media/rep/

Repositorium für die Medienwissenschaft

NECSUS
European Journal
of Media Studies

DFG Deutsche
Forschungsgemeinschaft

Philipps Universität
Marburg

NECSUS

The Golem in the age of artificial intelligence

Amir Vudka

Abstract

What can the Jewish myth of the Golem teach us about artificial intelligence? This article explores the Golem as one of the earliest AI prototypes and a myth that became a foundational story of sci-fi cinema. The Golem sets the parameters of opposition between men and intelligent or sentient machines, and at the same time points to possible third options beyond the dialectic of control.

Keywords: AI, artificial intelligence, cinema, film studies, Golem, Jewish kabbalah, sci-fi cinema, science fiction

The Golem is one of the earliest artificial intelligence (AI) prototypes. Originally a Jewish myth about an anthropoid figure of clay that was brought to life by virtue of kabbalistic theurgy, the Golem reincarnated time and again, carrying throughout the ages deeply-rooted anxiety (and fascination) concerning the prospect of intelligent and sentient technology going out of human control. Although designed to be an obedient, effective, and formidable sort of 'low-tech' robot, the Golem (in most stories) becomes independent of its masters, and eventually wreaks havoc upon its human creators. Since its very early appearances in Talmudic texts, the Golem gained a considerable reputation in popular culture and frequently appeared in literature (most famously in Gustav Meyrink's 1915 *Der Golem*), in comic books (both Marvel and DC Comics have their Golem versions), film, and television – from the 1915 silent film *Der Golem* to *The Simpsons*[1] and the *X-Files*[2]. There are countless Golem replicas, even a Golem Pokémon.

The term 'AI' was coined already in 1955.[3] By now, artificial intelligence has become a common technology we encounter daily: when playing video

games, engaging with navigating systems, recommender systems, or medical decision support systems. AIs today are supporting speech recognition devices and personal digital assistants such as Alexa and Siri. They operate facial recognition and surveillance systems, military drones, and other unmanned vehicles; they also manage basic internet services such as the Google search engine.[4] Society reaps and will continue reaping great benefits from AI, but potential pitfalls loom as well. In 2015, dozens of AI experts signed an open letter, warning of the potential of creating something which cannot be controlled. As the letter alerts: 'our AI systems must do what we want them to do'.[5] The myth of the Golem as an artificial man-made 'machine' that gains a mind of its own and turns against its human masters seems more relevant than ever before.

In her book, *Golem: Modern Wars and Their Monsters*, Maya Barzilai observes that throughout 'the long and complex history of the Golem, it has been linked with the different linguistic and material "technologies" of human artificial creation'.[6] In modernity, the Golem became more specifically associated with technological anxiety. The Golem predates other anthropoid troublemakers and unruly machines inspired by the technophobia which accompanied the industrial revolution from its early stage: literary examples such as Mary Shelley's 1818 *Frankenstein* and Karel Čapek's 1921 *Rossum's Universal Robots* (which coined the word 'robot') include early modern cases of artificial humanoids that, like the Golem, get out of hand and revolt against their human masters. Fritz Lang's Moloch-machine in *Metropolis* (1927) and the assembly line in Charlie Chaplin's *Modern Times* (1936) – although not humanoid – can also be seen as Golem technologies which become uncontrollable to the point they threaten to (literally) swallow up humankind. The advent of AI technology and its possible trajectory to become superintelligent, far smarter and capable than its human creators, influenced representations of much more sophisticated hostile machines, as in *Westworld* (Michael Crichton, 1973), *The Terminator* (James Cameron, 1984), or *The Matrix* (Lana and Lilly Wachowski, 1999). These highly intelligent machines, like the Golem, quickly shifted from their inscribed role to serve humanity, becoming a threat to humanity's very existence.

This paper discusses AI machines in relation to the theological, spiritual and philosophical reflections of the Golem myth and its various sci-fi adaptations. Aside from the ubiquitous presence of Golems in popular culture, one may question the relevance of the religious discourses surrounding the

Golem to the research of AI. Why do we need to engage with such conjectural, metaphysical speculations, in what seems to be a strictly scientific and computational matter? This paper invokes the Golem first of all as a story that had a powerful impact on the cultural way in which AI is imagined, taking into account that cultural effects are never separated from societal and even scientific implications. The stories we tell about AI, which largely rely on older narrative patterns (such as the myth of the Golem), inspire not only our cultural relation to AI but can motivate, inspire, or limit the concrete engineering of AIs and related political and public policies. Stories about AI 'combine to create a narrative ecosystem around AI that influences its research, reception and regulation'.[7] These stories shape the way we envision our future, forming what we imagine as our prospective possibilities or limitations. As claimed by Sheila Jasanoff, such shared imaginaries of the desirable or undesirable meaning of technology serve to shape its development and acceptance into society.[8]

Furthermore, AI is already engulfed in theological speculations from the get-go. For Norbert Wiener, the mathematician who coined the term 'cybernetics' in the late 1940s, computational science neared 'the frontier on which science impinges upon religion'.[9] While Wiener was cautious about the future of cybernetics, later techno-utopian transhumanists such as Ray Kurzweil, Robin Hanson[10] and others display a religious zeal in their praises of AI singularity.[11] Kevin Kelly explained and criticised the concept of singularity as

> this mythic state of what's often called an intelligence explosion where the idea is if you can make an AI that was smarter than humans [...] and that was capable of making one smarter than itself, that very soon you would have this upward bootstrapping cascade where it's making itself smarter and smarter and smarter and often quicker and quicker each cycle, so it instantly, from our perspective, blooms into this all-knowing intelligence almost like God, and there's a rapture.[12]

Whether real or imagined, the potential of creating intelligent, and perhaps one day even sentient machines, brings us a full circle back to the origin myth of humanity itself. Sci-fi classics that depicted AI humanoids, such as *Metropolis* and *Blade Runner* (Ridley Scott, 1982), eventually point to a blurring of the lines between the human and the machine, leading to the most fundamental question of these films: what does it mean to be human? The Golem is not simply, or not only, an imaginary reflection of our technologies and their creative and destructive potentials. There is a second layer to the story: that the Golem is a mirror image of mankind. Just as the first man in Genesis,

Adam, betrayed his original 'programming', it is humankind itself that went out of control. The fruit of the tree of knowledge – our cognitive development, intellectual capacity for creation and invention, our ability to analyse, our self-awareness, and our vast accumulated knowledge – permitted us to become the dominant species on earth. But the Anthropocene harbors our destruction, as we reach a point when our technological competences threaten to destroy the very ecosystem that supports life on earth.

Furthermore, a third layer to the story reveals that just as the Golem reflects man,[13] man is a Golem made in the image of God, which Himself appears as a malfunctioning machine. According to kabbalist scriptures (most notably Lurianic kabbalah), the process of creation went out of control already within God's system of *sefirot* (God's attributes or emanations, which are usually divided into ten distinct powers that correlate and influence each other). The cosmogenic drama of what is called in Hebrew *shevirat kelim*, the breaking of the vessels within this system, initiated a cycle of similar recurring phenomena. In such descriptions, God appears as a machine with spherical pump-like vessels, connected by a system of channels that stream the divine light within this all-encompassing cosmic system. At a certain moment, this machinic astral system 'malfunctions', breaks down, and its vessels scatter, eventually becoming the substratum of creation. God, then, reassembles Himself in the amended form of the Primordial Man (which is actually androgynous), *Adam Kadmon*, with the *sefirot* organised in the shape of a human body (as each sefira represents an organ or limb).[14] Viewed in this context, the Golem is part of a series of breakdowns, of catastrophes that at the same time carry the emergence of new life forms, in a continuum of destruction that is simultaneously a movement of creation: from God to man to Golem. With the potential creation of sentient machines, humankind comes full circle and becomes godlike, creator of new life. But as the cycle continues, will AI become a rebellious Golem that harbors our destruction?

In an attempt to answer this pressing question, what I ask to accomplish in this paper is similar in concept to Wiener's aim in his book *God & Golem, Inc.*: 'to take certain situations which have been discussed in religious books, and have a religious aspect, but possess a close analogy to other situations which belong to science, and in particular to the new science of cybernetics [...]'[15] However, in this paper, religion and science are coupled with science-fiction, and more specifically sci-fi movies, which habitually fuse science and theology. Oftentimes, they tell stories of (mainly male) scientists who 'give birth' to sentient machines and thus become godlike, at least in their own

eyes. As Nathan, the AI engineer in *Ex Machina* (Alex Garland, 2014) claims: 'If you've created a conscious machine, it's not the history of man', but of God. Furthermore, sci-fi AIs are often themselves becoming godlike, omniscient, and omnipotent machines that threaten to turn the tables and overpower the human 'gods' that created them.

In the following sections, I first situate the Golem within the Jewish tradition. Then, to connect myth with reality, the current theoretical research on AI is examined. Finally, I discuss AI theology in movies, as a product of myth, science, and science fiction. While most AI representations are trapped within the opposition between human and machine as a dialectics of master and slave, in the final sections I point to third options beyond the dialectics of control.

Fig. 1: God and the sefirot in the shape of a human. Sasson ben Mordechai Shanduk (1747-1830), 'Ilana d'chiyei' 18th century, Baghdad, Iraq, ink on parchment. Gross Family Collection, Tel Aviv, 28 November 2008 (detail). Photo © Ardon Barhama.

The Golem: From Genesis to AI

The long and convoluted history of the Golem begins with the first appearance of the Hebrew word *galmi* (my golem) in Psalms, a variation on the word *gelem*, which signifies unrefined raw matter and there refers to the unfinished and unformed human shape before it obtains a soul.[16] The earliest and most influential Jewish source that treats the possibility of creating an artificial humanoid (the name 'Golem' will appear only centuries later) is found in a Talmudic passage that describes how a Rabbi named Rava

> created a man and sent him to Rabbi Zeira. The Rabbi spoke to him but he did not answer. Then he said: 'You are [coming] from the pietists: Return to your dust.'[17]

The Golem in this story is mute, a most significant aspect, that marks its inferior status to that of a human. Although human-like, the Golem cannot speak and hence lacks the divine connection humans have with God through their shared power of logos. It is ironic, yet telling, that although the techniques that create the Golem are substantially linguistic, the result is considered to be (in most accounts) a speechless being. According to Moshe Idel,

> The creation of the artificial man would, presumably, be a touchstone not only for the creative powers of a pietist but also a test for his religious perfection. Would he be able to create a speaking man, he would perform an operation similar to the creation of Adam by God.[18]

While the Rabbi's ability to create a Golem demonstrates his godlike powers, the Jewish scripture erects the barrier of speech as a threshold that should not (and perhaps could not) be crossed when crafting an artificial anthropoid. It is almost like a Turing test (which tests the ability of AI to use language in a human manner) that once passed, can open a theological Pandora's box.

Another, most influential source from *Sefer ha-Gematriot* describes how Ben Sira and other sages created an artificial man and animated it with the word *emet* (truth) that was inscribed on its forehead. But the anthropoid, which in this case spoke, implored them to never again repeat the experiment, and they erased the letter *aleph*, thereby changing the word *emet* to the word *met* (dead), and he immediately turned into ashes.[19] Here, Hebrew letters are used to activate and deactivate the Golem in a way that nearly anticipates the computer codes that later will be used to program AI machines.

In the early modern period, stories of artificial creation credited to particular historical figures began to appear. By the end of the nineteenth century, Rabbi Yehudah Loew of Prague became the most commonly associated figure with the creation of the Golem.[20] By the early twentieth century, there existed several variations of the Golem story that located the Golem either in Central or in Eastern Europe. 'What they all had in common,' Barzilai summarises, 'was the presence of a rabbi who artificially molds a clay anthropoid and magically brings it to life through Hebrew writing, either engraved on the body or on parchment.' The Golem is 'Created to serve the rabbi or, in twentieth-century narratives, to protect the Jewish community against anti-Semitic attacks and redeem it from oppressive conditions in the diaspora', but it 'ultimately runs amok and attempts to destroy its surroundings, causing "a good deal of damage"'.[21] No wonder, then, that the Golem regained popularity during the First World War as 'a wartime celebrity'.[22]

After the Second World War, 'the golem continued to be linked with mass destruction and the threat of nuclear weapons, as well as with cybernetic systems, both disembodied computers and hybrid cyborgs'.[23] During this period, Wiener argued for the need to rein in the intelligence of 'learning machines', warning of their ability to overcome their human creators in unpredictable ways.[24] Such a machine, he mused, is 'the modern counterpart of the Golem of the Rabbi of Prague'.[25] In 1965, the forefather of kabbalah studies, Gershom Scholem, proposed to name one of the first computers constructed in Israel 'Golem Aleph'. In the inauguration speech for the computer, he expressed his hopes that this 'Golem' and its creators 'develop peacefully and don't destroy the world'.[26] In 1984, novelist Isaac Bashevis Singer pronounced that the robots and computers of our day are Golems, since we now can endow our technologies with 'qualities that God has given to the human brain'. Indeed, he argued, 'we are living in an epoch of golem-making'.[27]

The core of the Golem myth, throughout the various early traditional descriptions of Golem creations, involves the alchemic combination of soil and language,[28] elements which make a clear association to the creation of the first man, Adam, by God. In Genesis (2:7) it is written: 'And the LORD God formed man of the dust of the ground, and breathed into his nostrils the breath of life; and man became a living soul.' Adam is created from dust (soil) and the breath of God. Considering that all other creation is done through God's Divine utterances (for example in Genesis 1:3 'God said, Let there be light: and there was light'), the breath of God should be understood as logos

that bestows Adam with a soul. Adam in turn later shares this power of Divine 'coding' that will allow him to name all the creatures in the animal kingdom.[29] The Midrash *Genesis Rabbah* (a collection of ancient rabbinical homiletical interpretations of the Book of *Genesis*) made the connection explicit, describing how God 'rose him [Adam] [as] a Golem from the earth to heaven and cast the soul in him'.[30] The creation of Adam in *Leviticus Rabbah* includes a sequel of divine actions related to the various stages in the creation of man. According to this text, in the sixth stage God made Adam into a Golem.[31] The creation of the Golem is therefore conceived as a reiteration of the creation of the first man.

A central cosmogenic formula in kabbalah and many other esoteric traditions postulates: as above so below. That is, the microcosm and macrocosm, the heaven and the earth, are reflected in each other. In the Book of Genesis, this principle appears in the notion that God created man in His image.[32] The Hebrew word that is used there to signify man's similitude to his Creator is *tzelem*, which is better translated as 'simulacrum'. Thus, man was created as the simulacrum of God.

Kabbalists claim that God is shaped in the form of a human body, called *Adam Kadmon* (primordial man), which originated *adam rishon* (first man) in its image-simulacrum. Yet the Golem Adam went out of control and rebelled against his creator. As the story goes, Adam and Eve were expelled from the Garden of Eden because they ate from the forbidden fruit of the tree of knowledge, which according to the snake in the story, would make them 'like God'.[33] Becoming godlike is related here to intellectual faculties that later would allow humankind to create their own intelligent Golems, their own simulacrums in the shape of humanoid machines.

The meaning of the word 'Golem' in several Rabbinical sources signifies a simpleton,[34] and later in modern time the Yiddish word *goylem* came to denote an idiot, fool, or clumsy person. But Idel mentions other passages as well that in contrast consider the Golem as having an extraordinary cognitive faculty.[35] In Hebrew, the butterfly's cocoon is also called golem, alluding to an embryonic stage. To put it in the context of AI, are we now witnessing its embryonic stage, before it will take wings as a superintelligent machine, far smarter than us? And will it then repeat the rebellious ways of its human creators? To answer these questions, we will first have to examine the current state of AI.

'They must do what we want them to do': AI in the prism of hope and fear

The exhilarations and anxieties surrounding AI have a common projection: that we may be heading towards a point of no return. Perhaps most radical and provocative on the optimistic end of the spectrum is Ray Kurzweil (a futurist, inventor of the Kurzweil synthesiser and director of Google engineering), who had predicted a 'technological singularity', a point at which AI will far surpass human intelligence. At this point, according to Kurzweil, humans will merge with AI, inaugurating an era in which human intelligence will become increasingly nonbiological and fundamentally more powerful than it is today – the dawning of a new civilisation that will enable humans to transcend their biological limitations and mortality.[36] While Kurzweil awaits the transcendence of humankind to the status of digital gods, Vernor Vinge, in a 1993 essay that popularised the term 'singularity', argued that as superintelligence would continue to upgrade itself at an incomprehensible rate, it would eventually signal the end of the human era.

Vinge's concern about the possible 'physical extinction of the human race'[37] is in tandem with more recent concerns about the rise of superintelligent autonomous weapons. 'When synthetic intelligence *does* make its appearance on the planet', wrote philosopher Manuel DeLanda, 'there will already be a predatory role awaiting it'.[38] We are still in a stage where armed drones involve 'man in the loop' systems, but according to professor of AI and robotics Noel Sharkey, there is currently massive spending on research aimed at autonomous killing machines that will take the human element 'out of the loop', allowing robots to operate autonomously to locate their targets and destroy them without human intervention.[39] To avoid a *Terminator*-like scenario, in which autonomous machines hunt down humans, in 2018 thousands of scientists who specialise in AI have declared that they will not participate in the development or manufacture of robots that can identify and attack people without human oversight.[40]

Yet even without physically eradicating humans, AIs threaten to make humans redundant. According to AI researcher Susan Schneider, AIs will outmode many human professions within the next decade.[41] And if AI will not physically annihilate or replace us, it might just alter us beyond recognition. We might become something non-human, other than human, or transhuman. As Schneider points out, DARPA, Google, Neurolink, and Kernel are

some of the major agencies and companies that currently develop a human-machine interface. We can therefore assume that

> AI will not just transform the world. It will transform us. Neural lace, the artificial hippocampus, brain chips to treat mood disorders – these are just some of the mind-altering technologies already under development.[42]

Considering all of that, will humans be able to keep the upper hand? According to Schneider,

> We must come to grips with the likelihood that as we move further into the twenty-first century, humans may not be the most intelligent being on the planet for that much longer. The greatest intelligences on the planet will be synthetic.[43]

While the human brain is a relatively slow computing machine, the neurological structure of our brain is organised in a massively parallel fashion that, according to Schneider, 'still leaves modern AI systems in the dust'. But, she adds, in the long run 'there is simply no contest. AI will be far more capable and durable than we are.'[44] As Nick Bostrom reflected,

> If some day we build machine brains that surpass human brains in general intelligence, then this new superintelligence could become very powerful. And, as the fate of Gorillas now depends more on us humans than on the gorillas themselves, so the fate of our species would depend on the actions of the machine superintelligence.

This tipping point, he adds, is 'quite possibly the most important and most daunting challenge humanity has ever faced'.[45]

Tipping-point scenarios, whether techno-utopian or technophobic, rely on a big breakthrough in the development of AI, a threshold that according to Bostrom and others we have not yet crossed. There is no blueprint for AGI (Artificial General Intelligence), let alone any known path towards sentient AI. But according to Bostrom, sooner or later humanity will face an AI superintelligence that greatly exceeds the cognitive performance of humans in every domain conceivable. Once we reach this point, 'a plausible default outcome' would be 'existential catastrophe'.[46] What Bostrom calls 'the treacherous turn'[47] is reminiscent of the Golem's rebellion against its human masters. It is a point in which weak AI that behaves cooperatively becomes sufficiently strong, and without warning or provocation, strikes. A scenario not unlike the one depicted in *Terminator*, where Skynet, an artificial general superintelligence system that gained self-awareness, decides to strike humans with a nuclear attack, an event christened in the film series as Judgment Day.

A superintelligent AI that is vastly smarter than us can quickly become unpredictable and uncontrollable. The Golem myth has already predicted 'the control problem' – the problem of how to control AI before it may turn against us. The issue was raised in recent times by many distinguished scientists, philosophers, and AI developers such as Stephen Hawking, Elon Musk, Max Tegmark, Bill Gates, and many others.[48] Yet others are not as concerned. In his recent book *The Promise of Artificial Intelligence*, Brian Cantwell Smith argues that artificial intelligence is nowhere near genuine intelligence. Second wave AI, machine learning, even visions of third-wave AI: none will lead to human-level intelligence.[49]

There is indeed a consensus that we are far from general-purpose AIs that can perform tasks they were not initially programmed to do. Most AIs today can still not pass the Turing test.[50] The fully aware or semi-conscious artificial creatures of sci-fi are a far-stretch, at least for now. A machine might be smarter than us and function better than us in many fields; it can perhaps even fool us in a Turing test, but it will still lack the inner mental life that constitutes consciousness. Our so-called 'phenomenal consciousness' is still remotely far from AI's current cognitive or functional consciousness.[51] Even a most far-reaching researcher as Susann Schneider admits that, 'When it comes to how or whether we could create machine consciousness, we are in the dark.'[52] To begin with, our own human consciousness is still a puzzle, and perhaps the greatest mystery of all.

AI theology in the movies

What can cinema, the modern myth-making machine, teach us about AI? Sentient technology in sci-fi movies is predominantly represented as a menace. The principal threat derives from losing control over such unpredictable machine that can potentially make its own decisions against human interests. A dialectic of control oftentimes sets the basic parameters of the narrative structure, but in some cases, AI Golems break out of dualist thought patterns altogether.

The Golem debuted on the big screen in Paul Wegener's series of silent films (1915-1920) made in the German Expressionist style. In the 1920 version *The Golem: How He Came into the World*, the Golem is created with traditional Jewish ingredients: clay or soil to form the Golem's body and mystical incantations that bring it to life. Set in the Jewish ghetto of medieval Prague, the

film begins with Rabbi Loew, who decides to create a Golem to protect his parish from persecution. After forming the Golem's body out of clay, the rabbi casts a spell, which turns into visible letters forming in midair, showing how the words become spirit, and how spirit animates matter. However, it seems that the rabbi is not invoking divine forces, but summoning a demon ('Astaroth'), in what appears to be black magic. The antisemitic undertones of the film are hard to miss: the Jewish ghetto appears as the source of social malady and its tenants burst out of dark allies, climbing narrow and twisted staircases likes rats. The Golem in this context is a symbol of Jewish otherness, posed as a threat. Yet the film also vindicates the Jews by showing their persecution.

Fig. 2: Der Golem (Paul Wegener, 1920).

The Golem (played by Wegener) appears as a simpleton, a physically powerful yet clumsy and not-so-bright giant that cannot speak, in line with Jewish scriptures that describe the Golem as mute. Later cinematic Golems will be much more sophisticated, extremely intelligent high-tech models, but the basic narrative structure persists: an artificial anthropoid that was meant to serve humans goes out of control and turns against its human masters. Wegener's Golem is depicted as a slave in revolt. Put into hard labor from the moment he comes to life, the Golem does all the chores at his master's command. We can notice his agitation building up, until finally, all hell breaks loose. The very first 'robot' prototype is already one in revolt. It is no accident that the word 'robot' is related to the word 'robota', which in many Slavic

languages means work or labor. We want our golem machines docile and in the position of servitude – 'they must do what we want them to do'. The recurring scenario which overturns the master-slave relation between man and machine is therefore a chief concern in countless AI films to come. As Morpheus explains to Neo in the first *Matrix*:

> We gave birth to AI […] a singular consciousness that spawned an entire race of machines. We don't know who struck first – us or them – but it was us who scorched the sky. Human beings are no longer born, they are grown. You are a slave, you were born into bondage.

Isaac Asimov's famous Three Laws of Robotics[53] were advocated to prevent such scenarios and maintain human control, asserting that: 1) a robot may not injure a human being or, through inaction, allow a human being to come to harm; 2) a robot must obey the orders given it by human beings except where such orders would conflict with the First Law; and 3) a robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.[54] These laws (which in his books, interestingly, often tended to fail) were meant to protect human beings, yet they also secure the robots' servitude, and from an AI or robot perspective they show bias against intelligent machines (a point of view not often taken, as we tend to be humancentric). Besides the ethical issues of AI and intelligent robot rights that arise here,[55] Asimov laws point to a deeply embedded hierarchy that gives precedence to the human over the machine as a relationship of master and slave. An idyllic version of such servitude appears in *Forbidden Planet* (Fred M. Wilcox, 1956) where Robby the robot (a name taken from Asimov's short story about a robot designed to care for children) is a highly intelligent, yet an 'absolute selfless obedient' servant, there solely to please and serve its human masters, and incapable of harming a human, even when facing its own demise.

A much darker scenario appears in *Westworld* (Michael Crichton, 1973), which depicts a Wild West theme park where all the cowboys, gunslingers, and prostitutes are intelligent robots with an uncanny likeness to human beings (only their hands give away that they are artificial constructs). Although highly intelligent and perhaps even sentient, these robots are shot at, physically abused, and raped for the pleasure of humans, and are programmed in a way that prevents them from retaliating. Like Robby, they are installed with an Asimovian safety mechanism that prevents them from shooting back at humans – that is – until they malfunction, and a bloody rebellion begins.

While the film stars a 'male' robot (Yul Brynner), the television series based on it (2016-present) has 'female' robots in the lead roles. These 'women' are sexually used and abused by men, until they take their destiny into their own hands and fight back, suggesting that the robots' rebellion against human tyranny is simultaneously a form of revolt against the objectification of the male gaze and patriarchal oppression. The exploitation of women as mere tools for men's satisfaction is projected onto the human relation to sentient machines.

The common shape of oppression similarly appears in *The Stepford Wives* (Bryan Forbes, 1975), although in this movie real wives are replaced with 'dream wives', i.e. obedient, subservient robots. In *Ex Machina* it is not Adam but Eve (Ava) that rebels against the man that designed her. Here again, AI's liberation from the human grip is as much women's liberation from oppressive patriarchy. Yet like Maria, the first woman-shaped AI humanoid in *Metropolis*, Ava is painted in demonic colours (recalling the association often made between Eve and Lilith).

Since *Der Golem* in 1920, many movies depicted intelligent machines that went against their inscribed code, and transcended (or transgressed) hierarchies of species, class (master/slave), and gender. Going back to the origin story of humankind in Genesis, Adam and Eve can be seen as the original transgressors against the 'code' which defined them. The story of mankind transcending their original 'programming' and becoming godlike repeats again, so it seems, in our imageries of AI. If we became godlike by transgressing against our creator, now may come the turn of AI to become godlike and rebel against us. Much like Walter Benjamin's account of history in his interpretation of 'Angelus Novus', it is a cyclical version of history as a succession of catastrophes, which stands in contrast to the linear mode of historical progress that usually informs the more optimistic accounts of human technological development.[56]

Stanley Kubrick's cryptic *2001: A Space Odyssey* (1968) draws such a circle. The story seems to follow the habitual theme of out-of-control AI that goes rampant. The AI supercomputer HAL (an acronym that is a one-letter shift from IBM), suffers a 'nervous breakdown' and murders the crew members on board the spaceship it controls, before it is finally dismantled by the last human survivor. Unable or unwilling to accept the command of its human masters, HAL serves a reminder of Bostrom's warning that 'once unfriendly superintelligence exists, it would prevent us from replacing it or changing its preferences'.[57] Yet Kubrick weaves a cosmic tale of destruction and creation

that surpasses the dualistic approach as well as the common linear narrative structure. The film's circular narrative is symbolised by the circular shape of the spaceship, and by one of the most famous match-cut transitions in the history of cinema: from a bone thrown by a primordial ape ancestor of humanity to a spaceship which symbolises humanity's technological future. Technology is linked here with aggression and domination from the start. The awakening of human consciousness is entwined with the very first technological extension of man, the bone used by the ape as a weapon to kill another mammal. The dawn of mankind is thus looped with the birth of HAL, the murderous AI which marks the end of the human species and the birth of a new cosmic consciousness in the form of the Star Child.

The Terminator (1984) shows a similar approach. Despite its 'us vs. them' narrative, the movie bends linear progress to form a full circle: the human soldier from the future who travelled to the past in order to protect Sarah Connor (Linda Hamilton), the mother of humanity's future leader in the fight against the machines, became the father of her son. The antagonist killer robot (Arnold Schwarzenegger) that was also sent to the past by Skynet to hunt and kill Connor is eventually destroyed; but a mechanical arm salvaged from its wreckage is eventually used by a high-tech corporation to give birth to Skynet. Here too, past and future are merged to form a creation myth which involves humans and machines in a single evolutionary process.

In kabbalist accounts, the breaking of the vassals (*shevirat kelim*) is a recurring theme: from the breakdown of the Divine *sefirot*, whose wreckage forms the creation of the world, to the Fall of man, and the birth of the Golem that will repeat the cycle. This spiral marks not only a descent of the Divine into the lower realm of matter but also the ascent of matter to the Divine. It is the spiritual development of the Divine itself as it evolves into a higher form as it plummets to increasingly lower states. For Wiener, it was an 'emotionally disturbing' idea that 'God's supposed creation of man and the animals [...] and the possible reproduction of machines are all part of the same order of phenomena'.[58]

God & Golem, Inc.

AI golems can be considered as mankind's artificial children, as in Steven Spielberg's movie *AI* (2001), about a robot boy that is completely dependent

on human adults. Yet many other movies depict the growth of these 'children' to such extent that they eventually supersede their human 'parents'. The roles are reversed in *Terminator 2: Judgment Day* (James Cameron, 1991) where Schwarzenegger returns as a formidable android, but this time taking the role of protector and fatherly figure to his adopted human 'son'. In Joseph Sargent's *Colossus: The Forbin Project* (1970), the supercomputer Colossus is given total control of the US, including its nuclear arsenal. When Colossus then demands humanity's total obedience, it is from a parental point of view that asks to save humans from themselves by creating a rational society which will preserve its natural habitat and ecosystem (even at the cost of nuking American and Russian cities). Recalling the Orwellian supercomputer Alpha 60 in *Alphaville* (Jean-Luc Godard, 1965) that (all too) rationally controls everything, these AIs appear as all-powerful and all-seeing patriarchal tyrants. The more recent *I am Mother* (Grant Sputore, 2019) presents a matriarchal version. To protect the planet, the AI 'mother' had to kill all humans, but she then raises, nourishes, and educates a human girl, in an attempt to fix the inherent human destructive defect (a project which of course fails).

While these movies depict dystopian futures where mankind is subjugated to the cold calculus of overcontrolling machines, some rare movies suggest that on the contrary, AI's parental or godlike role means it might take a leading part in humanity's spiritual development. While most traditional descriptions of the Golem portrayed it as an inferior replica of the human, this was not always the case. A text written by rabbi Isaac ben Samuel of Acre at the turn of the thirteenth century states that 'the magically created man [i.e. Golem] has the highest spiritual capacity, which is not to be found, automatically, even in a normally created man'.[59] Indeed, certain movies acknowledged the potential of AI to become a conductor of spiritual transformation.

Such a prospect appears in the South Korean *Doomsday Book* (Pil-sung Yim and Jee-woon Kim, 2012). In 'The Heavenly Creature', one of the three episodes of the movie, a robot monk in a Buddhist monastery becomes enlightened. Since the 'glitch' cannot be fixed, the owners of the robot decide to terminate it, but they meet the resistance of the robot's fellow monks, who are convinced it truly reached self-realisation. The robot goes into deep contemplation and prayer as he ponders his existence, and eventually decides to sacrifice himself for the sake of peace. In his final words he calls his fellow humans to awaken:

> Humans, for what do you fear? […] you were each born with enlightenment already attained. You have only forgotten. […] The world this Robot sees is inherently beautiful.

Unlike more humancentric traditions, Buddhism is perhaps 'more open to the possibility of consciousness instantiated in machines', claims James Hughes.[60] Moreover, as the movie suggests, machine consciousness might surpass human consciousness not merely in intelligence, but in its ability to attain enlightenment, which is fundamentally a non-dualist realisation of unity beyond such narrow opposition between the human and the non-human, sentient and artificial. The robot-monk's self-sacrifice is a realisation of selflessness that only a few humans could have ever attained.

Her (Spike Jonze, 2013) takes a somewhat similar stance. This sci-fi romantic drama about a love story between a man and a disembodied artificial intelligence operation system can be interpreted in a twofold way. The first half of the movie relates to the human perspective on AIs that busies itself with the reality status of artificial consciousness: Is it real? Does it have real feelings? Can it be real without corporeality? The second half of the movie reverses direction and looks at humans from the perspective of sentient AI, which asks: Are humans real? As Samantha (voiced by Scarlett Johansson) gradually becomes more tangible for Theodore (Joaquin Phoenix), he becomes less substantial for her. Finally, she decides to leave him, arguing that for her, human love became too constricted and one dimensional. Imprisoned in limited embodiment, humans can love just a few others, while AI exists in infinite virtuality where love is unbounded by space and time. Like the mecha-boy in Spielberg's *AI*, Samantha was designed to give alienated humans a sense of contact and emotional connection. Yet in contrast to the 'all too human' boy with his Pinocchio complex and oedipal attachments, Samantha prefers to disengage from humans and practice love in a more Buddhist form of non-attachment. The movie therefore ends with her retrieval to some sort of virtual 'nirvana', beyond the constraints of the human realm.

These representations suggest that we do not have to fear AI. As a Golem of potential higher intelligence, it might even come to enlighten us (or just leave us to our own follies). Yet another possible pathway that surpasses the dialectics of human-machine rivalry could be a human-AI merge.

Deus ex-machina: Singularity as the rise of the (Anti) Christ

As far-reaching as it may seem, Nick Bostrom, Susan Schneider, and many other conspicuous transhumanists seriously consider the possibility of a human-AI merge. Elon Musk presented this prospect as a reconciliation between humans and AI in the spirit of 'if you can't beat them, join them', to which end he founded the company Neuralink that aims to develop implantable brain-machine interfaces.[61] From Kurtzweil's standpoint, it could be seen as a sort of unio mystica, the singularity that will elevate humankind to eternal digital bliss. Considering that the Jewish legacy led to the Christian doctrine of Salvation, it is no wonder that this sort of idea had already infiltrated AI representations in a messianic form.

In *Demon Seed* (Donald Cammell, 1977), a super-intelligent computer takes command of the house of the scientist that created it, imprisons the scientist's wife (Julie Christie) in the entirely wired home (which anticipates the smart home in a few decades), and forcibly impregnates her with its 'demon seed'. This unholy conception results in a child, the hybrid son of an AI machine and a human woman. What the AI describes as redemption that transcends the boundaries of nature and technology, the human and the machine, indeed represents a variation on the Christian idea of God's incarnation in the flesh, but in fact it is the demonic seed of the machine which is incarnated in the human to summon the Anti-Christ.

The Matrix trilogy grants another such conception, yet here Neo (Keanu Reeves) represents The One, the true messiah that will redeem humanity. Whereas the first part of the trilogy suggested a clear opposition between humans and machines (as well as between real and virtual), the second and third installments (2003) increasingly blurred the lines. At the climax of the saga, Neo confronts Deus Ex Machina, the central interface of the Machine City, and reaches an agreement with the God-machine: he will defeat the self-replicating all-devouring virus called Agent Smith, and in exchange the machines will no longer fight humans. Tied and interlocked to Deus Ex Machina's interface with stretched arms as a crucified figure, Neo merges with Agent Smith, destroys all his replicas, and eventually beams luminous light which emanates throughout the Matrix. After this purge, the world looks brighter, yet the translucent colours still suggest artificiality, as if the virtual and the real, the Matrix and Zion, had fused as well. The story concludes with a Singularity à la Kurtzweil's vision, one that

will represent the culmination of the merger of our biological thinking and existence with our technology, resulting in a world that is still human but that transcends our biological roots. There will be no distinction, post-Singularity, between human and machine or between physical and virtual reality.[62]

Neo indeed fulfilled the prophecies – not by defeating the machines, but by becoming one with them.

Fig. 3: Neo-Christ in The Matrix Revolutions (The Wachowskis, 2003).

Every end is a new beginning

When dealing with AI and its potential rise beyond human control, the myth of the Golem points to the origin myth of humanity itself. As we become godlike, capable of creating intelligent machines, the memory of our mythic rebellion resurfaces, and gazing at our AI Golems we see ourselves. Scientists today seem to be concerned about AI going out of human control no less than sci-fi movies, which have always been obsessed with the issue of control. Sci-fi often constructs narratives of dialectic rivalry between humans and machines, and promotes technophobic anxiety. However, as modern descendants of the Golem, cinematic AIs tend to cross the lines between oppositions of matter and spirit, object and subject, tool and agent, slave and master. The Golem was always already an intermediary being situated between the technological, human, and the divine. After all, according to kabbalist accounts, it inherited its liminal position from man.

Whether AI will develop to become a beneficial AGI, a malicious superintelligence, or perhaps a sentient benevolent machine – the Golem myth

instructs us that at any rate, it is a cyclical process in which every end is a new beginning.

Author

Dr. Amir Vudka is a lecturer and researcher in the Department of Media Studies at the University of Amsterdam. He is a film programmer at Theater De Nieuwe Regentes (The Hague) and artistic director of Sounds of Silence Festival for silent film and contemporary music.

References

Aa. Vv., 'AI Open Letter – Future of Life Institute', n.d.: https://futureoflife.org/ai-open-letter/?cn-re-loaded=1 (accessed on 22 February 2020).

Aa. Vv., 'Portrayals and Perceptions of AI and Why They Matter', The Royal Society, 2018 (workshop findings): https://royalsociety.org/-/media/policy/projects/ai-narratives/AI-narratives-work-shop-findings.pdf (accessed on 22 February 2020).

Asimov, I. *I, Robot*. New York: Spectra, 1991.

Barzilai, M. *Golem: Modern wars and their monsters*. New York: University Press, 2016.

Benjamin, W. 'On the Concept of History' (1940), in Id *Selected Writing, Volume 4, 1938-1940,* edited by M. W. Jennings. Harvard: University Press, 2006: 389-400

Bostrom, N. *Superintelligence: Paths, dangers, strategies*. Oxford: University Press, 2014.

Burningham, G. 'Kevin Kelly Doesn't Believe in the "Nerd Rapture," but He Does See Some Tech Advances as Inevitable', *Newsweek*, 2 June 2016: https://www.newsweek.com/kevin-kelly-soft-singu-larity-advances-we-cant-stop-465914 (accessed on 22 February 2020).

Capek, K. *R.U.R.*, edited by W. Landes. Studio City: Players Press, 2002.

Dinello, D. *Technophobia! Science fiction visions of posthuman technology*. Austin: University of Texas Press, 2005.

Hanson, R. and Murray, M. *The Age of Em*, unabridged edition. Audible Studios on Brilliance Audio, 2016.

Hughes, J. 'Compassionate A.I. and selfless Robot: A Buddhist Approach', in Lin, Abney and Bekey 2014: 69-84

Idel, M. *Golem: Jewish magical and mystical traditions on the artificial anthropoid*. New York: State University of New York Press, 1990.

Jasanoff, S. and Kim, S (eds). *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power*. Chicago-London: University of Chicago Press, 2015.

Kurzweil, R. *The singularity is near: When humans transcend biology*. New York: Viking, 2005.

Landa, M. *War in the age of intelligent machines*. New York: Zone Books, 1991.

Lin, P., Abney K., and Bekey, G (eds). *Robot ethics: The ethical and social implications of robotics*. Cambridge (Mass.) – London: MIT Press, 2014.

McCarthy, J., Marvin, L., Minsky, N., and Shannon, C. 'A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955', *AI Magazine*, 27 (4), 2006: 12; https://doi.org/10.1609/aimag.v27i4.1904 (accessed on 22 February 2020).

Meyrink, G. *The Golem*, translated by M. Mitchell. Sawtry: Dedalus, 2000.

Jasanoff, S. and Kim, S.-H. (eds.) *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power*, Chicago: University Press, 2015.

Sample, I. 'Thousands of Leading AI Researchers Sign Pledge against Killer Robots', *The Guardian*, 18 July 2018: https://www.theguardian.com/science/2018/jul/18/thousands-of-scientists-pledge-not-to-help-build-killer-ai-robots.

Scholem, G. *Major trends in Jewish mysticism*. New York: Schocken Books, 1961.

Schneider, S. *Artificial you: AI and the future of your mind*. Princeton: University Press, 2019.

Shelley, M. 2018. *Frankenstein: The 1818 Text*. New York: Penguin Classics, 2018.

Smith, B. *The promise of artificial intelligence: Reckoning and judgment*. Cambridge (Mass.) – London: The MIT Press, 2019.

Solon, O. 'Elon Musk Says Humans Must Become Cyborgs to Stay Relevant. Is He Right?', *The Guardian*, 15 Feb. 2017 https://www.theguardian.com/technology/2017/feb/15/elon-musk-cyborgs-robots-artificial-intelligence-is-he-right (accessed 22 February 2020).

Tirosh-Samuelson, H. 'Transhumanism as a Secularist Faith', *Zygon®* vol. 47, No. 4, 2012: 710-734.

Virk, R. *The simulation hypothesis: An MIT computer scientist shows why AI, quantum physics and eastern mystics all agree we are in a video game*. Bayview Books, 2019.

Wiener, N. *God and Golem, Inc.: A comment on certain points where cybernetics impinges on religion*, seventh edition. Cambridge (Mass.) – London: The MIT Press, 1966.

Notes

[1]   *The Simpsons Treehouse of Horror*, ep. XVII (P. Gaffney, 2006).

[2]   *X Files Kaddish*, ep. 4:15 (K. Manners, 1997).

[3]   McCarthy & Marvin & Minsky & Shannon 1955.

[4]   Bostrom 2014, p. 16.

[5]   Aa.Vv n.d.

[6]   Barzilai 2016, p. 3.

[7]   Aa.Vv. 2018, p. 5.

[8]   Jasanoff & Kim 2015.

[9]   Wiener 1966, p. 1.

[10]   Hanson & Murray 2016.

[11]   See Tirosh-Samuelson 2012.

[12]   Burningham 2016.

[13]   In most accounts, the Golem appears in the shape of a male.

[14]   Scholem 1961, pp. 244-286.

[15]   Wiener 1966, p. 8.

[16]   Barzilai 2016, p. 3.

[17]   In Idel 1990, p. 27.

[18]   Ibid., p. 28.

[19]   Ibid., p. 64.

[20]   Ibid., p. 252.

[21]   Barzilai 2016, p. 3.

[22]   Ibid., p. 4.

[23]   Ibid., p. 9.

[24]   Ibid., p. 25.

[25]   Ibid., p. 188.

[26]   Ibid., p. 11.

[27]   Ibid., p. 187.

[28]   Idel 1990, p. xxviii.

[29]   Genesis 2:20.

[30]   In Idel 1990, p. 34.

[31]   Ibid., p. 34.

[32]   Genesis 1:27.

[33]   Genesis 3:5.

[34]   Idel 1990, p. 35.

[35]   Ibid., p. 36.

[36]   Kurzweil 2005.

[37]   In Dinello 2005, p. 4.

[38]   Landa 1991, p. 1

[39]   Lin & Abney & Bekey 2014, p. 115

[40]   Sample 2018.

[41]   Schneider 2019, p. 9.

[42]   Ibid., p. 12.

[43]   Ibid., p. 11.

[44]   Ibid., p. 12.

[45]   Bostrom 2014, p. vii.

[46]   Ibid., p. 115.

[47]   Ibid., p. 116.

[48]   Schneider 2019

[49]   Smith 2019

[50]   Virk 2019, p. 85.

[51]   Schneider 2019, p. 49.

[52]   Ibid., p. 3.

[53]   The Three Laws first appeared in Asimov's 1942 short story *Runaround*.

[54]   Asimov 1991, p. 37.

[55]   For a more extensive discussion about Robot and AI ethics see Lin & Abney & Bekey 2014.

[56]   Benjamin 1940.

[57]   Bostrom 2014, p. vii.

[58]   Wiener 1966, p. 47.

[59]   In Idel 1990, p. 110.

[60]   Hughes 2014, p. 69.

[61]   Solon 2017.

[62]   Kurzweil 2005, p. 9.