Forum: Youth Assembly First Committee (YA1)

Issue: Mitigating the Spread of Misinformation through AI-generated Content

Chairs: Yara Temsah & Reemas Bdour

Introduction

With the rise of AI-generated content, misinformation can now spread faster and more convincingly. It is crucial to combat this issue, to ensure that users can identify and avoid false information. This document aims to go over the interrelation between spreading false information and AI-generated content and come forth with recommendations for mitigating this pressing issue.

Definition of Key Terms

Misinformation

False or inaccurate information, especially that which is deliberately intended to deceive.

AI-generated content

Type of artificial intelligence technology that can produce various types of content, including text, imagery, audio and synthetic data.

Background Information

As of 2023-2024, upgraded AI tools can now create highly realistic fake content rapidly and inexpensively, demonstrating a threat to information integrity and credibility. Many communities have exploited AI's capability to create generated content, and have used that to spread misinformation. The World Economic Forum's Global Risks Report 2024, released in January 2024, identified AI-powered misinformation as a primary immediate risk to the global economy as well as citizen safety over the next two years. AI-powered misinformation can go over and beyond defamation risks, privacy and harassment concerns, as well as reputational damage.

Major Parties Involved

China

China, aiming to be a leader in artificial intelligence, has rapidly developed a comprehensive framework for regulating AI, including regulations on AI recommendation algorithms and AI systems designed to synthetically generate images and video.

United Kingdom

The United Kingdom has been participating in international agreements like the Tech Accord to Combat Deceptive Use of AI in 2024 Elections, signed February 2024. In addition, the UK has also been proposing a contextual, sector-based regulatory framework based on existing networks of regulators and laws, updating their regulations to the advancement of technology.

European Union

The EU has taken forward-looking steps to defend against all formations of intentional disinformation, including Deep Fakes. They have published a procedure for tackling disinformation and stress public engagement to assist people in identifying trustworthy information.

Timeline of Key Events

Date	Description of Event
2018-2020	As Deep Fake technologies gained traction, initial detection methods began developing, including machine learning algorithms designed to spot anomalies like lighting inconsistencies and irregular facial movements in videos. These tools focused on image and video verification, laying the groundwork for future solutions in synthetic content detection.
2021 - 2022	Platforms began creating specific content policies in response to AI-based disinformation. For instance, Facebook and Twitter established guidelines on manipulated content, including labeling Deep Fakes and synthetic media. The European Commission’s strengthened Code of Practice on Disinformation also pushed platforms to address these emerging risks, particularly around elections.

2023	The European Union’s Digital Services Act (DSA) mandated transparency and responsibility from large platforms to counter misinformation risks, especially those from AI generated media. As a result, platforms like TikTok, Meta, and Youtube refined their policies to distinguish legitimate AI content from misleading AI manipulated media. These platforms adopted such strategies such as mandatory disclosure labels for Ai generated content and formed alliances like the Coalition for Content Provenance and Authenticity to standardize tracking media origin and authenticity.
2024	As generatie AI technology rapidly advances, platforms adopt more specific guidelines. Youtube requires explicit labeling for AI generated content, and TikTok automates content labeling for synthetic media. There’s an increasing focus on media literacy campaigns, encouraging users to be critical of digital content sources and recognize synthetic media cues. Additionally, voluntary industry frameworks are introduced, particularly for election related content, as part of a multi-platform response to Deep Fake and misinformation in democratic contexts.

Previous Attempts to Resolve this Issue

There were numerous efforts addressing this global issue however 3 main trials stand out :

1. **Early Detection Tools**

   Initial methods were developed to detect Deep Fakes by analyzing pixel and audio inconsistencies in videos. These techniques aimed to identify signs of manipulation in AI altered media;

2. **Platform Content Policies**

   Social media platforms including Facebook and Twitter (X), established content moderation policies to label and manage synthetic media, especially for high-stakes events like elections;

3. **EU Regulations - Code of Practice and DSA**

   The EU introduced the Code of Practice on Disinformation, followed by the Digital Services Act (DSA), which mandated transparency and content labeling requirements for large online platforms.

Possible Solutions to Resolve this Issue

1. **Targeting Deep Fakes**

   Passing laws addressing Deep Fakes, or creating comprehensive federal legislation to create stricter laws regarding AI-generated content;

2. **Criminalizing Deep Fake misinformation**

Criminalizing/penalizing the practice of creating Deep Fakes for misinformation purposes;

3. Transparent AI development

Promoting transparency, accountability, and fairness in AI systems, as highlighted by UNESCO's Recommendation on the Ethics of AI.

Bibliography

Allen, Mike. "AI and the Spread of Fake News Sites: Experts Explain How to Counteract Them." *News.vt.edu*, 22 Feb. 2024, news.vt.edu/articles/2024/02/AI-generated-fake-news-experts.html.

Bontridder, Noémi, and Yves Poullet. "The Role of Artificial Intelligence in Disinformation." *Data & Policy*, vol. 3, no. E32, 25 Nov. 2021, www.cambridge.org/core/journals/data-and-policy/article/role-of-artificial-intelligence-in-disinformation/7C4BF6CA35184F149143DE968FC4C3B6, https://doi.org/10.1017/dap.2021.20.

Chan, Kelvin. "AI-Powered Misinformation Is the World's Biggest Short-Term Threat, Davos Report Says." *AP News*, 10 Jan. 2024, apnews.com/article/artificial-intelligence-davos-misinformation-disinformation-climate-change-106a1347ca9f987bf71da1f86a141968.

Chu, Chu, and Jia An Lu. "Incorporating the UN Values: Artificial Intelligence and Information Integrity for the SDGs." *United Nations University*, 10 Sept.

2024, unu.edu/macau/blog-post/incorporating-un-values-artificial-intelligence-and-information-integrity-sdgs. Accessed 10 Nov. 2024.

Hornstein, Oscar. "Deputy PM Calls for International Collaboration on AI Deepfakes." *UKTN*, 19 Mar. 2024, www.uktech.news/ai/deputy-pm-collaboration-ai-deepfakes-20240319. Accessed 10 Nov. 2024.

Klepper, David, and Ali Swenson. "AI-Generated Disinformation Poses Threat of Misleading Voters in 2024 Election." *PBS NewsHour*, 14 May 2023, www.pbs.org/newshour/politics/ai-generated-disinformation-poses-threat-of-misleading-voters-in-2024-election.

Li, Cathy, et al. "How AI Can Also Be Used to Combat Online Disinformation." *World Economic Forum*, 14 June 2024, www.weforum.org/stories/2024/06/ai-combat-online-misinformation-disinformation/.

Lock, Oliver. "The Legal Issues Surrounding Deepfakes and AI Content." *Www.farrer.co.uk*, 9 Aug. 2023, www.farrer.co.uk/news-and-insights/the-legal-issues-surrounding-deepfakes-and ai-content/.

Miguel, Raquel. "Platforms' Policies on AI-Manipulated and Generated Misinformation." *EU DisinfoLab*, 4 June 2024, www.disinfo.eu/publications/platforms-policies-on-ai-manipulated-and-generated-misinformation/.

Milward, Hugh . “Combating AI Deepfakes to Safeguard the UK General Election - Source EMEA.” *Source EMEA*, 24 June 2024, news.microsoft.com/source/emea/features/combating-ai-deepfakes-to-safeguard-the-uk-general-election/. Accessed 10 Nov. 2024.

Oxford Languages. “Oxford Languages.” *Oxford Languages*, Oxford University Press, 2024, languages.oup.com/google-dictionary-en/.

Shoaib, Mohamed, et al. “Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models.” *Ar5iv*, 2023, ar5iv.org/abs/2311.17394. Accessed 10 Nov. 2024.