

Jewish Lessons from Artificial Intelligence

Just under 20 years ago during my academic studies in medical physics, researchers began asking questions about whether computers could be trained to learn how to analyse medical images such as X-rays, MRI, and PET scans to detect pathological signs. The 21st century has witnessed an explosion in the development of what is called Artificial Intelligence (AI), the ability for computer algorithms to learn how to perform a task for themselves and even learn from one another. Machine learning is no longer the stuff of science fiction, but a reality used in a variety of applications including financial technologies, autonomous cars and deciding whether convicts should be granted parole. The advance of AI has spawned a cornucopia of ethical and philosophical questions.

In 2016 Microsoft's Twitter chatbot called Tay was released. It was designed to learn from human users how to chat like a millennial on social media. It started so well. *"Can I just say that im stoked to meet u? humans are super cool."* Yet within 24 hours of exposure to twitter, Tay had become a fascist, antisemitic, women-hating bigot.

To understand the moral implication of creating AI, it is helpful to consider the moral responsibilities of creating natural intelligence, in the form of children. Every parent hopes to form and mould their kids in their own image. Parents instinctively define the boundaries between proper and improper behaviour, discipline their children to reject immoral actions and reward positive choices. This is why parents become concerned about the effects their children's peers and the media may have on them.

While developers are excited by the prospects of financial and economic advantages of AI, Tay has shown the importance of considering the ethical responsibility of developers to be able to monitor and filter how AI technologies actually learn. And therein lies the rub.

Another AI experiment was Google's AlphaGo which in 2016 challenged Lee Sedol, the world champion of the ancient strategy game Go to a series of matches. AlphaGo beat Sedol in the first game. Yet the most astounding spectacle was yet to come. AlphaGo made a move in the second round known as "Move 37" that deployed a tactic which left experts completely dumbfounded. Many thought it was a mistake, until AlphaGo went on to win that match and complete the five-match series 4-1.

AlphaGo was not only better than humans at playing strategy games, but the decision also it had made in Move 37 could not be understood by the most expert human players on earth. While humans can do a variety of different intelligent tasks such as hold conversations, make cups of tea, and contemplate the meaning of life, it is not beyond the range of possibility that a technology company could be able to artificially simulate general human intelligence. If the world expert Go players couldn't understand why AlphaGo made Move 37, what chance do we have in determining the decisions made by an Artificial General Intelligence? How could we devolve the responsibility of making moral choices to AI, faced with the possibility of not understanding their decisions?

מה שתשיב

www.mahshetashiv.com

Take a modern reworking of the famous trolley problem. An autonomous car is driving a child to school when a group of five other children walk out in front of the car. The AI algorithm driving the car can either decide to continue and kill five children, or swerve to avoid them, hit a tree, and kill the passenger. Children are precious and there is no other option, so presumably killing one child is better than killing five?

The *talmudic* sages discussed similar lose-lose scenarios involving conflicting moral principles. On the one hand, we acknowledge the sanctity of life, yet we cannot value one life over another – or even a group of others. The key principle often employed in such situations is known as *shev v'al ta'aseh*. When confronted with a lose-lose scenario, while there are many complicating factors, it is usually better to be passive than to act. Perhaps counter-intuitively even when one could save more lives by intervening, Jewish law would favour non-intervention over actively causing someone's death to save others.

Given these moral complexities, the parenting role of a human creator becomes even more critical. Yet there is an even greater complication. So much of our moral instruction when parenting centres on the capacity for children to be empathetic – for example, when we say “How would you like it if someone called you names?” Can we honestly hope to parent an artificial mind that we create, given the prospect that it could outsmart us and make unpredictable moral choices just like Move 37?

At the heart of this conundrum is what it fundamentally means to be human. Mankind has not faced such existential questions since Charles Darwin published his theory of evolution by natural selection in 1859. Neuroscience indicates that we are merely wet robots without free will. Theoretically, an artificial human mind should therefore be a possibility.

But when God created man in His image, it was not God's infinite intelligence He wanted us to emulate, but His capacity to love. Moral choices require the ability to understand, empathise and connect emotionally with other humans. This can only be accomplished by being human, precisely because each one of us is unique, irreplaceable, and vulnerable to loss. An artificial mind that could be copied and mass-produced is none of those things and could never experience them.

While God created man in His image, man can only recreate himself through love and love can never be artificial.

A version of this article first appeared in the Jewish Chronicle on the 8th of June 2018 under the title “*If machines can already beat us in games, do we need to worry?*”

If you have enjoyed this article and would like to donate to help the Mah Shetashiv team, please use your smartphone to scan the QR code, or visit our website at:

www.mahshetashiv.com/donate-to-mah-shetashiv

© Mah Shetashiv 2023

