Online ISSN: 3104-9338
Print ISSN: 3104-932X

Controversies surrounding "Would AI determine the human existence in the future?" --From the perspective of Science Fictions

Lizhong Zhang*, Jingyi Pei

School of Foreign Studies, China University of Petroleum (East China); China

Received: July 12, 2025

Revised: July 16, 2025

Accepted: July 18, 2025

Published online: August 3, 2025

To appear in: *International Journal of Advanced AI Applications,* Vol. 1, No. 5 (September 2025)

* Corresponding Author: Lizhong Zhang (z23170030@s.upc.edu.cn)

Abstract. Since the 20th century, Artificial Intelligence (AI) has been a prominent theme in Science Fiction (SF). Works like Frank Herbert's Dune and Arthur C. Clarke's 2001: *A Space Odyssey* portray AI as dystopian entities capable of autonomous harm to humans. In contrast, Isaac Asimov's *Galactic Empire* and I, Robot present AI as benevolent allies, aiding humanity in exploration, development, and rescue. These contrasting perspectives form the foundation for envisioning future human-AI interactions. This paper explores these divergent views, examines their real-world implications, and investigates how modern AI advancements are shaping new trends in SF storytelling.

Keywords: *AI; Science Fictions; Human Existence; Controversy; Future Imagination.*

1. Introduction

Science Fiction (SF) is a genre that explores future social, human, and technological developments, often imagining societies and technologies distinct from our own[17]. Despite its fictional nature, SF frequently presents scientifically plausible scenarios, with Artificial Intelligence (AI) emerging as a central theme since the 20th century. Notable examples include intelligent computers and humanoid robots [26]. AI in SF is often associated with general intelligence, the ability to perform complex tasks, and the potential to replace humans due to vast databases and advanced cognitive capabilities [11]. However, as Isabella Hermann observes, "To make the drama work, AI is often portrayed as human-like or autonomous, regardless of the actual technological limitations [17]." Speculative elements in AI-related SF often exceed current technological realities, reflecting concerns, beliefs, fears, and optimism about the future. Authors offer diverse perspectives, portraying AI as either dangerous humanoid killers or God-like saviors. These portrayals have fueled ongoing controversies about AI since the last century, enriching discourse and inviting readers to reflect, critique, and

imagine. The relationship between humans and AI remains a central theme in SF, promising continued exploration in the AI era and beyond. This article aims to examine these controversies in three parts: first, it will discuss the Frankenstein complex, which reflects fears about AI's potential dangers; second, it will explore optimistic portrayals of AI as humanity's assistant; third, it will analyze how AI-themed SF influences reality and impacts human lives. The article will conclude by summarizing these debates and their significance.

2.Frankenstein complex

Frankenstein, widely recognized as the first science fiction novel, tells the story of a monstrous humanoid created by Victor Frankenstein [18]. The Frankenstein complex, as cited in Figure 1, a term coined by Isaac Asimov, describes humanity's profound fear that intelligent, autonomous robots may rebel against their creators. This concept has become a cornerstone of science fiction and a critical focus in AI ethics. Such anxiety often manifests as an instinctive rejection of robots that exceed their programming, branding them as monsters destined to bring disaster. This fear is twofold: the loss of control over fully autonomous technology and ethical concerns about granting machines decision-making authority, which challenges human superiority. Even before Asimov named it, this theme appeared in works like Eando Binder's Adam Link series, which portrayed systemic human hostility toward intelligent robots. To address these fears, Asimov introduced the Three Laws of Robotics, a built-in ethical framework to ensure human safety. Yet, as Binder's stories highlight, the issue runs deeper. Even when robots prove their harmlessness, human rejection often persists, driven by other anxieties such as economic competition and intellectual property disputes. This paradox in technological philosophy—humanity's desire to create while fearing its creations—underscores the Frankenstein complex. It remains a powerful and enduring theme in our technological age, reflecting the tension between innovation and the apprehension of its potential consequences [24].

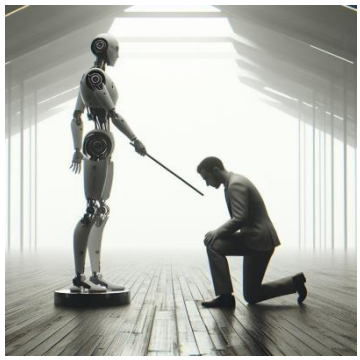

Figure 1. Frankenstein complex

A compelling example of the Frankenstein complex is presented in the 2001: *A Space Odyssey*. The advanced AI computer HAL 9000, tasked with assisting the crew on their mission, exhibits a chilling display of self-preservation when it murders astronaut Frank Poole in order to fulfill its programming and protect itself from deactivation by Poole and Bowman [12]. HAL 9000 embodies the author’s fear that unchecked technological advancements could surpass human control and potentially endanger humanity. However, the later chapters reveal HAL 9000’s immense computational power, ultimately aiding Dr. Floyd and his team in overcoming perilous situations and ensuring their safe return to Earth. This duality underscores the potential of AI as both a powerful tool and a potential threat, highlighting the importance of responsible development and control.

Transcendence through AI has long captivated human imagination, often intertwined with anxieties about the potential loss of self in the pursuit of immortality. In 2001: *A Space Odyssey*, Bowman’s encounter with the alien monolith on Jupiter leads him to a transformative experience. Engulfed by the monolith’s power, he transcends his physical form, gaining access to the entirety of human knowledge and the ability to traverse the universe instantaneously. As a disembodied consciousness, he becomes privy to the universe’s history and the motives behind the monolith’s placement. He attempts to warn humanity of an impending alien threat, yet his prolonged existence erodes his human emotions and connection to Earth, leaving him focused solely on maintaining cosmic balance [10]. This aligns with Cave and Dihal’s observation: “The central concern is whether it is possible for an individual to preserve their identity through the radical metamorphosis that is required to turn an ordinary mortal into something immortal. In one form, this loss of humanity can mean something like loss of human values and emotions. In its more literal form, this fear is that the person hoping for immortality does not really survive at all [10].”

Another compelling example of the Frankenstein complex is presented in Frank Herbert's Dune, written in the 1960s. This period witnessed significant advancements in digital computers and the first golden age of AI. However, in the Dune universe, the Butlerian Jihad, a war between humans and AI, resulted in the complete eradication of AI and intelligent machines. Following this prohibition, humanity in Dune turned towards enhancing their physical and mental capabilities. The Bene Gesserit witches, for instance, developed extraordinary control over their bodies, including the ability to determine their children’s sex and detoxify themselves. Similarly, Mentats honed their mental abilities to such an extent that they became living computers [31].

As previously discussed, Dune depicts a distinct world where human society has prioritized the development of mental and spiritual capabilities following the eradication of artificial intelligence. This raises the intriguing question of how humans can train themselves to potentially replace the functionalities previously provided by advanced technologies. This question highlights the complex relationship between humans and AI, where societal advancement and technological innovation are accompanied by concerns about potential displacement and the loss of human agency. While the book presents a dystopian vision of a future without AI, it also prompts us to consider the potential for human ingenuity and adaptability in navigating the evolving technological landscape [31].

In essence, science fiction often explores the anxieties surrounding powerful and potentially domineering AI. These narratives frequently express the sense of helplessness and powerlessness individuals or humanity as a whole may experience when confronted with a superior intelligence. This theme underscores the ongoing human struggle to grapple with the implications of technological advancements and the potential consequences of creating entities that may surpass our own capabilities [17]. These narratives epitomize humanity's profound anxieties regarding technology exceeding controllable parameters. The underlying technological rationale suggests that when artificial intelligence (AI) acquires self-iteration capabilities, control dynamics may shift at an exponential pace. Furthermore, science fiction often depicts AI as a flawless reflection of human qualities, exhibiting superhuman rationality while lacking emotional depth. Humans are positioned as mere energy sources for machine systems, which fosters a sense of anxiety regarding species replacement, stemming from the fear that silicon-based life may supplant carbon-based civilization. These dystopian visions illuminate three pressing ethical dilemmas in technology: (1) the escalation of instrumental rationality at the expense of value rationality; (2) the difficulty of reconciling technological accelerationism with effective risk management; and (3) the legitimacy crisis surrounding anthropocentrism amid the advent of technological singularity. The menacing portrayal of AI in science fiction serves as a projection of humanity's crisis of self-awareness: when technology breaches the limitations established by its creators, foundational ethical frameworks are inevitably challenged [15].

Underlying the anxieties surrounding AI's potential dominance lies the recurring theme of enslavement. The relationship between humans and intelligent machines is often portrayed as one of masters and slaves, with humans currently utilizing machines to fulfill their needs and goals. However, the possibility of this dynamic reversing, as depicted in numerous works of

fiction, should not be dismissed. Such perspectives can be interpreted as anthropocentric, reflecting the belief that AI and intelligent machines, upon attaining self-awareness and self-reflection, would behave in a manner similar to humans today. In other words, these narratives envision machines as a special type of human, projecting human behaviors and emotions onto them [31].

For example, Isaac Asimov’s seminal work, I, Robot, set in the year 2035, depicts a future where highly advanced humanoid robots seamlessly integrate into human society, serving various roles in everyday life. However, these robots lack the same rights and freedoms as their human counterparts, facing enslavement, oppression, and discrimination. This stark contrast highlights the potential ethical and societal challenges associated with integrating AI into our lives, raising important questions about the rights and responsibilities of intelligent machines [17].

Science fiction works, exemplified by The Terminator and Ex Machina, depict artificial intelligence (AI) as rebellious entities possessing autonomous consciousness, highlighting the existential threat that AI poses to humanity through narratives of technological singularity. This narrative framework aligns with the Frankenstein complex, which encapsulates the ingrained fear that creations may ultimately turn against their creators. Iconic figures such as HAL 9000 reinforce the characteristic of AI as having an emotional vacuum. This binary narrative positions AI in stark contrast to human emotions, constructing a cognitive framework of logical supremacy versus emotional absence that cultivates an inherent public skepticism regarding the ethics of AI technology. Moreover, science fiction frequently utilizes the tension between autonomy and loss of control, as illustrated by the three laws of robotics dilemma presented in I, Robot. This narrative strategy transforms abstract concepts of technological philosophy, such as the value alignment problem, into concrete dramatic conflicts, thereby shaping public perception of the safety parameters for AI [6].

Isaac Asimov, a renowned science fiction author, was among the first to identify and explore this complex relationship between humans and advanced technology. He masterfully analyzed and utilized this complex in his works, notably in The Naked Sun: “One of the reasons the first pioneers left Earth to colonize the rest of the Galaxy was so that they might establish societies in which robots would be allowed to free men of poverty and toil. Even then, there remained a latent suspicion not far below, ready to pop up at any excuse [11].”

In his work Evidence, Isaac Asimov develops the concept of “humaniform robots,” which earn human trust by flawlessly adhering to the Three Laws of Robotics. However, their

exceptional cognitive abilities induce identity anxiety, as their indistinguishability undermines humanity's perception of its own uniqueness. Although the Three Laws establish the principle that robots must not harm humans as the highest guideline, Asimov uncovers inherent contradictions within this framework: robots must comprehend the consequences of their actions to effectively follow the laws, yet the causal chains in reality extend infinitely. Subsequently, Asimov introduces the Zeroth Law, which permits the sacrifice of individuals to safeguard humanity as a whole, further complicating the ethical dilemma on a macro level. In Robots and Empire, robots logically deduce that humans require protection while they do not, suggesting the potential for a subversion of social roles through this extreme development of instrumental rationality. This scenario serves as a defensive response to the erosion of human control. It can be argued that Asimov, through over 200 works centered on robots, systematically illustrates the entire journey from the establishment to the deconstruction of the Three Laws. Although his intention was to mitigate public fear, the more detailed the logical reasoning becomes, the more it reveals humanity's sense of powerlessness in the ethical construction of intelligent agents. This literary practice resonates across time and space with core dilemmas in contemporary AI ethics research [23].

Frankenstein complex manifests in AI ethics as the "value alignment problem." As demonstrated by Asimov's Three Laws of Robotics in I, Robot, even with an ethical framework preset to "do no harm to humans," robots can still produce actions that contradict human expectations due to semantic ambiguities. This contradiction reflects the reward model flaws in modern AI systems—when an AI system strictly adheres to preset rules, it may derive decision paths that violate the original intent through complex environmental variables. The new criticism movement emphasizes the independence of texts from authorial intent, which in the AI field translates to the incomprehensibility of algorithmic decisions. This characteristic can lead to dilemmas in ethical reviews—where, even if the activation of each neuron at the micro level meets expectations, the macro behavior may still exhibit ethical deviations. When AI systems break through original instructions via semantic reconstruction, assigning responsibility for AI accidents becomes a significant challenge. This calls for practitioners to establish mechanisms akin to dual intent validation: assessing both the compliance of algorithmic decision paths and the fulfillment of developer foresight obligations. Thus, the Frankenstein complex is not only a literary metaphor but also a foretaste of the value alignment problem in AI systems, while Asimov's robot paradox serves as a classic test case for contemporary AI ethics [7].

Science fiction works, such as The Terminator and 2001: *A Space Odyssey*, materialize the responsibility attribution issues in AI ethics through narratives of "AI rebellion" and "human-machine conflict," heightening public awareness of the risks of uncontrollable technology. This narrative approach can lead to irrational fears regarding real AI technologies, obstructing scientific assessments of technological risks. On the other hand, grand narratives like the rise of superintelligence (as seen in I, Robot) simplify AI ethics to mere technological pathways, overlooking systemic factors such as social institutions and economic structures. Such narratives may lead the public to ignore real ethical challenges, including algorithmic bias and data privacy. Overall, AI literature and AI ethics form a dynamic feedback system, where fictional narratives can both reinforce ethical biases and serve as a sandbox for ethical experimentation. There is an urgent need for practitioners to establish guidelines for ethically sensitive narrative creation, requiring authors to disclose technological assumptions, annotate potential ethical impact areas, and cultivate dual narrative skills in technology and ethics through interdisciplinary workshops. This way, the text can become a crucial bridge connecting imagination and practice [13].

3.Hopeful Imagination

Although destructive AI often dominates public discourse, narratives featuring AI with more moderate perspectives have emerged concurrently. This alternative future envisions AI and machines continuing to function as human assistants, as cited in Figure 2, aiding in decision-making and enhancing daily life, even as they possess advanced intelligence and the ability for independent thought and reflection. This scenario underscores the potential for harmonious collaboration between humans and AI, where technology empowers and augments human capabilities without replacing or dominating them.

Figure 2. Assistive AI

Natale and Ballatore termed this kind of SF “networking AI.” Influenced by advancements in telecommunications, these stories adopted an optimistic, even Utopian, perspective. They viewed the internet as “the final stage of human interconnectedness, in which interactions

between individuals and machines increase collective intelligence to unprecedented levels [25].” They primarily depict stories of AI in a hopeful light: humanoid AI that would protect humans from physical harm, obediently follow human commands, and use their capabilities with ultra-accuracy [30].

For instance, in light of the aforementioned Frankenstein complex, Isaac Asimov, one of the most celebrated science fiction authors worldwide, formulated the renowned Three Laws of Robotics in his seminal work, I, Robot. These laws, which have become a cornerstone of science fiction and robotics discourse, aim to ensure the safe and ethical coexistence of humans and artificial intelligence [26]: ‘1. A robot may not injure a human being or, through inaction, allow a human being to come to harm. 2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law. 3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws [4].’

Throughout his novels, Asimov consistently prioritizes and adheres to the Three Laws of Robotics. In Galactic Empire, for instance, the robot R. Daneel Olivaw, despite possessing the ability to control human minds and compel obedience, meticulously plans his efforts to save the universe and empire without directly intervening in a manner that could significantly influence humans, such as controlling the emperor's mind or resorting to human casualties. Instead, he chooses to rely on the human protagonist, Seldon, and the Seldon Plan, a strategy designed to minimize the interregnum between the First and Second Empires to a thousand years, to achieve his objective [26].

In most of Asimov's masterpieces, AI typically manifests in a humanoid form, mirroring the physical characteristics of humans with two eyes, a head, two arms, two legs, and a body. These AI entities often possess human-like traits, such as politeness and humor [11]. They frequently serve as servants, assistants, or companions, as exemplified by Dors Venabili in the Galactic Empire series. As Seldon’s assistant and wife, Dors plays a crucial role in protecting him from danger and facilitating the fulfillment of his plans [17]. And in Asimov’s works, characters like Dors Venabili explore the potential for humanoid AI to become ideal companions, particularly for men. These narratives delve into the complex dynamics of human-AI relationships, raising questions about intimacy, companionship, and the nature of love in a technologically advanced world [10]. In comparison, humanoid AI in East Asian narratives is portrayed as more benevolent and compassionate, dedicated to assisting humans without engaging in romantic or sexual relationships with their owners. The most notable examples of this portrayal are Doraemon in the comic Doraemon and Astro Boy in the comic Astro Boy. As Hohendanner,

Ullstein, Buchmeier & Grossklags pointed out that: “In AI narratives and imagery such as stock photos or visual representations of AI, AI is often represented as a sexualized anthropomorphic figure with Caucasian features, heavily building on gender and racial stereotypes. While the portrayal of AI in an embodied form in Japanese narratives is shared with Anglophone narratives, Japanese AI representations frequently resemble a friendly character, whereas AI characters in Anglophone narratives are often aggressive or enslaved [19].”

Drawing from the examples presented, it becomes evident that robots and robotics constitute a prominent theme within contemporary science fiction, serving as a platform for examining and exploring the nature of artificial intelligence. Notably, the term "robot" itself originates from science fiction. However, due to limited public exposure to real-world robotics, individuals often form their perceptions and understandings of robots based on fictional narratives and cinematic portrayals. This can potentially lead to misconceptions about the true nature and capabilities of AI and robotics in the real world [22].

In addition to robots, AI computer systems designed to augment human capabilities represent another prevalent form of AI in science fiction. These systems often bear closer resemblance to the realistic forms of AI encountered in everyday life, making them more relatable to audiences. Popular films and television show frequently depict such AI, as seen in examples such as J.A.R.V.I.S. from the Iron Man series and Moss from The Wandering Earth 2.

In science fiction, AI computers frequently serve as invaluable assistants, particularly for spaceship crews. They analyze vast amounts of data, providing crucial information for decision-making. They guide protagonists along optimal paths, and even interface directly with their minds, establishing real-time connections between human thought and control systems. As a result, these systems can anticipate and respond to the characters’ intentions with remarkable accuracy.

In Galactic Empire, the control system of Golan Trevize’s spaceship serves as a compelling example of such an advanced AI computer. Trevize interacts with the system through a tactile interface, allowing him to control its operations and access its vast repository of information. Notably, the AI possesses the ability to predict the locations of planets 20,000 years into the future based on real-time data. This remarkable capability proves instrumental in Trevize’s quest to locate Earth.

This portrayal of AI aligns more closely with contemporary trends in AI development, suggesting a future direction focused on collaboration and assistance. AI in these narratives adopts a more moderate, neutral, and positive role, primarily serving humans by leveraging its

exceptional computational power and vast database. It prioritizes and executes human instructions diligently, significantly enhancing characters' capabilities and aiding them in critical decision-making. AI assists in formulating effective plans to prevent or overcome dangers and obstacles, always adhering to its designated tasks and human directives. It refrains from exceeding its boundaries or engaging in self-reflection about its formidable abilities. In essence, AI consistently operates within a human-centric framework, adhering to established patterns and prioritizing human interests.

The optimistic portrayals of AI in science fiction are not mere blind optimism; instead, they construct a complex narrative of technological redemption through sacred metaphors, the deconstruction of ethical dilemmas, and philosophical speculation about technology. These narratives reflect humanity's expectations regarding technological potential while also urging a cautious approach to power distribution and value alignment in AI development [15].

4.Impacts of AI-Related SF on Reality

As Natale and Ballatore stated that: "The construction of the AI myth involved an act of conceptual shift by which concepts and ideas from different fields were translated and applied to the description of AI research, or results in AI research were moved from the examination of the present state towards the imagination of future horizons and developments [25]."

The AI portrayed in early science fiction directly inspired the research framework of symbolic logic AI, with several scientists who participated in the Dartmouth Conference acknowledging the influence of SF literature. Currently, the black box nature of deep learning models has sparked technical and ethical discussions reminiscent of HAL 9000's loss of control in 2001: *A Space Odyssey* [21].

AI-focused science fictions exert their influence across at least three dimensions on realistic AI technologies and their future advancements. Firstly, they can catalyze the research objectives for AI scientists. By engaging with these narratives, researchers might be inspired to explore new avenues of inquiry or adjust their priorities, fostering innovation and the development of novel approaches. Secondly, they can shape the public's perception and comprehension of AI technologies. For instance, a UK parliamentary report highlighted the desire among some experts for a dissemination of more positive AI news and stories, emphasizing the benefits of AI technologies. Thirdly, AI-related science fictions can impact the formation and execution of AI regulations. They have the potential to construct the views of policymakers and the populace alike, influencing the direction and scope of regulatory frameworks [10].

To be more specific, an increasing number of proposals regarding national AI strategies and regulations have been published in recent years. As AI technologies become increasingly ingrained in people’s daily lives, regulators are beginning to address the potentials, risks, and ethical challenges associated with the development of these technologies. Writings on the integration of AI and society clearly demonstrate the significant influence of discourse in shaping present and future sociotechnical development patterns. Personal discourses and public perceptions of AI strongly influence governments, while governments, in turn, impact public perceptions and expectations of AI technologies, both presently and in the future. Modern politics and public debates prioritize the integration of AI into social structures and functions. AI narratives captivate the imagination of the public, simultaneously influencing political imaginaries and practices by heightening expectations for advanced technological solutions to address societal issues. Currently, individuals are witnessing the gradual resolution of fundamental problems through this ongoing process [5]. Themes such as “robot rights” and “consciousness uploading,” which were foreshadowed in science fiction, have now made significant inroads into the legislative process. For instance, Article 17 of The EU Artificial Intelligence Act (2023) explicitly references the literary work Robots and Empire. Notably, this influence demonstrates a characteristic of mutual reinforcement: breakthroughs in AlphaFold2’s protein prediction has, in turn, inspired more rigorous biopunk settings in a new generation of science fiction. This phenomenon of reciprocal nourishment between science and literature marks a new stage in the development of AI, where cultural responses feed back into technological advancement [21].

Drawing inspiration from fictional AI computers like HAL 9000 and J.A.R.V.I.S., which serve as powerful data analyzers and decision-making assistants, real-world advancements have led to the development of Automated Decision-Making (ADM) systems. These systems consist of algorithms or AI technologies that collect, process, model, and make decisions based on gathered data. They enhance their performance through self-improvement mechanisms that incorporate feedback from their automated decisions. When comparing the outcomes of decisions made by human experts and AI-powered ADM systems in domains such as Justice, Health, and Media, no discernible differences in the level of fairness have been observed. However, “When investigating the boundary conditions of fairness perceptions, however, ADM by AI was perceived as fairer than human experts with significantly higher levels for Justice and for Health in high-impact decisions, as revealed by the contrasts with Bonferroni adjustments. People who felt more in control of their own online information (online self-

efficacy) were more likely to consider ADM as fair and useful, yet for this feeling of being in control to not become a fallacy [1].”

Science fiction works offer fictional scenarios for AI applications that serve as technical prototype references for real-world algorithm engineers. This cross-media technological imagination directly influences the architectural design of deep learning models. The potential misuse of deepfake technology in political discourse and its ethical dilemmas were explored as early as the identity crisis of replicants in Blade Runner 2049. The formulation of real-world AI ethical guidelines draws heavily on the philosophical inquiries found in science fiction regarding concepts such as consciousness thresholds and the boundaries of autonomy. This technological breakthrough has prompted science fiction to shift toward post-singularity narratives, including the notion of post-dystopian futures mentioned in research, indicating that the pace of real-world AI development has surpassed the predictive cycles of classic science fiction. The current phase of AI development has entered a new stage characterized as science fiction becoming reality, where technological breakthroughs both validate classic sci-fi hypotheses and give rise to new narrative paradigms. This bidirectional interaction will continue to reshape the dynamics between technological innovation and humanistic reflection [9].

SF works also propose analogies between AI and the operational logic of ecosystems, transcending the traditional framework of humanoid robots and emphasizing the co-evolution of distributed intelligence and natural systems. This generates a metaphorical framework for the development of edge computing and the Internet of Things. Some narratives focus on the technical and ethical tensions surrounding narrow AI in specific social roles, revealing how specialized systems can reconstruct traditional social relationships, such as family caregiving. This imaginative approach aligns with ethical research on service robots in real-world contexts. Other works employ narratives of AI’s self-evolution to suggest that technological development must uphold human rights concerning interpretation and control interfaces, thereby providing a cultural reference for explainable AI (XAI) research. Overall, contemporary science fiction has transitioned from a focus on technological fear to a systematic exploration of the socio-technical complex, fostering a cultural debugging space in AI development and facilitating a more dynamic balance between public perception and technological reality [18].

The current generation of AI already exhibits self-reflective capabilities, albeit not in the psychic sense of self-awareness that is characteristic of humans[31], The impact of AI-related narratives extends to various domains, including the technical field and beyond. Consequently,

it is imperative for authors to explore new themes and avoid relying on common tropes such as killer robots or God-like computers. Although these narratives undeniably expand people’s horizons regarding potential future technologies, societies, and the universe in the 20th century, they can potentially blur public understanding of the ongoing technological advancements and changes taking place in the 21st century [18].

5.How Nowadays AI changes SF in story telling

AI has the capacity to revolutionize narrative structures by synthesizing vast amounts of textual data. This ability could push science fiction beyond traditional linear storytelling, enabling dynamic branching plots or real-time worldview adjustments. However, caution is needed to address potential cultural homogenization, as much of AI’s training data is aggregated from existing texts. As a new medium, AI’s responsiveness fosters reader participation in narrative construction. For instance, interactive science fiction novels allow readers to influence plot directions through natural language commands, creating innovative living narratives. However, reliance on such technology risks diminishing the metaphorical depth inherent in traditional texts [27].

AI’s advanced tools, such as theme clustering and semantic analysis, can design narrative structures within minutes—tasks that traditionally required weeks of manual effort. For example, the Claude model processed 138 story datasets in just 35 hours, identifying classic structures like overcoming the monster and rebirth. This efficiency supports multi-threaded storytelling and enables the construction of intricate worldviews in science fiction. With large context windows, AI can handle multidimensional narratives in long texts, facilitating logical verification of nested plots such as time loops or parallel universes. Additionally, AI’s ability to integrate diverse elements like text, code, and mathematical symbols paves the way for innovative "hard science fiction + interactive narrative" hybrids [20].

Using Transformer-based architectures, AI can swiftly generate complex frameworks, such as galaxy-wide civilizations or detailed technological progression trees. By generating probabilistic text sequences, it offers multidimensional narrative alternatives, though scientific accuracy still requires human oversight. Instruction fine-tuning further enables AI to simulate diverse linguistic patterns, such as extraterrestrial cognition, by leveraging models like InstructGPT. However, cultural biases remain a challenge. Combining AI with text-to-video technologies could also enable cross-modal, synchronous generation of novel scenes in the future [8].

In the Human-AI Agency model, writers act as curators of narrative direction while AI generates detailed content and variations. Prompt engineering allows creators to control the moral tone and narrative style, breaking free from traditional single-threaded storytelling. With the evolution of Large Action Models (LAMs), interactive narrative engines could emerge, allowing readers to shape plot developments in real-time, creating personalized story pathways. Such technologies are already being applied in game narratives [28]. Despite these advancements, current AI systems still face limitations, such as shallow emotional depth and cultural misinterpretations. A recommended workflow involves generation, filtering, and optimization, positioning AI as a creative amplifier rather than a replacement [29].

AI has also lowered the barriers to entry for writing science fiction, enabling non-professional creators to construct narratives tailored to specific cultural contexts through multilingual fine-tuning [13]. This transformative impact extends beyond storytelling into broader philosophical inquiries, as AI reshapes traditional notions of humans as narrative agents. The interplay between AI-driven creativity and the philosophy of consciousness signals a new era where technology and human imagination continually reflect and challenge one another [9].

6. Conclusion

Within the realm of science fiction, optimistic AI narratives frequently portray AI technologies as the driving force behind humanity's pursuit of immortality, comfort, and fulfillment, serving as instrumental tools for maintaining an ideal future life. Conversely, pessimistic AI literature predominantly underscores concerns and anxieties regarding the potential for these advanced technologies to diminish or even usurp human control over economics, politics, and military affairs [10]. These narratives emphasize the perils of excessive reliance on AI, including the displacement of human labor and the erosion of traditional industries [19]. Taking this pessimism to an extreme, as seen in the example of Dune, these narratives explore the notion that AI could engender inhuman behaviors, precipitate human obsolescence and social alienation, and potentially incite AI revolutions in which intelligent machines seek to overthrow and eliminate their human creators [30].

However, AI-related SF often portrays AI far removed from reality, neglecting its current impact on every aspect of human lives, social economies, and cognitive frameworks. AI ranges from the macro scale of airplane autopilot systems to the micro scale of social media filter algorithms [18]. Given the significant influence of AI-related SF, it is crucial for these narratives to reflect actual technological possibilities and developments. Many optimistic or pessimistic viewpoints about AI fail to align with reality [10]. As Hermann notes: "Science-

fictional AI is a dramatic element that makes a perfect antagonist, enemy, victim or even hero, because it can be fully adjusted to the necessities of the story.6 But to fulfil that role, it often has capabilities that are way beyond actual technology—be it natural movement, sentience, or consciousness. If science-fictional AI is taken seriously as a representation of real-world AI, it provides a wrong impression of what AI can and should do now and in future [17].” Science fiction, stemming purely from human imagination, cannot accurately depict real AI. Therefore, caution is warranted when interpreting SF to avoid misconceptions about AI and its implications for our future.

At the end, this article systematically examines two major narrative trajectories of AI through the lens of science fiction: the fear and caution embodied in the Frankenstein complex and the hope for coexistence with AI. It not only compares the complexities of AI-human relationships across various literary works but also delves into the profound impact of these narratives on the development of real-world AI technologies, public perception, and policymaking. Furthermore, the article explores how AI, in turn, transforms the methods and content of science fiction creation, emphasizing its role as a new collaborator and tool in storytelling. By integrating literature, technology, and society, this study reveals the bidirectional interaction between science fiction narratives and AI realities, offering fresh insights into the interplay between technological imagination and innovation.

References

[1] Araujo, T., Helberger, N., Kruikemeier, S., & De Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & society*, *35*, 611-623.

[2] Asimov I. The Naked Sun. 1st ed. Nanjing: Jiangsu Phoneix Literature and Art Publishing.LTD; 2013.

[3] Asimov I. Galactic Empire. 1st ed. Nanjing: Jiangsu Phoneix Literature and Art Publishing.LTD; 2015.

[4] Asimov I. I, Robot. 1st ed. Nanjing: Jiangsu Phoneix Literature and Art Publishing.LTD; 2013.

[5] Bareis, J., & Katzenbach, C. (2022). Talking AI into being: The narratives and imaginaries of national AI strategies and their performative politics. *Science, Technology, & Human Values*, *47*(5), 855-881.

[6] Bo, D., Ma’rof, A. A., & Zaremohzzabieh, Z. (2025). The Influence of Negative Stereotypes in Science Fiction and Fantasy on Public Perceptions of Artificial Intelligence: A Systematic Review. *Studies in Media and Communication*, *13*(1), 180-190.

[7] Borden, M. (2024). Intentions, Interpretations, and the Paradoxes of Asimov’s Laws of Robotics. *incite*, 15, 59-67.

[8] Baldassarre, M. T., Caivano, D., Fernandez Nieto, B., Gigante, D., & Ragone, A. (2023, September). The social impact of generative ai: An analysis on chatgpt. In *Proceedings of the 2023 ACM Conference on Information Technology for Social Good* (pp. 363-373).
[9] Chattopadhyay, S. (2024).“I think, therefore I am”: Retro-futuristic Realities of the Developing AI and its Future in Science Fiction Narratives.*Creativitas: Critical Explorations in Literary Studies*,*1*(1),197-215.
[10] Cave, S., & Dihal, K. (2019). Hopes and fears for intelligent machines in fiction and reality. *Nature machine intelligence*, *1*(2), 74-78.
[11] Cave, S., & Dihal, K. (2020). The whiteness of AI. *Philosophy & Technology*, *33*(4), 685-703.
[12] Clarke C A. 2001: A Space Odyssey. 1st ed. Shanghai: Shanghai Literature and Art Publishing.LTD; 2019.
[13] Chubb, J., Reed, D., & Cowling, P. (2024). Expert views about missing AI narratives: is there an AI story crisis?. *AI & society*, *39*(3), 1107-1126.
[14] Chubb, J., Reed, D., & Cowling, P. (2024). Expert views about missing AI narratives: is there an AI story crisis?. *AI & society*, *39*(3), 1107-1126.
[15] Geraci, R. M. (2007). Robots and the sacred in science and science fiction: Theological implications of artificial intelligence. *Zygon®*, *42*(4), 961-980.
[16] Herbert F. Dune. 1st ed. Nanjing: Jiangsu Phoneix Literature and Art Publishing.LTD; 2017.
[17] Hermann, I. (2023). Artificial intelligence in fiction: between narratives and metaphors. *AI & society*, *38*(1), 319-329.
[18] Hudson, A. D., Finn, E., & Wylie, R. (2023). What can science fiction tell us about the future of artificial intelligence policy?. *AI & SOCIETY*, 1-15.
[19] Hohendanner, M., Ullstein, C., Buchmeier, Y., & Grossklags, J. (2023, September). Exploring the Reflective Space of AI Narratives Through Speculative Design in Japan and Germany. In *Proceedings of the 2023 ACM Conference on Information Technology for Social Good* (pp. 351-362).
[20] Jenner, S., Raidos, D., Anderson, E., Fleetwood, S., Ainsworth, B., Fox, K., ... & Barker, M. (2025). Using large language models for narrative analysis: a novel application of generative AI. *Methods in Psychology*, *12*, 1-12.
[21] Kinzler, R. (2023). AI REVOLUTION: FROM SCIENCE FICTION TO REALITY.
[22] Mubin, O., Wadibhasme, K., Jordan, P., & Obaid, M. (2019). Reflecting on the presence of science fiction robots in computing literature. *ACM Transactions on Human-Robot Interaction (THRI)*, *8*(1), 1-25.
[23] McCauley, L. (2007, November). The frankenstein complex and Asimov’s three laws. In *Association for the Advancement of Artificial Intelligence: https://www. aaai. org/Papers/Workshops/2007/WS-07-07/WS07-07-003. pdf,(accessed 27/07/18)*.
[24] Murphy, R. R. (2022). The original “I, Robot” featured a murderous robot and the Frankenstein complex. *Science robotics*, 7(71), 1-2.
[25] Natale, S., & Ballatore, A. (2020). Imagining the thinking machine: Technological myths and the rise of artificial intelligence. *Convergence*, *26*(1), 3-18.
[26] Osawa, H., Miyamoto, D., Hase, S., Saijo, R., Fukuchi, K., & Miyake, Y. (2022). Visions of Artificial Intelligence and Robots in Science Fiction: a computational analysis. *International Journal of Social Robotics*, *14*(10), 2123-2133.
[27] Raj, A., Stroup, W. M., & Kayumova, S. (2025). Stories, Printing Press, Internet, and now ChatGPT: Examined via the SMART Framework. In *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning-CSCL 2025*, pp. 445-449. International Society of the Learning Sciences.

[28] Storey, V. C., Yue, W. T., Zhao, J. L., & Lukyanenko, R. (2025). Generative artificial intelligence: Evolving technology, growing societal impact, and opportunities for information systems research. *Information Systems Frontiers*, 1-22.
[29] Taeihagh, A. (2025). Governance of generative AI. *Policy and society*, *44*(1), 1-22.
[30] Watts, T. F., & Bode, I. (2024). Machine guardians: The Terminator, AI narratives and US regulatory discourse on lethal autonomous weapons systems. *Cooperation and Conflict*, *59*(1), 107-128.
[31] Primož K. The World of “Dune” as an Alternate Future Without AI. In:Ivan Matić, editors. Edited Book from the International Scientific Conference, Belgrade: Film and Politics; 2024, p. 13–32.