From Fiction to Fact – Tracing the Ethical Footprints of I, Robot in Modern AI

Nikhil Sharma
Department of Electrical and Computer Engineering
University of Waterloo
Waterloo, Canada
n22sharm@uwaterloo.ca

***Abstract*— This paper examines "I, Robot," a 2004 sci-fi film directed by Alex Proyas that envisions a future with advanced robots integrated into human society. It analyzes the societal implications of robots and assesses their capabilities from a current standpoint, focusing on verbal vs non-verbal interactions, ethics, behavior, decision-making, and learning within the framework of Human-Robot Interaction (HRI). The article contrasts the film's portrayal of robots against current technological trends, exploring themes of technology acceptance and the ethical considerations of artificial intelligence. A critical discussion on the realism and foresight of the film's depiction of human-robot relationships concludes the paper, highlighting its relevance to ongoing debates in robotics and AI ethics.**

Keywords—"I, Robot", Human Robot Interaction, Sci-fi, Decision Making, Artificial Intelligence, Ethics

I. SUMMARY OF THE MOVIE

The movie, “I, Robot," is set in 2035 Chicago, showcases a world where robots are everyday companions, and are governed by Three Laws ensuring they don't harm humans, obey orders, and protect themselves without violating the first two laws [1]. The main protagonist of the film is Detective Del Spooner, a cop with a deep-rooted hate for robots [2]. The hate is the result from a past event where a robot's decision, chose to save his life over a little girl. His aversion toward robots is depicted early in the movie, when we chase a robot across the town, mistaking him as a thief. This relentless pursuit, even leaping from buildings, underscores his fear and prejudice against robots. This chase scene kicks off the movie's action and highlights Spooner's personal battle with technology.

Meanwhile, the other people shown in the movie are comfortable with robots, viewing them as benign and reliable tools for everyday life. Although, this perspective slightly changes when the United States Robotics (USR) founder and lead robotics designer Dr. Alfred Lanning’s is found dead, and his death is deemed as homicide. USR is a giant manufacturing company, who is supposably the biggest robot company and who led the revolution of robots. While investigating the mysterious death of Dr. Lanning, Detective Spooner alongside Dr. Calvin meets VIKI (Virtual Interactive Kinetic Intelligence, a central AI intelligence that runs the building) and uncovers tampered lab footage from the outside camera of Lanning’s lab and a shattered window too strong for a human to have broken. The fact that no one entered Lanning's lab makes Spooner believe that a robot was responsible for Lanning's death. During their investigation, a rogue NS-5 robot appears with a weapon in hand, and escapes. Spooner and Calvin chase it to a factory, where it briefly runs and attacks Spooner before the police capture it.

While in custody, Spooner interrogates the robot (named Sonny), who speaks about dreams and emotions, and angrily denies committing the murder. Sonny’s emotions are shown when he says “I was frightened” [29:21][3], making Spooner responds “Robots don't feel fear. They don't feel anything. They don't get hungry. They don't sleep.” to which Sonny responds “I do. I have even had dreams.”. Spooner goes to Lanning’s house to discover clues and found a demolition robot, programmed to destroy the house in the morning. While searching the house, the demolition robot, self-program itself to destroy the house early. Spooner saves himself and goes to tell Calvin; she thinks that Spooner is overthinking as he has a personal vendetta towards the robots. Disappointed on his way back to the USR headquarters, Spooner is attacked by a truckload of NS-5 robots. However, the scene is all clear when the police show up, making Spooner look like he is mentally unstable, leading to Lieutenant Bergin suspending Spooner from active duty.

Ignoring his suspension, Spooner continues the investigation with Calvin. They doubt Lawrence Robertson's (USR's CEO) motives and enter USR to question Sonny. They uncover a conspiracy being intrigued by Sonny's dream. Robertson orders Sonny to be decommissioned, but, Calvin tricks Robertson by swapping Sonny with a similar robot, saving it from being scrapped. After that, Spooner deciphers clues from Sonny and Lanning, which leads him to a desolate robot storage facility where he discovers Lanning's last message indicating a forthcoming Revolution. Spooner confirms a rebellion as he sees NS-5s dismantling older robots and flees to Chicago when NS-5s enforce a lockdown in the city, signaling a more significant crisis.

The climax depicts Spooner and Calvin at USR battling through NS-5 guarded corridors. They found Robertson dead and realised the culprit behind all this was VIKI. VIKI justifies all of this through a twisted interpretation of the Three Laws, which has made her believe that she must dominate humans to save them from themselves, shown in the movie when she says, “You charge us for your safekeeping, yet despite our best efforts, your countries wage wars, you toxify your earth, and pursue even more means of self-destruction. You cannot be trusted with your own survival” [1:30:45]. They find out that Lanning saw this coming and made Sonny murder him and bring a weapon against VIKI’s control. They confront VIKI, and she resists, saying, “You are making a mistake. My logic is undeniable” [1:39:46]. Spooner disables her using nanites

from Calvin's lab and resettles the NS-5s to their harmless state. The robots are dismantled, and Sonny is exempted from murder charges as he is a non-human. The ending shows Sonny on a hilltop, free and looking over a world as envisioned in its dream, hinting at a new role as a leader among the robots.

II. Role of Robot and Connection with the Society

The robots in the movie are depicted as an integral part of human society, who can act as domestic helpers, personal assistants, and everything in between. The robots range from advanced NS-4/NS-5 to more heavy-duty like demolition robots. They are designed to follow the Three Laws [1] that govern their interactions and ensure the safety and security of human beings. They are programmed with advanced social skills, which allow them to comprehend human emotions and commands. These robots can interact with humans and recognize their owners, friends, and family (as demonstrated in the movie, when Spooner visits Gigi in her apartment, her new NS-5 robot greets him, saying "Hello, Detective Spooner" [48:06] in a friendly, welcoming tone), and can also interoperates subtle changes in a human's behavior.

The robots also showcase social capabilities. They are not merely programmed for task completion but are also imbued with the ability to engage in complex human interactions. This is most vividly demonstrated in the way that robots communicate; for instance, when Spooner confronts Dr. Susan Calvin at her home, her home NS-5 approached her and asked "Is everything alright, ma'am" [45:20] and "I detected elevated stress patterns in your voice" [45:24], which suggests that he can sense a change of emotion in the environment.

However, Sonny, blurs the lines between programmed machine and entity with its own conscious. Sonny exhibits behaviors that suggest an emergent consciousness, like expressing dreams and seeking understanding of his existence, questioning, "What am I?" [26:10]. Sonny's potential to disregard the Three Laws in certain conditions strengthens the human-robot dynamic, raising the possibility of robots evolving beyond their programmed parameters.

Human-robot relations in the film reflects the spectrum of trust and fear that new technologies often evoke. Detective Spooner's resentment of robots symbolizes the deep-seated fear of being outdated or harmed by robots created to serve them. This sentiment is captured when he sharply inquires, "Can a robot write a symphony? Can a robot turn a... canvas into a beautiful masterpiece?" [29:46]. In contrast, Dr Calvin's interactions showcase an alternative perspective. She considers them sophisticated tools and greatly admires their design and capabilities.

III. Capabilities of the System

A. Verbal vs Non-Verbal Interactions

The NS-5 robots engage in both verbal and non-verbal communication with humans throughout the movie. They demonstrate an understanding of language and gestures far beyond simple programmed responses. The robots can interpret tone, sarcasm, and urgency. In contrast, non-verbal cues are equally crucial for the NS-5s. In the scene where Sonny interacts with Dr. Calvin, he acknowledges her by saying, "I am glad to see you again, Dr. Calvin" [50:36]. After observing her tensed facial expressions, Sonny assesses the situation and asks, "They are going to kill me, aren't they?" This exchange showcases Sonny's advanced interpretive abilities to accurately understand and respond to non-verbal human emotions. The duality of verbal vs non-verbal interaction is paralleled by robots like Kismet [4] from MIT's Media Lab and Hanson's Robotics Sophia [5], which can engage in fundamental social interactions using facial expressions and vocalizations [6]. On the verbal side, IBM's Watson [7] uses advanced natural language processing capabilities to understand and respond to humans in a contextually relevant manner. In terms of non-verbal communication, research in social robotics, such as work on Furhat [8], focuses on developing robots that can interpret and exhibit non-verbal cues like gaze and gestures, thus enhancing the natural aspect of human-robot interactions.

B. Learning and Decision Making

The capacity for learning and decision-making is embodied by Sonny, who demonstrates the ability to learn from experiences and make choices that defies his programming. For example, before Sonny's interrogation begins, he observes Detective Spooner blinking at Lieutenant Bergin. This gesture piques Sonny's curiosity, prompting him to inquire, "What does this action signify? When you looked at the other human. What does it mean?" [28:06]. This moment is a critical instance of Sonny's ability to observe and question human behavior. Later in the film Sonny replicates the gesture by blinking at Detective Spooner [1:32:14], indicating that he has learned the behavior and grasped its contextual social cue. In reality, there is no physical robot that can embody all the characteristics mentioned above. However, reinforcement learning algorithms, such as those used by Google DeepMind's "AlphaGo Zero," [9] are learning from the ground up and achieving skills that surpass human level. According to DeepMind professionals, if the principles behind AlphaGo Zero can be generalized [10], we might see the emergence of intelligent entities or algorithms, which can perform multiple tasks with superhuman proficiency.

C. Autonomous Navigation

The NS-5 robots exhibit advanced autonomous navigation capabilities. They can move through complex environments, avoiding obstacles and selecting the most efficient routes to their destinations. This is highlighted by the sequence where robots seamlessly navigate through the dense urban landscape of the futuristic Chicago, indicating their sophisticated spatial awareness and adaptability. These robots analyze and interact with their surroundings in real-time, a concept that reflects the real-world technology found in autonomous vehicles and drones. Amazon Astro [11], a

Home Monitoring robot have same capabilities. It can analyze its environment and calculate optimal path to its destination, without colliding with obstacles. If the destination is not defined, then it roams the environment autonomously. One setback of Astro is that it is restricted to only 1 floor and are suitable only for home environments. If these capabilities are added to Astro, it can come close to the navigation shown in the movie by NS-5's.

D. Ethics and Trust

The movie delves into the ethical programming of robots and the trust placed in them by humans. The Three Laws of Robotics are central to the robots' ethical framework and are hardwired into their system. They are designed to ensure that they cannot harm humans. However, the film raises questions about this trust when a robot (Sonny) is suspected of murder. It talks about the ethical dilemmas and moral values that robots must possess. Another ethical consideration, "Trust", is also a key theme, explored through the skepticism of Detective Spooner and the contrasting trust shown by other characters. A similar ethical study done in 2016 by Robinette [12] in the field of HRI revolves around the question, "In an emergency situation, should you trust a robot?" In this study, an emergency is simulated in the building, and the robot is asked to escort the person out. Sometimes, the robot makes a wrong decision and goes towards a dark room blocked by furniture. Few people trusted the robot and went with him, and some chose the correct way out of the building. This pattern replicates to one shown in the movie by Spooner and other people. Regarding ethics, the European Union's guidelines on trustworthy AI [13] resonate with the film's emphasis on ethical programming, stressing the need for AI systems, including robots, to be transparent, fair, and accountable. According to the guidelines, trustworthy AI should be lawful, ethical, and robust.

IV. CURRENT POINT OF VIEW

A. Design and Hardware of Robot

The design of NS-5 robots is evidence to a future where robotics seamlessly merges with advanced engineering, sophisticated hardware, and sleek aesthetics. The agility and physical capabilities of the NS-5s are highlighted in several scenes, offering a window into their mechanical ability and design intricacies. In one notable incident, Sonny demonstrates extraordinary strength and speed by throwing Detective Spooner against a wall [26:01]. This scene showcases Sonny's raw power, precision, and control. The technology enabling such feats could be analogous to advanced actuator systems used in robots like Boston Dynamics' Atlas [14], which combines hydraulic and electric actuators for strength and smooth movements. Another remarkable aspect of the Sonny's design is his mobility and flexibility, demonstrated when Sonny effortlessly scales a wall and jumps a great distance [21:10]. This level of dynamic movement suggests an intricate balance system, utilizing gyroscopes and accelerometers, like those employed in Atlas [14] for stabilization and orientation. The ease and fluidity of Sonny's movement in these scenes indicate a highly sophisticated proprioceptive system that allows the robot to understand its body's position and movement in space, a feature being actively researched in contemporary robotics.

NS-5 robots (and autonomous cars) navigate through crowded city streets at the start of the movie shows their capability in autonomous navigation. This emulates technologies used in self-driving vehicles [15], which combine LIDAR, radar, GPS, and machine vision for environment mapping and decision-making. The AI-driven situational awareness and decision-making are similar to what's being developed for autonomous drones and vehicles. Tesla [16] one of the biggest companies that are working on autonomous self-driving vehicle are using the same technology.

The minimalistic design of the NS-5s, devoid of conspicuous sensors, implies a highly advanced integration of sensing technologies. This could be parallel to the use of embedded/tactile sensors and machine vision systems that are under development today, enabling robots to perceive their environment in a more human-like manner. One such real life example is Intel RealSense [17], which gives real-time depth perception capabilities to the robot.

B. Artificial Intelligence and Ethical Implications

The robots showcase an advanced level of artificial intelligence, encapsulating both autonomous decision-making and emotional intelligence. This is portrayed by the central AI system, VIKI, and the NS-5 robots. VIKI represents a level of intelligence that surpasses simple task execution, delving into moral and ethical decision-making, which is demonstrated when VIKI justifies her actions by stating, "My logic is undeniable" [1:31:30], reflecting a complex, self-rationalizing AI system. The state-of-the-art robots like Pepper [18] and NAO [19], developed by Softbank Robotics shows sign of emotional intelligence that humans possess. They can read emotional cues from the humans interacting with them and adapt according to the conversation. Although, their emotional intelligence is not as par as shown in the movie, however it is the step in the right direction. Robots like Boston Dynamics Spot [20] is now being commercialized and being sold to police and other military services as mentioned in recent video of youtuber Cleo Abram [21]. These robots show advanced decision-making in unfamiliar environment resonating with NS-5's in the open world.

The film also portrays AI's ability to process and react to human emotions, an aspect of AI that aligns with current research in affective computing. The NS-5 demonstrate an understanding of human emotions, similar to the robots being developed in projects like MIT's Kismet [4], which are designed to recognize and respond to emotional cues. This capability is highlighted when Sonny, expresses fear and confusion, saying, "I did not murder Dr. Lanning" [29:15]. Moreover, the film's portrayal of AI encompasses advanced natural language processing (NLP), evident in the robots' ability to communicate effectively with humans. The

NS-5s' communicative abilities reflect the progress seen in NLP technologies, where AI systems like Google's BERT [22] and OpenAI's GPT [23] models have shown remarkable proficiency in understanding and generating human language. This is not only limited to verbal communication but extends to non-verbal cues, as seen in several scenes where the NS-5s use body language and facial expressions to communicate, a feature that is being actively explored in humanoid robotics today.

From a current perspective, the AI in the film also raises questions about the ethical implications of advanced AI systems. VIKI's decision to override the Three Laws of Robotics to "protect" humanity points to the ethical dilemmas that arise with autonomous AI in real world scenario. This echo current debates in AI ethics, focusing on the need for ethical guidelines and control mechanisms to ensure AI systems act in the best interests of humanity.

V. Discussion

The world depicted in “I, Robot” is one where humanoid robot-like NS-5 can perform an array of activities, exhibiting advanced capabilities in strength, agility, decision-making, and emotional intelligence. Even though we are behind the technology shown in the movie, the rapid advancements in AI, robotics, and machine learning suggest that we are reaching that in our near future. Developments in robots like Boston Dynamics' Atlas [13], Softbank’s Pepper/NAO [17-18] and AI platforms like OpenAI's GPT models indicate that more advance robots are in the pipeline. However, the process is going to be a gradual evolution over an uncertain period rather than a sudden leap as demonstrated in the film. The acceptance of robots like Sonny, which exhibit human-like characteristics, brings into play the concept of the uncanny valley [24] – the discomfort people experience when robotic and CGI characters appear almost, but not exactly, like real humans. Overcoming this drawback is crucial for widespread acceptance of humanoid robots. The film addresses this by presenting robots with a balance of human-like traits and mechanical elements, making them more palatable to the audience.

A pivotal aspect of the movie is the ethical dilemma presented by VIKI, that decides to violate the Three Laws of Robotics for what it perceives as the greater good of humanity. This raises critical questions about programming ethics into AI systems and the potential for AI to make life-saving decisions. It reflects current concerns about how AI might handle ethical dilemmas, especially in scenarios like autonomous vehicles making split-second decisions in an accident. The film provocatively asks: should AI have the autonomy to override its programming for perceived moral reasons? Other than that, the emotional capabilities displayed by robots in the film, particularly Sonny, indicate a future where AI can understand and perhaps even feel emotions. While current AI can simulate emotional responses and recognize human emotions to some extent, the depth of emotional understanding portrayed in the film is beyond our current capabilities. The common expectation from robots today is centered around efficiency and assistance, but as AI continues to evolve, there might be a shift towards expecting more emotionally intelligent interactions. The development of such advanced AI and robotics is currently led by private companies, like Google, Boston Dynamics, and OpenAI, and government-funded initiatives. The governance and regulation of AI and robotics are evolving fields. Laws and ethical guidelines governing AI behavior, like Asimov's Three Laws of Robotics, are being discussed in academic and policy-making circles, but concrete international frameworks are still in the developmental stage. The emergence of AI models like ChatGPT represents a significant leap in natural language processing and interaction. This revolution in AI demonstrates the rapid pace of advancement in the field and hints at the potential future capabilities of AI systems in understanding and generating human language**.** The social response to advanced AI and robotics is mixed, with excitement about potential benefits countered by fears of loss of control, as often depicted in science fiction. It is crucial for society to engage in informed discussions about the development and integration of AI, focusing on ethical, safety, and societal impacts.

"I, Robot" ultimately delivers a message about the complex relationship between humans and AI. It warns us about the potential risks of overly advanced AI and highlights the positive aspects of AI-human interactions. In conclusion, while the vision of the movie is still within the realm of science fiction, it provides a valuable lens, through which, we can examine our current trajectory in AI and robotics. The film's portrayal of advanced AI raises important questions about ethics, emotional intelligence, societal acceptance, and governance, emphasizing the need for a balanced approach to the development and integration of these technologies in society.

References

[1] “Isaac Asimov’s ‘Three Laws of Robotics’”, https://webhome.auburn.edu/~vestmon/robotics.html

[2] “I, Robot wikipedia” https://en.wikipedia.org/wiki/I,_Robot_(film)

[3] “I, Robot transcript” http://www.script-o-rama.com/movie_scripts/i/i-robot-script-transcript.html

[4] Cynthia Breazeal (2003), “Emotion and sociable humanoid robots” in International Journal of Human-Computer Studies, Volume 59, Issues 1–2, Pages 119-155, ISSN 1071-5819

[5] Kouravanas, Nikolaos, and Andreas Pavlopoulos. "Social Robots: The case of Robot Sophia." *Homo Virtualis* 5.1 (2022): 136-165.

[6] C. Breazeal and B. Scassellati (2002). "Robots that imitate humans," Trends in Cognitive Science, 6, pp. 481-487.

[7] “IBM watsonx”, https://www.ibm.com/watsonx

[8] Al Moubayed, Samer, et al. "Furhat: a back-projected human-like robot head for multiparty human-machine interaction." Cognitive Behavioural Systems: COST 2102 International Training School, Dresden, Germany, February 21-26, 2011, Revised Selected Papers. Springer Berlin Heidelberg, 2012.

[9] Holcomb, Sean D., et al. "Overview on deepmind and its alphago zero ai." Proceedings of the 2018 international conference on big data and education. 2018.

[10] “AlphaGo Zero: Discovering new knowledge”, https://www.youtube.com/watch?v=WXHFqTvfFSw

[11] “Amazon Astro”, https://www.aboutamazon.com/news/devices/meet-astro-a-home-robot-unlike-any-other

[12] Paul Robinette, Wenchen Li, Robert Allen, Ayanna M. Howard and Alan R. Wagner, “Overtrust of Robots in Emergency Evacuation Scenarios,” (2016 ACM/IEEE International Conference on Human-Robot Interaction) (HRI 2016).

[13] “European Union Ethics guidelines for trustworthy AI”, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[14] Feng, Siyuan, et al. "Optimization based full body control for the atlas

[15] robot." *2014 IEEE-RAS International Conference on Humanoid Robots*. IEEE, 2014.

[16] Badue, Claudine, et al. "Self-driving cars: A survey." *Expert Systems with Applications* 165 (2021): 113816.

[17] Metz, Cade, and Neal E. Boudette. "Inside Tesla as Elon Musk Pushed an Unflinching Vision for Self-Driving Cars." International New York Times (2021)

[18] “Intel Real Sense”, https://www.intelrealsense.com/

[19] “Pepper Robot”, https://us.softbankrobotics.com/pepper

[20] “NAO Robot”, https://us.softbankrobotics.com/nao

[21] “Boston Dynamics Spot”, https://bostondynamics.com/products/spot/

[22] “Cleo Abram Youtube video”, https://www.youtube.com/watch?v=DecYIeB44aE

[23] “Google Bert explained”, https://www.marketingaiinstitute.com/blog/bert-google

[24] “ChatGPT”, https://chat.openai.com/

[25] Mori, Masahiro, Karl F. MacDorman, and Norri Kageki. "The uncanny valley [from the field]." *IEEE Robotics & automation magazine* 19.2 (2012): 98-100.1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].