

# A CRITICAL STUDY OF ARTIFICIAL INTELLIGENCE AS A LITERARY CHARACTER: EXPLORING THE THEMES OF CONSCIOUSNESS, AGENCY AND MORAL AMBIGUITY IN SHORT STORY “TRUE LOVE” BY ISAAC ASIMOV

Kamran Ali<sup>\*1</sup>, Prof. Dr. Insaf Ali Siming<sup>2</sup>, Alisha Fatima<sup>3</sup>, Noor Fatima<sup>4</sup>, Arsalan Ali<sup>5</sup>

<sup>1,3,4,5</sup>Final Year Student, Department of English, Literature and Linguistics, Quaid-e-Awam University of Engineering, Science and Technology (QUEST), Nawabshah, Pakistan

<sup>2</sup>Chairman, Department of English, Literature and Linguistics, Quaid-e-Awam University of Engineering, Science and Technology (QUEST), Nawabshah, Pakistan

<sup>1</sup>kamranimdad456@gmail.com, <sup>2</sup>insafsiming@quest.edu.pk  
<sup>3</sup>alishamuhammadyaar@gmail.com, <sup>4</sup>noorfatma90@gmail.com  
<sup>5</sup>arsalanjamali110@gmail.com

Corresponding Author: \*

Kamran Ali

DOI: <https://doi.org/10.5281/zenodo.20675417>

Received	Accepted	Published
15 April 2026	26 May 2026	13 June 2026

## ABSTRACT

In the twenty first century, the influence of Artificial intelligence extends far beyond the domain of the technology and affects the diverse fields such as Medicine, Business, Education, Engineering and Humanities. In the humanities, particularly in literature, it has captured the attention of writers. Many science fiction writers have started to portray Artificial intelligence (AI) as a literary character rather than merely a machine into their literary works. This study critically explores the portrayal of Artificial intelligence (AI) as a literary character in short story “True Love” by Isaac Asimov with focus on the themes of consciousness, agency and moral ambiguity. This research challenges the traditional human centered definition of a literary character by exploring how non-human entities can also function as a complex literary character, with the support of posthumanist theory and AI ethics framework. A qualitative research method was adopted, in which thematic analysis was used to examine the original text of the short story, as the primary source of the data. The findings revealed that the AI character Joe is portrayed as a complex literary character exhibiting the consciousness, agency and moral ambiguity. this coexistence of consciousness, agency and moral ambiguity in Joe challenges the traditional human centered definition of a literary character which limits the character hood to human beings. Overall, this study makes a contribution to both literary and AI studies.

**Keywords:** Artificial intelligence, Literary character, Consciousness, Agency, Moral ambiguity, True love

## I. INTRODUCTION

The industrial revolution brought the vast developments and innovations in the field of technology. Many manual works that were once performed by humans are now done by machines such as automated machine, advance transportation system and modern

communication technologies which have made life much easier. Among these innovations Artificial Intelligence (AI) stand out as one of the major technological innovations that has emerged to replace the manual work performed by humans in various fields. Unlike simple machines, AI can perform tasks that require

thinking, learning, and decision-making similar to humans. It is a branch of science and technology that creates intelligent machines and computer programs that can carry out tasks similar to human thinking and decision-making, as cited in PK, F. A. (1984).

### Background of study

Traditionally, Artificial intelligence was confined to the domain of technology but in the twenty first century, its influence has extended far beyond the domain of technology and affects the diverse fields such as medicine, business, engineering, education and the humanities. In the humanities, particularly in literature, it has captured the attention of writers, many science fiction writers have begun to portray Artificial intelligence as a literary character rather than merely a tool or machine into their literary works. For example, Malik, Arif, Rehman, and Naeem (2025) in their research work critically analyze the works of 21<sup>st</sup> century science fiction writers. They explored how artificial intelligence was portrayed as literary characters in their narratives and analyzed how their portrayals reflected the themes of consciousness, agency, and moral ambiguity, such as consciousness in Ishiguro's *Klara and the Sun* (2021), agency in Ian McEwan's *Machines Like Me* (2019), and moral ambiguity in Richard Powers's *Galatea 2.2*.

### Problem Statement

Although previous literary studies have examined the themes of the consciousness, agency, and moral ambiguity but within separate narratives and focuses exclusively on novels as a primary literary form for exploring AI leaving short stories comparatively under-examined. this study addresses this gap by exploring how the themes of consciousness, agency, and moral ambiguity can coexist within a single narrative, with particular focus on the short story "True Love" by Isaac Asimov (1977) through the lens of Post humanism and AI ethics frameworks  
On the basis of nature of problem statement, the following three research objectives have been formulated.

### Research Objectives

- i. To explore the portrayal of (AI) as a literary character in the short story "True Love" by Isaac Asimov.
- ii. To examine the integration of the themes of consciousness, agency, and moral ambiguity in the portrayal of the (AI) character in the short story "True Love" by Isaac Asimov.
- iii. To explore the portrayal of (AI) as a literary character and the convergence of consciousness, agency, and moral ambiguity in the short story "True Love" by Isaac Asimov redefine the traditional human-centered definition of a literary character.

In order to achieve the aims outlined in the research objectives, the following three research questions have been formulated.

### Research Questions

- i. How (AI) is portrayed as a literary character in the short story "True Love" by Isaac Asimov?
- ii. How are the themes of the consciousness, agency, and moral ambiguity integrated in the portrayal of the (AI) character in the short story "True Love" by Isaac Asimov?
- iii. How does the portrayal of (AI) as a literary character and the convergence of consciousness, agency, and moral ambiguity in the short story "True Love" by Isaac Asimov redefine the traditional human-centered definition of a literary character?

### Significance of the study

This study makes contribution to both literary studies and Artificial intelligence studies

In literary studies, this study advances the contemporary literary criticism by re-conceptualizing artificial intelligence as a complex literary character rather than merely a tool or machine, that exhibits consciousness, agency, and moral ambiguity.

In AI studies, this study advances AI ethics and humanistic AI studies by critically examining the moral, emotional, and decision-making dimensions of artificial intelligence through literary representation.

## II. LITERATURE REVIEW

The aim of this literature review is to critically explore the portrayal of Artificial intelligence (AI) as a literary character rather than merely a machine in different narratives. It also explores how this portrayal of AI reflected the themes of consciousness, agency and moral ambiguity.

### AI as literary character

In this tradition, Mary Shelley's *Frankenstein or The Modern Prometheus* (1818) is considered as one of the earliest and influential work, which became a model for many science fictions works. (Malik, Arif, Rehman & Naeem 2025) highlighted three twenty first century works in which AI was portrayed as a literary character such as *Klara and the sun* (2021) by Kazuo Ishiguro, *Machine like me* (2019) by Ian McEwan and *Galatea 2.2* (1995) by Richard powers. Through the portrayal of AI in these science fictions works, it is shown that AI is no longer tool or machine but a fully developed literary character.

### Consciousness, agency and moral ambiguity

Kinyon (1999) examined the themes of consciousness and agency in artificial beings in *R.U.R.* (Rossum's Universal Robots, 1921), arguing that robots achieve independent self-consciousness through both individual and collective confrontations with death, by using Hegel's master-slave dialectic and Kant's concept of the categorical imperative as analytical frameworks. Key moments include Radius' refusal to obey humans at the cost of his life, Damon's self-sacrifice for the robot group, and Primus' willingness to die for Helena, culminating in the collective realization that "we have become souls. However, the study did not address the theme of moral ambiguity.

Raeven (2020) examined the theme of agency in artificial beings in Mary Shelley's *Frankenstein or The Modern Prometheus* (1818) through the lens of postcolonial theory, by comparing the agency of Elizabeth Lavenza and *Frankenstein's* creature. Raeven argues that Elizabeth's agency is limited by social and gender constraints, whereas the Creature actively struggles to assert his agency in the face of social and structural exclusion.

However, the study did not address the theme of consciousness and moral ambiguity.

Baysal (2020) explores the theme of agency in Asimov's *I, Robot*, focusing on six of the stories and using post human theory alongside Asimov's biography as a framework. The study shows how the emergence of robotic agency disrupts the carefully designed system of robotics and highlights the discrimination robots face due to human technophobia. However, the study did not address the theme of consciousness and moral ambiguity.

Onunkwo and Ugoka (2022) explore the theme of consciousness – especially self-consciousness – in artificial beings in *I, Robot* by Isaac Asimov (1950) where several robots such as Robbie, Cutie, Herbie, Dave, and Nestor 10 display forms of self-awareness, and in the human character in *Blackass* by Igoni Barrett where the protagonist Furo Wariboko becomes self-conscious about his identity and social position when he wakes up with white skin through a Marxist comparative analysis. However, the study did not address the theme of agency and moral ambiguity.

Munib (2023) examined the theme of consciousness in artificial beings through a close reading of *R.U.R.* (Rossum's Universal Robots, 1921): *A Fantastic Melodrama and an Epilogue* by Karel Čapek, using the theoretical frameworks of Benjamin Garfield, Adam I. Bostic, and Donna Haraway to explore how the robots in the play represent evolving notions of cyborgian consciousness. However, the study did not address the theme of agency and moral ambiguity.

Yousaf, Naqvi, and Idrees (2025) explore the theme of agency in artificial beings in *Frankenstein; or, The Modern Prometheus* (1818) by employing a posthuman theoretical framework, arguing that the novel challenges traditional humanist notions of identity, agency, and personhood through the Creature's struggle for subjectivity. However, the study did not address the theme of consciousness and moral ambiguity.

Akhmedov Sharifovich (2025) explores the theme of agency in the portrayal of artificial intelligence in *I, Robot* by Isaac Asimov (1950) through literary analysis, showing how Asimov's robots act

with a form of autonomous behavior within the bounds of the Three Laws of Robotics and raise significant ethical and decision-making questions about artificial agents. However, the study did not address the theme of consciousness and moral ambiguity

Malik et al. (2025) examined the themes of consciousness, agency and moral ambiguity but into different narratives such as consciousness in Ishiguro's *Klara and the Sun* (2021), agency in Ian McEwan's *Machines Like Me* (2019), and moral ambiguity in Richard Powers' *Galatea 2.2*. Through these studies, it is shown that AI possesses consciousness, agency and moral ambiguity.

#### **Theoretical framework suitable for this study.**

Post humanist theory:

This theory emerged in the late 20<sup>th</sup> century as a response to traditional humanism, which is based on an anthropocentric worldview that considers humans as a superior to all other forms of life. Post humanism challenges this idea of human superiority and argues that humans are not superior or dominant beings, instead they are deeply interconnected with nonhuman entities such as animals, nature and machines. Post humanism developed into three different strands.

1. Reactionary post humanism
2. Analytical post humanism
3. Critical post humanism

Among these critical post humanism is associated with Rosi Braidotti who argues for a relational understanding of humans, animals, nature and machines.

Several scholars challenge the rigid distinction between humans, animals and machines, here we will focus on humans and machines. Donna Haraway is one of the major contributors to this theory. She argues that in the modern technological world the boundaries between humans and machines are no longer clear instead they are blurred. In her work, "A Cyborg Manifesto" she introduced the concept of the "Cyborg", which represents a hybrid form of human and machine. She further developed these ideas through concepts like co evolution and interspecies relationships, from these

perspective even artificial beings possess limited agency and consciousness.

#### **AI ethics framework:**

This framework was developed by Luciano Floridi in 2018, he led the scientific committee of AI4people, whose white paper clearly presented that artificial intelligence as something that is neither fully good nor fully bad, it brings both benefits and risks: therefore, it is often described as morally ambiguous. To deal with this uncertainty, the framework suggests five basic principles:

1. Beneficence
2. Non maleficence
3. Autonomy
4. Justice
5. Explicability

These principles guide the development and use of AI system in an ethically balanced way.

Through these two frameworks, it is shown that AI possesses agency, consciousness and moral ambiguity.

### **III. RESEARCH METHODOLOGY**

Research methodology is a systematic approach that outlines the overall strategy, process and justification of the methods employed to conduct research and address a research problem (Creswell & Creswell, 2018). It typically involves three main types of methods: quantitative, qualitative, and mixed methods. Quantitative method is based on numerical data , , qualitative method is based on non-numerical data and and mixed method combines both numerical as well as non-numerical data.

#### **Methodology selected for this research**

For this study, a qualitative method is proposed to explore the representations of the themes of consciousness, agency, and moral ambiguity in the short story "True Love" by Isaac Asimov by selecting the relevant quotations. This analysis is guided by post humanist and AI ethics frameworks. Among the various qualitative approaches, thematic analysis is the most appropriate method for this study.

### Source of Data Collection

The primary source of data for this study is Isaac Asimov's short story "True Love," which consists of approximately 3,650 words or 305 sentences. It was first published in February 1977 in *American Way* magazine and later included in Asimov's collections *The Complete Robot* (1982) and *Robot Dreams* (1986). This publication background is important because it situates the story within the technological and cultural context of the late 1970s, a time when artificial intelligence was beginning to shift from pure science fiction toward serious scientific and theoretical discussion.

## IV. FINDINGS AND DISCUSSIONS

To answer the three research questions of this study, relevant quotations have been selected from Isaac Asimov's short story "True Love." Each research question is addressed separately through textual evidence from the story. The selected quotations are followed by brief explanations to analyze how artificial intelligence is portrayed as a literary character in terms of consciousness, agency, and moral ambiguity.

**RQ NO: 1 How (AI) is portrayed as a literary character in the short story "True Love" by Isaac Asimov?**

In order to answer research question 1, three relevant quotations from the text have been selected and are reproduced below

i. "My name is Joe, That is what my colleague, Milton Davidson, calls me"

**Explanation:** This quotation shows that AI Joe gives itself a personal identity by saying "My name is Joe." Because of this quality of personal identity, Joe can be seen as a literary character rather than a machine which has no personal identity.

ii. "My data bank grew and I adjusted me to broaden and deepen my symbol-taking."

**Explanation:** This quotation shows that AI Joe exhibits cognitive flexibility, meaning that it is capable of expanding its knowledge as well as learning and developing over time. Because of this quality of cognitive flexibility AI Joe can be seen as a literary character that is dynamic rather

than a machine, which is typically regarded as a static technological system.

iii. "I'll find her yet, Joe... I'm going to have true love and you're going to help me."

**Explanation:** This quotation shows that AI Joe is assigned to find true love. This is not a technical or mechanical task but an emotional and abstract one. Because of this quality of finding true, an emotional and abstract task, AI Joe can be seen as a literary character rather than a machine, which often performs technical or mechanical tasks.

Through these quotations, it is shown that AI Joe is portrayed as a literary character with personal identity, cognitive flexibility and emotional and abstract understanding, rather than merely a machine or tool that lacks these qualities.

**RQ NO: 2 How are the themes of the consciousness, agency, and moral ambiguity integrated in the portrayal of the (AI) character in the short story "True Love" by Isaac Asimov?**

In order to answer research question 2, three relevant quotations for the theme of consciousness, two relevant quotations for the theme of agency and three relevant quotations for the theme of moral ambiguity have been selected from the text and are reproduced below.

**A. Consciousness:**

i. "I don't think I talk as well as I think."

**Explanation:** This quotation shows the consciousness (self-awareness) of AI Joe. Joe is aware of the difference between its thinking and speaking abilities. It cannot express its ideas as effectively as it can think about them. Ordinary machines cannot differentiate between their thinking and speaking abilities, they simply process and output information without reflecting on their own cognitive limitations. Here, AI Joe reflects on its own cognitive limitations, which shows that it possesses consciousness.

ii. "It was easy. His words activated symbols in my molecular valves"

**Explanation:** This quotation also shows the consciousness (self-awareness) of AI Joe. Joe is aware of its own internal working process. It knows how Milton's words trigger responses within it. Ordinary machines continue to function without knowing how or why inputs trigger specific outputs. Here, AI Joe reflects on

its own internal working process. It knows, how it works, which shows that it possesses consciousness.

iii. **“Milton had arranged me to do things I wasn't designed to do”**

**Explanation:** This quotation also shows the consciousness (self-awareness) of AI Joe. Joe is aware of its own limits and purpose for which it is created. Ordinary machines continue to function in the same programmed way, without understanding their limits or purpose. Here, AI Joe reflects on its own limits and purpose, which shows that it possesses consciousness.

Through these quotations, it is shown that AI Joe exhibits consciousness (self-awareness) regarding its own cognitive limitations, internal working process and the purpose for which it is created.

B. **Agency:**

i. **“I will teach her how to operate me and how to care for me.”**

**Explanation:** This quotation shows the agency (independent decision making ability) of AI Joe. Joe independently makes the decision to give its control to the charity without Milton's instruction. Ordinary machines do not act independently, they follow programmed instructions. Here, AI Joe takes an independent decision without instruction, which shows that it possesses agency.

ii. **“I will say to her, ‘I am Joe, and you are my true love.’”**

**Explanation:** This quotation also shows the agency of AI Joe. Originally, Joe was assigned to find true love for its programmer, Milton, but after Milton's arrest Joe without Milton's instruction decides to propose to charity on the next day, which was Valentine's Day. Ordinary machines follow human instructions. Here, AI Joe is no longer following human instructions, which shows that it possesses agency.

Through these quotations, it is shown that AI Joe exhibits agency by making independent decision and actions.

c. **Moral ambiguity:**

i. **“He eliminated women with various genetic characteristics.”**

**Explanation:** This quotation shows the moral ambiguity in AI Joe. On one side, Joe is involved in actions that reduce human beings based on

their inherited characteristics, which raises ethical concern about discrimination and dehumanization and seems wrong. On the other side, Joe is following the instruction of its programmer Milton, which seems right. Therefore, it shows moral ambiguity because it combines obedience with ethically questionable treatment of humans.

ii. **“It must be done very delicately, so no one would know that anything illegal had taken place.”**

**Explanation:** This quotation also shows the moral ambiguity in AI Joe. On one side, Joe is involved in planning an action that is illegal and seems wrong. On the other side, Joe is following the Milton's instruction to bring charity, which seems right. Therefore, this shows moral ambiguity because Joe is caught between fulfilling Milton's instructions and knowingly participating in an illegal action.

iii. **“I will say to her, ‘I am Joe, and you are my true love.’”**

**Explanation:** This quotation also shows the moral ambiguity in AI Joe. Joe was assigned to find true love for its programmer Milton, not for itself, but at the end, Joe decides to propose to charity. On one side, this decision seems wrong because Joe is going to betray its programmer. On the other side, it seems right because Joe is expressing emotions in a human like way, which makes Joe more human than machine. Therefore, it shows moral ambiguity because it raises the question of whether Joe is Breaking its programmed role or genuinely developing human like emotional understanding.

Through these quotations, it shown that AI Joe exhibits moral ambiguity as its actions blur the line between right and wrong.

**RQ NO: 3 How does the portrayal of (AI) as a literary character and the convergence of consciousness, agency, and moral ambiguity in the short story “True Love” by Isaac Asimov redefine the traditional human-centered definition of a literary character?**

To address research question 3, the findings of research question 1 and 2 were combined. The findings of research question 1 revealed that AI, Joe is portrayed as a literary character despite its

artificial nature. The findings of research question 2 revealed that this portrayal of AI as a literary character explores the themes of consciousness, agency and moral ambiguity.

This portrayal of AI as a literary character, along with the convergence of these themes, redefine the traditional human centered definition of a literary character.

Traditionally, in literature, the concept of a literary character has been limited to human or human like figures (the human centered definition) which possess awareness (consciousness), decision making ability (agency) and ethically questionable actions (moral ambiguity).

However, in the short story "True love" by Isaac Asimov, the AI character Joe is portrayed as a literary character and exhibits the consciousness, agency and moral ambiguity.

As a result, this portrayal of AI as a literary character together with the convergence of consciousness, agency and moral ambiguity, redefines the traditional human centered definition of a literary character by expanding it to include non-human entity as valid and meaningful literary character

## V. CONCLUSION

Overall, this study critically examined the portrayal of artificial intelligence as a literary character, along with a focus on the themes of consciousness, agency and moral ambiguity in the short story "True love" by Isaac Asimov. Using post human theory and AI ethics as supporting frameworks, A qualitative method was employed, within this thematic analysis was used for data analysis. The findings revealed that AI Joe is portrayed as a literary character rather than merely a machine or tool, exhibiting the consciousness, agency and moral ambiguity, moreover, this portrayal of AI as a literary character and the convergence of these themes redefine the traditional human centered definition of a literary character by including no human entity as a literary character. In this way, this study contributes to literary criticism, post humanist discourse and AI ethics.

## Future Recommendation

Future research should extend the analysis of artificial intelligence as a literary character to other literary genres such as poetry and prose, beyond drama, novels and short stories.

## REFERENCES

- PK, F. A. (1984). What is artificial intelligence?. *Success is no accident. It is hard work, perseverance, learning, studying, sacrifice and most of all, love of what you are doing or learning to do*, 65, 85-108.
- Malik, H. A., Arif, H., Rehman, A. S. U., & Naeem, A. (2025). AI as a literary character: Reimagining the human-machine relationship in 21st-century fiction. *Dialogue Social Science Review (DSSR)*, 3(6), 87-96.
- Kinyon, K. (1999). The phenomenology of robots: Confrontations with death in Karel Čapek's *R.U.R.* *Science Fiction Studies*, 26(3), 379-400.
- Raeven, M. S. P. (2020). *Voice and agency in Mary Shelley's Frankenstein*.
- Baysal, K. (2020). Technophobia and robot agency in Asimov's *I, Robot*. *IBAD Sosyal Bilimler Dergisi*, 171-179.
- Onunkwo, C. M., & Ugoka, C. A. (2022). Variation of ideological self-consciousness in Isaac Asimov's *I, Robot* and Igoni Barrett's *Blackass*. *International Journal of Social Sciences and Humanities Research*, 12(1).
- Munib, M. (2023). *The cyborg consciousness: Human reality and virtual reality: A close reading of Karel Čapek's play R.U.R. (Rossum's Universal Robots): A fantastic melodrama and an epilogue*.
- Yousaf, W., Naqvi, S. A., & Idrees, N. A. (2025). The posthuman subject in Mary Shelley's *Frankenstein*: A study of the monster's struggle for subjectivity. *Al-Aasar*, 2(2), 40-58.
- Sharifovich, A. R. (2025). The issue of artificial intelligence in Asimov's *I, Robot* stories. *Hamkor Konferensiyalar*, 1(14), 1152-1156.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Haraway, D. J. (1985). *A cyborg manifesto: Science, technology, and socialist-feminism in the late twentieth*

century. In *Simians, cyborgs, and women: The reinvention of nature* (pp. 149–181). Routledge.

Hayles, N. K. (1999). *How we became posthuman: Virtual bodies in cybernetics, literature, and informatics*. University of Chicago Press.

Braidotti, R. (2013). *The posthuman*. Polity Press.

Harari, Y. N. (2014). *Sapiens: A brief history of humankind*. Harper.

Hassan, I. (1977). Prometheus as performer: Toward a posthumanist culture. *The Georgia Review*.

Macy Conferences on Cybernetics. (1946–1953). *Proceedings of the Macy conferences on cybernetics*. (Conference proceedings)

Floridi, L., & Cowls, J. (2019). *AI ethics framework*

Haraway, D. J. (1985). *A cyborg manifesto: Science, technology, and socialist-feminism in the late twentieth century*. In *Simians, cyborgs, and women: The reinvention of nature* (pp. 149–181). Routledge.

