COLLABORATIVE DISCUSSION 3: DEEP LEARNING

Initial Post

Murthy Kanuri

Intelligent Agents

University of Essex

Initial Post

The rapid development of deep learning models has brought about an era where "new" content can be produced with unprecedented ease, as demonstrated by tools like DALL·E and ChatGPT. While these technologies are revolutionary, they also raise important ethical dilemmas that require a thoughtful and cooperative response.

Deepfakes and false information present a growing challenge. AI can produce convincing text, images, and videos that are hard for humans to tell apart (Floridi, 2023). Addressing this requires strong digital literacy and enhanced detection technologies to differentiate between authentic and fabricated content (Pei et al., 2024; Gupta et al., 2024).

Additionally, algorithmic bias is another significant challenge. When AI systems are trained on data that reflects past inequalities, those biases can affect the predictions. The consequences are particularly troubling in sensitive areas like hiring, Judicial proceedings, or access to financial services (Crawford, 2021; Rigotti & Fosch-Villaronga, 2024).

Copyright and intellectual property are also problems that have yet to be solved. However, the concept of fair use for the original creator remains unclear when training models on large amounts of copyrighted material (Borges, 2023). We must address ownership of AI-generated content and fair recognition and compensation for writers and artists who contribute to the training of these technologies.

Ultimately, these are as much sociological as technological issues. Building frameworks for ethical innovation that support common values like truth, fairness, and creativity requires collaboration between governments, business, and civil society (Fjeld et al., 2020).

References

- Borges, J. (2023). The legal battle over AI and copyright. *Journal of Intellectual Property Law*, **28**(2), pp. 112–134.
- 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. & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI*. [Report]. 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.
- Gupta, G., Raja, K., Gupta, M., et al. (2024). A comprehensive review of deepfake detection using advanced machine learning and fusion methods. *Electronics*, **13**(1), p. 95. Available at: https://doi.org/10.3390/electronics13010095.
- Pei, G., Zhang, J., Hu, M., et al. (2024). *Deepfake generation and detection: A benchmark and survey*. arXiv preprint. arXiv:2403.17881. Available at: https://arxiv.org/abs/2403.17881 [Accessed 23 September 2025].
- Rigotti, C. & Fosch-Villaronga, E. (2024). Fairness, AI & recruitment. *Computer Law & Security Review*, **53**, p. 105073. Available at: https://doi.org/10.1016/j.clsr.2024.105073.