Anthropic, Arm, ElevenLabs, Google, IBM, Inflection AI, LinkedIn, McAfee, Meta, Microsoft, Nota, OpenAI, Snap Inc., Stability AI, TikTok, Trend Micro, Truepic, and X (AI Elections Accord, 2024a).

Table 1: Definition of the seven principle goals

Prevention	Researching, investing in, and/or deploying reasonable precautions to limit risks of deliberately deceptive AI election content being generated.
Provenance	Attaching provenance signals to identify the origin of content where appropriate and technically feasible.
Detection	Attempting to detect deceptive AI election content or authenticated content, including with methods such as reading provenance signals across platforms.
Responsive Protection	Providing swift and proportionate responses to incidents involving the creation and dissemination of deceptive AI election content.
Evaluation	Undertaking collective efforts to evaluate and learn from the experiences and outcomes of dealing with Deceptive AI Election Content.
Public Awareness	Engaging in shared efforts to educate the public about media literacy best practices, in particular regarding Deceptive AI Content, and ways citizens can protect themselves from being manipulated or deceived by this content.
Resilience	Supporting efforts to develop and make available defensive tools and resources, such as AI literacy and other public programs, AI-based solutions (including open-source tools where appropriate), or contextual features, to help protect public debate, defend the integrity of the democratic process, and build whole-of-society resilience against the use of deceptive AI election content.

Source: AI Elections Accord (2024b)

To achieve the above mentioned seven principle goals, tech companies have committed to the following eight actions (AI Elections Accord, 2024a):

1. Developing and implementing technology to mitigate risks related to deceptive AI election content, including open-source tools where appropriate
2. Assessing models in scope of this accord to understand the risks they may present regarding deceptive AI election content
3. Seeking to detect the distribution of this content on their platforms
4. Seeking to appropriately address this content detected on their platforms
5. Fostering cross-industry resilience to deceptive AI election content
6. Providing transparency to the public regarding how the company addresses it
7. Continuing to engage with a diverse set of global civil society organisations, academics
8. Supporting efforts to foster public awareness, media literacy and all-of-society resilience

Table 2 presents the measures adopted by tech companies and developers of social media, AI chatbots and image generators against the seven principle goals of the tech accord.

Table 2: Comparison of regulatory efforts taken by tech companies

Seven Principle Goals	**Tool**	Tik Tok	Meta			X	Google		Microsoft		OpenAI	
			Instagram	Facebook	Threads		YouTube	Google Gemini	Microsoft Copilot	Microsoft Designer	Dall-E 3	ChatGPT
Prevention	Provide trusted and authoritative information	✓	✓	✓			✓	✓	✓			✓
Provenance	Label AI-generated content	✓	✓	✓	✓		✓					
	Content Credentials								✓	✓	✓	✓
Detection	Partnership with fact-checking organisations/ External evaluators	✓	✓	✓	✓	✓	✓					
	Self-regulation/community regulation					✓	✓				✓	
Responsive Protection	Contents/accounts to face some restrictions	✓	✓	✓	✓		✓					
Evaluation	Setting up of a committee										✓	
Public Awareness	AI media literacy campaigns	✓	✓	✓	✓		✓		✓			
Resilience	Sharing research and innovation		✓	✓	✓							
	Setting up a resilience fund								✓		✓	

To achieve the principle goal of *prevention*, some tech companies direct users to authoritative and trusted information. For instance, TikTok and Meta have in-app features that point users to trusted information sources. TikTok partners with electoral commissions to provide in-app "Election Centres" that are information pitstops for users to access for trustworthy election-related content (Loftus, 2024; Morgan, 2024; TikTok, 2024a; TikTok, 2024b). In the current lead-up to the US elections, Facebook users receive in-app notifications that provide them with details on voter registration and elections and have access to a Voting Information Center; while Instagram has stickers that point users to authoritative voting information (Meta, n.d.-a). YouTube has been training its algorithm to recommend "authoritative" elections content and has launched features that signpost users to authoritative election-related news (e.g., its "watch page" and a panel under the search bar direct users to authoritative sources) (YouTube, n.d.-a).

Instead of replying to elections-related queries, some AI chatbots direct users to reliable sources for information. Two examples are Google's Gemini and Microsoft's Copilot, which are restricting their responses for electoral-related queries and directing users to the Google search engine and Bing search results, respectively (O'Donovan & Zakrzewski, 2024; Robins-Early, 2024). OpenAI's ChatGPT is gradually integrating attribution and links in their responses to users and efforts are made to improve users' access to authoritative voting information. For instance, ChatGPT partners the National Association of Secretaries of State (NASS) to direct users to CanIVote.Org when they pose questions relating to US election procedures (OpenAI, 2024a).

Second, tech companies like TikTok, Meta, YouTube, OpenAI and Microsoft have introduced measures to clarify the *provenance* of content. TikTok, Meta and YouTube require users to label AI-generated and AI-manipulated content (Meta, 2024a; TikTok, 2024c; YouTube, 2024). Developers and tech companies like OpenAI, TikTok, Meta and YouTube have incorporated or are planning to incorporate Content Credentials in their tools and platforms (The Business Times, 2024; TikTok, 2024d). Developed by the Coalition for Content Provenance and Authenticity (C2PA), Content Credentials are tamper-evident meta data that are cryptographically attached to digital content; they provide verifiable trails of content's origins and edits (Adobe, 2023).

For example, images generated with ChatGPT and DALL-E 3 contain metadata that signals if the content is AI-generated (OpenAI, 2024b). Microsoft has also introduced Content Credentials to content generated by Microsoft's Copilot and Microsoft Designer (Microsoft Create team, 2023; Parsons, 2023). Additionally, Microsoft is making Content Credentials available to politicians and electoral officials through its Content Integrity tools (Hutson, 2024a). Labelling their content enables politicians and electoral officials to signpost the trustworthiness and authoritativeness of their content. Meta is currently in the process of building a tool that identifies invisible markers at scale so that AI-generated images from Google, OpenAI, Microsoft, Adobe, Midjourney and Shutterstock will be automatically labelled as these companies work on incorporating metadata to images created by their tools (Clegg, 2024).

Third, TikTok, Meta, X and YouTube have developed policies to assist with *detection*. TikTok and Meta partner fact-checking organisations while X adopts a community approach to checking published content (Clegg, 2024; TikTok, n.d.-a). TikTok has also been advancing its detection methods to keep pace with the evolution of AI. Besides partnering experts and fact-checkers, TikTok detects AI-generated content through using technology to search for clips or keywords relating to known AI-generated content (TikTok, n.d.-a). On the other hand, X relies on Community Notes — notes written by users to posts that might be misleading — to safeguard against misinformation using a ground-up approach (X, n.d.). Meanwhile, YouTube relies on internal teams to monitor and detect (i) emerging trends relating to inappropriate content and problematic behaviours, and (ii) coordinated influence operations (YouTube, n.d.-b). It also partners external evaluators and uses their inputs as training data for machine learning systems to build models that review large volumes of videos, to detect borderline content and potentially harmful misinformation (YouTube, n.d.-c). Notably, OpenAI has been working on developing detection classifiers, tools that use AI to evaluate the likelihood that the content is AI-generated. Since May 2024, OpenAI made its image detection classifier open to groups like research labs and research-oriented journalism non-profits to invite assessments on the effectiveness of the tool (OpenAI, 2024c).

The above-mentioned efforts undertaken by TikTok, Meta and YouTube overlap with their actions for *responsive protection*. Once a piece of content is detected to be false,

altered, harmful or unverified, TikTok and Meta respond by either removing the content or restricting its reach across different social networking sites (Meta, n.d.-c; TikTok, n.d.-b). TikTok endeavours to do the same for misleading video or audio content that are AI-generated (TikTok, n.d.-a). Meta's existing policies aims to protect users from different types of misinformation, inclusive of misinformation that (i) risks interfering with voter registration or census participation and (ii) are AI-generated or AI-manipulated falsehoods (Meta, n.d.-b). To tackle repeat offenders on Instagram, Meta imposes an additional penalty by restricting the pages, groups, profile, websites and Instagram accounts that repeatedly share misinformation (Meta, n.d.-c). YouTube limits the spread of potentially harmful misinformation and terminates the channels and accounts that have coordinated influence operations (YouTube, n.d.-b; YouTube, n.d.-c).

The fifth principle goal, *evaluation*, is defined as "collective efforts to evaluate and learn from the experiences and outcomes of dealing with deceptive AI election content". At the time of writing this paper, based on publicly available information, only OpenAI has formed a Safety and Security Committee that attempts to incorporate the principle goal of evaluation in its first task. From 28 May 2024, over 90 days, the committee's first task is to evaluate and further develop OpenAI's current processes and safeguards, including its practices to safeguard election integrity (OpenAI, 2024d).

The sixth principle goal, *public awareness*, has been part of TikTok's, Meta's, Google's and Microsoft's strategies in anticipation of the elections. TikTok states that it will deliver media literacy campaigns about misinformation and how to identify AI-generated content, by partnering with experts and fact-checking organisations around the world (Loftus, 2024). Meta seeks to raise public awareness by conducting media literacy campaigns. For instance, Meta has been rolling out anti-hate speech and misinformation campaigns on its platforms since April 2024 in preparation for the South African Elections (Siddo, 2024).

At present, Google has launched its prebunking campaign in some EU states. Through the campaign, Google hopes to educate citizens on how to spot common manipulation techniques to prepare them for the EU elections (Walker, 2024). Microsoft is in the process of rolling out its public awareness campaign called "Check. Recheck. Vote".

The campaign has launched in the 27 EU states and will be rolled out in other countries, including the US. Through the campaign, Microsoft aims to equip citizens with the first line of defence against deepfakes, i.e., to recognise warning signs, critically evaluate digital content and know what are authoritative election sources (Badanes, 2024; Linde, 2024).

The seventh principle goal, *resilience*, refers to building whole-of-society resilience against deceptive AI election content. Microsoft and OpenAI have jointly launched a USD2 million Societal Resilience Fund in May 2024 to advance AI education and literacy among voters and vulnerable communities (Hutson, 2024b). Building whole-of-society resilience requires collaboration among various stakeholders. To promote research and development, Meta shares its Meta Content Library, a collection of publicly accessible content collected from Facebook and Instagram and the Application Programming Interface (API) with researchers so they can search for data in the library programmatically (Clegg, 2023; Meta, 2024b). Meta has also made publicly available its invisible watermarking technique, Stable Signature, to solicit feedback from the AI research community on the technology can be built, operated and used responsibly (Meta, 2023).

4. KEY UPDATES TO THE INTERNATIONAL AI POLICY LANDSCAPE

The IPS Working Papers No. 52 reviewed the evolving AI regulatory landscape in countries like Australia (Artificial Intelligence Ethics Framework), China (Interim Administrative Measures for Generative Artificial Intelligence Services), Japan (Social Principles of Human Centric AI), the EU (Artificial Intelligence Act), the UK (Pro Innovation Approach to AI Regulation) and the US (Blueprint for an AI Bill of Rights).
Since the publication of the working paper, the AI policy landscape has developed further. At the supranational level, G7 leaders adopted the Hiroshima Process International Guiding Principles for Organisations Developing Advanced AI Systems and the Hiroshima Process International Code of Conduct for Organisations Developing Advanced AI Systems to promote safe, secure and trustworthy AI (European Commission, 2023a). The voluntary Code of Conduct builds on the existing Organisation for Economic Cooperation and Development (OECD) AI Principles and was drafted to help AI developers apply the International Guiding Principles into the

design, deployment and use of advanced AI systems (European Commission, 2023a; European Commission, 2023b; European Commission, 2023c).

The other developments include the United Nations General Assembly's landmark resolution on Artificial Intelligence that emphasises the overarching principles of international law, in particular human rights law, as the central pillar to regulating AI for both the public and private sectors (Blinken, 2024; UN News, 2024).

As a superpower, US' response to the opportunities and threats posed by AI is notable. In October 2023, President Biden issued an Executive Order on safe, secure and trustworthy AI.[2] Though the Executive Order does not directly safeguard threats from AI use in elections, an update given under "new standards for AI safety and security" potentially has some implications on elections. It states that the Department of Commerce is working to ensure that provenance mechanisms are in place for the labelling of AI-generated content. This is to assure Americans that the communications received from the government is authentic. In response to President Biden's executive order, the US National Institute of Standards and Technology (NIST) had released four draft publications to offer guidance in mitigating generative AI risks, promoting transparency for digital content and propose for the development of global AI standards (Department of Commerce, 2024).

In Australia, while there have been no significant updates in the AI policy landscape after its implementation of the AI Ethics Framework, the Australian government has published an Interim Response in January 2024 (Department of Industry, Science and Resources, 2024a). This interim response does not address the threats of AI use in elections directly (Department of Industry, Science and Resources, 2024b). However, it articulates the government's commitment to working with industries to develop a voluntary AI Safety Standard to implement risk-based guardrails for industries, develop options for voluntary labelling and watermarking of AI-generated materials,

[2] It highlights: (i) new standards for AI safety and security, (ii) protection of American's privacy, (iii) advancement of equity and civil rights, (iv) standing up for consumers and workers, (v) supporting workers, (vi) promotion of innovation and competition, (vii) advancement of American leadership around the world, and (viii) ensuring responsible and effective government use of AI (The White House, 2023).

and establish an expert advisory body to support the development of options for further AI guardrails.

The aforementioned either addresses the use of AI technologies in general or recommends actions to promote transparency (e.g., provenance mechanisms). The sections below focus on the efforts undertaken by South Korea, Brazil and some US states to regulate generative AI in the electoral context. We also discuss the EU's Artificial Intelligence Act (AIA), the UK Response Paper to the UK White Paper and the US White House Office of Management and Budget (OMB) policy that briefly discussed what responsible use of generative AI in elections entails.

4.1 South Korea: Ban on the Production and Dissemination of Election-Related Deepfakes 90 Days Before Election Day

An amendment to South Korea's Public Official Election Act came into force in January 2024, banning the use of AI-generated deepfakes for 90 days leading up to election day. Article 82(8) of the law states: "No one may produce, edit, distribute, screen or post deepfake videos for election campaigning purposes from 90 days before the election day to the election day" (National Election Commission of the Republic of Korea, 2024). The ban penalises violators to a jail term of up to seven years or a fine of up to 50 million won (The Korea Herald, 2024). However, lawmakers had earlier decided that AI-generated content to promote voter participation and intraparty activities is allowed. Such content includes AI-generated campaign mottos, song lyrics and speeches (Chakravarti, 2024; Lee, 2024).

4.2 Brazil: Resolution Regulating the Use of Artificial Intelligence in Political Campaigning

In preparation for the election on 6 October 2024, Brazil's Superior Electoral Court has approved a set of resolutions that regulate the use of AI during elections campaign (Farrugia, 2024). They prohibit the use of deepfakes — defined as "the use, to harm or favour a candidacy, of synthetic content in audio, video format or a combination of both, which has been generated or manipulated digitally, even with authorisation, to create, replace or alter the image or voice of a living, deceased or fictitious person"); restrict the use of chatbots and avatars to simulate dialogue between candidate and another real person; and require labelling of AI-generated and AI-manipulated content

(Superior Electoral Court, 2024). Additionally, the resolution calls for greater accountability and responsibility from tech companies to promptly remove content that represents risk (Mari, 2024).

Unlike South Korea's ban on deepfakes, Brazil's rule has no clear time frame and does not mention exclusions of certain deepfake content that is deemed acceptable. Overall, the rule seems to be targeted at ensuring that candidates do not produce and distribute electoral propaganda deepfakes. Candidates who breach the rule may lose their right to run for candidacy or have their mandate revoked if they win the election (Paulo, 2024).

4.3 US States: Ban on Deepfakes or Requirement for Disclosures About Use of Generative AI

Several states in the US have implemented bills that safeguard against the deceptive use of generative AI in the election context. Minnesota was one of the first states to ban the use of deepfake technology in elections (Swanson, 2024). The law prohibits the production and distribution of deepfakes within 90 days of election day, if the content was created without consent, with the goal of hurting a candidate or influencing election (Minnesota Legislature, 2023). More recently, Minnesota has made amendments to the bill. One of the amendments include the expansion of the original timeline of 90 days before the election to include early voting ("30 days before a political party nominating convention, or after the start of the absentee voting period prior to a presidential nomination party, state primary, local primary, special primary, or special election") (Cook, 2024; Gorham, 2024; Minnesota Legislature, 2024).

Besides Minnesota, other US states are passing laws relating to the creation and distribution of deepfakes. For instance, Arizona passed a bill that disallows the sponsorship, creation and distribution of deceptive deepfakes within 90 days before an election, unless there is clear disclosure that the content is generated by AI (State of Arizona, 2024).

At the Congress level, the US Senate Committee on Rules and Administration is advancing three bills that address the use of AI during elections, namely, (i) the Preparing Election Administrators for AI Act, (ii) the AI Transparency in Elections Act

and (iii) the Protect Elections from Deceptive AI Act (Branum & Harper, 2024). The first Act requires the US Election Assistance Commission (EAC) and the US NIST to report to Congress the voluntary guidelines for state and local election offices regarding the use and risks of AI in the administration of elections (Congress.gov, 2024). The second Act requires the disclosure of AI-generated images used in political advertisements (Branum & Harper, 2024). The third Act aligns closely with the regulation in some states, banning "materially deceptive" AI-generated content that influences elections (Branum & Harper, 2024; Tech Policy.Press, n.d.).

4.4 EU: Artificial Intelligence Act

The Artificial Intelligence Act (AIA) was approved by the European Council on 21 May 2024 (European Council, 2024). As presented in IPS Working Papers No. 52, the EU AIA regulates AI systems based on a proportionate risk-based approach.

The EU AIA includes a provision that guards against AI systems that might influence election processes. The EU AIA states that even if an AI system does not qualify as high-risk, it remains subjected to specific transparency obligations for high-risk AI systems as long as it interacts with natural persons or generate content that may pose risks of impersonation or deception. This requirement seeks to "identify and mitigate systemic risks that may arise from the dissemination of AI-generated or manipulated content in particular risk of negative effects on democratic processes, civic discourse and electoral processes, including through disinformation."

4.5 UK: Response Paper to the Pro-Innovation Approach to AI Regulation Paper

In March 2023, the UK government published the Pro-Innovation Approach to AI Regulation Paper (UK White Paper) which received negative feedback on its failure to consider how the AI regulatory framework would translate into safeguarding systemic risks like disinformation and electoral interference. The UK government has since responded to the public feedback with a response paper in February 2024 (Gov.uk, 2024). In the response paper, the government included a section on "Safeguarding democracy from electoral interference". In this section, the government listed three strategies to safeguard electoral integrity from generative AI.

First, the UK's defending Democracy Taskforce will increase its engagement with partners in 2024 by bringing together a wide range of expertise across government to work towards reducing the threat of foreign interference. Second, the government has revised pre-existing laws to safeguard against new generative AI threats. For instance, the Elections Act 2022 introduced a new digital imprints regime. Under this regime, digital campaigning materials targeted at the UK electorate are required to include imprints of the creator's name and address. This increases the transparency of digital political advertising during elections (AI-generated materials inclusive), informing voters about who is promoting the political material online and on whose behalf. Third, the government is looking to watermark election-related content so that citizens feel confident about the content they are viewing. Overall, these strategies reflect the UK government's attempt to increase transparency in the materials that surface online during an elections period.

4.6 US: White House Office of Management and Budget (OMB) Policy

In March 2024, US Vice President Kamala Harris announced the White House Office of Management and Budget (OMB) government-wide policy to combat AI risks and harness its benefits (The White House, 2024). Aligned to the aforementioned executive order, the new policy sets out four overarching actions for agencies: (i) address risks from the use of AI, (ii) expand transparency of AI use, (iii) advance responsible AI innovation, (iv) grow the AI workforce and (v) strengthen AI governance.

The OMB policy stipulates that agencies should practise greater scrutiny on their use of AI that is determined to be either "rights-impacting" or "safety-impacting" (e.g., influence voting and electoral integrity). For AI use that is determined to be rights-impacting or safety-impacting, agencies must apply "minimum practices" by 1 December 2024 or cease the AI use until they attain compliance. Otherwise, they may also obtain a one-year extension from OMB to continue the use or secure a waiver on the practice from OMB (Executive Office of the President, 2024; Fjeld et al., 2024).

To date, the most direct regulations addressing the threats of generative AI use in elections are seen in South Korea, Brazil and some US states. In the next section, we examine new approaches undertaken by Singapore towards responsible AI use.

5. KEY UPDATES TO SINGAPORE'S APPROACH TOWARDS RESPONSIBLE AI

The IPS Working Papers No. 52 reviewed the efforts undertaken by various stakeholders towards responsible AI development and deployment in Singapore.[3] Since the publication of the working paper in August 2023, Singapore has been continuing its efforts to promote responsible AI development and deployment. They include the continuous refinement of AI guidelines, increasing funding support for AI research, development of an AI evaluation toolkit and governance framework. At present, only the Model AI Governance Framework for Generative AI (MGF-Gen AI) addresses the threats of generative AI use in elections. The other initiatives tackle the threats of AI technologies more broadly.

5.1 AI Guidelines

5.1.1 PDPC's Advisory Guidelines on Use of Personal Data in AI Recommendation and Decision Systems

In March 2024, the PDPC issued new advisory guidelines on the use of personal data in AI recommendation and decision systems (PDPC, 2024). The guidelines aim to: (i) provide organisations with clarity on how they should utilise personal data to develop and deploy AI systems and (ii) encourage organisations to adopt baseline guidelines and best practices that help consumers understand how their personal data is being used.

The advisory guidelines target three phases of implementation of an AI system. During the first phase (the development, testing and monitoring phase), the PDPC encourages organisations to practise data minimisation, pseudonymise or de-identify personal data that is used as training data.[4] In this first phase, organisations are required also to update their policies regarding use of personal data to develop AI

[3] They include: (i) PDPC's Discussion Paper on AI and Personal Data — Fostering Responsible Development and Adoption of AI (2018), (ii) Funding for a Research Programme on the Governance of AI and Data Use (2018), (iii) Advisory Council on the Ethical Use of AI and Data (2018), (iv) National AI Strategy (2019), (v) Model AI Governance Framework (2020), (vi) AI Verify (2022), (vii) Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector (2019) and (viii) Public Sector AI Playbook (n.d.).

[4] If pseudonymisation is not possible, PDPC encourages organisations to conduct a Data Protection Impact Assessment before proceeding with the use of raw personal data.

systems and establish practices accordingly, as per the accountability obligation under the PDPA. For the second and third phase (deployment and procurement), organisations are required to be mindful of the three PDPA obligations, namely, consent, notification and accountability. Broadly, the goal is to ensure that consumers are provided with sufficient information to give meaningful and informed consent.

5.1.2 National AI Strategy 2.0 (NAIS 2.0)

In response to the proliferation of AI, in particular generative AI, Smart Nation Singapore launched its second National AI Strategy (NAIS 2.0) in 2023 (Smart Nation Singapore, n.d.). This comes four years after it unveiled the first National AI Strategy in 2019.

The NAIS 2.0 is underpinned by a pro-innovation approach and reinforces the government's commitment in enforcing appropriate guardrails for the innovation and deployment of AI. These objectives are to be achieved through: (i) frequently reviewing and updating the AI governance frameworks to address novel risks, for instance, the Model AI Governance Framework and AI Verify, (ii) adopting a risk-based approach to design interventions when governing AI, and (iii) broadening the current standards and laws in support of responsible AI use. As laid out in NAIS 2.0, Singapore will continue to work with partners on researching and developing appropriate technical standards, tools and services (e.g., watermarking), and contribute actively to international discussions on AI governance.

5.2 Funding for AI Research

The Singapore government continues to fund research that pioneers new AI capabilities and support national-level strategies. For instance, in March 2024, the National University of Singapore (NUS) established the new NUS AI Institute (NAII) (NUS News, 2024). It will focus on research in foundational AI research, policy and societal implications of AI, and real-world applications across various domains.

Another research initiative is the SGD70 million National Multimodal LLM Programme that was announced in December 2023 (IMDA, 2023). The programme is an initiative launched by the Singapore's IMDA in partnership with AI Singapore (AISG) and the Agency for Science, Technology and Research (A*STAR). Given that most LLMs

originate in Western countries, the initiative aims to build a multimodal and localised LLM for the Southeast Asia region.

In addition to these AI research initiatives that seek to establish Singapore as a strategic hub for AI, the Ministry of Digital Development and Information (MDDI) launched the Online Trust and Safety (OTS) research programme (MDDI, 2024a). The programme houses CATOS that seeks to help Singapore combat online harms through developing tools for detecting harmful content, including deepfakes, and through testing technologies like watermarking and content authentication.

5.3 LLM Evaluation Toolkit

To help build a trustworthy AI ecosystem, AI Verify Foundation and IMDA developed Project Moonshot along with partners like DataRobot (an AI company) and investment firm Temasek. An LLM evaluation toolkit, Project Moonshot integrates benchmarking, red teaming and testing baselines (Abdullah, 2024b, AI Verify Foundation, n.d.). It is an open-source toolkit that is made available to businesses so they can rigorously evaluate their LLMs (Abdullah, 2024b; IMDA, 2024d).

5.4 Governance Frameworks

5.4.1 Model AI Governance Framework for Generative AI

On 30 May 2024, Deputy Prime Minister Heng Swee Keat launched Singapore's MGF-Gen AI that was developed by AI Verify Foundation and IMDA (Abdullah, 2024a). The MGF-Gen AI builds on the existing Model Governance Framework that covers traditional AI and an earlier discussion paper on generative AI[5] (IMDA, 2024a). Broadly, the Model AI framework identifies nine key dimensions of a trusted AI ecosystem: (i) accountability, (ii) data, (iii) trusted development and deployment, (iv) incident reporting, (v) testing and assurance, (vi) security, (vii) content provenance, (viii) safety and alignment R&D and (ix) AI for public good (AI Verify Foundation, 2024).

Under content provenance, the framework lists pre-existing technical solutions, namely, digital watermarking techniques (e.g., Meta's Stable Signature) and

[5] Generative AI Discussion Paper: IMDA's 2023 discussion paper on generative AI was published to provide senior leaders in government and businesses with practical pathways to mitigate generative AI risks while tapping on opportunities afforded by generative AI (AI Verify Foundation, 2023).

cryptographic provenance solutions (e.g., C2PA). Given how malicious actors may find loopholes to circumvent provenance tools, the framework calls for: (i) the involvement of key parties in the content life cycle (e.g., publishers, since they are critical in providing users with provenance information), (ii) the standardisation of the types of edits to be labelled to help users tell apart AI-generated content from non-AI generated content and (iii) raising of awareness among users regarding content provenance so that users know how to utilise tools to authenticate content.

5.4.2 AI Governance Playbook for Digital Forum of Small States

In May 2024, it was announced that Singapore will co-lead the development of a Digital Forum of Small States (Digital FOSS) AI Governance Playbook (IMDA, 2024c). Spearheaded by Singapore and Rwanda, the playbook will be launched end of 2024. The playbook takes into consideration the constraints unique to small states when it comes to the development and adoption of AI technologies.

6. REGULATIONS THAT SAFEGUARD ELECTORAL INTEGRITY IN SINGAPORE

At present, the laws that safeguard the integrity of the elections in Singapore do not specifically address the use of AI or generative AI during elections. The existing laws and advisories target three key types of threats — mis- and disinformation, political advertising that influences election outcomes and foreign interference.

First, to tackle mis- and disinformation, the Singapore Police Force (SPF) and MDDI have released an Advisory on Online Conduct in the lead up to the Presidential Election (PE) in 2023. The SPF and MDDI encourage members of the public to evaluate and fact-check information before sharing. Perpetrators who spread misinformation and disinformation online are liable for an offence under Section 14D of the Miscellaneous Offences (Public Order and Nuisance) Act 1906 and POFMA 2019 (MDDI, 2023). Additionally, Cybersecurity Singapore Agency (CSA) has published an Advisory on Detecting and Responding to Deepfake Scams. CSA aims to build the resilience of the whole-of-society against deepfakes by equipping them with a three-step approach to detect deepfakes (Cybersecurity Agency of Singapore, 2024).

Second, to prevent political advertising from becoming a tool used by malicious actors to influence elections, Singapore has laws to promote transparency and accountability of political advertisements that include disclosure requirements for internet and electronic political advertising published in the lead up to the elections. They include Acts and advisories that require the labelling of the source and who is the payer for political advertising (e.g., the Presidential Elections Act 1991, Parliamentary Elections Act 1954and the Advisory on Party Political Films and the Use of Paid Internet Election Advertising).

Third, to safeguard against foreign interference, the Singapore government introduced the Foreign Interference Countermeasures Act (FICA) in 2021. FICA aims to safeguard against foreign interference in domestic elections. However, at this point, FICA does not directly address the AI-generated or AI-manipulated foreign interference campaigns. Table 3 provides a summary of the laws and advisories that uphold the integrity of the Singapore elections.

Table 3: Summary of the laws and advisories that uphold the integrity of the Singapore elections

Title of Laws/Advisories	Purpose	Updates	How this safeguards the integrity of Singapore elections
Laws			
Miscellaneous Offences (Public Order and Nuisance) Act 1906	The Miscellaneous Offences (Public Order and Nuisance) Act 1906 addresses various public order and nuisance related issues in Singapore. Overall, the act helps to maintain societal norms and insure the general welfare of the population.	Revised as of 17 June 2024.	Under the Act, Section 14D states that any person who communicates a false or fabricated message will be liable to a fine not exceeding SGD10,000 or to imprisonment for a term not exceeding three years or to both.
Parliamentary Elections Act (PEA) 1954	The Act aims to ensure fairness, transparency and integrity of the electoral process in Singapore. The Act informs both the electoral members of parliament and electors. For instance, the Act lists a set of conditions that disqualifies electors from voting at the election.	Revised as of 14 June 2024. To operationalise the amendments made in the PEA, the Subsidiary Legislation (SL) under PEA has been revised and came into effect on 14 June 2024 (ELD, 2024).	Sections 61A-61S outline regulations for the publication of election advertising (EA) in terms of: (i) transparency and accountability requirements, (ii) campaign controls: candidates and political parties, (iii) controls on third party campaigning and foreigners and (iv) supplementary provisions. As part of transparency and accountability, Section 61B highlights disclosure requirements for an EA, i.e., the identity particulars of every person who authorised, approved, directed and printed the EA (if print form), and whether it was paid for and by whom. To uphold free and fair elections, provisions in Sections 61C, 61F, 61G, 61K and 61L seek to prevent misleading and unauthorised election communications. Sections 61F and 61G disallow candidates and political parties to publish or cause the publishing of an EA (online and traditional) on behalf of an individual or political party, Sections 61K and 61L disallows unauthorised third parties to

			fund an EA (both monetary and non-monetary terms) and Section 61C disallows persons from publishing EAs during the cooling-off period election advertising ban. Overall, Section 61N gives authority to the Returning Officer to make corrective directions if Sections 61C, 61F, and 61K are bridged. This way, the PEA can safeguard against the use of generative AI to create an online EA that may influence the integrity of elections.
Presidential Elections Act 1991	The Presidential Elections Act 1991 includes provisions that safeguard the integrity and fairness of the presidential election process in Singapore. For instance, Section 71 allows the election of a candidate as President to be void under certain grounds. Some of these grounds include general bribery, corrupt or illegal practices by the candidate, or if majority of electors were prevented from electing their preferred candidate.	Revised as of 4 September 2023.	One of the key revisions of the Presidential Elections Act 1991 include an update to election advertising (EA) laws (ELD, 2023). Given the developments in media and communication, the amplification of online EA (e.g., reposting, resharing/sharing) will be subject to the same requirements as the publication of new EA. Similar to the PEA 1954, Section 42B of the Presidential Elections Act 1991 requires disclosure of the identity particulars of every person who authorised, approved, directed or paid for the EA to be publicly displayed, or printed the EA if it is in print form.
Political Donations Act 2000		Repealed on 29 December 2023; existing obligations under the Act will be ported over to FICA (MHA, 2023).	
Protection from Online Falsehoods and Manipulation Act (POFMA) 2019	The purpose of POFMA is to prohibit the communication of false statements of fact in Singapore. The POFMA Office, situated within the IMDA, is responsible for the administration of the act (POFMA Office, n.d.).	Revised as of 31 December 2021.	The Act criminalises the communication of false statements of fact in Singapore, covering the different forms in which false statements are communicated: (i) individual communication, (ii) the making or altering of bots for communication of false

			statements and (iii) providing services for communication of false statements. POFMA gives ministers the authority to instruct the competent authority (i.e., statutory board or holder of office in service of government or statutory board, e.g., POFMA office) to issue a direction (POFMA Office, n.d.). For example, under Section 40, a government minister can instruct the competent authority to issue an account restriction direction to internet intermediaries and providers of mass media services if the minister decides it is in the public interest to do so, like when the minister finds an account, inauthentic or controlled by a bot, to be disseminating falsehoods or carrying out coordinated inauthentic behaviour. The restriction direction will then require the internet intermediary to either disallow its services from being used to communicate in Singapore or disallow any person from using the accounts to interact with others in Singapore.
Foreign Interference (Countermeasures) Act (FICA) 2021	FICA deals with foreign interference taking either the form of Hostile Information Campaigns (HIC) or local proxies (Politically Significant Persons, PSPs).	Revised as of 14 June 2024. The provisions to counteract foreign interference via HICs came into force on 7 July 2022. The provisions to counteract foreign interference via local proxies came into force on 29 December 2023 (MHA, 2023).	FICA deals with foreign interference taking either the form of Hostile Information Campaigns (HIC) or local proxies (Politically Significant Persons, or PSPs). The Minister for Home Affairs can issue directions to entities (i.e., social media services, electronic services, internet services, and individuals who own or run websites, blogs or social media pages) to help the authorities investigate and counter hostile communications activity that is of foreign origin (MHA, n.d.).

			Additionally, FICA empowers a competent authority (appointed by Ministry of Home Affairs (MHA)) to designate individuals or organisations as PSPs. Once designated, PSPs are subject to countermeasures to mitigate the risk of foreign interference. These countermeasures apply to these means of foreign interference: (i) donations, (ii) volunteers, (iii) leadership, (iv) membership and (v) affiliations. If there are increased risks of foreign interference, one example of a stepped-up countermeasure is the requirement for PSPs to divulge whether they have been given migration benefits by foreign governments (MHA, n.d.).
Advisories			
Advisory on Party Political Films and the Use of Paid Internet Election Advertising (2020)	The advisory was drafted by IMDA in the lead up to the 2020 General Elections to address: (i) the surge in the number of politically themed online videos produced by political parties, socio-political entities and individuals, and (ii) socio-political entities and individuals (non-political parties or prospective candidates) engaging in paid internet election advertising (IEA). Overall, the advisory aims to ensure that (i) political films produced do not misinform voters during the elections and (ii) EA, both paid and non-paid, are labelled for transparency.	The advisory was published on 30 June 2020.	IMDA emphasises the importance of keeping political films rational and grounded on facts. It calls for producers of political films to adhere to the Films Act and the Internet Code of Practice (ICOP). IMDA also highlights that all EA (paid and non-paid) must contain the name of the publisher and person for whom or at whose direction the EA is published. The paid EA will be required to include additional labelling by using words like “sponsored by” or “paid for by” on the IEA. This ensures accountability and prevents the use of paid advertisements to interfere with elections processes.
Advisory on online conduct during Presidential Election (2023)	The advisory was drafted by the SPF and MDDI prior to the 2023 Presidential Election. It reminds members of the public to adopt appropriate online	The advisory published on 12 August 2023.	The SPF and MDDI emphasise the prevalence of misinformation and disinformation in the online space, especially in the lead up to the Presidential Election. They highlight that deepfakes no longer

	conduct during the Presidential Election, cautioning against: (i) misinformation and disinformation in the online space, (ii) online harassment and (iii) online posts with racial or religious connotations.		require sophisticated tools and techniques. Thus, they encourage members of the public to verify content they receive against official sources before posting or sharing the information. Additionally, they highlight that persons who spread misinformation and disinformation online may be liable for an offence of communicating false message under Section 14D of the Miscellaneous Offences (Public Order and Nuisance) Act or manipulated content under POFMA.
Advisory on Detecting and Responding to Deepfake Scams (2024)	This advisory was issued by CSA given recent deepfake videos of Senior Minister Lee Hsien Loong and Prime Minister Lawrence Wong; and a USD25 million scam that a finance worker at a multinational firm had been tricked into paying. The advisory outlines a three-step approach that enables members of the public to detect deepfakes.	The advisory was published on 21 March 2024.	Members of the public are advised to evaluate the content they receive via three steps: (i) assess the message, (ii) analyse audio-visual elements and (iii) authenticate content using tools. Overall, this advisory seeks to build the resilience of the whole of society against the threat of deepfakes. Citizens are advised to cross-reference trusted sources when in doubt about the content they come across.

7. RECOMMENDATIONS FOR SINGAPORE

At present, the government's strategies to maintain and protect the integrity of Singapore elections focus mainly on mitigating the possible influence of external actors on the election outcome and combating mis- and disinformation. For instance, the Presidential Elections Act and Parliamentary Elections Act impose disclosure requirements for political advertisements to prevent malicious actors from misusing them to influence voters. In the following, we propose what else can be done to guard against the threats of AI use in elections in a rapidly evolving landscape.

1. **Regulation of deepfakes**

   The Singapore government is considering how to regulate deepfake content and the possibility of a temporary ban during election time (Chia, 2024). As presented in Section 4, the existing regulations relating to the use of deepfake content during elections in South Korea, Brazil and some US states vary in terms of prohibitions and timeframes. Deepfake content comes in *different formats* — audio, video and images — and usually mimics a person's voice and facial features. Deepfake content has been used for humour, entertainment and artistic expression. However, in the context of elections, deepfake content has the potential to mislead voters and influence votes and voter turnout, even if the content is created without malicious intent. Notable examples of deepfake content that sway voters and election outcome include the robocall that mimicked US President Biden, and fake videos of President Yoon Suk-yeol apologising for his government's corruption and incompetence, and of a Turkish politician that linked him to a Kurdish group classified by the country's State Department as a foreign terrorist organisation.

   Any regulation of deepfake should safeguard election integrity (e.g., deceiving voters and influencing their perceptions of parties and candidates, swaying their voting decision) while allowing for non-harmful use that may yield positive effects (e.g., entertainment, education and promoting voter interest and participation). What should also be clear are the definition of *what constitutes a deepfake* (content that is generated by machine learning or "deep learning"), and what is permissible and not permissible in terms of *types of content* (if it is

election-related and the formats) and *purpose and intent* (as interpreted by reasonable minds). For instance, in the case of South Korea, the government bans the use of AI-deepfakes leading up to election day but it permits the use of AI-generated content to promote voter participation and intraparty activities.

2. **Expand pre-existing laws to combat threats of generative AI use in elections**

   Singapore could expand some of its current laws to include AI-manipulated and AI-generated content, to respond to the evolving campaigning tactics. In the UK, the government recently revised its Elections Act 2022 to include a new digital imprints regime. Producers of digital campaign materials are required to imprint the creator's name and address to increase transparency of digital political advertising (AI-generated materials included). The Singapore government could expand the disclosure requirements under current laws like the Parliamentary Elections Act and Presidential Elections Act to include AI-generated and AI-manipulated election materials. Such a move is aligned with the international movement among governments and tech companies to enhance content provenance and user transparency in order to mitigate risks of impersonation or deception.

3. **Increase collaboration with tech companies to combat threats of AI use**

   Tech companies play a pivotal role in the elections for two key reasons. First, social media platforms are spaces for the dissemination of AI-generated and manipulated mis- and disinformation. Second, AI chatbots may produce either inaccurate or biased electoral information, thereby negatively influencing their voting decisions. Recent collaborations between the government and tech companies include the participation of tech companies in AI Verify Foundation's Governing Committee to deepen the latter's involvement and commitment to promoting responsible AI use in Singapore (PDPC, n.d.). Adobe is also working with the government-funded CATOS to develop content provenance technologies for images and news articles (MDDI, 2024b).

   During Singapore's presidential election in 2023, TikTok directed users to authoritative information from the ELD using content labels and search guides

(Koh, 2024). Moving forward, the government should expand its collaboration with tech companies to target the potential threats posed by AI during elections.

One possibility is to create in-app election centres that push out promptly curated election-related information to users referencing, similar to what TikTok has rolled out in countries like the UK, US, South Africa and in the EU. The government can also work with tech companies to ensure their AI chatbots direct users to ELD's official website when users submit election-related queries. For example, OpenAI's ChatGPT directs US voters to the authoritative website on the country's voting information (e.g., CanIVote.Org). Tech companies can also leverage their AI chatbots to promote trustworthy sources of news to members of the public, like the licensed online news sites under the Online News Licensing Scheme Framework (IMDA, 2024b).

4. **Widen and deepen engagement with the public on challenges of AI in the context of elections.**

   To increase voter awareness of the potential uses of AI during election time and the attendance pitfalls, government agencies should step up their dialogues and engagement with members of the public. Critical content includes: (i) how and when AI is used in the electoral process, (ii) potential risks of AI and (iii) strategies adopted to mitigate the mishaps. This is key because it prepares members of the public for what is to expect.

   At present, several initiatives seek to engage citizens on the potential risks of AI in a few forms, like CSA's advisory on detecting and responding to deepfake scams and National Library's (NLB) talks on Zoom on guarding against misinformation and scams and applying tools and techniques to identify AI-generated content. These are notable efforts but remain insufficient considering what could be limited reach. Election officials could share these content through a wider array of platforms to reach different demographic groups (Angel et al., 2024). For instance, election officials could engage youth on the risks of AI via platforms like Telegram channel, Instagram and TikTok, and middle-aged and senior voters via Facebook. Discussions on the impact of AI on elections could be conducted using a variety of formats, for instance, Instagram reels that are

both engaging and educational. Current advisories like the one issued by the SPF and MDDI could be updated to highlight new threats brought about by AI technologies, like the increased complexity and difficulty in discerning facts from falsehoods and the possibility of AI chatbots providing inaccurate responses to queries.

5. **Evaluate AI tools used in electoral processes.**

If election officials use AI tools in electoral activities, they should ensure that the AI tools promote quality, privacy and security. To ensure quality, the election officials should test and/or pilot the efficacy of the AI tool before its usage in the electoral process (Angel et al., 2024). Where possible, election officials could also request for service providers to provide a demonstration of their technology using local voter data. These prerequisites will provide election officials a clearer picture of how accurately the AI tool will perform in the elections. For privacy, election officials can consider taking direction from PDPC's Advisory Guidelines on use of Personal Data in AI Recommendation and Decision Systems. The advisory can be used to evaluate the product or service that relies on AI by observing how the service providers apply PDPA when using personal data to develop and train AI systems. Most importantly, for security, election officials should ensure that they maintain control and ownership of all the data the AI system processes.

Though Singapore has not implemented AI tools in electoral processes at this point, it is possible that in the upcoming GE there might be discussions to do so. In fact, AI tools have become tools that help inform citizens about Singapore's law-making system. For instance, Singapore launched an AI-powered search engine that helps members of the public comb through parliamentary records, enabling them to understand how issues evolve in Singapore's law-making system (Goh, 2024).

Rapid advancements in AI and the growing prevalence of AI tools deployed in election campaigns surface a new set of challenges for governments around the world. While there is no silver-bullet solution to tackle the threats posed by AI to election integrity, this working paper proposes a suite of measures that can be implemented in the immediate term and long term.

8. REFERENCES

Abdullah, Z. (2024a, May 30). S'pore launches new governance framework for generative AI. *The Straits Times*. https://www.straitstimes.com/tech/s-pore-launches-new-governance-framework-for-generative-ai

Abdullah, Z. (2024b, May 31). S'pore rolls out new toolkit to test gen AI safety, lays out plans to shape global conversations. *The Straits Times*. https://www.straitstimes.com/singapore/s-pore-rolls-out-new-toolkit-to-test-gen-ai-safety-lays-out-plans-to-shape-global-conversations?login=true&close=true

Abrams, S. J., & Khalid, A. (2020, October 21). *Are colleges and universities too liberal? What the research says about the political composition of campuses and campus climate.* American Enterprise Institute. https://www.aei.org/articles/are-colleges-and-universities-too-liberal-what-the-research-says-about-the-political-composition-of-campuses-and-campus-climate/

Adami, M. (2024, March 15). *How AI-generated disinformation might impact this year's elections and how journalists should report on it*. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/news/how-ai-generated-disinformation-might-impact-years-elections-and-how-journalists-should-report

Adobe. (2023, September 14). *Content Credentials*. Adobe Help. https://helpx.adobe.com/sg/creative-cloud/help/content-credentials.html

AI Elections Accord. (2024a). Technology industry to combat deceptive use of AI in 2024 elections [Press release]. https://www.aielectionsaccord.com/uploads/2024/02/Press-Release-AI-Elections-Accord-16-Feb-2024.pdf

AI Elections Accord. (2024b). *A tech accord to combat deceptive use of AI in 2024 elections.* https://www.aielectionsaccord.com/

AI Verify Foundation. (n.d.) Project moonshot — an LLM evaluation toolkit. Retrieved on 14 June 2024, from https://aiverifyfoundation.sg/project-moonshot/

AI Verify Foundation. (2023). *Generative AI: Implications for trust and governance*.

Infocomm Media Development Authority. https://aiverifyfoundation.sg/downloads/Discussion_Paper.pdf

AI Verify Foundation. (2024, May 30). *Model AI governance framework for generative AI: Fostering a trusted ecosystem.* Infocomm Media Development Authority. https://aiverifyfoundation.sg/wp-content/uploads/2024/06/Model-AI-Governance-Framework-for-Generative-AI-6-June-2024.pdf

Alkaissi, H., & McFarlane, S. I. (2023, February 19). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus 15*(2), *e35179*. https://www.cureus.com/articles/138667-artificial-hallucinations-in-chatgpt-implications-in-scientific-writing#!/

Angel, S., Eisen, N., & Lee, N. T. (2024, March 11). *8 best practices for state election officials on AI.* The Brookings Institution. https://www.brookings.edu/articles/8-best-practices-for-state-election-officials-on-ai/

Angwin, J., Nelson, A., & Palta, R. (2024, February 27). *Seeking reliable election information? Don't trust AI.* The AI Democracy Projects. https://www.ias.edu/sites/default/files/Angwin-Nelson-Palta_SeekingReliableElectionInformationDontTrustAI_2-27-24.pdf

Badanes, G. (2024, July 5). Combatting AI deepfakes: Our participation in the 2024 political conventions. *Microsoft Blog*. https://blogs.microsoft.com/on-the-issues/2024/07/05/combatting-ai-deepfakes-our-participation-in-the-2024-political-conventions/

Bagri, V. S. (2023, May 13). Using AI to analyze voter sentiment during election campaigns. *The Times of India Blog*. https://timesofindia.indiatimes.com/blogs/voices/using-ai-to-analyze-voter-sentiment-during-election-campaigns/

Baum, J., & Villasenor, J. (2023, May 8). *The politics of AI: ChatGPT and political bias*. The Brookings Institution. https://www.brookings.edu/articles/the-politics-of-ai-chatgpt-and-political-bias/

BBC. (2024, February 17). Big tech vows action on "deceptive" AI in elections. https://www.bbc.com/news/technology-68316683

Blinken, A. J. (2024, March 21). Consensus adoption of US-led resolution on artificial intelligence by the United Nations General Assembly [Press statement]. US Department of State. https://www.state.gov/consensus-

adoption-of-u-s-led-resolution-on-artificial-intelligence-by-the-united-nations-general-assembly/

Brand, D. (2024, March 19). *The use of AI in elections*. Eurac Research. https://www.eurac.edu/en/blogs/eureka/the-use-of-ai-in-elections

Branum, B., & Harper, T. (2024, June 6). *Senate rules committee advances bills to address harmful AI in elections*. Center for Democracy and Technology. https://cdt.org/insights/senate-rules-committee-advances-bills-to-address-harmful-ai-in-elections/

Channel News Asia. (2024, May 16). Dance videos of Modi, rival turn up AI heat in India election. https://www.channelnewsasia.com/asia/india-modi-election-ai-deepfake-dance-video-misinformation-4340341

Chakravarti, J. (2024, February 20*). AI-generated deepfakes flood South Korean election campaigns.* Bank Info Security. https://www.bankinfosecurity.asia/ai-generated-deepfakes-flood-south-korean-election-campaigns-a-24399

Chen, H. (2024, February 11). AI “resurrects” long dead dictator in murky new era of deepfake electioneering. *CNN*. https://edition.cnn.com/2024/02/12/asia/suharto-deepfake-ai-scam-indonesia-election-hnk-intl/index.html

Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review, 107*(6), 1785–1786. https://doi.org/10.15779/Z38RV0D15J

Chia, O. (2024, July 9). Temporary deepfake ban discussed as way to tackle AI falsehoods during Singapore election. *The Straits Times*. https://www.straitstimes.com/singapore/temporary-deepfake-ban-discussed-as-way-to-tackle-ai-falsehoods-during-singapore-elections

Clayton, J. (2023, July 23). Intel’s deepfake detector tested on real and fake videos. *BBC*. https://www.bbc.com/news/technology-66267961

Clegg, N. (2024, February 6). *Labeling AI-generated images on Facebook, Instagram and Threads*. Meta Newsroom. https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads/

Clegg, N. (2023, November 21). *New tools to support independent research. Meta Newsroom.* https://about.fb.com/news/2023/11/new-tools-to-support-independent-research/

Congress.gov. (2024). S. 3897 — Preparing election administrators for AI Act. https://www.congress.gov/bill/118th-congress/senate-bill/3897/text

Cook, M. (2024, February 21). *Lawmaker says "aggressive" action needed to combat use of AI, deepfakes in MN elections.* Minnesota House of Representatives. https://www.house.mn.gov/SessionDaily/Story/18085

Cornia, A., Kunert, J., & Thurman, N. (2016). *Journalists in the UK*. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/research/files/Journalists%2520in%2520the%2520UK.pdf

Curry, R. (2023, December 17). How 2024 presidential candidates are using AI inside their election campaigns. *CNBC*. https://www.cnbc.com/2023/12/17/how-2024-presidential-candidates-are-using-ai-in-election-campaigns.html

Cybersecurity Agency of Singapore. (2024, March 21). Advisory on detecting and responding to deepfake scams. Singapore Government. https://www.csa.gov.sg/alerts-advisories/Advisories/2024/ad-2024-006#:~:text=9.,sources%20and%20using%20available%20tools.

Department of Commerce. (2024, April 29). Department of Commerce announces new actions to implement President Biden's executive order on AI [Press release]. https://www.commerce.gov/news/press-releases/2024/04/department-commerce-announces-new-actions-implement-president-bidens

Department of Industry, Science and Resources. (2024a, January 17). Supporting responsible AI: Discussion paper. Australian Government. https://consult.industry.gov.au/supporting-responsible-ai

Department of Industry, Science and Resources. (2024b). Safe and responsible AI in Australia consultation: Australian Government's interim response. Australian Government. https://storage.googleapis.com/converlens-au-industry/industry/p/prj2452c8e24d7a400c72429/public_assets/safe-and-responsible-ai-in-australia-governments-interim-response.pdf

Elections Department Singapore. (2023, February 6). First reading of the presidential elections (amendment) bill and the parliamentary elections (amendment) bill [Press release]. https://www.eld.gov.sg/press/2023/FIRST%20READING%20OF%20THE%20

PRESIDENTIAL%20ELECTIONS%20AMENDMENT%20BILL%20AND%20THE%20PARLIAMENTARY%20ELECTIONS%20AMENDMENT%20BILL.pdf

Elections Department Singapore. (2024, May 31). Updating election processes to ensure free and fair parliamentary elections in Singapore [Press release]. https://www.eld.gov.sg/press/2024/PR%20on%20Updating%20Election%20Processes%20to%20Ensure%20Free%20and%20Fair%20Parliamentary%20Elections%20in%20Singapore.pdf

Executive Office of the President. (2024, March 28). Advancing governance, innovation, and risk management for the agency use of artificial intelligence. https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf

European Commission. (2023a, October 30). Commission welcomes G7 leaders' agreement on Guiding Principles and a Code of Conduct on artificial intelligence [Press release]. https://ec.europa.eu/commission/presscorner/detail/en/ip_23_5379

European Commission. (2023b, October 30). *Hiroshima process international guiding principles for advanced AI system.* Policy and Legislation. https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-guiding-principles-advanced-ai-system

European Commission. (2023c, October 30). *Hiroshima process international code of conduct for advanced AI systems.* Policy and Legislation. https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-code-conduct-advanced-ai-systems

European Council. (2024, May 21). Artificial intelligence (AI) act: Council gives final green light to the first worldwide rules on AI [Press release]. https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/

Farrugia, B. (2024, May 29). *Regulating the use of AI for Brazilian elections: What's at stake.* DFRLab. https://dfrlab.org/2024/05/29/regulating-the-use-of-ai-for-brazilian-elections-whats-at-stake/

Fridman, L. (2023, March 26). Sam Altman: OpenAI on GPT-4, ChatGPT, and the

future of AI | Lex Fridman Podcast #367 [Video]. *YouTube*. https://www.youtube.com/watch?v=L_Guz73e6fw

Fjeld, C. T., Gambhir, R., Hecht, A., & Sokler, B. D. (2024, April 9). *OMB issues guidance to federal agencies on the use of artificial intelligence — AI: The Washington report*. Mintz. https://www.mintz.com/insights-center/viewpoints/54731/2024-04-08-omb-issues-guidance-federal-agencies-use-artificial

Giansiracusa, N., & Panditharatne, M. (2023, June 13). *How AI puts elections at risk — and the needed safeguards*. Brennan Center For Justice. https://www.brennancenter.org/our-work/analysis-opinion/how-ai-puts-elections-risk-and-needed-safeguards?gclid=Cj0KCQjw-pyqBhDmARIsAKd9XIM7iPmtI-3Y8V22nZ4C-3dStTVRI8Xu1FhBKqlwBhaAyvmqGUFwuaIaAq4uEALw_wcB&ms=gad_ai%20voting%20system_665268346943_8628877148_150749630603

Goh, Y. H. (2024, May 6). AI-powered search engine makes S'pore parliament debates more accessible. *The Straits Times*. https://www.straitstimes.com/singapore/ai-powered-search-engine-makes-s-pore-parliament-debates-more-accessible?close=true

Goldstein, J. A., & Lohn, A. (2024, January 23). *Deepfakes, elections, and shrinking the liar's dividend*. Brennan Center For Justice. https://www.brennancenter.org/our-work/research-reports/deepfakes-elections-and-shrinking-liars-dividend

Gorham, Q. (2024, July 4). Minnesota expands elections-related deepfake prohibitions. *KTTC*. https://www.house.mn.gov/SessionDaily/Story/18085

Gov.uk. (2024, February 6). *A pro-innovation approach to AI regulation: Government response*. Department for Science, Innovation & Technology. https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response

Hartmann, J., Schwenzow, J., & Witte, M. (2023, January 5). The political ideology of conversational AI: Converging evidence on ChatGPT's pro-environmental, left-libertarian orientation [Working paper]. Cornell University Library, arXiv.org. https://doi.org/10.48550/arXiv.2301.01768

Hutson, T. (2024a, April 22). *Expanding our content integrity tools to support global*

elections. Microsoft Blog. https://blogs.microsoft.com/on-the-issues/2024/04/22/expanding-our-content-integrity-tools-to-support-global-elections/#:~:text=By%20attaching%20secure%20%E2%80%9CContent%20Credentials,with%20since%20it%20was%20created

Hutson, T. (2024b, May 7). *Microsoft and OpenAI launch Societal Resilience Fund.* Microsoft Blog. https://blogs.microsoft.com/on-the-issues/2024/05/07/societal-resilience-fund-open-ai/

Infocomm Media Development Authority. (2024a, January 16). Singapore proposes framework to foster trusted generative AI development [Press release]. https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2024/public-consult-model-ai-governance-framework-genai

Infocomm Media Development Authority. (2024b, May 23). *Online news licensing scheme (ONLS) — Computer online service licence.* Regulations and licensing. https://www.imda.gov.sg/regulations-and-licensing-listing/online-news-licensing-scheme

Infocomm Media Development Authority. (2024c, May 30). Singapore launches model AI governance framework (Gen AI) and AI governance playbook for digital forum of small states (Digital FOSS) [Press release]. https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/factsheets/2024/gen-ai-and-digital-foss-ai-governance-playbook

Infocomm Media Development Authority. (2024d, May 31). Singapore launches project moonshot – a generative artificial intelligence testing toolkit to address LLM safety and security challenges [Press release]. https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2024/sg-launches-project-moonshot

Infocomm Media Development Authority. (2023, December 4). Singapore pioneers S$70m flagship AI initiative to develop Southeast Asia's first large language model ecosystem catering to the region's diverse culture and languages [Press release]. https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2023/sg-to-develop-southeast-asias-first-llm-ecosystem#:~:text=Funded%20by%20the%20National%20Research,multi%2Dmodal%20Large%20Language%20Models

Koh, F. (2024, June 1). In focus: As a crucial election looms, Singapore braces for torrent of 'highly problematic' online threats. *Channel News Asia*. https://www.channelnewsasia.com/singapore/singapore-election-deepfakes-online-misinformation-social-media-4373906

LaChapelle, C., & Tucker, C. (2023, November 28). *Generative AI in political advertising*. Brennan Center For Justice. https://www.brennancenter.org/our-work/research-reports/generative-ai-political-advertising

Lee, S. H. (2024, May 13). *AI and elections: Lessons from South Korea*. The Diplomat. https://thediplomat.com/2024/05/ai-and-elections-lessons-from-south-korea/

Linde, N-L. (2024, May 23). *Addressing the deepfake challenge ahead of the European elections.* Microsoft Blog. https://blogs.microsoft.com/eupolicy/2024/05/23/generative-ai-deepfakes-european-elections/

Loftus, S. (2024, January 18). Protecting election integrity in 2024. *TikTok Newsroom*. https://newsroom.tiktok.com/en-us/protecting-election-integrity-in-2024

Mari, A. (2024, February 28). Brazil outlines rules for AI use during elections. *Forbes*. https://www.forbes.com/sites/angelicamarideoliveira/2024/02/28/brazil-outlines-rules-for-ai-use-during-elections/

Meta. (n.d.-a). Meta's approach to the US 2024 elections. Retrieved June 14, 2024, from https://about.meta.com/file/605259258347828/US-2024-Election-Fact-Sheet_2.pdf/

Meta. (n.d.-b). *Misinformation*. Facebook Community Standards. Retrieved 14 June 2024 from https://transparency.meta.com/en-gb/policies/community-standards/misinformation/

Meta. (n.d.-c). *About fact-checking on Facebook, Instagram and Threads*. Meta Business Help Centre. Retrieved 14 June 2024, from https://www.facebook.com/business/help/2593586717571940?id=673052479947730

Meta (2024a, July 4). How AI-generated content is identified and labeled on Meta. https://www.meta.com/help/artificial-intelligence/how-ai-generated-content-is-identified-and-labeled-on-meta/

Meta. (2024b, May 1). *Meta Content Library and API*. Meta Transparency Center.

Updated on 1 July 2024, https://transparency.meta.com/en-gb/researchtools/meta-content-library/

Meta. (2023, October 6). *Stable signature: A new method for watermarking images created by open source generative AI.* Meta Blog. https://ai.meta.com/blog/stable-signature-watermarking-generative-ai/

Microsoft. (2024, February 16). Technology industry to combat deceptive use of AI in 2024 elections. https://news.microsoft.com/2024/02/16/technology-industry-to-combat-deceptive-use-of-ai-in-2024-elections/

Microsoft Create team. (2023, December 4). Building trust with content credentials in Microsoft Designer. https://create.microsoft.com/en-us/learn/articles/designer-content-credentials

Ministry of Digital Development and Information. (2023, August 12). Advisory on online conduct during presidential election [Press release]. https://www.mddi.gov.sg/media-centre/press-releases/advisory-on-online-conduct-during-presidential-election/

Ministry of Digital Development and Information. (2024a, January 10). Online Trust and Safety Research Programme and Centre for Advanced Technologies in Online Safety. https://www.mddi.gov.sg/media-centre/press-releases/online-trust-and-safety-catos/

Ministry of Digital Development and Information. (2024b, May 16). Centre for Advanced Technologies in Online Safety (CATOS) to receive S$50 million in funding. https://www.mddi.gov.sg/centre-for-advanced-technologies-in-online-safety-catos-to-receive-s-50-million-in-funding/

Ministry of Home Affairs. (2023, December 12). Provisions in the Foreign Interference (Countermeasures) Act for countering foreign interference via local proxies [Press release]. https://www.mha.gov.sg/mediaroom/press-releases/provisions-in-the-foreign-interference-countermeasures-act-for-countering-foreign-interference-via-local-proxies/

Ministry of Home Affairs. (n.d.). Introduction to foreign interference (countermeasures) act (FICA). Retrieved on 11 July 2024, from https://www.mha.gov.sg/fica

Minnesota Legislature. (2023). 2023 Minnesota Statutes, Section 609.771. https://www.revisor.mn.gov/statutes/cite/609.771

Minnesota Legislature. (2024). SF 4729. Office of the Revisor of Statutes.

https://www.revisor.mn.gov/bills/text.php?number=SF4729&session=ls93&version=latest&session_number=0&session_year=2024

Mirza, R. (2024, February 16). *How AI deepfakes threaten the 2024 elections*. Shorenstein Center on Media, Politics and Public Policy. https://journalistsresource.org/home/how-ai-deepfakes-threaten-the-2024-elections/

Morgan, K. (2024, February 14). *Our work to prepare for the 2024 European elections.* TikTok Newsroom. https://newsroom.tiktok.com/en-eu/our-work-to-prepare-for-the-2024-european-elections

Mowshowitz, Z. (2024, March 28). How AI chatbots become political. *The New York Times*. https://www.nytimes.com/interactive/2024/03/28/opinion/ai-political-bias.html

National Election Commission of the Republic of Korea. (2024, January 9). 90 days until National Assembly elections: Election campaigns using deepfake restricted [Press release]. https://www.nec.go.kr/site/eng/ex/bbs/View.do?cbIdx=1270&bcIdx=226657

NUS News. (2024, March 25). *NUS sets up AI institute to accelerate frontier AI research and boost real-world impact for public good*. https://news.nus.edu.sg/nus-sets-up-ai-institute/

O'Donovan, C., & Zakrzewski, C. (2024, June 16). Voice assistants, AI chatbots still can't say who won 2020 election. *The Washington Post*. https://www.washingtonpost.com/technology/2024/06/16/ai-chatbots-alexa-2020-election-results/

OpenAI. (2024a, January 15). *How OpenAI is approaching 2024 worldwide elections*. Safety and Alignment. Updated 14 May 2024, https://openai.com/index/how-openai-is-approaching-2024-worldwide-elections/

OpenAI. (2024b). *C2PA in Dall-E 3*. https://help.openai.com/en/articles/8912793-c2pa-in-dall-e-3

OpenAI. (2024c, May 7). *Understanding the source of what we see and hear online*. https://openai.com/index/understanding-the-source-of-what-we-see-and-hear-online/

OpenAI. (2024d, May 28). OpenAI board forms Safety and Security Committee.

Updated 18 June 2024, https://openai.com/index/openai-board-forms-safety-and-security-committee/

Parsons, A. (2023, October 10). *Adobe Max 2023: Milestone wave of Content Credentials adoption with industry partners Microsoft, Leica Camera, Nikon, Publicis Groupe, and more.* Adobe Blog. https://blog.adobe.com/en/publish/2023/10/10/new-content-credentials-icon-transparency

Paulo, S. (2024, March 8). Brazil seeks to curb AI deepfakes as key elections loom. *France 24*. https://www.france24.com/en/live-news/20240308-brazil-seeks-to-curb-ai-deepfakes-as-key-elections-loom

Personal Data Protection Commission Singapore. (2024, March 1). *Advisory guidelines on use of personal data in AI recommendation and decision systems.* https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/advisory-guidelines/advisory-guidelines-on-the-use-of-personal-data-in-ai-recommendation-and-decision-systems.pdf

Personal Data Protection Commission Singapore. (n.d.). *Singapore's approach to AI governance*. Retrieved on 11 July 2024, from https://www.pdpc.gov.sg/help-and-resources/2020/01/model-ai-governance-framework#:~:text=AI%20Verify%20Foundation%20has%20seven,development%20roadmap%20of%20AI%20Verify.

POFMA Office. (n.d.). *About us*. Retrieved on 14 June 2024, from https://www.pofmaoffice.gov.sg/about-us/

Ramer, H. (2024, February 27). Political consultant behind fake Biden robocalls says he was trying to highlight a need for AI rules. *AP News*. https://apnews.com/article/ai-robocall-biden-new-hampshire-primary-2024-f94aa2d7f835ccc3cc254a90cd481a99

Robins-Early, N. (2024, March 12). Google restricts AI chatbot Gemini from answering questions on 2024 elections. *The Guardian*. https://www.theguardian.com/us-news/2024/mar/12/google-ai-gemini-2024-election

Robins-Early, N. (2023, July 19). Disinformation reimagined: How AI could erode democracy in the 2024 US elections. *The Guardian*. https://www.theguardian.com/us-news/2023/jul/19/ai-generated-disinformation-us-elections

Rozado, D. (2023, March 2). The political biases of ChatGPT. *MDPI Soc. Sci., 12*(3), 148–155. https://doi.org/10.3390/socsci12030148

Seitz-Wald, A. (2024, February 26). Democratic operative admits to commissioning fake Biden robocall that used AI. *NBC News*. https://www.nbcnews.com/politics/2024-election/democratic-operative-admits-commissioning-fake-biden-robocall-used-ai-rcna140402

Siddo, B. I. (2024, April 12). *How Meta is preparing for the 2024 South African elections*. Meta Newsroom. https://about.fb.com/news/2024/04/how-meta-is-preparing-for-the-2024-south-african-elections/

Skelding, C. (2021, December 4). Twitter employees give to Democrats by wide margin: data. *New York Post*. https://nypost.com/2021/12/04/data-shows-twitter-employees-donate-more-to-democrats-by-wide-margin/

Smart Nation Singapore. (n.d.). *National AI strategy*. Retrieved 14 June 2024, from https://www.smartnation.gov.sg/nais/

Soon, C., & Tan, B. (2023, August). Regulating artificial intelligence: Maximising benefits and minimising harms. *IPS Working Papers No. 52*. Institute of Policy Studies, National University of Singapore. https://lkyspp.nus.edu.sg/docs/default-source/ips/ips-working-paper-no-52_regulating-artificial-intelligence-maximising-benefits-and-minimising-harms.pdf

Sparrow, T., & Ünker, P. (2023, May 24). Fact check: Turkey's Erdogan shows false Kilicdaroglu video. *DW News*. https://www.dw.com/en/fact-check-turkeys-erdogan-shows-false-kilicdaroglu-video/a-65554034

State of Arizona. (2024). Arizona Senate Bill 1359. https://www.azleg.gov/legtext/56leg/2R/bills/SB1359S.pdf

Sullivan-Paul, M. (2023, June 15). How would ChatGPT vote in a federal election? A study exploring algorithmic political bias in artificial intelligence [Master's thesis]. Graduate School of Public Policy, University of Tokyo. https://www.pp.u-tokyo.ac.jp/wp-content/uploads/2016/02/10_51218255_SULLIVANPAUL_Michaela.pdf

Superior Electoral Court. (2024, February 27). Resolution No. 23,732. https://www.tse.jus.br/legislacao/compilada/res/2024/resolucao-no-23-732-de-27-de-fevereiro-de-2024

Suraksha P. (2023, December 21). PM Modi taps Bhashini AI platform to bridge

language divide. *The Economic Times*. https://economictimes.indiatimes.com/tech/technology/pm-modi-uses-bhashini-at-kashi-tamil-sangamam-and-smart-india-hackathon/articleshow/106161438.cms?from=mdr

Swanson, K. (2024, March 8). *Minnesota banned the use of deepfakes in elections. Now dozens of other states are, too.* KSTP. https://kstp.com/5-investigates/minnesota-banned-the-use-of-deepfakes-in-elections-now-dozens-of-other-states-are-too/

Tech Policy.Press. (n.d.). Tracker Detail: Protect elections from deceptive AI Act – S.2770. Retrieved on 17 July 2024, from https://www.techpolicy.press/tracker/protect-elections-from-deceptive-ai-act/

The Business Times. (2024, May 9). TikTok to label AI-generated content from OpenAI and elsewhere. https://www.businesstimes.com.sg/companies-markets/telcos-media-tech/tiktok-label-ai-generated-content-openai-and-elsewhere

The Korea Herald. (2024, February 23). Deepfake risks in election [Editorial]. https://www.koreaherald.com/view.php?ud=20240222050761

The White House. (2024, March 28). Fact sheet: Vice President Harris announces OMB policy to advance governance, innovation, and risk management in federal agencies' use of artificial intelligence. https://www.whitehouse.gov/briefing-room/statements-releases/2024/03/28/fact-sheet-vice-president-harris-announces-omb-policy-to-advance-governance-innovation-and-risk-management-in-federal-agencies-use-of-artificial-intelligence/

The White House. (2023, October 30). *Fact sheet: President Biden issues Executive Order on safe, secure, and trustworthy artificial intelligence.* Briefing Room. https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/

TikTok. (2024a, April 20). *TikTok reinforces commitment to platform integrity with key initiatives ahead of South Africa elections*. TikTok Newsroom. https://newsroom.tiktok.com/en-africa/zaelections

TikTok. (2024b, June 4). *TikTok launches UK General Election Centre*. TikTok

Newsroom. https://newsroom.tiktok.com/en-gb/tiktok-launches-uk-general-election-centre

TikTok. (2024c, May 17). *Integrity and Authenticity.* TikTok Community Guidelines. https://www.tiktok.com/community-guidelines/en/integrity-authenticity?cgversion=2024H1update

TikTok. (2024d, May 9). *Partnering with our industry to advance AI transparency and literacy*. TikTok Newsroom. https://newsroom.tiktok.com/en-us/partnering-with-our-industry-to-advance-ai-transparency-and-literacy

TikTok. (n.d.-a). Supporting responsible, transparent AI-generated content. Retrieved on July 4, 2024, from https://www.tiktok.com/transparency/en/supporting-responsible-transparent-ai-generated-content/

TikTok. (n.d.-b). Combating harmful misinformation. Retrieved on 10 July 2024, from https://www.tiktok.com/transparency/en/combating-misinformation/

UN News. (2024, March 21). General Assembly adopts landmark resolution on artificial intelligence. https://news.un.org/en/story/2024/03/1147831

Vadde, S. (2024, January 1). *Unmasking hallucinations in AI chatbots: A critical guide to navigating fact and fiction* [Post]. LinkedIn. https://www.linkedin.com/pulse/unmasking-hallucinations-ai-chatbots-critical-guide-navigating-vadde-04jye

Walker, K. (2024, February 16). *Working together to address AI risks and opportunities at MSC*. Google — The Keyword. https://blog.google/technology/safety-security/working-together-to-address-ai-risks-and-opportunities-at-msc/

Weaver, D. H., Willnat, L., & Wilhoit, G. C. (2019). The American journalist in the digital age: Another look at US news people. *Journalism & Mass Communication Quarterly, 96*(1), 101–130. https://doi.org/10.1177/1077699018778242

Wirtschafter, V. (2024, January 30). *The impact of generative AI in a global election year*. The Brookings Institution. https://www.brookings.edu/articles/the-impact-of-generative-ai-in-a-global-election-year/

World Economic Forum. (2024, January 10). *The Global Risks Report 2024* [19th Edition: Insight Report]. https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf

X. (n.d.). About community notes on X.

https://help.x.com/en/using-x/community-notes

YouTube. (2024 March 18). New disclosures and labels for generative AI content on YouTube. https://support.google.com/youtube/thread/264550152/new-disclosures-and-labels-for-generative-ai-content-on-youtube?hl=en

YouTube. (n.d.-a). How does YouTube raise election information from authoritative sources? Retrieved on 10 July 2024, from https://www.youtube.com/howyoutubeworks/our-commitments/supporting-political-integrity/#election-news-and-information

YouTube. (n.d.-b). Getting ahead of emerging issues. Retrieved on10 July 2024, from https://youtube.com/howyoutubeworks/our-commitments/supporting-political-integrity/#early-detection-of-violative-content

YouTube. (n.d.-c). How does YouTube limit the spread of borderline content and potentially harmful misinformation? Retrieved on 10 July 2024, from https://www.youtube.com/howyoutubeworks/our-commitments/fighting-misinformation/#reducing-the-spread-of-borderline-content

APPENDIX: ABOUT THE AUTHORS

Carol **SOON** is Principal Research Fellow at the Institute of Policy Studies (Lee Kuan Yew School of Public Policy, National University of Singapore) where she heads the Society and Culture department. Her research interests include false information, media regulation, digital inclusion and public engagement. She has published her research in books and peer-reviewed journals such as the *Journal of Computer-Mediated Communication*, *Asian Journal of Communication* and *Public Integrity*. Her most recent book, "Mobile communication and online falsehoods: Trends, impact and practice" published by Springer Nature, addresses existing gaps in research and practice in the management of online falsehoods on instant messaging platforms in Asia. She is also Associate Director of the Asia Journalism Fellowship and Vice Chair of Singapore's Media Literacy Council.

Samantha **QUEK** is a Research Assistant at the Institute of Policy Studies (Lee Kuan Yew School of Public Policy, National University of Singapore). Her research interests include the social and policy implications of digital media and the internet, including artificial intelligence. She holds a bachelor's degree in Communications and New Media at National University of Singapore.

About IPS Working Paper Series

The IPS Working Papers Series is published in-house for early dissemination of works-in-progress. This may be research carried out by IPS researchers, work commissioned by the Institute or work submitted to the Institute for publication.

The views expressed in the Working Papers are strictly those of the author(s) alone and do not necessarily reflect the views of the IPS.

Comments on the Working Papers are invited. Please direct your comments and queries to the author(s).

IPS Working Papers are available from the IPS at $7.00 each (before GST). Postage and handling charges will be added for mail orders.

For more information, please visit www.lkyspp.nus.edu.sg/ips or
contact email: ips@nus.edu.sg or tel: 6516-8388.

Institute of Policy Studies
Lee Kuan Yew School of Public Policy
National University of Singapore
1C Cluny Road House 5
Singapore 259599

Tel: (65) 6516 8388
Web: www.lkyspp.nus.edu.sg/ips
Registration Number: 200604346E