Ethical Implications of Deep Learning Technologies

Initial Post

by Pavlos Papachristos - Saturday, 18 October 2025, 1:49 AM

Number of replies: 0

Deep-learning tools such as DALL-E and ChatGPT raise a number of ethical questions that deserve serious attention. Their ability to create convincing text, images, and other media at scale is changing the way information circulates, with consequences that reach far beyond the purely technical.

One of the hardest problems concerns accountability when AI-generated material causes harm. As the accompanying research observes, large language models can display *emergent behaviours that even their designers cannot always anticipate* (document, p. 1). When such systems spread false information or reproduce copyrighted material, it becomes unclear who should be held responsible—the developers, those who deploy the systems, or the end users (Dignum, 2019). The debate over AI-generated artwork winning creative competitions illustrates this confusion: if the image comes from an algorithm, who can really claim authorship?

Generative models also intensify the problem of misinformation. They can create coordinated disinformation or social-engineering campaigns at a speed and scale that human producers cannot match (Goldstein et al., 2023). Because these systems generate content tailored to individual contexts, fact-checking efforts often lag behind. The resulting flood of persuasive falsehoods threatens democratic decision-making, public-health messaging, and basic trust in shared information sources.

The document notes that *job displacement is becoming a genuine concern rather than just speculation* (document, p. 2). Creative workers—writers, designers, illustrators—face particularly strong pressure as generative tools encroach on their professions. Optimists suggest that automation will release people from routine tasks and open space for more meaningful work, but that view tends to overlook the short-term upheaval and growing inequality that may follow (Eloundou et al., 2023).

Training these systems requires enormous datasets, often assembled from publicly available material without explicit permission. This raises important questions about data rights and personal consent: should individuals have a say in how their images, words, or artistic styles are used to build commercial AI models? (Weidinger et al., 2023).

The ethical issues linked to generative AI are already here, not distant possibilities. Meeting them will require clear governance, greater transparency in how models are built, and open dialogue among researchers, policymakers, and the public. Only with such structures in place can these technologies be guided toward genuinely collective rather than narrowly commercial ends.

References

- Dignum, V. (2019) *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Cham: Springer.

- Eloundou, T., Manning, S., Mishkin, P. and Rock, D. (2023) 'GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models', *arXiv preprint* arXiv:2303.10130.

- Goldstein, J.A., Chao, J., Grossman, S., Stamos, A. and Tomz, M. (2023) 'How Persuasive is AI-Generated Propaganda?', *arXiv preprint* arXiv:2301.05291.

- Weidinger, L. et al. (2023) 'Taxonomy of Risks posed by Language Models', in *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. Chicago, IL: ACM, pp. 214-229.

Reply to Forum

Initial Post

by Martyna Antas - Sunday, 28 September 2025, 8:10 PM

Number of replies: 2

The rise of deep learning has enabled systems such as DALL·E and ChatGPT to generate text and images with remarkable ease. While these technologies enhance creativity and productivity, they also present ethical concerns requiring scrutiny.

One pressing issue is the environmental impact of training and deploying large models. Training a single large natural language processing model can emit as much carbon dioxide as five cars over their lifetimes (Strubell et al., 2019). Even the fine-tuning of smaller models consumes significant energy, suggesting sustainability remains a challenge across the pipeline (Wang et al., 2023). It has been argued that researchers and practitioners must reduce the carbon footprint of model training, with efficiency treated as an ethical responsibility (McDonald et al., 2022).

Another challenge is accountability and transparency. There is little consensus on governance across global AI ethics guidelines, particularly regarding responsibility when harm occurs (Jobin et al., 2019). The black box nature of deep learning undermines explainability, making it difficult for users or regulators to trace outputs (Liu et al., 2021). Without robust oversight, accountability for misuse such as misinformation or harmful content remains unclear.

There is also the risk of over reliance on generative AI. While AI can enhance trust when designed responsibly, excessive dependence may erode human judgement and decision-making (Liu et al., 2021). Creativity and problem-solving may decline if human contributions are continually replaced by automated outputs (Shneiderman, 2020).

Finally, these concerns are not merely technical but socio technical in nature. Principles such as fairness, transparency and accountability are best achieved through collaborative frameworks involving governments, industry and civil society (Fjeld et al., 2020). Ethical deep learning therefore requires both technical safeguards and wider governance measures.

In conclusion, while deep learning technologies open exciting opportunities, they also raise ethical questions related to sustainability, accountability and human agency. Addressing these challenges requires transparency, regulation and public dialogue to ensure innovation aligns with societal values.

References:

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/handle/1/42160420 (Accessed: 28 September 2025)

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: 28 September 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: 28 September 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: 28 September 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: 28 September 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: 28 September 2025)

Wang, X., Na, C., Strubell, E., Friedler, S. and Luccioni, S. (2023) *Energy and carbon considerations of fine-tuning BERT*, arXiv preprint. Available at: **https://arxiv.org/abs/2311.10267** (Accessed: 28 September 2025)

In reply to Martyna Antas

Peer response

by Jaafar El Komati - Wednesday, 1 October 2025, 9:13 AM

Martina, your post was standing out because you used to discuss environmental dimensions - a concern that often looks at prejudice or misinformation debate. The fact that training a large NLP model can be equal to the lifetime emissions of many cars (Strubell et al., 2019). It shows AI morality not only as a social issue but also as a climate.

I also got your point about accountability compelling. The "black box" problem is not just technical; This reduces faith. If users cannot understand why the AI system takes a decision, even the appropriate results may still feel illegitimate (Burrell, 2016). It makes sense to connect it with governance, as we are looking at initial experiments with AI regulation in the AI Act of the European Union, although critics argue that it still struggles with enforcement and global scope (Veale & Borgesius, 2021).

Your reflection on dependence on AI was actually a hit at home as well. It is attractive to tilt on tools such as chatgpt for every task, but as you mentioned, it risk reducing human creativity. I think a combination of human input with AI aid-Shneiderman (2020) says "human-focused AI"-we should be balanced that we need.

Your closing call seems significant for the associate governance. Without shared standards, stability and accountability are risk patchwork solutions.

References

- Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), 1–12.

- Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *Proceedings of ACL*, 3645–3650.

- Veale, M., & Borgesius, F. Z. (2021). Demystifying the draft EU Artificial Intelligence Act. *Computer Law Review International, 22*(4), 97–112.

- Shneiderman, B. (2020). *Human-centered AI*. Oxford University Press.

In reply to Martyna Antas

Re: Initial Post

by Pavlos Papachristos - Saturday, 18 October 2025, 1:57 AM

Your essay raises important issues that don't always get the attention they deserve, especially the environmental toll of deep-learning systems. The carbon-emission figures you reference are striking. Public debate tends to revolve around misinformation or creative misuse, yet the physical side of AI—the data centres, the cooling, the constant power draw—warrants the same level of ethical concern. As

models keep ballooning in size, with GPT-4 reportedly using far more parameters than its predecessors (OpenAI, 2023), those concerns become impossible to dismiss.

Accountability also extends beyond environmental impact. A quieter but equally troubling issue is bias amplification. Any model trained on the open internet inevitably mirrors the social hierarchies and prejudices embedded in that data (Bender et al., 2021). Each time DALL-E produces an image or ChatGPT writes a passage, traces of those inherited assumptions may surface. The real question, then, is whose stories and values get reinforced when these systems speak on our behalf. Your observations on over-reliance touch another nerve: skill erosion. As Brynjolfsson and McAfee (2014) warned, routine dependence on intelligent systems can dull human expertise. The phenomenon reaches well beyond the creative fields. When lawyers offload document review or clinicians rely on algorithmic triage, their ability to judge critically may fade over time. That gradual loss of intuition and context awareness is easy to miss until it's too late. Governance, meanwhile, remains the thorniest challenge.

In theory, multi-stakeholder frameworks sound promising; in practice, they run into stark power asymmetries. The major AI labs still make key design choices behind closed doors, citing trade secrets while influencing public life on a massive scale (Whittaker et al., 2018). Transparency helps, but true accountability requires checks that can counterbalance private interests when they conflict with the public good. Your conclusion gets to the core of it. These systems are not neutral tools but socio-technical infrastructures that shape behaviour, work, and even imagination. Meeting their ethical challenges will take more than clever regulation. It calls for sustained collaboration among engineers, policymakers, and citizens to ensure that AI development serves shared human ends rather than narrow economic goals.

References

- Bender, E.M., Gebru, T., McMillan-Major, A. and Shmitchell, S. (2021) 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?', in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. Virtual Event: ACM, pp. 610-623.

- 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.

- OpenAI (2023) 'GPT-4 Technical Report', arXiv preprint arXiv:2303.08774. –

- Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., West, S.M., Richardson, R., Schultz, J. and Schwartz, O. (2018) AI Now Report 2018. New York: AI Now Institute.

Initial post

by Ali Yousef Ebrahim Mohammed Alshehhi - Monday, 13 October 2025, 7:39 PM

Number of replies: 1

The rapid progress of deep learning technologies has transformed how we produce and use content. Generative output models like DALL·E for image creation and ChatGPT for natural language can provide highly authentic and original content that often mimics human creativity. However, while these technologies offer some creativity, they come with a myriad of ethical questions we must navigate carefully. One key ethical concern revolves around authenticity and misinformation. Deep learning can generate content that creates legitimate confusion between real and artificial information. As an illustration, deepfake technology can produce realistic but completely fabricated video or image media that can cause reputational harm, or alter public perception (Chesney and Citron, 2019). A second ethical issue is bias / fairness. These models are often trained on large datasets that may be impacted by historical / societal biases that these same models perpetuate in their outputs (Buolamwini and Gebru, 2018).

As a final note, there are also questions regarding intellectual property / ownership. If an AI generates example art based on existing data, it is unclear who holds the copyright, the user, the developer of the model, or the owners of the training data (Floridi and Chiriatti, 2020). There may also be privacy concerns, as models trained on personal data may inadvertently expose personal or sensitive information.

Thus, while deep learning has much to offer the potential user, it is important to have ethical considerations and regulatory measures in place to ensure responsible use. By addressing these items, it may be possible to balance innovation with societal trust and safety.

References

Buolamwini, J. & Gebru, T. (2018) 'Gender shades: Intersectional accuracy disparities in commercial gender classification', *Proceedings of Machine Learning Research*, 81, pp. 1–15.

Chesney, R. & Citron, D.K. (2019) 'Deep fakes: A looming challenge for privacy, democracy, and national security', *California Law Review*, 107(6), pp. 1753–1819.

Floridi, L. & Chiriatti, M. (2020) 'GPT-3: Its nature, scope, limits, and consequences', *Minds and Machines*, 30(4), pp. 681–694.

In reply to Ali Yousef Ebrahim Mohammed Alshehhi

Replying to Ali Yousef Ebrahim Mohammed Alshehhi

by [Pavlos Papachristos](#) - Saturday, 18 October 2025, 2:22 AM

Generative AI creates ethical problems that go far beyond deepfakes. There's a growing issue we might call "synthetic reality pollution"—basically, there's now so much AI-generated content online that people struggle to figure out what's real and what's fake (Goldstein et al., 2023). When you can't trust what you're seeing anymore, it erodes confidence in everything online. The bias problem makes things even worse.

When AI systems produce skewed images or biased text, that content often ends up being used to train the next generation of AI models, which then repeat the same problems (Bender et al., 2021). Take image generators, for example—if they keep showing company leaders as white men, they're not just reflecting existing stereotypes, they're actively reinforcing them for everyone who uses these tools. Then there's the messy question of who owns what. Artists are now taking companies like Stability AI and Midjourney to court, arguing that their copyrighted artwork was used without permission to train these AI systems (Butterick, 2023).

The courts are still trying to work out whether this counts as fair use, and whether creators deserve credit or compensation. Perhaps most troubling is that when something goes wrong—when ChatGPT gives dangerous advice or DALL-E creates harmful images—it's genuinely unclear who should be held responsible. Is it the company that built the model? The organization deploying it? Or the person who used it? (Dignum, 2019). As Fjeld et al. (2020) point out, we can't wait for problems to happen before we act.

We need governments, companies, and communities working together now to make sure AI develops in ways that actually serve people, not just profit.

References

- Bender, E.M., Gebru, T., McMillan-Major, A. and Shmitchell, S. (2021) 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?', in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. Virtual Event: ACM, pp. 610-623.

- Butterick, M. (2023) 'Stable Diffusion Litigation'. Available at: https://stablediffusionlitigation.com (Accessed: 18 October 2025). - Dignum, V. (2019) Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way. Cham: Springer.

- 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', Berkman Klein Center Research Publication, 2020-1.

- Goldstein, J.A., Chao, J., Grossman, S., Stamos, A. and Tomz, M. (2023) 'How Persuasive is AI-Generated Propaganda?', arXiv preprint arXiv:2301.05291.