

Yet another article on AI

by GWYNNE DYER

‘Sometimes I think it’s as if aliens have landed and people haven’t realised because they speak very good English,’ said Geoffrey Hinton, the ‘godfather of AI’ (Artificial Intelligence), who resigned from Google and now fears his godchildren will become ‘things more intelligent than us, taking control.’

AI pioneer Eliezer Yudkowsky, co-founder of the Machine Intelligence Research Institute, warns that ‘If somebody builds a too-powerful AI, under present conditions, I expect that every single member of the human species and all biological life on Earth dies shortly thereafter.’

And 1,100 people in the business, including Apple co-founder Steve Wozniak, cognitive scientist Gary Marcus, and engineers at Amazon, DeepMind, Google, Meta and Microsoft, signed an open letter in March calling for a six-month time-out in the development of the most powerful AI systems (anything ‘more powerful than GPT-4?’).

There’s a media feeding frenzy about AI at the moment, and every working journalist is required to have an opinion on it. I turned to the task with some reluctance, as you can tell from the title I put on the piece.

My original article said they really should put the brakes on this experiment for a while, but I didn’t declare an emergency. We’ve been hearing warnings about AI taking over since the first Terminator movie 39 years ago, but I didn’t think it was imminent.

So I did just enough research to convince myself that I knew what was going on, and churned out a complacent 800 words saying that this is not a ‘Skynet’ moment. Mea culpa. It is. Or at least close enough to get the donkey’s attention.

Luckily for me, there are some very clever people on the private distribution list for this column, and one of them instantly replied telling me that I’m wrong. The sky really is about to fall.

He didn’t say that. What he said was that the ChatGPT generation of machines ‘can now ideate using Generative Adversarial Networks (GANs) in a process actually similar to humans.’ That is, they can have original ideas ‘and, being computers, they can generate them orders of magnitude faster, drawing on a far wider knowledge base, than humans.

The key concept here is Artificial General Intelligence (AGI). Ordinary AI is software that follows instructions and performs specific tasks well, but poses no threat to humanity’s dominant position in the scheme of things. Artificial General Intelligence, however, can do intellectual tasks as well as or better than human beings. Generally, better.

If you must talk about the Great Replacement, this is the one to watch. Six months ago, no AGI software existed outside of a few labs. Now, suddenly, something very close to AGI is out on the market ‘and here is what my informant says about it.

‘Humans evolved intelligence by developing ever more complex brains and acquiring knowledge over millions of years. Make something complex enough and it wakes up, becomes self-aware. We woke up. It’s called ‘emergence’.

‘ChatGPT loaded the whole web into its machines ‘far more than any individual human knows. So instead of taking millions of years to wake up, the machines are exhibiting emergent behavior now. No one knows how, but we are far closer to AGI than you state.’

A big challenge that was generally reckoned to be decades away has suddenly arrived on the doorstep, and we have no plan for how to deal with it. It might even be an existential threat, but we still don’t have a plan. That’s why so many people want a six-month time-out, but it would make more sense to demand a year-long pause starting six months ago.

ChatGPT launched only last November, but it already has over 100 million users and the website is generating 1.8 billion visitors per month. Three rival 'generative AI' systems are already on the market, and commercial competition means that the notion of a pause or even a general recall is just a fantasy.

The cat is already out of the bag: anything the web knows, ChatGPT and its rivals know too. That includes every debate that human beings have ever had about the dangers of AGI, and all the proposals that have been made over the years for strangling it in its cradle.

So what we need to figure out urgently is where and how that AGI is emerging, and how to negotiate some form of peaceful coexistence with it. That won't be easy, because we don't even know yet whether it will come in the form of a single global AGI or many different ones. (I suspect the latter.)

And who's 'we' here? There's nobody authorized to speak for the human race either. It could all go very wrong, but there's no way to avoid it.