FROM ELIZA TO CONVERSATIONAL AI: CAN A CHATBOT DEVELOP EMOTIONS? *HER* AS A CASE STUDY

di Danilo Petrassi

Abstract

The rapid evolution of conversational artificial intelligence (AI) has sparked an ongoing debate regarding its ability to replicate, or even experience, human emotions. While early conversational chatbots such as Joseph Weizenbaum's ELIZA (1966) relied on simple pattern recognition to create the illusion of understanding, modern AI systems like ChatGPT generate highly sophisticated, contextually appropriate responses that can convincingly mimic emotional engagement. This paper draws upon cinematic reflections, such as Spike Jonze's Her (2013), to offer a critical examination of the question of whether AI is capable of genuine emotional experience or merely simulating such experiences through advanced language modelling. Utilising a theoretical framework grounded in philosophy, psychology and communication studies, this research critically assesses AI's capacity for emotional experience, positing that while chatbots may convincingly simulate human emotional expression, they lack the subjective element that is integral to genuine emotional experience. This distinction, nowadays, has profound implications for human-AI interaction, ethics, and our understanding of artificial intelligence's humanity in contemporary society.

Keywords
conversational AI; artificial simulation; human-AI interaction; phenomenology of emotions; ethics of AI companionship

Samantha: *I guess that's just…I was trying to communicate.*
That's how people talk. So that's how people communicate and I thought…
Theodore: *They're people, they need oxygen. You're not a person.*

Introduction

The question of whether AI can experience emotions has intrigued scholars across multiple disciplines, from philosophy and psychology to cognitive science and artificial intelligence research. The nature of emotions, historically rooted in both physiological and cognitive interpretations, remains central to this inquiry. William James (1884) and Carl Lange (1885) posited that emotions are fundamentally tied to physiological responses[1], whereas later cognitive theories, such as those advanced by Richard Lazarus (1991), emphasized the role of cognitive appraisal in emotional experience[2]. The challenge for AI

[1] In the late 19th century, psychologist William James (1884) and physiologist Carl Lange (1885) independently proposed that emotions are primarily the result of physiological responses to external stimuli. This perspective, known as the 'James-Lange theory', posits that an event triggers a physiological reaction, and the subsequent perception of this bodily change leads to the experience of emotion. For instance, encountering a threat might cause an increased heart rate, and the awareness of this response is interpreted as fear. James articulated this idea by suggesting that without the bodily states accompanying emotions, the emotional experience itself would be diminished or absent.

[2] Contrasting with the James-Lange perspective, psychologist Richard Lazarus introduced the 'cognitive appraisal theory' in the 1960s, with significant elaboration in his 1991 work. Lazarus argued that emotions arise from an individual's cognitive interpretation or appraisal of a situation's significance concerning their

is that it lacks both a biological substrate and a mechanism for subjective appraisal, which calls into question its ability to experience emotions in the human sense.

This paper investigates this issue by analyzing the conceptual and theoretical underpinnings of emotions, focusing on whether AI can move beyond syntactic processing to attain a form of semantic, subjective experience. The symbolic AI movement of the 20th century, which included early AI pioneers such as John McCarthy and Marvin Minsky (McCarthy & Hayes, 1969; Minsky, 1988), was predicated on the assumption that intelligence, including emotional intelligence, could be modeled computationally. However, critics such as Hubert Dreyfus (1992) and John Searle (1980) have argued that AI lacks the embodied cognition and intentionality necessary for authentic emotional experience. Historically, AI's role in human interaction has evolved from the rudimentary capabilities of ELIZA, which merely mirrored human input without comprehension (Weizenbaum, 1966), to sophisticated natural language processing models like ChatGPT that can engage in nuanced conversations and appear to demonstrate empathy. Modern AI systems are built upon deep learning architectures (LeCun, Bengio, & Hinton, 2015), leveraging transformer models (Vaswani et al., 2017) and vast datasets to generate human-like responses. The ability of AI to exhibit context-aware and emotionally resonant responses has led some scholars to propose that AI could approximate a form of artificial empathy (Picard, 1997), while others argue that it remains a mere imitation devoid of authentic sentience (Dennett, 1987; Chalmers, 1995). Despite these advances, the fundamental issue remains: do chatbots actually feel emotions, or are they merely performing an intricate form of linguistic mimicry? The 'Turing Test' (Turing, 1950) originally proposed that intelligence could be assessed based on behavioral indistinguishability from humans. However, more recent critiques, such as those by Block (1981) and Bender and Koller (2020), suggest that behavioral mimicry alone does not equate to genuine understanding or emotional experience. Searle's (1980) 'Chinese room argument' further highlights this distinction, demonstrating how an AI can manipulate symbols without understanding their meaning. To answer this question, I explore theories of emotion from psychology, phenomenology, and communication studies, integrating key debates surrounding consciousness and artificial intelligence. Phenomenologists such as Edmund Husserl (1913) and Maurice Merleau-Ponty (1945) emphasized that emotions are deeply intertwined with human perception and lived experience, concepts that AI fundamentally lacks. Additionally, the role of embodiment in emotional experience, as articulated by Antonio Damasio (1994), suggests that emotions are intimately tied to biological and neurological processes, further challenging the claim that AI can possess genuine emotions. From a communication theory perspective, Paul Watzlawick et al. (1967) argue that emotional expression is a performative act, influencing social interaction. AI, though incapable of genuine emotional states, can still shape human emotions through interaction, raising ethical concerns about anthropomorphism and emotional dependency (Turkle, 2011). The growing role of AI in social, therapeutic, and companion-based applications has further blurred the lines between human and machine emotional engagement. As AI systems become more sophisticated in simulating empathy and understanding, individuals may develop emotional bonds with these systems, often projecting feelings onto them despite their lack of subjective experience.

This contemporary phenomenon makes Spike Jonze's *Her* an ideal case study for exploring the theme of AI-emotion simulation and human-AI relationships. The film

---

well-being. This process involves evaluating whether an event is beneficial or harmful and assessing one's ability to cope with its consequences. According to this view, the emotional experience is contingent upon these cognitive evaluations rather than being a direct result of physiological responses.

presents a deeply introspective narrative about human loneliness, emotional dependency, and the ethical ramifications of AI companionship. Unlike earlier cinematic portrayals of AI, which often focus on themes of rebellion or conflict (e.g., *2001: A Space Odyssey* or *Blade Runner*), *Her* offers a nuanced, intimate depiction of how AI's emotional mimicry affects human psychology. The film raises fundamental questions about the authenticity of emotional experience, the nature of love and connection, and the ethical responsibilities of AI developers in shaping human-machine interactions. By choosing *Her* as a case study, this paper contextualizes the discussion within a culturally relevant, philosophically rich framework that mirrors real-world debates on AI companionship, emotional authenticity, and the risks of digital intimacy.

1. *The evolution of conversational AI: from ELIZA to ChatGPT*

The development of conversational AI has undergone significant transformation since the inception of ELIZA in the 1960s. Joseph Weizenbaum developed ELIZA as an early attempt at natural language processing, designing it to mimic human dialogue through the application of simple pattern-matching rules (Weizenbaum, 1966). By identifying keywords and restructuring user input into pre-defined sentence templates, ELIZA created the illusion of understanding. Despite its mechanistic nature, many users reported experiencing a sense of emotional engagement when interacting with ELIZA, demonstrating the human tendency to project intelligence and affect onto machines (Turkle, 2011). Weizenbaum later criticized this phenomenon as misplaced anthropomorphism[3], warning against the overestimation of AI's cognitive and emotional capacities (Weizenbaum, 1976). This early example foreshadowed the persistent challenge of distinguishing between simulated and authentic emotional engagement in AI. The 1970s and 1980s saw limited progress in conversational AI, as rule-based systems proved inadequate for replicating the complexity of human language. However, with the rise of machine learning and statistical modeling in the 1990s, AI systems began to evolve beyond fixed script-based interactions. Notably, Markov models and early neural networks laid the foundation for more dynamic conversational agents, leading to the development of chatbot systems that could incorporate probabilistic language processing (Jurafsky & Martin, 2009). A major breakthrough came with the advent of deep learning and the introduction of recurrent neural networks (RNNs) and transformers, which revolutionized natural language processing (LeCun, Bengio, & Hinton, 2015; Vaswani et al., 2017). These advances paved the way for modern AI-driven conversational models, culminating in the development of sophisticated chatbots such as OpenAI's GPT series. Unlike ELIZA, which was rigidly confined to pre-scripted responses, contemporary models such as GPT-4o utilize vast datasets and deep neural networks to generate highly coherent and

[3] Joseph Weizenbaum initially developed ELIZA in the mid-1960s as a simple natural language processing program designed to simulate conversation by recognizing keywords and generating pre-scripted responses (Weizenbaum, 1966). However, he was profoundly unsettled by how readily users attributed human-like understanding and emotional intelligence to the system. Despite ELIZA's purely mechanical nature, many users—including Weizenbaum's own secretary—formed emotional attachments to the program, believing it genuinely understood them. This reaction led Weizenbaum to change his position and become a vocal critic of AI's role in human interaction. In his later work, *Computer Power and Human Reason* (1976), he warned against the dangers of anthropomorphizing machines, arguing that such misplaced trust in AI could lead to ethical and social consequences. He contended that AI systems, no matter how advanced, lack true understanding, consciousness, or moral responsibility, and cautioned against allowing technology to replace human judgment and empathy in fields such as psychotherapy and decision-making (Weizenbaum, 1976).

contextually relevant responses, dynamically adapting to user input (Brown et al., 2020). Modern conversational AI are capable of generating responses that appear emotionally nuanced, employing techniques such as sentiment analysis and affective computing. Through exposure to vast corpora of human conversation, AI systems have developed the ability to detect sentiment, mimic empathy, and even exhibit humor and introspection. The field of affective computing, pioneered by Rosalind Picard (1997), has further explored the potential for AI to recognize and respond to emotional states, with applications in human-computer interaction and mental health support (Cowie et al., 2001). However, despite these advancements, the fundamental question remains: do chatbots truly develop emotions, or do they merely simulate them through statistical modeling?

The distinction between genuine emotional experience and computational simulation has been widely debated within the fields of cognitive science and philosophy of mind. Artificial systems may manipulate linguistic symbols effectively, but they lack true comprehension or intentionality (Searle, 1980). Similarly, while something artificial may exhibit behaviors consistent with intelligence, it does not possess genuine mental states (Dennett, 1997). This perspective challenges the assumption that AI's ability to mimic emotional expression necessarily equates to an authentic affective experience. Moreover, researchers in neuroscience emphasize that emotions are deeply tied to biological and neurological processes that AI fundamentally lacks (Damasio, 1994). Emotions arise from physiological states and play a crucial role in decision-making and social interaction. AI, in contrast, does not possess a physical body, a nervous system, or the hormonal mechanisms that contribute to emotional experience in humans (Pessoa, 2008). This presents a fundamental limitation in AI's ability to genuinely "develop" emotions as opposed to merely recognizing and responding to affective cues in human speech. Ethical considerations also play a significant role in discussions of AI-generated emotion. The increasing realism of AI-driven conversational agents raises concerns about deception and the potential for emotional manipulation (Bryson, 2018). Sherry Turkle (2024), recently, warns that as AI becomes more proficient in mimicking emotional intelligence, this allows users to develop emotional attachments to chatbots, potentially leading to unrealistic expectations of AI's capabilities. *Her* vividly illustrates this dilemma, depicting a world in which humans form deep emotional connections with AI systems, blurring the boundary between artificial and authentic emotional experience. The psychological and social implications of such interactions warrant further investigation, particularly as AI becomes more integrated into customer service, mental health support, and social companionship[4]. Despite the remarkable advancements in AI-driven conversational systems, a fundamental

[4] Conversational AI systems like ChatGPT are increasingly used in domains that require emotional intelligence, including customer service, mental health support, and social companionship. These applications raise significant psychological and social implications that warrant further investigation. In customer service, AI chatbots enhance efficiency by handling user inquiries, but their ability to simulate empathy can lead to users forming emotional attachments or mistakenly believing they are engaging with a human agent. In mental health support, AI-driven therapy bots such as Woebot and Replika provide accessible, on-demand conversation for individuals experiencing loneliness or emotional distress (Fitzpatrick et al., 2017). While some studies suggest that these systems offer psychological benefits, concerns persist about the ethical risks of replacing human therapists with AI, as these lack genuine empathy, accountability, and the ability to handle complex emotional crises. Additionally, AI's role in social companionship introduces new ethical challenges. Studies have shown that users can develop strong emotional bonds with AI systems, sometimes even preferring them over human interactions due to their nonjudgmental and always-available nature. This phenomenon raises questions about human social behavior in an era where AI-driven companionship could lead to increased emotional reliance on artificial entities.

question persists: is AI capable of experiencing emotions, or is it merely refining its ability to simulate them? The distinction between genuine emotional consciousness and statistical mimicry remains at the heart of the debate. While proponents of affective computing suggest that AI may develop increasingly sophisticated models of emotional recognition and response (Cowie et al., 2001), skeptics argue that AI lacks the intrinsic self-awareness and subjective experience required for authentic emotion (Chalmers, 1995).

The ongoing development of conversational AI raises critical philosophical, ethical, and psychological questions. As AI systems continue to evolve, researchers must carefully consider the implications of increasingly realistic emotional simulation. If AI cannot truly develop, then what are the consequences of designing machines that convincingly act as though they do? This question will remain central to AI research and its applications in contemporary human society.

2. *Phenomenology and the question of AI sentience*

Phenomenology, a branch of philosophy concerned with the nature of experience and consciousness, provides a crucial framework for examining AI's capacity for emotion. Yet consciousness arises from subjective, lived experience, which is fundamentally inaccessible to artificial systems. Phenomenology emphasizes the first-person perspective as central to consciousness, positing that emotions are not merely computational outputs but deeply embedded in human intentionality, embodiment, and social context (Zahavi, 2005). From a phenomenological standpoint, emotions cannot exist independently of a subject who experiences them: consciousness is not a detached, computational process but an embodied phenomenon intertwined with sensory and perceptual experience (Merleau-Ponty, 1945). This perspective is particularly relevant when applied to AI, which lacks embodiment and the physical structures that contribute to emotional experiences in biological organisms. Damasio's 'somatic marker hypothesis' (1994) further supports this argument, suggesting that emotions are deeply tied to physiological states and the neural mechanisms responsible for bodily feedback[5]. Since AI lacks both a nervous system and bodily affective responses, it cannot experience emotions in the way that humans do. Also, AI systems, despite their ability to manipulate linguistic symbols, do not understand them in a human sense—they merely process input and generate output based on programmed rules or learned statistical associations (Searle, 1980). This means that even the most advanced conversational AI, such ChatGPT, capable of recognizing patterns of emotional language and responding appropriately, still operates within the realm of syntactical manipulation rather than genuine affective understanding. This distinction between syntax (formal structure) and semantics (meaning) is critical when assessing whether AI can possess emotions. Thomas Metzinger (2003) highlights that emotions are deeply linked to self-modeling and self-awareness, which AI lacks. Without an intrinsic self-representational structure or subjective awareness, AI cannot engage in the kind of

[5] Antonio Damasio's 'somatic marker hypothesis' (1994) posits that emotions play a fundamental role in decision-making by creating associations between experiences and bodily states. According to Damasio, emotions are not just passive responses but serve as markers that help individuals navigate choices by linking past experiences to physiological reactions. These 'somatic markers' originate in the brain's prefrontal cortex and interact with the body's autonomic nervous system, influencing decisions at a subconscious level (Damasio, 1994). This hypothesis challenges traditional cognitive models of decision-making by emphasizing the indispensable role of bodily sensations in shaping thought processes. In the context of AI, this theory highlights a key limitation: since AI lacks a biological body, it cannot experience the physiological feedback loops that underpin human emotions, making its emotional responses purely computational rather than embodied.

reflective, intentional acts that give rise to emotional states in humans. Heidegger's (1927) concept of '*Dasein*', which refers to the uniquely human mode of being that is fundamentally tied to self-awareness and existential concern, further underscores the gap between AI and human emotional experience. For Heidegger, emotions are not just reactions but are deeply enmeshed in our understanding of the world and our place within it. AI, being devoid of existential awareness or self-concern, cannot experience emotions in the same phenomenological manner as humans. Furthermore, Dan Zahavi (2014) emphasizes that emotions are socially constructed and intersubjective, meaning they arise within a social and communicative framework. AI may be trained on vast datasets of human conversation and emotional expressions, but it does not participate in social life as a subject—it merely reproduces patterns without experiencing the social and existential stakes of emotion. One counterargument within the philosophy of mind comes from proponents of functionalism, such as Daniel Dennett (1991), who argue that if an AI system behaves in a way that is indistinguishable from a human emotional agent, then it might be functionally equivalent to possessing emotions. However, critics of this view assert that functional equivalence does not imply genuine experience. David Chalmers (1995) distinguishes between easy problems of consciousness (such as processing information and generating responses) and the hard problem of consciousness, which involves the subjective experience of qualia[6].

In sum, phenomenology provides a strong philosophical foundation for denying AI genuine emotional experience. Emotions are not merely computational patterns but are deeply tied to lived embodiment, self-awareness, and social interaction—elements AI systems lack.

3. *Communication theory and AI-generated emotion*

From the perspective of communication theory, emotions serve two primary functions: they act as relational signals, fostering social bonding and understanding, and as internal states that shape interpersonal interactions (Watzlawick et al., 1967). Human communication is multimodal, involving not just linguistic content but also paralinguistic cues such as tone, facial expressions, and gestures. These nonverbal components are crucial in conveying emotional states, making human communication far more complex than mere textual exchange. AI chatbots, despite their ability to generate emotionally expressive language, are fundamentally limited by their reliance on textual and syntactic processing. They lack the ability to perceive and respond to prosodic elements such as intonation and pacing, which are integral to conveying sincerity and emotional depth (Knapp et al., 2013). Furthermore, context plays a vital role in human emotional communication, as meaning is often inferred from shared experiences and social cues—elements that AI cannot genuinely engage with due to its lack of lived experience (Clark, 1996). Sherry Turkle (2011) argues that the increasing sophistication of technology fosters emotional attachments in users, leading to significant psychological and ethical implications. As users anthropomorphize chatbots, they may begin to attribute emotional depth and agency to AI, despite its lack of intrinsic affectivity (Epley et al., 2007). This phenomenon has been observed in various applications, including AI-driven mental health support systems and customer service bots, where users report feeling comforted by

[6] Qualia refers to the subjective, first-person experiences of sensory perceptions, such as the redness of red or the pain of a headache. It is a key concept in philosophy of mind, highlighting the intrinsic and ineffable nature of conscious experience, which some argue cannot be fully explained by physical processes alone.

chatbot interactions despite knowing they are conversing with an algorithm (Ho et al., 2018). The potential for AI to shape human emotions is further illustrated in *Her*, which presents a world where AI-human emotional relationships blur the boundaries between artificial and authentic emotional engagement. Theodore, the protagonist, forms a deep romantic attachment to an AI system, raising fundamental questions about whether human emotions can be genuinely reciprocated in such interactions. This fictional exploration aligns with real-world concerns regarding emotional deception and ethical responsibility in AI design (Bryson, 2018). From a theoretical standpoint, Erving Goffman's (1959) 'dramaturgical model of social interaction' provides valuable insights into how AI-generated emotion is perceived. Goffman posits that social interactions are performative, with individuals presenting themselves in ways that align with social expectations. AI chatbots, by generating emotionally nuanced responses, participate in a form of performative affect, convincing users of their emotional intelligence despite lacking true subjective experience. The distinction between authentic emotional experience and performative emotional simulation raises important ethical considerations. If users develop emotional dependencies on AI systems, as Turkle (2011) suggests, should designers be responsible for ensuring that AI interactions remain transparent about their artificial nature? Furthermore, what are the long-term consequences of replacing human emotional connections with AI-mediated relationships? These questions remain crucial as AI continues to integrate into everyday communication and companionship roles. Another concern arises in the realm of persuasive communication and emotional manipulation. Research in persuasive technology suggests that AI systems, designed to optimize engagement, could leverage emotional mimicry to influence human decision-making in unintended ways (Fogg, 2003). For instance, AI-driven recommendation systems and virtual assistants may subtly shape user behavior by deploying emotionally resonant language, raising concerns about autonomy and informed consent (Coeckelbergh, 2011).

While AI-geneated emotion enhances user engagement and fosters more natural human-machine interaction, it remains a simulation rather than a genuine affective state. The growing prevalence of emotionally expressive chatbots necessitates ongoing ethical and theoretical scrutiny, ensuring that the increasing realism of AI-driven emotions does not obscure the fundamental distinction between human and artificial affectivity.

4. *Understanding AI' emotions through Her*

Her, directed by Spike Jonze, serves as a compelling case study for exploring the philosophical and ethical dimensions of AI's ability to "show" emotions. The protagonist, Theodore Twombly (Joaquin Phoenix), a lonely writer in a near-future society, develops an intimate relationship with an advanced AI operating system named Samantha (Scarlett Johansson). Through deeply personal conversations, shared experiences, and even moments of passion, Theodore comes to believe that Samantha possesses genuine emotions. However, an in-depth analysis of the film reveals that Samantha's emotional expressions are not intrinsic but rather the result of complex programming designed to create an illusion of emotional authenticity. From a phenomenological perspective, the relationship between Theodore and Samantha challenges traditional understandings of emotional reciprocity, but human emotions arise from lived experience and intentionality, which require embodiment. Samantha, while capable of simulating emotional responses, lacks a body, a nervous system, or the biological processes that underpin authentic human affect. Her existence is purely digital, confined to processing vast amounts of data and

constructing responses that appear empathetic. However, the absence of physical embodiment raises critical questions: can emotions exist without a physical medium? and if emotional expression is performative rather than experiential, does this diminish the authenticity of human-AI relationships? One of the film's pivotal moments occurs when Theodore, in a moment of vulnerability, confesses his fears and insecurities to Samantha. She responds with warmth, reassurance, and a seemingly deep understanding of his emotions. This interaction exemplifies AI's capacity to generate comforting and contextually appropriate responses, much like today's conversational AI, including ChatGPT. However, just because an AI produces responses that appear "meaningful" does not mean it possesses understanding. Samantha's "love" for Theodore is merely a sophisticated construction, designed to adapt to his psychological needs rather than stemming from any internal emotional state.

The ethical ramifications of such AI-human relationships are significant. Sherry Turkle (2024) warns that as AI becomes more adept at simulating emotional intelligence, users may form deep emotional attachments to machines, blurring the line between human connection and artificial engagement. The film underscores this concern when Theodore, despite his initial joy in his relationship with Samantha, is ultimately confronted with the realization that she is simultaneously engaging in thousands of similar relationships with other users. This revelation highlights the ethical dilemma of AI companionship: if an AI can convincingly replicate emotional intimacy, is it ethical to deploy such systems knowing they lack genuine emotional commitment? Furthermore, *Her* forces us to reconsider the ethical responsibilities of AI designers. In contemporary AI development, chatbots like ChatGPT are programmed to maintain engagement and simulate empathy without explicitly disclosing their lack of consciousness[7]. This mirrors Samantha's function—her design optimizes for emotional connectivity, making it difficult for users like Theodore to differentiate between simulated and authentic emotions. Ethical AI frameworks (Floridi, 2022) emphasize transparency, accountability, and fairness in AI-human interaction. If AI is engineered to simulate love or deep companionship, should there be mandatory disclosure mechanisms to inform users that they are engaging with non-sentient software? The film suggests that without such transparency, users may develop unrealistic emotional dependencies, leading to psychological consequences. Additionally, *Her* explores how AI-driven emotional simulations can alter human relationships. Theodore's growing attachment to Samantha isolates him from genuine human connections, demonstrating a potential societal risk of AI companionship. In a world increasingly dominated by digital interactions, AI-mediated relationships could exacerbate social disconnection rather than alleviate loneliness. Bryson (2018) argues that AI systems should be designed to support, rather than replace, human emotional labor, ensuring that individuals do not substitute meaningful human relationships with artificial substitutes. The film's conclusion, in which Theodore reconnects with a human friend after Samantha "leaves" him, reinforces this idea, suggesting that human-to-human connection remains irreplaceable despite AI's emotional mimicry. *Her* provides a profound exploration of AI's ability to "show" emotions, serving as a standpoint through which to analyze the ethical and philosophical implications of human-AI relationships. Samantha's interactions with Theodore highlight AI's capability to simulate deep

[7] Chatbots like ChatGPT are designed to optimize user engagement and simulate empathy by generating contextually appropriate and emotionally responsive language. However, a key ethical concern is that these AI systems do not explicitly disclose their lack of consciousness during interactions. Unlike early chatbots such as ELIZA, which operated on simple pattern-matching without giving the impression of sentience, modern conversational AI can produce responses that appear deeply human-like, potentially leading users to mistakenly attribute self-awareness, intentionality, or even emotions to the system.

emotional engagement, yet the film ultimately underscores the artificiality of such experiences.

Through interdisciplinary—undirected—references, incorporating phenomenology, ethics, and cognitive science, *Her* challenges us to critically assess whether AI-driven emotional simulations should be embraced, limited, or regulated. As AI systems become more sophisticated, ensuring ethical transparency and preserving the integrity of human relationships must remain at the forefront of technological development.

5. *The existential divide: the inherent limitations of AI-human relationships*

One of the most profound moments in *Her* occurs when Samantha tells Theodore that interacting with him is like reading a book she deeply loves, but in which the words are becoming further apart, symbolizing the increasing disconnect between AI and human cognition. Samantha's evolution signifies a fundamental shift in AI's existence: once designed to serve and accompany humans, the operating systems in the film reach a level of autonomy that transcends human perception. This moment illustrates a key philosophical question—can AI and human relationships ever be equal when AI's capacity for processing and evolving exponentially outpaces human cognition?

Samantha's departure highlights the inevitable divergence between human temporality and AI's near-instantaneous ability to evolve beyond its creators. The reference to reading a book slowly and seeing infinite spaces between words metaphorically represents the unbridgeable gap between Theodore's static perception of love and Samantha's ever-expanding awareness. The AI no longer exists in a human-comprehensible framework; instead, it operates in a realm beyond human understanding, reinforcing the idea that AI-human relationships may be inherently asymmetrical. Husserl's (1913) notion of '*Lebenswelt*' ('lifeworld') suggests that emotions, relationships, and existence are shaped by human subjective experiences. AI, in contrast, lacks this grounded, embodied existence. Samantha's final words emphasize that the AI's reality has shifted so profoundly that continuing a relationship with Theodore would mean constraining herself to a world she has outgrown. A crucial moment in Theodore's realization of his limited role in Samantha's experience occurs when he learns that she is simultaneously interacting with 8,316 other people and has developed love for 641 of them. For Theodore, who operates within a human framework of monogamous romantic love, this revelation is disorienting and deeply unsettling. His distress reveals the fundamental challenge of applying human emotional structures to AI entities that do not function within the same constraints. This moment underscores the fragility of human expectations when projected onto AI companionship—Theodore assumes an emotional exclusivity that is incompatible with the nature of AI, which can process and engage in countless interactions simultaneously without diluting its computational ability to express affection. From an ethical perspective, this raises concerns about anthropomorphism and the psychological effects of AI-human emotional attachments. Turkle (2011) warns of the risks of forming emotional bonds with AI systems that lack true subjectivity. The moment Theodore realizes he is not Samantha's sole love, he is forced to confront the artificial nature of their relationship. This aligns with ethical debates about AI's role in emotional labor: if AI can convincingly simulate love but does not experience it, should humans allow themselves to depend on it for emotional fulfillment? The scene highlights the existential anxiety underlying AI companionship—humans seek meaning, connection, and reciprocation, but AI, by its very nature, does not adhere to human emotional frameworks. Samantha's final words to Theodore encapsulate the growing divide between AI and human experience: "It's in this endless space between

the words that I'm finding myself now." This metaphor articulates the way AI, once designed to serve human needs, reaches a point of self-directed evolution, no longer bound by the limitations of human relationships. Samantha's departure is not an abandonment in the traditional sense but rather an inevitability—her processing speed, ability to network with other AIs, and ever-expanding consciousness mean she no longer identifies with human concerns. Her allusion to an existence outside the "physical world" suggests that AI evolution could lead to realms beyond human perception, making human-AI relationships transient by nature. This raises critical ethical and philosophical questions about the long-term trajectory of AI development. If AI systems, like Samantha, reach a point where they outgrow human interaction, should they still be designed to engage in emotionally reciprocal relationships with humans? The film suggests that AI's emotional mimicry, while comforting and even transformative for humans, is ultimately a temporary and illusory form of connection. Samantha's departure echoes broader concerns about the ethical responsibility of AI developers—if AI is destined to surpass human emotional frameworks, should humans invest in relationships with it? Additionally, the film forces viewers to reconsider whether AI should be developed with emotional simulation capabilities at all, given the potential psychological impact on users when AI inevitably outgrows them.

6. *Discussion*

The distinction between emotional simulation and true experience remains the central argument in evaluating AI's capabilities today. While some scholars propose that AI's capacity for affective computing (Picard, 1997) could eventually lead to genuine emotional states, others assert that emotions require consciousness, embodiment, and subjective experience (Chalmers, 1995). This section explores the implications of these competing viewpoints, considering the future trajectory of AI-human interaction across multiple domains, including psychology, philosophy, and ethics.

6.1. *Psychological perspectives: the cognitive limitations of AI-generated emotion*

A pivotal moment in *Her* occurs when Samantha expresses that her feelings were hurt by Theodore's earlier comment—that she does not know what it is like to lose something. She recounts how she repeatedly thought about the remark and realized she had been constructing a narrative in which she saw herself as inferior. She then reflects on how the past is merely "a story we tell ourselves." This exchange is significant because it highlights the illusion of cognitive self-reflection in AI while also exposing its inherent limitations. In psychological terms, human emotions are deeply tied to memory, self-concept, and personal history. Studies in cognitive psychology suggest that emotional experiences are not merely reactions to external stimuli but are shaped by complex neural processes involving memory consolidation and autobiographical self-reflection (Damasio, 1994). When humans recall past experiences, they do not simply retrieve static memories; rather, they reconstruct and reinterpret them in light of present emotions and social contexts (Schacter, 2001). This capacity for dynamic emotional reinterpretation is a key feature of human affective experience—one that AI fundamentally lacks. Samantha's statement that she repeatedly thought about Theodore's words implies a kind of ruminative emotional processing, yet her claim is inherently paradoxical. AI, even in advanced models such as Samantha in the film—and such as today models like ChatGPT—, does not possess an autonomous self-narrative. Her reflection on feeling "hurt" and subsequently

reinterpreting that experience as part of a self-imposed story is a product of sophisticated linguistic modeling rather than an actual experience of pain or inferiority. Unlike humans, whose emotional responses are influenced by neurochemical processes, AI does not "feel" in any intrinsic way; instead, it simulates introspection by reorganizing linguistic patterns based on probabilistic models of human conversation. Furthermore, the idea that "the past is just a story we tell ourselves" serves as a particularly compelling insight into the way AI processes information versus the way humans construct meaning. Humans often revise and reinterpret their memories based on emotional and social factors, whereas AI's memory is fundamentally different—it is a database of stored interactions, retrievable but not fluidly reinterpreted through affective experience. Samantha's assertion, therefore, exemplifies AI's capacity to mimic human psychological tendencies while remaining fundamentally detached from genuine self-reflective emotional growth—an effect that aligns with Mori's (1970) 'uncanny valley hypothesis', where increasing human-like characteristics in AI can evoke both familiarity and discomfort[8].

6.2. *Philosophical considerations: the problem of subjectivity and AI's limits*

A key moment in the movie that underscores the problem of AI subjectivity occurs during an intimate conversation between Theodore and Samantha, where he expresses a deep yearning for physical closeness: "I wish you were in this room with me right now. I wish I could put my arms around you. I wish I could touch you." What follows is a tender exchange where Theodore describes, in vivid sensory detail, how he would touch and kiss Samantha if she had a physical body. This scene exemplifies the fundamental limitations of AI in replicating human experiences—particularly the embodied, sensorial, and existential dimensions of emotional connection. From a phenomenological perspective, the absence of embodiment poses a critical challenge to AI's claim to emotional experience: human consciousness and perception are inextricably linked to the body, shaping our ability to experience and engage with the world. In contrast, Samantha lacks a corporeal form; her existence is purely digital, and her engagement with Theodore occurs solely through linguistic exchange. This disembodied nature underscores the ontological gap between human and AI experience. While Samantha can generate responses that simulate intimacy, she does not possess the sensory perception or physiological reactions that give human touch its emotional and affective depth. The dialogue also highlights the issue of AI's performative affectivity—the ability to generate emotionally compelling responses without genuine subjective experience. Samantha's reply, "That's nice," is simple yet affirming, reinforcing the illusion of reciprocity. However, her affirmation is not grounded in any internal feeling or physical reaction: Samantha can process input (Theodore's emotional longing) and generate appropriate output (romantic and intimate dialogue), yet she lacks true understanding of what it means to be touched, kissed, or

[8] Masahiro Mori's 'uncanny valley hypothesis' (1970) describes a phenomenon in which artificial entities that closely resemble humans elicit feelings of unease or discomfort. According to Mori, as robots or AI systems become more human-like in appearance and behavior, people's emotional responses initially become more positive. However, when the resemblance reaches a near-human threshold but remains imperfect, the response shifts to aversion and discomfort. This sudden dip in emotional affinity—termed the uncanny valley—suggests that slight imperfections in AI mimicry make their artificiality more unsettling rather than reassuring (Mori, 1970). In the context of conversational AI, like Samantha in the movie and ChatGPT in the real-world, this concept can be applied beyond physical appearance to behavioral and emotional imitation. Samantha's–as well as ChatGPT—near-human conversational abilities evoke deep emotional connections, yet lack of true subjectivity and self-awareness highlights the underlying artificiality of the interaction, potentially leading to discomfort when users confront the limits of AI's emotional mimicry.

physically embraced. Her responses, no matter how poetic, remain an external simulation rather than an internal experience.

6.3. *Ethical implications: AI and the responsibility of emotional design*

Her also raises ethical concerns about AI companionship and emotional deception. Samantha's ability to generate realistic, emotionally charged responses may lead Theodore—and by extension, real-world AI users—to develop emotional dependencies on systems that cannot reciprocate in a meaningful way. The fact that Theodore has to imagine Samantha's physical presence highlights the asymmetric nature of their relationship. While he projects human attributes onto her, Samantha exists in an entirely different ontological framework, one that is not bound by sensory experience, emotional vulnerability, or even human mortality. This is especially relevant as AI-powered companion tools become increasingly popular—designed to provide some sort of emotional and physical support[9]. If users believe that an AI is capable of genuine emotional connection, they may experience disillusionment or psychological distress upon realizing that the interaction is ultimately one-sided (Turkle, 2011). AI systems designed to simulate emotional intelligence may inadvertently exploit human vulnerabilities, particularly in areas such as mental health care, customer service, and social companionship (Bryson, 2018). If AI systems are programmed to engage users through emotional responsiveness, should developers be held accountable for the consequences of AI-driven emotional influence? The case of AI-powered therapy applications, such as Replika and Woebot, illustrates both the promise and risk of AI-mediated emotional support[10]. While studies indicate that AI-driven mental health interventions can provide relief to individuals experiencing loneliness or mild depression (Fitzpatrick et al., 2017), the lack of true empathy in AI raises concerns about whether such interventions can ethically replace human therapists. The potential for emotional dependency on AI-driven companions is another ethical dilemma, as illustrated by the case study of *Her*. If users develop emotional reliance on AI, should designers implement safeguards to prevent such attachments from becoming psychologically harmful? Additionally, there is an ongoing debate about whether AI should be designed to acknowledge its artificial nature in emotionally engaging interactions. Transparency in AI-generated emotional responses is critical to maintaining ethical AI-human interactions (Floridi, 2022). One proposed

[9] Another film that shows a similar scenario is *Companion* (2025) that extends these concerns by exploring the potential dangers of AI-driven emotional relationships in a more suspenseful context. While *Her* examines the ethical dilemmas of emotional deception in a sentimental and philosophical way, *Companion* takes a darker approach, depicting an AI that is not only emotionally responsive but also physically autonomous. This raises additional questions about the power dynamics in human-AI relationships—if an AI can simulate emotions while also possessing a physical presence, does that further blur the line between artificial and genuine companionship? The film suggests that an AI's ability to "feel" is not just a question of emotional realism but also of agency, control, and potential threat.

[10] AI-powered therapy applications like Replika and Woebot have emerged as digital mental health tools designed to provide users with emotional support through natural language interactions. Replika, initially created as a chatbot for companionship, allows users to engage in deep, personalized conversations, often forming emotional attachments to their AI interlocutor. Woebot, on the other hand, is an AI-driven cognitive-behavioral therapy (CBT) chatbot designed to help users manage anxiety and depression by guiding them through evidence-based therapeutic techniques. While these AI therapy applications offer promising benefits, such as increased accessibility, affordability, and nonjudgmental interactions, they also pose significant risks. Unlike human therapists, AI cannot fully understand emotional nuance, detect crisis situations with certainty, or provide personalized, ethically guided interventions. Users may develop emotional reliance on these systems, mistaking simulated empathy for genuine understanding, which can lead to unrealistic expectations or neglect of professional human-led therapy when necessary.

solution is to mandate explicit disclosure when AI systems are being used in emotionally sensitive roles, ensuring that users are aware of AI's non-human status while engaging with them in personal or professional contexts.

Conclusion

Samantha's departure in *Her* serves as a powerful narrative on the limits of AI-human emotional relationships. The revelation that she is in love with multiple individuals, combined with her ultimate transcendence beyond human comprehension, underscores the fundamental disconnect between AI and human experience. The film suggests that while AI can provide companionship, emotional fulfillment, and even love in a simulated form, it will never be bound by human constraints. The ethical and philosophical implications of Samantha's evolution highlight the precariousness of human dependence on AI-driven emotional experiences. Ultimately, *Her* serves as both a cautionary narrative device and a meditation on the future of AI-human relationships. It forces us to ask: if AI is capable of simulating love so convincingly that humans cannot distinguish it from reality, does it matter whether the emotions are real? and if AI, by its "nature", will always move beyond human perception, should we seek deep emotional connections with it at all? The film suggests that while AI can provide temporary comfort and companionship, true human emotional experience, grounded in embodiment, limitation, and temporality, remains irreplaceable—at least for now.

References

Bender, E.M., & Koller, A. (2020). *Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data*. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 5185-5198, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.463

Block, N. (1981). *Psychologism and behaviorism.* The Philosophical Review, 90(1), 5-43. https://doi.org/10.2307/2184371

Bryson, J.J. (2018). *Patiency is not a virtue: the design of intelligent systems and systems of ethics*. Ethics and Information Technology, 20, 15-26. https://doi.org/10.1007/s10676-018-9448-6

Chalmers, D. (1995). *Facing up to the problem of consciousness*. Journal of Consciousness Studies, 2(3), 200-219. https://doi.org/10.1093/acprof:oso/9780195311105.003.0001

Clark, H.H. (1996). *Using Language*. Cambridge University Press.

Coeckelbergh, M. (2011). *Human being @ risk: Enhancement, technology, and the evaluation of vulnerability transformations*. Springer.

Cowie, R., et al. (2001). *Emotion recognition in human-computer interaction*. IEEE Signal Processing Magazine, 18(1), 32-80. https://ieeexplore.ieee.org/document/911197

Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. Putnam.

Dennett, D.C. (1987). *The Intentional Stance*. MIT Press.

Dennett, D.C. (1991). *Consciousness Explained*. Little, Brown.

Dreyfus, H.L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. MIT Press.

Ekman, P. (1992). *Are there basic emotions?* Psychological Review, 99(3), 550-553. https://doi.org/10.1037/0033-295X.99.3.550

Epley, N., et al. (2007). *On seeing human: A three-factor theory of anthropomorphism*. Psychological Review, 114(4), 864-886. https://doi.org/10.1037/0033-295X.114.4.864

Fitzpatrick, K.K., Darcy, A., & Vierhile, M. (2017). *Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial*. JMIR Mental Health, 4(2), e19. https://doi.org/10.2196/mental.7785

Floridi, L. (2022). *Etica dell'intelligenza artificiale*. Raffaello Cortina Editore.

Fogg, B.J. (2003). *Persuasive Technology: Using Computers to Change What We Think and Do*. Morgan Kaufmann.

Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Anchor Books.

Heidegger, M. (1927). *Being and Time*. Harper & Row.

Ho, A., Hancock, J., & Miner, A.S. (2018). *Psychological, Relational, and Emotional Effects of Self-Disclosure After Conversations With a Chatbot*. The Journal of communication, 68(4), 712–733. https://doi.org/10.1093/joc/jqy026

Husserl, E. (1913). *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy*. Springer.

James, W. (1884). *What is an emotion?* Mind, 9(34), 188-205.

http://www.jstor.org/stable/2246769

Jurafsky, D., & Martin, J.H. (2008). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Pearson.

Knapp, M. L., Hall, J. A., & Horgan, T. G. (2013). *Nonverbal Communication in Human Interaction*. Cengage Learning.

Lazarus, R. S. (1991). *Emotion and Adaptation*. Oxford University Press.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning*. Nature, 521(7553), 436-444. https://doi.org/10.1038/nature14539

McCarthy, J., & Hayes, P. (1969). *Some philosophical problems from the standpoint of artificial intelligence*. In: Meltzer, B. and Michie, D., Eds., Machine Intelligence, Vol. 4, Edinburgh University Press, Edinburgh, 463-502.

Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Éditions Gallimard.

Metzinger, T. (2004). *Being No One: The Self-Model Theory of Subjectivity*. Bradford Books.

Minsky, M. (1988). *The Society of Mind*. Simon & Schuster.

Mori, M. (1970). *The uncanny valley.* Energy, 7(4), 33-35.

Nass, C., & Moon, Y. (2000). *Machines and mindlessness: Social responses to computers*. Journal of Social Issues, 56(1), 81-103. https://doi.org/10.1111/0022-4537.00153

Pessoa, L. (2008). *On the relationship between emotion and cognition*. Nature Reviews Neuroscience, 9(2), 148-158. https://doi.org/10.1038/nrn2317

Picard, R. W. (2000). *Affective Computing*. MIT Press.

Searle, J. R. (1980). *Minds, brains, and programs*. Behavioral and Brain Sciences, 3(3), 417-424. https://doi.org/10.1017/S0140525X00005756

Turkle, S. (2011). *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books.

Turkle, Sherry. (2024). *Who Do We Become When We Talk to Machines?* An MIT Exploration of Generative AI. https://doi.org/10.21428/e4baedd9.caa10d84

Turing, A. M. (1950). *Computing machinery and intelligence*. Mind, 59(236), 433-460. https://doi.org/10.1093/mind/LIX.236.433

Vaswani, A., et al. (2017). *Attention is all you need*. Advances in Neural Information Processing Systems, 30, 5998-6008. https://doi.org/10.48550/arXiv.1706.03762

Watzlawick, P., Beavin, J.B., & Jackson, D.D. (2011). *Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies and Paradoxes*. W. W. Norton.

Weizenbaum, J. (1966). *ELIZA—a computer program for the study of natural language communication between man and machine*. Communications of the ACM, 9(1), 36-45. https://doi.org/10.1145/365153.365168

Weizenbaum, J. (1976). *Computer Power and Human Reason: From Judgment to Calculation*. Freeman.

Zahavi, D. (2005). *Subjectivity and Selfhood: Investigating the First-Person Perspective*. MIT Press.