Initial Post

by Saleh Almarzooqi - Friday, 3 October 2025, 4:51 AM

In the recent developments of Deep Learning, intelligent agents have emerged with even greater potential to generate completely new text, images, and even sound. DALL-E and ChatGPT are just two examples of how algorithms trained on massive datasets can be used to automate creative work previously considered the preserve of humans (**Leslie, David, et al.** 2024). Although such systems have a lot of potential, they present some urgent ethical concerns.

One of the key problems is related to veracity and manipulation. Since models can produce content that appears or sounds real, this also means the risk of fake news, forged images, or misleading messages being propagated at a mass level (Floridi, 2023). This will erode trust in the populace and complicate the ability to fact-check information sources.

The second problem is the embedded bias. Models are copied by their training material and have the potential to perpetuate stereotypes or discriminatory actions without intent to do so by generating outputs (Bender et al., 2021). These prejudices can be devastating in such fields as hiring, health, or justice.

The ownership and consent are another issue. Generators of images and text may recreate styles, pieces of data, or copyrighted content without any form of attribution, which poses challenging questions concerning intellectual property and fair remuneration (Sahil et al. 2024).

Last but not least is privacy. Big data can be filled with personal data, which can be repeated in model outputs unless there are effective security measures (**Shahzad, Aamir, et al.** 2023).

The concerns can be resolved through the enforcement of transparency, labelling of AI-generated content, stringent data governance, and incorporation of ethical considerations into design and implementation (Ienca, Buchholz, and Vayena, 2025). By doing so, the gains of Deep Learning will be achieved without undermining the people and their rights.

References

Bender, E.M., Gebru, T., McMillan-Major, A., and Shmitchell, S., 2021, March. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623).

Floridi, L., 2023. AI as agency without intelligence: On ChatGPT, large language models, and other generative models. *Philosophy & technology*, *36*(1), p.15.

Ienca, M., Buchholz, O. and Vayena, E., 2025. AI Ethical Principles: The Debate. *A Companion to Digital Ethics*, pp.101-112.

Leslie, D., Rincon, C., Briggs, M., Perini, A., Jayadeva, S., Borda, A., Bennett, S.J., Burr, C., Aitken, M., Katell, M., and Fischer, C., 2024. AI ethics and governance in practice: An introduction. *arXiv preprint arXiv:2403.15403*.

Sahil, Pratham, Neha, Sadotra, P., Chouksey, P., Chopra, M., and Thakur, A., 2024, March. Generative AI for the Creation of Images. In *International Conference on Information Technology* (pp. 193-207). Singapore: Springer Nature Singapore.

Shahzad, A., Kayani, H.U.R., Malik, A.A., Raza, M.A., and Saleem, A., 2023. Big data security, privacy protection, tools, and applications. *Pakistan Journal of Science*, *75*(02), pp.353-372.