philosophynews.com

WHiP-The Philosophers: The Robots Are Coming

Jeff Hawley, August 22, 2022

Should we fear a future in which the already tricky world of academic publishing is increasingly crowded out by super-intelligent artificial general intelligence (AGI) systems writing papers on phenomenology and ethics? What are the chances that AGI advances to a stage where a human philosophy instructor is similarly removed from the equation? If Jobst Landgrebe and Barry Smith are correct, we have nothing to fear.

One of the *What's Happening in Philosophy (WHiP)-The Philosophers* areas of exploration is current events and their relationship to the role of philosophers in society. This seems like a straightforward topic to cover. But as philosophers often tend to do, we can analyze what we mean by the terms 'role', 'philosophers' and 'society' to form a more rigorous description and perhaps uncover new and interesting correlated questions along the way.

So what exactly is our role in society as philosophers? One way we might refine this function is to follow Shelly Kagan's lead and leverage his term 'P functioning', which he defines as "a body that is functioning in the right way, a body capable of thinking and feeling and communicating, loving and planning, being rational and being self-conscious" (2012, p. 170). For Kagan's purposes, he is describing the physicalist account of what a 'person' is in the broadest terms. This account seems to cover all living persons, so some refinement is necessary to limit our scope to only philosophers within society. So where do we end up? This series explores current events related to bodies that are functioning in the right way, capable of thinking and feeling and communicating, loving and planning, being rational and being self-conscious, thinking philosophically and discussing philosophy with other P functioning bodies.

Last month's WHiP highlighted a recent *Scientific American* article from Almira Osmanovic Thunström, *We Asked GPT-3 to Write an Academic Paper about Itself—Then We Tried to Get It Published* (2022). Is GPT-3 fulfilling the role of a philosopher in society? Does GPT-3 meet the definition of a body which is P functioning in the right way? There are certainly some views within philosophy of mind that would grant that GPT-3 is indeed functioning in the right way and progressing toward AGI (Kurzweil, 2005). There are other views which would not grant the capability for thinking philosophically and discussing philosophy—*intelligence*—to GPT-3 or any similar computer system (Searle, 1992).

Turning to the question of our job security, Jobst Landgrebe and Barry Smith argue that we have little to fear. As you might guess (since the main thesis is right on the tin), their new book, *Why Machines Will Never Rule the World: Artificial Intelligence without Fear* (2022), goes into to painstaking detail as to why AGI is impossible. As a sort of spiritual descendent to Dreyfus' *What Computers Can't Do: A Critique of Artificial Reason* (1972), Landgrebe & Smith also look at the core assumptions about whether artificial intelligence is possible and make the case that it is not. As Landgrebe & Smith note in the foreword to their book, Dreyfus "explains that symbolic (logic-based) AI … was bound to fail, because the mental processes of humans do not follow a logical pattern" (p. x). While they think that Dreyfus ultimately "had been right from the beginning … he did not provide the sort of arguments [they] give in [their] book, which are grounded not on Heideggerian philosophy but on the mathematical implications of the theory of complex systems" (p. x).

In standing up the idea that we needn't live in fear of the "dark scenarios projected by Nick Bostrom* (2003), Elon Musk, and others", they turn to "a non-reductivism commitment to the existence of physical, biological, social, and mental reality, combined with a realist philosophy" that closely resembles the 'primary theory' of Horton (1982) and what analytic philosophers often refer to as 'folk psychology' (p. 3). According to Landgrebe & Smith, their thesis is essentially "about systems, and about how systems can be modeled scientifically" (p. 3). They aren't merely arguing that we should all just relax a bit and take the potential of Bostrom's 'malignant failure modes' in AGI as being an *unlikely* scenario. They instead look out from their philosophical vantage point and claim that AGI is a mathematical *impossibility*. If the problem of designing and implementing true AGI is framed as a problem of mathematics, the solution "cannot be found; not because of any shortcomings in the data or hardware or software or human brains, but rather for *a priori* reasons of mathematics" (Landgrebe & Smith, p. 9).

To present a small taste of these sorts of arguments, we can turn to their analysis of Chalmers' emulation argument in favor of the eventual existence of AGI. As Chalmers (2010) states:

1. The human brain is a machine.
2. We will have the capacity to emulate this machine (before long).
3. If we emulate this machine, there will be [AGI].
4. Absent defeaters [like major catastrophes which would halt further computer hardware evolution, etc.], there will be [AGI] (before long).

(Chalmers, 2010, p. 13 [p. 8 in linked PDF version]; Landgrebe & Smith, 2022, p. 196)

Landgrebe & Smith attack these premises one by one, beginning with the claim that Chalmers has failed to differentiate between inanimate (*physei*) and animate (*techne ontes*) driven systems in P1. According to Landgrebe & Smith, this failure to properly draw out this 'machine' distinction "means that he fails to recognize that the drivenness of animate complex systems prevents the modeling of the laws ultimately governing their behaviour" (p. 196). What sort of laws might there be? Living animate driven systems (from single-celled *archaea* to humans beings) can

- Autonomously produce energy-storing biomolecules from sunlight, inorganic, or organic compounds via oxidation-reduction reactions, which allow them
- to survive, and
- to reproduce, i.e. produce genetic descendants.

(Landgrebe & Smith, p. 196)

Importantly for their claim, "Inanimate drivenness … requires external energy supply and does not induce survival or reproduction" (p. 197). They go on to note that that "we cannot model the way in which the most primitive living organism type—an *archaeum*— functions, because it is a driven complex system in which more than 100,000 biomolecules dynamically interact and … change their molecular properties in order to maintain the life of the organism and enable its reproduction" (p. 198). While we may be able to "predict some of the behaviour," the authors claim that "neither virtually (via explicit or implicit mathematical models) nor physically (by creating synthetic organisms in the lab) can we create models that emulate … a system of this kind in a way that would allow prediction or emulation, and neither can any other sort of mathematics we can currently conceptualize" (p. 198). Put simply, the math just ain't there.

Premise two meets a similar fate as Landgrebe & Smith point out a possible false equivalence between evolution and human activity. This move is followed by an interesting voyage through Kurzweil (2005), Searle (1992), Lucas (1961), Penrose (1994a; 1994b), Bringsjord (2015), Block (1995), and back around to Chalmers (1996). For fans of the vast treasure troves of compelling theories and rebuttals within philosophy of mind, Landgrebe & Smith don't disappoint.

While not necessarily a casual read for those without a sufficient background in the core works within the philosophy of mind canon, *Why Machines Will Never Rule the World* succeeds at delivering a compelling and well presented case for their core thesis. Rather than accept the inevitability of a coming future world of AGI, they knock down the idea before it ever can really get off the ground. As a body which is functioning in the right way, capable of thinking philosophically and discussing philosophy with other P functioning bodies like you, I recommend giving the book a close read.

[*] While Bostrom acknowledges that the idea of existential catastrophe arising from the eventual existence of AGI does contain "degrees of uncertainty," he nonetheless doesn't think that we should be afforded the right to "safely or reasonably ignore the prospect" (2014, p. vii).

References

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18(2), 227–247. https://doi.org/10.1017/s0140525x00038188

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies* (Reprint, 2016 Paperback ed., Vol. 5). Oxford University Press.

Bringsjord, S. (2015). A refutation of Searle on Bostrom (re: malicious machines) and Floridi (re: information). *APA Newsletter on Philosophy and Computation*, 15(1), 7–9.

Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Chalmers, D. (2010). The singularity: a philosophical analysis. *Journal of Consciousness Studies*, 17, 7–65. http://www.consc.net/papers/singularity.pdf

Dreyfus, H. L. (1972). *What Computers Can't Do: A Critique of Artificial Reason*. Harper & Row.

Horton, R. (1982). Tradition and modernity revisited. In M. Hollis & S. Lukes (Eds.), *Rationality and relativism* (pp. 201–260). MIT Press.

Kagan, S. (2012). *Death*. Yale University Press.

Kurzweil, R. (2005). *The Singularity Is Near*. Viking Press.

Landgrebe, J., & Smith, B. (2022). *Why Machines Will Never Rule the World* (1st ed.). Routledge.

Lucas, J. R. (1961). Minds, Machines and Gödel. *Philosophy*, 36(137), 112–127. https://doi.org/10.1017/s0031819100057983

Penrose, R. (1994a). Mathematical intelligence. In J. Khalfa (Ed.), *What is intelligence?* (pp. 107–136). Cambridge University Press.

Penrose, R. (1994b). *Shadows of the Mind*. Oxford University Press.

Searle, J. R. (1992). *The Rediscovery of the Mind* (First Edition). A Bradford Book.

Thunström, A. O. (2022, June 30). We Asked GPT-3 to Write an Academic Paper about Itself—Then We Tried to Get It Published. *Scientific American*. Retrieved July 14, 2022, from https://www.scientificamerican.com/article/we-asked-gpt-3-to-write-an-academic-paper-about-itself-then-we-tried-to-get-it-published/