

Apocalypse not now? AI's benefits may yet outweigh its very real dangers

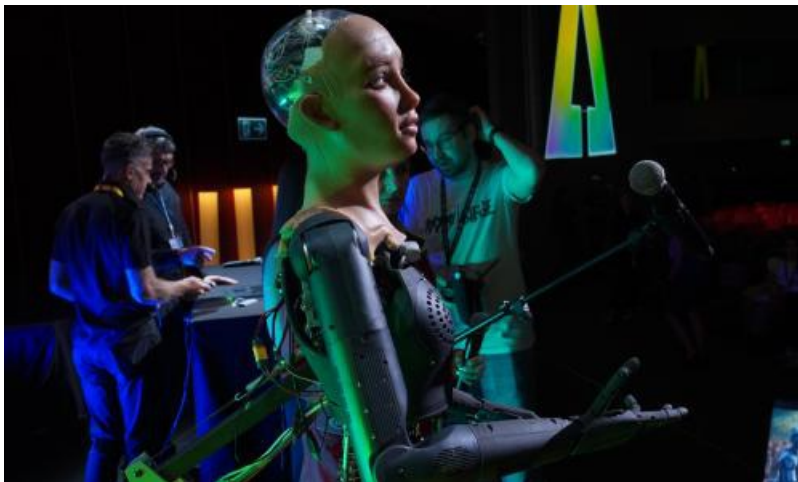
Publication Date: 2023-07-09

Author: Robin McKie

Section: Technology

Tags: Artificial intelligence (AI), The Observer, Science policy, Science (Education), University of Cambridge, Computing, Women, Research, features

Article URL: <https://www.theguardian.com/technology/2023/jul/09/ai-artificial-intelligence-dangers-benefits-cambridge-university>



Stephen Cave has considerable experience of well-intentioned actions that have unhappy consequences. A former senior diplomat in the foreign office during the New Labour era, he was involved in treaty negotiations which later – and unexpectedly – unravelled to trigger several international events that included Brexit. “I know the impact of well-meant global events that have gone wrong,” he admits. His experience could prove valuable, however. The former diplomat, now a senior academic, is about to head a new Cambridge University institute which will investigate all aspects of artificial intelligence in a bid to pinpoint the intellectual perils we face from the growing prowess of computers and to highlight its positive uses. An appreciation of the dangers of unintended consequences should come in handy. “There has been a lot of emphasis in the media on AI leading to human extinction or the collapse of civilisation,” says Cave. “These fears are exaggerated but that does not mean AI will not cause harm to society if we are not careful.” Possible perils include widespread unemployment, as machines take over jobs in education, journalism, law and academia; the spread of disinformation; the illicit hoarding of personal data; the use of facial recognition software to track protesters; and the pernicious influence of AI chatlines. An example of this last danger was illustrated last week when a UK court was told that an AI chatbot was involved in encouraging Jaswant Singh Chail in an attempt to kill the late Queen with a crossbow. AI may not have apocalyptic outcomes but its potential for disruption is clearly considerable. “Power is being concentrated in the hands of a few major corporations who have a monopoly over the way that AI is being built,” says Eleanor Drage, who will be leading a team of researchers within the new institute. “That’s the kind of thing we should be afraid of, because that could result in the misuse of AI.” The Cambridge Institute for Technology and Humanity will amalgamate three university establishments: the Leverhulme Centre for the Future of Intelligence; the Centre for the Study of Existential Risk, which is dedicated to studying all threats that could lead to human extinction or civilisational collapse; and the newly created Centre for Human Inspired Artificial Intelligence, which will focus on finding ways to advance AI for the benefit of humanity. The resulting institute, which will open later this year, will tackle the AI threats and will also focus on its prospects of bringing benefits to the world. This will be done by combining a wide array of talent – from writers to computer scientists and from philosophers to artists, adds Cave. “The institute will have a very interdisciplinary, outward-looking focus,” he insists. A crucial point emphasised by Cave and Drage has been the impact of past technological transformations on societies. “Steam power and the agricultural revolution were incredibly disruptive. Some people did well but many others lost their jobs and homes. “AI has the potential to do that, and we will

have to be very careful to ensure that the latter effects are kept to a minimum. However, the changes it is bringing are arriving at a far faster rate than those of previous technological revolutions." One major problem outlined by Drage is the heavy preponderance of men in the AI industry. "Only 22% of AI professionals are women," she told the Observer. "Nor is there any media encouragement for this to get better. In the media, in films, only 8% of AI scientists are portrayed as women. Women are seen as having no place in the industry." Instead, depictions are dominated by characters such as Tony Stark, the alter ego of Iron Man in the Marvel films. Supposedly a Massachusetts Institute of Technology graduate at the age of 17, Stark entrenches the cultural construction of the AI engineer as a male visionary, says Drage. He, rather than the Arnold Schwarzenegger Terminator image normally used to illustrate AI threats, is the real personification of its dangers. "It's not a trivial point. If women are depicted as having no effective role to play in AI at any level, then the products and services that the industry produces could easily end up actively discriminating against women." Both Cave and Drage stress that the new Cambridge institute will not just issue warnings about AI but will also work to seek out its benefits. "AI allows us to see patterns in data that humans cannot grasp, and that will have benefits for all sorts of fields: from drug discovery to improved energy use, and from personalised medicine to increasing efficiency in water crops," adds Cave. "We have a lot to gain – and a lot to lose unless we are careful."