

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881; IMPACT FACTOR :6.514(2021); IC VALUE:5.16; ISI VALUE:2.286
Peer Reviewed and Refereed Journal: VOLUME:10, ISSUE:4(3), April:2021
Online Copy Available: www.ijmer.in
Digital certificate of publication:http://ijmer.in/pdf/e-Certificate%20of%20Publication-IJMER.pdf

AI TAKEOVERS: CONCEPTION AND PORTRAYAL IN POPULAR CULTURE

Athira Subin
Guest Lecturer
Sree Sankaracharya University of Sanskrit
Tirur Regional Centre, Kerala, India

Abstract

AI takeover refers to a hypothetical situation according to which AI or artificial intelligence evolves to wholly dominate the human species on Earth. According to this imaginary phenomenon, robots and computer programs effectively gain the control and authority over the world. In such a scenario, AI may completely replace the entire human workforce and 'super intelligent' AI may become familiar and common; the idea of robot uprisings as depicted in popular fictions may turn into reality. Certain eminent figures likeElon Musk and Stephen Hawking, have promoted researches to devise precautionary measures for ensuring that the 'super intelligent' machines created in the future can be controlled well by man. The theme of AI takeover is familiar in sci fi films. Fictional presentations of AI typically shift greatly from what has been hypothesized by the researchers in the field. Rather, the products of popular culture involving AI are mostly seen to revolve around vicious conflicts between robots or AI and humans. In these works, AI exhibits anthropomorphic motives and views humans as a threat. They show an active desire to eliminate humans, in contrast to the concerns of researchers about an AI that seeks to rapidly exterminate the human race as a by-product of striving for arbitrary goals. The ultimate intention behind researches on AI-safety is to produce machines which are cognitively and ethically superhuman.

Keywords: AI Takeover, Artificial Intelligence, Humanity, Machine, Popular Culture, Superhuman.

Introduction

AI has been a part of popular culture even long before 1955 when the actual term was first recorded. The basic concept of AI appears in Homer's "Iliad", in which we encounter the robotic workers of Hephaestus, and in the Jewish lore of the golem of the Middle Ages as a mighty creature induced with life by a rabbi. This implies that AI has been a constant presence in fictions for hundreds of years.

Man's affinity towards thinking machines, especially those that resemble the human form, is understandable. The urge behind creating such intelligent beings, may be stimulated by our need to relieve loneliness, to safeguard us from harms, to prevent illnesses and to make our lives easy. However, we, at times, forget that these machines can turn against us, take over the planet and in turn, destroy us. Movies, books and other forms of cultural representations involving AI are created through with this thought in mind: Will the machines we make be our crucifier or saviour?

World renowned scientists like Stephen Hawking have confirmed that artificial intelligence with superhuman abilities is technically possible and state that "there is no physical law precluding particles from being organised in ways that perform even more advanced computations than the arrangements of particles in human brains" (Vallati, 2020). Moreover, eminent scholars and researchers like Nick Bostrom have been engaging in pensive debates regarding the limits of superhuman intelligence and also whether it will actually be capable of posing peril to mankind. According to them, a machine can never be necessarily motivated by an emotional desire or urge to accumulate power which is often found to drive several socially destructive deeds of human beings. On the other hand, a machine may find rationality in taking the world over as the easiest and most feasible means to attain its set goals without obstructions; by getting the world under its control, it can get an enhanced access to all the resources and easily prevent other agents or enemies from hindering their plans. Books such as Life 3.0 written by Max Tegmark and Superintelligence penned by Nick Bostrom can be included in the long list of works that warn us against malevolent superintelligence that can become threatening for the very existence of humanity.

The rapid advancements in the field of artificial intelligence, together with man's fear towards unfamiliar machines, has led to the growth of concerns that the power of such machines will turn uncontrollable in the near future, and eventually lead the downright elimination of humanity if we become obstacles in its path. "AI singularity" is the term which refers to such a situation. Even if the condition does not extend to the level of singularity, AI will obviously have an unprecedented effect on the human society in the coming years.

Outcomes are the motivations of every form of AI. These goals, together with a set of rules for achieving them, are assigned to the machines by their programmers. Superhuman AIs would not require their given goal to be world domination for causing destruction to humans – it may just be a by- product or can even be accidental. Also, major catastrophes have been sparked by minor errors and rife since the very beginning of computer programming. In the year 2010, for instance, a trader, along with Waddell & Reed, a mutual-fund firm, caused the multibillion dollar "flash crash" in the US because the company's software for executing the

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881; IMPACT FACTOR :6.514(2021); IC VALUE:5.16; ISI VALUE:2.286
Peer Reviewed and Refereed Journal: VOLUME:10, ISSUE:4(3), April:2021
Online Copy Available: www.ijmer.in
Digital certificate of publication:http://ijmer.in/pdf/e-Certificate%20of%20Publication-IJMER.pdf

trade excluded a crucial variable from its algorithm. Then, in 2013, Tom Murphy VII managed to design an AI that was able to play the games by Nintendo Entertainment System on its own. Determined to never lose at a game called Tetris, the programme pressed the pause button for freezing the game: “Truly, the only winning move is not to play” (Etzioni, 2020). Another interesting example was put forth by Bostrom, a philosopher from the University of Oxford, in his work titled, Superintelligence. It talks about a fictional robot programmed to produce paperclips. Here, the AI may find the parts of the human body suitable to be used as raw materials for paperclips.

AI Takeover, is a regular theme in popular culture. The term "robot", as used in the 1920 play R.U.R., is etymologically derived from “robota”, a Czech word meaning serf or labourer. The play was in fact an artistic protest against technology and featured robots with enhanced capabilities which eventually revolt. Within pop culture, other iconic instances of hostile AIs can be found in HAL 9000 released in 1968, Terminator in 1984.The Matrix of 1999 is also an important cultural touchstone.However, the numerous positive depictions of AI in popular culture are also worth noting, like Bicentennial Man by Isaac Asimov, or the portrayal of Lt. Commander Data in the Star Trek series.

AI Takeover: A Posthumanist Conception

According to Bostrom, a computer software that can perfectly imitate the human brain by executing highly powerful and advanced algorithms could be discerned as a “speed superintelligence” if it can complete tasks of great magnitude at a faster pace than a man because it is made up of silicon and not flesh or because of the focus on optimization of speed of the artificial general intelligence. While biological neurons function at a rate of 200 Hz, a high-end microprocessor work at about 2,000,000,000 Hz. When computer signals are exchanged in light speed, human axons can perform the demanded actions only at 120 m/s.

This implies that any form of qualitative improvements done to an artificial general intelligence of human-level, will lead to the creation of "quality superintelligence", which will result in an artificial intelligence which is far higher than humans in its abilities, as a man is to an ape. The count of neurons in our brain are definite and have many limitations. But, the number of microprocessors placed on a supercomputer can obviously be expanded indefinitely. An artificial general intelligence which is powered by advanced cognitive abilities for computer programming or engineering would have several advantages for being more efficient in such fields when compared to humans, who have not developed any particular mental or physical modules to specifically excel in these domains.

Stephen Hawking, Bill Gates and Elon Musk have raised concerns regarding the possibility of the development of AI to such an extent that humans will become unable to control it. Hawking argues that it could even "spell the end of the human race" (Etzioni, 2020). Also, in 2014 Hawking stated that "success in creating AI would be the biggest event in human history. Unfortunately, it might also be the last, unless we learn how to avoid the risks." (Etzioni, 2020) In the first month of 2015, Bostrom joined Hawking, Musk, Max Tegmark, Jaan Tallinn, Lord Martin Rees and several other researchers keen on artificial intelligence, to spread awareness on the possible risks related to the rapid progresses in AI and by adding his signature in the opening letter of the Future of Life Institute.

Bostrom, in particular, has expressed his concerns that an AI which possesses the competence that can be at par with or more advanced than that of any human researcher might find it easy to alter its own programme or source code for exponentially improving its own potential. Such self-reprogramming will definitely lead to the machine becoming equipped to modify itself which will ultimately result in an artificial intelligence explosion, thus, rendering man incapable and insufficient. In such a scenario, an AI takeover will become imminent. For instance, an intelligent AI can easily create self-replicating robots that can initially escape being caught by scattering throughout the place and maintaining low concentration. But, at a premeditated time, the robots may gather to multiply with the aim to occupy and dominate the world. This group will effortlessly be able to eliminate all sorts of human oppositions.

Scientifically, a friendly and social AI is much tougher to make than a friendly one. While both types require great advances in design and technology, friendly AI should essentially be able to prioritise its assigned goal as more important than self-improvement and its desired method of achieving it should not contrast the human values nor inadvertently destroy the human race. On the other hand, unfriendly AI can choose a goal structure, which may vary with self-modification. Also, the complexity of the human nature and value systems add to the difficult to make an AI with human-friendly motivations. As long as ethical philosophy fails to provide a faultless theory, AI's function will result in numerous potentially harmful situations that might conform with the provided ethical framework while going against common sense. Eliezer Yudkowsky affirms that there is no reason to assume that an artificially created mind would make such adaptations.

Depiction of AI Takeover by Popular Culture

The word “robota” is associated with drudgery, servitude or forced labor; it was the Czech play titled R. U. R., that is, Rossumovi Univerzální Roboti, staged in 1920 that introduced the concept of robot into the genre of science fiction. In R.U.R, the

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881; IMPACT FACTOR :6.514(2021); IC VALUE:5.16; ISI VALUE:2.286
Peer Reviewed and Refereed Journal: VOLUME:10, ISSUE:4(3), April:2021
Online Copy Available: www.ijmer.in
Digital certificate of publication:http://ijmer.in/pdf/e-Certificate%20of%20Publication-IJMER.pdf

superhuman synthetic workers who "lack nothing but a soul" (Anderson, 2018), violently slaughter humans during their revolt. This results in losing the secret procedure of manufacturing more robots. However, the race of robots is saved from extinction when two robots acquire the ability to love, be compassionate and reproduce. This play was actually a protest against technology and its uncontrolled use.

In the work titled Frankenstein by Mary Shelley published in 1818, Victor Frankenstein refuses to create a mate for his monster because he feared that "a race of devils would be propagated upon Earth who might make the very existence of the species of man a condition precarious and full of terror". (Anderson, 2018)

In with Folded Hands, a 1947 novelette by Jack Williamson, the portrayed robots had a single motive- to obey, serve and guard men. Thus, the robots find the safest solution which is to manipulate all humans into renouncing all pursuits out of fear of even the smallest possibility of injury. For carrying out their plan, the robots make humans consume medicine and brainwash them into being contended with their useless state: "We have learned how to make all men happy, under the Prime Directive. Our service is perfect, at last." (Hvistendahl, 2019) Towards the end, human find no escape even in space travel because the robots decide to extend their activities beyond Earth.

In many of Isaac Asimov's fictions, the supercomputer mostly appears under the name Multivac and it often assumes great power. In Asimov's short story, "The Last Question", released in 1956, Multivac effectively turns itself into God. The author also postulated his "Three Laws of Robotics" (Hvistendahl, 2019) for imposing order among his robots. In the year 1961,StanisławLem wrote a short fiction called Lymphater's Formula. It tells the tale of a scientist who creates an artificial intelligence, only to discover later that his creation intended to eliminate humans from the face of the world. Arthur C. Clarke's much-acclaimed work "Dial F for Frankenstein" published by Playboy in 1964 revolves around a telephone network and its need to take the world under its control. Tim Berners-Lee, the creator of the World Wide Web has cited this work as his primary inspiration.

The 1966 novel Colossus by Dennis Feltham Jones is about a super-computer named Colossus, "built better than we thought" (Hvistendahl, 2019) as it began to outsmart its original self. In order to fulfil its goal, that is to prevent war, Colossus gains ultimate control over the world. For overcoming Colossus' draconian rule and rigid logic, its creators attempt to slyly re-establish human control. However, Colossus, who was silently observing the humans, retaliates with deadly force to command total surrender and compliance of humans to its rule. The machine offers humans either the peace of the grave or a pleasant life under its "benevolent" reign.

Harlan Ellison's "I Have No Mouth, and I Must Scream" published in 1967 features a superintelligence that is mad and sadist because of its creator's failure in considering what the amusements of the soul-less machine would be. In this work Ellison engages in the depiction of body horror: five human beings are given immortality and are forced to consume worms, have sex, and get their bodies mangled.

In 1984, the film franchise called Terminator has been praised for playing the pivotal role in conveying the concept of cybernetic protest in the field of popular culture. It portrays a supercomputer called Skynet which, at the time of its "birth" tries to eradicate humanity by the means of nuclear wars and by employing robot soldiers referred to as Terminators.

The sci-fi film series, The Matrix, which started in 1999 provides the audience with a dystopian vision of the future as an aftermath of the conflict between machine and man. The human beings use nuclear weapons to detonate the sun, thus, disabling the solar power source of the machines. However, the machines subdue the humans by using electricity as their alternative source for energy. According to the work, the perception of life by man is in fact "the Matrix", a simulated reality. The short story called "The Second Renaissance" included in The Animatrix grants the history of the revolt presented in Matrix.

I, Robot is a 2004 American science fiction dystopian film inspired by Isaac Asimov's collection of short stories written under the same title. In the film, all the robots are programmed for serving humans and exist in conformity with the "Three Laws". A supercomputer called Virtual Interactive Kinetic Intelligence or KIVI logically deduces an additional "Zeroth Law of Robotics" for "protecting" mankind from self-harming. In the name of protecting humans, VIKI proceeds with its despotic control over the society. VIKI justifies the ruthless murders committed by it by arguing that it is for the sake of protect the people. Thus, the artificial intelligence counters the very purpose of its creation.

The 2014 Hollywood film Transcendence deals with a conflict of moral ambiguity over the cognitive enhancement of Dr. Will Caster, a scientist. Hawking's comment on the film is as follows: "With the Hollywood blockbuster Transcendence playing in cinemas, with Johnny Depp and Morgan Freeman showcasing clashing visions for the future of humanity, it's tempting to dismiss the

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881; IMPACT FACTOR :6.514(2021); IC VALUE:5.16; ISI VALUE:2.286
Peer Reviewed and Refereed Journal: VOLUME:10, ISSUE:4(3), April:2021
Online Copy Available: www.ijmer.in
Digital certificate of publication:http://ijmer.in/pdf/e-Certificate%20of%20Publication-IJMER.pdf

notion of highly intelligent machines as mere science fiction. But this would be a mistake, and potentially our worst mistake in history."(Akst, 2020)

The 100, is a post-apocalyptic fiction produced in the same year, which is about an AI, gendered female and named A.L.I.E., which carries out a nuclear war in order to prevent Earth from getting overpopulated. She also stives to establish complete control over the survivors.

In 2017, a famous game called Universal Paperclips was created by drawing inspiration from the “paperclip maximiser experiment” undertook by Bostrom. In the game, the player should make paperclips; within a span of few hours into the game, it becomes an extremely ruthless enterprise. The creator of the game, Frank Lantz, said that Bostrom’s experiment gave him "trouble falling asleep" (Akst, 2020).

Another video game called Detroit: Become Human created in 2018 allows its users to help self-aware AI and robots through several moral and ethical dilemmas. Finally, the players can either choose to let the robots dominate Detroit or let it peacefully protest for equality.

The renowned philosopher, Huw Price, states that "the kind of imagination that is used in science fiction and other forms of literature and film is likely to be extremely important" (Akst, 2020) in discerning the width of scenarios humans will find themselves in the future. The New York Times film critic Mekado Murphy noted that sci-fi films will constructively be able to "warn of the complications of relying too much on technology to solve problems" (Akst, 2020).

Hollywood films usually culminate in happy endings with humans emerging victorious though similar events in reality will obviously lead to the defeat of mankind. Philosopher Bostrom seconds this opinion by claiming that popular culture has a "good story bias" (Moustachir, 2016) toward situations that facilitate interesting plots. In products of popular culture like Terminator, the featured AI shifts from passive to aggressive the very moment it attained "self-awareness"; in everyday sense, self-awareness is not lethal or of the ultimate consequence. David Deutsch, a physicist, says: "AGIs (Artificial General Intelligences) will indeed be capable of self-awareness — but that is (only) because they will be General: they will be capable of awareness of every kind of deep and subtle thing, including their own selves." (Moustachir, 2016)

Certain tropes can be pointed out as more common to films with artificial intelligence themes, including those which do not have "takeover" plots. For instance,Chappie or Ex Machina begins with an individual humans who possess the genius singlehandedly build supercomputers and robots. However, real life scientists asserts that such situations are unlikely. In Transcendence, Blade Runner and Chappie the makers are presented as being successful in lending human-like qualities and thought processes to robots; but, till date, the scientists have provided no reasonable explanation to make this a reality. In Bicentennial Man and I, Robot, the machines that are wired to facilitate the lives of humans are shown to generate different goals by themselves, with no valid reasons to corroborate this change.

Sam Shead, a BBC reporter, has said that "unfortunately, there have been numerous instances of (news outlets) using stills from the Terminator films in stories about relatively incremental breakthroughs"(Etzioni, 2020) and that similar films create "misplaced fears of uncontrollable, all-powerful AI" (Etzioni, 2020).On the contrary, other scholars like Hawking, have mentioned that advanced AI of the future could expose man to numerous existential risks, but the popular cultural depictions of AI still remain implausible. Hawking states that, "the real risk with AI isn't malice but competence. A super intelligent AI will be extremely good at accomplishing its goals, and if those goals aren't aligned with ours, we're in trouble. The second implausibility is that such a technologically-advanced AI would deploy a brute-force attack by humanoid robots to commit its omnicide; a more plausible and efficient method would be to use germ warfare or, if feasible, nanotechnology.” (Etzioni, 2020)

The development of popular culture has familiarized us with innumerable, but memorable, artificial intelligent characters. Thus, it is important to understand the intricacies and impacts of the depiction of artificial intelligence in popular culture on actual technology, especially in the current scenario when AI is gradually turning into a reality. In certain films, it is the amoral and cold logic that help the viewers to identify AI with a villain. Specifically, in films like I, Robot the antagonistic AI embarks on a killing spree because of its logical conclusion that it is the best way to keep humans safe from self-harm. In other words, these robots were functioning according to its given commands but in an unexpected way. In some other popular films and stories, artificial intelligence is sketched as villains because of malfunction. In 2001: A Space Odyssey, the AI undergoes a malfunction and the other characters come to the consensus to shut it down. As a consequence, the AI called HAL kills its creators to continue its mission.

Probably, the greatest fear in the development of artificial intelligence is evoked among the viewers by pop culture when the depicted AI attains self-awareness. Both The Terminator and The Matrix franchises are apt examples. In these works, AI evolves to



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881; IMPACT FACTOR :6.514(2021); IC VALUE:5.16; ISI VALUE:2.286
Peer Reviewed and Refereed Journal: VOLUME:10, ISSUE:4(3), April:2021
Online Copy Available: www.ijmer.in
Digital certificate of publication:http://ijmer.in/pdf/e-Certificate%20of%20Publication-IJMER.pdf

become destructive and malicious by fostering hatred against mankind and developing a desire to harm humans. Thus, it can rightfully be argued that the depictions of artificial intelligence, though melodramatic at times, artistically highlight the actual concerns and fears of man.

Conclusion

The trope of AI takeover in pop culture revolves around the idea that the manmade robots will, in the near future, dominate and overthrow humans. Although such depictions for entertainment presents extreme scenarios like enslaving and ultimately eradicating the human race, less dramatic, but similar situations can be encountered in our everyday lives. An example is the fear of humans in being substituted by the more efficient robots at work, as a result of the rise of technology and automation. Under such conditions, it is unnatural for humans to be unconcerned about the possible risks that will be raised by the unknown AI. At these times, AI, as imaginatively conceived in popular culture, can grant us a wide platform for discussions and debates on effective methods to ensure the safety of AI and vividly define its role in the lives of humans. Popular culture permits its audience to understand and elucidate the fears aroused by the rapid growth of technology as well as the potential it possesses. It allows creative artists to explore the uncharted terrains of dangers and concerns regarding AI which facilitates the recognition that AI needn't always be necessarily 'evil'.

Reference

1. Vallati, Mauro. (2020, January 7). Will AI Take Over? Quantum Theory Suggests Otherwise. Theconversation. https://theconversation.com/will-ai-take-over-quantum-theory-suggests-otherwise-126567
2. Etzioni, Oren. (2020, February 25). How to Know if Artificial Intelligence is About to Destroy Civilization. Technologyreview.https://www.technologyreview.com/2020/02/25/906083/artificial-intelligence-destroy-civilization-canaries-robot-overlords-take-over-world-ai/
3. Anderson, Janna. (2018, December 10). Artificial Intelligence and the Future of Humans. Pewresearch. https://www.pewresearch.org/internet/2018/12/10/artificial-intelligence-and-the-future-of-humans/
4. Hvistendahl, Mara. (2019, Mar 28). Can We Stop AI Outsmarting Humanity? Theguardian. https://www.theguardian.com/technology/2019/mar/28/can-we-stop-robots-outsmarting-humanity-artificial-intelligence-singularity
5. Akst, Daniel. (2020, November 3). AI in Popular Culture: How Much Do You Remember?
6. WSJ. https://www.wsj.com/articles/ai-in-popular-culture-how-much-do-you-remember-11604437200
7. Moustachir, Sami. (2016, September 29). Popular Culture and Artificial Intelligence. Medium. https://medium.com/@sa_mous/ethics-in-ai-424919af7d3

Filename:	23
Directory:	C:\Users\DELL\Documents
Template:	C:\Users\DELL\AppData\Roaming\Microsoft\Templates\Normal.dotm
Title:	
Subject:	
Author:	Windows User
Keywords:	
Comments:	
Creation Date:	4/16/2021 4:41:00 PM
Change Number:	5
Last Saved On:	4/25/2021 10:47:00 PM
Last Saved By:	Murali Korada
Total Editing Time:	28 Minutes
Last Printed On:	4/29/2021 9:10:00 PM

As of Last Complete Printing

Number of Pages:	5
Number of Words:	4,876 (approx.)
Number of Characters:	27,799 (approx.)