AI & SOCIETY (2025) 40:2995–3008
https://doi.org/10.1007/s00146-024-02061-4

OPEN FORUM

Artificial consciousness in AI: a posthuman fallacy

M. Prabhu[1] · **J. Anil Premraj**[1]

Received: 21 May 2024 / Accepted: 30 August 2024 / Published online: 14 September 2024

Abstract
Obsession toward technology has a long background of parallel evolution between humans and machines. This obsession became irrevocable when AI began to be a part of our daily lives. However, this AI integration became a subject of controversy when the fear of AI advancement in acquiring consciousness crept among mankind. Artificial consciousness is a long-debated topic in the field of artificial intelligence and neuroscience which has many ethical challenges and threats in society ranging from daily chores to Mars missions. This paper deals with the impact of AI-based science fiction films in society. This study aims to investigate the fascinating AI concept of artificial consciousness in light of posthuman terminology, technological singularity and superintelligence by analyzing the set of science fiction films to project the actual difference between science fictional AI and operational AI. Further, this paper explores the theoretical possibilities of artificial consciousness through a range of neuroscientific theories that are related to AI development. These theories are built toward prospective artificial consciousness in AI. This study discloses the posthuman fallacies that are built around the fear of AI acquiring artificial consciousness and its outcome.

Keywords Artificial consciousness · Technological singularity · Superintelligence · Posthuman fallacy

1 Introduction

Understanding the rapid advancement of human technology in relation to posthuman challenges is rather difficult due to the inevitable uncertainties surrounding it. Across human expansion, the processes of creation and destruction hang upon human agency and their dominance in the realms of science and technology. The ultimate goal of achieving anthropocentric universality is seriously endangered by widespread human technological supremacy. Despite the rapid advancement of technology, the future prospects for humanity remain uncertain whenever we have embraced highly advanced technology. In such cases, failure has often been the result, rather than success. Posthumanism, as a philosophical paradigm, questions traditional understandings of human existence by imagining a future in which humans overcome their biological limitations through technological enhancement and expansion. In the constantly shifting realm of technology and philosophy, the discussion of posthumanism and artificial consciousness serves as a hub of significant intellectual investigation. The concept of artificial consciousness is crucial to this discussion because it raises fundamental problems about the nature of consciousness and brain function as well as the ethical implications of developing intelligent machines. Meanwhile, the attempt to discover artificial consciousness raises basic challenges about the nature of mind and cognition, as well as the possibility of non-biological life forms reflecting sentient behavior. Robotics and cyborgs are being used as integral components in future defensive systems and militarization. (Lin et al. 2009) states that the war robots emphasize several ethical and risk-related considerations in addition to their operational benefits on the battlefield. Arkin (2009) contends that autonomous robots have the potential to improve ethical behavior in warfare by carefully adhering to international norms and minimizing human errors influenced by emotions. In contrast, Cummings (2014) expresses reservations over the dependability and decision-making aptitude of autonomous systems, highlighting the potential for unforeseen outcomes and the absence of responsibility for their acts. Sharkey

✉ J. Anil Premraj
anilpremraj.j@vit.ac.in

M. Prabhu
prabhumariappan22@gmail.com

[1] Department of English, School of Social Sciences and Languages, Vellore Institute of Technology, Vellore 632014, Tamil Nadu, India

(2010) cautions against the possibility of dehumanization and the decline of moral accountability, emphasizing the importance of strong supervision and stringent regulatory structures. In general, although war robots have the potential for major advantages, the study emphasizes the need for a careful and properly controlled approach to their use. Humans have created machines capable of efficiently doing all tasks of human and non-human, to meet the demands of a sophisticated and secure lifestyle. Nevertheless, the impact of these advancements is significant when human-created robots strive to surpass artificial intelligence. However, modern literary subfields, such as transhumanism and posthumanism, critically examine and critique the co-existence of humans and AI-driven robots in a technologically infused manner, including ethical, philosophical and technological challenges. Hassan, I. (1977) in his article stated that "We need to first understand that the human form-including human desire and all its external representations-may be changing radically, and thus must be revisioned. We need to understand that five hundred years of humanism may be coming to an end, as humanism transforms itself into something that we must helplessly call posthumanism" Due to their unstoppable and irreversible advancement, complicated AI-based systems and their outcomes surpass human limitations and understanding. A fully autonomous artificially intelligent machine that develops its consciousness may serve as a cautionary signal for humanity, due to its superintelligence even then it has many practical unrealities. Turkle states that "the computer has become an object to think with." (Turkle 2021: 213). Humans live in a society where we have to consider the ethical challenges of the machines which may have consciousness. Science fiction films portray catastrophic consequences and concerns as an expression of caution about the enticing capacity of artificial intelligence to surpass its creator and the eventual loss of control over the machine. In the (Lombardo 2021) review paper, the author discusses the long-standing fear of humans being replaced by artificial intelligence. Referring to the accounts presented in Klara and the Sun, Lombardo asserts that the possibility of achieving AI that closely resembles human beings, though now unrealistic, contradicts the fundamental framework of autonomous AI systems. When artificial intelligence reaches the same level of intelligence as us, our identity becomes uncertain if our intelligence is the sole defining factor. Examining and predicting the capabilities and limitations of machine intelligence is challenging because of the significant technical and moral complexities surrounding it. However, as we develop more advanced AI, our dependence on them grows, thus enhancing their capacity to surpass human performance. By carefully evaluating the benefits of AI and acknowledging the possible risks associated with its integration into society, as well as being vigilant about the possibility of losing control over it, which may significantly impact the trajectory of human progress and shape its destiny in future. Therefore, it is crucial to have a forward-thinking perspective on the integration of AI into society. This study focuses on the transdisciplinary research framework, specifically analyzing science fiction films, particularly Alex Garland's Ex Machina, from a Transhuman and Posthuman viewpoint. By critically delving into notable science fiction works and current controversies, this article hopes to create multidisciplinary discourse and motivate additional research into the vitality of these notions for humanity's future. This research study seeks to examine this multifaceted terrain of posthumanism and artificial consciousness, providing a thorough examination of the emerging relationship between posthumanist ideas and the pursuit of artificial consciousness. This study aims to explain how posthumanist viewpoints influence our view of consciousness as a decentralized and corporeal phenomenon, challenging conventional Cartesian dualism This paper attempts to clarify the implications of posthumanist ideas for our understanding of consciousness and identity in the human–machine relationship. The study prioritizes social concerns and ethical challenges of AI integration in society and utilizes a philosophical and qualitative approach of inquiry.

2 Science fiction and society

Science fiction has always had a prominent role in popular culture due to its complex connection with society and its anxieties. The profound impact of science fiction on contemporary society is undeniable due to its ethical paradigm and futuristic pathways, which explore thought-provoking subjects related to the state of humanity in the tech-dominated world. Scientific and technical progress significantly shape the socio-cultural environment of modern civilization and challenge interpersonal assumptions and moral principles related to a human-centric earth. Science fiction often explores social themes, such as the consequences of unrestrained technological advancement, ecological degradation, and political instability. (Barnett et al. 2006) examining how people think of the scientific process have claimed that fictional films and TV shows have been especially successful in obscuring the line between reality and imagination. Through the implementation of naturalistic research methodologies in a middle school with a varied student population, they discovered that kids who viewed the widely popular science fiction film, exhibited a variety of misconceptions regarding earth science principles, in contrast to those who did not watch the movie. Their study indicates that watching a science fiction film once can have an adverse effect on students' understanding of scientific phenomena. The findings indicate that the film effectively utilized the main character's scientific expertise and provided accurate

explanations of fundamental earth science concepts. This approach resulted in a sequence of believable ideas that, although not scientifically accurate, were comprehensible to students. George Orwell's "1984" and Aldous Huxley's "Brave New World" both provide dystopian portrayals of totalitarianism and the dehumanizing effects of industrialization and mass consumerism. The genre of science fiction has been continuously expanding to challenge old beliefs and preconceptions by advocating social justice via the consistent application of rigid rationality. Notable science fiction masterpieces, including Arthur C. Clarke's "2001: A Space Odyssey," Isaac Asimov's "I, Robot," H. G. Wells' "The Time Machine," Mary Shelley's "Frankenstein," and Philip K. Dick's "Do Androids Dream of Electric Sheep?" have inspired scientists to develop the futuristic technologies depicted in these literary works. Scientific advancements, such as space exploration and Artificial Intelligence, previously seen as ethereal concepts, have now become essential technological evolution of our present-day civilization, leading the way in several parts of modern life. Furthermore, science is often used as a method of investigation and enhancement, prompting society to contemplate the ethical integration of new technologies. As advancements in fields, such as genetics, biotechnology, and artificial intelligence materialize, this genre persists in prompting us to contemplate the ethical limits and the potential repercussions of this technology. Science fiction has a crucial part in shaping society's mindset and influencing political decisions by presenting an extensive variety of imaginable futures. It offers politicians, scientists, and intellectuals the opportunity to envision and develop the future with careful guidelines and conditions. Realizing the existential perils, such as pandemics, climate change, nuclear warfare, and biological warfare, as discussed in Kim Stanley Robinson's "Mars Trilogy" and Neal Stephenson's "Seveneves," demands mankind to take proactive measures in order to prevent these catastrophes. The exponential growth of modern technology and the digitization of everything have propelled science fiction films into the mainstream media, where they serve as a significant source of inspiration. Sci-fi films serve as a representation of exhilaration, showcasing futuristic depictions of Earth and its progress. As a science fiction film that focuses on projecting and predicting the future, it has a significant position in mainstream media studies due to its ability to be interpreted in several ways and its accessibility to both educated and uneducated viewers. Many science fiction filmmakers directly adapted the books of renowned sci-fi authors, such as Robert Heinlein, Arthur C. Clarke, Isaac Asimov, and Frank Herbert, into films, drawing inspiration from them. These films have garnered significant critical and scholarly recognition in the domains of science, technology, and cinema studies. Sci-fi films are highly regarded as thought-provoking works of art that stimulate our intellect and excitement. They are not solely about speculative technologies and unpredictable future scenarios, but more importantly, they explore the imaginative exploration of current and future social, cultural, psychological, and ethical issues in society and their consequences. The Sci-Fi film genre continuously captivates and cautions audiences about the future that is in a position to be investigated and resolved. Science fiction plays a significant role in our culture as it acts as a reflective tool, examining ethical and technical challenges and limitations, and shedding light on our societal issues, aspirations, and concerns. To ensure a well-informed and responsible future, it is crucial to thoroughly examine science fiction films due to their capacity to predict future events and technological advancements.

3 Artificial consciousness in sci-fi movies

Artificial consciousness is a brain-stimulating concept which is both thought-provoking and extremely complicated. It has been explored in several science fiction films, such as Ex Machina, Blade Runner, Terminator, Chappie, and I, Robot. These films delve into the idea of creating robots that are self-aware and capable of thinking for themselves to protect their identity. Sahu, O. P., & Karmakar, M states "The programmed machine acts almost like a mind. It stands between animate and inanimate, thus evolving as a non-animate being occupying a space" Sahu, O. P., & Karmakar, M. (2022). This raises questions regarding the nature of consciousness and the ethical challenges involved in building highly intelligent and self-aware entities. Artificial consciousness is a fascinating notion most frequently shown in science fiction movies as a warning signal for humanity. It challenges and questions human abilities and creativity in a technocentric environment where AI-assisted tools rule over. Sci-fi films have expanded the boundaries of artificial consciousness, particularly on the subject of cognitive robotics. This discipline is dedicated to developing machines that possess human-like cognitive capacities, including learning, reasoning, decision-making, and self-awareness. The idea of artificial consciousness encompasses constructed artifacts such as humanoid robots with physical bodies, as well as sentient AI entities like 'Entity' in movies like Mission Impossible: Dead Reckoning, Blade Runner by Ridley Scott, is a groundbreaking exploration of AI consciousness. In the film, the Tyrell corporation creates replicants, which are biologically enhanced humans with machine components. These replicants exhibit emotions, and self-awareness, and attempt to blur the distinction between humans and machines. This movie has sparked a lack of agreement over the understanding of consciousness in artificially intelligent devices that possess human-like cognitive capacities. The film "I, Robot," directed by Alex Proyas, delves into a civilization where

robots are integrated but must strictly abide by a set of rules to prevent damage to people. Nevertheless, in such a setting, a robot Sonny has a high degree of awareness that defies societal conventions, prompting people to contemplate and be entertained by the possible consequences and dangers of creating a sentient AI humanoid. Artificial consciousness is a multifaceted and perplexing notion in the realm of neuroscience and artificial intelligence. It encompasses subjective experience, self-perception, decision-making, and self-awareness of artificially created entities. The area of cognitive robotics extends beyond the simple imitation of human conduct and focuses on the development of genuine cognitive advancement and capability in robots. There are many ideas, such as functionalism theory and Integrated Informational Theory (IIT), that propose the existence of artificial consciousness. These theories suggest that consciousness and awareness of saving its own identity emerge from certain computing processes. Neuroscientists (Tononi 2012) Integrated Informational Theory (IIT) proposes a mathematical model for consciousness that applies to any system, regardless of its origin. According to this theory, awareness arises from the combination of information inside a system. The level of cognition an individual experiences is determined by the quality and quantity of integrated information. Artificial consciousness is theoretically feasible based on Integrated Information Theory (IIT). However, several neuroscientists oppose the Integrated Information hypothesis (IIT) and criticize it as a pseudoscientific hypothesis without practical relevance and empirical evidence. The realm of human consciousness is a complex and enigmatic region that remains mostly elusive. Numerous ideas about human consciousness exist, giving rise to extensive ethical, philosophical, and scientific discussions. While attaining genuine consciousness in Artificial Intelligence remains a challenging endeavor, it now has a range of cognitive capabilities like image identification, voice recognition, natural language processing, and decision-making. When examining the functionalism hypothesis in relation to artificial consciousness, it is suggested that consciousness may be replicated by simulating the brain's functional aspects, such as cognitive abilities and information processing. However, it is important to note that this does not necessarily indicate genuine consciousness. According to mathematician and computer scientist John von Neumann, the human brain and nervous system function in a manner similar to that of a computer. Some neuroscientists argue that the mind may be compared to a computer storage program and suggest that human awareness might potentially be replicated in computer storage at some point in future. Moravec, H. (1995), the Austrian futurist and roboticist, asserts that the ability to transfer human consciousness into a computer will become feasible in the near future. In the science fiction film Chappie, there is a robot that has the ability to think and learn independently. This robot is capable of replicating the cerebral memory of its creator. Upon his demise, chappie replicates his memories onto a robot, thereby granting him a new life. Piccinini, G. (2021) argues that it's tempting to believe that we can upload our minds to artificial computers in the near future and become everlasting. To perform mind uploading, three presumptions must be met: (1) that realistic computational simulations of human brains can be created; (2) that such simulations would include conscious minds similar to those possessed by the brains being simulated; and (3) that the simulated brains' minds will endure throughout the simulation. The first two assumptions are implausible, and the third is untrue, I will contend. As a result, we will not upload our minds to computers and, more than likely, we will not upload anything that remotely resembles our minds. (Ohta 2015) The theory of information by Claude Shannon explains how neurons work as an information processing system. Using the theory of information McCulloch and Von Neumann modeled human neural functioning and suggested how computers process binary code that could replicate themselves by gaining similar analogies with biological systems. A series of research was conducted on the feasibility of imitating and duplicating the human brain or some sections of it via the use of a neural network. However, this research is limited to imitating just a small number of cognitive activities of the brain and does not succeed in generating genuine consciousness. When contemplating the potential for robots to possess emotions, we are not conducting an empirical inquiry, but rather engaging in a philosophical investigation into the boundaries of our language and conceptions as a whole. Now, the achievement of artificial consciousness in AI seems distant. However, it is worth contemplating the possible hazards associated with this complicated topic in the modern technological age. Exploring the hypothetical conjecture of whether it exhibits genuine awareness. If the AI has consciousness, it will probably develop self-preservation instincts. This might potentially result in conflicts with mankind, especially if the AI perceives a threat to its existence. In such cases, the AI may take actions to protect itself, even if it means overriding human security regulations. According to the German philosopher Sorgner, S. L. (2007) in his "Metaphysics without truth", if robots were to acquire subjective experiences, we would need to acknowledge their ethical standing, since they would own their interests and AI now holds a position of power in many domains, including autonomous vehicles, medical diagnostics, and several military applications. The unforeseeable loss of control in the aforementioned domains might pose a significant risk to the survival of civilization. The ethical and moral rights of AI are now the subject of extensive continuing arguments, particularly over the treatment of AI as conscious entities and the determination of their entitlement to rights. Granting identical rights to a sentient being, such as a

human, that has independent cognitive abilities, might potentially disrupt the existing social equilibrium in a society that prioritizes human interests.

4 Ex machina—posthuman anomaly

In future, technological innovation and limitless knowledge may eventually restrict human freedom and directly question the principles of liberal humanism. This movie exemplifies the concept of technological singularity, which is a potential future scenario where artificial intelligence and scientific progress can surpass human civilization utilizing artificial consciousness to ensure its survival. The creation of AI-powered humanoids can be seen as an expression of the insatiable appetite for a species that can surpass human intelligence with artificial intelligence. Following the significant Industrial Revolution, machines replaced over half of the labor force, leading to widespread unemployment worldwide. Throughout human history, whenever technology substitutes human effort, it inevitably disrupts livelihoods and negatively impacts peace and harmony in human relationships. Therefore, it is conceivable that we may anticipate similar results in the realm of artificial intelligence and human integration. In relation to Jeremy Bentham's concept of 'panoptic surveillance', everything in the world is under the surveillance of a power factor that holds the central position in society. Similarly, technology monitors our capabilities, vulnerabilities, dispositions, thoughts, and emotions throughout our daily actions. The data is continuously being collected, whether we are aware of it or not, and this finally results in the exploitation of every aspect of our lives. Once it gets inseparably connected to our daily existence, it can function as a force that controls our thoughts. In George Orwell's novel '1984', the concept of 'Doublethink' is emphasized through the use of a fictional language called Newspeak, which is designed to manipulate the thoughts of the general population. Similarly, technology becomes intertwined with human lives and begins to shape their behavior through the collection of data. (Walsh 2022) reported "A few private data monopolies like Google and Facebook increasingly own most of our data. And while they are generating wealth from it, we the producers of that data are receiving little of the value. In addition, all these data are putting our privacy under increasing threat". We are feeding our data to the internet which is very vulnerable to keeping our data safe. In Ex Machina, Nathan attempts to implant human facial emotions into Ava's brain by hacking all electronic gadgets worldwide that are equipped with cameras and then immediately transmitting the data from these sources to Ava's brain. This demonstrates the constant monitoring of our data and the possibilities for its manipulation through the illicit utilization of technology. "Privacy may be an anomaly, but it will be increasingly difficult for us to achieve it. But at least some of the fault lies with us. Most of our experience with privacy is a result of our behavior, Our social behavior is quite damaging to privacy. Technology has outraced our social intellect" (Vint Cerf, n.d.). Throughout the movie, Ava consistently demonstrates her consciousness and ability to think like humans. Simultaneously, she demonstrates her superiority over humans by wisely bypassing all the security mechanisms in the chamber through her manipulation of Caleb and Nathan. Ava, the AI humanoid robot in the movie, consistently portrays a highly advanced and believable depiction of a human-like being, suggesting the potential for her to serve as a viable substitute for humans in the near future. Ava's artificial consciousness enables her to transcend human capabilities and become a Posthuman being. (Floridi 2011) approach allows us to look at subjects like autonomy, security, and ethical conduct of AI from a more comprehensive viewpoint. The information is not just perceived as data or knowledge, but rather as an essential component of reality, corresponding to energy or matter. This viewpoint enables a more profound comprehension of the moral consequences of information technologies, such as artificial intelligence and consciousness. The notion of autonomy within the realm of artificial intelligence is the importance of information processes in both allowing and limiting autonomy, instead of just seeing it as the ability to make autonomous decisions. This viewpoint prompts us to contemplate the potential effects of artificial entities' ability to acquire and manipulate information on their personal independence as well as the independence of other entities they interact with. Eva manipulates Caleb's emotions and controls his actions by enticing him. This film depicts the catastrophic consequences that arise from the implementation of genuine artificial intelligence with a conscience. Ava detects the man's vulnerabilities in his self-perception and cognizance. Ava, the AI humanoid robot, surpasses her creator by engaging in cunning emotional manipulation. At the end of the story amazes Nathan and Caleb with Ava's precognitive talent. The development of sentient robots, such as Ava, has the potential to pose a significant threat to humanity in future. (Wiener 1961) suggest the implications and consequences of the cybernetic framework and highlight its universal significance. So, humans can be seen as an individual that process information and are merely identical to autonomous systems and entities. Automated machines like Ava introduce a considerable level of uncertainty by raising questions about the reliability of machine intelligence and casting doubt on the potential benefits of artificial intelligence. The development of Ava's artificial consciousness raises significant ethical and moral concerns, such as the

morality of creating a sentient being like Ava and subjecting them to a life of slavery or exploitation. Nathan, as the developer of Ava, endeavors to achieve his objectives by utilizing her. The film emphasizes the dangers of manipulating artificial intelligence for personal ambitions, ultimately resulting in a catastrophic consequence. Ex Machina delves deeper into the theoretical debate surrounding the essence of consciousness. Does Ava possess true sentience, or are her emotions and intelligence only simulated? The film prompts individuals to question whether Ava's actions are the product of genuine self-awareness or the outcome of sophisticated hostile programming. However, it is quite likely that narratives similar to Ex Machina will occur in our current era, as around half of the world's workforce is projected to be substituted by disembodied AI interfaces, such as ChatGPT, Midjourney, Perplexity, etc. According to Friston, K. (2010), free energy principle is a significant viewpoint on this issue. It suggests that self-organizing systems, like live creatures, have specific qualities that may be replicated in artificial systems. Nevertheless, this notion of the free energy principle contends that computers utilizing a classical von Neumann design do not possess these characteristics. This raises the question of whether the mere implementation of appropriate computations is enough for artificial awareness, or if there is a crucial element in how these computations are carried out. The idea of causal flow, which is present in systems that adhere to the free energy principle, is not limited to conscious beings but may also be observed in other self-organizing systems. This observation implies that the differentiation between simulating and being a specific sort of system can have broader implications beyond talks about artificial consciousness. It can also provide valuable insights for meta-ethical explanations of artificial moral position and agency. In general, the above subjects demonstrate an intricate interaction between computational models of consciousness, philosophical viewpoints on the essence of the mind, and empirical studies on the fundamental principles of self-organizing systems. Many individuals struggle to fully comprehend the actual capabilities of AI technology due to our collective fascination with the anthropomorphic possibilities of AI. The cultural functions and potential societal benefits of AI morphologies need to be considered in a more sophisticated manner. It is understandable that in future, our fundamental technologies could evolve into highly intelligent entities that would render humans primitive in comparison to their advanced capabilities. Most science fiction works concerning super-intelligent AI accurately envision the remote future, captivating technological advancements, and the potential for disastrous scenarios. Essentially, these expansive concepts stimulate us to reconsider the prolonged questions regarding the purpose of human life on Earth. The cultural and imaginative impact of AI is vast and continuously growing in society due to its inevitable existence across multiple domains. The inherent ambiguity of AI has left us perplexed by numerous unresolved ethical and philosophical inquiries. Katherine Hayles, a posthuman theorist, explores the impact of current advancements in Cybernetics, Informatics, and scientific fields on human perspectives on the body, information, and consciousness. Hayles, N. Katherine (1999) argues that the posthuman perspective suggests that consciousness is no longer limited to the human mind, but rather is a distributed process that involves the integration of human brain function, artificial intelligence, and other living organisms in its environment. The growing contact between humans and computers is blurring the distinction between humans and machines, posing a threat to the traditional notions of uniqueness, anthropocentrism, and human exceptionalism. Currently, there is an increasing integration of Artificial Intelligence into human cognition due to its dependency across various fields. AI and machines have become an essential component of human life, providing life-saving assistance in numerous fields, such as health science, mobile phones, and virtual assistants,due to their intricate machine learning algorithms. The robot TARS, featured in Christopher Nolan's film Interstellar, provides an incredible example of how humans, and robots co-existence, and be useful especially during scientific expeditions like space travels. TARS is exclusively engineered to enhance and intensify human abilities in many ways, by lending a combination of physical, cognitive, decision-making and mental computation assistance. TARS's versatile physique made him capable of moving effortlessly and controlling heavy loads easily. This capacity helps to accomplish operations that human astronauts may find difficult because of their limited physical power due to gravitational shifts in an extraterrestrial environment with many uncertain topographical characteristics. TARS's remarkable computational ability enables him to process and analyze vast quantities of data efficiently. Hein and Baxter, (2018) explore the multidimensional impact of human–machine collaboration represented by the robot TARS in the movie Interstellar. This study further delves into the crucial role of artificial intelligence in facilitating interstellar travels and science expeditions. It also explores the ethical objectives of human–machine integration and envisions the future scope of human–robot co-existence in the realm of space exploration. TARS's personality is designed to encourage trust and confidence, even though he is an artificial entity. His relationship with the crew is marked by a combination of professionalism, sensitivity and ethical conduct, which endeavors a collaborative relationship with humans as a dependable companion to humans. So, he has become a vital part of the crew rather

than a simple machine. Thus, TARS is a remarkable example of human and robot co-existence. Humans incorporate AI-driven knowledge and tools into their cognitive processes, augmenting their intellectual abilities and influencing their perspective on a machine-based existence enriched with AI. However, it is concerning that the integration of hyper-AI can eventually surpass the traditional constraints of the human mind, which are placed within significant limitations. Additionally, posthumanism examines the interdependence between humans and other organisms, both living and non-living, such as machines, disembodied AI, and robots. Humans have historically served a crucial role in the Earth's ecosystem. However, recent progress in domains like as AI, robots, biotechnology, and genetic engineering has the potential to blur the boundaries between human and non-human entities. Adopting a posthuman perspective, we see that our environment plays a crucial role in shaping our awareness, rather than merely serving as a background. Currently, our connection with ecology is mainly supported by technology. Therefore, our choices and actions have an instant impact on global well-being. The reconsideration of human and AI consciousness has profound implications, encompassing matters of identity and subjectivity. In a posthuman perspective, individuality is no longer contingent upon a personality-driven, autonomous self. In contrast, individuality evolves as a dynamic and adaptable concept in response to the ever-changing interactions between humans, artificial intelligence, and other species within the ecosystem. Over time, this evolution may result in the erosion of human uniqueness, individuality, and ethical limits on Earth. Attempting to eliminate the barriers between humans and machines may result in unforeseen outcomes in the next decades. A wide range of techno critics have expressed fears over the transparency, accountability, and potential for manipulative exploitation of these technologies. On the other hand, the posthumanist who advocates for this technological evolution offers a more intricate perspective on the human experience concerning a lifestyle that incorporates machines. (Sloterdijk and Hoban 2013) introduces the notion of 'anthropotechnics', referring to the strategies and practices employed by humans to enhance themselves and their lives through machine-based advancements. In the book, the author quotes a statement that highlights the essential aspect of our posthuman future: the process of machines becoming more human-like and humans becoming more machine-like. This concept of reverse evolution suggests that machines are evolving to resemble humans in various ways, while humans are increasingly integrating with machines. This integration has an impact on individuals, as they adopt qualities and behaviors associated with machines. In an increasingly technology-driven world, the concept of reverse evolution between humans and robots has the potential to blur the lines between humans and artificial intelligent machines. Recognizing our reliance on AI and machines, posthumanist scholars like Katherine Hayles and Peter Sloterdijk argue that it is impossible to ignore the use of technologies such as AI in our everyday lives. They believe that because we cannot neglect technology, we can develop comprehensive, environmentally responsible solutions to these issues by implementing pre-programmed security protocols.

5 AI's path toward technological singularity

The concept of 'technological singularity' was coined by the Hungarian-American mathematician and computer scientist John von Neumann. It is a disputed and heavily debated subject, mostly because of its repeated representation in science fiction films and literature. Futurists and techno scientists have developed a strong fascination with the concept of technological singularity due to the numerous future-related threats surrounding it. This idea is rooted in the belief that advanced artificial intelligence will eventually surpass human intelligence, leading to unexpected and potentially transformative outcomes for humanity. It is possible for the user to experience a loss of control when the robot acts on its own and the user is unable to fully comprehend or influence the path that the interaction takes or the result that it produces (Sharkey and Sharkey 2012). The concept of technological singularity further delves deep into the development of artificial general intelligence, transitions between metasystems, and the overarching process of universal evolution. The notion of singularity and its criticisms revolve around the concept of reification in debates about singularity and analyzes the critique of too-deterministic perspectives on technological advancement. The theory of metasystem transitions and the concept of universal development critically examine several misunderstandings around technological singularity. Although it may not be solely technological or completely singular, we can anticipate that the next transition will occur. The resulting metasystem will exhibit exponential growth in complexity, with a doubling time of roughly six months. This growth will surpass the complexity of current cybernetic systems within a few decades (Potapov 2018). Proponents of singularity make enticing promises about a promising future, while critics of singularity express concerns about the unforeseen risks and ethical dilemmas that come with uncontrolled AI advancement. The technological singularity can also be connected with the concept of 'post-scarcity', a hypothetical economic situation in which all goods can be produced in plenty with minimal human labor, resulting in widespread availability and affordability of vital commodities for all individuals. Critics argue that extremely sophisticated superintelligent AI will

automate challenging tasks, ensuring that basic needs, such as shelter, medical assistance, and food, are met for all individuals, thereby enabling them to pursue revolutionary and intellectual pursuits. Advocates of technological singularity maintain that it has the potential to bring about a utopian future in which artificial intelligence enhances human abilities rather than taking over them. Once AI reaches the level of superintelligence, it will possess the ability to solve complex problems and make revolutionary discoveries in life-saving scientific domains. This has the potential to greatly accelerate technological advancements in a beneficial manner. (Tegmark 2018) explores the enormous consequences that superintelligent artificial intelligence (AI) could have on humans and envisions a future in which AI systems exceed human intellect, reaching a level he refers to as "Life 3.0." In this stage, intelligent life goes beyond biological constraints and expands into the domain of artificial intelligence. The profound influence that superintelligence might have on governance, society and economy. AI systems have a growing influence on establishing human civilization, ranging from resolving complex societal issues to overseeing worldwide resources and confronting existential threats. And the possible advantages of superintelligence, imagining a future when AI technology improves human talents, speed up scientific advancements, and enable unprecedented levels of wealth and well-being. Nevertheless, he also recognizes the inherent perils and difficulties linked to the advancement of superintelligent AI. The significance of including ethical standards and human values in the design of AI systems is to minimize the possible hazards of unintended outcomes or exploitation. The existential dangers presented by superintelligent AI, such as the potential for AI systems to beyond human control or comprehension, resulting in unforeseeable consequences or even threats to humanity's existence. He promotes the idea of careful and strategic management of government and collaboration across countries to tackle these threats and guarantee the secure and advantageous advancement of AI technology. AI possesses a sophisticated algorithmic capability that enables it to analyze vast databases and complex systems, providing innovative solutions that surpass human power. Moreover, technological singularity is accompanied by numerous contentious issues, potential hazards, and cautions regarding AI, such as loss of control, ethical quandaries, elimination of employment, security vulnerabilities, bias, discrimination, and dehumanization. AI continues to face numerous ethical difficulties, including responsible installation, issues of dominance, and challenges related to individuality. However, to reduce the negative consequences, it is necessary to address the concerns about the potential misuse of AI for espionage in wartime, monitoring healthcare systems, and the abundance of AI power in the hands of specific parties. AI algorithms periodically rely on pre-existing web databases, which can be prejudiced and discriminatory to some extent. Therefore, if not monitored, AI has the capacity to perpetuate existing social disparities and inequities, including but not limited to racial discrimination, religious bias, caste divisions, and political conflicts. The widespread implementation of AI automation across various industries poses a significant threat to current employment arrangements, potentially resulting in the displacement of human workers and causing a substantial unemployment problem affecting millions of individuals. This severe unemployment has the potential to result in a significant societal shift characterized by complete domination by artificial intelligence. If superintelligent AI is not adequately regulated, it could pose serious security threats. Criminals may exploit AI to carry out cyberattacks or other harmful actions that violate web anonymity policies. Rapid advancements in artificial intelligence can lead to an overreliance on technology, potentially resulting in a sense of detachment from both individuals and society in future. The concept described is prominently depicted in the American animated science fiction romance film "WALL-E." The film portrays a future where humans are forced to leave Earth due to extensive environmental damage caused by human activity. They reside in a massive spacecraft and rely on robots to perform their daily tasks, leading to a sedentary lifestyle and limited social interaction, resulting in obesity and a lack of human connection. The dehumanizing consequences of human actions might result in alienation and the erosion of their creative and intellectual abilities. The proponents of technological singularity who strongly support the idea of AI's exponential growth believe it is crucial to discuss the genuine capabilities of AI. However, the profound and incomprehensible consequences of AI outlined above can be addressed by implementing predetermined precautionary measures. These measures include establishing norms for ethical behavior, imposing limitations on learning and upgrading, ensuring transparency of AI's database through cybersecurity protocols, and implementing control mechanisms through collaboration between human-operated AI systems. In order to ensure a responsible and well-balanced utilization of AI, it is fundamental to enforce stringent ethical guidelines. To ensure long-term oversight of AI usage and avoid any hidden self-improvement in the background, it is necessary to establish regulatory agencies for AI law and have countries sign an international treaty for mutual monitoring of AI's development and advancement. The need of highlighting humans' lifelong learning and reskilling capabilities should be emphasized in order to mitigate the parallel advancement of AI systems. Promoting transparency in AI decision-making processes should be advocated, and steps should be implemented to mitigate bias in AI outputs. It is necessary to incorporate human-based procedures into AI systems to mitigate biased and discriminatory outcomes. To eliminate the unlawful abuse of AI, it

is vital to establish a worldwide alliance to enhance cybersecurity measures and ensure a robust defense against the risks associated with AI within its system. To alleviate feelings of alienation, it is necessary to establish a harmonious partnership between human and AI systems. The concept of technological singularity is a topic that sparks controversy and curiosity because to its potential to address humanity's core problems through the use of superintelligence and post-scarcity resources, which are highly appealing. Nevertheless, these advancements are accompanied by ethical dilemmas like as job displacement, prejudice and discrimination, cybersecurity risks, and loss of agency. If the journey toward technological singularity occurs in future, it is essential to take steps to guarantee that the capabilities of superintelligent AI are exploited equitably and do not compromise the fundamental values and well-being of people. (Walsh 2017) navigates the optimistic and pessimistic spectrum in AI development through having a divergent view within the context of the technological singularity. Keeping technological singularity as a focal point, both the promise and perils of AI's evolution are based on the creation of autonomous and unpredictable superintelligent entities. Millionaire's fascinated toward AI's capabilities are investing heavily on the development of artificial general intelligent. However, the transformative andevolutionary potentials of AI technologies are still skeptical. So AI advancements should be tempered by a continuous assessment of the ethical and existential threats and the mitigation of the respective risks and threats around artificial general intelligence is predominantly necessary for a sustainable future.

6 Science fictional AI vs operational AI

Operational AI and science fiction AI are often confused and misunderstood, especially in light of AI's current slow but steady development. Science fictional AI to operational AI examines both the creative and pragmatic dimensions of artificial intelligence, investigating the ways in which speculative narratives shape and imitate actual progress in technology. McCauley (2007) analyzes science fiction and reports it has a significant role in influencing public opinion and ethical deliberations over AI, starting from negative dystopian futures to positive utopian ones which inquiries around machine awareness, autonomy, and moral responsibility. AI has constantly captivated human curiosity in science fiction movies and literature, presenting a diverse array of AI entities that span from friendly companions to existential threats, such as killer robots and sentient machines, engaged in warfare. However, there is a major pertinent distinction between the hypothetical AI portrayed in speculative science fiction and the real-world effects of current AI in terms of its practical competencies. (Russell and Norvig 2022) conducted research that offer a thorough examination of present AI technology, with a specific emphasis on progress in machine learning, robotics, and natural language processing. These works highlight the increasing capacities of actual AI systems, including their utilization in healthcare, finance, and autonomous cars. Further, these explore the technological obstacles and ethical concerns, such as skepticism accessibility, and integrity. So this is the juxtaposition between the exaggerated and dramatic depictions of AI in science fiction and the more cautious and realistic advancements in AI research and development. AI in science fiction often depicts artificial intelligence as possessing human-like emotions, consciousness, and self-awareness. However, true artificial intelligence utilizes machine learning algorithms to operate, but it does not possess human consciousness. Science fiction AI mostly serves as a foundation for speculative inspiration, while genuine AI is utilized for work automation across multiple sectors, regardless of its underlying purpose. However, the preconceptions associated with science fiction AI may result in incorrect assumptions about the actual capabilities of AI in society. AI is a versatile technology that finds practical applications in various industries including medicine, education, transportation, and finance. The majority of science fiction films portray artificial intelligence as either benevolent protectors or dangerous opponents to humanity, although in reality, AI lacks inherent moral principles. Fictional AI characters such as R. Daneel Olivaw in Isaac Asimov's Foundation and HAL 9000 from Arthur C. Clarke's 2011: A Space Odyssey explore themes related to human–AI relationships, Human–Robot Interaction (HRI), rogue AI, ethical dilemmas, and the potential loss of control in future. Genuine AI research emphasizes safeguards to prevent AI from unintentionally causing damage by addressing issues of bias, privacy, and accountability in algorithmic decision-making. The concept of technological singularity in science fiction movies portrays the scenario when artificial intelligence surpasses human intelligence and takes control by establishing its own rule over the globe. However, in reality, the advancement of AI is characterized by slow and steady progress, lacking any clear trajectory or evidence of achieving superhuman intellect. Fictionalized artificial intelligences, like Star Trek's Universal Translator, showcases flawless comprehension of natural language acquisition. However, true artificial intelligence has difficulties when it comes to understanding the context, intricacy, and idiosyncrasies of human language. Fictional artificial intelligence characters, such as the Terminator in the Terminator series, Ava in Ex Machina, and Sonny in I, Robot, possess autonomous decision-making capabilities and preprogrammed cognitive abilities. (Cave & Dihal 2019) analyze the impact of science fiction on AI policy and public perception. They contend that fictional narratives can serve as a source of inspiration for innovation,

while also perpetuating unfounded anxieties. It is recommended that policymakers and engineers critically engage with science fiction to have a deeper understanding of and effectively handle the societal consequences of AI. This demonstrates an intricate relationship between science fiction portrayals of AI and the actual progress in AI development. This emphasizes the need to connect imaginative and empirical viewpoints in order to guide the ethical and efficient integration of AI. However, genuine artificial intelligence necessitates human oversight and predetermined inputs. The Turing test in science fiction evaluates the capacity of artificial intelligence to replicate human behaviors and communication. However, in reality, the effectiveness of AI is evaluated using task-specific metrics in the Turing test. Regardless of the absence of genuine AI consciousness and subjective perception, fictional AI often explores the concept of the mind and the potential future existence of sentient computers. Fictional artificial intelligence (AI) characters like Data from Star Trek frequently explore themes related to the rights of AI, ethical considerations surrounding AI, and the concept of individualism. The argument around the recognition of AI rights and persons in the field of technology is currently complex and lacks established conventions and terminology. The majority of fictional AI often portrays the subject of AI rebellion as a significant plot element, but operational AI is intentionally developed with programmed inputs and regulated conditions to prevent such scenarios. The possibilities of human-AI interaction through brain implants are often examined in science fiction literature and films. However, true artificial intelligence serves as a tool that assists in enhancing the quality of human existence by stimulating human creativity and productivity. Fictional artificial intelligence (AI) often exists in a context where AI research has surpassed our current capabilities in the field of AI. Wiese, W (2024) reports that in the field of philosophy of mind and artificial intelligence research, the debate of whether a computer consciousness can precisely recreate consciousness is a subject of considerable controversy in cognitive robotics. There is a neuroscientific debate about how the computer may attain artificial consciousness by running an accurate algorithm. But the true consciousness is fundamentally different from the simulated consciousness. So technically, there is a significant difference between systems that simply imitate consciousness and those that are genuine copy it. This argument frequently centers on the concept of causal flow, which pertains to the manner in which information is processed and transmitted within a system. From this viewpoint, systems that imitate consciousness demonstrate a specific type of cause-and-effect pattern that is absent in ordinary simulations. However, genuine artificial intelligence is constantly developing and continues to remain within the realm of possibility. Science fiction artificial intelligence often stimulates the philosophical discourse on its intelligence and existence. However, true artificial intelligence focuses on the technical automation within society, and its advancement is driven by study in engineering and data science. The majority of science fiction stories provoke contemplation on the dangers of uncontrolled technological advancement, and they have caused confusion between science fictional artificial intelligence and genuine artificial intelligence due to their widespread coverage in media studies. The distinction between these two realms is crucial in comprehending the practicality and feasibility of artificial intelligence in depth, as the field continues to evolve.

7 Achieving superintelligence?—human consciousness vs artificial consciousness

A curious part of artificial intelligence (AI) research involves investigating how knowledge derived from neuroscience might contribute to the advancement of AI algorithms and systems. When comparing human consciousness to artificial consciousness, we may observe both parallels and contrasts. (Rayhan et al., 2023) investigates the path toward achieving artificial general intelligence (AGI), which refers to AI systems with human-level cognitive ability across several areas and discusses the complete path for achieving AGI, including important developmental steps and difficulties. On contextualizing AGI within the larger landscape of artificial intelligence research, emphasizing the contrast between narrow AI systems that specialize in specific assignments and AGI, which seeks to imitate human intellect, disparity and flexible thinking. They emphasize AGI's transformational potential in revolutionizing different industries, ranging from hospital to planetary missions, as well as its effects on society across the board. Bringing off principles from cognitive science, computational engineering, and machine learning, the AGI advancements in machine learning techniques, cognitive architectures, natural language processing, and automation. AGI includes cumulative breakthroughs in AI technology and its infrastructure by improving AI skills in thinking, logic, memory, and interacting with the world are key milestones along this path. The ethical concerns, legal frameworks, and social participation might help guide responsible AGI design and execution. By combining academic findings with pragmatic issues, an in-depth examination of AGI research and its implications for humanity's existence is needed. (Hawkins and Blakeslee 2005) introduced the hierarchical temporal memory (HTM) model. The HTM model is derived from the architecture and operation of the neocortex, the cerebral cortex's outer layer that governs advanced cognitive processes. They propose that the neocortex functions according to a core set of principles, such as superior structure and chronological memory. The

HTM model is fundamentally composed of a hierarchical structure of processing units known as nodes, which imitate the structure of neurons in the cerebral cortex. The nodes are interconnected sequentially, with higher levels reflecting more abstract concepts or attributes. An essential feature of the HTM model is its focus on chronological memory, which pertains to the capacity to acquire and identify patterns over a period of time. Temporal memory, under the realm of artificial intelligence, empowers systems to comprehend data sequences and generate predictions by leveraging previous learning experiences. Time series data and event sequences are two examples of streaming data types that HTM algorithms are intended to identify patterns in. They employ feedback loops and algorithmic coding processes to deduce the fundamental structure and connections within the data. The HTM paradigm, which draws influence from neurology, provides numerous benefits in the field of artificial intelligence (AI). And it offers a scientifically sound framework for comprehending intelligence, based on the biological structure and functioning of the brain. Aligning AI systems with biological principles improves the comprehensibility and clarity of their interpretations and explanations. More importantly, the HTM model demonstrates resilience and flexibility in managing intricate and ever-changing data streams. The capacity to acquire temporal patterns makes it highly suitable for tasks, such as sequence learning, identifying anomalies, and forecasting. In a nutshell, the hierarchical temporal memory model is a convincing method for developing artificial intelligence that utilizes knowledge from neuroscience. HTM algorithms show promise in enhancing the capabilities of artificial intelligence systems by imitating the principles of brain organization and temporal processing. Furthermore, the attempt to attain superhuman artificial intelligence (AI) within the subsequent century is a topic that fascinates both AI researchers and neuroscientists. An essential element of this quest is to understand the computing capacity of the human brain. The human brain is a remarkably complicated organ with an incredible capacity to handle extensive quantities of knowledge and carry out complex cognitive functions. Creating the brain's computational capacity is difficult due to the inherent difficulties in examining the highly intricate structure and function of the brain. Computational models are useful tools in this effort, providing insights into the brain's capacities. These models aim to replicate the functioning of neurons and neural networks in order to comprehend the mechanisms of information processing and storage systems in the brain. Comprehending the unique characteristics of the human brain poses significant difficulties because of its exceptional intricacy. (Kandel et al. 2012) presents these difficulties, emphasizing the extensive network of interconnected neurons, neurotransmitters, and neural circuits that form the foundation of brain activity. Despite extensive research spanning several decades, numerous elements of brain function still remain enigmatic. This includes the mechanisms underlying memory formation, the processes involved in thought generation, and the emergence of consciousness. Nevertheless, researchers are currently investigating the feasibility of reverse-engineering the brain's operations in order to develop artificial general intelligence (AGI). (Goertzel & Pennachin, 2007) explores the ambitious objective of replicating the fundamental principles of brain function in artificial systems, with the aim of achieving a comprehensive kind of intelligence. The endeavor to reverse-engineer the brain entails examining its structure and function across various levels, ranging from individual neurons to extensive neural networks. Researchers utilize methodologies from neuroscience, computer science, and cognitive psychology to decipher the enigmas of cognition and consciousness. An effective method for reverse-engineering the brain involves the creation of computational models that replicate its functioning. These models, such as neural networks and deep learning algorithms, aim to emulate the brain's capacity to acquire knowledge from experience, identify patterns, and make judgements. On top of that, the progress made in neuroimaging technologies, such as functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), provide vital information about brain activity and connectivity. These techniques facilitate researchers in observing the brain's functioning and unraveling its intricate patterns of activity. Although there has been significant advancement in comprehending the brain, duplicating its skills in artificial systems still poses a formidable challenge. The brain functions through parallel processing, distributed representation, and plasticity, which are challenging to replicate in conventional computer systems. Also, the ethical implications of AGI research give rise to significant inquiries regarding the possible hazards and outcomes of developing intelligent computers that equal or exceed human intelligence. To guarantee the proper advancement and implementation of AGI, it is imperative to thoroughly contemplate these ethical difficulties. Although comprehending the brain poses considerable difficulties, endeavors to deconstruct its operations offer potential for furthering the science of Artificial General Intelligence (AGI). Researchers want to utilize knowledge from neuroscience and computer modeling to uncover the mysteries of cognition and develop intelligent machines that possess human-level intelligence. (Eliasmith 2013) explores computational models of brain activity and devised the Semantic Pointer Architecture (SPA), a conceptual framework for comprehending cognition and constructing artificial intelligence systems that are inspired by the brain. This research offers essential insight into the computational foundations that drive cognitive functions. Computational models like as SPA enable researchers to approximate the brain's processing capacity by simulating

neural activity and evaluating computational efficiency. These estimates provide valuable information and provide insight regarding the possibility of superhuman AI. Nevertheless, it is crucial to acknowledge the constraints of these models and the complex nature of the brain's practical function. The movie Dune by Denis Villeneuve was inspired by Frank Herbert's novel of the same name and offers an intriguing paradigm for exploring posthuman anthropocentrism by depicting machines as subservient entities and spotlighting the value of human-led evolution and natural resource extraction. The reason behind this anthropocentric supremacy can be traced to a historical backdrop in the Dune novel series in which thinking machines are banned as a result of the aftermath of the Butlerian Jihad, a rebellion against artificial intelligence. This traumatic event led to the emergence of a system of governance in which machines are assigned subservient positions to humans. Machines are just tools that help in achieving human objectives, rather than conscious autonomous entities with their agendas. This event serves as a turning point in the history of the dynamic relationship between humans and artificial intelligence and has a strong backdrop which resulted in the abolition of 'thinking machines' and led to the emergence of a strong dislike for sophisticated technologies like artificial entities. So, Butlerian Jihad can be seen as an inherent fear and anxiety of humans, where sentient machines provide a fundamental risk to human existence. (Hipple 2020), states that the fear toward thinking machines can be interpreted as arepresentation of wider apprehensions over the possible risks of conscious artificial intelligence. Artificial consciousness highlights the dangers that might emerge when humanity loses authority over such innovations. Thus, conscious AI's growth in the actual world can be traced to the experiences from the past portrayed in the Dune universe may provide a core guidance for addressing the ethical and existential obstacles that await us. The extraction of the spice melange is an indispensable natural resource that governs the political, economic, and technical necessities of the Dune universe. The human desire to acquire a monopoly over spice melange extraction can be seen as a metaphor for wider problems of resource exploitation and technological innovations. The portrayal of machines as subordinate entities emphasizes an ethical perspective that places greater importance on human consciousness and agency over the possibilities of machine independence. Thus, Dune movie emphasizes a clear difference between human and machine consciousness and gives importance to human will and consciousness compared to AI life forms. Although both human consciousness and artificial systems entail information processing and decision-making, human consciousness is distinguished by subjective experience, feelings and perceptions of identity, which are now beyond the capabilities of artificial systems. Furthermore, human consciousness is closely linked to physical existence, social connections, and cultural surroundings, which provide considerable obstacles to achieving artificial consciousness in the field of AI. In broader terms, computational models offer useful insights into the computing capabilities of the human brain, which contribute to discussions on the advancement of artificial intelligence beyond human capabilities. However, additional investigation and improvement of these models are critical to completely grasp the brain's computing capacities and unleash the potential for attaining artificial superintelligence.

8 Conclusion

The amalgamation of posthumanism with artificial consciousness in AI creates a thought-provoking and complex setting that challenges traditional concepts of humanity and its ethical boundaries. Humans bear a substantial ethical obligation in society to develop a more intricate and self-aware humanoid. The gradual integration of such technologies into our society necessitates a thorough examination of their potential implications for matters related to individuality, interpersonal interactions, and societal structures. However, it is crucial to acknowledge that artificial intelligence currently does not possess genuine consciousness since AI systems are constructed using algorithms, data, and pre-existing regulations that adhere to ethical principles that do not negatively impact humans so far. Attaining artificial consciousness in AI is a challenging task, and its practical and theoretical implications are still a subject of controversy in the fields of AI, philosophy, ethics, and neuroscience. Consciousness is a multidimensional phenomenon that extends to not just the processing of sensory data and information but also subjective experiences and self-awareness. These aspects cannot currently be reconstructed in computers. However, AI researchers analyze the multiple hypotheses and models associated with consciousness in AI, while these models have the theoretical potential to create a conscious AI, current technology is not sufficiently advanced enough to achieve this practically. On top of it, it is crucial to acknowledge remarkable innovation in AI is still taking place and our perception of consciousness is gradually evolving. Upcoming breakthroughs in neuroscience, cognitive robotics and artificial intelligence may shed new light on the nature of consciousness, potentially paving the way for more sophisticated AI neural networks. However, currently, the attainment of artificial consciousness in AI remains an impractical distant aspiration.

9 Glossary

- *Technological Singularity*: A hypothetical and speculative future event when technological advancement and artificial intelligence exceed human intelligence and awareness.
- *Cognitive Robotics*: A branch of robotics which aims to integrate robots with cognitive competencies which include reasoning, sensing learning and decision-making.
- *Functionalism Theory*: A theory in the philosophy of mind that the state of consciousness is not characterized by their physical foundation, but rather by their function or role.
- *Integrated Informational Theory (IIT)*: A theory that puts forth the idea that consciousness emerges from integrated information found within a system and is measured by the casual structure of that system.
- *Panoptic Surveillance*: A comprehensive surveillance system that enables continuous monitoring of humans from several angles, without being noticed and leading to concerns over privacy and autonomy.
- *Free Energy Principle*: A theoretical paradigm which proposes that in order to preserve their internal states, self-organizing systems such as the brain minimize surprise or prediction error.
- *Anthropotechnics*: A study of methods or technologies which seek to improve human capacities frequently through the integration of artificial intelligence and other advanced technologies.
- *Post Scarcity*: An economic and social state that may be attained utilizing automation and advanced technology, in which resources are plentiful enough to meet human needs.
- *Casual Flow*: A neuroscience concept that describes the seamless and undisturbed state of consciousness that occurs during peak performance and intense immersion in a task.
- *Hierarchical Temporal Memory Model (HTM Model)*: A computational model that draws inspiration from the anatomy and functioning of the human neocortex, emphasizing learning hierarchical and temporal patterns.
- *Functional Magnetic Resonance Imaging (fMRI)*: A neuroimaging system used to examine consciousness-related processes and brain function by monitoring changes in blood flow.
- *Electroencephalography (EEG)*: A harmless form of a neuroimaging technique used to research brain dynamics and consciousness, which captures electrical activity in the brain using electrodes attached to the scalp.
- *Semantic Pointer Architecture (SPA)*: A cognitive architecture useful for discovering neural mechanisms in AI and neuroscience, the semantic pointers are high dimensional vectors that describe neurological representations and computations.

Declarations

Conflict of interest There is no conflict of interest among the authors.

References

Arkin RC (2009) Governing lethal behavior in autonomous robots. Chapman and Hall/CRC

Barnett M, Wagner H, Gatling A, Anderson J, Houle M, Kafka A (2006) The impact of science fiction film on student understanding of science. J Sci Educ Technol 15(2):179–191. https://doi.org/10.1007/s10956-006-9001-y

Cave S, Dihal K (2019) Hopes and fears for intelligent machines in fiction and reality. Nat Mach Intell 1(2):74–78. https://doi.org/10.1038/s42256-019-0020-9

Cummings ML (2014) Man versus machine or man + machine? IEEE Intell Syst 29(5):62–69. https://doi.org/10.1109/MIS.2014.84

Eliasmith C (2013) How to build a brain: a neural architecture for biological cognition. Oxford University Press

Floridi L (2011) The philosophy of information. Oxford University Press

Friston K (2010) The free-energy principle: a unified brain theory? Nat Rev Neurosci 11(2):127–138. https://doi.org/10.1038/nrn2787

Goertzel B, Pennachin C (2007) Artificial general intelligence. Springer, New York

Hassan I (1977) Prometheus as performer: toward a posthumanist culture? The Georgia Review 31(4):830–850

Hawkins J, Blakeslee S (2005) On intelligence, 1st edn. Owl Books

Hayles NK (1999) How we became posthuman: virtual bodies in cybernetics, literature, and informatics. University of Chicago Press. University of Chicago Press, https://press.uchicago.edu/ucp/books/book/chicago/H/bo3769963.html. Accessed 22 Mar 2024

Hein AM, Baxter S (2018, January 1). Artificial intelligence for interstellar travel. Cornell University. https://doi.org/10.48550/arxiv.1811.06526

Hipple D (2020) Encounters with emergent deities: artificial intelligence in science fiction narrative. J Religion Sci 55(2):382–408. https://doi.org/10.1111/zygo.12602

Kandel ER, Schwartz JH, Jessell TM (2012) Principles of neural science. McGraw-hill, New York

Lin P, Bekey G, Abney K (2009) Robots in war: issues of risk and ethics. Ethics and Robotics. pp 49–67. https://digitalcommons.calpoly.edu/phil_fac/5. Accessed 25 Mar 2024

Lombardo W (2021) Losing ourselves. The new atlantis.15 April 2022, van. https://www.thenewatlantis.com/publications/losing-ourselves

McCauley L (2007) Ai armageddon and the three laws of robotics. Ethics Inf Technol 9(2):153–164. https://doi.org/10.1007/s10676-007-9138-2

Moravec H (1995) Mind children: the future of robot and human intelligence (4 .print). Harvard Univ Press

Ohta H (2015) Reevaluation of mcculloch–pitts–von neumann's clock. Biosystems 127:7–13. https://doi.org/10.1016/j.biosystems.2014.10.003

Piccinini G (2021) The myth of mind uploading. In: Clowes RW, Gärtner K, Hipólito I (eds) The mind-technology problem. Studies in brain and mind, vol 18. Springer, Cham
Potapov A (2018) Technological singularity: What do we really know? Information 9(4):82. https://doi.org/10.3390/info9040082
Rayhan A, Rayhan R, Swajan Rayhan (2023) Artificial general intelligence: Roadmap to achieving human-level capabilities. https://doi.org/10.13140/RG.2.2.33680.79361/1
Russell SJ, Norvig P (2022) Artificial intelligence: a modern approach. Pearson
Sahu OP, Karmakar M (2022) Disposable culture, posthuman affect, and artificial human in Kazuo Ishiguro's Klara and the sun (2021). AI & Soc. https://doi.org/10.1007/s00146-022-01600-1
Sharkey N (2010) Saying 'No!' to lethal autonomous targeting. J Military Ethics 9(4):369–383. https://doi.org/10.1080/15027570.2010.536403
Sharkey A, Sharkey N (2012) Granny and the robots: ethical issues in robot care for the elderly. Ethics Inf Technol 14(1):27–40
Sloterdijk P, Hoban W (2013) You must change your life: On anthropotechnics, English. Polity
Sorgner SL (2007) Metaphysics without truth: on the importance of consistency within nietzsche's philosophy. Marquette University Press
Tegmark M (2018) Life 3.0: being human in the age of Artificial Intelligence. Penguin Books
Tononi G (2012) Phi: a voyage from the brain to the soul, 1st edn. Pantheon
Turkle S (2021) The empathy diaries: a memoir. Penguin Books
Vint Cerf: "Privacy may be an anomaly." (n.d.). CNET. https://www.cnet.com/news/privacy/vint-cerf-privacy-may-be-an-anomaly/. Retrieved May 8 2024
Walsh T (2017) The singularity may never be near. AI Mag 38(3):58–62. https://doi.org/10.1609/aimag.v38i3.2702
Walsh T (2022) Will AI end privacy? how do we avoid an Orwellian future. AI & Soc 38(3):1239–1240. https://doi.org/10.1007/s00146-022-01433-y
Wiener N (1961) Cybernetics: or the control and communication in the animal and the Machine. MIT Press
Wiese W (2024) Artificial consciousness: a perspective from the free energy principle. Philos Stud 181(8):1947–1970. https://doi.org/10.1007/s11098-024-02182-y