Article type: Essay
Reviewer: Stephen Cave and Kanta Dihal
Word count: 1278

TECHNOLOGY

Ancient dreams of intelligent machines

Stephen Cave and Kanta Dihal look back at three millennia of cultural responses to automata and AI.

The French philosopher René Descartes was reputedly fond of automata: they inspired his view that living things were biological machines that function like clockwork. Less known is the fact that he had a daughter, Francine, who died of scarlet fever at the age of five. After the philosopher's own death in 1650, a strange story began to circulate — that a distraught Descartes had had a clockwork Francine made, a walking, talking simulacrum. When Queen Christina invited the philosopher to Sweden in 1649, he sailed with the automaton concealed in a casket. Suspicious sailors forced the trunk open; when the mechanical child sat up to greet them, the horrified crew threw it overboard.

The story is probably apocryphal. But it sums up the hopes and fears humanity has associated with human-like machines for nearly three millennia. Those who build them do so in the hope that they will overcome natural limits — in Descartes' case, death itself. But this very unnaturalness terrifies and repulses others. In our era of advanced robotics and artificial intelligence, those polarised responses persist, with pundits and the public applauding or warning against each advance. Digging into the deep history of intelligent machines, both real and imagined, we see how these attitudes evolved: from fantasies of trusty mechanical helpers to fears that runaway advances in technology may lead to creatures that supersede humanity itself.

Arguably the oldest known story of something approximating AI can be found in Homer's eighth-century BC *Iliad*, his epic poem of the Trojan War. In it Hephaestus, disabled god of metalworking, creates golden handmaidens to help him in his forge: "In them is understanding in their hearts, and in them speech and strength, and they know cunning handiwork" (book XVIII). Hephaestus was also supposedly responsible for the first 'killer robot', Talos. A mechanical bronze colossus featuring in the third-century BC epic *Argonautica*, it patrolled the shores of Crete lobbing boulders at invaders.

These fictions were grounded in reality: ancient Greek technologists were astonishingly skilled in mechanics and metalwork. In her forthcoming book *Gods and Robots*, classicist Adrienne Mayor describes bronze automata that featured at the Olympic Games two centuries before the *Argonautica* — a leaping dolphin and eagle in apparent flight. In his first century AD treatise *On Automaton-Making*, the mathematician-engineer Hero of

Alexandria describes a fully automated puppet theatre that, through a combination of displaced grain, axles, levers, pulleys and wheels, could enact an entire tragedy.

These classical-era stories reveal how humanoid machines were mostly conceived as representing straightforward hopes — the ideal servant who always obeys, the perfect soldier who never tires. But as the influence of Greece declined over the first centuries AD, the Latin West entered a millennium in which the skills of automaton-making were lost, along with the aspirations associated with them. As historian E.R. Truitt describes in *Medieval Robots* (2015), it was the Byzantine Empire and Arab world that preserved the mechanical arts over those centuries. Around 850, for instance, the Banu Musa brothers in what is now Iraq published the *Book of Ingenious Devices*, which featured automata such as a hydro-powered organ. Thus the 'otherness' of mechanical simulacra was compounded in the West. Associated there with the exotic and the idea of an 'infidel' East, they were treated with awe and suspicion.

The thirteenth century saw a resurgence of Western interest in automata, which began appearing at courts as showpieces designed to wow visitors. The Count of Artois at the château of Hesdin in what is now northern France commissioned a clutter of mechanised beasts and androids that spoke to guests, or even soaked them with water. At the same time, hints of darker themes began to emerge. A number of great medieval scholars, such as Roger Bacon, Albertus Magnus and Gerbert of Aurillac (Pope Sylvester II), were rumoured to have created bronze heads that could answer any question — a proto-Siri. These stories end badly with the oracle's destruction, often by people who mistrust it. In such cautionary tales, the creation of AI is an act of Promethean hubris, semi-divine power no mortal should possess — presaging the story of Descartes' daughter and Mary Shelley's *Frankenstein* (see https://www.nature.com/articles/535490a).

The march of mechanical humanoids nevertheless persisted into the great cultural and technological flowering of the Renaissance. Evoking widespread wonder, hydraulic, spring-powered and clockwork automata proliferated in Europe, from mechanized angels in churches to Neptune automatons in grottos. Their mystical connotations began to give way to reflections of cutting-edge scientific developments: Leonardo da Vinci, for instance, drew up plans for a robotic knight operated internally by weights and pulleys.

Over subsequent centuries, Descartes' view of living things as complex machines inspired new heights in manufacture, and Europe reached peak automata in the seventeenth to early nineteenth centuries. Master craftsmen built marvels of art-imitating-life, such as Jacques de Vaucanson's 1739 Digesting Duck. Sporting more than 400 moving parts in each wing and internal rubber tubing, it appeared to eat, drink and defecate.

But the duck's digestive process was a trick: its 'droppings' were actually pre-fabricated pellets ejected from a hidden compartment. A few decades later in 1770, inventor Wolfgang von Kempelen's infamous chess-playing Mechanical Turk debuted. Though many suspected it was a hoax, nobody could quite figure out how they were fooled until a revelatory article in 1857. Awe over the transgressive nature of such mechanisms mixed with fears of deception, as reflected in the fiction of the time. In Prussian Romantic author E.T.A. Hoffmann's 1816 short story 'The Sandman', for example, the protagonist Nathanael is bewitched by the beauty of a woman called Olimpia. The discovery that she is an automaton drives him to suicide.

The imaginative resonance of intelligent machines began to reach its fullest in the twentieth century, when industrialisation replaced the rhythms of nature with those of the production line. It was also a time of revolution and mechanised warfare. Against that backdrop, the term 'robot' was born in Karel Čapek's 1920 play *R.U.R. (Rossum's Universal Robots)*. Famously, in the very work coining the term, the robots rebel against and destroy their creators. And that narrative of rebellion has proven to be the most potent of all our AI fears, retold repeatedly as technology evolves.

During the cold war space race, *2001: A Space Odyssey* (1968) gave us HAL 9000, the murderous spaceship computer. With the rise of the Internet, we got the Terminator films' Skynet, a defence network that becomes self-aware, and *The Matrix*, featuring intelligent machines that farm humans whose minds unknowingly inhabit a virtual reality. Now, with AI dominating headlines, we have sophisticated robots again overthrowing wetware masters, from Ava in the 2015 film *Ex Machina* to the latest instalment of the *Westworld* franchise.

The persistent trope of robot revolts reveals the paradox at the heart of our relationship with intelligent machines. We want to create clever tools that can do everything we can do and more. They will be the perfect oracles, servants, soldiers, even lovers. To fulfil our hopes, they must have attributes such as intellect and agency — minds of their own, superior to ours. But paradoxically, that is also why we fear HAL and Skynet. The tension lies in our conflicted desire to create beings superhuman in capacity, but subhuman in status.

But though our hopes continually threaten to collapse into such fears, we hope nonetheless. For every robot rebel, there is a benevolent counterpart such as Star War's C3PO or the android child David in Steven Spielberg's 2001 *AI: Artificial Intelligence.* Both kinds of stories, the hopeful and the fearful, reveal to us our complex emotional responses to AI. Understanding these and their deep history is crucial to making the most of life with intelligent machines.

Stephen Cave is Executive Director of the Leverhulme Centre for the Future of Intelligence (CFI) at the University of Cambridge.

Kanta Dihal is the postdoctoral researcher on the AI Narratives project at CFI.