

COLLABORATIVE DISCUSSION 3: DEEP LEARNING

Peer Response
Murthy Kanuri
Intelligent Agents
University of Essex

Table of Contents

<i>Peer Response from Jaafar El Komati</i>	<i>3</i>
<i>Peer Response from Jaco Espag</i>	<i>4</i>
<i>Response to Jaafar El Komati.....</i>	<i>5</i>
<i>Response to Jaco Espag</i>	<i>6</i>

Peer Response from Jaafar El Komati

Murthy, your post really resonated with me, especially the way ethical concerns were linked to sociological concerns. I think the one regarding deepfakes is really pressing. We've seen how fast manipulated content spreads during elections and wars, and with AI now capable of producing realistic "evidence," the very nature of truth is open to attack (Chesney & Citron, 2019).

Your description of copyright as unsettled ground also holds true. The "fair use" question of AI training data is hot and courts are only just beginning to figure out what sort of compensation might be paid to creators (Hutson, 2023). It is one where policy will be likely to lag behind technology unless governments intervene quickly.

I also appreciated your insistence on collaboration. Regulating AI is not something that can be entrusted to the technology sector solely. As you said, civil society has to be brought into the equation. Indeed, some researchers argue that it is as vital to create digital literacy among the population as to enhance detection capabilities—because an educated population is more able to resist manipulation (Wardle, 2020).

Your entry leaves me wondering: what do you believe would most likely be the first realistic action—tighter copyright legislation, better detection tech, or education?

References

- Chesney, R. & Citron, D.K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), pp. 1753–1820.
- Hutson, M. (2023). Who owns AI-generated art? *Nature*, 621, 22–24.
- Wardle, C. (2020). Understanding information disorder. *Journal of Applied Journalism & Media Studies*, 9(1), pp. 13–26.

Peer Response from Jaco Espag

I fully agree with your assessment. While these generative artificial intelligence (AI) tools offer exciting possibilities, their use could lead to ethical situations that are hard to ignore. The ability to generate convincing content may provide those without the ability to do so a valuable tool, but it is the grey or, in some cases, even criminal areas that are of great concern.

This ability to generate content that is highly convincing is eroding public trust in what we read and see online (Williamson and Prybutok, 2024). While the issue of not being able to trust online content is problematic, there is also the risk that criminals will use the same technology to impersonate, social engineer or otherwise coerce individuals in an attempt to circumvent cybersecurity systems. As noted by Schmitt and Flechais (2024), the incidents of using generative AI in phishing scams are on the rise.

Along with the growth of our collective online presence, the internet has become a source of income for many. In many cases, these works require talent or other creative abilities that may not be commonplace. Training models on this content and allowing anyone to produce similar content may destroy these creatives' source of income and potentially limit the entry of new talent into the market due to fear of plagiarism. As noted by Vlaad (2024), those who create are often not credited or compensated for their work being used to train these models.

References

- 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>.
- 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>.
- Williamson, S.M. and Prybutok, V. (2024) 'The Era of Artificial Intelligence Deception: Unraveling the Complexities of False Realities and Emerging Threats of Misinformation', *Information*, 15(6), p. 299. Available at: <https://doi.org/10.3390/info15060299>.

Response to Jaafar El Komati

The essay by Jaafar El Komati raises important ethical issues related to generative AI, specifically those pertaining to authorship, bias, and abuse. Ownership is a major concern since models that have been trained on large datasets frequently use creative works without permission. This illustrates the current legal controversies of whether work done by AI can be considered derivative or even original (Borges, 2023; ScienceTimes, 2025). These issues strongly underscore the urgent requirement for updated copyright standards.

Systemic bias in AI systems is also a concern. As discussed above, models like OpenAI Sora exemplify where and how bias can be exacerbated and disseminated (Wired, 2025). This is consistent with other research that also finds AI applications have a notable impact on fairness and transparency, especially in decision-making, law enforcement and recruitment (Rigotti & Fosch-Villaronga, 2024; Fjeld et al., 2020). Poorly governed AI is as likely to reinforce disparities as to diminish them.

Misuse presents the gravest danger. The advent of deepfakes (i.e., media that seem authentic but are manipulatively generated content) and the known generation of malign content through AI underpin how rapidly such innovations may become weaponised (AP News, 2024; Gupta et al., 2024; Pei et al., 2024). These dangers are not only the stuff of theory, but immediate societal challenges.

Nevertheless, there is still room for AI to enhance accessibility and originality. For thinkers like Floridi (2023), an ethically principled governance strategy is also crucial to guarantee that innovation at significant works for rather than against human flourishing. So, in addition to regulation, accountability, and transparency, ongoing collaboration between governments, developers, and researchers will allow us to see that generative AI grows in a way that fosters creativity while keeping society safe from harm.

References

- AP News (2024) AI-generated child sexual abuse images are spreading. AP News. Available at: <https://apnews.com/article/42186aaf8c9e27c39060f9678ebb6d7b> [Accessed 2 October 2025].
- Borges, J. (2023) 'The legal battle over AI and copyright', *Journal of Intellectual Property Law*, 28(2), pp. 112–134.
- 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*. [Report]. Harvard University. Available at: <https://dash.harvard.edu/bitstreams/c8d686a8-49e8-4128-969c-cb4a5f2ee145/download> [Accessed 2 October 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> [Accessed 2 October 2025].
- 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> [Accessed 2 October 2025].

- 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 2 October 2025].
- Rigotti, C. and 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> [Accessed 2 October 2025].
- ScienceTimes (2025) 'Ethics in generative AI: addressing bias, misinformation, and intellectual property challenges', ScienceTimes. Available at: <https://www.sciencetimes.com/articles/60309/20250223/ethics-generative-ai-addressing-bias-misinformation-intellectual-property-challenges.htm> [Accessed 2 October 2025].
- Wired (2025) 'OpenAI's Sora is plagued by sexist, racist, and ableist biases', Wired. Available at: <https://www.wired.com/story/openai-sora-video-generator-bias> [Accessed 2 October 2025].

Response to Jaco Espag

Jaco Espag spotlights several serious ethical concerns surrounding generative AI, including how it influences bias, questions of authorship, and its potential for misuse. Bias is perhaps the most pressing issue, since stereotypes encoded in AI-generated work could exacerbate social divisions. Research indicates that models like DALL·E frequently replicate damaging stereotypes in their outputs without corrective mechanisms (Stock et al., 2022; Sandoval-Martin and Martínez-Sanzo, 2024). Recruitment has faced a similar issue of copying or even entrenching prejudices within society in AI systems (Rigotti and Fosch-Villaronga, 2024).

The matter of authorship and originality is also key. AI's increasing presence in cultural production, as Espag points out, presents challenging questions about things such as credit, fair payment for labour and authenticity. Vlaad (2024) raises comparable considerations and suggests that the more human creativity is progressively entangled with mechanical processes, the harder it is to assess the value of artistic production or human society (where art is concerned). Faith in digital media is also being eroded by generative AI's capacity for creating photorealistic but false or defamatory material (Bendel, 2025).

Privacy risks compound these challenges, since training datasets are often scraped in bulk from online sources without consent. This raises important questions about the fair use of individuals' images and creative contributions.

Espag is right to emphasise that openness, regulation, and human control are essential to minimising damage. Mandatory labelling of AI-authored content and restrictions on its use in defined contexts may mitigate misuse (Szadeczky and Bederna, 2025). Not just any governance model will do, but as Floridi (2023) stresses, only governance models rooted in ethics can ensure that generative AI will enrich human life, not undermine it.

Ultimately, innovation must be balanced with ethical responsibility. Generative AI could revolutionise human creative expression, but only if we can trust the systems behind it to align with our values and protect what makes us fundamentally human.

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> [Accessed 2 October 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> [Accessed 2 October 2025].
- Rigotti, C. and 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> [Accessed 2 October 2025].
- 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> [Accessed 2 October 2025].
- 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> [Accessed 2 October 2025].
- 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> [Accessed 2 October 2025].
- 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> [Accessed 2 October 2025].