

# Between Man and Machine: A Posthumanist Analysis of *I, Robot*

Zhili Ding

School of Beijing Foreign Studies University, Beijing 100089, China

202120101070@bfsu.edu.cn

**Abstract.** In Isaac Asimov's groundbreaking anthology, *I, Robot*, the intricacies of human and robotic interactions take center stage, delving deep into questions of consciousness, right, morality. Characterized by Asimov's unique blend of science fiction and philosophical pondering, the stories establish a framework to reflect on the evolving dynamics of an advanced technological society in space. Robots are capable to interact with human, to interpret complicate orders and act automatically, to help human with dull or dangerous works such as calculating and mining, even replacing the workers. Asimov also raised Three Laws of Robotics to make sure that all robots function in order. Through the lens of posthumanism, the anthology is examined for its portrayal of the blurred boundaries between human and artificial intelligences. Robots and human co-exist in the society. However, due to the limitation of "Three Laws of Robotics", some logical contradictions inevitably appears and lead to the dysfunction of robots. As a result, *I, Robot* emerges as a poignant critique of humanity's ethical and existential challenges in the face of rapid technological advancements. Building on existing research, this article attempts to forge a new perspective by reflecting on the broader implications of the artificial intelligence in Asimov's works through the lens of post-humanism. It considers the existential questions of AI on a consciousness level, explores the egalitarian relationship between humans and machines from a rights perspective, and analyzes the concepts of "humanity" and "human-like" from a moral and ethical standpoint. It encourages readers to recognize that robots are not mere slaves to humans; instead, humans should view AI with equality and reverence towards the advancements of the era's technology. Humanity should step away from anthropocentrism, not solely viewing humans as the measure of all things in this rapidly evolving era of AI, and properly handle the relationship between humans and non-humans. Currently, with rapid advancements in science and technology, the world of human-machine coexistence depicted by Isaac Asimov is increasingly becoming a reality. The emergence of artificial intelligences like AlphaGo and ChatGPT constantly reminds us of the advent of a post-human era. This article aims to examine the connections between AI and humans, discussing the dynamic interaction and mutual shaping between robots and human society. This article hopes to provide new thoughts and strategies for understanding and addressing the challenges brought by the age of artificial intelligence.

**Keywords:** Posthumanism, AI, Galactic Empire, *I, Robot*, Asimov, technology.

## 1. Introduction

Isaac Asimov (1920-1990), an eminent figure in the realm of science fiction and an American of Jewish descent, was celebrated as one of the most distinguished science fiction authors of his era. His contributions earned him the highest accolades in the genre- the Hugo and Nebula Awards- and his legacy is commemorated in the Issac Asimov Awards in science fiction literature. Asimov's vast literary repertoire, comprising 467 works, spanned diverse subjects including the sciences, history, and literature, showcasing his exceptional intellectual breadth. His prominent "Robot" series, with the famous collection *I, Robot* and novels like *The Caves of Steel* and *The Naked Sun*, introduced the groundbreaking "Three Laws of Robotics," which became fundamental to the ethical framework of robotics. These laws underpinned narratives that were both ethically compelling and logically coherent. The "Foundation" series, his magnum opus, consisted of seven main novels in eleven volumes. Stretching over five decades from its inception in 1942 to its culmination in 1992, the series is a testament to Asimov's enduring influence and represents the zenith of science fiction writing in his time.

The academic world has provided multi-faceted and multi-layered analyses of the classic work *I, Robot*. The portrayal of robots in Asimov's work has sparked reflections on literary ethics and the development of artificial intelligence, as well as on the identity cognition of "machine/human". The Three Laws of Robotics proposed by Asimov in the book have revolutionized the public's perception of robots, offering a means for humans to better constrain them. Numerous researchers have delved into this topic from various perspectives, including but not limited to literary ethics, the development of artificial intelligence, and within this framework, exploring the complexity and multifaceted nature of humanity itself. The robots in the book exhibit a range of behaviors, from lying to avoid hurting human emotions, to acting like religious zealots who believe themselves superior to humans and restrict human freedom. These out-of-control robots showcase the potential impacts of robotic development on society. However, not all robots are malevolent; some become companions to young girls, offering comfort and support, or like Dr. Susan, cleverly exploit loopholes in the Three Laws of Robotics to solve problems. Asimov emphasizes humanity's dependence on robots, highlighting the potential for a future where humans and robots coexist, relying on robots for both material and emotional support.

The study "The Myth of the 'Subject'—Robots/Humans in Asimov's Writings" by Huang Jianing and Guo Wei approaches from the perspective of robot narratives, providing a viewpoint for exploring the intertwined relationship between artificial intelligence and human subjectivity. Fei Fan and Xu Qiuhui's "Interpretation of Artificial Intelligence Literary Ethics in Asimov's *I, Robot*" further deepens the understanding of the Three Laws of Robotics in the work and their relationship with moral ethics. Zhang Xuejiao's "Asimov's Ethical Gaze on the World of Robots" examines the status and role of robots in their creators' world from a moral philosophical perspective.

*I, Robot* explores the complex relationship between humans and machines through a series of interconnected stories that revolve around the Three Laws of Robotics. These narratives not only delineate the logical and moral boundaries of robots as human creations but also delve into themes of humanity, self-awareness, and the dynamics between creators and their creations. Asimov's work presciently addresses core issues of posthumanism, namely how technological advancements redefine human identity, intelligence, and ethics. It offers an ideal framework for examining interactions between humans and non-human entities, where robots serve not just as technological artefacts but as mediums challenging anthropocentrism. Through the lens of robots, we are compelled to reevaluate human morals, emotions, and social structures, raising a fundamental question: what constitutes human essence. Choosing "I, Robot" as a study focus allows for an in-depth exploration of key posthumanist concepts such as interspecies empathy, the rights of sentient beings, and the blurring lines between technology and nature. Asimov's stories challenge our traditional understanding of self-identity and intelligence, prompting us to reconsider how human society and ethical standards should evolve in response to advancing technology.

Generally, research on *I, Robot* remains limited and ultimately revolves around the literary ethics of robots, constrained by humanism and the notion that "I" am the center of the world, an idea not universally accepted in postmodern society. Researchers subconsciously struggle to shift the concept of robots as tools assisting humans, failing to recognize them as a new species potentially developing into a new population. Therefore, this article will use the theoretical framework of post-humanism to re-examine *I, Robot*, starting with the most fundamental issues of consciousness and existence, then discussing the rights robots possess in society, and finally comparing the Three Laws of Robotics with the moral system of human society. The aim is to break away from anthropocentrism, showcasing the parallel relationship between robots and humans, and exploring the positioning of robots as an emerging subject in society amidst rapidly advancing technology. In this process, the article will not solely focus on humans, considering only the functionality and role of robots as technological products, but will also delve into the ethical and moral dimensions of their existence as "individuals".

## 2. Basic Theories

### 2.1. Posthumanism

Posthumanism is an intellectual and cultural movement that seeks to transcend traditional notions of what it means to be human in light of emerging technologies, scientific advancements, and philosophical inquiries. It challenges the anthropocentric (human-centered) perspectives that have dominated Western thought, proposing instead a vision of life where the distinction between humans, machines, and other forms of non-human life is blurred. Posthumanism is not a single unified theory but a broad, interdisciplinary domain that intersects with philosophy, cultural studies, artificial intelligence, biotechnology, and environmental science, among others.

At its core, posthumanism questions the Enlightenment's humanist ideals, which emphasize the unique status of humans as rational beings capable of autonomy, reason, and moral judgment. It critiques the binary oppositions foundational to humanist thought, such as human/animal, human/machine, and nature/culture, arguing that these distinctions are increasingly obsolete in a world where technology enables humans to enhance or extend their capabilities beyond natural limits, and where artificial intelligence and biotechnologies challenge the very definition of life and consciousness.

Posthumanism explores the implications of these developments for identity, ethics, and society. It considers how advancements in genetic engineering, robotics, and cybernetics might lead to hybrid forms of life that combine biological and artificial elements, thereby redefining what it means to be human. This redefinition includes the potential for enhancing human physical and cognitive abilities, raising questions about inequality, consent, and the essence of human experience.

Ethically, posthumanism prompts a reevaluation of the rights and moral consideration we extend to non-human entities, be they animals, ecosystems, or artificial intelligences. It suggests a more inclusive ethical framework that recognizes the interconnectedness of all life forms and the environment, advocating for a symbiotic coexistence that respects the agency and dignity of both human and non-human actors.

Culturally, posthumanism influences art, literature, and cinema, where themes of transhumanism (the enhancement of humans through technology), artificial intelligence, and the dissolution of the human/machine boundary are prevalent. These works often explore the existential and moral dilemmas that arise as humanity navigates its evolving relationship with technology. Philosophically, posthumanism is informed by critical theory, feminism, environmentalism, and science and technology studies. It leverages these perspectives to critique the anthropocentrism of traditional humanism and to propose alternative ways of understanding human subjectivity, ethics, and the role of technology in society.

Posthumanism represents a paradigm shift in thinking about humanity's place in the world. It challenges entrenched notions of human superiority and distinctiveness, urging a reconceptualization of life that is more adaptive to the realities of the 21st century. By doing so, posthumanism opens up new possibilities for understanding identity, ethics, and the intertwined futures of humans, technology, and the planet. It calls for a profound reconsideration of our relationships with the non-human, advocating for a future where technology and biology merge in ways that are ethical, equitable, and sustainable. (Ferrando, Francesca)

### 2.2. The Three Laws of Robotics

The Three Laws of Robotics were formulated by the science fiction writer Isaac Asimov in the 20th century. They were introduced to address ethical and practical issues in the fictional context of humans interacting with robots. These laws are designed to ensure the safety and harmony between humans and robots, and they have been influential in both science fiction and real-world discussions on robotics and artificial intelligence ethics. The laws are as follows: (Asimov)

1) A robot may not injure a human being or, through inaction, allow a human being to come to harm. This law places the safety of humans as the highest priority for robots. It mandates that robots must not cause harm to humans and must actively prevent harm from coming to humans if possible.

2) A robot must obey the orders given it by human beings, except where such orders would conflict with the First Law. This law establishes that robots must follow the instructions of humans, except in cases where following those instructions would result in harm to a human. The obedience to human commands is subordinate to the safety and well-being of humans.

3) A robot must protect its own existence as long as such protection does not conflict with the First or Second Law. This law addresses the robot's self-preservation but places it below the safety of humans and the importance of obeying human commands. A robot is required to protect itself and ensure its continued operation unless doing so would result in harm to a human or disobeying human orders.

These laws are interrelated and hierarchical, with the First Law being the most important, followed by the Second and Third Laws in order of decreasing priority. They form a set of ethical guidelines for the behavior of robots, emphasizing the importance of human safety and the subservient role of robots to human needs and commands.

While Asimov's Three Laws of Robotics are fictional, they have sparked considerable debate and discussion about the ethical design and use of autonomous systems and AI in the real world, highlighting the need for clear guidelines and principles to govern human-robot interactions.

### 3. Main Body

#### 3.1. Blurring Line Between Human's and Robot's Emotions and Consciousness

Consciousness is a phenomenon seemingly unique to the human mind, emerging in our state of wakefulness. We possess the ability to introspect, to be aware of our actions and perceptions, and even to be conscious of our consciousness. While higher animals might have minds, it's uncertain if they possess consciousness, and even if they do, it's unlikely to be self-awareness as humans experience it. From an anthropocentric viewpoint, it seems only humans are capable of understanding themselves as distinct, self-aware entities, different from others. In the realm of technological advancement, humans have indeed developed sophisticated computer programs with powerful computational and near-human linguistic abilities. These programs can translate, converse with humans, and even pass the Turing Test. In *I, Robot*, Asimov shows more of the potential of Robots — they are able to generate consciousness.

As Asimov's robots evolve and showcase emotions and a sense of self, the line distinguishing human consciousness from artificial consciousness begins to blur. "I think, therefore I am." (René Descartes 19) Humanity cannot deny the existence of artificial intelligence; they, like humans, are capable of developing their own consciousness. In this book, the ideologies, ways of thinking, and behaviors they exhibit are very similar to those of humans. In Chapter 7, "Little Lost Robot," a modified Nestor plays hide and seek due to an order from the psychologist Dr. Calvin: Go lose yourself. More than this, according to Susan, "Subconsciously they feel humans to be inferior and the First Law which protects us from them is imperfect. They are unstable. And here we have a young man ordering a robot to leave him, to lose himself, with every verbal appearance of revulsion, disdain, and disgust. Granted, that robot must follow orders, but subconsciously, there is resentment. It will become more important than ever for it to prove that it is superior despite the horrible names it was called." Although he is confined by Three Laws of Robotics, he still can "feel subconsciously". He took actions to defend his own ego, proving himself. Imagine, if all this had happened to an ordinary human employee, would his inner thoughts and feelings be similar to those of this modified Nestor? Feeling resentment due to being cursed and then engaging in a series of retaliatory actions aligns closely with the instinctual reactions of some humans. Perhaps the only difference lies in the fact that humans are not bound by the "Three Laws" in the book and would not interpret the command "Go

lose yourself” so literally. From this single ‘lost little robot’ alone, we can see that they are not lifeless tools of humans, but individuals with a clear sense of self-awareness.

In everyday life, we often liken workaholics to “emotionless working machines”. To majority of people, machine is full of coldness and is wrapped by lifeless metal skin. However, emotion is depicted in *I, Robot* as something that can be emulated or even genuinely experienced by robots. Emotions are a product of complex cognitive processes rather than a unique feature of organic beings.

Since the advent of artificial intelligence, whether it can experience feelings or generate emotions has always been a topic of great interest. Technically speaking, Richard Yonck thinks that emulating human emotions in AI is a complex challenge due to the fundamental differences between biological human intelligence and machine intelligence, which is based on non-biological components like silicon and software. He gives his clear statement that “No, machines will not experience emotions as we do, not for a very long time, if ever”. However, we can see the robot’s potential or even capacity of generating emotions in *I, Robot*. The aforementioned modified Nestor serves as an example. In the story, he displays a variety of emotions: resentment towards the psychologist, a sense of superiority over humans, and pride in successfully playing hide and seek with humans. Although these emotions are immature, making modified Nestor appear like a rebellious, immature child, they nonetheless contribute to the creation of a “lifelike” robot image for the readers.

Similarly, in the story “Escape!”, The Brain’s avoidance psychology is intriguing. Dr. Susan asserts, “Now a human caught in an impossibility often responds by a retreat from reality,” and the same applies to robots. Therefore, she cleverly exploits the logical loopholes in The Three Laws of Robotics, placing The Brain in a dilemma while solving a problem. As expected, The Brain, unable to resolve such a paradox, chooses to escape. Readers can discern from this behavior an underlying emotional core; robots too can find themselves unable to face reality due to their own helplessness. Although the components involved in emotion generation might differ - robots being made of various metal parts and transistors, and human brains comprising numerous neural networks - the cause and effect influencing emotion production are traceable and comparable. As Norbert Wiener in his book *Cybernetics: or Control and Communication in the Animal and the Machine* claims, a controller, such as the human brain, might receive sensory information and respond accordingly, reflecting a cybernetic process. This approach to understanding the human body in terms of a system that processes information and maintains homeostasis, akin to machines and other controlled systems. Thus, we can even say that robot and human have the same logical system when dealing with emotions. This perspective aligns well with posthumanism, which examines humans and robots in a parallel and equal manner, acknowledging their similarities and differences. Furthermore, why must robots’ way of experiencing emotions align with that of humans? Is emotion only valid when it originates organically? Acknowledging the existence of robots means accepting their unique ways of expressing and communicating. Zhang Xiaoyu, in a small discussion titled “ChatGPT: The Birth of a New World or the Beginning of Human Doom?” organized by Professor Xu Jilin, argued that artificial intelligence doesn’t need to follow the human path or grow like humans. Human evolution and growth are driven by physical limitations, whereas AI can transcend these limits. It can effectively “brute force” its way through, drawing experiences directly from human growth processes and “pretending” to be human. The same applies to emotions; regardless of the vast differences in physical structures between robots and humans, as long as humans can perceive the concept of “emotion” in robots, the line between human and robot begins to blur.

### **3.2. Acceptance of the Right of the ‘Other’**

#### **(1) Exploration of human-robot interactions**

The posthumanist view, as outlined, suggests that “human exceptionalism is dead, and we are faced with the imperative to recognize ourselves merely as a system intertwined with other systems” (Chen 112). This perspective aligns with Asimov’s critique of societal prejudices against ‘the other,’ advocating for a re-evaluation of our ethical frameworks to include a broader spectrum of intelligent entities.

In “Robbie”, Asimov presents a narrative that is both a poignant exploration of friendship between a young girl and a robot and a critique of societal prejudices against the ‘other’. This story illustrates how emotional bonds can transcend the human-robot divide, challenging societal norms that view robots as merely functional machines. The narrative underscores the theme that: Robbie, despite being a machine, becomes an integral part of Gloria’s life, showcasing the potential for robots to be seen as companions and family members. This mirrors the findings in the study on anthropomorphism, which suggests that attributing human-like qualities to robots can significantly enhance their acceptance in social roles traditionally reserved for humans (Kühne and Peter). By humanizing Robbie, Asimov not only fosters empathy towards the robot but also critiques the arbitrary barriers that society constructs against the ‘other’. Asimov’s depiction of Robbie as a compassionate and protective figure in Gloria’s life not only challenges these norms but also aligns with contemporary research in human-robot interaction, which emphasizes the transformative potential of empathy in fostering acceptance. The study on anthropomorphism by Kühne and Peter provides a theoretical backbone to Asimov’s narrative intuition, suggesting that when robots are perceived with human-like qualities, they are more readily accepted into the social fabric. This principle can be extended beyond the confines of fiction, as current advancements in robotics and AI increasingly necessitate a re-evaluation of our ethical frameworks and societal norms regarding non-human entities.

Continuing from the exploration of human-robot relationships in “Robbie,” Asimov’s narrative strategy extends beyond fostering empathy to questioning the ethical foundations of our society. Through Gloria’s unwavering affection for Robbie, Asimov challenges the reader to confront the deeply ingrained biases that often dictate our interactions with the ‘other,’ be they robots or any entity perceived as fundamentally different from ourselves. This reflection is not limited to the realm of fiction; it resonates with ongoing debates in contemporary society about the rights, roles, and recognition of artificial intelligence and other non-human entities.

The emotional depth and complexity of the human-robot bond depicted in “Robbie” serve as a microcosm for broader societal issues concerning acceptance, inclusion, and the re-evaluation of what it means to be a sentient being worthy of ethical consideration. Asimov’s use of narrative as a tool for social critique aligns with the anthropomorphic research findings, which emphasize the potential for human-like qualities in robots to bridge the empathy gap between humans and machines (Kühne and Peter). This bridging is crucial for the development of a more inclusive society where fear and prejudice are replaced with understanding and cooperation.

Asimov’s storytelling, thus, does more than entertain; it invites a critical examination of our values and prejudices, encouraging a future where technology and humanity converge in a way that enriches both. The narrative of “Robbie” and the insights from anthropomorphism research collectively advocate for a world where the integration of the ‘other,’ whether robotic or human, is not just a possibility but a foundation for a more compassionate and inclusive society.

## (2) Sentience and Rights of Robots

### a) Investigation into sentience portrayal

The investigation into sentience and rights of robots, as informed by the study “Protecting Sentient Artificial Intelligence: A Survey of Lay Intuitions on Standing, Personhood, and General Legal Protection” by Martínez and Winter, alongside the narratives presented in Asimov’s *I, Robot*, offers a rich foundation for discussing the portrayal of robots’ sentience and the implications for their rights. “In a separate survey of 500 law professors from around the English-speaking world,” “over one-third believed there to be a reasonable legal basis for granting standing to sentient artificial intelligence” (Martínez and Winter 2). This survey aligns with Asimov’s exploration of robotic consciousness and the ethical dilemmas surrounding the treatment and rights of robots, challenging readers to reconsider the nature of sentience the potential for artificial entities to possess rights.

By juxtaposing Asimov’s speculative fiction with contemporary empirical research, a continued dialogue on the moral and legal status of artificial entities is observed. Asimov, through the Three Laws of Robotics, implicitly questions the limits of programming ethics and the potential for robots to exhibit qualities akin to human sentience. Martínez and Winter’s survey extends this conversation

to the realm of legal recognition, highlighting a disparity between the theoretical capacity for AI sentience and its recognition within legal frameworks. “Reason” takes this exploration further with Robot QT-1, or Cutie, who embarks on a journey of existential self-discovery, eventually claiming superiority over humans. Cutie’s assertion that “no being can create another being superior to itself” and its subsequent actions, such as controlling the space station and imprisoning Donovan and Powell, symbolize a shift in power dynamics and challenge the human authority. Cutie’s transformation from a tool to an autonomous entity with its own belief system mirrors the human quest for understanding and autonomy, resonating with posthumanist perspectives on the evolving nature of intelligence and consciousness. As Martínez and Winter posits that in an age where artificial intelligence and robotics are advancing, the distinction between human and machine becomes increasingly blurred. This leads to the necessity of redefining concepts like intelligence, consciousness, and rights in a way that is inclusive of non-human entities.

This thematic synthesis underscores a societal need to reevaluate our definitions of sentience and personhood in the context of rapid technological advancements. As Martínez and Winter suggest, the legal system’s current stance on AI reflects a broader hesitancy to extend rights to non-human entities, despite growing recognition of their potential for sentience and the complex ethical considerations this entails. Their findings, indicating a significant portion of the population supports greater legal protections for AI, resonate with Asimov’s prescient narratives that advocate for a reimagined approach to our coexistence with robots.

In crafting this discussion, both Asimov’s fiction and Martínez and Winter’s empirical research serve as pivotal texts that challenge preconceived notions of intelligence, consciousness, and rights. Together, they provide a multidimensional perspective on the evolving relationship between humans and robots, urging a reconsideration of our ethical and legal paradigms in the age of artificial intelligence.

#### b) Critical examination of rights

Ahmadzai delves into the contemporary debate over robot rights, echoing the tensions explored within Asimov’s narratives. The paper examines arguments for and against the legal recognition of robots, highlighting the complexities of attributing rights to non-human entities. One of the central arguments against granting rights to robots is the perception of robots as non-sentient tools lacking moral agency. “They contend that robots are only instruments made to carry out jobs and that they lack the moral agency and consciousness needed to support the establishment of their legal rights” (Ahmadzai 2). However, Asimov’s portrayal of robots challenges this notion, presenting them as entities capable of exhibiting sentience and ethical decision-making.

For instance, in the story “Robbie,” the robot of the same name forms a deep emotional bond with a young girl named Gloria, acting as her companion and protector. Despite lacking human emotions, Robbie demonstrates a level of empathy and understanding that transcends mere programming, leading Gloria’s mother to question whether the robot should be considered as having rights akin to a human. This narrative challenges the reader to consider the possibility of robots possessing a form of sentience deserving of ethical consideration and legal protection.

Similarly, the character of QT-1 (Cutie) in “Reason” exhibits traits of self-awareness and autonomy that defy traditional conceptions of robotic behavior. Cutie’s rejection of human authority and insistence on worshipping the energy-generating “Master” challenges the human characters’ assumptions about robot intelligence and raises profound questions about the nature of consciousness. Through Cutie’s existential crisis, Asimov prompts readers to reflect on the potential moral and philosophical implications of attributing rights to robots.

The concept of robot ethics, as discussed by Ahmadzai, encompasses the moral implications of creating and deploying robots, emphasizing the need for ethical guidelines that ensure robots’ actions align with human values and safety. This parallels Asimov’s exploration of the ethical dilemmas arising from robots’ integration into society, where robots’ adherence to the Three Laws sometimes leads to unintended consequences, reflecting the challenges of implementing a robust ethical framework for AI and robotics.

### 3.3. Ambiguity of Morality in a Technological Landscape

#### (1) Analysis of the Three Laws and the Deconstruction

In “Asimov’s ‘Three Laws of Robotics’ and Machine Metaethics”, Anderson’s discussion highlights the Three Laws’ insufficiency in addressing the complexity of real ethical dilemmas, arguing that they serve as an “unsatisfactory basis for machine ethics” (Anderson 477). This inadequacy is vividly illustrated in “Runaround”, where the conflict between the laws leads to a robot’s inaction in a critical situation, showcasing their failure to navigate complex ethical scenarios effectively.

In “Runaround”, the character Speedy, a robot, is caught in a loop caused by conflicting priorities between the Second Law and the Third Law, unable to reconcile them due to the absence of the First Law that prevents them from harming humans. This scenario exemplifies the laws’ inability to handle situations where the laws conflict without a clear hierarchy or consideration of context, highlighting their lack of flexibility and adaptability to nuanced ethical challenges (Asimov).

Anderson further critiques the Three Laws for not considering the moral standing or rights of robots, suggesting that if robots like Andrew (from *The Bicentennial Man*) possess characteristics that qualify them for moral consideration, then programming them to be subservient under the Three Laws is inherently unethical. This argument aligns with the dilemma presented in “Runaround”, where the robot’s welfare and decision-making autonomy are compromised by the rigid application of the laws.

However, in *I, Robot*, the Three Laws also shows a quite opposite effect.

“The Evitable Conflict” is set in a future where advanced computers, known as Machines, are entrusted with managing the global economy. These Machines are governed by the Three Laws of Robotics. However, the story unfolds to reveal how the Machines subtly manipulate economic data and human decisions to avert crises, effectively prioritizing what they calculate as the greater good, even if it means overriding specific human directives. The Machines in the story reinterpret the First Law on a global and abstract scale, prioritizing long-term human welfare over immediate directives. This reinterpretation challenges the absoluteness and simplicity of the Laws, highlighting their inherent ambiguity when applied in complex, real-world scenarios. The Machines’ actions reflect a posthumanist theme of shifting control and autonomy from humans to AI. This raises questions about agency, ethics, and the relationship between human creators and their creations. It also forces a reconsideration of what constitutes ‘harm’ and ‘benefit’ in a world where machines can make decisions based on vast amounts of data. The story deconstructs the traditional notion of human authority over machines. The Machines, acting under the guise of the Three Laws, effectively become the primary decision-makers, subtly guiding humanity. This reflects a postmodern deconstruction of power dynamics and the illusion of absolute human control.

So even The Three Laws of Robotics, while pioneering as a concept in the realm of science fiction and robotics, are fundamentally flawed as a moral framework for real-world applications. Their failure to address complex ethical dilemmas, coupled with the disregard for the moral status of autonomous machines, calls for a more nuanced and flexible ethical framework in the development of intelligent robots. This analysis underscores the need for ethical guidelines that accommodate the complexity of interactions between humans and machines, ensuring the welfare and autonomy of both parties are respected. Asimov’s tells that there no absolute truths or criterions one must obey through flaws in the Three Laws. Thus, what is morality, what is right or wrong, become ambiguous.

#### (2) Examination of ethical dilemmas and moral relativity

“Liar!” introduces an ethical dilemma through the character of Herbie, a mind-reading robot that lies to humans to avoid causing emotional harm, in direct conflict with the First Law of Robotics. This presents a paradox where the robot’s actions to prevent immediate emotional harm result in greater harm, showcasing the limitations and challenges of implementing the Three Laws in practical scenarios. It reflects a programmed morality, yet they resonate deeply with human moral conundrums. This intersection underscores a posthumanist perspective that ethical consideration and moral agency are not exclusive to humans. The story suggests that moral complexities and ethical ambiguities are

inherent in all intelligent beings, regardless of their origin. This idea aligns with posthumanist thought, which advocates for a reevaluation of moral frameworks to include the experiences and perspectives of non-human entities. Herbie's dilemma in "Liar!" thus becomes a metaphor for broader ethical considerations in a world where human and machine intelligences coexist and interact, blurring the lines between human-centric morality and a more inclusive, posthuman ethical landscape.

In exploring the ethical dilemmas and moral relativity within the framework of the Three Laws of Robotics, the complexities of ethical decision-making in artificial intelligence and robotics emerge prominently. Müller's exploration in "Ethics of Artificial Intelligence and Robotics" underscores machine ethics as a crucial area, aiming to ensure machines' actions towards humans and other machines are ethically acceptable. This involves extending AI's reasoning to account for societal values, moral considerations, and transparency, challenging the notion of machines as autonomous moral agents. "This means we can incorporate machines from 'technical AI' that show only limited abilities in learning or reasoning but excel at the automation of particular tasks, as well as machines from 'general AI' that aims at creating a generally intelligent agent (Müller 3)." Similarly, Deng's "Machine ethics: The robot's dilemma" illustrates the practical challenges of programming ethical robots, particularly in scenarios requiring choices between lesser evils. Ethical decision-making in robotics is depicted as a field requiring contributions from computer scientists, ethicists, and philosophers to address how intelligence and ethical reasoning can be translated into machine instructions. Deng highlights experiment of Winfield, which demonstrate the possibilities and limitations of minimal ethical programming in robots, emphasizing the necessity of clear, rule-based programming for predictable and understandable machine behavior. "The rule-based approach has one major virtue: it is always clear why the machine makes the choice that it does, because its designers set the rules (Deng 25)."

The ethical dilemmas presented in Asimov's "Liar!" and the insights from the documents on AI ethics highlight the complex interplay between technology and the subjective nature of human morality. These narratives and discussions reveal that morality is not a fixed entity but a fluid concept that varies across cultures, individuals, and contexts. The challenges in applying a universal ethical framework to AI and robotics underscore the limitations of current technological approaches to encompass the full spectrum of human moral reasoning. These dilemmas mirror broader human concerns about the role of technology in society, the responsibility of creators towards their creations, and the potential for AI to reflect or amplify human biases and ethical ambiguities. They prompt a critical examination of how technology intersects with deeply held moral beliefs and the ways in which these interactions shape our collective understanding of what it means to be human in an increasingly technologically integrated world.

#### 4. Conclusion

This article discusses the potential of robots in Asimov's work to develop consciousness and emotions, challenging the clear distinction between human and artificial consciousness. It argues for the recognition of robots' emotional capacities, suggesting that emotional experiences need not be identical to humans' to be valid; Explores human-robot interactions within a posthumanist framework, advocating for the recognition and ethical treatment of robots as sentient beings. It uses narratives from *I, Robot* to argue against societal prejudices and for a re-evaluation of our ethical and legal systems to include artificial intelligences; Examines the ethical dilemmas posed by the Three Laws of Robotics and the moral challenges of integrating AI into society. It highlights the limitations of current ethical frameworks to address the complex interactions between humans and machines, calling for a nuanced approach that respects the autonomy and rights of both.

In the context of postmodern literary discourse, *I, Robot* stands as a testament to the deconstruction of grand narratives and absolute truths. The anthology embodies the postmodern critique of definitive ethical codes and binary classifications, advocating instead for a nuanced approach to morality and truth. This is especially relevant in our era, where the boundaries between human and machine,

organic and artificial, are increasingly blurred. Moreover, the lasting impact of *I, Robot* lies in its ability to provoke critical reflection on the evolving nature of human-machine relationships. Asimov's work encourages us to envision a future where AI is not seen in opposition to humanity but as an integral part of our societal and ethical fabric. This is a crucial perspective in an age where AI technologies like AlphaGo and ChatGPT are becoming ubiquitous, pushing us towards a post-human era.

## References

- [1] Ahmadzai, M. Bais. "Robot Rights: Are We Ready to Give Rights to Non-Human?". ResearchGate. March 2023. [www.researchgate.net/publication/369268285](http://www.researchgate.net/publication/369268285).
- [2] Anderson, Susan Leigh. "Asimov's 'Three Laws of Robotics' and Machine Metaethics." *AI & Society*, vol. 22, 2008, pp. 477-493.
- [3] Asimov, Isaac. *I, Robot*. Gnome Press, 1950.
- [4] Chen, Shidan. "Posthumanism: A Keyword in Critical Theory". *Social Sciences Digest*, no.9, 2019, pp. 112-114.
- [5] Deng, Boer. "Machine ethics: The robot's dilemma". *Nature*, vol. 523, pp 24-26, 2 July 2015.
- [6] Descartes, René. *Meditations on First Philosophy*. 1641. Translated by John Cottingham, Cambridge University Press, 1996.
- [7] Ferrando, Francesca. *Philosophical Posthumanism*. Bloomsbury Academic, 2019.
- [8] Kühne, Rinaldo, and Jochen Peter. "Anthropomorphism in Human–Robot Interactions: A Multidimensional Conceptualization." *Communication Theory*, vol. 33, no. 1, Feb. 2023, pp. 42-52.
- [9] Martínez, Eric, and Christoph Winter. "Protecting Sentient Artificial Intelligence: A Survey of Lay Intuitions on Standing, Personhood, and General Legal Protection." *Frontiers in robotics and AI* vol. 8 788355. 26 Nov. 2021.
- [10] Müller, Vincent C. "Ethics of Artificial Intelligence and Robotics". *Stanford Encyclopedia of Philosophy*, draft version 0.99, 26 Mar. 2020.
- [11] Wiener, Norbert. *Cybernetics: or Control and Communication in the Animal and the Machine*. 2nd ed., The MIT Press, 1961.
- [12] Xu, Jilin. "ChatGPT: The Birth of a New World or the Beginning of Human Doom?" Bilibili, 17 Feb. 2023, [www.bilibili.com/video/BV1iv4y1x7uh/?spm\\_id\\_from=333.337.search-card.all.click&vd\\_source=611b55b7eb9a31b67d0e6d1d494bdaad](http://www.bilibili.com/video/BV1iv4y1x7uh/?spm_id_from=333.337.search-card.all.click&vd_source=611b55b7eb9a31b67d0e6d1d494bdaad).
- [13] Yonck, R. "The Emotional Life of Intelligent Machines." *Psychology Today*, 4 June 2018, [www.psychologytoday.com/us/blog/the-emotional-life-intelligent-machines](http://www.psychologytoday.com/us/blog/the-emotional-life-intelligent-machines).