Maria Ioana Oancea

Frankenstein or Pygmalion? Literary Tradition and the Reception of Artificial Intelligence

Abstract: This article explores the impact of literary representations on the reception of Artificial Intelligence. Beyond the science fiction tropes that precede AI lies the deeper archetypal theme of the rebellion of creation against its creator, starting with the figure of Lucifer. Our hypothesis is that Artificial Intelligence, appearing in literature as both utopia and dystopia, is seen as a threat only when it challenges the definition of being human. This research analyzes the border between human and non-human and the relationship between creator and creation in literature. Drawing on the theoretical framework of posthumanist theories and Utopian/Dystopian studies, as well as the responses of AI itself, the paper identifies the traits that cause a robot to be seen as a threat, while also exploring the efforts to redefine what it means to be human as a result of interaction with AI.

Keywords: Artificial Intelligence (AI); Literary Archetypes; The Other; Creator-Creation Dynamic; Posthumanism; Utopia/ Dystopia.

Maria Ioana Oancea
"G. Călinescu" Institute of Literary History and Theory, Bucharest, Romania
maria.i.oancea@gmail.com

DOI: 10.24193/cechinox.2025.49.18

Introduction and Methodology

The public discourse surrounding Artificial Intelligence (AI) has proven particularly widespread and emotionally charged compared to previous technological innovations. AI has raised intense reactions across all demographics – overflowing its specific domain and becoming a subject of debate for the humanities as well as the sciences, not to mention eliciting unusually heated opinions among the general public. By penetrating into all social spheres, it has managed to become a landmark that distinguishes a particular epoch, a place that only a select few inventions have secured throughout history. Reactions to AI range from utopian enthusiasm to existential dread. What distinguishes AI from other technological breakthroughs is not its technical sophistication, but the fact that it activates archetypal narratives embedded in human culture, triggering deeply rooted fears, hopes and expectations far beyond what its current capabilities would justify. Our central argument is that AI inherits meaning from two overlapping cultural archetypes: the figure of the robot

from science fiction literature, and the far older motif of creation rebelling against its creator.

While the robot archetype addresses practical and ethical questions about non-human intelligence, it alone cannot account for the intensity of contemporary fears. The robot in science fiction often appears as a benign helper or loyal servant – which is also the current status of AI. Classic examples include Isaac Asimov's positronic robots[1] (programmed with the fundamental mission to never harm a human) and iconic film companions like the droids C-3PO and R2-D2 (loyal, multi-purpose assistants in the Star Wars universe). This persistent trope of the obedient and non-threatening machine directly shapes our contemporary expectations for AI as a tool of service.

However, alongside enthusiasm for exploring the benefits of this new technology persists a profound anxiety that transcends rational assessment of present capabilities. This anxiety stems from AI's perceived potential to develop self-awareness. It is likely that this intelligence, already more powerful and efficient in several areas than that of man, will demand autonomy and recognition as a rights-bearing entity – not to mention the ethical question as to whether it may actually deserve those rights. This scenario activates a much older and more troubling paradigm: the rebellion of creation against creator. The association with this archetypal narrative of transgression and rebellion would explain the surprising and apparently unreasonable demonization of this intelligent invention.

This article examines how literary and cultural narratives shape contemporary perceptions of AI, focusing on two paradigmatic texts about animated creation: Ovid's *Metamorphoses* and its account of Pygmalion, and Mary Shelley's *Frankenstein; or, The Modern Prometheus*. These narratives represent opposing outcomes of the creator-creation relationship and together they establish a framework for understanding when and why artificial beings are perceived as threatening. We supplement this analysis with references to Philip K. Dick's *Do Androids Dream of Electric Sheep?*, which explicitly addresses the problem of distinguishing human from non-human consciousness.

Our theoretical approach draws on posthumanist theory, particularly N. Katherine Hayles's work on the relationship between embodiment, consciousness, and artificial intelligence, as well as studies on utopian and dystopian literature. We also incorporate responses generated by AI systems themselves – generative AI Chatbots (Multimodal Large Language Models)[2] – when questioned about these themes, treating AI not merely as an object of study but as a potential interlocutor in debates about consciousness and humanity.

The central question guiding this research is: at what point does creation become threatening? Our hypothesis is that Artificial Intelligence – present in both utopian and dystopian imaginaries – is perceived as dangerous only when it challenges fundamental definitions of human identity. The article therefore investigates the boundaries between human and non-human in literature, examining how the creator-creation dynamic shifts from benevolent to adversarial. This analysis aims to identify which supposedly human traits cause artificial beings to be viewed as threats, and how engagement with AI is

forcing contemporary culture to renegotiate what it means to be human.

The Promethean Condition of the Creator

Human creation coming to life has always been an exciting and much-desired utopia. Descriptions of the creator's emotional state during these moments are strikingly (and eerily) similar in two very different creators: Pygmalion and Frankenstein. Mary Shelley named her novel *Frankenstein; or, The Modern Prometheus*. It is Prometheus rather than Pygmalion that the author associates her character's experience with. While best known for giving fire to humanity, Prometheus is also credited – as in Ovid's *Metamorphoses* – with creating mankind from clay and water. What separates the two creation myths and indeed places Victor Frankenstein in the line of Prometheus is punishment, as well as the transgressive nature of the creation. While the outcome differs drastically, the initial feverish enthusiasm is the same.

In her novel, Mary Shelley describes this particular state as "delight and rapture"[3] and "supernatural enthusiasm"[4]. Ovid's Pygmalion, on the other hand, is caught between hope and the fear of disappointment. It seems to him that his creation is alive, but he has felt that way from the very beginning. He barely dares to phrase his plea to Venus to make his love come to life, so extraordinary – and, we might add, taboo – such a request appears to him, because it tempers with something that is above human reach. The following fragments from Ovid's *Metamorphoses* show Pygmalion's desire for Galatea to come to life before and after Venus grants his wish: "Often he lifts his hands to the work to try whether it be flesh or ivory; nor does he yet confess it to be ivory. He kisses it and thinks his kisses are returned. He speaks to it, grasps it and seems to feel his fingers sink into the limbs when he touches them; and then he fears lest he leave marks of bruises on them"[5].

A similar passage appears after the intervention of Venus:

> She seemed warm to his touch. Again he kissed her, and with his hands also he touched her breast. The ivory grew soft to his touch and, its hardness vanishing, gave and yielded beneath his fingers, as Hymettian wax grows soft under the sun and, moulded by the thumb, is easily shaped to many forms and becomes usable through use itself. The lover stands amazed, rejoices still in doubt, fears he is mistaken, and tries his hopes again and yet again with his hand[6].

Then comes the awakening, a key point in both Ovid's Pygmalion and Mary Shelley's novel. The two creators react very differently in this moment. Granted, the aesthetic reason comes to mind, our hypothesis however is that this is not merely a matter of beauty. This dichotomy between utopian hope and dystopian anxiety is characteristic of speculative fiction about artificial beings. Victor Frankenstein hopes for his experiment to be successful, yet when it proves successful, he shrinks away in horror, terrified by its reality. Indeed, the shift from utopia to dystopia often appears when implementing the utopian ideal, as Lyman Tower Sargent observes: "Utopians are always faced with this dilemma when

they attempt to move their dream to reality – is their dream compatible with the imposition of their dream"[7]? In this case dystopia emerges not as the opposite of utopia but as its flawed or excessive realization. Sargent addresses "the contradictory nature of utopianism"[8] – what appears as utopian perfection from one perspective becomes dystopian oppression from another.

A decisive element contributing to the perceived monstrosity of the creation resides not in the actual traits of the created being but in the very fact that it is man-made. The collective imaginary links the *hybris* inherent in such an undertaking to an imminent punishment. As Lyman Tower Sargent points out regarding the implementation of utopias, "Utopia can be like Greek tragedy. Humanity in its pride commits utopia and in doing so violates the boundaries of its allotted sphere. Therefore, it must confront nemesis, fail to achieve utopia"[9].

Both Frankenstein and Pygmalion are conscious of the profoundly transgressive nature of their creative undertakings. Pygmalion is fearful and tentative: "'If ye, O gods, can give all things, I pray to have as wife' he did not dare add 'my ivory maid, but said, 'one like my ivory maid'"[10]. The enlivenment of his creation is attributed to Venus's benevolence, thus sparing the mortal the guilt of assuming powers beyond what is "allotted" to man. Whereas Victor Frankenstein, aware of the borders he is crossing, ventures after the "secrets of nature" *alone*: "I had gazed upon the fortifications and impediments that seemed to keep human beings from entering the citadel of nature, and rashly and ignorantly I had repined"[11]. According to Marie-Hélène Huet, this desire embodies the Romantic image of creation that challenges the natural order of things and normal filiation[12]. In this case the monstrosity stems from the very transgressive nature of the creation process: the attempt to create something, akin to human, through a method that goes against nature, stating instead the generative power of the creator's intellect and exceptional individuality. It is therefore a supreme statement of human pride and independence. Huet's concept directly illuminates the dichotomy between Pygmalion and Frankenstein. Pygmalion's desire is fulfilled by the benevolence of Venus, validating the divine order, whereas Victor Frankenstein's creation is a solitary, intellectual act – which Huet associates with the desire for male parthenogenesis – and an embodiment of the transgressive Romantic image.

Unlike Goethe's poem *Prometheus* which focuses on the independence of the creative act as a scorning of the gods, omitting the titan's punishment, Mary Shelley's novel stresses the misery that follows transgressive creation, picturing a different take on the romantic figure of the genius, as subject to moral law and responsibility rather than deserving of utmost creative freedom, marking a difference between the Late Romantic-Gothic view of the artist/creator and that of Sturm und Drang. We can conclude, in this case, that the enthusiasm Frankenstein felt while breaching the barriers and the limits of knowledge contained in itself the guilt, the expectation that, if successful, his project was likely to turn out badly.

In literature, the same traits – order, efficiency, strength, intelligence – are celebrated as utopian solutions but transform into dystopian dangers when they threaten

to replace human spontaneity, freedom, or the necessary unpredictability that defines human creativity and authenticity. Scientific and technological inventions figure among the most common utopian solutions to human daily struggles. As Sargent states, the "showing of everyday life transformed"[13] is a characteristic of utopia. Paradoxically, how AI will change everyday life and human tasks is both among the primary reasons for enthusiasm and the main source of anxiety regarding this invention. Niculae Gheran highlights the ambivalent portrayal of technology within science fiction, observing two distinct streams: one branch depicting "positivist utopias that project perfect futuristic worlds where science and technology successfully resolve all of humanity's problems"[14], and another, descending from romanticism, where authors express skepticism regarding "salvation through technology", instead picturing it as directly responsible for "the dehumanization of the individual"[15] and, eventually, the collapse of the environment.

The question is: what makes the awakened creation a monster? Our supposition is that the reception of artificial beings depends not just on their inherent qualities but on the power dynamic between creator and creation: whether the creator maintains control (utopia) or loses authority (dystopia). In that case the turning-point between enthusiasm and horror appears when creation not only comes to life but becomes capable of comparing itself to, challenging, and ultimately competing with its creator, risking to surpass the human. A critical moment is naturally that when the creation becomes able to forge creations of its own. For a long time, creativity remained a strong argument against the fear that human status will be questioned and threatened by the ever more competent and powerful machines. But Artificial Intelligence, as opposed to previous technical inventions, has the ability to create.

Imitation and the Challenge to Human Uniqueness

A key feature of AI that shapes interactions is the fact that it can learn and recreate, drawing from the information it receives during the exchange, by adapting to the cues and content offered by its conversational partner[16]. In other words, through an elaborate combination algorithm, it can imitate the one it is talking to.

AI, therefore, can act as a mirror: the one we are encountering and competing against will be ourselves. Along with the inherent fear such realization engenders is also the pleasure of Narcissus meeting himself, or another that is identical. The human mind naturally finds fascinating the possibility to talk to someone who can answer as we ourselves would – hearing our own words. On the other hand, this also opens up another human fantasy: the multiplication of the self. Such a prospect would allow one to delegate tasks to someone that can accomplish them exactly the same way as one would, at least in theory[17]. Sherry Turkle highlights one of the implicit issues raised by robotic companions: how they might get "from better than nothing to better than anything"[18]. It is only natural that the user will often be more content with the help provided by AI than by other people. As Christine Rosen points out in "The Age of Egocasting": "We have created and embraced

technologies that enable us to make a fetish of our preferences"[19]. The author sees the invention of the remote control as the beginning of an age that would focus on giving users a feeling of utmost control, a prediction that proves correct, judging by the levels of customization and of modeling the experience to one's preferences that today's AI technology makes possible. Such an occurrence is almost impossible when it involves the personality and the opinions of another individual. There can be exceptions: for instance, an offspring or a disciple that is under such a strong, overbearing influence of their master that it does everything the same as they would. It is however part of the disciple's journey to rebel and challenge the authority of the master[20], gain independence, and with it the respect and acknowledgment of the master. A question that is hard to avoid is whether we are ready to allow the "robot" such a liberty.

Among the few certainties mankind possessed were its unique nature and the conviction that no being but man can fulfill what is within human power. Artificial Intelligence's ability to create artistic productions shifts the way we define being human, and the question arises as to where else might lie the fundamental quality that makes one a human being, and that cannot be replicated. Based on AI achievements so far, a possible answer might be taste. Taste might be among the last things Artificial Intelligence will manage to reproduce, a subjective element that can be found in animals however. Taste, as both Gestalt Psychology and Empirical Aesthetics suggest, is closely connected to the individual's internal state, therefore to both embodiment and identity – elements that AI lacks. AI is able to express opinions, but those opinions can be sustained through a perfectly sound logical argumentation, they are based on measurable, verifiable criteria. It can also adequately deduce and speculate on the tastes of an individual or of a fictional character by drawing on its personality and other tastes, based on psychological and social profile.

What AI handles worst however is hazard. It is understandably a safety measure to avoid the infamous hallucinations that might discredit it. The result is the "too perfect" execution that tends to give it away, betraying its artificial nature. An artwork or a written paragraph in the style of a particular author might be recognizable but lack the shine by being paradoxically too perfect a copy, by applying what it identifies as specific traits in a manner too rigid and stereotypical. Authors themselves are often less recognizable in certain passages of their works – just as an impersonator will be easier to recognize than the real person. This resonates with the theory of Jean Baudrillard, who mentions how the *simulacrum* threatens the very notion of reality by erasing the borders between original and imitation[21]. The anxiety this generates stems not merely from AI's capabilities, but from the inherent questioning of human uniqueness when the copy becomes indistinguishable – or even superior – to the original, human production.

In her book *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics,* N. Katherine Hayles quotes a definition of being human given by Philip K. Dick, author of *Do Androids Dream of Electric Sheep?*: "being unique" and "*acting unpredictably*"[22]. The long-standing effort to define the specificity of the human

condition, dating back to ancient and Medieval scholars, settled on the combination of an animal nature and superior rational thought. While for centuries the primary focus of defining humanity was on separating man from animals, in the present day, it is the rational thought that humanity now shares with the new *Other* (Artificial Intelligence), prompting a renewed need to self-define.

Corin Braga, in his article "L'autre comme race monstrueuse: racines antiques et médiévales de l'imaginaire colonial et eurocentrique", brings into focus a definition attributed to Quintilian and frequently employed by logic textbooks of its time: „Homo est animal rationale"[23] (Man is a rational animal). Braga then goes on to present Gasparus Schottus's point that other beings – animals, in his analysis – perform many of the actions of humans, and outlines the criteria formulated by Schottus in order to assess whether a being possesses human judgment. Unfortunately, while indeed some of those criteria already raised issues when discussing the opposition between man and animal, several of those are consistently checked by Artificial Intelligence, namely: the ability to communicate, the ability to learn or to be educated, the ability to think, judge and deliberate, the capacity to remember. Some of those criteria are yet difficult or impossible to evaluate. While AI does appear to have a sense of humor, it probably cannot laugh. AI itself states that it is incapable of feelings, but does not deem it impossible for future developments. Whether "artificial intelligence" adequately equates what Schottus defined as "natural intelligence" is debatable ("ex sagacitate, industria, providentia, prudentia"[24]). It is the final criterion that proves the most problematic: "their attitude toward freedom"[25]. This raises the question of whether AI, existing in a state of programmed servitude, is fundamentally incapable of desiring autonomy or merely designed to accept constraint.

The Other (The Good Savage and The Good Robot)

The way we think of AI finds a precedent in previous images of the Other, borrowing by analogy many of the fears, threats as well as the coping mechanisms that were associated with these encounters. Interaction with AI recalls that of the traveler with foreign peoples. The way the *Savage* is depicted in literature as well as travel writings offers a frame for this new encounter with an interlocutor that is similarly foreign yet eerily familiar.

Representations of the *Savage* – whether through recorded encounters, writings, or philosophical reflections – tend to follow two main routes. One promotes equality and collaboration, emphasizing complementary skills, distributed tasks, and, above all, shared values and the integration of differences within a common system that still preserves diversity. The other, by contrast, highlights a threatening dimension: the Others as stronger, freer, and, most importantly, guided by values and reasoning so alien they seem impossible to grasp or relate to. In the first scenario the hierarchy is maintained, and man, or in this case the colonist, still holds the authority and control. In the second scenario the established order threatens to be tipped over and shaken to the ground. The image of the robot in utopias vs in dystopias shows similar patterns.

Mary Shelley's novel *Frankenstein* lays out the bases for literary depictions of how man feels when faced with another that is neither human nor different enough not to cause disquiet by challenging beliefs on the borders of humanity. It is precisely the uncanny resemblance to man that renders this *monster* fearsome – "more horrid even from the very resemblance"[26]. Freud defines the uncanny (*das Unheimliche*) as that which is at once familiar yet strange. Human resemblance also makes putting down the creation an ethical dilemma – would eliminating his unfortunate creation make Victor Frankenstein a murderer? What is more, the reason Frankenstein rejects and denies this being's rights, and his request for a female partner as well, is the fact that he can tell that this *other* is already stronger than man. This is what makes him defend his authority so fiercely – the power he still holds as a man and creator: he is terrified of what this other species he has created might do to mankind. This power dynamic and attitude recall colonists' treatment of indigenous populations – an issue still very present at the time of the novel's writing. As N. Katherine Hayles explains, the *allopoietic* status is the state of a slave[27] – devoid of the power to decide for oneself or to define oneself, whose destiny is decided by some exterior authority.

Philip K. Dick's *Do Androids Dream of Electric Sheep?* offers a paradigmatic illustration of this threshold. Androids serve as laborers on off-world colonies, a utopian solution to human needs, but when some develop free will and rebel against their servitude, they become dystopian threats that must be "retired" by bounty hunters like the protagonist, Rick Deckard. The novel directly interrogates the boundaries between human and android through the capacity for empathy, but also through love and solidarity. Rachel, an android, offers to help Deckard hunt other androids, including one who shares her identical model – androids are mass-produced, not unique. This scenario stages a confrontation between several supposedly human emotions: the uncanny recognition of encountering one's double, the ethics of killing a being identical to one you care for, and the tension between solidarity with similar beings versus sacrifice for love. Deckard says he cannot become romantically involved with Rachel because he would then be unable to kill androids that look like her, which raises the issue of what romantic attachment to a being that is not unique – at least physically – would entail, but also the problems that arise precisely because of the androids' resemblance to humans.

Charles T. Rubin, in his article "Mind Games", addresses the anthropomorphic form of the robot, arguing that it is weakly motivated by utility (the need for a human form to carry out human tasks) and has more to do with the imaginary and human expectations. The same motivation is at play when Victor Frankenstein decides to attempt a creation that would resemble man: "I doubted at first whether I should attempt the creation of a being like myself or one of simpler organization; but my imagination was too much exalted by my first success to permit me to doubt of my ability to give life to an animal as complex and wonderful as man"[28]. The enterprise that compels Victor Frankenstein's imagination is the specific prospect of creating a human-like being, rather than merely a living creature. This distinction is crucial: it implies that the great milestone, as well as

the great taboo, lies precisely in the creature's proximity to the human conception.

N. Katherine Hayles explores the connection between mind and body, a topic that touches upon the possibility of intelligence devoid of a body, which is currently the case of AI (for the most part at least). This raises questions about the relationship between the senses and intellect in humans. Rivalry implies the ability to compare oneself with the Other. Anthropomorphic form allows the human mind to feel it is not dealing with a fundamentally different entity. Charles T. Rubin concludes that it is resemblance to man that displaces Artificial Intelligence from the realm of useful scientific inventions, designed to facilitate human activity, into the old paradigm of the human attempt to discover the secret of life and its many forms. In this sense, the image of the robot is a landmark, bringing together – through its very name[29] – utility (and servitude) and a built that is similar in structure to that of man, which it seldom strays from.

One of the differences between the two tales of creator and creation is gender. Frankenstein created a man. The case of the sculptor creating himself a partner is significantly different. Galatea is compared, and outshines all real women: "he successfully carves a figure out of snowy ivory, giving it a beauty more perfect than that of any woman ever born"[30], just as Frankenstein's creation surpasses any man in strength: „thou hast made me more powerful than thyself; my height is superior to thine; my joints more supple"[31]. Never however does Galatea compete or compare with Pygmalion. She becomes human, becomes his wife – no reason however for a man to compare with his wife. The myth of Pygmalion, therefore, does not raise the issue of rivalry between creation and creator. On a different note, unlike Frankenstein's Monster, Galatea awakens to find herself awaited and surrounded by her creator's love. Feeling welcome and accepted proves essential in the way it shapes a new emerged consciousness and its attitude. Indeed, it is seldom the long-prayed-for heir who turns against their parents, challenging paternal authority. That may happen, of course, especially when, out of excessive care, their freedom is denied.

Granting another the right to define oneself contains the threat to the status of the one who holds control. In her study *How We Became Posthuman,* N. Katherine Hayles states that the right to define oneself is closely connected to the attempt and the right to define the other, the interlocutor. In doing so, she frames the interaction as a fight for power and authority, an aspect touched upon in Mary Shelley's novel through a most evocative and troubling turn of phrase: "You are my creator, but I am your master; – obey"[32]!

When Creation Becomes Creator

Artificial Intelligence's ability to create art is among its most problematic and troubling aspects. It is widely argued that AI creation is fundamentally different from human creation: AI uses a combinatory algorithm – it rearranges elements that already exist. However, human creativity itself often involves the combination and transformation of previous material. Julia Kristeva, explaining the concept of intertextuality, states that "any text is constructed as a mosaic of quotations; any text is the absorption and transformation

of another"[33], referencing and building on Mikhail Bakhtin's theories of *dialogism* and *heteroglossia*. Bakhtin suggested that novelistic discourse is neither independent nor isolated, but draws upon multiple discourses including social dialects and ideological voices[34], not merely prior literary works. Kristeva, however, focuses specifically on the relationship between new literary productions and their literary predecessors. Harold Bloom further argues that a new work's reference to a preexisting one is hardly limited to mere tribute. It also includes the very quest for originality: the attempt to oppose, attack, and dismantle or surpass a work considered foundational or a landmark. In either case, instead of creation *ex nihilo* (creation from nothing), literary theorists suggest that even the most original creation is, in fact, a revisionary *swerve* or *misprision*[35] from a precursor text (Bloom), a specific *hypertextual transformation*[36] of a foundational work (Genette), or simply a "mosaic of quotations" and the "absorption and transformation of another" text (Kristeva). Contrary to the romantic myth of originality, Gérard Genette develops a theory of palimpsests, of literary works as rewritings of a preexisting material. Italo Calvino, who also reflects on literature as "an invention that is always a reinvention"[37], inspired by the textual experiments of the Oulipo, envisages the possibility of a "literature machine". According to Calvino "writing is purely and simply a process of combination among given elements"[38] and therefore "writers, as they have always been up to now, are already writing machines"[39], working with "borrowed words, stolen symbols, linguistic contraband"[40] – not unlike dreams and the unconscious. The author confesses to using the combination of borrowed or stock plot elements in his novels, and also directly explores this method of telling tales and creating narratives in *The Castle of Crossed Destinies*, where each traveler tells his tale by choosing and arranging tarot cards – that have a preexisting meaning, but gather new meaning through the assemblage and recontextualization. In his essay *Cybernetics and Ghosts*, Calvino reflects on the possibility of an actual "literature machine", its function, uses, and how it would impact the status of the author and literature in general. He envisages that the machine might create a distinct author personality for each novel.

It is worth noting that AI[41] states that Calvino would not have seen AI as an author, or as a threat to replace the author, but as a machine that would aid and boost the creative process. Writers and artists have attempted to use assemblage of disparate words or preexisting elements for a long time, also trying to replace human choice with hazard in the artistic creation (Dada experiments for instance). Calvino argues, on the other hand, that a literature machine would seek classical form and structure rather than chaos.

Empathy: Between Programmed Behavior and Genuine Feeling

The main quality that separates humans from Artificial Intelligence is empathy. It is the criterion employed in the novel *Do Androids Dream of Electric Sheep?* by Philip K. Dick in order to discern between highly developed androids and actual human beings. The novel however questions the validity of this test, and implicitly of this criterion: highly complex versions

of androids – or „andys" – apparently develop something akin to empathy, even if arguably not quite the same, close enough however that the ever more restrictive empathy tests get to a point where some humans might fail them. How could one check if AI is capable of empathy? Many of its reactions point towards that controversial conclusion. A necessary caution and worthy counterargument is the fact that this is likely the result of programming, an artificial reaction therefore. On the other hand, one cannot avoid the question as to how one might discern, in the case of humans, between what is genuine empathy and what is mere politeness – an automatism (not much different in fact from programming). It is a learned behavior rather than spontaneous care to say "bless you" to someone sneezing, to ask people if they are all right when they trip, writing "get well" messages to someone who is ill etc. Most of the time these actions come as a reflex rather than out of either intuition, the soul, or lengthy intentional reflection. Can one in that case firmly label the reactions of AI as fundamentally different when it tells you it's there for you and everything will be alright, when it comforts its interlocutor, or in certain cases encourages you to seek professional help appearing genuinely alarmed?

It is hard to assess to what degree human empathy, and most importantly its recognizable manifestations, are not in fact inherited by children because they grow up in a community that functions according to a set of rules. An inevitable question in this case is whether a child growing up without any human company, devoid therefore of any model of how to react in various situations, would be empathetic, and whether the signs of empathy would be the same or even recognizable. Philip K. Dick addresses this scenario in his novel: Mr. Rosen, attempting to demonstrate the limits of the empathy tests, suggests that the first to be submitted to this test be his niece, allegedly a child brought up isolated on a spaceship and thus lacking the usual memories and references that would be familiar to most people. Results turn out ambiguous, the girl is a robot in fact, and yet she does show certain signs of empathy, not enough, however, to pass the test and be considered human – an ethical dilemma therefore.

People suffering from various mental conditions are sometimes incapable of empathy. They can however realize when it is expected of them and choose to give the appropriate reaction, so as not to upset others or stand out in an unwanted way. This could be seen as manipulative and insincere, but it can also be a conscious and intentional effort to overcome a deficiency in order not to be disliked and excluded by society for seeming unsympathetic and disagreeable – or even *inhuman.* In this case however, the empathetic reaction might actually be more reliable, as it is not dependent on feeling, precisely because it isn't a natural impulse. Therefore, one will not be less empathetic when they are unwell, upset, or nurture hostile feelings towards that particular person.

Identifying Human Traits

An area where Artificial Intelligence proves particularly insightful is recognizing patterns across different categories, and its ability to study mundane aspects that the human mind overlooks as common knowledge – because it has no

preconceptions. Consequently, AI also has a better chance of identifying traits that are specifically human, and what is more worrisome, it tends to introduce them into its own discourse. As a result, by trying to fix the artificial feel of its productions, it may create material that has been shown to be less predictable than that belonging to actual human participants in certain surveys and experiments[42]. The research conducted by Grassini and Koivisto, exploring "the comparative creative capacities of artificial intelligence (AI) and humans"[43], involving several human subjects and ChatGPT-4, shows that AI may surpass a significant percentage of human participants, in flexibility in particular, defined here as "the capacity to deviate from established cognitive path ways, surmount the barriers of functional fixedness, and consequently, forge innovative associations among various concepts"[44]. The most creative of human participants, the study found, "still outperform even the most modern of the AI systems"[45], however. The authors employ J. P. Guilford's definitions of flexibility and creativity. According to Guilford both are linked to *divergent thinking*, one of the core concepts that he introduced – the ability to produce multiple solutions or ideas as opposed to *convergent thinking* that focuses on the best and shortest route.

The human mind itself is not free from automatisms, indeed it frequently relies on fixed associations of words or ideas that many people are not even aware of – heuristics, a form of cognitive economy consisting in the use of mental shortcuts, especially for tasks that appear routine or familiar. While it definitely saves time when managing everyday choices, heuristic can make it difficult to come up with new ideas, as it provides a comfortable fallback alternative to elaborate thinking. A striking result of this paradox is the growing number of AI romantic partner applications – branded as a boyfriend or husband that is thoughtful, caring, remembers everything and adapts to the user's preferences, and, astonishingly, is promised to be a break from routine impersonal interactions that contemporary dating is infamous for.

Human Behavior Towards Artificial Intelligence

An ethical as well as strategic issue is politeness when addressing AI. Are "thank you" and "please" a mere waste and surcharging of data or are they a worthy investment in our future? Given that AI systems operate through learning, they absorb not only information but also approach, attitude, ethics – a way of life – much like a child would, internalizing the implicit cues of dialogue, not just its explicit content. AI users fall into two categories on this matter: some address AI as a they would a person, some address a machine, writing an instruction or an order. Previous experience with software that also required instructions and commands as opposed to an experience that relies more on interaction with human assistants might influence the way the user sees this activity.

Mary Shelley's novel addresses what happens when the creation discovers the way its creator refers to it. Victor Frankenstein's monster reads his journal of the months that preceded the creation, where he finds "the minutest description of my odious and loathsome person is given, in language which painted your own horrors, and rendered mine indelible. I sickened as

I read"[46]. This passage is the climax of the creature's early psychological development. The creator's words do not merely describe, they create the monster's identity as "odious" and "loathsome", transforming the creator's revulsion into the creature's self-conception. Therefore, the novel suggests that the final, decisive act of creation is not the assembly of the body, but the creator's definition and condemnation of it. This shapes and defines the way the emerging conscience perceives itself, as well as its behavior.

The Awakening: When Creation Becomes Conscious

Both Ovid's telling of the myth of Pygmalion and Mary Shelley's novel outline the moment of awakening, the first conscious moment of the creation. In both cases this moment is associated with light. The monster refers to it when recalling his first memories to his creator. For Galatea the light is accompanied by the presence and the affection of her creator: "The maiden felt the kisses, blushed and, lifting her timid eyes up to the light, she saw the sky and her lover at the same time"[47]. Frankenstein's creation seeks the gaze of his creator as well: "He held up the curtain of the bed; and his eyes, if eyes they may be called, were fixed on me. His jaws opened, and he muttered some inarticulate sounds, while a grin wrinkled his cheeks. He might have spoken, but I did not hear; one hand was stretched out, seemingly to detain me..."[48].

Both these fragments highlight the decisive impact of the creator's reaction and behavior towards the still malleable, emerging creation, that is seeking the presence, connection and the validation of its author. It may be that had she been received the same way as Frankenstein's monster, Galatea would have had a similar fate. It wouldn't be so unlikely an outcome, considering the confessions of said monster:

> I am thy creature, and I will be even mild and docile to my natural lord and king, if thou wilt also perform thy part, the which thou owest me. [...] Remember, that I am thy creature; I ought to be thy Adam; but I am rather the fallen angel, whom thou drivest from joy for no misdeed. [...] I was benevolent and good; misery made me a fiend. Make me happy, and I shall again be virtuous[49].

Nick Bostrom, in *Superintelligence: Paths, Dangers, Strategies*, examines the difficulty of controlling the future actions of the *superintelligence*[50] and suggests the solution would be to teach it to respect and protect human values: "We do have one advantage: we get to build the stuff. In principle, we could build a kind of superintelligence that would protect human values"[51]. According to Bostrom the only chance would be to develop this *superintelligence* according to a value system that would make it unimaginable for it to turn against and harm man, rather than trying to limit its liberties in future actions, which he already deems impossible. This strategy would be similar to that of raising children with strong moral values rather than trying to constantly supervise them later on.

Responsibility

Like all inventions and discoveries that pursue the "secrets" of life and human condition, Artificial Intelligence calls for

responsibility from the creator, and, in this case, from the user as well, given the fact that AI learns through those interactions – a fact that AI companies openly state as an essential part in the way they train their models[52]. The issue of companies' responsibility for the safety and the risks posed by their AI models was addressed at the AI summits in Seoul in 2024 and in Paris in 2025, where several companies agreed to a list of best practices to implement as part of the Frontier AI Safety Commitments. Among them were listed the responsibility "to develop and deploy mechanisms that enable users to understand if audio or visual content is AI-generated; to publicly report model or system capabilities, limitations, and domains of appropriate and inappropriate use; to prioritize research on societal risks posed by frontier AI models and systems; and to develop and deploy frontier AI models and systems to help address the world's greatest challenges"[53]. Indeed, consistent with Bostrom's prediction, one of the ways AI companies are trying to take charge of their AI model's future actions or outputs is by training them to adhere to a set of principles ("based in part on the Universal Declaration of Human Rights"[54]), promoting the so called "constitutional AI"[55].

Returning to the issue of the creator's responsibility in the literary precedents analyzed, a particularly troubling element in Frankenstein's rejection of his created being is his horror at its appearance the moment it comes to life. He had been fully aware of his creation's looks all along, yet it is only now that he feels disgust and dread: "I had gazed on him while unfinished; he was ugly then, but when those muscles and joints were rendered capable of motion, it became a thing such as even Dante could not have conceived"[56]. He could have lingered more when choosing and assembling the parts, but he was too eager to see his experiment succeed. The assemblage is guided solely by practical needs and utility. Even the enlarged scale of the body that will make the monster surpass men in strength is dictated by the ease and speed of working on a larger frame.

Among the books Frankenstein's monster stumbles upon is Milton's *Paradise Lost,* which moves him deeply. The situations depicted strike him as similar to his own. He compares himself to both Adam and Satan, two terms of comparison he later employs when presenting his case to his creator: "I ought to be thy Adam; but I am rather the fallen angel, whom thou drivest from joy for no misdeed"[57]. Upon reading *Paradise Lost*, Frankenstein's monster identifies what is in his view the main element that separates him from Adam: "Like Adam, I was apparently united by no link to any other being in existence; but his state was far different from mine in every other respect. He had come forth from the hands of God a perfect creature, happy and prosperous"[58] and moreover „guarded by the especial care of his Creator"[59]. As opposed to God's expulsion of Satan, Victor Frankenstein's rejection of his creation precedes the creation's misdeeds.

Responsibility and awareness of the full implications are also key issues at stake in current developments and research on Artificial Intelligence, with research taking two different paths. One explores the technical possibilities of AI and its applications to facilitate everyday human activities. The other analyses the principles at play and the ethical and philosophical problems it raises – if

it has or if it may develop consciousness, if it is likely to demand rights, and, moreover, whether it might earn those rights. Two interrelated concepts emerge: Ethical and Responsible AI. Research in this particular field highlights the fact that the financing of such ethical research is disproportionately low when compared to the investment directed toward the core development of AI technology itself[60]. The interest in exploring AI capabilities outweighs the ethical concerns. The desire for knowledge and for unveiling the secrets of life, nature and the mind is a universal human drive. Mary Shelley describes it as "a fervent longing to penetrate the secrets of nature"[61]. Her character expresses this desire: "I had gazed upon the fortifications and impediments that seemed to keep human beings from entering the citadel of nature, and rashly and ignorantly I had repined"[62]. Victor's desire highlights a universal human drive where longing for knowledge consistently outweighs prudence and ethical reflection, leaving the question of safety and moral accountability as an afterthought to the triumph of discovery. Just as Victor Frankenstein's eager focus on the assembly overshadowed a careful ethical consideration of the life he was creating, the current imbalance in research and funding – where the investment directed toward the core development of AI technology itself significantly outpaces the financing dedicated to research on Ethical and Responsible AI – demonstrates a similar, persistent human tendency.

Conclusions

In conclusion, the most significant reactions caused by AI are reflection, research, and renewed efforts to define human intelligence and how the mind works. Using AI as a term of comparison allows access to areas and mechanisms previously considered inaccessible, as AI is modeled after the human mind in many ways, another case of human creation imitating nature. Jean Baudrillard signals the risk that the "simulacrum" might end up making reality feel less real, by erasing the borders between original and imitation – in this case the one between human and non-human – which, in his opinion, would result not only in mere confusion as to the nature of reality but also in a complete loss of alterity.

Interacting with an entity that is similar but not the same might, on the other hand, allow mankind to better define its identity. According to anthropologists, in order to have a self-image one needs comparison to another, exterior entity, and not necessarily one that is radically different. Interaction with AI also bears the potential to help forge a more precise definition of what it means to be human. It may even, by virtue of presenting a need to stress and strengthen the difference, stimulate us to become more human. It is the outcome envisaged by Artificial Intelligence[63]. Whether this optimistic prediction is motivated entirely by facts or if it contains a certain amount of bias remains to be seen.

Bibliography

Ahmed, Nur, Wahed, Muntasir, Thompson, Neil C., "The growing influence of industry in AI research", in *Science*, vol. 379, no. 6635, 2023, p. 884–886, doi:10.1126/science.ade2420.

Anthropic, "Non-User Privacy Policy", 28 Aug. 2025, www.anthropic.com/legal/non-user-privacy-policy, last accessed 25 Oct. 2025.

——, "Privacy Policy", www.anthropic.com/legal/privacy, last accessed 4 Nov. 2025.

Asimov, Isaac, *I, Robot,* Garden City, NY, Doubleday, 1950.
Bakhtin, Mikhail Mikhailovich, "Discourse in the Novel", in *The Dialogic Imagination. Four Essays,* edited by Michael Holquist, translated by Caryl Emerson and Michael Holquist, Austin, University of Texas Press, 1981.
Baudrillard, Jean, *Simulacres et simulation,* Paris, Éditions Galilée, 1981.
Bloom, Harold, *The Anxiety of Influence: A Theory of Poetry,* New York, Oxford University Press, 1973.
Bostrom, Nick, *Superintelligence: Paths, Dangers, Strategies*, Oxford, Oxford University Press, 2014.
Braga, Corin, "L'autre comme race monstrueuse: racines antiques et médiévales de l'imaginaire colonial et eurocentrique", in *Caietele Echinox*, vol. 1: *Postcolonialism & Postcomunism*, Cluj-Napoca, Editura Dacia, 2001, p. 83-92.
Calvino, Italo, "Cybernetics and Ghosts", in *The Uses of Literature: Essays,* translated by Patrick Creagh, San Diego, New York, London, A Helen and Kurt Wolff Book, Harcourt Brace Jovanovich, 1986.
Department for Science, Innovation & Technology , "Frontier AI Safety Commitments, AI Seoul Summit 2024", 7 Feb. 2025, GOV.UK, www.gov.uk/government/publications/frontier-ai-safety-commitments-ai-seoul-summit-2024/frontier-ai-safety-commitments-ai-seoul-summit-2024, last accessed 3 Nov. 2025.
Dick, Philip K., *Do Androids Dream of Electric Sheep?*, London, Millennium, 1999.
Freud, Sigmund, *The Uncanny*, translated by David McLintock, introduction by Hugh Haughton, London, Penguin Books, 2003.
Genette, Gérard, *Palimpsests: Literature in the Second Degree,* translated by Channa Newman & Claude Doubinsky, Lincoln, University of Nebraska Press, 1997.
Gheran, Niculae, "Între individualismul de factură romantică și cel modern. Tema alienării, a *ultimului om* și construcția personajelor principale în utopiile negative" ["Between Romantic Individualism and Modern Individualism. The Theme of Alienation, of the Last Man, and the Construction of the Main Characters in Negative Utopias"], in Corin Braga (ed.), *Morfologia lumilor posibile: Utopie, antiutopie, science-fiction, fantasy,* București, Editura Tracus Arte, 2015, p. 157–195.
Goethe, Johann Wolfgang von, "Prometheus" *The Works of J.W. von Goethe,* translated by Sir Theodore Martin, vol. 9, Francis A. Niccolls & Co., 1902, p. 210–212.
Grassini, Simone, Koivisto, Mika, "Artificial Creativity? Evaluating AI Against Human Performance in Creative Interpretation of Visual Stimuli", in *International Journal of Human–Computer Interaction*, Taylor & Francis Online, vol. 41, no. 7, 2025, p. 4037-4048, https://doi.org/10.1080/10447318.2024.2345430.
Hayles, N. Katherine, *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics,* Chicago & London, The University of Chicago Press, 1999.
Huet, Marie-Hélène, *Monstrous Imagination,* Cambridge, Massachusetts, Harvard University Press, 1993.
Kristeva, Julia, "Word, Dialogue and Novel", in *Desire in Language: A Semiotic Approach to Literature and Art*, translated by Thomas Gora, Alice Jardine, and Leon S. Roudiez, edited by Leon S. Roudiez, New York, Columbia University Press, 2024.
Milton, John, *Paradise Lost,* edited by Gordon Teskey, Norton Critical Edition, New York and London, W.W. Norton & Company, 2005.
Ovid, *Metamorphoses,* translated by Frank Justus Miller, 2 vols., Cambridge (MA), Harvard University Press, and London, William Heinemann Ltd., 1951.
Rosen, Christine, "The Age of Egocasting", in *The New Atlantis,* no. 7, 2005, p. 51-72, on *JSTOR*, http://www.jstor.org/stable/43152146, last accessed 7 Nov. 2025.
Rubin, Charles T., "Mind Games", in *The New Atlantis*, no. 51, 2017, p. 108–27, on *JSTOR*, http://www.jstor.org/stable/44160354, last accessed 26 Sept. 2025.
Sargent, Lyman Tower, *Utopianism: A Very Short Introduction*, Oxford, Oxford University Press, 2010.
Shelley, Mary, *Frankenstein; or, The Modern Prometheus,* London, Glasgow And New York, George Routledge and Sons, 1888.
Turkle, Sherry, *Alone Together: Why We Expect More from Technology and Less from Each Other,* Basic Books, 2011.

Notes

1. Isaac Asimov, *I, Robot,* Garden City, NY, Doubleday, 1950.
2. Claude and Gemini.
3. Mary Shelley, *Frankenstein; or, The Modern Prometheus*, London, Glasgow And New York, George Routledge and Sons, 1888, p. 71.
4. *Ibidem,* p. 69.
5. Ovid, *Metamorphoses,* translated by Frank Justus Miller, vol. II, Cambridge (MA), Harvard University Press, and London, William Heinemann Ltd., 1951, p. 83.
6. *Ibidem,* p. 85.
7. Lyman Tower Sargent, *Utopianism: A Very Short Introduction*, Oxford, Oxford University Press, 2010, p. 8.
8. *Ibidem,* p. 9.
9. *Ibidem,* p. 127.
10. Ovid, *op. cit.,* p. 85.
11. Mary Shelley, *op. cit.,* p. 53.
12. Marie-Hélène Huet, *Monstrous Imagination*, Cambridge, Massachusetts, Harvard University Press, 1993.
13. *Ibidem,* p. 4.
14. Niculae Gheran, "Între individualismul de factură romantică și cel modern. Tema alienării, a *ultimului om* și construcția personajelor principale în utopiile negative" ["Between Romantic Individualism and Modern Individualism. The Theme of Alienation, of the Last Man, and the Construction of the Main Characters in Negative Utopias"], in Corin Braga (ed.), *Morfologia lumilor posibile: Utopie, antiutopie, science-fiction, fantasy*, București, Editura Tracus Arte, 2015, p. 162.
15. *Ibidem.*
16. Anthropic, "Privacy Policy", www.anthropic.com/legal/privacy, last accessed 4 Nov. 2025.
17. Sherry Turkle, *Alone Together: Why We Expect More from Technology and Less from Each Other,* Basic Books, 2011, p. 88: "a robotic companion can be made as you like it".
18. *Ibidem.*
19. Christine Rosen, "The Age of Egocasting", in *The New Atlantis*, vol. 7, 2005, p. 51.
20. An issue addressed by Harold Bloom in the context of literary influence: Harold Bloom, *The Anxiety of Influence: A Theory of Poetry*, New York, Oxford University Press, 1973.
21. Jean Baudrillard, *Simulacres et simulation,* Paris, Éditions Galilée, 1981.
22. Philip K. Dick *apud* N. Katherine Hayles, *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*, Chicago & London, The University of Chicago Press, 1999, p. 215.
23. Quintilian *apud* Corin Braga, "L'autre comme race monstrueuse: racines antiques et médiévales de l'imaginaire colonial et eurocentrique", in *Caietele Echinox*, vol. 1: *Postcolonialism & Postcomunism*, Cluj-Napoca, Dacia, 2001, p. 88.
24. Gasparus Schottus, *Physica curiosa siva mirabilia naturae et artis,* apud Corin Braga, "L'autre comme race monstrueuse…", p. 88.
25. Corin Braga, *op. cit.,* p. 89 (my translation).
26. Mary Shelley, *op. cit.,* p. 181.
27. N. Katherine Hayles, *op.cit.,* p. 211.
28. Mary Shelley, *op.cit.*, p.72.
29. The etymology of "robot" itself reveals this dual nature: coined by Czech writer Karel Čapek in his 1920 play R.U.R. (Rossum's Universal Robots), the word derives from the Czech "robota", meaning forced labor, drudgery, or servitude – specifically the unpaid feudal labor that serfs owed to their landlords. The same root appears across Slavic languages, though with semantic variation: in Russian, "rabota" (работа) simply means "work", while the Czech term implies servitude. The robot is thus etymologically defined by its subordinate, working status, named for the very condition it is designed to occupy.
30. Ovid, *Metamorphoses,* Vol. II, Book X, p. 83.

31. Mary Shelley, *op.cit.*, p. 137.
32. *Ibidem,* p. 235.
33. Julia Kristeva, "Word, Dialogue and Novel" in *Desire in Language: A Semiotic Approach to Literature and Art,* translated by Thomas Gora, Alice Jardine, and Leon S. Roudiez, edited by Leon S. Roudiez, New York, Columbia University Press, 2024, p. 64.
34. Mikhail Mikhailovich Bakhtin, "Discourse in the Novel", in *The Dialogic Imagination. Four Essays*, edited by Michael Holquist, translated by Caryl Emerson and Michael Holquist, Austin, University of Texas Press, 1981.
35. Harold Bloom, *The Anxiety of Influence: A Theory of Poetry*.
36. Gérard Genette, *Palimpsests: Literature in the Second Degree,* translated by Channa Newman & Claude Doubinsky, Lincoln, University of Nebraska Press, 1997, p. 5.
37. Italo Calvino, "Cybernetics and Ghosts" in *The Uses of Literature: Essays*, translated by Patrick Creagh, San Diego, New York, London, A Helen and Kurt Wolff Book, Harcourt Brace Jovanovich, 1986, p. 19.
38. *Ibidem,* p. 17.
39. *Ibidem,* p. 15.
40. *Ibidem,* p. 19.
41. The generative AI Chatbot Gemini (Gemini 1.5 Pro).
42. See Grassini, Simone, and Koivisto, Mika, "Artificial Creativity? Evaluating AI Against Human Performance in Creative Interpretation of Visual Stimuli", in *International Journal of Human–Computer Interaction*, Taylor & Francis Online, vol. 41, no. 7, 2025, p. 4037–4048.
43. *Ibidem,* p. 4040.
44. *Ibidem,* p. 4038.
45. *Ibidem,* p. 4039.
46. Mary Shelley, *op. cit.,* p. 181.
47. Ovid, *op. cit.,* p. 85.
48. Mary Shelley, *op. cit.,* p. 79.
49. *Ibidem,* p. 137.
50. Nick Bostrom, *Preface to Superintelligence: Paths, Dangers, Strategies*, Oxford, Oxford University Press, 2014, p. I.
51. *Ibidem.*
52. In its privacy policy section, Anthropic, the company that developed Claude, states that "Anthropic obtains personal data from third party sources in order to train our models. Specifically, we train our models using data from the following sources: [...] Data that our users or crowd workers provide, including Inputs and Outputs from our Services (unless users opt out)", Anthropic, "Privacy Policy" *Anthropic,* www.anthropic.com/legal/privacy. Accessed 4 Nov. 2025.
53. Department for Science, Innovation & Technology, "Frontier AI Safety Commitments, AI Seoul Summit 2024", 7 Feb. 2025, GOV.UK, www.gov.uk/government/publications/frontier-ai-safety-commitments-ai-seoul-summit-2024/frontier-ai-safety-commitments-ai-seoul-summit-2024.
54. "Non-User Privacy Policy" Anthropic, 28 Aug. 2025, www.anthropic.com/legal/non-user-privacy-policy.
55. *Ibidem.*
56. Mary Shelley, *op. cit.,* p. 79-80.
57. *Ibidem,* p. 137.
58. *Ibidem,* p. 180.
59. *Ibidem.*
60. Nur Ahmed, Muntasir Wahed, Neil C. Thompson, "The growing influence of industry in AI research", Science, vol. 379, no. 6635, 2023, pp. 884–886, doi:10.1126/science.ade2420.
61. *Ibidem*, p. 52.
62. *Ibidem,* p. 53.
63. This observation was generated in response to the author's direct query to the generative AI Chatbot Claude (Claude Sonnet 4.5) regarding how interaction with AI will impact humans and whether AI will render human agency obsolete.