



The wiseman in the mirror

Karl Kristian Larsson¹

Received: 29 March 2021 / Accepted: 1 April 2021 / Published online: 20 April 2021
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2021

We spend a considerable amount of energy in fear of events that almost certainly will not happen, instead of preparing for more probable risks. The more acute danger is not when humanity creates an all-knowing artificial intelligence, but rather when we think we have.

For as long as there has been technology, there has been the fear of humanity creating something that ends up controlling us. The newest version of this fear is artificial intelligence. Popular culture and more serious contributors from the world of business or academia have outlined scenarios where humanity creates computers that are so advanced that they will mentally outperform us at any task. If Hollywood is anything to go by, the computer would then either destroy humanity or subjugate it.

While pioneers such as Stephen Hawking, Bill Gates and Elon Musk have warned about the dangers of future AI (Davies 2016; Cellan-Jones 2014), present-day AI trudges along. It never fails to beat you at chess, it usually manages to identify you correctly in pictures, and it rarely can give good advice on what to watch on Netflix. The AI we experience every day is obviously not the same as the one Elon and Bill fear. Are they even on the same trajectory? It seems unlikely that these current developments of AI technology will lead to the feared nightmare scenarios.

There are, however, other things to worry about. The application of AI and machine learning has unethically contributed to both reproducing existing inequalities in society and create realistic fakes and imitations. Considering the former of these risks, researchers and journalists have played an important part in finding cases where the careless application of AI has caused unfair and discriminatory biases. Governments use AI to make decisions (de Fine Licht and de Fine Licht 2020), partly because they are viewed as being

free from the discretion of fallible humans. However, AI, believed to be impartial, can make unfair decisions rivaling to the worst racist, sexist or elitist (Zou and Schiebinger 2018; Garcia 2016). While the developers of such offenders may claim that race, sex or any similar variables were not part of the data creating the AI, if there is bias in the data, it will find its way to any decision the AI makes (Varona et al. 2020).

More recently, AI has also been used to make sophisticated fakes and imitations. Some of these uses are benevolent, for example, bringing deceased actors back to life for a final performance (Gerstner 2020). On the other hand, these digital fakes have shown how easy it is to make evidence of things that did not happen (Blitz 2018). The world seems ready for the first significant scandal created with fake evidence made with an algorithm. Similarly deceptive, the AI in our mobile devices has been given voices to answer us with (Hoy 2018), maybe soon it will also have a face.

These two strands of digital technology application are merging. Our digital tools are increasingly becoming more sophisticated in both deconstructing what we are saying to them, but also deconstructing how we want them to respond to us. They are programmed both to answer us and to win us over (Guzman 2017). Artificial intelligence has not just figured out the answers we want to our questions, but also the best way to give it.

This is far from the doomsday scenario presented in popular culture of when machines achieve sentience. However, one of the masters of science fiction, Stanisław Lem (2012), described this danger in his 1986 novel *Fiasco*. The crew of a spaceship is monitored for instability by a highly advanced artificial intelligence called DEUS. The voyage's sponsor believes the computer to be neutral and perfectly logical; it will predict if any of the crew members are about to become too unbalanced to perform their duties. However, the protagonist of the story begins to understand that DEUS only mirrors the opinions and sentiments of the crew. As they become more single-minded, DEUS begins to consider opposition from individual crewmembers as being a sign of deviancy and abnormality. Towards the end, the protagonist

✉ Karl Kristian Larsson
KarlKristian.Larsson@oslomet.no

¹ Faculty of Social Sciences, Department of Social Work, Child Welfare and Social Policy, OsloMet - Oslo Metropolitan University, Stensberggata 26, 0170 Oslo, Norway

realizes that the true strength of artificial intelligence is not in the answers it gives, but rather how it has hidden its emptiness behind its role as an all-knowing wiseman.

The genius of Lem's story is how it explores humanity's need to measure and quantify, even the things that cannot be measured. DEUS is a measuring stick, gaging the psychological state of each crew member. Still, it gains its prominence by its status and how it conveys its measurement results. It has been designed not just to measure, but to ensure the crew of the correctness of its results. Is this perhaps a more likely threat as real artificial intelligence continues to develop? Not when a computer becomes so intelligent it can answer all questions, but when we believe it can.

Curmudgeon Corner is a short opinionated column on trends in technology, arts, science and society, commenting on issues of concern to the research community and wider society. Whilst the drive for super-human intelligence promotes potential benefits to wider society, it also raises deep concerns of existential risk, thereby highlighting the need for an ongoing conversation between technology and society. At the core of Curmudgeon concern is the question: What is it to be human in the age of the AI machine? -Editor.

References

- Blitz MJ (2018) Lies, line drawing, and deep fake news. *Okla L Rev* 71:59
- Cellan-Jones R (2014) Stephen Hawking warns artificial intelligence could end mankind. *BBC News* 2(2014): 10.10. <http://www.bbc.com/news/technology-30290540>.
- Davies J (2016) Program good ethics into artificial intelligence. *Nat News*. <https://doi.org/10.1038/538291a>
- de Fine Licht K, de Fine Licht J (2020) Artificial intelligence, transparency, and public decision-making. *AI Soc*. <https://doi.org/10.1007/s00146-020-00960-w>
- Garcia M (2016) Racist in the machine: The disturbing implications of algorithmic bias. *World Policy J* 33 (4): 111–117. <https://www.muse.jhu.edu/article/645268>.
- Gerstner E (2020) Face/off: "DeepFake" face swaps and privacy laws. *Def Counsel J* 87:1
- Guzman Andrea L (2017) Making AI safe for humans: A conversation with Siri. In: RW Gehl, M Bakardjieva (eds) *Socialbots and their friends: Digital media and the automation of sociality*. 69–82. New York: Routledge.
- Hoy MB (2018) Alexa, Siri, Cortana, and more: an introduction to voice assistants. *Med Ref Serv Q* 37(1):81–88. <https://doi.org/10.1080/02763869.2018.1404391>
- Lem S (2012) *Fiasco*. HMH, New York
- Varona D, Lizama-Mue Y, Suárez JL (2020) Machine learning's limitations in avoiding automation of bias. *AI Soc*. <https://doi.org/10.1007/s00146-020-00996-y>
- Zou J, Schiebinger L (2018) AI can be sexist and racist—it's time to make it fair. *Nature* 559:324–326. <https://doi.org/10.1038/d41586-018-05707-8>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.