



Open Research Online

Citation

Pell, David (2024). Commentary by Dr D J Pell on: 'The Age of the Intelligent Machine: Singularity, Efficiency and Existential Peril', Dr A. Amigud. *Philosophy & Technology*, 37, article no. 53.

URL

<https://oro.open.ac.uk/97715/>

License

(CC-BY-NC-ND 4.0) Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Policy

This document has been downloaded from Open Research Online, The Open University's repository of research publications. This version is being made available in accordance with Open Research Online policies available from [Open Research Online \(ORO\) Policies](#)

Versions

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding

Commentary by David J Pell on:

‘The Age of the Intelligent Machine: Singularity, Efficiency and Existential Peril’, Dr A. Amigud

NB. Here, artificial general intelligence (AGI) means: ‘...AI systems that are generally smarter than humans...’ (OpenAI, 2024).

Dr Amigud’s thought-provoking discussion concludes with the big question ‘...whether the path we are taking is going to lead us to eudaimonia or if we better change tack?’

Homo sapiens, evolved around 300,000 years ago at a time of massive climate change and our human ancestors, say two million years before that. Both likely faced many near extinctions (see Doncaster, 2024, Hu et al, 2023 and Prothero, 2018) and sapiens remained at just a few million souls until very recently when we have done dramatically better. It took all human history for us to reach one billion in 1800 but then, powered by our phenomenally rapid technological development, only until about 1928 to reach double that. In less than 100 years since we have more than quadrupled to over 8 billion (Roser & Ritchie, 2023). In this evolutionary blink of an eye, the Anthropocene, we have become a plague on our planet whilst securing the opportunity for improving health, wealth, and happiness for all given the political will. We have, however, come up well short of delivering a state of Aristotelean eudaimonia as the highest human good enabling all to flourish (Duignan, 2024). Incautious use of our industrial technological power over the rest of Nature, and the potential of our burgeoning AGI now places society at what Amigud concludes to be a ‘peculiar junction in history’.

At this junction we face those natural existential perils our species has always faced, for example asteroid impacts and disease, plus those from our application of industrial technology, including climate change, 6th mass biological extinction, pandemics of animal-to-human diseases and mechanised wars including nuclear holocaust and now, also AGI. We teeter on the cusp of an abyss in which these self-created ‘monsters’ seem to compete to threaten our existence. As claimed by Google Deep Mind chief Demis Hassabis: ‘We must take the risks of AI as seriously as other major global challenges, like climate change’ (Milmo, 2023). Other ‘godfathers’ of AGI such as Geoffrey Hinton (Lambert, 2023) and developers such as Elon Musk (Landi, 2023, Narayan et al, 2023) and the ChatGP creators (OpenAI, 2024) have also called for a pause in development, the latter claiming that it ‘...would also come with serious risk of misuse, drastic accidents, and societal disruption.’

In response, calls have been made for rule utilitarian, world regulation as for all the major threats e.g. environment – COPs, pandemics - WHO, nuclear – IAEA, war – UN. So now for AI, The Bletchley Declaration (Gov.uk, 2023), the EU AI Act (European Parliament, 2024) and the USA / UK agreements (Gov.uk, 2024) are examples. Unfortunately, such summits and conventions tend to be cumbersome, slow, and often full of Greta Thunberg’s ‘blah, blah, blah’ (Carrington, 2021). They also tend to be ill-enforced and much contravened because of the natural free-riding, self-maximising behaviour of signatory and non-signatory nation states. The result is the social dilemma of a tragedy of the commons (Hardin, 1968) where here the commons is our shared safety and any available eudaimonia.

Some dispute the scale of the risk. LeClun (in Barr, 2024) for example, another godfather of AI, reassures that: ‘Intelligence has nothing to do with a desire to dominate. It’s not even true for

humans...If it were true that the smartest humans wanted to dominate others, then Albert Einstein and other scientists would have been both rich and powerful, and they were neither.'

We also worry because, unlike the other existential perils, it is more personal. As Harari (2016) told Harris and Raskin (2023), '...the dilemma is that AI is to the virtual and symbolic world what nukes are to the physical world.'. Perhaps this is a clue for control by endowing AGI entities with consciousness, 'minds' of their own and a conscience? Here, Amigud's starting point of the Cartesian 'thinking thing' as the closest approximation to AI, is useful because we can assume that the thinking which Descartes refers to requires consciousness i.e. 'I think therefore I am'. What consciousness is and where it is located in humans and other creatures, is not understood, however. This is the 'hard problem' after Chalmers (1996). We cannot yet create artificial consciousness. Nevertheless, AI/human hybrids already exist as a 'new us' marrying evolved human consciousness with AI e.g. arguably with mobile 'phone dependence. Full scale conscious cyborgs or at least conscious humans with specific 'superpowers' are likely trajectories. Unlike the Zombie like 'unconscious technology' discussed earlier, conscious AGI will get to make its own choices and, as with us, is sometimes likely to run amok through malfunction or psychopathy. If we can also build in a 'conscience', however, then we might secure motivational control (See Meissner, 2019).

As Amigud observes in an extension of the runaway trolley thought experiment (Thompson, 1976), however, we would then face a moral dilemma were a conscious entity to plead for its 'life'. The degree of that dilemma would depend on how conscious we perceive it to be, but we routinely kill conscious animals including some with a sense of self e.g. chimpanzees (Gallup, 1970). Moreover, we justify killing conscious humans with moral justifications, especially self-defence. So why not conscious AGI entities on the same grounds?

The most pessimistic response to Dr Amigud's question is that we will inevitably evolve to a point which causes us to be culled, at best condemned to struggle around in a post- apocalyptic, Hobbesian, state of nature. More optimistic scenarios are available. According to Lovelock (2019), for example, we have moved from 300 years of the Anthropocene into his Novacene, where hyperintelligence dwells and where we might partner with it and achieve at least some measure of Stoic eudaimonia as 'living in agreement with Nature' (Stephens, n.d.) This provides a good case for no major change of tack but an acceptance and response to the risks of all our existential perils including AGI, as we clamber back off the cusp of the abyss.

(998 words)

References

Barr, K. (2023, December 19). *Meta's Chief AI Guy Says Don't Regulate Us, AI Just Isn't That Smart*, Gizmodo, <https://gizmodo.com/meta-s-chief-ai-guy-says-isn-t-that-smart-1850940808> (Accessed 07 April 2024).

Carrington, D. (2021, September 28). *'Blah, blah, blah': Greta Thunberg lambasts leaders over climate crisis*, The Guardian, <https://www.theguardian.com/environment/2021/sep/28/blah-greta-thunberg-leaders-climate-crisis-co2-emissions> (Accessed 07 April 2024).

Doncaster, C. Patrick (2024, April 06). *Timeline of the Human Condition*, University of Southampton, <https://www.southampton.ac.uk/~cpd/history.html> (Accessed 06 April 2024).

Chalmers, D., J. (1996). *The conscious mind in search of a fundamental theory*, Oxford University Press Incorporated, E book ISBN 9780198026532.

Duignan, Brian. (2024). *Eudaimonia*, Britannica, <https://www.britannica.com/topic/eudaimonia> (Accessed 06 April 2024).

European Parliament News (2024, March 13). *Artificial Intelligence Act: MEPs adopt landmark law*, <https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law#:~:text=On%20Wednesday%2C%20Parliament%20approved%20the,fundamental%20rights%2C%20while%20boosting%20innovation> (Accessed 07 April 2024).

Gallup, G. (1970) *Chimpanzees: Self-Recognition*, *Science* 167(3914):86-87, DOI: [10.1126/science.167.3914.86](https://doi.org/10.1126/science.167.3914.86)

Gov.uk (2023). *The Bletchley Declaration by Countries Attending the AI Safety Summit 1-2 November 2023*, UK Government <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023> (Accessed 07 April 2024).

Gov.uk (2024, April 01). *UK & United States announce partnership on science of AI safety*, UK Government, <https://www.gov.uk/government/news/uk-united-states-announce-partnership-on-science-of-ai-safety>

Harari, Y. N. (2016). *Homo Deus: A Brief History of Tomorrow*. McClelland & Stewart.

Hardin, G. (1968). *The Tragedy of the Commons*, *Science*, New Series, Vol, 162, No.3859, American Association for the Advancement of Science, pp. 1243-1248.

Harris, T., & Raskin, A. (January 2021, January 15). *The AI Dilemma*, Center for Humane Technology, <https://www.humanetech.com/podcast/the-ai-dilemma> (Accessed 07 April 2024).

Hu, W., Hao, Z., Du, P., Di Vincenzo, F., Manzi, G., Cui, J., Fu, Y., Pan, Y., Li, H., (2023). *Genomic inference of a severe human bottleneck during the Early to Middle Pleistocene transition*, *Science* (American Association for the Advancement of Science), 381(6661), pp. 979–984. Available at: <https://doi.org/10.1126/science.abq7487>

Lambert, H., (2023, June 21). *Is AI a danger to humanity or our salvation?*, *The New Statesman*, <https://www.newstatesman.com/long-reads/2023/06/men-made-future-godfathers-ai-geoffrey-hinton-yann-lecun-yoshua-bengio-artificial-intelligence>

Landi, M. (2023, November 01). *Elon Musk says AI one of the 'biggest threats' to humanity*, *Independent*, <https://www.independent.co.uk/news/uk/politics/elon-musk-rishi-sunak-government-spacex-tesla-b2439715.html>

Lovelock, J. (2019). *Novacene, The Coming Age of Hyperintelligence*, Penguin Random House UK.

Meissner, G. (2020). *Artificial intelligence: consciousness and conscience*. *AI & Society*, 35(1), 225+, Springer, <http://dx.doi.org.libezproxy.open.ac.uk/10.1007/s00146-019-00880-4>

Milmo, D. (2023, October 24). *AI risk must be treated as seriously as climate crisis, says Google DeepMind chief*. *The Guardian*, <https://www.theguardian.com/technology/2023/oct/24/ai-risk-climate-crisis-google-deepmind-chief-demis-hassabis-regulation> (Accessed 07 April 2024).

Narayan, J., Hu. K., Coulter, C & Mukherjee, S. (2023, April 05). *Elon Musk and others urge AI pause, citing 'risks to society'*, *Reuters*, <https://www.reuters.com/technology/musk-experts-urge-pause-training-ai-systems-that-can-outperform-gpt-4-2023-03-29/>

OpenAI Blog (2024). *Planning for AGI and Beyond*, <https://openai.com/blog/planning-for-agi-and-beyond> (Accessed: 06 April 2024).

Prothero, D. R. (2018) *When Humans Nearly Vanished: The Catastrophic Explosion of the Toba Volcano*, Smithsonian Books, ISBN 13: 1588346358.

Roser, M. & Ritchie. H., (June, 2023). – 'How has world population growth changed over time?' *OurWorldInData.org.*, <https://ourworldindata.org/population-growth-over-time>. (Accessed 16 December 2023).

Stephens, W. O. (n.d.). *Stoic Ethics*, *Internet Encyclopedia of Philosophy*. <https://iep.utm.edu/stoiceth/#:~:text=Stoicism%20is%20known%20as%20a,the%20Stoics%2C%20and%20so%20their> (Accessed 06 April 2024).

Thomson, J. J. (1976). *Killing, letting die, and the trolley problem*. *The Monist*, 59(2), Article 2.