AI Snake Oil

Chapter 4 Part II - The Long Road to Generative AI

This Chapter continues the discussion on Generative AI with a focus on its history, including how the same base algorithms used for images can be applied to textual training data, and how the models moved from being primarily research tools to becoming chatbots that the general public can use. While the benefits are mentioned, there is more attention given to the risks, including being bullshit generators, the troublesome deepfake use cases, and the cost of improvement.

Some questions to consider for this section.

- This chapter goes into some depth describing how GPTs work, implying this will help users avoid Snake Oil. Is this a sound conclusion to draw? Arthur C Clarke famously said that "*Any sufficiently advanced technology is indistinguishable from magic*". Will the general public appreciate what a GPT is really doing behind the scenes, or will it appear to be 'magic'?

- The authors state that informing people that deepfakes and deceptive information exist will be sufficient warning to people of the dangers of trusting the Internet. Yet, when even news sources like CNET are guilty of publishing misleading and incorrect information, is this enough? How will people be able to tell what to trust and what is suspect when pretty much everything *could* be suspect? Do you agree that this may lead to "*The Liar's Dividend"?*

- This chapter once more raises the risks of copyright infringement, the ethics of training on other people's data, the political implications of electoral interference, and the exploitation of cheap labour in the categorisation of training data. Discuss.

- Generative AI Models "*inherit the toxicity of the Internet*" and therefore require training to reduce their propensity to output toxic content. But who decides what is toxic? Are facts about Tiananmen Square toxic? (DeepSeek thinks so). And with the amount of toxicity in thee training data, and the variety of views of what is offensive, how effective can this really be?

- I hadn't realised the amount of human effort that goes into training the models. Is this approach sustainable as the volumes of potential training data increase? Today, increasing amounts of data are themselves being generated by AI. When this forms the input to the next iteration of training data, what effect will this have on the moderation process and the quality of the data?

- The moderation work is described as "*Precarious Work"* and the authors quote a report that suggests what's required is "*unionisation, transnational organisation, and solidarity between highly paid tech workers and their lower paid counterparts*". Is this likely, and would it work?