Help! The mindless AI zombies are running the establishment!

This paper takes the position that artefactual ethical agents are not, despite widespread claims to the contrary, anywhere on the horizon – never mind, as some would have it, already with us. Neither are existing AI agents conscious nor, in any useful sense of the word, cognising. The massive hype around generative AI – the current AI bubble – obscures the real dangers:

- The way that generative AI systems are designed to come across as smoothly polished and supremely self-confident in a way that mimics the behaviour of aggressively dominant individuals who tend to get their way.
- The way that these systems are known to reflect biases inherent in the corpora on which they are trained and which they essentially summarise over in giving their responses.
- The increasing dependence of people – even highly trained researchers in cognitive science, cognitive neuroscience and AI – on tools that are known to be both biased and otherwise flawed, both factually and structurally, in their output.
- The increasing tendency of such researchers to talk in uncritical terms of these tools as intelligent, autonomous meaning makers – even as potentially conscious agents.
- The increasing dependence of increasingly consolidated media outlets, already pressed to the wall for profits, on these tools to "assist" them in their work.
- The tireless capacity of generative AI systems to produce misinformation and misleading information that floods social media.
- The increasing reliance of many people, partly in response to the growing media deserts and partly due to scepticism toward traditional media, to rely on social media and other non-traditional sources as their primary news sources.
- The way the technology encourages people to stop thinking critically, allowing them to be guided by pseudo-agents that merely seem wise and benevolent.
- The way the technology seduces them into thinking of their conversation partner as a conscious, emotional, human-like agent – even those who, on reflection, would conclude otherwise.
- The lack of adequate guardrails on the technology and the fierce resistance to such guardrails from the industry.
- The way that, in pursuit of profits, human workers are being replaced by AI models that cannot work to the same standard of performance (witness automated filtering of CVs) or being forced to work for lower pay with less certain security.
- The truly mindboggling energy consumption of this technology – most significantly but hardly exclusively for training of the models – in a way that puts increasing pressure on an environment already buckling from human-mediated climate change.

Many have claimed that generative AI is either near to, or already crossing over into, conscious intelligence – capable of reflection, capable of authorship, perhaps capable of feelings. If that claim is, as I believe, a mistake, what gives rise to it?

AI has, since its inception in the 1950s, pursued two distinct paths: one focusing on AI as ***artificial*** intelligence with a focus on reproducing behaviours that previously were thought to require (human) intelligence but turn out not to. This is AI as engineering and automation, as

described by Blay Whitby (2003). The archtype, perhaps, is Joseph Weizenbaum's (1966) Eliza which, to Weizebaum's shock and dismay, convinced many that they were engaging in a meaningful conversation with a thoughtful, intelligent partner – even to point of being upset if the program was suddenly terminated.

The other path has pursued artificial ***intelligence***: which is to say, the (re-)creation of mind ("artificial general intelligence"), often coupled to the idea that mind is computational with rarely, if ever, any attempt to define what "computational" means except to commit to a mechanistic view of mind. Typically, there is a commitment to one of the two following positions:

- Intelligence is all about a ***lot*** of explicit rules, together with symbol manipulation (the originally dominant approach), or…
- Intelligence is all about a ***few*** explicit rules and a great deal of pattern matching based on a training set (the connectionist or machine-learning approach).

Generative AI represents the first kind of AI research masquerading as the second. The reasoning appears to be, make the presentation sufficiently convincing, and for all practical purposes one ***is*** conversing with a thoughtful, intelligent partner. The problem is that, with a bit of knowledge of what to look for, the illusion quickly can be made to break down.

Generative AI models are poor at solving problems that require reasoning about one's own reasoning: so-called meta-reasoning. They are poor at solving novel problems, even if people find them relatively easy to understand. After all, novel problems are reflected nowhere in the training corpora. They are likewise poor at solving novel formulations of familiar problems, if not sufficiently close to the original formulation to be "recognisable" through statistical analysis. They cannot distinguish between real and "hallucinated". Most tellingly, perhaps, they struggle to develop an argument or structure a longer text in a way that makes more than superficial sense. It sounds good, but what (on reflection) are they really saying?

It is telling that, as much as these systems are widely described as intelligent in a human-like way or even conscious, few if anyone is ready to commit to calling them alive. Given that all our experience of human-like (as opposed to artificial) intelligence and all our experience of consciousness is derived from living creatures, it behooves the proponents of these claims to give some appropriate explanation for why intelligence – or consciousness – should, nevertheless, be attributed. In setting out his "semiotic hierarchy",

Jordan Zlatev (2018, 2009) has described true language fluency – as opposed to automated language processing – as dependent on semiosis, itself dependent on consciousness – which is dependent on cognition, which is dependent on life. From language down to life, each level presupposes the next. Unsurprisingly, Zlatev is highly concerned about where generative AI is going.

If the arguments presented here and by Zlatev and others are correct, the ethical implications are huge. Generative AI models are the palest reflection of their creators – imitators *par excellence* that are utterly mindless. Students are using them to write papers without the means to tell whether those papers are meaningful or gibberish. Academics are under intense pressure to use them: to write their lesson plans, grade student submissions, even help them write their own papers. We are dumbing ourselves down, re-creating ourselves in the image of our own creation, even as our species faces an existential threat from climate change.

References

Weizenbaum, J. (1966) Eliza: A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, ***9***(1): 36-45. https://dl.acm.org/doi/pdf/10.1145/365153.365168.

Whitby, B. (2003) *The myth of AI failure*. https://www.sussex.ac.uk/informatics/cogslib/reports/csrp/csrp568.pdf.

Zlatev, J. (2018) Meaning making from life to language: The semiotic hierarchy and phenomenology. *Cognitive Semiotics*, ***11***(1): 20180001. https://lucris.lub.lu.se/ws/portalfiles/portal/76609558/Zlatev2018CogSem.pdf.

Zlatev, J. (2009) The semiotic hierarchy: Life, consciousness, signs and language. *Cognitive Semiotics*, ***4***(supplement): 169-200. https://d1wqtxts1xzle7.cloudfront.net/93186069/download-libre.pdf