Safer Internet Lab

Snapshot

Countering AI Disinformation

Lessons from Taiwan's 2024 Election Defense Strategies

Summer Chen

SAIL Snapshots is a peer-reviewed article series featuring key research findings from Safer Internet Lab researchers and its associates. The views presented are exclusively those of the authors and do not reflect the official stance of SAIL, CSIS, Google, or any other affiliated organisation.

Countering AI Disinformation

Lessons from Taiwan’s 2024 Election Defense Strategies

Summer Chen[1]

Editor: Noory Okthariza[2]

INTRODUCTION

The presidential and legislative elections in Taiwan are set for January 2024, and deepfake and AI-generated disinformation have already made appearances. Although this type of disinformation isn’t the primary tactic of information operations in this election, it’s evident that malicious actors are experimenting with AI technology to accelerate content production and reduce costs. As these actors become more adept with AI tools, we can foresee a shift where disinformation not only grows in volume but also in sophistication, posing a greater threat to the integrity of the information ecosystem. Taiwan's preparations and responses to this election offer valuable lessons for review and learning.

The deepfake and AI-generated disinformation circulating in Taiwan’s 2024 elections mainly falls into three categories. The first type involves editing existing footage with AI-cloned voices to distort the original message—for instance, interviews with presidential candidates or U.S. lawmakers are altered, adding voice overs with entirely opposite messages. The second type is AI-generated audio clips falsely attributing endorsements to a specific

[1] Co-Founder, FactLink
[2] Researcher, Department of Politics and Social Change, CSIS Indonesia

candidate by public figures to manipulate voter intentions. The two types spread through coordinated networks of fake accounts on platforms like Facebook, YouTube, and TikTok. The third type includes fully AI-generated fabricate content, such as an e-book titled *The Secret History of Tsai Ing-wen* and videos alleging a certain candidate has illegitimate child, using a conspiratorial, sensational tone.

However, the three types of AI disinformation have limited reach and have not penetrated Taiwan’s information ecosystem. The main tactics of information operations in the election include fabricating fake polls, baiting disinformation or misleading information on forums or social media to bait mainstream media into amplifying it in order to influence public opinion. However, the three types of AI-generated disinformation observed so far remain limited to dissemination by fake accounts on social platforms. They have not penetrated mainstream media or triggered significant public engagement, failing to influence Taiwan's information environment.

So far, AI-generated videos of presidential candidates have not been used in this election. According to AI detection experts from Taiwanese universities, National Institute of Cyber Security and Academia Sinica, as of late 2023, AI-generated footages still had noticeable flaws, while edited videos combined with AI-cloned audio appeared more realistic and natural, and are harder to detect with current tools. This provides insight into the technological progression of disinformers, though observations need continual updates as AI technology advances.

This research covers the period from the campaigning phase of Taiwan’s presidential and legislative election up to the voting Day, spanning from September 2023 to January 13, 2024. During this time, this research tracked AI-generated and deepfake disinformation tackled by law enforcement, fact checking organizations, media outlets, and research institutions. The data was collected through field observations, as the author served as the Chief Editor at Taiwan FactCheck Center, leading a team to monitor and debunk online rumors during the election. The report also references documents from law enforcement, findings from research institutions, and some interviews with professionals in the field.

REGULATORY PREPAREDNESS, GAPS AND CURRENT INTERVENTIONS AND LIMITATIONS

In defending against deepfake and AI-generated imagery in Taiwan's presidential and legislative elections, one main actor is the government's law enforcement agencies, while the other is civil society. Taiwan's law enforcement agencies and the Central Election Commission (CEC), in order to maintain neutrality, operate independently from political parties and adhere strictly to legal procedures. For fact-checking and media literacy organizations, public trust and credibility are built on their independence and non-affiliation with any political party. Consequently, while government and civil society operate through different mechanisms, they work independently but effectively complement each other's efforts.

MAIN ACTOR: LAW ENFORCEMENT

To address potential AI-generated and deepfake disinformation in the 2024 elections, Taiwan amended its laws on May 26, 2023. Provisions targeting AI-generated and deepfake disinformation are added to the **Presidential and Vice Presidential Election and Recall Act** and the **Public Officials Election and Recall Act**.

Under the new regulations in the Presidential and Vice-Presidential Election and Recall Act and the Public Officials Election and Recall

Act, creating or disseminating AI-generated and deepfake voices, images, or records with the intent to influence elections can lead to up to 7 years in prison. Additionally, tech platforms and media outlets are required to restrict or remove flagged content within two days of notification, or face fines from NT$200,000 to NT$10 million.

These regulations apply from the announcement of the election date until the day before voting. Candidates or citizens can request content verification; if content is confirmed as deepfake or AI-generated disinformation, the CEC contacts tech platforms and media outlets for removal or restricted access. In this election, no cases have been removed or taken down using this approach.

To strengthen enforcement, Taiwan has appointed dedicated prosecutors in six major cities and established a nationwide "AI-Generated and Deepfake Disinformation Case Processing Center." Three prosecutors work in shifts around the clock in the month leading up to the election.

During the election, the CEC has established communication channels with relevant departments of tech companies, including Google, Meta, and LINE, although TikTok communicates through an outsourced PR firm. The CEC and the tech companies. However, the CEC and tech companies still operate on a case-by-case reporting basis, without any established routine mechanisms or regular meetings in place.

Some AI-generated and deepfake disinformation was proactively monitored and investigated by law enforcement at the start of the rumors' circulation, such as videos falsely portraying one presidential candidate praising their opposition and the video claiming one candidate has an illegitimate child. However, law enforcement officials noted that after reporting such cases to tech platforms, their responses were often inefficient and unclear.

Law enforcement officials reflected that there were still limitations and challenges in the enforcement process, including:

- **Inefficient Platform Response**: Platforms have slow response times and low feedback rates.
- **Lack of AI Verification Tools:** AI Verification technologies remain underdeveloped.
- **Tracing Difficulties**: Rumor-mongers often use VPNs or free online forums, making it hard to trace origins.

In this election, the public did not express concerns about whether the removal of AI or deepfake disinformation would interfere with freedom of speech. However, one case occurred where a legislative candidate, after private sex videos were made public, immediately condemned the videos as a deepfake and called for an investigation by law enforcement. Citing concerns that the disclosure of verification results could trigger political controversy and impact the election, the results were withheld on the grounds of protecting sexual privacy.

OTHER ACTOR: CIVIL SOCIETY

During the election period, another key actor in countering AI-generated disinformation is the collective effort of fact-checking organizations, media literacy groups, media outlets, and research institutions specializing in information operations.

For example, to build AI literacy knowledge and skills, Taiwan FactCheck Center (TFC) established a list of AI detection and verification experts. With support from the Institute for Information Industry, TFC connected with AI experts who are willing to collaborate with the media and form a specialist community. The National Institute of Cyber Security provided AI verification tools and technical expertise, assisting TFC in developing AI verification methodologies, tools and AI literacy materials.

TFC promoted two main AI literacy initiatives projects: one targeted at media professionals through capacity-building workshops for media executives, journalists, and fact-checkers, creating a frontline against disinformation. The other initiative project aimed at students, teachers, and community residents, delivering AI literacy through explanatory articles, interactive on-line quizzes, lectures, and workshops to raise awareness of potential deepfake and AI-generated content during elections.

Another organization, Taiwan Media Watch Foundation, trained volunteers and teachers to lead students in a board game called *Election Wind Indicator*. This board game simulates information manipulation and opinion shaping during elections, allowing participants to practice identifying and countering disinformation and opinion tactics through gameplay.

In election campaigns, civil society organizations operate independently of political parties. When malicious disinformation targets candidates, parties, or the administration, fact-checking and debunking by non-partisan groups gain more public trust than self-issued clarifications.

Since 2019, Taiwan’s fact-checking, media literacy, and information research efforts have raised public awareness and built a robust civil society defense. According to the 2023 Annual Misinformation Survey, conducted by National Taiwan University's Department of Journalism, 58.6% of respondents are aware of fact-checking organizations like the Taiwan FactCheck Center (TFC), MyGoPen, and CoFacts and 50% have heard of TFC. 83.8% of respondents who heard of TFC regarded the Center as credible or very credible. It indicates a growing level of public trust in these organizations.

Due to the limitations of AI detection technologies and tools, and the time-consuming nature of verification work, it is even more essential for civil society, such as TFC to take preventive and proactive measures during the election. These measures include building an AI expert community, organizing workshops for media professionals, and educating readers on AI literacy. By fostering AI literacy, it aims to enhance the media's gatekeeping capabilities and raise readers' awareness of AI-related risks, thereby mitigating the potential impact of AI-generated disinformation outbreaks.

CONCLUSION AND RECOMMENDATION

In summary, the main findings of this research are as follows:

1. Deepfake and AI-generated disinformation have appeared in Taiwan's elections, though they have yet to become primary information operation tactics. This disinformation falls into three categories: AI-cloned audio files, edited videos with AI-cloned voices, and fully AI-generated fabricated content. Edited videos combined with AI-generated audio appear realistic, but no cases of AI-generated candidate videos being used to influence the election have been observed.

2. Taiwan’s law enforcement agencies and civil society have proactively prepared defenses against AI-generated and deepfake disinformation. Law enforcement actively monitors and investigates, while civil society promotes AI literacy, preventing this misinformation from gaining influence. Government agencies and civil fact-checking organizations operate independently but effectively complement each other.

3. In May 2023, Taiwan amended its laws to impose up to seven years in prison for deepfake and AI-generated misinformation intended to influence elections, requiring tech platforms and media outlets to remove verified content or face fines ranging from NT$200,000 to NT$10 million. During the election period, dedicated prosecutors were appointed in six cities, and an "AI-Generated or Deepfake Disinformation Case Processing Center" was established one month before the voting day. In practice, the Central Election Commission interacts with tech platforms on a case-by-case basis, with no routine meetings in place.

4. Civil society actors, including fact-checking organizations and media, enhance public awareness of AI disinformation through fact checks and AI literacy. Given the rapid advancement of AI-driven disinformation, experts agree that countering AI disinformation with AI literacy to build information resilience is a crucial and effective strategy.

We can foresee that AI and deepfake technologies are being explored by rumor-mongers. The rapid advancement of AI will soon accelerate disinformation production, create more varied and harder-to-trace claim versions, reduce production costs, and produce images so realistic that flaws are indistinguishable. The presence of AI and deepfake disinformation during elections is going to pose a major threat, bringing severe challenges to democracy in our society.

Based on the experience from Taiwan's recent election, the following recommendations are proposed to maintain a healthy information ecosystem amid rapid AI development, ensuring election integrity and defending democracy:

1. **Legislation and Policy for AI:** Enact regulations requiring AI products to be publicly trustworthy technologies. Legislation should mandate that AI products include watermarks, original source data, or similar measures, to prevent AI technology from being exploited for fraud, scam and disinformation.
2. **Independent Mechanisms for Law Enforcement and Civil Society**: During elections, law enforcement and civil society should handle AI-generated and deepfake disinformation on their own mechanisms. Law enforcement should investigate autonomously to uphold electoral integrity through judicial independence, while civil society, serving as a government watchdog without ties to political parties or figures, should focus on debunking rumors and promoting AI literacy. This independent role enables public trust and allows civil society to act as a guardian of truth in politically turbulent times.
3. **Establish an AI Verification and Detection Expert Community**: Create a community of stakeholders with high demand for AI verification and detection technology, including AI experts, law enforcement, journalists, fact-checkers, and AI tools developers. Through collaborative effort, this community can learn AI techniques, develop counter-tools, and refine verification methods for AI-generated information.
4. **Capacity Building for Law Enforcement**: At the onset of elections, enhance the capabilities of verifying personnel in law enforcement and facilitate knowledge exchange with other countries to learn the latest verification techniques. During the election period, establish a standard enforcement process for handling AI-generated

and deepfake disinformation, with internal training for relevant law enforcement personnel.

5. **Enhance AI Literacy Education**: Invest resources in training and upskilling media professionals and fact checkers in AI literacy. For the public, extend long-established media and digital literacy programs to include AI literacy, helping people understand AI technology, recognize the potential risks of AI-generated disinformation, and develop skills to filter and find the credible resource of information.

Safer Internet Lab

saferinternetlab.org

Jl. Tanah Abang III no 23-27
Gambir, Jakarta Pusat. 10160

Find Us On

CSIS Indonesia | Safer Internet Lab