Peer response 2 – Dinh Khoi Dang

In his initial post, Dinh Khoi Dang provides a comprehensive description of the potential benefits and risks of AI writers within different possible settings, including administrative tasks, academic and professional writing, and creative work. The main benefits outlined were increased efficiency and accessibility, while the main risks encompassed ethical issues, lack of originality in the produced content, and increased and perpetuated biases.

Dinh Khoi Dang correctly highlights the need for mitigating approaches to help ensure that the benefits of using these models end up outweighing their risks, such as ethical development frameworks, human oversight, and permanent assessment of their use and impact on society. Other important dimension is the implementation of enhanced development approaches, specifically aimed at overcoming these risks from the onset. These include the possibility of instilling moral or ethical reasoning during development, and fine-tuning models to ensure that outputs are aligned with broader societal values (Hutson, 2021; Perrigo, 2024; Weidener and Fischer, 2024). Another area of focus is the curation and improvement of model training data so that pre-existing biases can be dampened and are not perpetuated (Hutson, 2021). Finally, adequate oversight must be put in place to ensure that users are made fully aware of the risks and pitfalls of using these models, and model developers held accountable by potentially harmful or wrong outputs, which requires establishing innovative legal frameworks and strong oversight and regulatory bodies (Muller et al., 2021; Food and Drug Agency, 2024).

References:

Food and Drug Agency (2024) 'Artificial Intelligence & Medical Products: How CBER, CDER, CDRH, and OCP are Working Together'. Available from: https://www.fda.gov/media/177030/download?attachment.

Hutson, M. (2021) 'Robo-writers: the rise and risks of language-generating AI', *Nature*, 591(7848), pp. 22–25. Available from: https://doi.org/10.1038/d41586-021-00530-0.

Muller, H., Mayrhofer, M.T., Van Veen, E.-B. & Holzinger, A. (2021) ‘The Ten Commandments of Ethical Medical AI’, *Computer*, 54(7), pp. 119–123. Available from: https://doi.org/10.1109/MC.2021.3074263.

Perrigo, B. (2024) ‘Inside Anthropic, the AI Company Betting That Safety Can Be a Winning Strategy’, *Time Magazine*, 30 May. Available from: https://time.com/6980000/anthropic/ (Accessed: 21 December 2024).

Weidener, L. & Fischer, M. (2024) ‘Role of Ethics in Developing AI-Based Applications in Medicine: Insights From Expert Interviews and Discussion of Implications’, *JMIR AI*, 3, p. e51204. Available from: https://doi.org/10.2196/51204.