Human after all: How cinema uses AI to extend moral and ethical dilemmas

The cinematic staple of giving machines human characteristics reveals a lot about our complex relationship with AI and morality. By Chris Speed

It's not hard to see why AI is an interesting starting point for a movie. Beyond the obvious storylines that explore the threat to our perception of being the dominant intelligent species, AI has become a lens through which to consider more existential questions – a way to interrogate the very condition of 'being human'.

In order to do this, a persistent habit in cinema has been to cast AI in the form of a human body. Whether it's from as far back as Fritz Lang's Metropolis (1927) with Maria's robot double, or more recent examples such as the childlike android David in Steven Spielberg's A.I. Artificial Intelligence (2001), the question of what it is to be human is explored through the decision making of a more-than-human. But what do these embodiments of artificial intelligence tell audiences about our own moral and ethical condition?

Before we dive into cinema's role in presenting these issues, it is worth noting that cinema is still struggling to overcome significant challenges in casting AI into gendered forms. In most cases, manifestations of AI in

a male form demonstrate a desire to exert power and seek intellectual superiority. Female embodiments may seek to explore the same issues but come with an added dimension of sexualisation, a trait which exemplifies the biases that lie behind some large-scale datasets.

'Good morning, Dave'

While cinema audiences of the 1960s were contemplating the power of Alpha 60, a sentient computer system that has complete control of the city of Alphaville in the Jean-Luc Godard film of the same name, or the onboard computer HAL 9000 in Stanley Kubrick's 2001: A Space Odyssey, that prioritises its own 'life' and the spacecraft's mission over the lives of the crew, academics were developing thought experiments to explore moral and ethical dilemmas.

Of the numerous experiments that emerged, the 'trolley problem' resonates with many of the cinematic plots through which audiences explore human deliberation and the logic of machines.
The trolley problem is relatively simple. There is a

runaway trolley (or train), ahead of which there are five people tied to the tracks. On a sidetrack is one person who is also tied down. You stand at a lever on the train and are faced with two options: do nothing and allow the train to continue on its path and kill five people, or pull the lever, divert the train toward the sidetrack and kill only one person.

As AI has crept into our lives this thought experiment has become less abstract. In the hands of scientists, it has been aligned with the grand challenge to "help [the scientists] learn how to make machines moral".

Studies such as Moral Machine, developed by the Scalable Cooperation group at the MIT MediaLab, place viewers in a series of scenarios in which the trolley is swapped for an autonomous vehicle. The moral dilemma is complicated through the introduction of more information about the consequences of a decision: that you might kill subjects of different ages, genders, physical health and species (human or cat).

Computer says 'no'

Of course, these dilemmas make for good plots in movies involving AI, immersing the viewer in a moral quandary where the decision-making of an AI in human form is in conflict with a human protagonist or a community that they represent.

Most recently we see it used in the Netflix film Outside the Wire which places a human alongside an AI, in what appears initially to be collaborative circumstances. As the story unfolds, the scriptwriters put the duo in increasingly contradictory moral dilemmas where the AI and human have differing views.

The opening scenes see our human hero Harp, a drone pilot based in a ground control station in the US, in the first of a series of these dilemmas. He is monitoring an incident involving peacekeeping American troops stationed in Eastern Europe, fighting pro-Russian insurgents. Harp decides to disobey his commanders and deploys a Hellfire missile killing Americans and Russian ground troops but ending the incident. During the subsequent military trial, Harp justifies his actions by stating, "There were 40 men on the ground, and I saved 38."

Harp is punished for ignoring a direct action to hold fire, and is sent into action where he is assigned to Captain Leo, an advanced AI masquerading as a human officer. The scriptwriters construct a moral bond between the pair as Captain Leo asserts that Harp had made the right decision at the time, revealing that he had more data about the circumstances of the incident than both the troops on the ground and the senior officers in command. Tension is built throughout the story, as the audience is put in situations that place stress on the relationship between the human and the AI, as moral decisions change according to the politics of each scene.

However, as the story moves towards its conclusion, the intentions that inform Captain Leo's decisions become more clouded and Harp struggles to follow the logic. As we approach the final dilemma, the audience and Harp are led to understand Leo's reasoning behind his decision-making process - that he sees his kind (autonomous robots) as an inevitable cause of future conflict and that the correct moral action is to launch a nuclear warhead at the USA to prevent them from using AIs in the future.

Moral machines

Literally targeting American audiences with a moral dilemma that places them on the railway tracks of the 'trolley problem', Harp pleads with Leo, arguing that humanity must learn to design better AI to avoid the unnecessary deaths of millions of innocent people. I'll let you watch the movie to find out what our all-American hero does next.

Outside the Wire may not be a great movie. But what is particularly interesting is the decision of the scriptwriters to place the responsible development of AI in the hands of the viewer. It suggests that AI won't be going away anytime soon, but it's likely we will have to play a part in an increasing amount of moral and ethical decisions to manage its outcomes.

———