

## Peer Response to Ali Alzahmi

Ali, your post addresses ethical challenges involving deep learning systems, such as accountability, cultural appropriation, and authenticity. Several measures involving regulations, design, and institutional frameworks could help alleviate the challenges you mentioned.

First, the implementation of accountability frameworks should be mandatory. There should be accountability measures at every step of the AI development lifecycle. This includes the principles of explainability-by-design and the creation of model documentation (model cards or data sheets for datasets) that describes the data collection, usage, and auditing processes (Mitchell et al., 2019). Legal frameworks surrounding accountability, especially harmful and biased outputs, could close the gaps of liability ambiguity.

Finally, the appropriation of culture calls for the construction of more inclusive and representative datasets. Cross-border ethical boards and community engagement in dataset construction and creation could help to respect the contextual integrity of the datasets and the artistic and linguistic elements. Protections should be built around the use of art and language. Even the use of dissenting participatory governance and culturally ignorant documentation should be mitigated. Exploitation of indigenous and minority art and language should be sanctioned (UNESCO, 2023).

The public digital literacy initiatives I have suggested are meant to help users tell the difference between human and AI-generated content while appreciating the work and creativity of humans. The Content Authenticity Initiative and Adobe (2024) proposed some trust and authenticity preservation tools that include watermarking and verification.

Governments can use ethical auditing, transparent governance, and involving communities to promote human-centric values rather than stifling innovation.

## References

- Adobe (2024) *Content Authenticity Initiative: Building trust in digital content*. Available at: <https://contentauthenticity.org/> (Accessed: 8 October 2025).
- Mitchell, M. et al. (2019) 'Model cards for model reporting', *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT)\**, pp. 220–229. Available at: <https://doi.org/10.1145/3287560.3287596> (Accessed: 8 October 2025).
- UNESCO (2023) *Recommendation on the Ethics of Artificial Intelligence*. Paris: UNESCO. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000381137> (Accessed: 8 October 2025).

## Peer Response to Hashim Korimbocus

Hashim, you highlight critical ethical challenges surrounding deep learning, specifically concerning bias, authorship, misinformation, and disruption of creativity. Each of these challenges has specific preventive strategies that prioritize transparency and accountability at every stage of AI development and integration.

To counter algorithmic bias, developers should focus on bias mitigation and inclusive dataset construction. Approaches such as Google's "Model Cards" and IBM's "AI Fairness 360 Toolkit" offer accountability about data origins and facilitate scrutiny of output discrimination over time (Mitchell et al., 2019). Furthermore, the formation of diverse annotation teams and the practice of "stereotype" impact assessments at the regular intervals will help avoid entrenched stereotypical biases in training datasets (Bender et al., 2021).

Concerning authorship and originality, proposing ethical data licensing frameworks, which offer equitable attribution and compensation to dataset contributors, will help address concerns. The Creative Commons framework and the EU's AI Act (2024) both advocate ownership clarity through consent-based data collection practices and AI output watermarking.

To address misinformation and deepfakes, the use of digital provenance methods, like the Content Authenticity Initiative, will verify legitimate media origins and identify AI-induced alterations (Adobe, 2024). Coupling this with the promotion of critical media literacy will encourage users to analyse various digital materials.

Ultimately, as you point out, the most important thing is embedding responsible innovation principles—ensuring that the advancement of AI is augmenting human creativity and trust, rather than eliminating it.

## References

- Adobe (2024) Content Authenticity Initiative: Building trust in digital content. Available at: <https://contentauthenticity.org/> (Accessed: 8 October 2025).
- Bender, E.M. et al. (2021) 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?', Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, pp. 610–623.
- European Union (2024) Artificial Intelligence Act. Available at: <https://digital-strategy.ec.europa.eu/en/policies/european-ai-act> (Accessed: 8 October 2025).
- Mitchell, M. et al. (2019) 'Model cards for model reporting', Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT)\*, pp. 220–229. Available at: <https://doi.org/10.1145/3287560.3287596> (Accessed: 8 October 2025).

## Peer Response to Abdulla Alshaibani

Abdulla, you've mentioned the ethical issues pertaining to deep learning technology, especially the issues of misinformation, privacy of personal information, intellectual property, and sustainability. The problem could go away, or at the very least, be minimized with the provision of appropriate technical safeguards, statutory controls, and, of course, cooperative governance at the broader level.

Since deepfakes pose a serious threat to misinformation, policy and technology developers should focus on implementing smarter content authentication strategies like the C2PA (Coalition for Content Provenance and Authenticity) standards. These standards embed metadata into artificial content to prove its authenticity. With such technology, people could be able to tell fake content from real content (Adobe, 2024). Other balanced media literacy programs would help people learn to take a critical approach and assess content more deeply.

With regard to the regulation of intellectual property, more stringent governance over data and the setting of appropriate privacy controls will be necessary to reduce unauthorized document access and use. Transparency will come with the adoption of ethical data use agreements, licensing arrangements, and open access copyrights (Henderson et al., 2023). These arrangements would help solve the problem of copyrighted material unauthorized use.

To protect sensitive information, control over disclosure of personal information during model training can be solved with federated learning and differential privacy. Those methods will allow models to learn from decentralised datasets and not to reveal personal data (Abadi et al., 2016).

Lastly, there is the need to implement eco-friendly approaches to Artificial Intelligence training, develop standards around carbon accountability, and implement effective systems around the aforementioned components. The establishment of "Green AI" standards could guarantee the aforementioned eco-friendliness regarding the accounting of computational resources (Schwartz et al., 2020).

As a last comment, embedding the Responsible Innovation Framework you mentioned into the legal, educational, and technical systems you outlined would help guarantee the ethical and sustainable evolution of deep learning.

## References

- Abadi, M. et al. (2016) 'Deep learning with differential privacy', *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 308–318. Available at: <https://doi.org/10.1145/2976749.2978318> (Accessed: 8 October 2025).
- Adobe (2024) *Content Authenticity Initiative: Building trust in digital content*. Available at: <https://contentauthenticity.org/> (Accessed: 8 October 2025).
- Henderson, P. et al. (2023) 'Towards datasets governance in machine learning', *Nature Machine Intelligence*, 5(3), pp. 228–240. Available at: <https://doi.org/10.1038/s42256-022-00576-5> (Accessed: 8 October 2025).
- Schwartz, R. et al. (2020) 'Green AI', *Communications of the ACM*, 63(12), pp. 54–63. Available at: <https://doi.org/10.1145/3381831> (Accessed: 8 October 2025).