

# ‘A race it might be impossible to stop’: how worried should we be about AI?

Publication Date: 2023-05-07

Author: John Naughton

Section: Technology

Tags: Artificial intelligence (AI), The Observer, Computing, Consciousness, Psychology, Robots, Google, features

Article URL: <https://www.theguardian.com/technology/2023/may/07/a-race-it-might-be-impossible-to-stop-how-worried-should-we-be-about-ai>



Last Monday an eminent, elderly British scientist lobbed a grenade into the febrile anthill of researchers and corporations currently obsessed with artificial intelligence or AI (aka, for the most part, a technology called machine learning). The scientist was Geoffrey Hinton, and the bombshell was the news that he was leaving Google, where he had been doing great work on machine learning for the last 10 years, because he wanted to be free to express his fears about where the technology he had played a seminal role in founding was heading. To say that this was big news would be an epic understatement. The tech industry is a huge, excitable beast that is occasionally prone to outbreaks of “irrational exuberance”, ie madness. One recent bout of it involved cryptocurrencies and a vision of the future of the internet called “Web3”, which an astute young blogger and critic, Molly White, memorably describes as “an enormous grift that’s pouring lighter fluid on our already smoldering planet”. We are currently in the grip of another outbreak of exuberance triggered by “Generative AI” – chatbots, large language models (LLMs) and other exotic artefacts enabled by massive deployment of machine learning – which the industry now regards as the future for which it is busily tooling up. Recently, more than 27,000 people – including many who are knowledgeable about the technology – became so alarmed about the Gadarene rush under way towards a machine-driven dystopia that they issued an open letter calling for a six-month pause in the development of the technology. “Advanced AI could represent a profound change in the history of life on Earth,” it said, “and should be planned for and managed with commensurate care and resources.” It was a sweet letter, reminiscent of my morning sermon to our cats that they should be kind to small mammals and garden birds. The tech giants, which have a long history of being indifferent to the needs of society, have sniffed a new opportunity for world domination and are not going to let a group of nervous intellectuals stand in their way. Which is why Hinton’s intervention was so significant. For he is the guy whose research unlocked the technology that is now loose in the world, for good or ill. And that’s a pretty compelling reason to sit up and pay attention. He is a truly remarkable figure. If there is such a thing as an intellectual pedigree, then Hinton is a thoroughbred. His father, an entomologist, was a fellow of the Royal Society. His great-great-grandfather was George Boole, the 19th-century mathematician who invented the logic that underpins all digital computing. His great-grandfather was Charles Howard Hinton, the mathematician and writer whose idea of a “fourth dimension” became a staple of science fiction and wound up in the Marvel superhero movies of the 2010s. And his cousin, the nuclear physicist Joan Hinton, was one of the few women to work on the wartime Manhattan Project in Los Alamos, which produced the first atomic bomb. Hinton has been obsessed with artificial intelligence for all his adult life, and particularly in the problem of how to build machines that can

learn. An early approach to this was to create a “Perceptron” – a machine that was modelled on the human brain and based on a simplified model of a biological neuron. In 1958 a Cornell professor, Frank Rosenblatt, actually built such a thing, and for a time neural networks were a hot topic in the field. But in 1969 a devastating critique by two MIT scholars, Marvin Minsky and Seymour Papert, was published ... and suddenly neural networks became yesterday’s story. Except that one dogged researcher – Hinton – was convinced that they held the key to machine learning. As New York Times technology reporter Cade Metz puts it, “Hinton remained one of the few who believed it would one day fulfil its promise, delivering machines that could not only recognise objects but identify spoken words, understand natural language, carry on a conversation, and maybe even solve problems humans couldn’t solve on their own”. In 1986, he and two of his colleagues at the University of Toronto published a landmark paper showing that they had cracked the problem of enabling a neural network to become a constantly improving learner using a mathematical technique called “back propagation”. And, in a canny move, Hinton christened this approach “deep learning”, a catchy phrase that journalists could latch on to. (They responded by describing him as “the godfather of AI”, which is crass even by tabloid standards.) In 2012, Google paid \$44m for the fledgling company he had set up with his colleagues, and Hinton went to work for the technology giant, in the process leading and inspiring a group of researchers doing much of the subsequent path-breaking work that the company has done on machine learning in its internal Google Brain group. During his time at Google, Hinton was fairly non-committal (at least in public) about the danger that the technology could lead us into a dystopian future. “Until very recently,” he said, “I thought this existential crisis was a long way off. So, I don’t really have any regrets over what I did.” But now that he has become a free man again, as it were, he’s clearly more worried. In an interview last week, he started to spell out why. At the core of his concern was the fact that the new machines were much better – and faster – learners than humans. “Back propagation may be a much better learning algorithm than what we’ve got. That’s scary ... We have digital computers that can learn more things more quickly and they can instantly teach it to each other. It’s like if people in the room could instantly transfer into my head what they have in theirs.” What’s even more interesting, though, is the hint that what’s really worrying him is the fact that this powerful technology is entirely in the hands of a few huge corporations. Until last year, Hinton told Metz, the Times journalist who has profiled him, “Google acted as a proper steward for the technology, careful not to release something that might cause harm. “But now that Microsoft has augmented its Bing search engine with a chatbot – challenging Google’s core business – Google is racing to deploy the same kind of technology. The tech giants are locked in a competition that might be impossible to stop.” He’s right. We’re moving into uncharted territory. Well, not entirely uncharted. As I read of Hinton’s move on Monday, what came instantly to mind was a story Richard Rhodes tells in his monumental history *The Making of the Atomic Bomb*. On 12 September, 1933, the great Hungarian theoretical physicist Leo Szilard was waiting to cross the road at a junction near the British Museum. He had just been reading a report of a speech given the previous day by Ernest Rutherford, in which the great physicist had said that anyone who “looked for a source of power in the transformation of the atom was talking moonshine”. Szilard suddenly had the idea of a nuclear chain reaction and realised that Rutherford was wrong. “As he crossed the street”, Rhodes writes, “time cracked open before him and he saw a way to the future, death into the world and all our woe, the shape of things to come”. Szilard was the co-author (with Albert Einstein) of the letter to President Roosevelt (about the risk that Hitler might build an atomic bomb) that led to the Manhattan Project, and everything that followed. John Naughton is an Observer columnist and chairs the advisory board of the Mindereroo Centre for Technology and Democracy at Cambridge University.