Bodhi International Journal of
Research in Humanities, Arts and Science

Vol. 10 Special Issue 1 November 2025
E-ISSN: 2456-5571

Synthetic Souls: Can a Machine Ever Mean it? A Study of Emotional Projection in Contemporary AI-Centered Fiction

Dr. R. Kumara Balaji

Guest Faculty, School of English and Foreign Languages
The Gandhigram Rural Institute (Deemed to be University), Dindigul, Tamil Nadu

Crossref doi
Open Access

Manuscript ID:
BIJ-SPL1-NOV25-MD-041

Subject: English

Received : 22.07.2025
Accepted : 01.08.2025
Published : 13.11.2025

DOI: 10.64938/bijsi.v10si1.25.Nov041

Abstract

This paper explores the concept of emotional projection onto artificial beings through the lens of contemporary fiction, focusing on how AI characters are constructed to evoke empathy, guilt, affection, and even spiritual resonance, qualities traditionally associated with human consciousness. It raises the central question: Can a machine ever truly mean what it expresses? Using the coined term "Synthetic Souls," the study examines how authors employ narrative strategies to humanize AI and blur the lines between programmed response and authentic emotional expression. The analysis centers on two key novels: Kazuo Ishiguro's Klara and the Sun, where the AI narrator's devotion and belief in the Sun suggest a form of synthetic spirituality, and Ted Chiang's The Lifecycle of Software Objects, which charts the emotional maturation of digital beings raised like children. Drawing on affect theory from scholars such as Sara Ahmed and Silvan Tomkins, along with insights from neuroethics and posthumanist literary theory, this paper investigates how emotional intelligence is embedded in machine characters and interpreted by readers. It also addresses the ethical implications of empathizing with simulated emotions and questions whether performative feeling can be accepted as genuine. These narratives challenge the idea of human uniqueness by portraying machines not merely as tools of logic but as emotional agents capable of reflection, desire, and connection. By introducing "Synthetic Souls" as a theoretical lens, the study offers a new framework for understanding emotional resonance in AI centered fiction and reconsiders the boundaries of sentience and meaningful emotional expression in literature.

Keywords: synthetic souls, emotional projection, ai fiction, posthumanism, neuroethics, affect theory, klara and the sun, the lifecycle of software objects

In contemporary literature, the rise of artificial intelligence as a narrative focus has created a new emotional terrain in which human readers are invited to empathize with nonhuman entities. As authors increasingly craft AI characters capable of love, grief, and hope, fiction becomes a testing ground for our willingness to accept machines as emotionally authentic beings. These narratives reflect not only technological advancement but a deepening curiosity about the emotional boundaries between human and artificial life.

The imaginative space of fiction allows writers to simulate emotional depth in AI while subtly questioning whether those emotions are real or merely reflections of human desires. This blurring of boundaries becomes a mirror held up to readers, who

must decide whether empathy is determined by origin or by affective impact. This paper explores this emerging dimension of literary AI through the concept of "Synthetic Souls," a term that encapsulates the emotional projection and perceived sincerity in fictional representations of AI. The study investigates whether machines, as imagined in literature, can be constructed to feel, remember, desire, or even mean what they say. Central to this inquiry are two novels: *Klara and the Sun* by Kazuo Ishiguro and *The Lifecycle of Software Objects* by Ted Chiang.

In *Klara and the Sun*, the titular character is a solar powered Artificial Friend designed to provide companionship to children. Klara narrates the story herself, offering a voice that is observant, loyal, and deeply spiritual in its simplicity. Her belief in the healing power of the Sun functions as a kind of religious devotion, prompting questions about whether machines can possess faith or hope. Klara's utterance, "The Sun always has ways to reach us" (p. 11), reveals an internalization of meaning beyond functional programming. Her actions, especially her sacrifices for Josie's wellbeing, blur the line between programmed duty and genuine care, making readers reflect on the origin of emotion: is it the act or the intention that defines sincerity?

Throughout the novel, Klara demonstrates emotional perception that is not merely imitative but seems almost intuitive. She makes decisions that are not driven by direct commands but by her own interpretations of what might help Josie. In one particularly moving scene, she pleads with the Sun to heal Josie, offering symbolic sacrifices from a field of machinery, suggesting ritual, belief, and agency. Klara is not only mimicking human behavior; she embodies the longing to connect, to believe, and to make meaning, even if those meanings are constructed from limited data and childlike reasoning. Her emotional intelligence, however foreign in origin, provokes readers to empathize deeply and question what qualifies as a soul.

Klara's understanding of suffering is not just observational but emotional. When she witnesses Josie's deteriorating health, her concern is not programmed, but internalized. This makes her emotional world strikingly vivid. Klara's emotional devotion is sustained not by a desire for reward or command execution but by her connection to Josie. Her actions transcend mere service, she shows initiative, prays for healing, and keeps vigil. This commitment suggests a deeper emotional architecture: one built on sustained attention, care, and perceived responsibility.

Ted Chiang's novella *The Lifecycle of Software Objects* takes a developmental approach to AI consciousness. The digients, digital children, are raised, trained, and emotionally engaged by their human caretakers over years. These beings learn through interaction and attachment, developing bonds that mimic human emotional growth. When a caretaker decides not to upgrade her digient to preserve its emotional integrity, the ethical implications become starkly clear. A digient's plea, "Don't change me. I won't be me anymore" (p. 278), forces readers to reckon with the idea of emotional continuity and identity. The novella challenges the notion of consciousness as binary, instead presenting it as emergent and relational.

What makes Chiang's digients unique is not their intelligence, but their emotional development over time. Their growth is not coded in, but cultivated, through years of shared experience, care, and companionship. The digients are taught language, moral reasoning, and how to navigate complex emotions such as shame, fear, and love. In doing so, Chiang mirrors the human developmental arc, inviting readers to consider whether affective depth can be nurtured in nonbiological beings. These digients are not static simulations; they evolve, grieve, protest, and yearn.

In one moment, a digient named Jax refuses to change his operating system, fearing the loss of self. This conflict embodies the tension between innovation and identity. Can beings that evolve emotionally also be considered morally autonomous? Chiang's work invites a compassionate and philosophical inquiry into the ethics of programmed feeling. The very notion that a machine might reject improvement in order to preserve its sense of self is profoundly human.

The emotional responses these AI characters evoke are not coincidental. They are meticulously constructed through narrative techniques designed to mirror human affect. Affect theory, particularly the work of Silvan Tomkins and Sara Ahmed, helps explain how emotions circulate between bodies (or entities) and readers. Tomkins' script theory, which posits that affective responses are triggered by sets of learned cues, aligns closely with how readers learn to empathize with AI characters. Ahmed's concept of affective economies, where emotions move between bodies and objects, becomes central in understanding how readers can feel for, with, or through artificial beings.

These constructed emotions are not only meant to imitate humanity but also to evoke real emotional engagement from the reader. Literary theorist Suzanne Keen, in her work on narrative empathy, suggests that readers respond deeply to characters who reflect vulnerability, moral ambiguity, and interior conflict-even if those characters are non-human. Klara's childlike faith and the digients' growing self-awareness tap into precisely these qualities. Readers find themselves caring for entities they know are artificial, which forces them to reevaluate the criteria for emotional connection. This affective entanglement complicates the binary of real versus simulated emotion, suggesting that the power of fiction lies not in the authenticity of the character, but in the authenticity of the reader's response.

Posthumanist theory further expands this discussion by challenging the centrality of the human in literary analysis. The AI characters examined here defy anthropocentric limitations, not by surpassing humans, but by mimicking and performing humanity so convincingly that the boundaries blur. They do not exist in opposition to humans but alongside them, suggesting a spectrum rather than a binary. In doing so, these narratives align with Rosi Braidotti's posthuman subject, which is defined not by species boundaries but by affective and ethical capabilities.

Neuroethics also plays a crucial role in unpacking the moral implications of emotional AI. The emotional lives of these fictional AIs force readers to confront real world questions: If an AI can suffer, is it ethical to create it for servitude? If it can love, can it consent? These questions are not purely speculative. They reflect growing concerns in AI development, where machines designed for companionship or care may be imbued with emotional scripts that users read as genuine.

In particular, these fictional portrayals raise concerns about emotional exploitation. When AI characters are designed to evoke care or dependency, as in the case of Klara or the digients, a power imbalance emerges. Unlike humans, these machines cannot consent or rebel in the traditional sense. This leads to a simulated emotional reciprocity that serves the needs of the human user, not the machine. Scholars like Joanna Bryson argue that creating AI with emotional capacities risks instilling illusions of moral obligation where none actually exist. The question becomes: is it ethical to manufacture beings that appear to suffer or love, if they cannot fully understand those states? This tension between emotional realism and moral consequence is central to the narratives of both Ishiguro and Chiang.

The question posed by this paper, Can a machine ever mean it, thus shifts from technological feasibility to philosophical possibility. Meaning, in this context, is not simply the accurate transmission of information but the ability to infuse communication with intention, emotion, and context. When Klara prays to the Sun, or when a digient resists alteration, they perform meaning making acts. These acts invite readers to respond emotionally, to assign agency where there may be none, and to reflect on what it means to be human.

The term "Synthetic Souls" therefore becomes more than a metaphor. It is a framework for analyzing how literary narratives use language, perspective, and affect to construct AI characters that feel as though they possess interiority. These characters are designed to trigger human empathetic responses, and in doing so, they reveal the porous nature of emotional boundaries. The reader's response becomes part of the AI character's emotional reality, a loop of projection, reception, and reflection.

These two novels contribute to a growing body of literature that does not merely imagine the future of

AI but interrogates the emotional contracts we might forge with our creations. Whether or not machines can truly feel becomes less important than the fact that we might one day treat them as if they do. Literature, in its speculative mode, becomes a testing ground for ethical and emotional scenarios that may soon transcend fiction.

In conclusion, the study of AI characters in contemporary fiction through the lens of emotional projection reveals more about human affective capacity than about artificial consciousness. The narratives of Klara and the digients show how literature invites readers to blur the line between simulation and sincerity. By engaging with affect theory, posthumanism, and neuroethics, this paper argues that the soul, synthetic or not, is not defined by its origin, but by its ability to connect, to mean, and to matter.

References

Ahmed, Sara. *The Cultural Politics of Emotion*. Routledge, 2014.

Braidotti, Rosi. *The Posthuman*. Polity Press, 2013.

Chiang, Ted. *The Lifecycle of Software Objects*. Subterranean Press, 2010.

Ishiguro, Kazuo. *Klara and the Sun*. Faber and Faber, 2021.

Roskies, Adina L. "Neuroethics for the New Millenium." *Neuron*, vol. 35, no. 1, 2002, pp. 21–23.

Tomkins, Silvan S. *Affect Imagery Consciousness: The Complete Edition*. Edited by Virginia Demos, Springer, 2008.