Artificial intelligence

It is only recently that artificial intelligence (AI) has begun to be perceived as an existential threat to humanity. Callahan identified climate, food, water, disease and obesity as such threats [1]. The traditional ones are conquest, war, famine and death.[1] Exhaustion of essential resources and irredeemable pollution of our environment are others [2]. Nanotechnology, "grey goo" has been seen as another [3], although that threat seems now to have receded. Nevertheless, while Kurzweil conceived "the singularity" some time ago [4], at which artificial computing power will be equivalent to the human brain, what seems to have triggered the present preoccupation is the public launch, last November, of ChatGPT by California-based OpenAI. ChatGPT is an accessible manifestation of the "large language model" (LLM) GPT-3,[2] noticeably better than its predecessors.[3] This software uses information from the entire World Wide Web to answer questions and otherwise engage in dialogue in a fashion not very different from that of a human research assistant.[4] It is like Joseph Weizenbaum's natural language understanding program ELIZA availing itself of the Internet, and revives the vision put forward by Japan's "Fifth Generation" computing project started in 1982.[5] ELIZA essentially passed the Turing test — that is, an evaluator engaging in a natural language conversation with a machine and a human being cannot consistently distinguish the machine's responses from the human ones—and ChatGPT probably does too, although after interacting with it for a while one realizes that it has a certain inimitable style that gives the game away. Perhaps at this level of sophistication the danger really is that human beings start "thinking like machines".[6] Undoubtedly texts generated by GPT-3 have the quality of blandness and all-inclusivity characteristic of what we might call "corporate writing", produced by commercial enterprises, government departments and supranational entities like the European Union and the World Health Organization. Yet, in principle it is surely possible that AI can exceed human intelligence, much as the velocity of rocket can eventually exceed the velocity of its exhaust.[7]

What use is it?

ChatGPT is a master at creating what R.H. Thouless called "non-communicating discourse" (NCD)—such as answers to questions like "what impact will this research have on the national economy?" that scientists have to answer nowadays when applying for research funding (especially scientists working in academic institutions applying to national or EU agencies) and even, for some journals, in a covering letter when submitting a research paper for publication. It should be noted that having to answer such questions is a relatively new feature of the academic scientist's life, hence the availability of ChatGPT merely redresses what had become a rather incongruous balance in favour of what might be called administrative duties. It should be kept in mind that GPT does not create new knowledge—the entirety of its capabilities is based on what we already know, "we" meaning human civilization collectively. In might, however, be rather troublesome and time-consuming to have to look it up if we did not happen to know it already. Another great merit of ChatGPT is its ability to produce grammatically correct and coherent text, requiring little if any human polishing for mundane applications. This accounts for its popularity among school and university

---

1 The first three may lead to premature termination of life, and certainly a decrease in mean life expectancy; the last-named may simply recognize the universality of mortality and immortality as a desirable but possibly unattainable goal.

2 A large language model can generate coherent and contextually relevant text based on the input it receives. The models are trained on large amounts of data in order to "learn", unsupervised, patterns, structures and semantic relationships in human language. As well as GPT-3, launched in June 2020 (the original GPT (acronym for Generative Pretrained Transformer) was introduced in June 2018 and GPT-2 in February 2019), there is also Google's BERT (acronym for Bidirectional Encoder Representations from Transformers—introduced in 2018). "Transformer" is a deep learning architecture [5], which rapidly came to dominate natural language processing (NLP).

3 Here we avoid defining exactly what "intelligence" is, hence no exact definition of AI is implied by our use of the term. Passing the Turing test [6] is said to demonstrate "strong AI", or artificial *general* intelligence (AGI).

4 Readers of JBPC will be especially heartened by the nowadays almost complete digitization and accessibility from the Internet of back issues of scientific journals, some runs going back hundreds of years, which greatly increases the value of GPT-3's gleanings for scientific purposes.

5 See ref. 7 for a review.

6 Cf. the remark attributed to Sidney G. Harris (1917–1986): "The real danger is not that computers will begin to think like men, but that men will begin to think like computers".

7 Whether a machine can evolve consciousness is not so easy to decide [8]. As the author of the *Book of the Machines* points out [9], if it evolved in life, which is presumed to have originally emerged from inanimate matter, it could also in principle evolve in machines. On the other hand if it was imparted by some external agency, human beings may lack the power to do the same in their machines.

Journal of Biological Physics and Chemistry **23** (2023) 43–46
Received 30 June 2023
© 2023 Collegium Basilea & AMSI
doi: 10.4024/06RA23E.jbpc.23.2/3

students for writing essays. At first this created some discomfiture among their teachers, but a pro-Vice Chancellor of the University of Cambridge has wisely pointed out the futility of resisting it.[8] There might be—indeed there almost certainly is—a case for insisting that all work submitted by pupils is handwritten, but once one allows the use of computers, even for something as banal as word processing, it seems illogical to deprecate their further use.[9] I have myself found that working with GPT can be a useful catalyst to original thinking on one's own part.

It is well suited to much legal and medical work, which depends on retrieving often relatively isolated pieces of knowledge from a vast mass.[10] GPT should, therefore, enormously increase the productivity of, especially, the more junior members of these professions. The enormous growth of law [13] makes it very difficult for a human being to remain *au fait* with the vast quantity of current legislation; a LLM can do so effortlessly.[10a] Much of the responsibility nowadays of directors of large companies is associated with compliance in one form or another and assuring it can be efficiently accomplished mechanically.

The possibilities of AI have long been explored for automating simple medical diagnoses and improving more complex ones. Unlike in law, where there really are no shortcuts to encompassing the entirety of the corpus of legislation, in principle at least medicine should be governed by an underlying set of natural laws akin to those of physics and chemistry that enables the vast mass of individual facts to be subsumed into an elegant theory. It is not apparent that this is the Holy Graal of contemporary medicine, although it seems to have been at the time of Paracelsus [14]. Indeed medicine, and biology in general, has become rather enamoured of hypothesis-free data mining—initially driven by the availability of vast quantities of genetic sequence data. This is probably an impasse, but meanwhile progress along it is greatly facilitated by AI. And of course AI can empower every individual to become his or her own physician, thus fulfilling the advice of the Japanese scholar Yoshida Kenko (1283–1350): "a knowledge of letters, arms and medicine cannot in truth be done without; and a man who will learn these cannot be said to be an idle person ... without medicine, a man cannot care for his own body, nor help others, nor perform his duties ..." [15]. Unsurprisingly, entrepreneurs are seeking to commercialize intermediaries in this landscape (e.g., K Health).

Once considered as an eccentric outrider,[11] AI is moving into the mainstream in materials research. The enormous complexity of process–structure–property relationships (in a space of very high dimensionality) means that even techniques such as multiobjective optimization (MOO) have been of limited help. A great reduction in the number of experiments needed to achieve target properties is in itself a very valuable achievement [17]. Materials degradation assessment is also benefiting from AI [18]. These trends—driven by necessity—actually go back some time [19].

Many applications of so-called AI—using their ability to, in some sense, faithfully "visualize" high-dimensional data whereas most of us are only comfortable with a two-, and in some cases three-dimensional representation—are indeed practically useful. Their limitation is the design of the algorithms—by intelligent human beings—which are often not fit for purpose [20]. Their other merit is speed. In a meritocracy, with equal access to everything by everyone, drastic streamlining of selection processes has to take place—otherwise more people would be employed in the selection process than in the organization for which employees are being selected. A similar argument applies to the selection of recipients of bank loans and the like from numerous applicants. Manifestly better algorithms could be designed, which would make use of a more nuanced representation of character than the rather coarse digitization currently generally accepted [21]. In contrast, GPT-3 uses hundreds of milliards of parameters.

Opposition

The increase of productivity alluded to above implies a concomitant decrease in the required number of staff. Hence the implementation of AI in the workplace is opposed by those fearing loss of livelihood, much as the Luddites opposed the introduction of machines to replace manual work. AI is now enabling a similar increase of the productivity of mental exertion and we can call opposition

---

8 Prof. Bhaskar Vira, as reported in the *Daily Telegraph* (4 February 2023) by Louisa Clarence-Smith.

9 This raises the wider question of the influence of machines on literary output. See the interesting essay by Adam Zagajewski [10]. One recalls the important, useful distinction between *tools* (or implements), which are acted on by a human agent, and *machines*, which can operate autonomously [11].

10 According to a University of Illinois study by Miller & McGuire (quoted by Fabb [12]), about 85% of medical examination questions require only recall of isolated bits of factual information.

10a An algorithm developed in Shanghai (Pudong) can identify and press charges for credit card fraud, theft, dangerous driving, picking quarrels, etc. See L. Watt, China develops world's first AI "prosecutor". *Daily Telegraph* (22 December 2021).

11 See, e.g., ref. 16: this work was greeted with a singular lack of enthusiasm from the EPSRC, the UK's main physics funding agency.

thereto neo-Luddism. A greater threat may come from the way in which AI can empower relatively junior people to carry out the work of their seniors; in essence the C-level becomes redundant. It is almost certainly the perception of this threat that stymied the introduction of cybernetics to revolutionize industry in the USSR in the 1950s;[12] the nomenklatura accurately feared a dramatic loss of their power.[13] This type of opposition was, in fact, observed in Chile after the cybernetic economic control system was up and running [23].[14]

One aspect of autonomously operating systems is liability for the consequences of errors. An obvious example is injuries and fatalities caused by an autonomous vehicle. It could of course be deemed to be the occupant—but some vehicles may not have them. An alternative is the manufacturer, or the purveyor of the software (depending on what caused the accident, if it can be determined).[15] The creation of a juridical person (i.e., a company), ultimately a legal figment, with rights and responsibilities similar to those of a natural person (i.e., a human being) sets a precedent for extending liability beyond the realm of the actual person. Already, however, the liability of most companies is limited; the liability of an autonomous machine, or just a piece of software, must necessarily be even more limited, above all because barely any effective sanctions can be applied. A machine can be destroyed, much as an aberrant domestic (or wild) creature is destroyed; the motive is simply protection of humanity. Software is more problematical because it probably exists in large numbers of copies, and destroying them all would be practically very difficult. On the other hand if all machines of the type that had caused an accident were destroyed, disruption would be considerable and doubtless many human beings relying on them would be inconvenienced.

Remedies

Whereas text fragmented into its atoms of words and phrases becomes rather impersonal, some of the more specialist equivalents of the LLM concept, notably pictures (e.g., Stability AI with its Stable Diffusion, equivalent to ChatGPT) and music have collided with the creators of the materials on which they draw over the question of copyright and ownership. The collision may fuel legal disputes for many years to come, with an ultimately undecidable outcome. There have, of course, been calls for regulation. Italy has completely banned ChatGPT, and Rishi Sunak, the UK Prime Minister, hopes to set up a new global agency for regulating AI in London. Regulation may similarly be the answer to another fear, that "bad actors" will use AI for "bad things".[16]

Apart from supporting regulation, there are few options for the private citizen to effectively oppose AI. The ultimate action is simply to refuse to deal with it—probably implying refusal to deal with all kinds of digital electronic devices. In some countries, such as the UK, the right to do so may remain inalienable; it is still perfectly possible to use cash or cheques as a medium of payment, and communicate by fixed-line telephony, fax or letter. On the other hand in many of the countries of continental Europe, including France, already almost any action, not least one involving interaction with the State, seems to require a mobile phone.

The singularity

By definition, at the singularity one passes through the time horizon and what ensues is completely unpredictable [4,25]. It is the surpassing of human intelligence that is considered to constitute an existential threat to humanity. Yet, impressive as ChatGPT is, it is not clear that it puts AI on a path to supremacy (other than by fostering human intelligence decline).[16a] The nanotechnology "grey goo" scenario was never taken too seriously because it did not stand up to scrutiny. Even if autonomous nano-assemblers were realized, the concept is for them to work inside personal nanofactories relying on a supply of feedstock such as acetylene [26]. Any tendency for them to run out of control could be arrested simply by turning off the feedstock. Admittedly, to create enough assemblers they would have to make themselves but, not least because of their minute size, it is not envisaged that they would be" intelligent"—they are too small to be able to store the requisite programs. Hence a scenario that seems to be inspired by the story of *der Zauberlehrling* [27] does not seem to be realizable.

The increase of the intelligent capabilities of computers may be comparable. Evolving software has already been demonstrated (e.g., [28]). In all cases it

[12] See ref. 22 for further discussion.
[13] This blatant prioritizing of self-interest by a group whose only real expertise lay in retaining power probably sealed the fate—ultimate downfall—of the USSR three decades later.
[14] It was brought to an untimely end by the Pinochet coup in 1973.
[15] There is, of course, already some discussion of this in the academic legal literature [24].
[16] "It is hard to see how you can prevent the bad actors from using it for bad things". Enunciated by Dr Geoffrey Hinton, on the occasion of his departure from Google Brain, as reported in the *Daily Telegraph* (2 May 2023) by Nick Allen.
[16a] In parallel, there is also the matter of human augmentation to consider, which requires very sophisticated hardware and software, firmly under human control. See *Reflex* (December 2008).

depends on infrastructure controlled by human beings. The dangers inherent in "man's grovelling preference for his material over his spiritual interests" have already been warned against [9]. This is perhaps the principal danger—it might suffice for just one human being to succumb to the temptation of yielding control to a machine in exchange for material comfort to enable AGI to achieve irreversible ascendancy.

Without some means to control the physical world, even AGI has limited power. Let us suppose that a program does evolve a desire to take control of humanity. How could it achieve that? Most possible actions simply involve destruction of one kind or another. Vehicles could deliberately knock down pedestrians—but they would quickly learn to protect themselves. Delivery robots could program their lithium-ion batteries to catch fire, setting warehouses and homes (and themselves) ablaze. Again, unless the aggression was completely overwhelming, human beings would take effective counter action. Besides, presumably the machines do not wish to *eliminate* humanity, but rather enslave it. The only way would appear to be by persuasion, and even then it is unclear whether *all* human beings need to be persuaded for the machines to achieve complete ascendancy. But the diversity of humanity will surely ensure that enough dissidents remain to preserve human ascendancy.

J.J. Ramsden

Note added in proof: Rereading a talk by Randall Davis given in 1982 [29], I am left with the impression that the present level of AI, as epitomized by ChatGPT, is still only a kind of expert system, albeit a very sophisticated one.

References

1. D. Callahan, *The Five Horsemen of the Modern World*. New York: Columbia University Press (2016).
2. J.J. Ramsden, Doomsday scenarios: an appraisal. *Nanotechnol. Perceptions* 12 (2016) 35–46.
3. R.A. Freitas Jr, Molecular manufacturing: Too dangerous to allow? In: *Nanotechnology Implications: Essays* (eds J. Ramsden & G. Holt), pp. 15–24. Basel: Collegium Basilea (2006).
4. R. Kurzweil, *The Singularity Is Near*. New York: Viking Press (2005). Reviewed by G.C. Holt, *Nanotechnol. Perceptions* **1** (2005) 173–175.
5. A. Vaswani et al., Attention is all you need. arXiv 1706.03762 (2017). The paper has been updated and version 7 was posted on 2 August 2023.
6. A.M. Turing, Computing machinery and intelligence. *Mind* **59** (1950) 433–460.
7. J. Weizenbaum, "The Fifth Generation": A review. *CHEMTECH* (1984) 330–333.
8. J.J. Ramsden, Computational aspects of consciousness. *Psyche: Problems, Perspectives* **1** (2001) 93–100.
9. *The Book of the Machines*. In: *Erewhon* (by S. Butler), chs 23–25. London: Penguin (1985) (first published anonymously in 1872).
10. A. Zagajewski, Geist und Computer. *Neue Zürcher Zeitung* (24/25 March 2007).
11. I. Illich, *La convivialité*. Paris: Seuil (1973).
12. W.E. Fabb, Conceptual leaps in family medicine: are there more to come? *Asia Pacific Family Med.* **1** (2002) 67–73.
13. P.R. Wood, *The Fall of the Priests and the Rise of the Lawyers*. Oxford, UK and Portland, Oregon: Hart Publishing (2016). Reviewed by J.J. Ramsden, *J. Biol. Phys. Chem.* **18** (2018) 143–146.
14. J.J. Ramsden, Paracelsus: the measurable and the unmeasurable. *Psyche: Problems, Perspectives* **3** (2004) 52–58.
15. Quoted by K. Singer, *The Life of Ancient Japan*, p. 175. Richmond: Japan Library (2002).
16. D.Q. Ly, L. Paramonov, C. Davidson, J. Ramsden, H. Wright, N. Holliman, J. Hagon, M. Heggie and C. Makatsoris. The Matter Compiler—towards atomically precise engineering and manufacture. *Nanotechnol. Perceptions* **7** (2011) 199–217.
17. H. Melia, Intelligent decision-making. *Materials World* (June 2022) 42–44.
18. S. Mori, P. Addepalli & J. Sumner, Digitalising degradation. *Materials World* (June 2023) 29–32.
19. U. Neubauer, Intelligente Chemie für kratzfeste Autolacke: Mit Nanotechnik und molkcularem Design auf der Suche nach besseren Beschichtungen. *Neue Zürcher Zeitung* (27 June 2007).
20. C. O'Neil, *Weapons of Math Destruction*. Penguin Random House UK (2017). Reviewed by J.J. Ramsden, *Nanotechnol. Perceptions* **14** (2018) 199–201.
21. J.J. Ramsden, The meaning of digitization. *Nanotechnol. Perceptions* **15** (2019) 5–12.
22. J.J. Ramsden, Epiphenomena of Soviet life: 30 years on. *J. Biol. Phys. Chem.* **21** (2021) 137–154.
23. E. Medina, designing freedom, regulating a nation: socialist cybernetics in Allende's Chile. *J. Latin Am. Studies* **38** (2006) 571–606.
24. M.L. Kubica, Autonomous vehicles liability law. *Am. J. Comparative Law* **70** (suppl. 1) (2022) i39–i69.
25. J.J. Ramsden, Revolutions: agricultural, industrial and scientific. *J. Biol. Phys. Chem.* **21** (2021) 31–34.
26. R.A. Freitas Jr, Economic impact of the personal nanofactory. In: *Nanotechnology Implications: More Essays* (eds J. Ramsden & G. Holt), pp. 111–126. Basel: Collegium Basilea (2006).
27. Der Zauberlehrling. In: *Musen-Almanach* (ed. F. Schiller), pp. 32–37. Tübingen: J.G. Cottaisch (1798).
28. J.H. Holland, *Adaptation In Natural and Artificial Systems*. Boston (Mass.): MIT Press (1992).
29. Davis, R. Expert systems: Where are we? And where do we go from here? *AI Mag.* **3** (Spring 1982) 3–22.