EDITORIAL

Making the Right Decision: The Ethics of AI

Johann Frederick A. Cabbab
Associate and Issue Editor, PhJLIS

Iyra S. Buenrostro
Editor-in-Chief, PhJLIS

Our reels and timelines now have occasional posts about self-driving cars – some even delivering pizza. The anxiety is real with chatbots carrying on conversations and text prompts generating beautiful artwork (Leffer, 2023). What if machines start passing the Turing test? What if they become self-aware? What if singularity, or the rapid technological expansion characterized by computers becoming smarter than humans and with repercussions for humanity, happens? What will be our Skynet?

DEUS EX MACHINA

The anime entitled *Neon Genesis Evangelion* (which was released in 1995) was polarizing compared to others in the genre. The robots bled. Other than the bloodshed, one scenario was surprising for the time because it involved a trio of supercomputers called a Magi System based on the three personas (thought processes) of its maker ("Liliputian Hitcher," 2023). Ideally, the three had to come to an agreement before an instruction was carried out.

Nearly thirty years later, when that series was released, the possibilities put forth by the anime seem to be crawling and dawning closer and closer.

THE BLUE PILL

For instance, rewind to 2018, at a hawker center in Singapore, one could come face-to-face with a smart food tray return robot making the rounds between the chairs and tables, neatly avoiding anything in its path. Stand in front of it, and it stopped. A year later, in 2019, there was a robosweeper at the airport. It was as big as a janitorial cleaning cart or trolley, was closed off on all sides by metal and plastic sidings, and it also stopped when you stood in front of it. Funny was that it had a jolly cartoonish face painted on the hood, maybe to humanize it and not scare kids approaching it. Speaking of anthropomorphism, that same year marked the popularity of apps like My Talking Pet, which could make photos of your pets talk. Cats back then were finally able to counter allegations that they were fat and lazy.

THE RED PILL

In 2020, at the height of the pandemic, anyone could face swap (a.k.a. deepfake) their features onto celebrities' bodies in videos using an app called Reface (formerly Doublicat). Now, Reface is facing lawsuits stemming from the use of images and likenesses of said celebrities without their consent (Bandara, 2023). On a spookier note, a year later, in 2021, one could animate a long-dead person's portrait using MyHeritage.com's Deep Nostalgia service, making them move as if they were alive. It was disturbing. DALL·E, a text-to-image model developed by OpenAI, also came out that year.

The following year, 2022, had a boom in text-to-image AI art generators and chatbots using large

language models. Image generators Midjourney and Stable Diffusion, and chatbot ChatGPT all rolled out their releases during the year. Early 2023 saw Adobe include generative fill capabilities in their products. Late 2023 saw DALL·E integration into ChatGPT Plus (McAuliffe, 2023). That chatbot generates images for you when you engage it in conversation. Graphics application Krita has Stable Diffusion integration and live AI painting capabilities on more powerful systems (Schmitt, 2023). Text-to-video and image-to-video became within arm's reach (especially for Tiktok aficionados). Google and Facebook joined the fray. Meta posted updates on AI image generation (Edwards, 2023), language translation (Nuñez, 2023), and voice cloning (Batt, 2023). Google (2023), on the other hand, unveiled the multimodal Gemini. This means having a chatbot/generator with language, audio, code, and video understanding, so with the input you provide, it can summarize text, explain a chart, or explain a video to you (Kerner, 2023).

What probably takes the cake would be OpenAI's ChatGPT unveiling the feature of being able to churn out GPTs (OpenAI, 2023). Instead of typing in the chatbox, "Help me format this code," you can instruct it to "Make me a software engineer who helps format my code," (ChatGPT's own in-app example). It then constructs a chatbot specifically for that purpose. So technically, that's AI making bespoke AI for you, much like Karel Čapek's play *R. U. R.* ("Rossum's Universal Robots," 1923), where robots are mass-produced by other robots in factory assembly lines.

ROBOT SPD 13

Ashok et al. (2022) have observed that AI is everywhere, and their ability to pass the Turing Test seems inevitable, maybe in the next few decades or even years. They also noted how AI has not been infallible; there are times when AI apps show biases and make questionable autonomous decisions. Case in point: early on, trying to have ChatGPT write your paper will yield made-up references complete with non-existent links. Ashok et al. (2022) echo what many in the global community believe as well: there must be ethical considerations. They also put forth a conceptual model that may have academic and professional implications.

The United Nations Educational, Scientific and Cultural Organization or UNESCO (2022) has recommended a human rights approach to AI that includes the following: proportionality and do no harm, safety and security, right to privacy and data protection, multi-stakeholder and adaptive governance and collaboration, responsibility and accountability, transparency and explainability, human oversight and determination, sustainability, awareness and literacy, and fairness and non-discrimination. In the library and information field context, the International Federation of Library Associations and Institutions (IFLA, 2020) believes the use of AI technologies in libraries should be subject to clear ethical standards. For example, AI applications that rely on extensive data collection must not override patron privacy. Libraries, however, can and must educate users about AI; support high-quality, ethical AI research; and have the laws, infrastructure, and technologies necessary to adapt to and serve growing AI communities (IFLA, 2020). The OCLC has also commissioned the development and release of a research agenda, *Responsible Operations: Data Science, Machine Learning, and AI in Libraries,* to help the library community raise awareness and understanding towards the responsible use of these technologies and to actively promote transparency, explainability, and accountability in their operations (Padilla, 2019). Hence, in navigating the growing ethical issues related to the use of AI, Cox (2022) emphasizes the importance of analyzing the ethics of AI within the milieu of its potential benefits and challenges to information services and various industries.

As far as academic implications go, Contact North | Contact Nord (2018), a not-for-profit online education corporation funded by the Government of Ontario, posits that "AI is paving the way for personalized, adaptive learning" (No. 2). EDUCAUSE, a nonprofit association which champions advancing higher education through the use of IT, reiterates though that "…as AI develops more human-like capability, ethical questions surrounding data use, inclusivity, algorithmic bias,

and surveillance become increasingly important to consider" (Alexander et al., 2019, p. 27). They further point out that there is a misconception about what AI can and cannot do, resulting in "inflated expectations and a risk that users could assign inappropriate kinds and amounts of authority to AI systems" (EDUCAUSE, 2017, p. 2). The only answer to this would be transparency regarding the sources and use of data, the kind of errors possible, and the limits of the value of the output (EDUCAUSE, 2017).

Isaac Asimov (1950) stated three laws of robotics in his works, which have impacted ethics for AI. Like transparency, the law that a robot may not injure a human is intended to safeguard humankind. Hopefully, we will not reach a volatile state similar to Rossum's Universal Robots versus the League of Humanity (Čapek, 1923), where they feel the need to take a stand regarding all of this: think like the factory and treat them as appliances or liberate them from the chains of slavery.

Hopefully, with our own intelligence, we will make the right decision.

REFERENCES

Alexander, B., Ashford-Rowe, K., Barajas-Murphy, N., Dobbin, G., Knott, J., McCormack, M., Pomerantz, J., Seilhamer, R., & Weber, N. (2019). *EDUCAUSE horizon report: 2019 higher education edition.* EDUCAUSE. https://library.educause.edu/-/media/files/library/2019/4/2019horizonreport.pdf

Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for artificial intelligence and digital technologies. *International Journal of Information Management*, *62*, Article 102433. https://doi.org/10.1016/j.ijinfomgt.2021.102433

Asimov, I. (1950). Runaround. In *I, Robot* (p. 40). Doubleday.

Bandara, P. (2023, June 6). *AI face-swap app reface is sued for 'exploiting' photos of celebrities.* PetaPixel. https://petapixel.com/2023/06/06/ai-face-swap-app-reface-is-sued-for-exploiting-photos-of-celebrities/

Batt, S. (2023, December 12). *Meta reveals its AI voice cloning tool, and you can try it for yourself*. XDA Developers. https://www.xda-developers.com/meta-ai-voice-cloning-tool-demo/

Čapek, K. (1923). *R. U. R. (Rossum's Universal Robots): A fantastic melodrama* (P. Selver, Trans.). Doubleday, Page & Company.

Contact North | Contact Nord. (2018, September 28). *Ten facts about artificial intelligence in teaching and learning.* Teachhttps://teachonline.ca/tools-trends/ten-facts-about-artificial-intelligence

Cox, A. (2022). The ethics of AI for information professionals: Eight scenarios. *Journal of the Australian Library and Information Association, 71*(3), 201–214, https://doi.org/10.1080/24750158.2022.2084885

EDUCAUSE. (2017, April 12). *7 things you should know about artificial intelligence in teaching and learning*. https://library.educause.edu/~/media/files/library/2017/4/eli7143.pdf

Edwards, B. (2023, December 7). *Meta's new AI image generator was trained on 1.1 billion Instagram and Facebook photos*. Ars Technica. https://arstechnica.com/information-technology/2023/12/metas-new-ai-image-generator-was-trained-on-1-1-billion-instagram-and-facebook-photos/

Google. (2023). *Welcome to the Gemini era*. Google DeepMind. https://deepmind.google/technologies/gemini/#introduction

International Federation of Library Associations and Institutions. (2020). *IFLA statement on libraries and artificial intelligence.* https://repository.ifla.org/handle/123456789/1646

Kerner, S. M. (2023, December 13). *What is Gemini? google's most advanced AI model explained*. WhatIs. https://www.techtarget.com/whatis/definition/Google-Gemini

Leffer, L. (2023, October 2). 'AI anxiety' is on the rise--here's how to manage it. *Scientific American.* https://www.scientificamerican.com/article/ai-anxiety-is-on-the-rise-heres-how-to-manage-it/

Lilliputian Hitcher. (2023, December 2). In *Wikipedia.* https://en.wikipedia.org/wiki/Lilliputian_Hitcher

McAuliffe, Z. (2023, September 21). OpenAI brings ChatGPT and Dall-E together for next-gen image creation. *CNET.* https://www.cnet.

com/tech/services-and-software/openai-brings-chatgpt-and-dall-e-together-for-next-gen-image-creation/
Nuñez, M. (2023, December 1). *Meta AI unveils 'seamless' translator for real-time communication across languages.* VentureBeat. https://venturebeat.com/ai/meta-ai-unveils-seamless-translator-for-real-time-communication-across-languages/
OpenAI. (2023, November 6). *Introducing GPTs*. https://openai.com/blog/introducing-gpts
Padilla, T. (2019). *Responsible operations: Data science, machine learning, and AI in libraries*. OCLC Research. https://doi.org/10.25333/xk7z-9g97
Schmitt, N. [Acly]. (2023, November 17). *AI live painting in Krita - interactive updates* [Video]. YouTube. https://www.youtube.com/watch?v=AF2VyqSApjA
UNESCO. (2022). *Recommendation on the ethics of artificial intelligence.* https://unesdoc.unesco.org/ark:/48223/pf0000381137.locale=en

The PhJLIS is published by the
School of Library and Information Studies,
University of the Philippines Diliman.
ISSN 2719-0471 (Online)