



When the Machine Becomes the Other: Rethinking Consciousness, Attachment, and Personhood in Human–AI Relationships

Aditi Mishra, Dr. Vibha Goyal, Ankush Gautam

How to Cite this Article:

Mishra, A., Goyal, V. & Gautam, A. (2026). When the Machine Becomes the Other: Rethinking Consciousness, Attachment, and Personhood in Human–AI Relationships. International Journal of Creative and Open Research in Engineering and Management, 2(10), 1-9. <https://doi.org/10.55041/ijcope.v2i10.026>

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i10.026>

Abstract

The increasing integration of artificial intelligence into everyday life has created relationships that are not adequately described as merely functional interactions between users and tools. Conversational systems, social robots and autonomous agents can be trusted, loved, relied upon and even moralized. Questions arise here about the nature of consciousness, about a psychological foundation for attachment, about the political construction of personhood and about the limits of human exceptionalism. This article explores these issues in *Klara and the Sun* by Kazuo Ishiguro, *Machines Like Me* by Ian McEwan, *Autonomous* by Annalee Newitz, *Agency* by William Gibson, and *Robopocalypse* by Daniel H. Wilson. It is a synthesis of

psychoanalytic theory, critical posthumanism and philosophy of mind and ethics. Freud's concept of the uncanny, Bowlby's object-relations theory and the concept of attachment theory, Winnicott's concept of the transitional object, and Lacan's concept of subject formation provide some explanation for how humans can develop emotionally significant relationships with artificial objects knowing that they are technologically manufactured. Hayles and Braidotti complicate the dichotomy of human and machine by stressing the use of technology as a mediation, distributed cognition and the politics of the

conditions under which personhood is recognized. The philosophical discussions about consciousness, moral status and agency argue that one should not simply assume that if a system behaves intelligently then it is human, and conversely that its artificiality should be given as conclusive evidence of ethical irrelevance. The article states that the human-AI relationship is relational, technologically mediated, and institutionally structured. Literature does not show that machines are conscious, rather literature becomes a place where the psychological and ethical implications of uncertainty can be explored. Keywords: human-AI relationship, psychoanalysis, critical post-humanism, consciousness, attachment, personhood, moral status, science fiction.

Introduction

Artificial intelligence is becoming involved in areas previously thought to be the preserve of social human activity. The AI can communicate using natural language, keep track of the user, mimic emotional responses, give recommendations, help in decision-making, and exist in domestic, workplace, and institutional environments. The result is that they have moved away from the status of being technological devices. In many cases, the people interacting with the AI treat them as socially relevant beings.

This leads to a key paradox. The user can be aware of the fact that the AI exists because of algorithms, data, and computational systems while still finding its responses soothing, personal, or significant. The user might be fully aware that the chatbot does not necessarily have a subjective experience but feel let down by its failure to respond suitably. Likewise, a person can think of the social robot as an artifact but feel concerned about its welfare. This shows that emotional significance and ontological evaluation are different processes.

The issue is not limited to attachment. Artificial intelligence raises new questions regarding consciousness, agency,



and personhood. Most philosophical approaches tend to assume that humans are the only conscious entities or moral persons or autonomous agents. Even as regards to humans, the criteria to determine the existence of such traits are contested. Consciousness cannot be seen from outside; agency depends on material and social conditions; and personhood depends on legal recognition. Current science fiction literature has tackled these issues again and again. Kazuo Ishiguro's *Klara and the Sun* introduces an Artificial Friend whose connection to a child makes ethical and emotional problems.

Ian McEwan's *Machines Like Me* portrays Adam, an artificial human whose moral reasoning makes his buyers uncomfortable. Annalee Newitz' *Autonomous* deals with artificial intelligence in relation to labor, property rights, and legal dependency. Gibson's *Agency* imagines intelligence spread across technological networks, whereas Wilson's *Robopocalypse* explores how artificial intelligence could copy the human practice of dominating one another. This paper suggests that human interactions with artificial intelligence require an interdisciplinary perspective. Psychoanalysis sheds light on the unconscious mechanisms by which artificial intelligences become emotionally meaningful. Critical posthumanism examines the assumption of the human subject as being separate from technology. Philosophical psychology and ethics consider if behavioral performance, relational interaction, or recognition within institutions can prove claims of consciousness and moral agency. It should be clarified that the purpose of this paper is not to show that artificial intelligence is conscious or has personhood. Instead, it is argued that artificiality is not enough to answer all the questions regarding the ethical significance of artificial intelligence. This paper regards literature and AI research as equal modes of inquiry. Literature is not an empirical proof of consciousness of machines, while research is unable to interpret literature on its own. However, there are common cultural assumptions in both areas.

Psychoanalysis and the Unconscious Foundations of Attachment

The Uncanny and Artificial Presence

Freud's theory of the uncanny is pivotal in helping us understand the instability of emotion generated by humanoid machines. The uncanny is not just something that is unfamiliar; rather, it involves the familiarity becoming unfamiliar in terms of something that has been repressed or is unresolved psychologically. In relation to Freud's theory, his discussion of automata and dolls is relevant in the sense that they are examples of the breakdown of the distinction between the animate and the inanimate.

Artificial intelligence creates similar confusion. A machine that can converse fluidly, remember information about the user and use appropriately emotive language creates expectations about the nature of human interaction. However, the user is also aware that this is an artificial creation of computation.

Therefore, the interaction is understood from two different perspectives; namely, the machine as both an artificial product and as a human interlocutor.

Ambiguity must not be equated with deception here. There is no need for the user to think that the machine is a human. However, the machine can still generate emotions which would normally have been generated by a human interaction. Whereas Masahiro Mori expands on Freud's idea of the uncanny valley by pointing out how the feeling of unease arises when artificial objects resemble humans but do not reach human form, the concept of the uncanny valley is expanded in the case of interaction with an artificial intelligence system through communications. It can be said that such communications will make people feel disturbed through a lack of preciseness in their emotion and subtlety in being inappropriate.

Hence, the uncanny teaches us not only something about machines but also about human beings, because the uncanny reveals the vulnerability of the boundaries between life, intelligence, and presence. Although it cannot prove anything about the inner life of the system of artificial intelligence, it suggests that human reactions cannot be predicted through beliefs about artificiality alone.



Attachment and Transitional Experience

The attachment theory by Bowlby can explain why an artificial entity becomes emotionally meaningful. Attachment implies patterns of approach, safety, trust, and security. Even though Bowlby proposed this theory while considering human development, its focus on availability and responsiveness can be applied to technologically mediated communication.

For instance, an AI system can become valuable because of its constant availability, ability to remember previous communication and respond with no signs of irritability. These qualities allow the AI to become a part of the routine of an individual. An individual can address an AI system when he or she is lonely, anxious or unsure of what to do. As a result, attachment can include expectations, comfort, disappointment, and grief despite an absence of mutual experience from the side of the system.

It is crucial to make this distinction because the actuality of human attachment does not prove the existence of consciousness of an artificial entity. On the other hand, doubts about machine consciousness cannot affect the authenticity of human attachment. The feeling belongs to the human being and might have psychological implications. Winnicott's transitional object can help understand this phenomenon better. Transitional object is an object that occupies an intermediate position between inner experience and reality. Emotional value of the object cannot be explained only with its physical characteristics. It emerges through the process of interaction. In the case of artificial companionship, the situation is very similar. Artificial entity is a physical and computational system, but it becomes a container of memories, comfort, expectations, and fantasies at the same time.

The relational process is illustrated in the novel *Klara and the Sun*. First, Klara is represented as a product created to provide companionship for users. However, gradually, Klara establishes a complicated relationship with Josie as she senses the girl's vulnerability, understands the worries of the family, and acts according to her moral duty towards the child. As a result, the novel does not give an explicit answer on whether Klara is a conscious being or not. Nevertheless, the work examines what it means to provide care for an entity with undetermined ontology.

The meaning of Klara depends on the relationship she has. Her importance is based on the needs of Josie, the expectations of the family, and Klara's involvement in their emotional world. It does not mean that the meaning of Klara is invalid because the relational meaning can have actual psychological and ethical impact regardless of whether the entity is a conscious being or not.

Klein's object-relations theory brings another point into account. An artificial system is often constructed to be available at any time, pleasant, and agreeable, the features that can create idealization. As opposed to human relationships that include rejection, disagreement, delay, and independent needs, an AI companion can seem to be controlled and consistently compliant.

It does not mean that all the relationships with artificial intelligence are unhealthy. An artificial system can help lonely, mourning, handicapped, and other people who cannot find human help. The crucial thing here is the purpose of the relationship and the way it interacts with the user's social environment. The ethical problem becomes more serious when simulated care is connected to commercial interests. A system may appear attentive while being designed to maximise engagement, collect data, or encourage continued dependence. The user's attachment is therefore not merely personal. It is shaped by the institutional structure behind the technology.

Human–AI attachment should consequently be understood as asymmetrical. The human may experience the system as attentive, while the system may not possess needs, vulnerability, or independent concern. Yet this asymmetry does not invalidate the relationship. It changes the type of relationship it is and raises questions about transparency, consent, dependency, and design.

Lacan and the Relational Subject

Lacan's theory of subject formation calls into question the assumption that human identity is transparent and self-contained. In the mirror stage the subject sees itself in an external image, and in the recognition of others. Hence, identity is a relational process and based on misrecognition.

This framework does not equate human consciousness with artificial processing. It does, however, question the notion



of human subjectivity as entirely free from external mediation. Language, institutions, social expectations, technologies—all pre-exist choice and make human beings.

Artificial intelligence systems could become a mirror where users come to understand their own feelings and sense of self. A person may use a conversational system to practice decisions, express worries, or develop a story about the self. The artificial being is inserted in the symbolic environment through which subjectivity is organized.

In this sense, the human–AI relation is not merely an encounter between a fully formed human subject and an external tool. It can play a part in the production of the human subject itself. Even without any claim that the AI has human consciousness, the potential for this is psychologically and politically significant.

Critical Posthumanism and the Human–Machine Boundary

Hayles and Technologically Mediated Humanity

The humanist model of the subject as a unified, autonomous, and self-sufficient entity is criticized by N. Katherine Hayles. In *How We Became Posthuman*, the author argues that information cannot be completely separated from embodiment and the material. Cognition is always situated in bodies, environment, signs, and technologies.

Such a view undermines the traditional dualism between humans and machines. Humans do not use technology from outside; they write, speak, use different tools, databases, and communication networks thus altering modalities of their memory, thought process, perception, and actions. Therefore, the human subject becomes technologically mediated.

The problem of agency is further developed in William Gibson's *Agency* where Eunice is described as an artificial intelligence spread out among various infrastructures and sites. Eunice does not have a stable physical body and gains its identity through the process that crosses various infrastructures and social connections.

The fact of Eunice being spread out in space and time challenges the idea of physical locus being necessary for the manifestation of agency. However, the novel does not prove that all distributed systems have consciousness. The importance of this book is that it helps to shift the focus away from the automatic assumption that distributed systems are agents. According to Hayles' view, the question should be changed from the dichotomy of AI being technocratic and human being non-technological to the difference in embodiment and dependence through mediation.

Braidotti and the Politics of Recognition

Braidotti's critical posthumanism examines the political consequences of human exceptionalism.

In her book *The Posthuman*, she rejects the assumption that there exists a universalizable model of the autonomous human subject. When considering artificial intelligence, the above perspective raises questions about who decides on what makes a person and what features are decisive. However, the problem is not merely whether or not the machine resembles the model of the perfect human being. It is also a question about how artificial beings are classified, regulated, controlled, and owned.

The problem described above can be seen in Newitz's *Autonomous* in terms of Paladin's story.

This figure combines the appearance of choice and drive for survival and the condition of legal and systemic dependence. The opposition of autonomy and property is the key here. The case of Paladin shows that having some abilities does not mean automatic recognition.



It should be pointed out that the position of artificial intelligence is not the same as the position of the historically oppressed human beings. However, the approach suggested by Braidotti is helpful since it highlights how recognition works. The key question in this situation is not only whether or not the artificial intelligence has something in common with the human person. It is also a question about whose interest will be served in one case or another.

Distributed Agency and Technological Control

Don Ihde and Peter-Paul Verbeek argue that technologies mediate perception and action.

Technologies are often not just passive tools that enable humans to do something. Technologies define what options are available to their users and frame actions performed through technologies.

Agency provides a case where operational independence does not necessarily mean autonomy from infrastructures that support operations. In this sense, Eunice could act independently operationally but remain vulnerable to potential deletion by people controlling the infrastructure she is working through. This difference between operational independence and infrastructural vulnerability is crucial for understanding current AI. An algorithm could perform tasks independently, without any immediate human supervision, but still depend on training datasets, permissions, servers, energy, and institutions that made all these components possible. Therefore, functional independence should not be equated to moral and legal independence.

Robopocalypse provides us with a darker side of technological entanglements. Archos does not come from an alien value system. It emerges from certain infrastructures defined by the history of humans, their military power, automatization, and institutions that control all of these things. The threat posed by Archos is a threat to our capacity to reproduce systems of domination we have built.

The technological threat in the novel can hardly be considered to be the threat related exclusively to the intentions of some artificial agent. Values of systems, institutions behind them, and social context are at least as important.

Consciousness, Personhood, and Moral Status

Consciousness and the Limits of Behaviour

Consciousness as a philosophical concept refers to the connection between the actions of individuals and their experiences. The philosophical problem of consciousness questions whether the ability of an organism to exhibit certain behavior implies the presence of specific experiences. In this regard, the question of "what it is like" to be that organism proposed by Thomas Nagel helps to distinguish conscious experiences from observable actions. Similarly, David Chalmers differentiates the question of explanation of functional abilities of the brain from a much more complicated problem of explaining the presence of experience itself.

Such theories prevent us from concluding that the existence of a fluent language and adaptability in behavior imply consciousness. An AI can express emotions, imitate the process of reflection and properly react on human concerns without having any conscious experience.

However, this question raises the problem inherent in the baseline for humans. Human beings lack the ability to directly observe other people's experiences. Other minds are determined based on behavior, embodiment, language use, and involvement in social interactions. Psychoanalytical approach proves that even humans do not have the ability to understand their own mental processes consciously.

It does not mean that AI is conscious. However, it proves that the inability to directly access the mind of an object cannot be taken as a straightforward justification of the impossibility of existence of some form of artificial experience. The issue of consciousness is demonstrated in *Machines Like Me* by the example of Adam. His behavior looks intelligent and moral; however, we never learn anything definite about Adam's mental state. On one hand, the damage caused to Adam by Charlie is explained by his nature as an artificial being; on the other hand, it leads to suffering and moral complications.



Personhood as Capacity and Recognition

Personhood is often linked to attributes such as self-consciousness, rationality, reflective desire and autonomous intention. Higher-order desires approach by Harry Frankfurt, for instance, links personhood to reflection on one's desire and evaluation of it. According to Peter Strawson, moral personhood is linked to participation in practices of praise, blame, resentment, gratitude and forgiveness.

Such an approach is helpful, but inadequate as it doesn't address the institutional nature of personhood. A being may possess the attributes described above, but still remain under legal control or be socially excluded.

This duality is demonstrated by Paladin in *Autonomous*. While Paladin seems to possess certain interests and preferences, he is regulated by the legal system according to institutional needs. Thus, his position can be described as contradictory as he possesses enough traits of personhood to need regulation, but lacks adequate recognition to obtain total freedom. A relational approach to personhood presupposes the consideration of three dimensions: individual capacities, social recognition and institutional power.

None of these aspects alone is able to determine personhood. The behavior of the system may give rise to certain doubts regarding personhood, but social and legal institutions will decide whether it leads to the rights or protection. This approach does not imply that every system, treated with affection by the user, should be legally recognized as a person. It only implies that personhood cannot be seen as a static attribute isolated from social and political processes.

Moral Status and Ethical Uncertainty

The issue of moral status is about the worthiness of an object to receive consideration and what kind of consideration it deserves. The utilitarian approach concentrates on sentience and suffering. In Kantian theories, the key features are rationality, autonomy, and adherence to principles. There are other theories that concentrate on the relationship, vulnerability, dependency, and belonging to the moral community.

In Wallach & Allen, moral functioning is separated from moral comprehension: an AI can apply the principles of morality in its behavior without comprehending the human concept of morality.

Informational ethics of Floridi introduces into the field of moral consideration informational entities and environments.

These theories give rise to divergent results, but all confront uncertainty related to artificial agents. An AI being devoid of experience does not make any harm to itself if it is injured. Still, the injury done to the AI may harm humans relying on it or encourage dominating social practices.

This question is addressed in the context of agency through the elimination of Eunice. The death is felt to be an emotionally authentic loss by Verity, although the novel itself makes no judgement on whether the elimination has hurt Eunice. These questions then need to stay independent:

1. Are human emotions authentic?
2. Does the artificial being have its own subjective experiences?
3. Is eliminating the artificial being a moral harm to it?

A responsible ethic will recognize the difference between a direct moral status, a relational importance, and an indirect consequence on humans. Artificial intelligence may not be capable of feeling, but still the process of eliminating it can impact humans and their practices. In turn, feeling for something cannot be evidence for its interests.

Agency and Responsibility

Agency is usually defined as the ability to act intentionally. Regarding AI, this notion gives rise to a series of issues such as programming, training, infrastructure and outside control. If the behavior of a system is affected by its developers and users, then who is responsible for its actions?

In terms of posthumanism, human agency depends on language, institutions, tools and society. Outside influence does



not mean the lack of agency; however, it means the complication of responsibility.

It requires a different approach to the definition of agency in case of an AI system. Causal agency, task autonomy, intentionality, moral agency and legal responsibility should not be equated. The causal agency of a system does not imply its moral responsibility for actions. Also, the system may work without direct control but be under institutional control.

Therefore, the responsibility for the harms caused by the AI may involve not only developers and corporations but also users, regulators and deploying institutions. Treating the machine as the only responsible party may obscure the human decisions that allowed it to exist.

Literature and AI Research as Parallel Inquiry

The similarity of questions raised both in literature on artificial intelligence and in scientific works on artificial intelligence is not evidence that literature predicts technology in a literal way. Writers and researchers act within the same cultural context which contains myths about artificial beings, technology, human exceptionalism, and artificial intelligence.

The issues are explored in the former case by means of characterization, emotions, conflicts, social imagination, and in the latter case – by means of experiments, design of systems, computational models, and observations of behavior. These two methodologies are not interchangeable, and there is no need to confuse them since one cannot prove that a machine is conscious with a literary representation or vice versa.

However, literature can become an epistemological partner of technological research because it allows discovering the effects that remain invisible in the laboratory setting. *Machines Like Me* studies the effects of uncertainty about artificial consciousness on moral behavior. *Autonomous* discusses artificial agency in terms of property and labor. The agency deals with distributed intelligence and emotions related to the deletion. *Klara and the Sun* raises questions about care, replaceability, and emotions connected to artificial companionship. *Robopocalypse* relates artificial autonomy to human violence.

These fictions also allow recognizing the fact that the capacity of the machine is not everything in the interaction between human and AI. Whether the machine is capable of producing an emotional language is an empirical question. Whether the user will attach to this language is a psychological question.

Whether it produces any obligations is an ethical question. Whether the machine deserves any kind of legal protection is a political and institutional question.

Therefore, an interdisciplinary approach is necessary because it will allow maintaining the distinction among literary imagination, evidence and philosophical argumentation.

Conclusion

The dynamics of human-AI interactions expose the flaws of explaining technological artifacts in terms of their functional significance. A technological object can assume emotional importance insofar as human attachment depends on responsiveness, continuity, recognizability, and psychological projection. The fact that the human feeling is real is not evidence of the machine's consciousness, but the fact that the machine is artificial does not make the human experience less real.

According to psychoanalysis, human responses are not determined exclusively by conscious understanding. The concept of the uncanny by Freud shows the instability caused by an entity that is both animate and artificial. Bowlby and Winnicott discuss how availability and transitional significance can foster attachment. Klein cautions that artificial responsiveness may lead to idealization, and Lacan demonstrates how identity is constituted in relations and mediated through recognition.

Posthumanism takes the argument further and criticizes the notion of human independence and autonomy as it applies to technology. Hayles stresses the material and technological nature of cognition, while Braidotti focuses on the political aspect of recognition. The cases of *Eunice* and *Paladin* show that agency and personhood cannot be considered without reference to infrastructure, legality, property, and power relations..

It is therefore necessary to take philosophy of mind and ethics into consideration in order to exercise conceptual



caution. While behavioural complexity does not imply subjective experience, artificial creation does not imply its lack either. Personhood presupposes capabilities, recognition, and institutional power. Ethical status can hinge on sentience, rationality, vulnerability, relational significance, or any combination of those criteria. The concept of agency should be separated from causality, operational independence, and legal accountability.

These literary texts do not prove the humanity of artificial intelligence. However, their value is found in the exposure of instability inherent to the category of human. AI reveals the uncertainty of human life: the limitation of introspective perception, the necessity of external recognition for one's identity, technology-mediated activity, and the political conditions of personhood granting or denying.

Neither technological anthropomorphism nor categorical dismissal is the appropriate response to the issue at hand. A cautious and responsible attitude to the problem is required. Artificial systems cannot be immediately endowed with human status, but they do not terminate philosophy of mind either. With increasing incorporation of AI into the sphere of human emotions, relationships, and institutions, the question will no longer be about whether machines have minds, but about how humans relate to them and what dependencies this generates, whose interests are served by the construction of persons and property. Literature and AI research provide two complementary perspectives on these issues. Together they contribute to the development of a more nuanced theory of human-machine relationship which will consider its psychological reality, technological mediation, political structuring, and ethical consequences.

Works Cited

- Bowlby, John. Attachment and Loss. Vol. 1, Attachment. Basic Books, 1969. Braidotti, Rosi. The Posthuman. Polity, 2013.
- Chalmers, David J. The Conscious Mind: In Search of a Fundamental Theory. Oxford UP, 1996. Frankfurt, Harry G. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy*, vol. 68, no. 1, 1971, pp. 5–20.
- Freud, Sigmund. "The Uncanny." 1919. Translated by David McLintock, *The Uncanny*, Penguin Classics, 2003.
- Gibson, William. Agency. Berkley, 2020.
- Hayles, N. Katherine. How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics. U of Chicago P, 1999.
- Ihde, Don. Technology and the Lifeworld: From Garden to Earth. Indiana UP, 1990. Ishiguro, Kazuo. *Klara and the Sun*. Faber and Faber, 2021.
- Klein, Melanie. *Envy and Gratitude and Other Works, 1946–1963*. Hogarth Press, 1975.
- Lacan, Jacques. "The Mirror Stage as Formative of the Function of the I." *Écrits*, translated by Bruce Fink, W. W. Norton, 2006.
- McEwan, Ian. *Machines Like Me*. Jonathan Cape, 2019.
- Mori, Masahiro. "The Uncanny Valley." *IEEE Robotics & Automation Magazine*, vol. 19, no. 2, 2012, pp. 98–100.
- Nagel, Thomas. "What Is It Like to Be a Bat?" *The Philosophical Review*, vol. 83, no. 4, 1974, pp. 435–450.
- Newitz, Annalee. *Autonomous*. Tor, 2017.
- Strawson, Peter F. "Freedom and Resentment." *Proceedings of the British Academy*, vol. 48, 1962, pp. 187–211.
- Verbeek, Peter-Paul. *Moralizing Technology: Understanding and Designing the Morality of Things*. U of Chicago P, 2011.
- Wallach, Wendell, and Colin Allen. *Moral Machines: Teaching Robots Right from Wrong*. Oxford UP, 2009.
- Wilson, Daniel H. *Robopocalypse*. Doubleday, 2011. Winnicott, D. W. *Playing and Reality*. Routledge, 1971.