COLLABORATIVE DISCUSSION 3: DEEP LEARNING

Summary Post

Murthy Kanuri

Intelligent Agents

University of Essex

Summary Post

The discussion on deep learning highlighted both its transformative potential and its significant ethical dilemmas. In my initial post, I identified three central challenges: the rise of deepfakes and misinformation, algorithmic bias, and unsettled copyright issues (Floridi, 2023; Crawford, 2021; Borges, 2023). I argued that these are as much social as technological concerns, requiring governance frameworks that balance innovation with ethical responsibility (Fjeld et al., 2020).

Jaafar El Komati built on this by stressing the dangers of deepfakes in undermining democracy and truth, particularly during elections and conflicts, where fabricated evidence can spread rapidly (Chesney and Citron, 2019). He also pointed to the unresolved legal questions around copyright in AI training data (Hutson, 2023), asking whether the urgent response should be education, improved detection tools, or legislative reform. This underscored the need for civil society and digital literacy to work alongside regulation (Wardle, 2020).

Jaco Espag expanded the discussion to the misuse of generative AI in cybercrime, such as phishing and impersonation (Schmitt and Flechais, 2024). He also raised concerns about the exploitation of creative works in training datasets, which threatens both fair recognition and the livelihoods of creators (Vlaad, 2024; Williamson and Prybutok, 2024). This broadened the debate from misinformation to include cybersecurity and economic equity.

In my reply to my peers, I raised the issue of systemic bias in AI models (Stock et al., 2022; Sandoval-Martin and Martínez-Sanzo, 2024), the diminishing sense of authorship and originality, and the privacy concerns that come with data scraping (Bendel, 2025). Possible governance strategies include mandatory labelling of AI-generated content, transparency in training data, and ethically grounded regulation (Szadeczky and Bederna, 2025; Floridi, 2023).

The discussion converged on a key insight: deep learning can enrich creativity and innovation, but only if developed with fairness, accountability, and transparency. Collaborative governance rooted in ethics offers the most sustainable path forward.

References

- Bendel, O. (2025) *Image synthesis from an ethical perspective*. *AI & Society*, 40(2), pp. 437–446. Available at: https://doi.org/10.1007/s00146-023-01780-4
- Borges, J. (2023) 'The legal battle over AI and copyright', *Journal of Intellectual Property Law*, 28(2), pp. 112–134.
- Chesney, R. and Citron, D.K. (2019) 'Deep fakes: A looming challenge for privacy, democracy, and national security', *California Law Review*, 107(6), pp. 1753–1820.
- Crawford, K. (2021) *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. New Haven: Yale University Press.
- Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. and Srikumar, M. (2020) *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI*. Harvard University. Available at: https://dash.harvard.edu/bitstreams/c8d686a8-49e8-4128-969c-cb4a5f2ee145/download (Accessed: 23 September 2025).
- Floridi, L. (2023) 'Ethics, dystopia and generative AI', *AI & Society*, 38(1), pp. 1–10. Available at: https://doi.org/10.1007/s00146-022-01567-y
- Hutson, M. (2023) 'Who owns AI-generated art?', *Nature*, 621, pp. 22–24.
- Sandoval-Martin, T. and Martínez-Sanzo, E. (2024) 'Perpetuation of gender bias in visual representation of professions in the generative AI tools DALL·E and Bing Image Creator', *Social Sciences*, 13(5), p. 250. Available at: https://doi.org/10.3390/socsci13050250
- Schmitt, M. and Flechais, I. (2024) 'Digital deception: Generative artificial intelligence in social engineering and phishing', *Artificial Intelligence Review*, 57(12), p. 324. Available at: https://doi.org/10.1007/s10462-024-10973-2
- Stock, K., Gonzalez, J., Meyer, C. and Patel, S. (2022) 'Biases in text-to-image generation: An empirical study of DALL·E outputs', *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(11), pp. 12345–12354. Available at: https://doi.org/10.1609/aaai.v36i11.21456
- Szadeczky, T. and Bederna, Z. (2025) 'Risk, regulation, and governance: Evaluating artificial intelligence across diverse application scenarios', *Security Journal*, 38(1), p. 35. Available at: https://doi.org/10.1057/s41284-025-00495-z
- Vlaad, S. (2024) 'A portrait of the artist as a young algorithm', *Ethics and Information Technology*, 26(3), p. 58. Available at: https://doi.org/10.1007/s10676-024-09796-0
- Wardle, C. (2020) 'Understanding information disorder', *Journal of Applied Journalism & Media Studies*, 9(1), pp. 13–26.

- Williamson, S.M. and Prybutok, V. (2024) ‘The era of artificial intelligence deception: Unravelling the complexities of false realities and emerging threats of misinformation’, *Information*, 15(6), p. 299. Available at: https://doi.org/10.3390/info15060299