e-Portfolio Activity: Reflective Activity 1 – Ethics in Computing in the age of Generative AI (1005 words)

Generative AI (GenAI) tools have significantly shifted how the general public interact with AI (Ipsos, 2023), as ChatGPT saw 100 million users in the first two months of its launch, surpassing the most popular social media applications such as Instagram and TikTok (Hu, 2023). Notably, a sizeable proportion of government employees in the UK are already using the tools in their work (Bright et al., 2023), signifying a critical shift towards maximising productivity gains and increased efficiency, whilst balancing mistrust in AI (Leyer & Schneider, 2021). Given these rapid developments, an additional spotlight is placed on ensuring ethical practices.

Corrêa et al. (2023) reviewed over two hundred governance policies and ethics guidelines that have been published across the globe to draw an overall consensus on what are considered ethical practices in AI usage. The most strongly held principle focuses on ensuring transparency and explainability, followed by trustworthiness and security, and with fairness and non-discrimination following this. It can be argued that the growing need for explainability mirrors the increasing complexity and "black-box" nature of AI, particularly GenAI models (Hagendorff & Wezel, 2020). The report further highlights the differences in distribution of published research in AI ethics vs. general papers on AI. Europe has published the most papers in AI ethics, followed by the United States, even though China and United States are leading in AI publications overall (Draux, 2024). Critically, Africa, South America and South East Asia produce little with respect to both ethics and general AI papers. The implications of the country trends will be assessed now.

China, United States and Europe are constantly in competition with each other in what is deemed as the AI arms race (Elliott, 2025; Hagendorff, 2020). The competition therefore results in a conflict in balancing regulation and innovation. Focusing specifically on Europe, Europe has taken an effective stance in regulation via the introduction of the Regulation (EU) 2016/679 (2016), also known as the General Data Protection Regulation (GDPR), which successfully protects the data privacy of individuals. It has been drawn upon in other countries outside of Europe such as India as a benchmark for AI regulation (Fjeld et al., 2020). Regulation of GenAI models are presented with a unique challenge as the training of these models demand vast amounts of data, resulting in the ingestion of personal and private data in the process (Fui-Hoon Nah et al., 2023). Hacker et al. (2023) argue that technology-neutral regulations such as GDPR are much better equipped at managing the risks of GenAI for several reasons. Firstly, given the rapid evolution of AI, particularly the progress in GenAI models, technology-specific regulations quickly become outdated. Secondly, it may fail to sufficiently capture emerging AI-based technologies, such in the case of AI-specific regulations failing to capture challenges posed by GenAI models.

Meskó & Topol (2023) discuss this in detail with respect to healthcare and the failure of the United States' Food And Drug Administration (FDA) to accommodate the new wave of GenAI-powered technology. They argue that pre-approved devices which use traditional AI and are upgraded with GenAI technology fall at risk of unaccounted ethical and legal challenges in protecting patient privacy and human rights. The authors highlight the potential harm created via model bias due to unrepresentative data, lack of transparency in model predictions and data breaches due to leakage.

Hacker et al. (2023) propose a new regulatory framework based on high-risk applications of GenAI rather than the models itself, and Meskó & Topol (2023) propose creating a regulation specifically for companies who plan on deploying GenAI in existing devices and services. Despite this, returning to the initial point on “AI arms race”, the United States overall are pushed to accelerate development in GenAI for the sake of national security, in spite of recognising the need for a dialogue between ethicists, philosophers and technologists (Vincent, 2023).

Surprisingly, given limited output in AI ethics research, China have been the first country to specifically target GenAI with dedicated legislation, which came in force in 2023, signalling China’s ambition towards global leadership in AI (Roberts & Hine, 2023). However, the law was met significant challenges within the country itself, as researchers noted the potential detriment to speed of innovation and high costs as a result of lack of returns and increased risk (Abiri & Huang, 2023). Since the initial draft, a relaxation of measures has taken place, however, China’s aim to control misuse of GenAI to gain unfair political advantage is met with this law.

Critically, as the global powers in AI continue their race, a large proportion of the developing world are suffering at the expense of GenAI acceleration, raising the issue of due responsibility towards these nations. For example, workers in developing nations are employed in labelling “farms” to generate training data for these models (Perrigo, 2023) earning low wages, revealing the challenge of hidden costs and exploitation (Hagendorff & Wezel, 2020). Additionally, Hagerty & Rubinov (2019) argue that people from low-income countries may be more vulnerable to discriminatory AI systems due to their race, religion, gender and sexuality due to inherent biases in the

data. Additionally, given the computing power required by GenAI systems, additional attention is required towards ensuring equal resource allocation and waste reduction to ensure developing nations do not suffer from environmental consequences they have not been responsible for (Mannuru et al., 2023). It is vital to ensure that all nations are shaping ethical AI policies to benefit from local contexts and to drive “global fairness” (Jobin et al., 2019).

In summary, the rapid adoption of GenAI has brought significant benefits in productivity and efficiency but has also intensified ethical challenges as global powers lead the conversation in the context of competition vs. cooperation. Key issues include transparency, fairness, data privacy, and the risk of bias, with global disparities in research and regulation further complicating the landscape. As the leading powers prioritise both innovation and regulation to various extents, it remains essential to develop and enforce ethical frameworks that address these concerns, ensure accountability, and promote equitable outcomes for all nations.

References:

Abiri, G. & Huang, Y. (2023) A red flag? China’s generative AI dilemma. *Harvard journal of law & technology*, *37*: 1-20.

Bright, J., Enock, F.E., Esnaashari, S., Francis, J., Hashem, Y. & Morgan, D. (2024) Generative AI is already widespread in the public sector. *arXiv preprint arXiv:2401.01291*.

Corrêa, N.K., Galvão, C., Santos, J.W., Del Pino, C., Pinto, E.P., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E. & de Oliveira, N. (2023) Worldwide

AI ethics: A review of 200 guidelines and recommendations for AI governance. Patterns, 4(10).

Draux, H. (2024) Research on Artificial Intelligence – the global divides. Digital Science. Available from: https://www.digital-science.com/tldr/article/research-on-artificial-intelligence-the-global-divides/ [Accessed 3 April 2025]

Elliott, L. (2025) The west is already losing the AI arms race. The Guardian. Available from: https://www.theguardian.com/commentisfree/2025/jan/30/ai-arms-race-china-deepseek [Accessed 4 April 2025]

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. & Srikumar, M. (2020) Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center Research Publication*, (2020-1).

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K. & Chen, L. (2023) Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of information technology case and application research*, *25*(3): 277-304.

Hacker, P., Engel, A. & Mauer, M. (2023) Regulating ChatGPT and other large generative AI models. In *Proceedings of the 2023 ACM conference on fairness, accountability, and transparency* (1112-1123).

Hagendorff, T., (2020) The ethics of AI ethics: An evaluation of guidelines. *Minds and machines*, *30*(1): 99-120.

Hagendorff, T. & Wezel, K. (2020) 15 challenges for AI: or what AI (currently) can't do. *Ai & Society*, *35*(2): 355-365.

Hagerty, A. & Rubinov, I. (2019) Global AI ethics: a review of the social impacts and ethical implications of artificial intelligence. *arXiv preprint arXiv:1907.07892*.

Hu, K. (2023) ChatGPT sets record for fastest-growing user base - analyst note. Reuters. Available from: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/ [Accessed 1 April 2025]

Ipsos (2023) Gen AI goes mainstream. Ipsos. Available from: https://www.ipsos.com/en/almanac-2024/gen-ai-goes-mainstream [Accessed April 1 2025]

Jobin, A., Ienca, M. & Vayena, E. (2019) The global landscape of AI ethics guidelines. *Nature machine intelligence*, *1*(9): 389-399.

Leyer, M. & Schneider, S. (2021) Decision augmentation and automation with artificial intelligence: threat or opportunity for managers?. *Business Horizons*, *64*(5): 711-724.

Mannuru, N.R., Shahriar, S., Teel, Z.A., Wang, T., Lund, B.D., Tijani, S., Pohboon, C.O., Agbaji, D., Alhassan, J., Galley, J. & Kousari, R. (2023) Artificial intelligence in developing countries: The impact of generative artificial intelligence (AI) technologies for development. *Information Development*, 02666669231200628.

Meskó, B. & Topol, E.J. (2023) The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ digital medicine*, *6*(1): 120.

Perrigo, B. (2023) Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic. TIME. Available from: https://time.com/6247678/openai-chatgpt-kenya-workers/ [Accessed 5 April 2025]

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance)' (2016) Official Journal L 119, 1-88. http://data.europa.eu/eli/reg/2016/679/oj

Roberts, H. & Hine, E. (2023) The future of AI policy in China. East Asia Forum. Available from: https://eastasiaforum.org/2023/09/27/the-future-of-ai-policy-in-china/ [Accessed 5 April 2025]

Vincent, B. (2023) Pentagon CIO and CDAO: Don't pause generative AI development — accelerate tools to detect threats. DefenseScoop. Available from: https://defensescoop.com/2023/05/03/pentagon-cio-and-cdao-dont-pause-generative-ai-development-accelerate-tools-to-detect-threats/ [Accessed 5 April 2025]