

Jewish Contemporary Ethics Part 46: Artificial Intelligence 3 – The Ethics of Autonomous AI

by Rabbi Dr Moshe Freedman, New West End Synagogue



Artificial Intelligence (AI) is a broad term which describes computer programmes which can learn from the information they receive, respond appropriately and even make decisions without using explicit, pre-

programmed instructions. The previous article explored the challenge of understanding the decisions made by such AI algorithms. If AI can outsmart and exceed human ingenuity and initiative, it will inevitably mean that the reasoning behind its decisions will become indecipherable and inscrutable to its human masters.

This is of great concern. The applications of AI can be divided into two broad categories: the processing of vast amounts of data to detect trends, elicit useful information and advise future actions, versus autonomous machines which act without human supervision or intervention. If AI is restricted to the former advisory role, then its efficacy will be greatly diminished. However, while the prospect of developing autonomous machines is seductive, it opens a Pandora's box of ethical questions.

Take for example the Pentagon's Defense Advanced Research Projects Agency (DARPA) which is responsible for developing military technology. According to an article in the Economist (September 7th, 2019) DARPA has built software called Real-time Adversarial Intelligence and Decision-making (RAID) which "aims to predict the goals, movements and even the possible emotions of enemy forces five hours into the future." There is an obvious ethical problem in deploying autonomous AI decision-making technology in warfare. The rules of engagement in military scenarios revolve around "proportionality (between civilian harm and military advantage) and necessity. Software that cannot explain why a target was chosen probably cannot abide by those laws."

Additionally, for both scientific and legal reasons it is necessary to be able to audit decisions that are made, even when AI operates in a merely advisory role. Otherwise when something goes wrong it will be impossible to correct mistakes and ascribe responsibility. In a 2009 report entitled 'Autonomous Systems: Social, Legal and Ethical Issues', the Royal Academy of Engineering addressed the question by asking "are autonomous systems different from other complex controlled systems? Should they be regarded either as 'robotic people' – in which case they might be blamed for faults; or machines – in which case accidents would be just like accidents due to other kinds of mechanical failure."

From a Jewish legal perspective, autonomous AI could be equated to the ownership of an animal or the parental responsibility over a child who does not independently have legal competence. While *adam hamazik* – direct damage caused by a legally competent agent – is liable, *mamon hamazik* – damage caused by one's possessions such as an animal they own, their child or an AI machine – is far more complex.

Based on the Gemara in Bava Kama, Rabbi Moshe Sternbuch of the Edah HaChareidit in Jerusalem rules that parents may be liable for the damage their child does only if they were negligent in watching their child (Teshuvot VeHanhagot 3:477). Yet if a claim was brought against the actions of an autonomous AI machine, without understanding the motive for an action, it might be impossible to ever ascertain grounds for negligence in any scenario.

The next two articles will examine the possibility and implications of creating human-like Artificial General Intelligence.