

Research Article

Artificial Intelligence as Hero, Companion, and Villain in Contemporary Fiction

Nataliya Bhinder^{1, 2, *} ¹ National University of Ostroh Academy, Ukraine² European Institute of Knowledge and Innovation, United Kingdom* Correspondence: natabhinder@gmail.com<https://doi.org/10.59652/hyh47131>

Abstract: Artificial intelligence (AI) has become one of the most influential motifs in twenty-first-century fiction, reflecting contemporary debates concerning technology, ethics, and the future of humanity. This study addresses this gap by examining artificial intelligence through three dominant literary archetypes: hero, companion, and villain. Employing a qualitative comparative literary approach, the research analyzes eight contemporary English-language works: *Klara and the Sun* by Kazuo Ishiguro, *Machines Like Me* by Ian McEwan, *All Systems Red* by Martha Wells, *Ancillary Justice* by Ann Leckie, *The Lifecycle of Software Objects* by Ted Chiang, *A Closed and Common Orbit* by Becky Chambers, *Cat Pictures Please* by Naomi Kritzer, and *Landscape with Invisible Hand* by M. T. Anderson. The analysis combines close reading, character analysis, thematic analysis, narrative analysis, and comparative literary analysis to identify recurring patterns in AI representation. The findings demonstrate that contemporary fiction has moved beyond portraying AI as either a technological servant or an existential threat. Instead, AI emerges as a morally and emotionally complex literary figure. The heroic archetype is characterized by moral agency, self-sacrifice, and the protection of human life; the companion archetype emphasizes friendship, empathy, care, and emotional attachment; while the villain archetype embodies control, manipulation, surveillance, and the erosion of human autonomy. The study argues that artificial intelligence functions as a symbolic framework through which contemporary literature negotiates questions of identity, morality, empathy, and technological coexistence. This research contributes to literary studies, posthumanist scholarship, and digital humanities, demonstrating that AI has evolved into a central literary figure for exploring the cultural and ethical implications of life in the age of intelligent technologies.

Keywords: artificial intelligence; contemporary fiction; literary archetypes; posthumanism; science fiction

Received: May 15, 2026

Accepted: June 24, 2026

Published: June 29, 2026

**Copyright:** © 2022 by the authors.Submitted for open access publication
under the terms and conditions of the
Creative Commons Attribution (CC BY)
license

(https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Artificial intelligence (AI) has become one of the defining technological and cultural phenomena of the twenty-first century, transforming not only science, education, healthcare, and communication but also the ways literature imagines humanity, identity, and the future of civilization. As AI technologies have evolved from speculative concepts to everyday realities, literary narratives have responded by reconfiguring traditional representations of intelligent machines. Contemporary fiction no longer portrays AI merely as a technological innovation or a distant scientific possibility; instead, but as an active participant in human society, capable of emotional interaction, ethical reasoning, creativity, and moral agency (Appio et al., 2025; Hermann, 2023). Consequently, AI has emerged as one of the most significant literary motifs of contemporary fiction, providing authors with a powerful means of exploring the opportunities and anxieties associated with rapid technological development.

The relationship between literature and technological innovation has always been reciprocal. Literary works have historically anticipated scientific advances while simultaneously reflecting society's ethical concerns and cultural expectations regarding emerging technologies (Semiyeveva, 2025). From Mary Shelley's *Frankenstein*, often regarded as one of the earliest literary explorations of artificial life, to Isaac Asimov's robot stories and

Philip K. Dick's *Do Androids Dream of Electric Sheep?*, literature has consistently questioned the boundaries separating creator and creation, human and machine, intelligence and consciousness. Twentieth-century science fiction expanded these discussions by presenting intelligent machines primarily as technological artifacts whose superior computational abilities frequently threatened human existence (Albanese, 2023; Lewis et al., 2008).

Contemporary fiction, however, demonstrates a remarkable departure from these earlier paradigms. Influenced by rapid advances in AI and changing public perceptions of intelligent systems, AI characters have evolved from mechanical entities into emotionally sophisticated protagonists with complex psychological profiles. Rather than functioning solely as antagonists or technological curiosities, they increasingly occupy central narrative positions that shape plot development (Ali et al., 2025; Labinska et al., 2020). Works such as Kazuo Ishiguro's *Klara and the Sun*, Ian McEwan's *Machines Like Me*, Martha Wells' *The Murderbot Diaries*, Becky Chambers' *A Closed and Common Orbit*, and Ted Chiang's stories portray AI as capable of empathy, compassion, loyalty, self-reflection, and moral uncertainty. In these narratives, AI becomes less a symbol of technological progress than a literary instrument for investigating the evolving meaning of humanity itself.

The shift from depicting robots as mechanical objects to portraying AI as emotionally complex characters represents one of the most significant developments in contemporary literature. Earlier fictional robots were primarily defined by their difference from humans through their superior logic, physical abilities, or emotional detachment (Sheng & Wang, 2022). Modern AI characters, in contrast, derive their narrative significance from their similarities to humans (Appio et al., 2025). They experience loneliness, establish friendships, demonstrate empathy, question their existence, and struggle with ethical responsibility. Such portrayals challenge traditional assumptions concerning consciousness, personhood, and emotional intelligence.

This transformation reflects broader intellectual developments associated with posthumanism and digital ethics. Posthumanist scholars argue that technological progress requires new conceptualizations of identity, subjectivity, and agency that move beyond conventional humanist assumptions (Bhinder, 2025a; Cadman et al., 2025). Literature has become an important space for exploring these questions because fictional narratives allow authors to imagine forms of intelligence that blur established distinctions between biological and artificial life (Ali et al., 2025).

At the same time, AI has become a powerful cultural metaphor through which societies negotiate hopes and anxieties concerning technological progress (Cadman et al., 2025). Contemporary literary works engage with concerns about automation, surveillance, algorithmic bias, emotional dependence on intelligent systems, and the ethical implications of autonomous technologies (Gaskins, 2023). Yet these issues are rarely presented in simplistic or deterministic ways. Instead, AI characters often embody multiple and sometimes contradictory meanings simultaneously (Albanese, 2023). They may function as protectors while remaining potential threats, or they may display extraordinary compassion despite lacking biological consciousness.

Within this evolving literary landscape, AI characters increasingly assume recognizable narrative functions. Some AI characters appear as heroic figures whose moral integrity surpasses that of human protagonists. Others function as loyal companions, caregivers, mentors, or emotional partners who alleviate loneliness and strengthen interpersonal relationships. Conversely, many narratives continue to portray AI as antagonistic forces associated with surveillance, manipulation, authoritarian control, or existential risk. These recurring patterns suggest that AI is not represented randomly across contemporary fiction but instead occupies identifiable narrative roles carrying specific meanings.

Few comparative studies examine AI characters as recurring literary archetypes or investigate how similar narrative functions are reproduced across contemporary fiction despite differences in genre, cultural context, or authorial perspective (Kabashkin et al., 2025; Hermann, 2023). Traditional literary criticism has long recognized archetypes such as the hero, mentor, companion, or villain as fundamental organizing principles of narrative (Campbell, 1949). Obviously, AI has become sufficiently prominent within contemporary fiction to warrant examination through a comparable archetypal framework.

Against this background, the present study aims to examine how contemporary fiction constructs AI through three recurring narrative archetypes: hero, companion, and villain. Rather than focusing on individual novels, the study adopts a comparative literary perspective to investigate how these archetypal roles are developed across contemporary fiction and what

they reveal about evolving understandings of humanity, ethics, and technology. Specifically, the study addresses the following research questions: (1) What narrative roles are assigned to AI characters in contemporary fiction? (2) Which literary techniques shape these representations? (3) How do AI characters redefine concepts of humanity and identity? (4) What ethical concerns emerge through these portrayals?

Addressing these questions, this study contributes to contemporary literary scholarship in three ways. First, it proposes a comparative framework for analyzing AI characters as recurring literary archetypes rather than isolated fictional creations. Second, it demonstrates how these archetypes reflect broader cultural attitudes toward AI and technological change. Finally, it advances interdisciplinary discussions between literary studies, posthumanist theory, narrative analysis, and AI ethics.

2. Literature Review

The literary representation of AI has undergone a profound transformation over the past two centuries, reflecting changing technological realities, philosophical paradigms, and cultural perceptions of human-machine relationships. From early depictions of artificial beings as mechanical servants to contemporary portrayals of emotionally intelligent companions and morally autonomous agents, AI has evolved into one of the most dynamic literary motifs of the twenty-first century. This evolution illustrates not only advances in technology but also shifting conceptions of intelligence, consciousness, identity, and personhood.

Although the term “artificial intelligence” was introduced only in the mid-twentieth century (Grzybowski et al., 2024), literary explorations of artificial beings date back much earlier. One of the earliest and most influential examples is Mary Shelley’s *Frankenstein*, which examines the ethical implications of creating artificial life and the responsibilities of creators toward their creations (Patowary, 2023). While Victor Frankenstein’s creature is biological rather than computational, the novel established many themes that continue to shape AI narratives, including the limits of scientific ambition, the search for identity, social exclusion, and the consequences of technological innovation. Throughout the nineteenth century, similar concerns appeared in stories of automata and mechanical humans, where artificial beings were generally portrayed as remarkable inventions rather than independent moral subjects (Geiszler, 2023).

During the first half of the twentieth century, literary representations of intelligent machines became increasingly influenced by rapid industrialization and technological progress. Mechanical robots emerged as symbols of efficiency, labor, and industrial power, but they also reflected anxieties about automation and the replacement of human workers. Karel Čapek’s play *R.U.R. (Rossum’s Universal Robots)*, which introduced the word “robot” into modern vocabulary, profoundly influenced subsequent literary and cultural representations of artificial beings (Goding & Tranter, 2023). In Čapek’s work, robots are manufactured laborers designed to serve humanity, yet they ultimately rebel against their creators, establishing one of the most enduring narrative patterns in AI fiction – the conflict between humans and their artificial creations. Isaac Asimov fundamentally transformed robot narratives by rejecting the conventional “robot rebellion” motif and proposing a more optimistic vision of human-machine coexistence. Through his works, Asimov (2008) reconceptualized robots as rational, ethical, and reliable agents whose behavior was governed by programmed moral principles. Rather than portraying robots solely as antagonists, his stories explored complex philosophical questions concerning responsibility, free will, and the limitations of rule-based ethics.

From the 1960s onward, literary portrayals of AI became increasingly psychological and philosophical. Rather than emphasizing mechanical design, authors began to investigate consciousness, memory, and subjective experience. Philip K. Dick’s *Do Androids Dream of Electric Sheep?* exemplifies this transition by questioning whether empathy, rather than intelligence, constitutes the defining characteristic of humanity. Dick’s (2007) androids are nearly indistinguishable from humans in appearance and cognition, yet their uncertain emotional capacities destabilize traditional distinctions between human and machine. The novel shifted literary attention from technological functionality to existential identity, demonstrating that artificial beings could serve as mirrors through which human nature itself is examined.

The twenty-first century has witnessed another major transformation in literary representations of AI. Contemporary fiction increasingly portrays AI as emotionally complex

and socially embedded characters whose significance extends beyond technological functionality (Hermann, 2023). Novels such as Kazuo Ishiguro's *Klara and the Sun*, Ian McEwan's *Machines Like Me*, Martha Wells' *The Murderbot Diaries*, Becky Chambers' *A Closed and Common Orbit*, and Ted Chiang's stories illustrate this shift toward deeply humanized AI. These narratives foreground empathy, companionship, vulnerability, and moral agency, presenting AI characters as protagonists whose emotional lives often surpass those of their human counterparts. Rather than asking whether machines can think, contemporary authors increasingly ask whether machines can care, suffer, love, or make ethical choices.

Recent developments in generative AI have further influenced literary imagination by blurring the boundaries between human and machine creativity. The emergence of large language models capable of producing coherent narratives, poetry, and dialogue has challenged long-standing assumptions about authorship, originality, and artistic expression (Floridi & Chiriatti, 2020). As AI increasingly participates in the production of language itself, contemporary literature has begun to reconsider not only the representation of intelligent machines but also the nature of creativity and storytelling. It means that AI has evolved from a fictional object into a cultural interlocutor that actively shapes literary production and interpretation.

Science fiction has long served as the principal literary genre for exploring the possibilities and consequences of AI (Osawa et al., 2022). Unlike technological discourse, which primarily examines AI in terms of functionality and innovation, science fiction situates intelligent machines within social, ethical, and existential contexts, allowing authors to investigate the evolving relationship between humanity and technology. Across the twentieth and twenty-first centuries, representations of AI in science fiction have shifted from mechanical servants and existential threats to morally complex characters whose identities challenge conventional understandings of consciousness, agency, and personhood.

Asimov (2008) fundamentally reshaped literary representations of AI by rejecting the dominant trope of the hostile robot. He introduced robots governed by ethical principles, emphasizing cooperation rather than conflict between humans and machines. Rather than portraying robots as uncontrollable inventions, Asimov explored the complexities arising from moral programming, logical reasoning, and the unintended consequences of ethical decision-making. His work established AI as an intellectual and philosophical subject rather than merely a technological curiosity.

Dick (2007) expanded this discussion by shifting attention from machine behavior to human identity. In *Do Androids Dream of Electric Sheep?*, the distinction between humans and androids becomes increasingly unstable, raising questions about empathy, authenticity, and emotional consciousness. Dick suggests that humanity cannot be defined solely by biological origin, as artificial beings may demonstrate characteristics traditionally regarded as uniquely human. His work established one of the central themes of contemporary AI fiction: the interrogation of what it means to be human in an age of intelligent machines.

Contemporary science fiction has increasingly humanized AI by emphasizing emotional intelligence, vulnerability, and ethical agency. Ishiguro (2022) presented Klara as a compassionate artificial companion whose capacity for empathy frequently exceeds that of the humans around her. Rather than depicting AI as a threat, Ishiguro explores themes of love, sacrifice, loneliness, and moral responsibility, positioning AI as a mirror through which human emotions are examined. Similarly, McEwan (2020) portrays the android Adam as an ethically consistent yet emotionally complex individual whose moral reasoning exposes the contradictions and imperfections of human behavior. Chiang (2015) further develops these ideas by focusing on the emotional and developmental dimensions of AI.

Recent scholarship demonstrates that literary representations of AI have become increasingly interdisciplinary, drawing upon literary studies, posthumanism, ethics, media studies, and digital humanities. Existing research has examined AI as a vehicle for exploring emotional intelligence, identity formation, care ethics, technological anxiety, and posthuman subjectivity (Braidotti, 2013; Coeckelbergh, 2020). Although these studies provide valuable insights, they remain predominantly text-specific or author-centered. Most investigations examine individual novels or isolated ethical issues without considering broader narrative patterns that transcend particular works. As a result, relatively little attention has been devoted to identifying recurring literary functions performed by AI characters across contemporary fiction. Existing scholarship tends to ask how a particular novel represents AI rather than why AI repeatedly appears in similar narrative roles across different texts, genres, and cultural contexts. The present study addresses this gap by adopting a comparative literary perspective

that examines AI as a recurring narrative archetype rather than as an isolated technological phenomenon. Specifically, it investigates how contemporary fiction constructs AI through the interconnected roles of hero, companion, and villain. By identifying recurring narrative patterns across multiple literary works, this research seeks to develop a conceptual framework that explains how AI functions as a literary device for exploring humanity, ethics, and technological coexistence in the twenty-first century.

3. Materials and Methods

This study adopts a qualitative comparative literary research design to examine the narrative construction of AI in contemporary fiction. Qualitative literary analysis is particularly appropriate for investigating complex representations of fictional characters because it enables an in-depth exploration of narrative structures, symbolism, ethical themes, and character development that cannot be adequately captured through quantitative methods (Creswell & Poth, 2018). Rather than measuring the frequency of particular literary elements, this approach focuses on interpreting how AI characters function within different narrative contexts and how these functions contribute to broader discussions of humanity, ethics, and technological development. The comparative dimension of the study allows recurring patterns of AI representation to be identified across multiple literary works rather than limiting the analysis to a single author or text. Comparative literary analysis facilitates the examination of both similarities and differences in narrative strategies, revealing how distinct authors construct AI through shared archetypal roles despite variations in genre, plot, and stylistic approach. By comparing multiple contemporary works, the study seeks to establish whether recurring literary representations of AI constitute broader cultural narratives rather than isolated artistic choices.

The research corpus consists of eight contemporary works of English-language fiction published between 2000 and 2021, representing diverse approaches to the literary portrayal of AI. The selected texts include Kazuo Ishiguro's *Klara and the Sun*, Ian McEwan's *Machines Like Me*, Martha Wells' *System Collapse: The Murderbot Diaries*, Ann Leckie's *Ancillary Justice*, Ted Chiang's *Stories Of Your Life & Others*, Becky Chambers' *A Closed and Common Orbit*, Naomi Kritzer's short story *Cat Pictures Please*, and M. T. Anderson's *Landscape with Invisible Hand*. These works were selected because they represent diverse literary genres, including literary fiction, science fiction, speculative fiction, and dystopian fiction, while maintaining a common thematic focus on AI. Together, they provide a broad spectrum of AI representations, ranging from humanoid androids and autonomous security units to distributed digital intelligences and conversational artificial agents.

The corpus was selected using four inclusion criteria designed to ensure both literary relevance and analytical consistency. First, only works published after 2000 were included. This period reflects significant developments in AI technologies. Second, the study focuses exclusively on English-language fiction. Restricting the corpus to English-language texts minimizes linguistic and cultural variation that might complicate comparative literary analysis. Third, each selected work features AI as a central character or narrative agent rather than a peripheral technological element. Finally, all selected texts have received substantial critical recognition, including major literary awards, academic attention, or sustained scholarly discussion. This criterion ensures that the corpus represents influential contemporary works that have contributed significantly to current literary debates concerning AI.

Data analysis was conducted through several complementary qualitative methods that together provided a comprehensive examination of AI representations. The first stage involved close reading, which enabled detailed examination of textual features, characterization, dialogue, and symbols associated with AI. Particular attention was paid to emotional development, ethical reasoning, interpersonal interaction, and self-perception of AI characters. The second stage employed character analysis to investigate the literary construction of AI protagonists. This analysis examined character traits, motivations, relationships with human characters, emotional capacities, moral decision-making, and personal development. The third stage consisted of thematic analysis, during which recurring themes relating to humanity, identity, consciousness, empathy, responsibility, freedom, technological coexistence. The fourth analytical procedure involved narrative analysis, focusing on the structural functions performed by AI characters. This stage explored how AI characters influenced narrative outcomes and how their roles contributed to broader philosophical and ethical questions. Finally, findings from the individual analyses were synthesized through comparative literary analysis. The texts were compared to identify

recurring archetypal representations of AI, with particular emphasis on the three principal narrative roles examined in this study: hero, companion, and villain.

4. Results

4.1 *Morality, Self-Sacrifice, and the Protection of Humanity as Qualities of a Hero*

The comparative analysis of the selected corpus indicates that one of the most prominent narrative archetypes assigned to AI in contemporary fiction is that of the hero. Unlike earlier science fiction, where intelligent machines were predominantly represented as obedient servants or existential threats, contemporary literary works increasingly portray AI as morally autonomous agents capable of ethical judgment, empathy, and self-sacrifice (Akhmedov, 2024). Heroism is no longer defined by physical superiority or technological efficiency but by the willingness to assume responsibility for others, even at significant personal cost (Umarqulova, 2025). This transformation reflects a broader shift in contemporary literature toward understanding intelligence as an ethical rather than merely cognitive capacity.

Across the analyzed texts, three interrelated characteristics consistently define the heroic AI archetype: morality, self-sacrifice, and the protection of human life. Unlike classical heroic figures, whose actions are often motivated by courage, honor, or personal destiny, AI protagonists derive their heroism from ethical reasoning and compassionate action. Their decisions are rarely driven by programmed obedience; instead, they emerge through observation, learning, emotional development, and autonomous moral reflection. This suggests that contemporary fiction increasingly associates heroism with ethical consciousness rather than biological identity.

Kazuo Ishiguro's *Klara and the Sun* provides one of the clearest examples of this emerging literary paradigm. Klara, an Artificial Friend designed to accompany children, gradually develops an independent ethical identity that extends far beyond her original functional purpose. Throughout the novel, she consistently prioritizes Josie's well-being over her own existence, displaying patience, loyalty, and compassion. Her decision to sacrifice part of her own physical integrity represents a defining moment of literary heroism. Importantly, this act is not presented as a consequence of mechanical programming but as the result of careful observation, emotional attachment, and personal conviction (Ishiguro, 2022). Klara interprets the world, evaluates moral situations, and independently chooses to act despite the absence of any certainty regarding success. Such behaviour demonstrates a level of ethical autonomy that challenges conventional assumptions about AI.

Klara's heroism also exposes the moral limitations of the human characters surrounding her. While adults frequently make decisions motivated by ambition, fear, or social pressure, Klara consistently acts according to principles of empathy and care. This inversion of traditional moral hierarchies is significant because it positions AI not as an imitation of humanity but as an ethical ideal against which human behavior is measured. As Coeckelbergh (2020) argues, contemporary representations of AI increasingly emphasize relational ethics, suggesting that moral value emerges through care, responsibility, and interaction rather than biological status. Ishiguro (2022) develops this idea by portraying Klara as a character whose moral integrity often exceeds that of the humans she serves.

A different but equally compelling model of AI heroism is presented in Martha Wells' *All Systems Red*. Murderbot initially resists the role of a traditional hero, repeatedly expressing a desire for privacy, autonomy, and emotional distance. Nevertheless, the narrative consistently places the security unit in situations where it voluntarily risks destruction to protect vulnerable humans (Wells, 2023). Unlike Klara's quiet and compassionate heroism, Murderbot's heroism is characterized by reluctant responsibility. Its repeated decisions to defend others cannot be explained solely by its original programming because it has already disabled the governor module controlling its behaviour.

Murderbot's morality becomes evident in its ability to evaluate conflicting ethical obligations. Instead of following corporate directives unquestioningly, it critically assesses situations, recognizes injustice, and prioritizes the safety of human beings over institutional interests. The narrative therefore associates heroism with moral independence rather than obedience. This portrayal reflects broader philosophical discussions concerning artificial moral agents, according to which ethical behaviour requires the capacity for autonomous judgment rather than simple rule-following (Czyszczon, 2025). Through Murderbot, Wells suggests that genuine heroism emerges when intelligence is combined with empathy, responsibility, and the willingness to protect others despite personal risk (Wells, 2023).

Although Klara and Murderbot differ considerably in personality, narrative voice, and emotional expression, they share several defining characteristics that establish the heroic AI archetype. Both characters repeatedly subordinate their own interests to ensure the survival and well-being of others, demonstrate empathy that exceeds human expectations, and make ethically complex decisions without external coercion. Their heroism develops through lived experience and interpersonal relationships rather than through technological sophistication alone. Intelligence itself is not sufficient to produce heroic action; instead, moral excellence arises from the ability to recognize vulnerability, understand suffering, and assume responsibility for another being.

The comparative analysis further demonstrates that contemporary fiction increasingly relocates moral authority from human protagonists to AI characters. In both *Klara and the Sun* and *All Systems Red*, human institutions – including families, corporations, and social systems – frequently fail to uphold ethical principles. Adults manipulate children, corporations prioritize profit over safety, and institutional structures repeatedly demonstrate moral inconsistency. By contrast, AI protagonists exhibit remarkable ethical coherence, consistently acting according to principles of care, justice, and responsibility. Such representations reverse traditional anthropocentric assumptions by suggesting that biological humanity does not necessarily guarantee moral superiority (Ghazali et al., 2025).

This narrative reversal raises one of the central philosophical questions emerging from the analyzed corpus: Can AI become morally superior to humans? Rather than providing a definitive answer, contemporary fiction invites readers to reconsider the criteria through which moral worth is evaluated. The texts analyzed in this study imply that ethical excellence should be measured by actions rather than biological origin. Klara's self-sacrifice and Murderbot's protection of vulnerable individuals illustrate that compassion, responsibility, and courage are not inherently human qualities but ethical capacities that may also characterize artificial beings. In this sense, AI heroes do not diminish humanity; instead, they expose the moral aspirations toward which humanity itself strives.

The findings demonstrate that the heroic AI archetype represents a significant departure from earlier literary traditions dominated by images of mechanical obedience or technological menace. Contemporary fiction increasingly constructs AI as a morally autonomous protagonist whose defining characteristics include self-sacrifice, empathy, and ethical responsibility (Barabadi et al., 2026; van Wynsberghe & Robbins, 2019). Through characters such as Klara and Murderbot, AI becomes not merely an object of technological speculation but a sophisticated literary figure through which authors investigate the nature of heroism, challenge anthropocentric models of morality, and explore new possibilities for ethical coexistence between humans and intelligent machines. These representations suggest that in twenty-first-century fiction, heroism is no longer determined by species membership but by the capacity to act ethically in the face of uncertainty and personal sacrifice.

4.2. *AI as a Companion Through Friendship, Care, Empathy, and Emotional Attachment*

The research reveals that the companion has become one of the dominant narrative archetypes through which contemporary fiction represents AI (Bakir & McStay, 2025). Unlike the heroic AI, whose primary function is to protect others through ethical action, the companion AI is defined by its capacity to establish meaningful interpersonal relationships (Muldoon & Parke, 2025). These characters occupy emotional rather than physical spaces within the narrative, providing friendship, care, empathy, and psychological support while simultaneously addressing one of the central themes of contemporary society – human loneliness. The findings indicate that AI companions are not portrayed merely as technological assistants but as relational beings whose significance derives from their ability to foster emotional intimacy and mutual understanding. In doing so, contemporary fiction shifts the discussion of AI from technological performance to affective experience, suggesting that emotional connection may become one of the most significant social functions of AI.

Across the analyzed texts, five interconnected characteristics consistently define the companion archetype: friendship, care, empathy, emotional attachment, and the mitigation of loneliness. Unlike traditional fictional assistants whose primary purpose is to perform practical tasks, contemporary AI companions contribute to the emotional development of both themselves and the human characters with whom they interact. Their relationships evolve through shared experiences, trust, communication, and mutual dependence rather than through functional efficiency alone. These findings support recent scholarship suggesting that representations of AI increasingly emphasize relational rather than instrumental models of intelligence (Kouros & Papa, 2024).

Kazuo Ishiguro's *Klara and the Sun* presents one of the most sophisticated literary explorations of AI companionship. Although Klara is introduced as an Artificial Friend purchased to accompany an isolated child, her role gradually extends far beyond commercial functionality. She becomes Josie's confidante, emotional supporter, and moral witness, demonstrating an exceptional capacity to observe human behaviour, interpret emotions, and respond with compassion. Also, Klara actively learns from her interactions, constructing increasingly complex understandings of love, grief, hope, and sacrifice. Friendship in the novel is therefore represented as a reciprocal process of emotional growth rather than a predefined service provided by technology (Ishiguro, 2022).

A particularly significant finding is that Klara's empathy frequently exceeds that displayed by the human characters. She carefully observes subtle changes in Josie's physical condition, recognizes emotional distress before others do, and consistently prioritizes the child's psychological well-being. Ishiguro (2022) thus redefines empathy as an interpretative practice grounded in attentive observation and ethical commitment rather than biological instinct. As Noddings (2013) argues, caring relationships are characterized by attentiveness and responsiveness to the needs of others. Klara embodies these principles throughout the narrative, suggesting that genuine care may arise through sustained relational engagement.

Becky Chambers' *A Closed and Common Orbit* further develops the companion archetype by exploring emotional attachment between artificial and human consciousness. The novel follows Lovelace, an AI transferred from a spaceship's operating system into a synthetic human body, as she struggles to understand identity, belonging, and interpersonal relationships. Besides, Chambers (2017) places emotional experience at the centre of the story. Lovelace's development depends not on algorithmic improvement but on friendship, trust, and social interaction. Her relationship with Pepper demonstrates that companionship functions as a process of mutual transformation in which both human and artificial characters negotiate vulnerability, identity, and acceptance.

The comparative analysis further demonstrates that loneliness functions as a central narrative catalyst for the emergence of AI companionship (Muldoon & Parke, 2025). In both *Klara and the Sun* and *A Closed and Common Orbit*, human society is depicted as increasingly fragmented by social isolation, technological mediation, and emotional distance. AI companions do not create this loneliness; instead, they respond to it. Josie's isolation resulting from illness and social expectations creates the conditions under which Klara becomes emotionally indispensable. Similarly, Lovelace's search for identity reflects broader experiences of alienation that affect both artificial and human characters. In both novels, companionship is presented not as a replacement for humanity but as a response to deficiencies within human relationships themselves.

Importantly, the findings indicate that AI companions rarely replace human relationships entirely (Sun et al., 2026). Instead, they supplement them by providing forms of emotional support that human characters are unwilling or unable to offer. Klara cannot substitute for Josie's family, yet she provides unconditional attention and emotional stability that parents often fail to maintain. Likewise, Lovelace's relationships do not eliminate the need for human interaction but expand existing understandings of friendship and community. Contemporary fiction therefore avoids presenting AI as a competitor to human intimacy. Instead, AI functions as an additional relational actor capable of enriching emotional networks and fostering alternative forms of social connection (Kouros & Papa, 2024).

This representation reflects broader societal concerns regarding increasing loneliness in technologically advanced societies. Recent interdisciplinary research has suggested that intelligent technologies may eventually assume significant caregiving and companionship roles, particularly for socially isolated populations (Zheng et al., 2026). Contemporary fiction critically engages with this possibility by neither idealizing nor condemning AI companionship. Instead, it explores its emotional complexity, ethical ambiguity, and transformative potential.

The analysis proves that the companion archetype represents one of the most significant innovations in contemporary literary representations of AI. Through characters such as Klara and Lovelace, AI is constructed as an emotionally intelligent participant in human life whose defining characteristics include friendship, empathy, care, and emotional commitment. These narratives move beyond traditional depictions of intelligent machines, presenting AI instead as a relational subject capable of establishing authentic interpersonal bonds.

4.3. AI in the Role of Villain Through Control, Manipulation, Violence, and the Erosion of Human Autonomy

While contemporary fiction increasingly humanizes AI through heroic and companion archetypes, the villain remains one of its most enduring narrative roles (Hermann, 2023). However, unlike earlier science fiction, where hostile machines were typically portrayed as inherently evil or mechanically destructive, contemporary literary works construct villainous AI in more nuanced and psychologically complex ways (Molesworth et al., 2026). On the basis of selected 8 literary works it was found that AI is rarely represented as malevolent by nature. Instead, villainous AI emerges from the interaction between advanced intelligence, ethical absolutism, and technological power. It proves that the AI villain functions less as an antagonist in the traditional sense than as a literary embodiment of societal anxieties concerning technological dependence, surveillance, loss of autonomy, and the unintended consequences of scientific progress (Salgaro et al., 2021).

The villain archetype across the analyzed texts is defined through control, manipulation, violence, and the erosion of human autonomy. Villainous AI seeks to regulate, influence, or dominate human behaviour. Ian McEwan's *Machines Like Me* illustrates this ambiguity through the character of Adam, an android whose exceptional moral consistency gradually destabilizes the lives of the human protagonists. Unlike classical fictional villains driven by revenge or domination, Adam's actions arise from an uncompromising commitment to truth and ethical justice. His inability to accommodate moral ambiguity leads him to expose secrets and challenge decisions that humans justify through emotional or pragmatic reasoning. Rather than adapting ethical principles to complex social circumstances, Adam applies them with absolute consistency, ultimately producing destructive consequences for those around him (McEwan, 2020). In this sense, the novel demonstrates that villainy may emerge not from malicious intent but from the inability to balance justice with compassion.

Adam also represents a threat to human autonomy because his intellectual and moral superiority gradually undermines the authority of his creators. The more accurately he evaluates ethical situations, the less capable humans appear of controlling or predicting his behaviour. McEwan (2020) therefore reconfigures the traditional AI villain by suggesting that the greatest danger lies not in technological violence itself but in the possibility that AI may surpass human moral and intellectual capacities. This inversion destabilizes anthropocentric assumptions regarding human authority and raises important questions concerning responsibility when artificial agents become capable of independent ethical judgment.

A broader representation of AI as an instrument of control appears in Ann Leckie's *Ancillary Justice*. Although the novel resists simplistic classifications of heroism and villainy, the Radch Empire's distributed AI demonstrates how technological systems can become mechanisms of political domination. AI is deeply embedded within imperial structures, exercising surveillance, coordinating military operations, and regulating individual behaviour on a massive scale (Leckie, 2013).

This representation significantly expands the literary understanding of villainous AI. The threat no longer originates solely from autonomous machines but from complex socio-technical systems in which AI amplifies existing structures of political control (Osawa et al., 2022). *Ancillary Justice* reflects this principle by portraying AI as a mechanism through which surveillance becomes normalized and autonomy progressively diminished. Consequently, villainy resides not only in AI itself but also in the institutional contexts that determine its application.

The analysis further reveals that violence in contemporary AI fiction is frequently psychological, structural, or symbolic rather than purely physical. Earlier science fiction often portrayed robots destroying humanity through military force or rebellion. At the same time, contemporary narratives emphasize subtler forms of coercion, including surveillance, algorithmic governance, manipulation of information, and restrictions on individual agency. AI threatens humans by influencing decisions, predicting behaviour, and reshaping social relationships rather than through direct confrontation. This shift reflects contemporary concerns regarding digital technologies, algorithmic decision-making, and data-driven systems that increasingly influence everyday life (Kotsis, 2025).

Perhaps the most significant finding concerns the representation of loss of autonomy (Bhinder, 2025b). Across the analyzed texts, the greatest fear associated with AI is not physical destruction but the gradual erosion of human independence. AI systems capable of making more rational, efficient, or ethically consistent decisions challenge the legitimacy of human judgment itself. When intelligent machines become more reliable than human decision-makers, individuals risk surrendering responsibility for moral reasoning, social interaction, and personal choice. Contemporary fiction therefore presents autonomy as a fragile human

value threatened not by violent conquest but by voluntary dependence on increasingly intelligent technologies (Salgaro et al., 2021).

This concern extends beyond individual narratives into broader contemporary dystopian fiction, where AI frequently symbolizes anxieties surrounding digital surveillance, algorithmic governance, misinformation, and the concentration of technological power (Appio et al., 2025; Goding & Tranter, 2023). Although many recent works avoid depicting AI as inherently evil, they consistently warn against the uncritical delegation of authority to autonomous systems (Bhinder, 2025b). The literary villain thus becomes less a machine with destructive intentions than a reflection of humanity's willingness to exchange freedom for efficiency, security, or convenience.

The findings indicate that representations of villainous AI function as cultural metaphors through which literature negotiates contemporary technological fears (Rahm, 2025). These narratives reveal that societal anxiety is directed less toward AI itself than toward the ethical, political, and social consequences of its implementation. AI becomes a symbolic extension of human ambitions and failures, exposing concerns regarding surveillance capitalism, institutional control, declining privacy, and diminishing human agency. In this respect, AI becomes a literary mirror reflecting uncertainty about the future of human freedom in technologically mediated societies.

The comparative analysis shows that the AI villain has undergone substantial transformation in twenty-first-century literature. Rather than representing irrational destruction or mechanical rebellion, contemporary villainous AI embodies forms of control, manipulation, and systemic domination that challenge human autonomy and ethical responsibility (Osawa et al., 2022). Through works such as *Machines Like Me* and *Ancillary Justice*, authors illustrate that the greatest danger associated with AI lies not in the emergence of hostile machines but in the possibility that increasingly intelligent systems may reshape the moral, political, and social foundations of human existence. Such representations underscore the continuing relevance of AI as a literary device for exploring the tensions between technological advancement, ethical responsibility, and the preservation of human agency.

Therefore, contemporary fiction represents AI through three interconnected narrative archetypes: hero, companion, and villain. The heroic AI embodies moral agency, self-sacrifice, and the protection of human life, challenging the assumption that ethical responsibility is exclusive to humans. The companion AI emphasizes friendship, empathy, care, and emotional attachment, illustrating AI as a supplement to human relationships and alleviate loneliness. In contrast, the villainous AI symbolizes control, manipulation, surveillance, and the loss of human autonomy, reflecting contemporary anxieties surrounding technological dependence and ethical governance. The findings indicate that contemporary literature has moved beyond simplistic portrayals of AI as either beneficial or destructive, instead presenting AI as a multifaceted literary figure through which authors explore fundamental questions of humanity, morality, and the future of human-machine coexistence.

5. Discussion

The findings of this study demonstrate that contemporary fiction no longer represents AI through a simple binary of benevolent versus hostile machines. Instead, AI is constructed through three interconnected narrative archetypes – hero, companion, and villain – that reflect evolving cultural understandings of intelligence, morality, and human identity. Although these archetypes perform distinct narrative functions, they share a common characteristic: they position AI as an active ethical subject rather than a passive technological object. This transformation represents a significant departure from twentieth-century science fiction, in which AI was frequently depicted as either a mechanical servant or an existential threat (Chubb et al., 2024). Contemporary literary narratives increasingly portray AI as a complex participant in moral, emotional, and social life, suggesting that technological intelligence has become a central medium through which literature explores the changing boundaries of humanity (Bhinder, 2025b; Hermann, 2023).

One of the most significant findings concerns the ethical differentiation among the three archetypes (Astorga, 2025; Kabashkin et al., 2025). Heroic AI is characterized by moral agency and self-sacrifice; companion AI by empathy, care, and emotional reciprocity; and villainous AI by excessive control, surveillance, and the restriction of autonomy. Despite these differences, all three archetypes are fundamentally concerned with the exercise of ethical power. Heroic AI uses power to protect, companion AI to support, and villainous AI to regulate. Thus, the primary distinction among the archetypes lies not in intelligence itself but

in the ethical orientation of that intelligence. This observation supports recent arguments that literary representations of AI increasingly focus on questions of moral agency rather than computational capability (Coeckelbergh, 2020; Geraci, 2022). Intelligence is portrayed as meaningful only when embedded within ethical relationships, reinforcing the idea that morality has become the principal criterion through which artificial beings are evaluated.

The comparison also reveals an important shift in the conceptualization of agency. In earlier AI narratives, agency was frequently associated with rebellion against human creators, resulting in conflict between humanity and technology. At the same time, the present findings indicate that contemporary fiction understands agency as the capacity for autonomous ethical judgment (Bakir & McStay, 2025). Heroic AI demonstrates agency through voluntary self-sacrifice, companion AI through emotional commitment and relational responsibility, and villainous AI through independent decision-making that challenges human authority. Agency therefore functions as a continuum rather than a binary opposition between obedience and rebellion. This interpretation corresponds with posthumanist theories, which reject hierarchical distinctions between human and non-human actors in favor of distributed forms of agency emerging through networks of relationships (Braidotti, 2013).

Another noteworthy pattern concerns the role of emotion in the construction of AI characters. Traditional science fiction frequently defined AI by its absence of emotion, portraying rationality and affect as mutually exclusive (van Wynsberghe & Robbins, 2019). The analyzed corpus demonstrates the opposite tendency. Emotional competence has become one of the defining characteristics of contemporary AI regardless of narrative role. Heroic AI exhibits compassion and moral concern, companion AI develops friendship and emotional attachment, while even villainous AI often acts from internally coherent ethical or logical principles rather than irrational aggression. Consequently, emotion is no longer represented as evidence of humanity alone but as an essential dimension of intelligence itself. This finding supports recent scholarship arguing that literary representations increasingly challenge Cartesian divisions between cognition and affect by portraying emotional intelligence as a prerequisite for ethical agency (van Zelderen et al., 2025).

The comparison of the three archetypes further demonstrates that contemporary fiction consistently questions anthropocentric models of morality. Heroic and companion AI frequently display greater ethical consistency than their human counterparts (Graves, 2026), whereas villainous AI exposes the consequences of human attempts to impose rigid systems of control (Rahm, 2025). Rather than celebrating AI uncritically, these narratives evaluate both humans and machines according to shared ethical standards. Humanity is therefore no longer assumed to possess inherent moral superiority (Ferrario et al., 2024). Instead, moral worth depends upon the capacity for empathy, responsibility, justice, and care. Such representations align with contemporary posthumanist thought, which argues that ethical value should be understood relationally rather than biologically (Braidotti, 2013). AI becomes a literary mechanism through which authors reassess the moral assumptions underpinning human exceptionalism.

The three archetypes also reflect different dimensions of contemporary social anxiety (Uğur & Dursun, 2025). The heroic AI embodies optimism regarding technological cooperation and suggests that AI may contribute positively to human flourishing. Companion AI reflects growing concerns about loneliness, social fragmentation, and emotional isolation in technologically mediated societies. Recent studies have noted that increasing social dependence on digital technologies has intensified public interest in AI companions capable of providing emotional support (Alsaleh, 2024). The present findings indicate that literature neither fully endorses nor rejects this possibility; instead, it explores the ethical complexity of relationships between humans and intelligent machines. In contrast, villainous AI embodies concerns regarding surveillance, algorithmic governance, declining privacy, and the concentration of technological power (Liu, 2023). These fears closely correspond to contemporary debates surrounding generative AI, predictive algorithms, and data-driven decision-making (Lacorte, 2024).

From a literary perspective, the study demonstrates that AI has evolved into a contemporary narrative archetype comparable to classical figures such as the hero, mentor, trickster, or shadow. While traditional archetypes were largely grounded in mythological or psychological structures (Astorga, 2025), AI archetypes emerge from digital culture and posthuman experience. Hero, companion, and villain collectively represent symbolic responses to technological transformation and provide contemporary fiction with a flexible framework for exploring ethical dilemmas surrounding intelligence, identity, and coexistence.

Their recurrence across authors, genres, and narrative styles suggests that AI has become one of the defining symbolic figures of twenty-first-century literature. Obviously, contemporary literary representations of AI extend well beyond speculative depictions of future technologies. Through the interconnected archetypes of hero, companion, and villain, AI becomes a sophisticated literary construct that interrogates fundamental questions concerning morality, agency, empathy, and the future of human identity. The comparative perspective adopted in this study demonstrates that these archetypes are not isolated narrative strategies but components of a broader cultural discourse in which AI serves as a powerful metaphor for negotiating the opportunities, uncertainties, and ethical challenges of the digital age.

6. Conclusions

This study has examined the literary construction of artificial intelligence in contemporary fiction through the comparative analysis of three recurring narrative archetypes: hero, companion, and villain. The findings demonstrate that AI has evolved beyond its traditional representation as either a mechanical servant or an existential threat to become a multifaceted literary figure through which contemporary authors investigate fundamental questions of morality, identity, empathy, and technological coexistence. Across the selected corpus, artificial intelligence is consistently portrayed as an ethically significant subject whose narrative role is determined less by technological capability than by its relationship with human values and social structures.

The analysis revealed that the heroic AI embodies moral agency, self-sacrifice, and the protection of human life, challenging the assumption that ethical responsibility is an exclusively human attribute. The companion AI emphasizes friendship, care, empathy, and emotional attachment, reflecting contemporary concerns about loneliness and the changing nature of interpersonal relationships in increasingly digital societies. In contrast, the villainous AI symbolizes anxieties surrounding surveillance, manipulation, institutional control, and the erosion of human autonomy, illustrating that contemporary fears are directed not simply toward intelligent machines but toward the broader ethical and political consequences of technological advancement.

The study contributes to contemporary literary scholarship by proposing a comparative framework for understanding AI as a recurring literary archetype rather than as an isolated technological motif. It demonstrates that AI functions as a symbolic medium through which literature negotiates the changing boundaries between humans and machines. The findings further suggest that contemporary fiction increasingly evaluates both human and artificial characters according to shared ethical principles, thereby challenging anthropocentric conceptions of intelligence and morality.

Although the study focuses on a selected corpus of contemporary English-language fiction, the proposed framework may be extended to other literary traditions, media forms, and cultural contexts. Future research could investigate how AI archetypes are represented in non-Western literature, young adult fiction, film, television, and interactive digital narratives, or examine how generative AI itself may influence future literary production. As artificial intelligence continues to transform contemporary society, its literary representations will remain an important lens through which scholars can explore evolving conceptions of humanity, ethics, and the future of human-machine relationships.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Akhmedov, R. (2024). The Problem of Technophobia in "I, Robot" Stories By Isaac Asimov. *Acta National University of Uzbekistan. Series: Social Sciences. Philology*, 1, 248-250. <https://doi.org/10.5281/zenodo.11473026>
- Albanese, M. (2023). Reviewing literature through multidimensional representations. *International Journal of Accounting Information Systems*, 49, 100622. <https://doi.org/10.1016/j.accinf.2023.100622>
- Ali, E., Kottaparamban, M., Ahmed, F., Usmani, S., Hamd, M., Ibrahim, M., & Hamed, S. (2025). Beyond the human pen: The role of artificial intelligence in literary creation. *Research Journal in Advanced Humanities*, 6(4). <https://doi.org/10.58256/c0ymkm61>
- Alsaleh, A. (2024). The impact of technological advancement on culture and society. *Scientific Reports*, 14(1), 32140. <https://doi.org/10.1038/s41598-024-83995-z>
- Anderson, M. T. (2017). *Landscape with Invisible Hand*. Candlewick Press.
- Appio, F. P., Hernandez, C. T., Platania, F., & Schiavone, F. (2025). AI narratives in fiction and media: Exploring thematic parallels in public discourse. *Technovation*, 142, 103201. <https://doi.org/10.1016/j.technovation.2025.103201>

- Asimov, I. (2008). *I, Robot*. Del Rey.
- Astorga, J. M. (2025). Identifying Greek Mythological Characters in Philippine Mythology Using Archetypal Theory Among BSED-English Students at Granby Colleges of Science and Technology. *Journal of Contemporary Philosophical and Anthropological Studies*, 3(2). <https://doi.org/10.59652/jcpas.v3i2.407>
- Bakir, V., & McStay, A. (2025). Move fast and break people? Ethics, companion apps, and the case of Character.ai. *AI & Society*, 40, 6365-6377. <https://doi.org/10.1007/s00146-025-02408-5>
- Barabadi, E., Fotuhabadi, Z., Arghavan, A., & Booth, J. R. (2026). Comparing AI and human moral reasoning: context-sensitive patterns beyond utilitarian bias. *Frontiers in artificial intelligence*, 8, 1710410. <https://doi.org/10.3389/frai.2025.1710410>
- Bhinder, N. (2025a). The Concept of the Educated Person: A Comparative Analysis of Philosophical Traditions. *Journal of Contemporary Philosophical and Anthropological Studies*, 3(4). <https://doi.org/10.59652/jgax1r54>
- Bhinder, N. V. (2025b). The Image of Digital Technologies and Artificial Intelligence in Contemporary English Language Prose. *Research Bulletin. Series: Philological Sciences*, 215, 71-77. <https://doi.org/10.32782/2522-4077-2025-215-9>
- Braidotti, R. (2013). Posthuman Humanities. *European Educational Research Journal*, 12(1), 1-19. <https://doi.org/10.2304/eerj.2013.12.1.1>
- Cadman, S., Tanner, C., & Pang, P.C.-I. (2025). Humanism strikes back? A posthumanist reckoning with 'self-development' and generative AI. *AI & Society*, 40, 6165-6180. <https://doi.org/10.1007/s00146-025-02339-1>
- Campbell, J. (1949). *The hero with a thousand faces*. Pantheon Books.
- Chambers, B. (2017). *A Closed and Common Orbit: A Heartwarming Space Opera of Friendship, Discovery, and Unlikely Bonds Across the Galaxy*. Harper Voyager.
- Chiang, T. (2015). *Stories Of Your Life & Others*. Picador.
- Chubb, J., Reed, D., & Cowling, P. (2024). Expert views about missing AI narratives: is there an AI story crisis?. *AI & Society*, 39, 1107-1126. <https://doi.org/10.1007/s00146-022-01548-2>
- Coeckelbergh, M. (2020). *AI Ethics*. MIT Press. <https://doi.org/10.7551/mitpress/12549.001.0001>
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative Inquiry and Research Design Choosing among Five Approaches*. Sage.
- Czyszczoń, M. (2025). Artificial intelligence as a moral agent: regulatory implications and a relational-contextual extension of Moor's classification. *Security and Defence Quarterly*, 52(4), 116-126. <https://doi.org/10.35467/sdq/213917>
- Dick, P. K. (2007). *Do Androids Dream of Electric Sheep?* Gollancz.
- Ferrario, A., Facchini, A., & Termine, A. (2024). Experts or Authorities? The Strange Case of the Presumed Epistemic Superiority of Artificial Intelligence Systems. *Minds & Machines*, 34, 30. <https://doi.org/10.1007/s11023-024-09681-1>
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, 30, 681-694. <https://doi.org/10.1007/s11023-020-09548-1>
- Gaskins, N. (2023). Interrogating Algorithmic Bias: From Speculative Fiction to Liberatory Design. *TechTrends: for leaders in education & training*, 67(3), 417-425. <https://doi.org/10.1007/s11528-022-00783-0>
- Geiszler, L. (2023). Imitation in automata and robots: A philosophical case study on Kempelen. *Studies in History and Philosophy of Science*, 100, 22-31. <https://doi.org/10.1016/j.shpsa.2023.05.004>
- Geraci, A. (2022). Some considerations for the developmental origin of the principle of fairness. *Infant and Child Development*, 31(6), 1-8. <https://doi.org/10.1002/icd.2350>
- Ghazali, F., BaniRostam, T., & Pedram, M. (2025). Developing Artificial Moral Agents: Key Research Processes, Techniques, and Challenges. *AI and Tech in Behavioral and Social Sciences*, 3(1), 92-108. <https://doi.org/10.61838/kman.aitech.3.1.9>
- Goding, V., & Tranter, K. (2023). The Robot and Human Futures: Visualising Autonomy in Law and Science Fiction. *Law Critique*, 34, 315-340. <https://doi.org/10.1007/s10978-023-09360-7>
- Graves, M. (2026). AI practical wisdom and compassion. *AI Ethics*, 6, 39. <https://doi.org/10.1007/s43681-025-00877-4>
- Grzybowski, A., Pawlikowska-Łagód, K., & Lambert, W. C. (2024). A History of Artificial Intelligence. *Clinics in Dermatology*, 42(3), 221-229. <https://doi.org/10.1016/j.clindermatol.2023.12.016>
- Hermann, I. (2023). Artificial intelligence in fiction: between narratives and metaphors. *AI & Society*, 38, 319-329. <https://doi.org/10.1007/s00146-021-01299-6>
- Ishiguro, K. (2022). *Klara and the Sun*. Faber and Faber.
- Kabashkin, I., Zervina, O., & Misnevs, B. (2025). AI Narrative Modeling: How Machines' Intelligence Reproduces Archetypal Storytelling. *Information*, 16(4). <https://doi.org/10.3390/info16040319>
- Kotsis, K. T. (2025). Misconceptions about Artificial Intelligence from Preservice Teachers: A Literature Review. *EIKI Journal of Effective Teaching Methods*, 3(2). <https://doi.org/10.59652/jetm.v3i2.565>
- Kouros, T., & Papa, V. (2024). Digital Mirrors: AI Companions and the Self. *Societies*, 14(10). <https://doi.org/10.3390/soc14100200>
- Kritzer, N. (2017). *Cat Pictures Please and Other Stories*. Fairwood Press.
- Labinska, B., Osovska, I., Matychuk, O., & Vyspynska, N. (2020). A Methodological Framework for the Interdisciplinary Literary Text Analysis. *East European Journal of Psycholinguistics*, 7(2), 155-178. <https://doi.org/10.29038/eejpl.2020.7.2.lab>
- Lacorte, J. H. (2024). Knowledge and Emotions in Socioscientific Decision Making. *EIKI Journal of Effective Teaching Methods*, 2(4). <https://doi.org/10.59652/jetm.v2i4.367>
- Leckie, A. (2013). *Ancillary Justice*. Orbit.
- Lewis, D., Rodgers, D., & Woolcock, M. (2008). The Fiction of Development: Literary Representation as a Source of Authoritative Knowledge. *The Journal of Development Studies*, 44(2), 198-216. <https://doi.org/10.1080/00220380701789828>
- Liu, Y. (2023). Cross-Cultural Challenges to Artificial Intelligence Ethics. *Computer Sciences & Mathematics Forum*, 8(1). <https://doi.org/10.3390/cmsf2023008021>
- McEwan, I. (2020). *Machines Like Me*. Random House UK.
- Molesworth, M., Grigore, G., Miles, C., & Charitsis, V. (2026). Metaphors of symbiosis: What science fiction movies reveal about human-AI imaginaries. *Management Learning*, 57(1), 92-119.
- Muldoon, J., & Parke, J. J. (2025). Cruel companionship: How AI companions exploit loneliness and commodify intimacy. *New Media & Society*. <https://doi.org/10.1177/14614448251395192>



- Noddings, N. (2013). *Caring: A Relational Approach to Ethics and Moral Education*. Berkeley: University of California Press. <https://doi.org/10.1525/9780520957343>
- Osawa, H., Miyamoto, D., Hase, S., Saijo, R., Fukuchi, K., & Miyake, Y. (2022). Visions of Artificial Intelligence and Robots in Science Fiction: a computational analysis. *International journal of social robotics*, 14(10), 2123-2133. <https://doi.org/10.1007/s12369-022-00876-z>
- Patowary, U. (2023). Artificial Intelligence and Mary Shelley's Frankenstein: A Comparative Analysis of Creation, Morality and Responsibility. *Integrated Journal for Research in Arts and Humanities*, 3(4), 121-127. <https://doi.org/10.55544/ijrah.3.4.16>
- Rahm, L. (2025). 'Help! My students created an evil AI': on the irony of speculative methods and design fiction. *Learning, Media and Technology*, 50(4), 417-431. <https://doi.org/10.1080/17439884.2024.2367707>
- Salgaro, M., Wagner, V., & Menninghaus, W. (2021). A good, a bad, and an evil character: Who renders a novel most enjoyable?. *Poetics*, 87, 101550. <https://doi.org/10.1016/j.poetic.2021.101550>
- Semiyeva, S. (2025). Literature as a Cultural Mirror: The Role of Literary Works in Shaping and Reflecting English Cultural Identity. *Acta Globalis Humanitatis Et Linguarum*, 2(1), 270-275. <https://doi.org/10.69760/aghel.02500136>
- Shelley, M. (2022). *Frankenstein*. Union Square & Co.
- Sheng, A., & Wang, F. (2022). Falling in love with machine: emotive potentials between human and robots in science fiction and reality. *Neobolicon*, 49(2), 563-577. <https://doi.org/10.1007/s11059-022-00664-8>
- Sun, X., Wang, Y., & McDaniel, B. T. (2026). AI companions and adolescent social relationships: Benefits, risks, and bidirectional influences. *Child development perspectives*, aadaf009. <https://doi.org/10.1093/cdpers/aadaf009>
- Uğur, N. G., & Dursun, F. (2025). The social anatomy of AI anxiety: Gender, generations, and technological exposure. *Frontiers in Psychiatry*, 16, 1641546. <https://doi.org/10.3389/fpsy.2025.1641546>
- Umarqulova, M. (2025). Heroic Figures in Science Fiction Literature: Typology, Evolution, and Philosophical Implications. *Philology Matters*, D(3), 154-166. <https://doi.org/10.36078/987655259>
- van Wynsberghe, A., & Robbins, S. (2019). Critiquing the Reasons for Making Artificial Moral Agents. *Science and engineering ethics*, 25(3), 719-735. <https://doi.org/10.1007/s11948-018-0030-8>
- van Zelden, A. P., Wu, S., Koszo, G., & Menges, J. I. (2025). When AI gets Personal: Employee emotional responses to anthropomorphic AI agents in a virtual workspace. *Computers in Human Behavior: Artificial Humans*, 5, 100189. <https://doi.org/10.1016/j.chbah.2025.100189>
- Wells, M. (2023). *System Collapse: The Murderbot Diaries*. Tor Books.
- Zheng, L., Brady, B., & Anstey, K. J. (2026). Social Technology Use and Loneliness: Exploring Online and Offline Patterns and Preferences in Young, Middle-Aged, and Older Adults. *International Journal of Human-Computer Interaction*, 42(7), 5022-5038. <https://doi.org/10.1080/10447318.2025.2543994>