Summary Post

The discussion on the ethical dimensions of deep learning revealed how interconnected issues such as sustainability, accountability, and human agency have become central to the responsible development of AI. In my original post, I highlighted three main concerns. First, the environmental cost of training large models remains substantial, with studies showing emissions comparable to multiple cars over their lifetimes (Strubell et al., 2019). This raises the need to treat efficiency as an ethical priority, not merely a technical one (McDonald et al., 2022).

Second, I discussed the problem of accountability in deep learning systems. The opacity of neural networks makes it difficult to trace responsibility when harm occurs, while governance frameworks across jurisdictions remain fragmented (Jobin et al., 2019; Liu et al., 2021). Third, I addressed the risk of over-reliance on AI tools, which can erode human creativity and decision-making—an issue that human-centred design aims to prevent (Shneiderman, 2020).

The peer responses developed these points further. Jaafar and Pavlos both expanded on the environmental perspective, connecting it to wider governance challenges and the European Union's AI Act, as well as social biases and data inequality (Bender et al., 2021; Veale and Borgesius, 2021). Pavlos also raised the concern of skill erosion, noting that dependence on AI can dull expertise across professions (Brynjolfsson and McAfee, 2014). Saeed reinforced the argument that energy efficiency is an ethical duty and suggested explainable AI as a practical step toward accountability and trust. Ahmed

echoed these points, emphasising that transparency and human agency are essential to ensure that AI remains a tool for empowerment rather than substitution.

Together, the discussion underscores that the ethics of deep learning extend beyond technical performance. Achieving sustainable, transparent, and human-aligned AI requires collective responsibility from researchers, policymakers, and society as a whole.

Reference:

Bender, E.M., Gebru, T., McMillan-Major, A. and Shmitchell, S. (2021) 'On the dangers of stochastic parrots: Can language models be too big?', *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.* Virtual Event: ACM, pp. 610–623. Available at: https://doi.org/10.1145/3442188.3445922 (Accessed: 20 October 2025).

Brynjolfsson, E. and McAfee, A. (2014) *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies.* New York: W.W. Norton.

Jobin, A., Ienca, M. and Vayena, E. (2019) 'The global landscape of AI ethics guidelines', *Nature Machine Intelligence,* 1(9), pp. 389–399. Available at: https://arxiv.org/pdf/1906.11668 (Accessed: 20 October 2025).

Liu, H., Wang, Y., Fan, W., Liu, X., Li, Y., Jain, S., Liu, Y., Jain, A.K. and Tang, J. (2021) *Trustworthy AI: A computational perspective.* arXiv preprint. Available at: https://arxiv.org/abs/2107.06641 (Accessed: 20 October 2025).

McDonald, J., Li, B., Frey, N., Tiwari, D., Gadepally, V. and Samsi, S. (2022) *Great power, great responsibility: Recommendations for reducing energy for training language models.* arXiv preprint. Available at: https://arxiv.org/abs/2205.09646 (Accessed: 20 October 2025).

Shneiderman, B. (2020) 'Human-centred artificial intelligence: Reliable, safe & trustworthy', *International Journal of Human–Computer Interaction,* 36(6), pp. 495–504. Available at: https://doi.org/10.1080/10447318.2020.1741118 (Accessed: 20 October 2025).

Strubell, E., Ganesh, A. and McCallum, A. (2019) 'Energy and policy considerations for deep learning in NLP', *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics,* pp. 3645–3650. Available at: https://aclanthology.org/P19-1355.pdf (Accessed: 20 October 2025).

Veale, M. and Borgesius, F.Z. (2021) 'Demystifying the draft EU Artificial Intelligence Act', *Computer Law Review International,* 22(4), pp. 97–112. Available at: https://doi.org/10.9785/cri-2021-220402 (Accessed: 20 October 2025**).**