Published on 24 July 2025

SLAYING THE HYDRA

Combating Misinformation and Disinformation in the Era of Generative Artificial Intelligence

[2025] SAL Prac 19

Gerald **TAN** Han Jie
BSocSc (National University of Singapore),
JD (Singapore Management University); CIPP/A, CIPP/E, CIPM, FIP;
Advocate and Solicitor (Singapore);
Senior Legal Counsel, Amadeus IT Group, Singapore.

I. Introduction

1 In Greek mythology, the half-god hero Heracles found a challenging adversary in the Hydra of Lerna. For each head of the Hydra that Heracles cut off, two heads would regrow in its place. After a fierce and prolonged battle, Heracles, with some help from his companion, managed to kill the Hydra by using a torch to seal the headless tendons of the Hydra's necks after cutting off its heads.

2 Misinformation and disinformation are the Hydra of our times. While misinformation is false information that is created and spread by mistakes, disinformation takes on a more deliberate character – false information that is created and spread by someone who knows full well of the falsity.[1]

3 The Global Risks Report 2024 published by the World Economic Forum placed misinformation and disinformation as the top-ranking global risk for the next two years, describing it as a "rapidly evolving risk".[2] Generative artificial intelligence

1 "Misinformation v Disinformation: What's the Difference?", *BBC* <https://www.bbc.co.uk/bitesize/articles/z3hhvj6> (accessed 20 June 2025).

2 World Economic Forum, *The Global Risks Report 2024* (19th Ed, January 2024) at pp 14 and 18 <https://www3.weforum.org/docs/WEF_The_Global_

(cont'd on the next page)

("Gen AI") models have led the charge in pushing misinformation and disinformation to its pole position.[3]

4 This article examines the challenges arising from misinformation and disinformation powered by Gen AI, and how best to combat those challenges.

II. Hydra presents

5 To understand the difficulty in combating this modern Hydra in misinformation and disinformation, there is a need to first understand what Gen AI is and how it has changed the game altogether.

6 Prior to the advent of Gen AI, AI was largely confined to machine learning based on predictive models, where the AI would observe and classify patterns in the relevant data set.[4] Gen AI was a breakthrough in that it could go beyond the observation and classification exercises; it could also create a high-quality image or text on demand, based on the data it was trained on and not merely based on classification of patterns.[5]

7 The coming of age of Gen AI has changed the game in two profound ways.

8 First, Gen AI has led to increased accessibility. Tools like OpenAI's GPT-4 and DALL-E have led to the democratisation of high-quality AI-generated text, images and other forms of media, by putting user-friendly tools in the hands of users without requiring sophisticated technical expertise. These tools and the like have lowered the barrier to content creation, and

Risks_Report_2024.pdf> (accessed 20 June 2025).

3 World Economic Forum, *The Global Risks Report 2024* (19th Ed, January 2024) at p 18 <https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf> (accessed 20 June 2025).

4 McKinsey & Company, "What Is Generative AI?" (2 April 2024) <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai> (accessed 20 June 2025).

5 McKinsey & Company, "What Is Generative AI?" (2 April 2024) <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai> (accessed 20 June 2025).

therefore directly contributed to the proliferation of AI-generated misinformation and disinformation media.

9 Second, Gen AI has brought about a sharp increase in content quality. Fake websites that appear legitimate can be generated based on non-existent data.[6] "Deepfake" generators can create highly convincing yet fake videos or audio recordings. The rise of deepfakes makes it difficult for consumers to discern manipulated media from authentic or original content.[7]

10 Like many other advances in technology, Gen AI and its output, such as deepfakes, function as a double-edged sword. While it provides certain benefits, it could easily be used as tools of misinformation and disinformation. It possesses the ability to undermine public trust in public figures and institutions, sowing discord in society or even provoking violence.[8] New forms of crimes will likely proliferate as well, with non-consensual deepfake pornography and content targeted at manipulating stock markets being some examples.[9]

11 The surrounding environment within which Gen AI is emerging is also a factor. There has been the observed decline of traditional journalism and people's trust in traditional news

6 Pranshu Verma, "The Rise of AI Fake News Is Creating a 'Misinformation Spreader'", *The Washington Post* (17 December 2023) <https://www.washingtonpost.com/technology/2023/12/17/ai-fake-news-misinformation/> (accessed 20 June 2025); Matthew Cantor, "Nearly 50 News Websites as 'AI-generated', a Study Says. Would I be Able to Tell?", *The Guardian* (8 May 2023) <https://www.theguardian.com/technology/2023/may/08/ai-generated-news-websites-study> (accessed 20 June 2025).

7 Thomson Reuters, "Practice Innovation: Seeing Is No Longer Believing – The Rise of Deepfakes" (18 July 2023) <https://www.thomsonreuters.com/en-us/posts/technology/practice-innovations-deepfakes/> (accessed 20 June 2025).

8 Thomson Reuters, "Practice Innovation: Seeing Is No Longer Believing – The Rise of Deepfakes" (18 July 2023) <https://www.thomsonreuters.com/en-us/posts/technology/practice-innovations-deepfakes/> (accessed 20 June 2025).

9 World Economic Forum, *The Global Risks Report 2024* (19th Ed, January 2024) at pp 14 and 18 <https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf> (accessed 20 June 2025).

outlets, and this has been associated as a driver of misinformation and disinformation.[10]

III. Cutting off Hydra's heads

12 Combating misinformation and disinformation brought about by Gen AI is a multifaceted challenge. Nonetheless, there are some ways to rise to this challenge.

13 A first port of call when thinking about solutions should involve holding stakeholders accountable via the law. Governments and other institutions should work towards a healthy online ecosystem, where social media platforms are obligated to have in place mechanisms to identify and prevent the spread of harmful misinformation and disinformation. This is an act that must be balanced with freedom of expression and privacy-related rights. Legislatures should also consider enacting laws that prohibit and punish individuals or organisations spreading disinformation, especially when public health, safety or democracy is at stake.

14 Several countries already have laws targeting misinformation and disinformation, which could also apply to deepfakes under certain circumstances. Broadly, such laws can be split into: (a) laws that would cover deepfakes but not specifically target them; and (b) laws that specifically target deepfakes.

15 For laws that cover deepfakes but not specifically target them, Singapore has, as a baseline, laws relating to malicious falsehoods and defamation that also serve as bases for civil claims in many common law jurisdictions. Some countries have also enacted laws to combat falsehoods generally but not deepfakes specifically. These include Singapore's Protection from Online Falsehoods and Manipulation Act[11] ("POFMA"), which

10 Jon Bateman & Dean Jackson, "Countering Disinformation Effectively: An Evidence-based Policy Guide", *Carnegie Endowment for International Peace* (31 January 2024) <https://carnegieendowment.org/research/2024/01/countering-disinformation-effectively-an-evidence-based-policy-guide?lang=en> (accessed 20 June 2025).

11 Act 18 of 2019.

came into effect in 2019. POFMA aims to prevent the electronic communication of falsehoods, and is agnostic as to whether the medium is a deepfake or another form of content. More recently in 2023, the Singapore Online Criminal Harms Act[12] ("OCHA") was passed to allow the Government to issue directions against online platforms to block accounts relating to scams or certain types of content from reaching Singapore-based users.[13] As with POFMA, OCHA is agnostic as to whether the medium is a deepfake or another form of content.

16 In more recent years, there has been a push towards more legislation across the globe to target deepfakes specifically. In the US, California and Texas were the first states to pass laws specifically targeting deepfakes in 2019, prohibiting distribution of deepfakes of political candidates close to election dates.[14] At the federal level, the DEEPFAKES Accountability Act, introduced in the US Congress in 2019, would have required deepfake creators to provide disclosures relating to the use of deepfakes.[15] The Chinese Government led the way in Asia, and enacted first-of-its-kind regulations on deepfakes in January 2023.[16] The Chinese regulations create a legal obligation on vendors of deepfake services to verify the real identities of its users, and outrightly prohibit content that endanger national security or disrupt the economy.[17] Singapore, for example, has employed a more targeted approach recently by passing a Bill which makes

12 Act 24 of 2023.

13 Infocomm Media Development Authority, "3 Things Singapore Is Doing to Take Action Against Deepfakes" (19 July 2024) <https://www.imda.gov.sg/resources/blog/blog-articles/2024/07/3-things-sg-do-to-take-action-against-deepfakes> (accessed 20 June 2025).

14 Amanda Lawson, "A Look at Global Deepfake Regulation Approaches" (24 April 2023) <https://www.responsible.ai/a-look-at-global-deepfake-regulation-approaches/> (accessed 20 June 2025).

15 Amanda Lawson, "A Look at Global Deepfake Regulation Approaches" (24 April 2023) <https://www.responsible.ai/a-look-at-global-deepfake-regulation-approaches/> (accessed 20 June 2025).

16 Arjun Kharpal, "China Is About to Get Tougher on Deepfakes in an Unprecedented Way. Here's What the Rules Mean", *CNBC* (22 December 2022) <https://www.cnbc.com/2022/12/23/china-is-bringing-in-first-of-its-kind-regulation-on-deepfakes.html> (accessed 20 June 2025).

17 Arjun Kharpal, "China Is About to Get Tougher On Deepfakes in an Unprecedented Way. Here's What the Rules Mean", *CNBC* (22 December 2022) <https://www.cnbc.com/2022/12/23/china-is-bringing-in-first-of-its-kind-regulation-on-deepfakes.html> (accessed 20 June 2025).

it an offence to publish deepfakes, but limits the scope of the law to deepfakes regarding any political candidate only from the time the Writ of Election is issued to the close of polls.[18]

17 More recently, the European Union (EU)'s new Artificial Intelligence Act[19] requires member states to pass laws to impose transparency obligations on deepfakes,[20] which is another way to regulate the use of deepfakes. Spain, a member of the EU, led the way in passing local legislation to impose fines for unlabelled deepfake content.[21] It is expected that more countries within the EU will follow suit.

18 Regardless of the specificity of the laws enacted or to be enacted, there are concerns on enforceability and effectiveness. Regulating information online has been known to be notoriously difficult, leading to a "cat and mouse" game,[22] akin to cutting off a head of the Hydra only to see two appear in its place.

19 Second, beyond the law, education is a crucial battlefront too. As part of education, teaching and promoting critical thinking and media literacy amongst the population is important. These would include teaching individuals how to recognise potential

18 Chin Soo Fang, "Bill Passed to Counter Digitally Manipulated Content, Deepfakes During Elections", *The Straits Times* <https://www.straitstimes.com/singapore/politics/bill-passed-to-counter-digitally-manipulated-content-deepfakes-during-elections> (accessed 20 June 2025).

19 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), [2024] OJ L 2024/1689.

20 See more exposition on this at "EU AI Act Unpacked #8: New Rules on Deepfakes", *Freshfields* (26 June 2024) <https://technologyquotient.freshfields.com/post/102jb19/eu-ai-act-unpacked-8-new-rules-on-deepfakes> (accessed 20 June 2025).

21 Mimansa, "Spain Leads EU in AI Regulation: Unlabled Deepfakes Could Cost Millions", *MediaNama* (12 March 2025) <https://www.medianama.com/2025/03/223-spain-leads-eu-in-ai-regulation-unlabeled-deepfakes-could-cost-millions/> (accessed 20 June 2025).

22 Alina Alimova, "Defending Against Deep Fakes Through Technological Detection, Media Literacy, and Laws and Regulations", *The International Affairs Review* (12 July 2025) <https://www.iar-gwu.org/print-archive/ikjtfxf3nmqgd0np1ht10mvkfron6n-bykaf-ey3hc-rfbxp-dpte8-klmp4-m2khf> (accessed 20 June 2025).

biases, understanding the provenance behind the content and evaluating the source of the content.

20 The difficulty with government-led education is that it takes time and a lot of resources to be effective in reaching most of the population.[23] Non-governmental organisations and individuals would therefore have to play a significant role in obtaining a greater reach in a shorter amount of time.

21 Perhaps the lower-hanging fruit here lies in funding reliable news outlets. For instance, the New Jersey Civic Information Consortium, a state-supported non-profit, receives money from governments and private donors and thereafter disburses grants to outlets that promote the "quantity and quality of civic information".[24] A similar set-up was adopted in Singapore in 2021, where the journalism arm of Singapore Press Holdings was reconstituted as SPH Media, a non-profit outfit drawing funds from the Government and private donors who share an interest in supporting "quality journalism and credible information".[25]

22 Third, technologies could be smartly employed to combat misinformation and disinformation, though there still are limitations. There are already deepfake detection algorithms in the market, for instance, Microsoft's Video Authenticator and Sensity AI – both use machine learning algorithms to identify deepfake-styled anomalies in videos.[26] The problem is, these technologies are not readily available for everyday use by most

23 Jon Bateman & Dean Jackson, "Countering Disinformation Effectively: An Evidence-based Policy Guide" (31 January 2024) <https://carnegieendowment.org/research/2024/01/countering-disinformation-effectively-an-evidence-based-policy-guide?lang=en> (accessed 20 June 2025).

24 Jon Bateman & Dean Jackson, "Countering Disinformation Effectively: An Evidence-based Policy Guide" (31 January 2024) <https://carnegieendowment.org/research/2024/01/countering-disinformation-effectively-an-evidence-based-policy-guide?lang=en> (accessed 20 June 2025).

25 Grace Ho, "SPH to Restructure Media Business into Not-for-profit Entity to Support Quality Journalism", *The Straits Times* (6 May 2021) <https://www.straitstimes.com/singapore/sph-to-restructure-media-business-into-not-for-profit-entity> (accessed 20 June 2025).

26 BBC, "Deepfake Detection Tool Unveiled by Microsoft", *BBC* (1 September 2020)<https://www.bbc.com/news/technology-53984114>(accessed 20 June 2025); Franklin Okeke, "7 Best AI Deepfake Detector Tools for 2025",

(cont'd on the next page)

of the population. For fake news detectors, such technology was still being trialled as recently as in November 2024, with proofs of concept still pending.[27]

23 Besides the above technologies, there have been developments relating to the setting-up of a new system for the provision of content. In this regard, a group of companies including Microsoft, Adobe and Intel have come together to form the Coalition for Content Provenance and Authenticity ("C2PA"). The C2PA aims to deal with misinformation and disinformation by "provid[ing] an open technical standard for publishers, creators and consumers to establish the origin and edits of digital content".[28] The driver behind the C2PA is to eventually require content credentials to be applied across media content, as a verification tool to safeguard against false information.[29] However, there remains inherent limitations to developing such a new system: those more likely to share false information are more likely to ignore or fail to take notice of information labels, and the C2PA is unable to compel deepfake-generation vendors to comply.[30]

IV. Conclusion

24 There is probably no guaranteed method for this modern Hydra to be slain. The law needs to be used in tandem with education and technologies. Perhaps the best way forward is to

Technopedia (23 January 2024) <https://www.techopedia.com/best-ai-deepfake-detectors> (accessed 20 June 2025).

27 Magda Osman, "How Close are We to an Accurate AI Fake News Detector?", *The Conversation* (6 November 2024) <https://theconversation.com/how-close-are-we-to-an-accurate-ai-fake-news-detector-242309> (accessed 20 June 2025).

28 Coalition for Content Provenance and Authenticity, "Overview" <https://c2pa.org/> (accessed 2 July 2025).

29 Ryan Heath, "Inside the Battle to Label Digital Content as AI-generated Media Spreads", *Axios* (8 February 2024) <https://www.axios.com/2024/02/08/google-adobe-label-artificial-intelligence-deepfakes> (accessed 20 June 2025).

30 Ryan Heath, "Inside the Battle to Label Digital Content as AI-generated Media Spreads", *Axios* (8 February 2024) <https://www.axios.com/2024/02/08/google-adobe-label-artificial-intelligence-deepfakes> (accessed 20 June 2025).

employ a combination of the above strategies through a holistic approach, working with multiple levels of society and having multi-stakeholder engagement.