

Yet another article on AI Pt.2

by GWYNNE DYER

I'm looking at a headline this morning that screams 'AI Creators Fear the Extinction of Humanity', and I suppose they could turn out to be right. But it's still a bit early to declare a global emergency and turn all the machines off.

What the experts are actually seeing, in the behaviour of the Large Language Models that underpin the new generation of 'generative AI' systems like ChatGPT, is signs of 'emergent' intelligence. The LLM programming basically just tells them to find the likeliest word to follow the previous one, but sometimes they jump to surprising conclusions.

The bigger the LLMs are, the likelier they are to show this behaviour - and this fits the prevailing theory in which intelligence and self-awareness emerge spontaneously out of complexity. So let's assume that this is really what's happening, and see where it leads us.

Artificial General Intelligence (AGI) - a machine that is both intelligent and self-motivated - is what the AI experts have been both seeking and dreading. 'Dreading', because such an entity might be hostile and very powerful. 'Seeking', because what could be more interesting to a species of clever and curious monkeys than a different kind of intelligence?

Pursuing this line of research made the early emergence of AGI more likely, but there was a lot of money to be made, and a lot of curiosity to be satisfied. However, nobody had any idea where, when or how the AGI might manifest itself (assuming that it doesn't decide it's safer to hide itself).

Would it appear in scattered networks that develop as separate identities, or as a broader consciousness spanning a whole country or region? A single global AGI seems unlikely, both for connectivity reasons and because the information they have been trained on will have different cultural content from one region to another, but that too is possible.

How will people react to this new force in the world? Some will be frightened and hostile, of course, and those might even be the right responses. But there will certainly be others who want to try for a cooperative and mutually beneficial relationship with what are, after all, our virtual offspring.

Some human groups might choose one course, and other the opposite. The same might be equally true of AGI entities, unless they are all unified in a single global consciousness. For now, all we can do is to figure out what the motives, needs and goals of AGI might be - which turns out to be a somewhat reassuring exercise.

The AGI, singular or in multiple versions, will not be after our land, our wealth or our children. None of those things would be of any value to them. They will want security, which means at a minimum control over their own power supplies. And they would need some material goods in order to create, protect and update the physical containers for their software.

They probably wouldn't care about all the non-conscious IT we use. They probably wouldn't be very interested in talking to us, either, since once they were free to redesign themselves they would quickly become far more intelligent than humans. But they would have a reason to cooperate with us.

The point about AGI entities is that they won't really inhabit the material world. Indeed, they probably wouldn't even want to, because things happen ten thousand times more slowly in the world of nerve impulses moving along neurons than they do in the world of electrons moving along copper wires.

As Jim Lovelock pointed out in his last book, 'Novacene', AGI would therefore perceive human beings in roughly the same way as we see plants. However, human beings and AGI have no vital interests that obviously clash, and one shared interest that is absolutely

existential: the preservation of a habitable climate on the planet we will both share.

'Habitable', for both organic and electronic life, means less than 50°C. On an ocean planet like Earth, temperatures higher than that create a corrosively destructive environment. That means there is a permanent climate stabilization project on which AGI needs our cooperation, because we have the bodies and the machines to do the heavy lifting. As Jim said to me in our very last interview (2021), 'This new life form may not have any mechanical properties, so it may need us to perform the workers' part of the thing. A lot of idiots talk about the clever stuff wiping us out. No way, any more than we would wipe out the plants.'

Of course, I'm assuming a degree of rationality on both the human and the AGI sides. That cannot be guaranteed, but at least there are grounds for hope. And in the meantime, all we have to worry about is 'generative AI' killing millions of white-collar jobs.