Universiti Teknologi MARA
Cawangan Negeri Sembilan

ACADEMY OF LANGUAGE STUDIES

Edition: 16/2025

APB REMBAU E-BULLETIN

e-ISSN: 2682-776X

Deepfake and AI Challenges in Malaysia

By: Muhammad Nabihan Abu Bakar & Masliya Yahya

As artificial intelligence (AI) evolves, there are growing concerns around the world and in Malaysia concerning its misuse, notably in the form of deepfake technology. Deepfakes employ deep learning to create hyper-realistic fake images or movies that are often indistinguishable from the genuine thing. While this technology provides potential in domains such as entertainment and education, it also poses significant risks, particularly on social media platforms. In Malaysia, the presence and implications of deepfake technology are tremendous, sparking debates about misinformation, cybersecurity, and ethical consequences.

The phrase 'deepfake' refers to synthetic media that manipulates reality with artificial intelligence, frequently producing videos that deceive viewers about real events or statements made by individuals. This technology, initially suggested in a Reddit post in 2017, has acquired attention due to its potential for abuse in a variety of industries, including political campaigning and media manipulation (Dasilva et al., 2021). One of the most significant worries in Malaysia is the damage that deepfakes can do to businesses. A phoney film of a company's CEO making provocative remarks, for example, might spark public outrage, a loss of trust, and substantial financial implications. In a market that is increasingly reliant on digital branding and online presence, such reputational issues cannot be neglected.

Deepfake identities on social media present serious privacy and safety concerns. Fake photos can be used in extortion, harassment, and even scams. Malaysians, particularly public figures and women, are vulnerable to their images being exploited without permission. These concerns about digital safety are especially pertinent in a country with one of Southeast Asia's highest social media consumption rates (Godulla, Hoffman & Seibert, 2021).

The political implications of deepfakes were highlighted during the 2018 elections in several nations, where misinformation was carefully propagated over social media, contributing to voter manipulation. Deepfake technology's accessibility worsened misinformation difficulties, allowing for the quick distribution of manipulated media that damaged public trust. During these electoral cycles, the concept of 'political deepfakes' gained traction, resulting in increased government and media attention (Momeni, 2024). Vividly, deepfake content could affect voters' impressions of candidates, raising ethical and legal concerns about election integrity.

Combating deepfakes demands a multifaceted approach. Researchers and

technology companies must collaborate to develop more effective detection tools that utilise AI to distinguish between genuine and fraudulent transactions (Hu et al., 2022). Public education is also important. Malaysians must be aware of the methods used to verify information and report suspicious things. Furthermore, more stringent legislation and increased regulatory oversight are necessary to hold abusers accountable and safeguard digital spaces.

While detection approaches have advanced significantly, they must continue to evolve to combat increasingly sophisticated synthetic media. Rapid developments in machine learning also democratise access to deepfake generation tools, prompting concerns that people with limited technical abilities may create damaging content, influencing public discourse (Mehta et al., 2023).

The growing challenges require robust mitigation strategies involving technological evolution, regulatory frameworks, and public education to navigate the complexities introduced by deepfakes effectively. In conclusion, while AI brings innovation, Malaysia must confront the darker side of deepfake technology. By prioritising awareness, regulation, and technological safeguards, the country can ensure a safer digital future.

References

Dasilva, J. P., Ayerdi, K. M., & Galdospín, T. M. (2021). Deepfakes on Twitter: Which actors control their spread? *Media and Communication*, 9(1), 301-312. https://doi.org/10.17645/mac.v9i1.3433

Godulla, A., Hoffmann, C. P., & Seibert, D. (2021). Dealing with deepfakes – an interdisciplinary examination of the state of research and implications for communication studies. *Studies in Communication and Media*, 10(1), 72-96. https://doi.org/10.5771/2192-4007-2021-1-72

Hu, J., Liao, X., Liang, J., Zhou, W., & Qin, Z. (2022). Finfer: frame inference-based deepfake detection for high-visual-quality videos. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(1), 951-959. https://doi.org/10.1609/aaai.v36i1.1997

Mehta, P., Jagatap, G., Gallagher, K., Timmerman, B., Deb, P., Garg, S., & Dolan-Gavitt, B. (2023). Can deepfakes be created on a whim? *Companion Proceedings of the ACM Web Conference 2023*, 1324-1334. https://doi.org/10.1145/3543873.3587581

Momeni, M. (2024). Artificial intelligence and political deepfakes: shaping citizen perceptions through misinformation. *Journal of Creative Communications*, 20(1), 41-56. https://doi.org/10.1177/09732586241277335