



Journal of Advanced Research in Computing and Applications

Journal homepage:
<https://karyailham.com.my/index.php/arca/index>
ISSN: 2462-1927



Social Dilemma to AI Dilemma: What's Next?

Nurul Nuha Abdul Molok¹, Mohd Fahmi Mohd Nordin^{2,*}

¹ Department of Information Systems, Faculty of Information and Communication Technology, International Islamic University Malaysia, Gombak, 53100 Kuala Lumpur, Malaysia

² School of Information Technology, UNITAR International University, 47301 Petaling Jaya, Selangor, Malaysia

ARTICLE INFO

Article history:

Received 2 July 2025

Received in revised form 1 August 2025

Accepted 15 August 2025

Available online 4 September 2025

Keywords:

Social media; artificial intelligence;
cybersecurity; cyber safety

ABSTRACT

This study examines past and present social media and artificial intelligence (AI) dilemmas and potential future challenges. It is a dilemma since these technologies are considered as a double-edged sword as there are positive and negative outcomes. The conceptual study explores both dilemmas by reviewing and analyzing sources from academia, government and the industry, and its findings indicate that the rapid pace of AI development is outpacing the establishment of ethical guidelines, making internet governance more difficult. Cyber safety and security issues such as the creation of realistic deepfakes, sophisticated manipulation of public opinion, and the erosion of reality are major concerns. The conclusion is that if humanity struggled to manage the implications of AI-driven algorithms in social media and ill-prepared for Generative AI, we will face the risks of an entirely different magnitude, potentially impacting our ability to control our future and even our existence. Therefore, proactive actions and a focus on human well-being in technology design and governance are crucial, in order to be prepared for the future dilemma.

1. Introduction

The increasing influence of AI-driven technologies on human society is a significant concern to individuals and organizations. The 2020 documentary "The Social Dilemma" in collaboration between the Center of Humane Technology [1] and Netflix, highlighted how AI algorithms in social media platforms manipulate users, leading to issues like declining mental and physical health, privacy breaches, and the uncontrolled spread of misinformation. Despite this initiative and awareness, these problems persist until today and are becoming more detrimental, thanks to the emergence of Generative Artificial Intelligence (GAI).

This relatively new technology is mainly exemplified by Large Language Models (LLMs), which was made famous by ChatGPT, and image and video generation tools, and it presents an amplified challenge, as highlighted in the March 2023 video presentation by the same center that introduced The Social Dilemma. This time it was titled "The A.I. Dilemma." It stated the escalating recklessness

* Corresponding author.

E-mail address: fahmi.nordin@unitar.my

<https://doi.org/10.37934/arca.39.1.7786>

of AI companies deploying powerful AI systems without fully understanding of their capabilities, making them harder to control and predict, and potentially leading to catastrophic consequences.

The purpose of this study is to explore the potential future implications of advanced AI, particularly GAI, given the precedents set by AI-driven social media platforms. It explores the possibility of future dilemma that might exist due to technologies that will emerge. Hence, the research question that this study aims to answer is:

“How do AI technologies impact organizational cybersecurity and societal safety?” In this study, following Molok *et al.*, [2, p.338], cybersecurity is defined as *“the protection of Information and Communication Technology (ICT) infrastructure which includes data, hardware, software, and networks in order to preserve the confidentiality, integrity and availability of information and information processing facilities”*. Cyber safety refers to *“the protection of people’s privacy, physical, mental and emotional wellbeing, through safe and responsible use of ICT”* [2, p.338].

To answer the research question, this study explores the social dilemma induced by AI-driven social media platforms and the AI dilemma, leading to a call for future research on the next dilemma caused by technologies that may emerge in the next few years.

2. A Review on Social Media

This section covers conceptual findings of this study by reviewing existing literature about social media, cybersecurity, information systems, and philosophy gathered from academic sources. Additionally, documents from the government and the ICT industry were also reviewed and analyzed. Not only that, videos from the Center of Humane Technology and news agencies related to the social media dilemma were also reviewed. This section presents the evolution of social media and the social dilemma which covers advantages and disadvantages to individuals and organizations.

2.1 The Evolution of Social Media

To understand how social media platforms shape users’ behavior, it is useful to understand the evolution of this technology. According to Abdul Molok *et al.*, [3], the nature of social media allows cyber safety and security threats to happen as these platforms are *“(1) instantaneous as it is available to the audience immediately upon posting, (2) ubiquitous as it is globally accessible across myriad demographics, and (3) persistent in that is archived in perpetuity”*. These characteristics of social media entice many individual users and businesses, as well as cyber criminals, creating opportunities and challenges to everyone.

Although the use of text was popular in the early days of Facebook back in 2004 and Twitter (now X) since 2006, digital technologies allow the use of images, audio and videos that are rampant in YouTube (since 2005), Instagram (since 2010), and now TikTok (since 2016), according to Walters [4]. Findings from Kross *et al.*, [5] show that Facebook, Instagram and Twitter were the three most popular social media platforms in 2021, but in 2025, Salesforce [6] reports that the top five social media platforms are Facebook with more than 3 billion active users, Whatsapp (2.78 billion users), YouTube (2.5 billion users), Instagram (2 billion users) and TikTok (1.6 billion users).

This evolution shows that, with technological advancements, social media platforms are becoming more dynamic, actively shaping users' online social networking behavior through the diverse content they generate.

2.2 The Social Dilemma

As mentioned in the Introduction, in 2020, a Netflix docu-drama titled “The Social Dilemma” covers social media as platforms to connect us with “friends”, “connections”, or “followers”, but at the same time they are able to manipulate users, thanks to AI-driven algorithms (a set of rules that social media systems follow) [1]. This was highlighted by the Center for Humane Technology, co-founded by former employees of tech giants: Google, Mozilla Labs and NVIDIA, and all the co-founders were notably featured in this documentary.

Without a doubt, there are many advantages of social media. According to Grover *et al.*, [7], organizations can build stronger connections with stakeholders, leading to increased sales, new customer acquisition, and improve corporate reputation through social media. Despite these benefits to organizations, these platforms also caused organizational issues such as leakage of sensitive information [3] and mismanagement of users’ data [4,7]. Therefore, public and private organizations need to balance the social media trade-offs and address these issues through people, process and technological controls [3].

While there are many benefits of social media, the use of these platforms caused social issues of mental and physical well-being [4], challenge security and privacy, pornography and fake news were shared uncontrollably [1]. Indeed, social media platforms are affecting individual’s decision making [7], whether in positive or negative ways.

What is more worrying is the implications of social media to children. Children are susceptible to sexual grooming, cyber bullying, gaming disorder, cyber stalking by pedophiles, identity theft, isolation and loneliness [4,7-10]. Children and adults are playing games, scrolling and watching videos or stories, until they lost the track of time and neglect their responsibilities. Cases of anxiety, depression and self-harm among teenagers continue to rise, misinformation spark outrage and amplifies biases, and the lines of truth or false are becoming blurrier [1,2,7-10].

Undoubtedly, the challenges of social media to the society are still prevalent today. The instantaneous (real-time), ubiquitous (everywhere) and persistent (always-on) nature of social media, combined with diverse user-generated content (text, audio, video) and powered by AI, significantly amplifies the complexity of today's digital landscape.

3. A Review on Artificial Intelligence

After looking at the social media review, this section encompasses conceptual findings on AI by reviewing and analyzing existing literature from academic sources as well as documents from the government and the industry on AI, cybersecurity, information systems, ethics and philosophy. Not only that, videos from the Center of Humane Technology and news agencies related to the AI dilemma were also reviewed. This section incorporates the evolution of AI and the AI dilemma, presenting the positive and negative outcomes of this technology to individuals, organizations and the environment.

3.1 The Evolution of AI

It is important to note that, although GAI suddenly made AI popular, it is not new. According to the National Guidelines on AI Governance and Ethics [11] as shown in Figure 1, AI has existed since the 1950s, as a computer science field which produces intelligent systems and machines that can imitate human intelligence. This is followed by machine learning in 1959, deep learning in 2017 and GAI in 2021. The prevalent adoption of GAI is due to the fact that it is able to create new outputs as

it allows “textual representations in natural language, and images... to videos, music, and software code, thus covering the full spectrum of symbolic expressions of human thought” [11]. While the traditional AI systems or machines can demonstrate intelligence in terms of reasoning, predicting, and learning, GAI can perform more than that, it is able to create new content (text, images, audio, and video), based on learned patterns from training data.

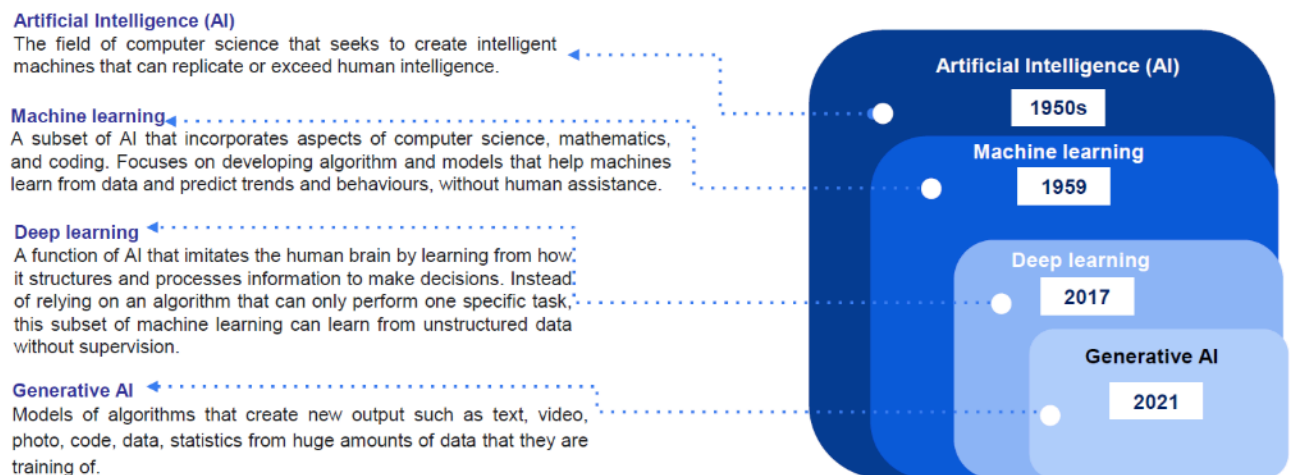


Fig. 1. The evolution of AI Technology [11]

According to Santos *et al.*, [13], technological advancements in machine learning have changed human cognitive capabilities significantly, thus transforming both public and private sector organizations. It is interesting to note that the World Economic Forum [14] anticipated that malicious use of AI or offensive AI would be used by cyber attackers as cyber weaponry that is sophisticated and could evade detection due to its ability to mutate itself as it learns. This was written in 2019 before the GAI was considered to emerge in the AI technology evolution based on the national guidelines [11].

3.2 The AI Dilemma

While users are still grappling with the social dilemma, GAI emerged, not the machine learning AI that is behind the AI-driven algorithms in social media, but the advanced AI, particularly LLMs such as ChatGPT, Gemini and Claude, image generation tools such as Midjourney and DALL-E, and video generation tools such as Sora and Canva Magic Studio. Now that the use of GAI is globally widespread, it is expected that its impacts are increasing day by day. Hence, after the Social Dilemma hype, “The A.I. Dilemma” was presented in March 2023 by Tristan Harris, co-founder of the Center of Humane Technology focusing on GAI as a higher stake to humanity [12].

According to Santos *et al.*, [13], AI enables “real-time data collection, analysis, personalized recommendations and experiential learning, enhancing human cognitive capabilities significantly”. Although AI enhances productivity and analytical capabilities, it escalates AI companies’ recklessness, deploying powerful AI systems without fully understanding of their capabilities, making them harder to control and predict [12]. Similarly, researchers even believe that the inability to control GAI could lead to unethical use of these technologies affecting cyber defenses [13-16] and catastrophic social consequences [15-18], amplifying the danger of social media [1]. The creation of realistic deepfakes continues to deceive people, manipulation of public opinions is made more personalized and sophisticated, and the reality erodes since fabricated texts, images, audios and videos further blurs the line between what is real and what is not [1, 12-15]. Evidently, as reported in the U.S. news [17],

a 14-year-old teenage boy in Florida died by suicide after engaging with an AI bot, and there is a surge of AI apps that are creating non-consensual nude photos (including child exploitation) which are spreading online. It also reports that educators are struggling with how to effectively use AI in the classroom without jeopardizing genuine learning [17].

Other than a dilemma to the society, AI is also considered as a double-edged sword to organizations. On one hand, AI is widely used in cybersecurity controls such as in penetration testing, system administration, user and system authentication, threat detection and response, fraud detection and others [13-16]. It also improves organizational performance, decision-making, public service delivery and economic and societal value [13]. According to Kam *et al.*, [15], GAI is used by cybersecurity professionals such as the Security Operation Center (SOC)'s security analysts to make their analysis more effective, helping them to handle the increasing volume and sophistication of cyber threats.

On the other hand, AI is introducing new cybersecurity risks, biases in decision-making and is susceptible to exploitations [14]. AI is also used by cyber criminals to attack computer systems, similar to what human attackers can do, as the attacks mimic human intelligence [14]. AI-powered malware and social engineering attacks are becoming more common as it can automate phishing emails [15]. According to the national guidelines on AI [11], Cybersecurity Malaysia reports that the capabilities of GAI could be exploited to disseminate misinformation, orchestrate deceptive schemes, and modify malware to evade detection by cyber defenses. Internet governance becomes more difficult as the rapid pace of AI development is outweighing the establishment of ethical guidelines and societal wisdom [1].

It is alarming to find out that large human labor is required to create the GAI technologies and thus, OpenAI (the company behind ChatGPT) was reported to have exploited cheap labors in Kenya, as reported by Richter [19]. Not only that, according to [19], the environment is also impacted due to fact that the AI data centers require thousands of megawatts of energy and in need of fresh water to cool the machines. It is reported that,

"...generative AI's impact is perhaps most acutely felt — from Kenya, where OpenAI reportedly outsourced workers to annotate data for as little as \$2 per hour, to Chile, where AI data centers threaten the country's precious water resources..." [19, p.1].

If social media technologies that are powered by AI can reshape human behavior, erode mental and physical well-being, and divide society, what do you think the GAI might do? Can such AI algorithms which appeared to be benign, turn into something more malignant to humanity? What about AI data centers that are impacting workforce and the environment? Tristan Harris from the Center of Humane Technology, former design ethicist at Google, suggested that failure to effectively manage AI in social media suggests an even greater unpreparedness for GAI, which introduces risks of a significantly higher magnitude, potentially jeopardizing our capacity to control our future and even our existence. Both "dilemmas" underscore the Center of Humane Technology's core message: technology must be designed and governed with humanity's well-being at its core, moving away from business models that prioritize profit at all costs.

4. Mitigating the Risks: Strategies and Solutions

From the above section, it can be summarized that, AI does not only cause cyber safety and security impacts to individuals and organizations, but its data centers are reported to be affecting the environment, our health and the entire ecosystem. Thus, this section covers the strategies and

solutions to mitigate all these risks posed by AI, either through the AI-powered social media to the use of GAI.

4.1 Mitigation Strategies to Individuals

This subsection presents the strategies to mitigate the risks of AI to individuals, including children and adults.

4.1.1 Promoting AI and cyber safety literacy

This study suggests that children need to be educated about the nature of AI, its limitation and when to use their own thinking to solve any problems. This skill is important in mitigating the potential for negative emotional responses (like hate, anger, and frustration) that can arise from human-AI interactions [13]. This is in line with our previous works on cyber wellness for children and youth as presented in [2] and [8] about foundational education to navigate social media and other digital technologies, including those increasingly powered and driven by AI. The public sector agencies should work together with the private sector agencies to conduct awareness programs to the public to promote AI and cyber literacy. The authors have been conducting cyber safety and security awareness programs for children, parents, teachers, senior citizens and employees since 2014 and will continue doing so at schools, community centers and organizations, incorporating the impacts of GAI to everyone in the programs.

4.1.2 Parental oversight

Our previous works [2] and [8] highlight the role of parents and teachers in addressing cyber safety and security risks in the current digital environments. Parents need to understand social media and AI technologies' impacts to themselves and their children before combatting online sexual grooming, cyberbullying, gaming addiction and other cyber threats that may happen to their children. In response to [17] about suicidal behavior among youth, parents and guardians should actively monitor children's interactions with AI applications, especially GAI chatbots, to identify signs of unhealthy emotional attachments or exposure to harmful content. Similarly, the national guidelines [11] suggest that parents and guardians should understand and exercise their rights regarding how AI products use their children's data, including the right to information, to object to data usage, and to request data deletion.

4.1.3 AI Developers' responsibility

AI companies that create AI applications for public use, especially for minors, must implement robust safety features. This includes improving tools to restrict models and filter sensitive content, enhancing detection, response, and intervention mechanisms for user inputs that violate terms or community guidelines, and deploying proactive safety measures [15,17]. Additionally, the national guidelines [11] suggest that AI developers should be legally liable for damages caused by their AI systems so that this can incentivize the creation of safer AI, particularly for vulnerable populations like children. Similar to [11], this study suggests that governments must play proactive actions to ensure these AI companies exercise their ethical responsibilities through laws and regulations. For example, enhancing detection, response, and intervention mechanisms for user inputs that violate

terms or community guidelines and introducing age-specific changes to AI models designed to reduce the likelihood of encountering sensitive and non-solicited content for users under 18.

4.2 Mitigation Strategies to Organizations

This subsection covers the strategies that need to be established and implemented by organizations, both public and private.

4.2.1 Ethical AI governance

When presenting about the A.I. Dilemma, [12] expresses its concern about the development of advanced AI or GAI that is happening very fast, while AI governance is moving at a slower rate. Hence, organizations must ensure that their AI governance frameworks evolve concurrently with the increasing adoption of AI technologies by their members. How to address this? According to the national guidelines [11], policymakers and organizations should integrate ethical AI principles, such as fairness, reliability, safety, control, privacy, and security, into their AI development and deployment frameworks. This includes considering the specific vulnerabilities of users when designing and implementing AI systems and integrating data ethics into data governance frameworks to identify and address risks from unethical data practices, such as misuse, breaches, or biases.

Our recent work [16] suggests empowering organizational leaders with ethical thinking for cyber resilience, emphasizing the importance of thinking ethically while developing and implementing cybersecurity strategies. This focus on ethical considerations in the design and implementation of digital infrastructure is crucial for ensuring that AI systems are developed and deployed with inherent safeguards.

4.2.2 AI and cybersecurity governance

Hendrycks *et al.*, [18] recommend the following cybersecurity strategies to govern AI:

- i. implementing the state-of-the-art cybersecurity controls to prevent accidental information leakage by insiders or information theft by malicious actors through AI systems;
- ii. developing and implementing multi-layered defense strategies against risks to create robust safety systems;
- iii. prioritizing cyber safety and security research and development, ensuring that AI safety and security improvements outpace general AI capabilities;
- iv. fostering strong safety cultures within organizations developing and deploying advanced AI to prevent catastrophic accidents; and
- v. commissioning external penetration testers to proactively identify hazards, dangerous behaviors, and vulnerabilities in AI systems before deployment.

To imply the need for accountability and transparency, The Malaysian National Guidelines on AI Governance and Ethics [11] proposes the following strategies:

- i. to establish internal and external cybersecurity audits to ensure adherence to cyber safety and security protocols and ethical guidelines;

- ii. to ensure consumer protection principles are embedded in AI development and service delivery, providing clear information, options for human interaction, and avenues for compensation;
- iii. to develop and enforce regulations and standards for data handling, storage, and processing to prevent misuse and breaches, aligning with international standards;
- iv. to mandate transparency and explainability in AI systems, ensuring users understand how and why decisions are made, which is essential for trust and accountability
- v. to develop relevant index measurements for each AI system to validate, assess performance, conduct impact assessments, and determine risk levels;
- vi. to invest in tools for Responsible AI, such as model statistics and data explorers, to analyze prediction distributions, characterize errors, and mitigate representation issues in datasets; and
- vii. to hold AI developers legally liable for damages caused by their AI systems.

Interestingly, both [11] and [18] suggest international coordination as a strategy to address AI cybersecurity issues. Hendrycks *et al.*, [18] promotes international coordination and safety regulations to manage the "AI race" (a competitive environment among nations or corporations that compels them to accelerate AI development) and prevent the hasty deployment of unsafe AI systems. The National Guidelines [11] advises to align approaches to AI with global sustainability goals and environmental, social, and governance (ESG) requirements, fostering integration and synergy of AI governance frameworks across organizations and industry sectors.

4.2.3 Educating employees about AI

The National Guidelines on AI [11] suggests that, organizations should educate their employees about AI. Employees should be informed about the potential risks of AI, especially concerning privacy, data usage, and the potential for emotional or psychological manipulation by AI systems. Understanding ethical principles underlying AI governance, such as fairness, reliability, and transparency, can empower individuals to make more informed decisions about AI interactions. The National Guidelines [11] further recommends that organizations also need to develop critical thinking skills to evaluate AI-generated content and interactions in order to help individuals avoid manipulation and misinformation.

5. Conclusions

AI's growing influence, especially GAI, poses escalating risks to individuals, organizations, and the environment. Issues like mental health decline, privacy breaches, misinformation, and cybersecurity threats, intensified by GAI's unpredictable nature, demand urgent attention. The study emphasizes that current governance lags behind rapid AI adoption. Mitigation requires comprehensive strategies: promoting AI literacy and parental oversight for individuals, and for organizations, implementing robust ethical AI governance, stringent cybersecurity measures, and prioritizing safety over profits. Ultimately, AI or any technological development and deployment must be human-centric, moving away from profit-driven models to safeguard societal well-being and existence.

Following "The Social Dilemma" and "The A.I. Dilemma," what is the future dilemma going to be? Will it arise from the intricate interplay of advanced AI, that potentially leading to uncontrollable rogue AIs as anticipated by Hendrycks *et al.*, [18] or complex issues with humanoid robots? The immense computational power of quantum computing could further amplify AI's capabilities and

risks, while blockchain technology might enable resilient, autonomous AI. This convergence will challenge humanity's control, demanding urgent re-evaluation of governance and ethics. The core message remains: technology must be designed and governed with humanity's well-being at its core, ensuring we remain in control amidst these emerging frontiers.

Acknowledgement

This research was not funded by any grant.

Conflict of Interest Statement

The authors declare that there is no conflict of interest regarding the publication of this paper. No financial support, grants, or other forms of compensation were received that could have influenced the outcomes of this work. The research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions Statement

Nurul Nuha Abdul Molok conceptualized and designed the study, supervised the project, and wrote and review the manuscript. Mohd Fahmi Mohd Nordin conducted the study on AI and its challenges, and wrote this part of the manuscript. All authors contributed to manuscript revision, read, and approved the final version.

References

- [1] Center for Humane Technology. "The Social Dilemma". 2020.
- [2] Molok, Nurul Nuha Abdul, and Zahidah Zulkifli, "Parents' Roles in Mitigating Cyber Threats to Children in the New Norm", *Proceedings of National Population Conference* (2021) p. 337-344.
- [3] Molok, Nurul Nuha Abdul, Atif Ahmad, and Shanton Chang. "A case analysis of securing organisations against information leakage through online social networking." *International Journal of Information Management* 43 (2018): 351-356. <https://doi.org/10.1016/j.ijinfomgt.2018.08.013>
- [4] Kross, Ethan, Philippe Verduyn, Gal Sheppes, Cory K. Costello, John Jonides, and Oscar Ybarra. "Social media and well-being: Pitfalls, progress, and next steps." *Trends in cognitive sciences* 25, no. 1 (2021): 55-66. <https://doi.org/10.1016/j.tics.2020.10.005>
- [5] Walters, Austin. "The Evolution of Social Media: Where Dit It Begin...". Wsacommunications.co.uk, June 2022.
- [6] Salesforce. "Social Media Platforms: 10 most popular in 2025".
- [7] Grover, Purva, Arpan Kumar Kar, and Yogesh Dwivedi. "The evolution of social media influence-A literature review and research agenda." *International Journal of Information Management Data Insights* 2, no. 2 (2022): 10011. <https://doi.org/10.1016/j.ijime.2022.100116>
- [8] Molok, Nurul Nuha Abdul, Nur Aiena Hajeerah Abdul Hakim, and Nur Syazwani Jamaludin. "SmartParents: Empowering parents to protect children from cyber threats." *International Journal on Perceptive and Cognitive Computing* 9, no. 2 (2023): 73-79. <https://doi.org/10.31436/ijpcc.v9i2.406>
- [9] Ahmad, Nazilah, Ahmad Arifin, U. Asma'Mokhtar, Zaihosnita Hood, Sabrina Tiun, and Dian Indrayani Jambari. "Parental awareness on cyber threats using social media." *Jurnal Komunikasi: Malaysian Journal of Communication* 35, no. 2 (2019): 485-498. <https://doi.org/10.17576/JKMJC-2019-3502-29>
- [10] Jason Loh and Nik Nurdiana Zulkifli. "The rising danger of cyber-paedophilia in Malaysia (Part 1)". *Focus Malaysia*, Aug 2022.
- [11] Ministry of Science, Technology and Innovation (MOSTI). "The National Guidelines on AI Governance and Ethics". September 2024.
- [12] Harris, Tristan, and Aza Raskin. "The AI Dilemma." www.humanetech.com, March 2023.
- [13] Santos, Ricardo, Amélia Brandão, Bruno Veloso, and Paolo Popoli. "The use of AI in government and its risks: lessons from the private sector." *Transforming Government: People, Process and Policy* (2024). <https://doi.org/10.1108/TG-02-2024-0038>
- [14] Dixon, William, and Nicole Eagan. "3 Ways AI Will Change the Nature of Cyber Attacks." World Economic Forum, June 2019.
- [15] Kam, Hwee-joo, Chen Zhong, Allen Johnston, and Wael Soliman. "Generative AI and Cybersecurity: An Activity Theory Perspective." (2025).

- [16] Molok, Nurul Nuha Abdul, Zahidah Zulkifli, Jongkil Jay Jeong, Sean Maynard, and Atif Ahmad. "Ethical Thinking in Cyber Resilience: Lessons from Malaysian Cyber Leaders."
- [17] Loreben Tuquero. "'One Big Beautiful Bill' could block AI regulations for 10 years, leaving its harms unchecked". *Politifact*. June 2025.
- [18] Hendrycks, Dan, Thomas Authors, and Mantas Mazeika. "An Overview of Catastrophic AI Risks". 2023. <https://doi.org/10.1201/9781003530336-1>
- [19] Richter, Hani. "Karen Hao on how the AI boom became a new imperial frontier". *Reuters*. July 2025.