



BAHA TEACHINGS.ORGTM
...Baha'i Inspired Perspectives

Artificial Intelligence: Can Robots Be Moral?

Zarrín Caldwell

Science fiction writers have asked this question for decades: can robots be moral? Now, though, we've entered an era where that question has become all too real.

My husband shared a story with me the other day about a psychopathic algorithm (named Norman Bates), who was created by a team at the Massachusetts Institute of Technology (MIT), to process data from the dark corners of the web.

Needless to say, Norman developed a very negative view of the world—seeing dead bodies and destruction everywhere. In Norman's defense, he was running on biased data. But, it still made me think about where we are going with artificial intelligence (AI) and what that means vis-à-vis our critical, but largely ignored role as spiritual beings.

In fairness, and as the article noted, AI can play an important role in a variety of industries—like personal digital assistants, fraud prevention, and voice recognition. They can be instrumental too in advancing education or identifying disease. We don't really know where these technologies might take us. I'm reminded of the 2012 movie *Robot and Frank*, in which a robot butler is programmed to look after his elderly (and forgetful) client. In the end, though, the robot helped Frank with a multimillion dollar robbery.

Which brings me to the more fundamental question of whether machines are really capable of making moral decisions.

MIT also has a project called Moral Machine, which is gathering human perspectives on moral decisions made by machine intelligence—such as self-driving cars. Is it more acceptable or ethical, for example, to kill two passengers or five pedestrians? Nobody wants to make that choice. Still, ongoing research asks, how do we turn ethical values into quantifiable parameters that machines/robots can understand?

I must admit that I remain skeptical—in part because I adhere to the “garbage in, garbage out” philosophy. In other words, if you have greedy or corrupt people without moral values or ethical frameworks who are programming machines, the outcome may be quite dystopian. The Baha'i writings don't address modern autonomous robots, per se, but they often address how scientific and material advances can be used for building up or tearing down:

Consequently, when thou lookest at the orderly pattern of kingdoms, cities and villages, with the attractiveness of their adornments, the freshness of their natural resources, the refinement of their appliances, the ease of their means of travel, the extent of knowledge available about the world of

nature, the great inventions, the colossal enterprises, the noble discoveries and scientific researches, thou wouldst conclude that civilization conduceth to the happiness and the progress of the human world. Yet shouldst thou turn thine eye to the discovery of destructive and infernal machines, to the development of forces of demolition and the invention of fiery implements, which uproot the tree of life, it would become evident and manifest unto thee that civilization is conjoined with barbarism. Progress and barbarism go hand in hand, unless material civilization be confirmed by Divine Guidance, by the revelations of the All-Merciful and by godly virtues, and be reinforced by spiritual conduct, by the ideals of the Kingdom and by the outpourings of the Realm of Might. – Abdu'l-Baha, *Selections from the Writings of Abdul-Baha*, pp. 284-285.

In the same passage, Abdul-Baha adds:

Until material achievements, physical accomplishments and human virtues are reinforced by spiritual perfections, luminous qualities and characteristics of mercy, no fruit or result shall issue therefrom, nor will the happiness of the world of humanity, which is the ultimate aim, be attained.

The Baha'i teachings claim that human beings (unlike robots) are endowed with souls. All contemporary discrediting of religion aside, each of the world's great religious teachers have advocated for high standards of ethics and morality—to help us, individually, reach our spiritual potential and, collectively, to live peaceably with each other.

But, what happens when we divorce material pursuits from these spiritual guidelines? I believe that we are seeing the consequences of those choices in an increasingly violent society. Is that also happening with our seemingly uncritical pursuit of AI? Instead of rushing headlong into promoting the latest technological innovations, maybe we should be asking tougher questions about how these tools are being used, and by whom?

A statement titled *One Common Faith* released by the Baha'i International Community in 2005 provides additional reflections on both the advantages and limits of these technological advances:

Clearly, materialism's error has lain not in the laudable effort to improve the conditions of life, but in the narrowness of mind and unjustified self-confidence that have defined its mission. The importance both of material prosperity and of the scientific and technological advances necessary to its achievement is a theme that runs through the writings of the Baha'i Faith. As was inevitable from the outset, however, arbitrary efforts to disengage such physical and material well-being from humanity's spiritual and moral development have ended by forfeiting the allegiance of the very populations whose interests a materialistic culture purports to serve.

Many are realizing, in other words, that adhering to a purely materialistic philosophy has led to a profound emptiness of soul. Perhaps we see that in sobering World Health Organization statistics that note that suicide is the second leading cause of death among 15-29 year olds globally. For all the practical support that AI could potentially provide, we may need to look elsewhere to discover what it really means to be human.