



Editorial

A Braver New World? Of chatbots and other cognoscenti

The 1920s mark the rise of robots in the human lexicon. Karel Čapek, a Czech playwright, wrote *R. U. R.*, which stands for *Rossumovi Univerzální Roboti* or *Rossum's Universal Robots*. The word for worker or labourer in Czech is *robotá*; Karel was given the idea for this word by his artist brother Josef, and the word robot for a human-manufactured humanoid entity was born in 1920. November 30, 2022, just over a century later, was when ChatGPT (Chat Generative Pre-trained Transformer), a chat robot, chatter bot or advanced chatbot, was made available by the company OpenAI for free download.

The angst expressed in *R. U. R.* is prescient:

“Harry Domin [General Manager of R. U. R.]: What the school books say about the *united* efforts of the two great Rossums is all a fairy tale. They used to have dreadful rows. The old *atheist* hadn't the slightest concept of industrial *matters*, and the end of it was that Young Rossum shut him up in some laboratory or the other with his monstrosities while he himself started on the business from an *engineer's* point of view... But in ten years Rossum's Universal Robots will produce so much *corn*, so much *cloth*, so much everything that things will be practically without price. There will be no poverty. All work will be done by living machines. Everybody will be free from worry and liberated from the degradation of labor. Everybody will live only to *perfect* himself... [*Better sense is starting to prevail among some humans.*] All the universities are sending in long petitions to restrict their [robot] production. Otherwise, they say mankind will become extinct through lack of fertility. But the R. U. R. shareholders, of course, won't hear of it. All the governments, on the other hand, are clamoring for an increase in production, to raise the standards of their armies... [*Then the robots go on a rampage.*]

Mr. Alquist [Architect and Head of Works Department of R. U. R.]: For our own selfish ends, for profit, for progress, we have destroyed mankind. Now we'll *burst* with all our greatness. ... [*Then comes an explanation for the robot rebellion.*]

Dr. Gall [Head of Physiological and Experimental Department of R. U. R.]: I changed the way of making them... Chiefly—chiefly, their—their irritability...

Radius [a rampaging robot]: Through me the governments of Robots of the world commands you to deliver Rossum's formula. [*The formula has been destroyed by Miss Helena Glory, a sensitive human being.*]...

Mr. Alquist: I have told you to find human beings.

Radius: There are none left.”

This extensive quotation from *R. U. R.* expresses aspirations and concerns that are still relevant a century later. The recent literature is rife with conversations that various humans have had with ChatGPT and other chatbots such as Microsoft's Bing, some of which have veered towards belligerence and have showed “aberrant” behaviour. A bug in the training software is the proffered explanation.

These latest chatbots based on Large Language Models (LLMs) are said to be poised to free humans from laborious tasks such as finding solutions to problems based on Big Data. However, programs such as ChatGPT could also be used to write essays in exams, scientific papers and reports, and herein lies the ethical dilemma and the need for vigilance and human intervention. Can the products of ChatGPT be distinguished from human endeavours? Can ChatGPT pass the Turing Test? How would a human know it is interacting with another human and not a chatbot?

There is a race on to compare human-derived products with chatbot-produced entities, and the results are mixed, with some saying that chatbots could pass exams and others reporting that chatbot-assisted essays were clearly distinguishable from student essays and received lower grades. But the jury is still out (Stokel-Walker and van Noorden 2023), and chatbots are going to get better and better. Witness the evolution of Deep Blue, the IBM

supercomputer that lost to the Grand Master and World Chess Champion Gary Kasparov in 1996 but came back with renewed training and software tweaking to win a six-set match against Kasparov in 1997.

While there are obvious advantages to these AI developments, there are concerns. Governments are already raising the need for treaties and safeguards in weaponry systems that would be controlled by AI-driven strategic systems. What human safeguards should be built in? What human overrides must exist to prevent a rogue AI program? How would one identify a tendency towards roguishness in an AI program?

There are also important concerns about the energy consumption of these AI systems. These data are not immediately available. One guesstimate puts the daily carbon footprint of ChatGPT at 23.04 kgCO₂e (kilograms of carbon dioxide equivalent) (Ludvigsen 2023). The Green Algorithm calculator has estimated that Microsoft's AI algorithm Meena emitted the equivalent of 164488 kgCO₂e in its training period, which is equivalent to 71 flights between New York and Melbourne (Lannelongue *et al.* 2021). It is also therefore imperative that in the exuberance of the benefits of AI, the carbon footprints of these innovations are also correctly calculated and factored into an energy budget (Budenny *et al.* 2022).

For academicians worried about plagiarism, it is believed that since ChatGPT is trained on a compressed or paraphrased version of sources available on the Internet, its plagiarism score will be low. This means that *à priori*, ChatGPT-written manuscripts will not throw up red flags on plagiarism, but they could rephrase the available literature in novel ways and pass them off as direct human endeavours.

The ramifications of these Generative Pre-trained Transformers will soon be known as they insert themselves into human existence. It is therefore important to research these ramifications so that corrective measures could be taken (van Dis *et al.* 2023).

Yet, the arrow of innovation will ever point upwards. From a Brave New World into one Braver still.

References

- Budenny SA, Lazarev VD, Zakharenko NN, *et al.* 2022 eco2AI: Carbon emissions tracking of machine learning models as the first step towards sustainable AI. *Doklady Rossiiskoi Akad. Nauk Matematika, Informatika* **508** 134–145
- Lannelongue L, Grealey J and Inouye M 2021 Green algorithms: quantifying the carbon footprint of computation. *Adv. Sci.* **8** 2100707
- Ludvigsen KGA 2023 The carbon footprint of ChatGPT. *Towards Data Sci.* (<https://towardsdatascience.com/the-carbon-footprint-of-chatgpt-66932314627d>)
- Stokel-Walker C and Van Noorden R 2023 What ChatGPT and generative AI mean for science. *Nature* **614** 214–216
- van Dis EA, Bollen J, Zuidema W, *et al.* 2023 ChatGPT: five priorities for research. *Nature* **614** 224–226

RENEE M. BORGES,
Editor-in-Chief, *Journal of Biosciences*,
Centre for Ecological Sciences,
Indian Institute of Science,
Bengaluru 560012, India
(Email, renee@iisc.ac.in)