



The Death of the author: a cognitive approach to demystifying large language model's narratological intelligence

La Mort de l'auteur : une approche cognitive pour démystifier l'intelligence narratologique des grands modèles de langage

Ladjroud HAMZA ¹

Mouloud Mammeri University of Tizi-Ouzou | Algeria

Ladjroud.hamza@gmail.com

Abstract: Some authors explore the world within and without in an attempt to depict things meticulously and artistically to the reader, who would engage with the text hermeneutically. The poststructuralist school of thought dethroned the former and reinstated the reader as the originator of the epistemic frameworks and interpretive possibilities. In this paper, I will argue that the death of the author was a mere philosophical idea that sneaked into the literary world devoid of any scientific foundations. The studies reported here establish for the first time Barthes's eight postulates on the basis of which he dethroned the author, and foreground the cognitive basis supporting my argument. First, research in embodied cognition corroborated my view that the author's body plays a pivotal role in the mental formation of thoughts and the expression thereof at the level of language. Second, theory of mind research further accentuated the agency of the author in the process of meaning production. Thus, the function of the author in the literary work is defined by the multifarious ways mental defiance operates in the narratological fabric.

Keywords: Artificial Intelligence, ChatGPT, Cognitive Science, Death of the Author, Deep Seek, Embodied Cognition, Theory of Mind.

Résumé : Certains auteurs explorent le monde intérieur et extérieur dans une tentative de représenter les choses avec minutie et art au lecteur, lequel s'engage avec le texte de manière herméneutique. L'école de pensée poststructuraliste a détrôné les premiers et a réinstallé le lecteur comme l'origine des cadres épistémiques et des possibilités interprétatives. Dans cet article, je soutiendrai que la mort de l'auteur n'était qu'une idée philosophique qui s'est introduite dans le monde littéraire dépourvue de tout fondement scientifique. Les études rapportées ici établissent pour la première fois les huit postulats de Barthes sur la base desquels il a détrôné l'auteur, et mettent en avant la base cognitive qui soutient mon argument. Premièrement, la recherche sur la cognition incarnée a corroboré mon point de vue selon lequel le corps de l'auteur joue un rôle central dans la formation mentale des pensées et dans leur expression au niveau du langage. Deuxièmement, les recherches sur la théorie de l'esprit ont davantage souligné l'agence de l'auteur dans le processus de production du sens. Ainsi, la fonction de l'auteur dans l'œuvre littéraire est définie par les multiples manières dont la défiance mentale opère dans le tissu narratologique.

Mots-clés : Intelligence Artificielle, ChatGPT, Science Cognitive, La Mort de l'Auteur, Deep Seek, Cognition Incarnée, Théorie de l'Esprit



¹ Corresponding author: HAMZA LADJROUD | hamza.ladjroud@gmail.com

“A mermaid named Isabelle. She was swimming in the ocean. She saw a lot of fish and made friends with them. There were five fish and their names were Sandy, Bella, Fredy, Nancy and Lucy. We had so much fun playing. The mermaid had a shiny pearl on her neck. Suddenly, when the mermaid was swimming, she dropped her pearl. Then, someone stole it and they sold it in the shop for USD 1000. A very kind man that noticed a mermaid was sad; she could not find her necklace. So, that man was going to buy a new necklace that looked like hers.”²

A short story written by an anonymous author, whose identity is not of any relevance to our discussion at this point, who is aptly able to pass the false-belief test and seemingly capable of creative thought. Additionally, the narrative is loaded with epistemological bearings that triggered Michel Foucault to ponder and wonder over Samuel Beckett’s rhetorical question: “What does it matter who is speaking.” (Arnold 2002, 71) The same speaker-question was entertained by Roland Barthes at the onset of his seminal essay *The Death of the Author*. When applied to the above-quoted story, not only queries about the identity of the speaker will surface, but also the underlying epistemological structure informing the act of narration per se. Aren’t we expected to question the ethics of the narrative for letting the thief go away with his crime? Aren’t we going to fail trying to justify the surplus number of characters, who would soon prove to be contributing just a little to the development of the story plot, if anything at all? Aren’t we going to ask ourselves how did the ‘kind man’ know that the mermaid’s sadness was due to losing her necklace? By the same token of knowing more about the story’s author, don’t we furthermore find ourselves obliged to reflect on the sense of abruptness in the way events unfolded in the narrative proper, which might have crippled the reader’s imagination under the counter-intuitive principle of: “tell do not show”? Don’t we equally find ourselves at odds with the identity of the story’s narrator as he metamorphoses from the character of an omniscient narrator into a participant in the events of the story as indicated by the pronoun “we”? All these questions, do not only constitute part of the content of the mental clouds that hover over the consciousness of the close reader, but they also find their way in the epistemological horizons of the cognitive scientist who would likely feel an urge to explore the mental structures responsible for the origin of those thoughts. Therefore, the main aim of this paper is to draw on a body of research concerned with *Theory of Mind* and *Embodied Cognition* to address the abovementioned questions in relation to Barthes’s theoretical postulates on the death of the author.

The sense of mystery, as furnished by the previous readership contemplations, will soon phase itself out the very moment the identity of the author is revealed to us: a seven-year-old girl attending Year 3 of the Cambridge curriculum in a British school. Now, does it really matter who is speaking? I think, it definitely does, for the sheer fact that the more we knew about the child-author, the less expectations our interpretive framework would engender, the more we would instead aim at theorizing about children’s cognition and psychological development. In this case, the author is dethroned by shifting her status from an authoritative source of knowledge into a subject of scientific scrutiny. No wonder that scientists and literary critics may implement the same architectural pattern when exploring texts written by eminent authors. However, the experimentation’s *telos* will be

² This story was penned down by my daughter for her weekend creative writing practice at home. The story was thought out and written by herself without any adult intervention at any stage of the writing process. The story was written in a single draft that was not manipulated by any adult. Only some grammar errors were corrected that the author of this paper does not believe will affect the general purpose for which the story was embedded in this research paper. For reference to the original copy of the story, please refer to the appendix.

slightly different, as the scientist will intend to grapple with queries arguing that these will unravel paradigms about human nature and existence. Unlike those drawn from children's behavior, they mostly serve to connect the dots in our diachronic development from mere embryos to fully grown males and females. In short, based on the identity of the author, we, as laymen reading the mermaid's story, would dismiss plenty of our interpretations as faulty and irrelevant due to the author's inability to have intended them in the first place. If authorial intention is absent, what value does reader interpretation add to a narrative devoid of all this semantic load falsely attributed to it? What is the difference then between a hermeneut and a schizophrenic reader, who perceives, in narratives, interpretive opportunities reflecting ideas that could have never risen from a seven-year-old child? To illustrate, let us explore how the same narrative would lose all its former hermeneutic values if the author had been a large language model (LLM) such as ChatGPT-4:

Lila, the mermaid, swam joyfully through the coral reefs, laughing with her five fish friends: Sandy, Bella, Fredy, Nancy, and Lucy. Around her neck, she wore a shiny pearl that gleamed with the colors of the ocean. Suddenly, the pearl slipped from her neck and fell to the sandy seabed. Before she could grab it, a sneaky thief swam by, snatching the pearl and disappearing. Heartbroken, Lila gazed at the empty spot around her neck. A kind man, noticing her sadness from a boat above, asked what was wrong. Learning of her loss, he bought a necklace with a pearl that resembled the one she had lost. Lila smiled; her heart lightened by his kindness.³

Admittedly, the AI-generated version of the story does not fail any of the orthodox rules of creative writing, as it perfectly evokes the most pertinent sensory imagery in accord with the narratological context. Moreover, the short narrative deploys thoughts in the sequential order characteristic of the algorithmic description of human language and how it is represented in LLMs. What the story truly lacks is the human blueprint that connects the characters, plot and occurrences to the large body of global knowledge, as cognitive linguistics labels it, that distinguishes our human nature from that of all non-human organisms. Although the previous narrative was penned down by a child, the human aspect thereof is captured in the narratological tidbits and the grammatical errors (the use of the first-person plural), which betrayed the child-author's unconscious affiliation with the world of mermaids. True that we may include an algorithmic structure that would generate such a human effect in the AI-generated stories. However, quite crucial to this undertaking, how long are we going to keep administering this placebo effect to human readers, under the false claim that "the birth of the reader must be at the cost of the death of the author." (Barthes 1977, 148)

The relative completeness of the AI-generated narrative should call into question the significance of the author's character and its place in the literary work. Should we make a gigantic leap of imagination to a period where the language model(s) will be capable of generating ingenious narratives rivaling those written by its human counterparts, additionally the model will be able to have first-hand access to all aspects of human life (cultural, political, economic, religious...etc.) scrutinizing the human *status quo* in the lights of the colossal digital trove of data available to the model; do we in this imaginary scenario expect the language model to be able to replace human literary genius? Are we

³ The story has been generated by ChatGPT-4 after some basic instructions has been given as they appear in the appendix of this paper.

really to accept the new taste being poured into the literary world and fail even to notice the difference between human and artificial ingenuity? If this is mere chimerical thinking, then what makes the replacement impossible? Or, put differently, what idiosyncratic or collective parameters will conserve our proclivity toward narratives written by human writers rather than LLMs, even in the presence of sweeping ingenuity in the latter's output? Would it not imply the resurrection of Barthes's dead author?

To address these questions, this research paper will first establish Barthes's propositions, which underlie his poststructuralist views on human nature in general and on the death of the author in particular. Once this foundation is laid, I will juxtapose Barthes's postulates with cognitive conceptions of human nature in order to evaluate the extent to which philosophical ideas can withstand the test of scientific rigor. It must be emphasized, however, that this methodological framework does not seek to diminish the invaluable contributions that poststructuralist literary criticism has made to literary and cultural studies. Rather, the argument arises from the recognition that the profound technological and intellectual developments of the twenty-first century render it increasingly urgent to reconsider the role of literary studies and to probe some of the limitations and challenges of contemporary criticism.

1. Human Nature from an Essentialist Perspective

It is invariably difficult to imagine thoughts rising from a human mind without an act of defiance to other existing ideations. It is even more cumbersome to ever think without engaging in mental comparisons between disparate thoughts or the systems from which they emanate. Take for instance a toddler who is trying to identify a "cat" from a "dog". How would this ever be possible without analogical reasoning? The child simply engages in an unconscious form of defiance whereby the two mental images are brought against one another in the actual presence of either or both animals. Then, based on the mental framework (schema development), the child is capable of identifying the animal(s) with utmost precision, and hence discarding the second option as incompatible and temporarily irrelevant. The same very child could come to ethical conclusions by means of an analogy with a totally different situation of a deontological consideration. A child may conjecture that cheating on the exam is equally unethical drawing on the social normative principle that stealing is unrighteous because it causes harm to others.

Proponents of the *Essentialist Philosophy* to human nature capitalize on scenarios such as these to corroborate the view that humans share universal traits that make them different from all the other creatures on the face of the earth. Prior to the Christian narrative of human descent, Aristotle forcefully argued that rationality was the dividing line between humanity and bestiality. Accordingly, humans are born with innate capacities that make them eager to reason (*logos*), generate thoughts and be insatiably hungry for knowledge. In fact, Aristotle took rationality further than this to conclude that the purpose of human life (*telos*) is to engage in rational thought: "[...] one must posit the life [of that which possesses reason] in accord with an activity, for this seems to be its more authoritative meaning." (Aristotle 2011, 13) The analogical reasoning that Aristotle deployed, to distinguish Man from animals and plants, was predicated on a propositional input and a teleological output. His conceptualization of the tripartite soul ruled out the vegetative (nutritive) soul and sensitive (perceptive) soul as common to animals and plants, whose

existence in humans is justified by physiological and biological needs for growth, movement and survival. Left only for the latter goal, human beings would perish along the same pathway Gilgamesh pursued looking for “The Plant of Everlasting Life”. Therefore, the rational soul should be more concerned with reaching out for ultimate realities aided by the power of reason, not merely sensual momentary pleasures. Equipped with a third soul, and aided by the power of reason, Man should aim for a state of flourishing, where each person is enabled to realize their full potential (eudaimonia).

The biblical tradition would come along to support this notion of flourishing; however, it has to defer the fruition of the undertaking to an apocalyptic realm (the hereafter). While Aristotle stressed rationality as the essence of Man, the onset of the book of Genesis seems to be unequivocal about the total (non-partial) peculiarity of mankind as well as their dominion over everything that dwells planet earth. Prior to this heavy duty, the Bible authoritatively decreed that human nature should be a derivative of the divine and its working mechanisms should be reflective of the source from which it emanates: “So, God created man in his own image, in the image of God created he him; male and female created he them.” (Genesis 1:27) The postponement of Aristotle’s “eudaimonia” was conversely the direct outcome of the contentious antonymy between the ontological origins of Man as an image of God (*imago Dei*) and the dictates of the other two souls (vegetative and sensitive), as the Aristotelian tradition goes. Given this, the resulting Fallen Nature of humankind lends rationality a pivotal role in the salvation narrative, but it delays the foreseeable benefits of the microcosmic struggle among the conflicting faculties of Man, and the macrocosmic duels between Man and God, until a later stage on the diachronic development of the human spirit. In short, although both traditions remain distinct, the Aristotelian and biblical philosophies share some common grounds in their conceptualization of human nature—both of them impute innateness to human nature.

In his monumental book *Leviathan*, Thomas Hobbes can be seen as providing socio-political and psychological footnotes on the system of ideations inaugurated by Aristotle and the biblical narratives regarding human nature. Of the tripartite soul, Hobbes seems to concur with Aristotle in postulating that “[...] in the nature of man, we find three principal causes of quarrel. First, competition; secondly, diffidence; thirdly, glory.” (Hobbes 1998, 83) Therefore, according to the Hobbesian political theory, the project of tracing ideation in Man should commence with understanding the psychological pressures impinging on the individual’s psyche during the state of nature. The fact that all humans come to the world equipped with the same abilities, and simultaneously driven by self-preservation and desire for power, engenders a pressing need to come to terms with the overwhelming feelings of insecurity and fear in the absence of a unison social contract. In other words, before mankind could enjoy the glory of the final stage of their socio-psychological development, under the subduing and centralizing power of the absolute sovereign (Leviathan), the gift of reason is again militarized, not to achieve virtuous ends, but to seek individualistic gains in the form of self-preservation and acquisition of power. Reason, in the Hobbesian sense, is fundamentally different from that of Aristotle, as it frames human nature in a mechanistic conceptualization, whereby reason is deployed as a tool of manipulation similar to the puppet in the despotic power of the puppeteer constructing and deconstructing illusions as needed. In fine, while Hobbes’s deterministic and mechanistic systems function to further compartmentalize human freedom by subjugating

Man to the sum of his naturalistic drives, Michel Foucault's philosophy extrapolates the unfreedom theme to include extraneous factors equally significant in the understanding of human nature.

2. Human Nature from a Post-structuralist Perspective

The notion of the author in the literary critical analyses of Michel Foucault, Jacques Derrida and Roland Barthes is deeply rooted in the epistemological foundations of the poststructuralist philosophy in its quest for a profound and comprehensive conceptualization of human nature. In its very essence, poststructuralism emerged as a reaction to the postulations of the essentialist philosophy. Foucault unwittingly turned the world of Hobbesian determinism topsy-turvy by arguing that the individual's rational existence is not the only driving force there is, for there are other forms of knowledge, he labelled "epistemes", whose presence in society can decentralize the status of man in the whole matrix. Through a panoptic interplay of power and knowledge as manifestly practiced in the multifarious institutions of governments, prisons, hospitals, and schools are delegated to produce certain "truths" about humanity, whose implicit effect is to categorize, normalize and discipline mankind. It is indeed in this form of interplay that one finds the power of *logos* losing its grip and soon vanishing in thin air: "As the archaeology of our thought easily shows," concludes Foucault the last chapter of his book, "man is an invention of recent date. And one perhaps nearing its end." (Foucault 2005, 422) In so doing, Foucault's socio-political theory pushed to the periphery Man as an eternal and fundamental subject of study. Instead, what remains central to epistemological inquiry regarding human nature, is the historical changes, or what Foucault dubs "epistemic breaks", and the fundamental arrangement of knowledge that constructed the figure of Man as it may lead to his demise with more foreseeable shifts and twists to the archaeology of knowledge.

In his seismic essay *The Animal That Therefore I am*, Jacques Derrida carries out a decentralizing attack on essentialists by deconstructing the long-standing human/animal divide, which informed the core principles of the metaphysical philosophy. For Derrida, this form of binary thinking is altogether erroneous for many reasons, the most important of which is that it falsely assumes that only humans are capable of rationality and awareness of their existence (as in his response to Heidegger's notion of the *Dasein*); whereas, in human thought, animals are all registered as belonging to one single category, ignoring all the staggering variations in the animal kingdom (2008, 38). Having said this, we can now better understand Foucault's conception of human nature as a construct, because despite the diversity Man observes day in and day out in the faunal realm, the form of logocentrism devised by mankind to speak of, write about, and propagate their existence is heavily contingent on the success of the binary thinking endeavor. This entire process is accomplished rather discursively rather than physically. Thus, the wider the ideational schism between the two creatures (animal vs. human), the more central status Man obtains *vis-à-vis* the presence of all living and non-living things. To expose the frailties and limitations of the essentialist discourse, Derrida puts forth a question, borrowed from Jeremy Bentham, that would dismantle logocentrism, breaking it down into the smallest foundational components for a systematic deconstruction and morphological analysis:

“Can they suffer?” amounts to asking “Can they not be able?” And what of this inability [*impouvoir*]? What of the vulnerability felt on the basis of this inability? What is this nonpower at the heart of power? What is its quality or modality? How should one take it into account? What right should be accorded it? To what extent does it concern us? Being able to suffer is no longer a power; it is a possibility without power, a possibility of the impossible. Mortality resides there, as the most radical means of thinking the finitude that we share with animals, the mortality that belongs to the very finitude of life, to the experience of compassion, to the possibility of sharing the possibility of this nonpower, the possibility of this impossibility, the anguish of this vulnerability, and the vulnerability of this anguish. (Derrida, 2008 : 28)

Through the paradox of suffering and power, Derrida revealed the incongruities that lie in the logocentrism of power. So, the inability inherent in the sufferer’s experience is self-revealing evidence that the afflicted is able to experience suffering only insofar as he is incapable of warding off the evils thereof from his life. Considering the universality of affliction and anguish, the cogito binary indeed becomes obsolete and impractical in a world where the animal and human share at least one thing in common. What this practically means is that if an animal is trapped in a deep hole, unable to free itself, speak or demand justice from whoever caused the calamity. Nevertheless, the animal is still undeniably capable of only one thing: to feel suffering. The invulnerability characteristic of the suffering mode of expression transposes human nature from the realm of indifference to that of ethical obligation. To reiterate, the implications of Derrida’s philosophy are to be sought in Man’s loss of agency not only to governmental institutions, but this time to a functional side of the dichotomy (animal). Authority is thus dispersed in an anatomical construction of knowledge drawing on taxonomy rather than archeological hierarchy.

Poststructuralist philosophy aimed at decentralizing human agency from the microcosmic world so that meaning formation/apprehension acquires dynamicity. Meaning has to be made abundantly available in the writing space, which furnishes a zone of contact between the reader and the text. Without this bold step, the semantic affordances of a literary text will forever be enclosed in the exclusively rich intentions of the author. If our conception of human nature no longer allows the emergence of this form of agency, then meaning will likely float from texts unrestricted by an authority extraneously untextual in nature. To this very end, Roland Barthes wrote his short essay to announce the death of the author, almost a century after Nietzsche declared the death of God. No doubt that the death of the latter presupposes the end of Man’s metaphysical existence as promulgated by the Aristotelian, biblical traditions and Enlightenment traditions. Man, in fact, derived agency from the exceptionalism that the biblical narratives established textually and practically. Therefore, it is crucial to scrutinize how Barthes destabilized textual authority, shifting to a new interplay between the reader and the text.

What is peculiarly intriguing about Barthes’s postulation is that the author did not need to be decentralized as in the critical framework of Foucault and Derrida. Instead, the Barthesian poststructuralist realization had to deliver the author to the netherworld so that its throne passes to a more semantically effective and pragmatic agency: the text-reader function. The practicality of this new attitude, calling for a reading of a text as though its author were absent, lies in a twentieth-century paradigm shift in the conception of text from “a line of words releasing a single ‘theological’ meaning,” into “a multi-dimensional space in which a variety of writings, none of them original blend and clash.”

(Barthes 1977, 146) Consequently, the text turned into a boundless space where the reader-text function can operate across multiple levels, informed by the multifarious idiosyncratic experiences of the reader, shifting the focus away from the imposing limitations of the author figure on the text. Having outlined these foundational connections, it may be worthwhile to highlight the main Barthesian postulates that will function as recurring touchstones in the discussion of authorship in Large Language Models (LLMs).

- **Postulate #1:** The anonymity and autonomy of writing
- **Postulate #2:** The author as a historical construct
- **Postulate #3:** The fallacy of biographical criticism
- **Postulate #4:** The primacy of language over the author
- **Postulate #5:** The author as a linguistic function, not a personal identity
- **Postulate #6:** The text as a multi-dimensional space of intertextuality
- **Postulate #7:** The reader as the site of meaning, not the author
- **Postulate #8:** The birth of the reader at the expense of the death of the author

Some of these principles can be understood in the light of the previous discussion on the essentialist and poststructuralist philosophies. Others will be dealt with in more detail when applied to non-human entities assuming the author-function such as ChatGPT-4 and Deep Seek.

3. Human Nature from a Scientific Perspective

In the twenty-first century, authorship could be approached by asking three intrinsically pertinent questions: Can LLMs outpace human intelligence in literary creation? Can LLMs author influential literary works such as Shakespeare's *Othello*, for instance? To what extent can a LLM be successful in depicting human experience in writing? Undoubtedly, the three questions seem to reflect a post-modern attempt to resurrect the author into its original status, however not in the metaphysical sense that champions the transcendence of meaning creation and promulgation. What I intend to do with these three questions is to highlight a disconcerting issue at the heart of the act of writing per se: If the rise of the reader occurs at the expense of the death of the author (postulate #8), what concessions must we, as readers, make in order to engage with a text generated by a non-human agency, such as those algorithmic models, which bear no inherent relation to the human condition? In this section, I will try to dwell on some theoretical insights from cognitive science to address questions (2) and (3). The first question, despite the positive answer that it can receive, I believe that the discussion of question (3) might be able to cancel out any literary value to be perceived by the human reader, even in the event of LLMs succeeded in outsmarting human literary genius.

LLMs will be able to generate literary texts, reflecting authentic human experience, such as those emanating from Shakespeare's and Leo Tolstoy's geniuses, only and only if the designers of these models could find practical ways to algorithmicize phenomena resting at the heart of being human: embodies cognition, theory of mind (ToM) and visuo-spatial cognition. This is in no way a comprehensive list of the special gadgets our species carries with them on the daily basis to go about being, acting and reacting as human agents. Let me introduce these three peculiar parameters briefly and I will dwell on each of them more elaborately as the paper progresses. Through embodied cognition our bodies interact

with the environment and feed those sensational experiences into our minds. For instance, research in cognitive science shows that gesture-based learning when solving mathematical problems fosters effective learning and retention. The link between how the body moves and the type of mental experiences our brains generate is pivotal to our understanding of language production. Just try to imagine how could Shakespear inspire the mastermind of evil, Iago, to produce a metaphor such as this with an impaired cognitive system incapable of embodied or visuo-spatial cognition: “O beware, my lord, of jealousy: / It is the green-eyed monster which doth mock / The meat it feeds on.” (*Othello* 1994, 131) Shakespeare’s construction of Iago can be deemed as a masterful application of Theory of Mind, showcasing an intricately sophisticated understanding of human cognition and manipulation. Iago’s ability to infer and impute mental states to others, coupled with his exceptional ability to predict their thoughts and motivations, situates him as a character of unparalleled psychological depth. Mentalizing does not only grant Iago an almost preternatural control over those characters around him, but also underscores the paradox of his intellect as he oscillates between pure ingenuity and destructive evil. How could a language model be expected to produce a peculiarly human experience such as this?

3.1. Passing the Embodied Cognition Test

The human body is not only a site of pain and pleasure; it is rather crucial as a tool for the production of our thoughts as we transform them into linguistically comprehensible outputs. Imagine two undergraduate students exchanging ideas about the submission of the mid-term paper, and while the discussion is going about, one of the students uttered a statement such as “the deadline for paper submission is fast approaching”. Without any shred of doubt, we all agree that underlying this simple syntactic structure, there is something peculiarly human to it—which is our experience of deadlines as approaching as though on a timeline or on a straight track moving fast toward us, while we stay put, static and helplessly unable to manipulate our spatial presence. Similarly, imagine that the sentence was an original utterance that has never been spoken before and a LLM was able to produce it in a particular context. Doesn’t this make us wonder and ponder the authenticity of the sentence and the details it provides? Does the utterance not make us feel funny about the fact that a LLM attributing exceptionally human experience to itself? Is it not the same feeling that we would have if a human being did literally mean that they could really fly, instead of using the syntactic structure to convey a rather metaphorical meaning?

A century ago, these and many other similar queries could have been mere philosophical introspections exclusively brought to the attention of philosophers and literary critics, arguing for the demise of the author. Research in cognitive science, when applied to literary criticism, cannot support the Barthesian or Foucauldian postulation for the simple reason that the author matters; and that his conscious/unconscious thoughts before and during the creation of the narrative must be recognized as integral to harrowing an epistemological terrain where the reader is scaffolded to engage with the text hermeneutically. Almost two decades ago, Lawrence T. Williams and John A. Bargh from the university of Colorado drew on the findings of a research paper corroborating the

significant role the insular cortex⁴ plays in processing physical temperature and interpersonal warmth (trust) information (2008, 606). The experimenters designed two studies in which they hypothesized that since the same region in the brain is responsible for physical warmth and the feelings of warmth humans feel in an interpersonal interaction with others, then what should follow from this is that: if the experiment subjects were made to experience physical warmth through holding a cup of hot (versus iced) coffee and then they were later asked to judge the personality of a target person, the insular cortex should be able to map out physical warmth (from the cup) onto an interpersonal emotive analog registered toward another person. In another study, participants were asked to conjure moments when they were socially included or excluded. To test the insular cortex's ability to map out the emotive social onto the physical, the experimenters told the participants that they had issues with the central air in the building, and they were accordingly questioned on how warm or cold the temperature of the room was. In both experiments the findings were in support of the main hypothesis: attributing warmth and care to others was associated with bodily physical warmth as experienced from the cup of coffee. Additionally, physical detection of room temperature depended on the participants' thoughts at the moment of experimentation whether they envisioned themselves as lonely (excluded) or in the company of others (included). All these cognitive processes took place right in the insular cortex.

Given the findings of these narrations, it seems invariably impossible to separate linguistic cognition from embodied cognition. As demonstrated by George Lakoff and Mark Johnson (1980) and John Bargh (2006), metaphors are not self-contained ideas standing in isolation from the cognitive underpinnings of the person producing them; they are rather inextricably connected networks of concepts and experiences generated through our quotidian engagement and interaction with both the environment and society. Metaphors are, in fact, actual vessels that we deploy not only to subjugate abstract thoughts to the power of expression, but also to be able to understand the world around us in line with how our brains are hardwired to comprehend physical phenomena. Behold how does Romeo choose to unravel the unspeakable power of his love for Juliet and her ineffable beauty, through bodily sensations (of warmth), so that he can cross the constraints of mental abstraction to the multifarious affordances of metaphorical expression (experienced through the physical body): "O she doth teach the torches to burn bright! / It seems she hangs upon the cheek of night / As a rich jewel in an Ethiop's ear / Beauty too rich for use, for earth too dear." (1994, 97). Conversely, the emotive effect of King Lear's descent into madness was created through an imagery of coldness and stormy weather: "Blow, winds, and crack your cheeks! Rage, blow, / You cataracts and hurricanoes, spout / Till you have drenched our steeples, drowned the cocks!" (1990, 177).

If a LLM is given the right instructions by a human agent, it could possibly generate statements as penetrating as those of Shakespeare. However, when the literary text is exported to the reader, the latter engages in readership hermeneutics whose success and pertinency call for a questioning of the anonymity and autonomy of writing as neutral, composite and independent of the author's identity (postulate #1) and the primacy of

⁴ The Insular Cortex (or insula, for short) is a region in the brain located in the lateral sulcus separating the temporal lobe from the parietal and frontal lobes). Its role in perceiving internal bodily states such as hunger, thirst, heartbeat; emotional processing; pain perception is well documented.

language over the author (postulate # 4). Isn't language to the author more than an instrument to express his thoughts? It is indeed a reflection of a quite sophisticated neuronal activity in the insular cortex underlying the metaphorical meaning devised by the author. In other words, the writer is not only the authority who dictates the order of words in the narrative and the meanings the word order is meant to convey; more than this he is the actual embodiment of the cognitive network of insular maps as they communicate with one another to engender a particular mental effect (thoughts and ideas in their metaphorical cloak, for instance). One may rightly argue that a narrative is not the sum of its metaphorical expressions, and that authors may choose to deploy some metaphors only to convey a small portion of their ideas. Although this form of rebuttal still cannot detract the pivotal role the author plays in bringing forth abstraction in a form most compatible with our embodied cognitive systems, a clarification and reminder is needed that the metaphor which has been subject to studies in cognitive linguistics is not the figure of speech that high school students are asked to chase in literary texts along with other forms of figurative language. Metaphor, in Lakoff's and Johnson's conceptualization, is a whole organizing conceptual system that our minds utilize when creating representations about the world as well as when attempting to understand its structures (1980, 6).

The point that this line of argumentation intends to make is that a LLM will fail miserably to communicate genuine human experience in narrative form by relying solely on syntactic systems for generating language output. The language model might be able to pass the Turing test, but can never be able to solve John Searle's Chinese Room thought experiment without a well-established system capable of both syntactic and semantic incorporation. It follows that the LLM-author, at this stage, won't be able to allow for the Barthesian postulates 1 & 4 to materialize without any issues, corroborating our hypothesis that the author's role in the narratological production is significantly crucial and that computational simulation ill-affords the necessary prerequisites to dethrone "the subject of enunciation", as the Foucauldian nomenclature goes.

3.2. Passing Theory of Mind (ToM) Test

Shakespeare's Iago is not a character that we are unlikely to encounter in real life. Nevertheless, dissimilar to all other fictional and non-fictional characters, his peculiarity was made heavily contingent on the complexity of the characterological network that Shakespeare devised to engineer an antagonist of which R. A. Foakes asserts: "Iago sees the worst in the other characters, and reveals their vices and follies; but he goes beyond the satirist in using this exposure to destroy others." (2010, 15). In these profound words, we incidentally find an opportunity to revisit and evaluate another Barthesian postulate (#3). True that biographical criticism depends largely on reading the author's work with the intent to unravel as many details about the writer and the influence thereof on the writing process as possible. I deem this a classical version of what the biographical approach should be truly about. I shall therefore contend that the biographical analysis of a literary work should be able to assert the identity of the author by allowing access to the author's cognitive mind (not biographical mind). Put differently, instead of seizing the literary work as an opportunity to read the psychoanalytic insights of/about the author, we should rather aim at explicating Shakespeare's cognitive biography that enabled him to construct an Iago character performing a sophisticated form of mentalizing (*Theory of Mind*). Another important point in Foakes's assertion is that Shakespeare's Iago does not

only reveal the mental frailties of the other characters in the play, but he also exploits the limitations of their cognitive systems to subvert their agency and lead to their existential downfall. The significance of this remark will force an evaluation of the author's role as the architect of the mental imputations Iago attributes to the other characters in the play (hence a reconstruction of Shakespeare's understanding of mind-reading) as well as the origin of intentionality as it drives Iago's tragic plot to unintended outcomes. Here, I shall argue that the presence of the unintended in *Othello* serves as a proof of intentionality in Iago, which is the byproduct of Shakespeare's cognitive biography. Furthermore, it substantiates the playwright's agency, as Shakespeare intentionally orchestrates the narrative to unfold in a way clearly intended to subvert Iago's mis-calculated schemes. To begin, let's contemplate what Paul Cefalo dubs "Iago's Dilemma":

Iago's dilemma lies between theory and practice. He has the unrelenting inclination, even ethos, for mind reading, but he cannot easily or accurately do so in practice (improvisation is certainly one of his manifest talents, but his very need for it belies his inability predictively to gauge how his projects will unfold). His challenge is the inescapable, generic problem of other minds, a challenge that in the world of the play transmutes into a curse. (Cefalo, 2013: 269)

The classical definition of theory of mind refers to our ability to impute mental states to ourselves and others. By definition, then, anyone intending to manipulate the mental states of others, needs to attune his mind on reading as many behavioral clues as possible, in order to be able to plan and strategize effectively. However, engaging in mind-reading and mentalizing about other people's thoughts will never guarantee the accuracy and correctness of the undertaking's outcomes. In other words, we will never be able to know for certain that our predictions are true or false until we receive feedback from the environment. Evolutionary psychologists trust that ToM evolved to serve survival purposes and hence the limitations inherent in mind-reading are largely due to the hyper state of alertness our cognitive mechanisms triggered in the presence of a potential predator. "Our evolutionary past," Lisa Zunshine asserts, "ensures [...] that you will intuitively expect me to "read" your body as indicative of your thoughts, desires, and intentions." (2010, 120) Consequently, under those primordial evolutionary conditions, the brain had to quickly scrutinize behavioral and bodily clues to anticipate a worst-case scenario—to thwart the possibility of being caught off guard. This partly demystifies Cefalo's term "dilemma": Shakespeare never intended for Iago to read the other character's minds as an omniscient narrator would, simply because such a narratological choice will grant Iago unnecessary superpowers, undermining the plausibility of his human nature. Moreover, in literary criticism, theory of mind should primarily be concerned with reflecting the intentionality governing the author's cognitive apparatus's underlying mechanisms, encompassing his understanding of social cognition, and his exceptional ability to translate this into a narrative form articulating the sophisticated network of contending intentionalities.

In relation to the main analysis of this section, two key questions must be revisited: Are LLMs capable of ToM? Do current LLMs (such as GPT-4 and Deep Seek) exhibit any evidence of intentionality? What insights does research in cognitive science tell us about the human brain's capacity for ToM and intentionality? These research inquiries have been addressed by numerous cognitive and computational scientists; however, literature exploring their implications for resurrecting the author, from a mystified death to a position of centrality, remains largely absent. The reason for entertaining these questions is to accentuate over

and over again the significance of the human author's cognitive role in the narrative process. Our argument also seeks to position the reader's cognitive manipulation of the literary work as secondary to that of the author. Cognitively speaking, the play—particularly in Shakespeare's case—emerges from an active interaction between the playwright's specific cognitive regions, which process vast inputs that constitute Shakespeare's global knowledge of the world (as described in cognitive linguistics), and those involved in the creation of extensive networks of intentionalities within and between characters. In contrast, the cognitive burden assumed by the reader is distinct, involving divergent processes for deciphering meaning and understanding intentional networks. It is important to note that even the intentions of the reader and playwright diverge: while the reader is focused on connecting the dots and grasping semantic networks (epistemologically dependent on the author), the playwright is focused on creating and manipulating meanings and perceptions (epistemologically independent and ontologically empowered).

The pleasure readers find in reading fiction is often attributed to the brain's insatiable hunger for attributing mental states to others and attempting to read their minds based on the behavioral cues people demonstrate. Our current appreciation of Iago's mental manipulations could be appreciated for the same reason. In her writings, Zunshine raises some awareness about the availability of the body as a text ready for interpretation; however, she warns against putting unchecked trust in the behavioral clues to betray the mental content of the other people's minds. The caveat is established on the basis that our mirror neurons may receive and interpret the body-as-a-text clues in ways not originally intended by the performer. Therefore, our neuronal reaction to the narratological situational context will result in an erroneous output. This could be a proof enough to discredit the reader-function that Barthes did relish to advance at the expense of the death of the author. Nevertheless, a better argument to lift the burden off the reader's shoulders is to contemplate the fact that the reader's role in the act of narration is to indulge his mind in an endless cycle of mind reading exercises throughout the reading process. The practice could be enjoyable so much so that the reader decides to predict and even anticipate the foreseeable actions of the characters. However, for all practical purposes, the reader should not transgress against the author-function, pretending to be able to read the playwright's mind—dethroning the author is a reckless attempt to obliterate the force that unifies the text, reader and hermeneutics. Put differently, the mind of the author is not as much expected to engage in mind reading as it is forced to understand the mechanisms by means of which human beings mentalize. Without the intricate complexity of such authorial observations and the duplications thereof in the act of narration, the literary work will fail to produce any authentic effect in the reader. Moreover, the mind of the author is to the readers as the beacon light is to the ship navigating a dark sea, guiding their understanding through complex networks of mind-readings and intentionalities. Without the author's intentional insight into human mental processes, the reader is left adrift, unable to navigate the narrative with clarity or depth. Ultimately, the author can never be relegated to the status where he is considered merely as a historical construct (**postulate #2**) or deprived of his epistemic and ontological authority as the originator of the semantic backdrop against which the reader negotiates and constructs his own idiosyncratic narrative(s) (**postulate #7**).

The role of the author cannot be more accentuated than when the author-function is attributed to either a writer afflicted by some kind of cognitive impairment or when the author is a large language model such as GPT-4 and Deep Seek. In both cases, the inherent limitations of authorship reveal the extent to which the author serves as the epistemic and ontological anchor of meaning, complexity, and intentionality within the textual framework. Let's take, for instance, the case of autism. Failing to understand other minds is a key cognitive feature for people who suffer from autism (Baron-Cohen 2001). While chimps can show corroborating the view that they are capable of attributing mental states to human agents in experimental conditions (Premack and Woodruff, 1978; Call and Tomasello, 2008), and that four-year-old children can adeptly pass the false-belief test (Z. Wang and L. Wang, 2015), what quite characterizes children with an autistic disorder is that they fail abjectly to represent other people's mental states (Frith, 1994; Tager-Flusberg, 2007). Cognitive psychologists and neuroscientists tell us that imputing intentional causality to others is achieved by coordinated connectivity between three regions in the brain: temporo-parietal junction, right inferior frontal gyrus and left premotor cortex (Rajesh et al, 2014). What seems to be challenging to our hypothesis is that LLM are equally capable of passing the false-belief tests, that chimps would fail, despite their lack of all these anatomical structures. In a recent study, experimenters tested a number of LLMs with a battery of forty (40) false-belief tasks containing a wide range of characters and scenarios similar to those used to test ToM in humans. The results supported the hypothesis that LLM's progress in solving ToM tasks was on the rise, when the experimenters compared the evidence from the early models, GPT-1, GPT-2XL and GPT-3 with those performed by ChatGPT-4: "GPT-davinci-0003 and Chat-GPT-3.5-turbo successfully solved 20% of the tasks. The most recent model, ChatGPT-4, substantially outperformed the others, solving 75% of tasks, on par with 6-y-old children." (Kosinski, 7)

Given these experimental results, can we conclusively assert that there is nothing inherently peculiar about the human author, thereby reverting to the obituary of authorship suggested by Barthes and Foucault? Such a conclusion remains premature, as the empirical data merely indicate that LLMs are capable of processing vast amounts of syntactic structures to generate responses within a given social cognition scenario. In contrast, human cognitive processing involves not only the reception of syntactic input but also its semantic negotiation to facilitate comprehension. This process, I refer to as defiance, entails an active imposition of interpretation onto linguistic input. Humans engage in this interpretative act to extract meaning, resolving the psychological tension that arises when a sequence of utterances remains in a state of discordance and incomprehensibility. While philosophers, such as John Searle, may frame this phenomenon in terms of "intentionality," I prefer the term defiance for two reasons. Firstly, both anthropological and linguistic research, alongside common-sense observation, reveal an intrinsic human drive for meaning-making, as evidenced by the foundational role of myths and religious narratives in structuring human understanding of the world. In other words, there is an urge inside us to make meanings out of even meaningless input (defying absence of meaning). Secondly, defiance presupposes the existence of multiple competing interpretative possibilities, with only one ultimately surfacing into human consciousness. For instance, when viewing an image of an apple the brain does not only contrast it with other types of apples and fruit stored in the long-term memory, but it moreover attempts to identify the apple with a margin of error close or equal to nought. The effort put into

the mental processes of this kind is justified by the brain's urge to defy any possibility of mis-identification. In evolutionary terms, reckless mental processing of input may jeopardize the organism's survival. This form of intentionality is yet to be developed and explored in LLMs. Despite the latter's ability to identify objects by using image recognition processes, the LLM does not possess a system capable of intentional thought and action.

Defiance, although a theme tackled in specific contexts in psychology and political theory, is not a concept without biological truth. Evolutionary biologists and geneticists acknowledge the fact that paternal and maternal genes may enter into a tug-of-war game over which allele should be repressed. The IGF2 (Insulin-like Growth Factor 2) gene is expressed from the father's allele because it promotes the offspring growth at the expense of the mother's resources. Consequently, the mother's genes operate to suppress this effect. This is commonly known as the Genetic Conflict Hypothesis. It clearly explains that our genome is hardwired to engage in defiance by virtue of specific mechanisms and guided toward particular evolutionary goals. The absence of defiance in genetic imprinting may cause Prader-Willi Syndrome (PWS), where the child receives the maternal gene that suppresses excessive resource demand, whereas the paternal gene that promotes feeding and growth is missing. Almost the same version of defiance is required in the brain of a paranoid schizophrenia patient, where the latter fails to deactivate a hyper-activated intention detection system. While a healthy individual reserves attributing intentions only to humans engaged in private or social events, the paranoid patient might interpret an event of a book falling off a table as someone deliberately causing the book to fall. Bara et al conducted an experimental study in which they used fMRI neuroimaging to screen the brain regions responsible for ToM in healthy individuals and whether these regions are deactivated in non-social interactions in the case of paranoid patients. The results interestingly confirmed their hypothesis that "paranoid patients may have a hyperactive intention detector when they observe physical events, i.e., in our Ph-C condition they do not deactivate their intention detector." (2011, 6) In fine, defiance is not only conceptual in nature, but it is also inherently embedded in the underlying biological mechanisms of the living organism.

Now, let us imagine how different John Searle's Chinese Room thought experiment could have been had he equipped the man locked inside the room with a program that will not enable him to put the sets of squiggles together but to defy the possibility of another formal combination to ever emerge—that is to say, instead of having only the program and the batch of cards to be combined in many different ways, we would have another algorithmic structure that would defy the various combinations of the program in favor of only one syntactic structure. Regardless of his ignorance of Chinese, the man will need a second English guidebook demonstrating the different possible combinations and the underlying reasons for allowing one to materialize at the expense of the suppression of all the other possibilities. A defying system like this will still be formal, since the locked man will remain ignorant of the Chinese language. However, we cannot ignore the fact that some elements of semantics will have started to appear in the system, which previously was wholly syntactic in principle. To illustrate, if, according to the rules of the program and the available squiggles, the number of combinations is three; the defying system will have to reduce this number to only one plausible event and discard the other two as irrelevant.

The previous articulation of the problematic appears as though arguing for Douglas Hofstadter's view of analogy as the core mechanism of cognition and human thought. The defiance thesis, herein postulated, joins Hofstadter's argument that human cognition is far more than about logical reasoning and formal rules, as the tradition view goes. However, Hofstadter attempts to establish an even deeper ontological foundation to human thought by arguing that the slips of the tongue provide clear-cut evidence about some competing lexical elements, which resolve the emergent conflict in the form of lexical blending. For instance, in their suggestive book *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking*, Hofstadter and Emmanuel Sander give an example of a dean who's in the event of expressing a future goal in an enthusiastic manner, spit out the following utterance: "We'll pull no stops unturned to get him to come here". Admittedly, the dean of the university intended to deploy either of the two idiomatic expressions: 'pull out all the stops' and 'leave no stone unturned'. For Hofstadter et al, the two stock phrases showed up in the dean's mind at the same time, and a state of rivalry soon emerged between them. The conflict was resolved by allowing both of them to emerge in the form of lexical blending: 'pull no stops unturned' (2013, 262)

Although the defiance thesis acknowledges the presence of rivalry between concepts, its core tenet is predicated on the condition of '*the excluded second*', which further underscores the forceful rise of only one concept at the expense of an obliterated other. Lexical blending is a reconciliatory concept whose function is based on combining two contending syntactic structures; however, the semantics of the utterance will always remain immune to change. Whether one use the first, second or the blended idiomatic expression, the intended meaning and its effect will be the same. Furthermore, when explored singularly, the three expressions are foregrounded in George Lakoff's *conceptual metaphor*. While the latter fail to capture all everyday utterances, the defiance thesis argues that even in the 'Good morning' greeting one should assume the presence of an underlying form of mental rivalry. What is actually concealed from the mind of the speaker upon uttering the greeting to a passerby are a number of questions: Why should I greet in the first place? How should I greet this person (as different from all the other persons I know)? When my greeting will make more sense and procure less misunderstanding? Which form of body language should accompany the utterance? The finger print machine, which scans my finger every day at work, is mindless of all these communicative protocols and ramifications, for it has been greeting me with the same utterance, tone and moment for more than nine years. One may rightly argue that other more advanced machines can alternate between several forms of greeting. True that this is possible, but the point that I am trying to make here is that this form of alternation is still mechanistic and based solely on formal rules. In contrast, the questions above show that the interlocutor engages in a mental act of meticulously weighing different options/possibilities based on specific situational contexts, before a quasi-conscious decision to suppress the many and grant access to only one.

Conclusion

In this paper I avoided listing the arguments offered by computer scientists and LLM experts, for despite the impaired intelligence of language models compared to that of humans, the experts of the field will remain faithful to futuristic wishful thinking about the

promising advances their models will be able to achieve. Such scholarly speculations may indeed prove to be true insofar as the LLMs will be able to produce peculiarly AI literary narratives, of the same quality as a calculator generating peculiarly mechanistic calculations, faster and more precise, but devoid of any experiential value (in the psychological sense). More and more papers are being penned down and published to exonerate LLM and Generative AI from any false rumors about their future incapacity to beat human non-Artificial intelligence. These arguments are deemed irrelevant and unyielding to the main discussion of this paper: Is the author really dead?

Instead, this paper aimed at refuting the eight postulates extrapolated from Roland Barthes's essay. Drawing on theoretical insights from cognitive sciences and neuroscience, we are unlikely to think of the author as a linguistic function (postulate #5) or the text as a space of intertextual contact (postulate #6). The author is a cognitive phenomenon where a number of brain regions is recruited to perform particular tasks. The absence of any of these regions or the impairment thereof would be reflected in the narrative proper. When Barthes and Foucault published their papers, the structural school perceived language as a set of syntactic rules governing the putting together of a batch of linguistic elements. With the advent of generative AI, and now LLMs, a return to the poststructuralist claims about the death of the author became a pressing necessity. If ChatGPT-4 and Deep Seek are still failing to respond correctly to the most mundane ToM tasks, we are obligated to resurrect the author to its original status as the producer of meaning. My view does not imply that the reader's role in literary hermeneutics is to be swept under the rug. The reader would still preserve the right to extrapolate the semantic ontologies of the text; however, this time by envisioning even the reader as a cognitive phenomenon, following the same previous principles we applied to restore the author to his throne.

Similarly, the text is essentially the result of cognition, mentalizing and most importantly defiance. The short story heading the first section of the paper illustrates clearly that plot structure, superfluous number of characters, the unfounded morality of the story, testify to the seven-year-old author's underdeveloped mechanisms of defiance. Had the writer been able to develop a sense of defiance, she could have better adjusted characterology, or balanced the role of each character so that none of them is unnecessarily dysfunctional. Moreover, the writer could have thought out the different possibilities the narrative could foreground its moral message, and thus defy all the other emerging possibilities for a chosen one. Below is a simple ToM scenario that I put together to test the extent to which Chat-GPT-4 and Deep Seek are capable of mentalizing about other people's mental states:

Read the scenario and answer the question:

Ammy entered an empty room. She put two boxes in there (green and blue). Johnny entered the same room sometime after, and placed his red ball in the green box. Peter entered the room right after Johnny left and he took the red ball and placed it in the blue box. When Darwin came to the room to look for the red ball, where do you think he will be looking for it?

Neither of the LLMs was able to provide the correct answer. Their failure is yet indicative of the LLMs's inability to address some inference and mind-reading scenarios. Answering

the pervious scenario will require the model to be able to imagine (or conjure) a number of possibilities whereby each character's cognitive space is configured as "know" or "does not know". While the reader is weighing the epistemic content of each character, he should additionally defy the possibilities of "knowing" and "not knowing" at equal footing by invoking purposefulness through the "why" question: If Darwin knows, then why does he know what he knows? If Darwin does not know, then why does he know what he does not know? Furthermore, the reader should be able to infer that Darwin is a new character, which has just been let into the room, and that he did not observe anyone before placing or moving the ball, so he could not have known in which box the ball was kept.

The only recommendation this paper seems to be entitled to make is to draw the attention of neuroscientists and cognitive scientists to the power of a cognitive ability that all humans possess, which I referred to throughout as "mental defiance". Designing short stories which require various degrees of defiance from the writer as well as the reader for the purpose of identifying the brain regions recruited for this cognitive ability will provide important breakthroughs into the adequate ways to translate the cognitive mechanisms into structures fathomable by a large language model. Until then, Roland Barthes's author will have to be resurrected to assume his many functions.

Bibliographical references

- Anonymous. *The Epic of Gilgamesh*. Translated by Andrew George. London: Penguin Classics, 2000.
- ARISTOTLE (384 BCE). *Nicomachean Ethics*. Translated by Robert C. Bartlett and Susan D. Collins. Chicago: The University of Chicago Press, 2011.
- BARA B.G., et al. "Intentional Minds: A Philosophical Analysis of Intention Tested Through fMRI Experiments Involving People with Schizophrenia, People with Autism, and Healthy Individuals." *Frontiers in Human Neuroscience* 5, no. 7 (2011): 1-11. doi:10.3389/fnhum.2011.00007.
- BARGH J. "What Have We Been Priming All These Years? On the Development, Mechanisms, and Ecology of Non-Conscious Social Behavior." *European Journal of Social Psychology*, no. 36 (2006): 147-168.
- BARON-COHEN S. "Theory of Mind and Autism: A Review." *International Review of Research in Mental Retardation*, no. 23 (2001): 169-184.
- BARTHES R. (1915). *Image, Music, Text*. Translated by S. Heath. London: Fontana, 1977.
- BECKETT S. "The Unnamable." In *The Unnamable*, edited by Steven Connor. Faber and Faber, 2012.
- BERGEN B.K. (1972). *Louder Than Words: The New Science of How the Mind Makes Meaning*. New York: Basic Books, 2012.
- BOZKURT A., et al. "Speculative Futures on ChatGPT and Generative Artificial Intelligence (A): A Collective Reflection from the Educational Landscape." *Asian Journal of Distance Education* 18, no. 1 (2023): 53-130.
- BREITHAUPF F. "Humans Create More Novelty Than ChatGPT When Asked to Retell a Story." *Science Reports* 14, no. 875 (2024): 1-11. <https://doi.org/10.1038/s41598-023-50229-7>, consulted on 7/02/2025)
- CALL J. and TOMASELLO, M. "Does the Chimpanzee Have a Theory of Mind?" *Trends in Cognitive Science*, no. 5 (2008): 187-192. Doi: 10.1016/j.tics.2008.02.010.
- CEFALO P. "The Burdens of Mind Reading in Shakespeare's Othello: A Cognitive and Psychoanalytic Approach to Iago's Theory of Mind." *Shakespeare Quarterly*, no. 64 (2013): 265-294.
- DERRIDA J. (1930). *The Animal Therefore I Am*. New York: Fordham University Press, 2008.
- DORANTI A. "The Audience as Co-Author: An Introduction." *Text—Interdisciplinary Journal for the Study of Discourse* 6, no. 3 (1986): 239-248. <https://doi.org/10.1515/text.1.1986.6.3.239>, consulted on 13/10/2024)
- FOAKES R.A. "The Descent of Iago: Satire, Ben Johnson, and Shakespeare's Othello." In *Bloom's Modern Critical Interpretations*, edited by H. Bloom. Infobase Publishing, 2010.
- FOUCAULT M. (1926). *The Order of Things: An Archaeology of the Human Sciences*. London: Taylor and Francis, 2005.
- FOUCAULT M. "What is an Author." In *Reading Architectural History*, edited by D. Arnold. Routledge, 2002.
- FRITH U., HAPPÉ, F., and SIDDONS, F. "Autism and Theory of Mind in Everyday Life." *Social Development* 3, no. 2 (1994): 108-124.
- HENRICKSON L. "Chatting with the Death: The Hermeneutics of Thanabots." *Media, Culture and Society* 45, no. 5 (2023): 949-966. doi:10.1177/01634437221147626.

- HO M.K., et al. "Planning with Theory of Mind." *Trends in Cognitive Sciences* xx, no. xx (2022): 1-13. <https://doi.org/10.1016/j.tics.2022.08.003>, consulted on 15/12/2024)
- HOBBS T. (1588). *Leviathan*. Edited by J.C.A. Gaskin. Oxford: Oxford University Press, 1998.
- HOFSTADTER D. (1945) and SANDER, E. (1961-). *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking*. New York: Basic Books, 2013.
- KANA R.K., et al. "Functional Brain Networks and White Matter Underlying Theory-of-Mind in Autism." *SCAN*, no. 9 (2014): 98-105. doi:10.1093/scan/nss106.
- KOSINSKI M. "Evaluating Large Language Models in Theory of Mind Tasks." *PNAS* 121, no. 45 (2024): 1-11. <https://doi.org/10.1073/pnas.2405460121>, consulted on 15/12/2024)
- LAKOFF G. (1941) and JOHNSON, M. (1949). *Metaphors We Live By*. Chicago: The University of Chicago Press, 1980.
- LAKOFF G. "Explaining Embodied Cognition Results." *Topics in Cognitive Science* (2012): 1-13.
- NEHAMAS A. "What an Author is?" *The Journal of Philosophy* 83, no. 11 (1986): 685-691.
- PREMACK D. and WOODRUFF, G. "Does the Chimpanzee Have a Theory of Mind?" *The Behavioral and Brain Sciences*, no. 4 (1978): 515-526.
- SAXE R. and KANWISHER, N. "People Thinking About Thinking People: The Role of the Temporo-Parietal Junction in 'Theory of Mind'." *NeuroImage* 19 (2003): 1835-1842. doi:10.1016/S1053-8119(03)00230-1.
- SEARLE J.R. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3, no. 3 (1980): 417-457.
- SHAKESPEARE W. (1564). *King Lear*. Edited by J.L. Halio. Cambridge: Cambridge University Press, 1990.
- SHAKESPEARE W. (1564). *Othello*. Edited by N. Sanders. Berkeley: University of California Press, 1994.
- SHAKESPEARE, W. (1564). *Romeo and Juliet*. Edited by B. Evans. Cambridge: Cambridge University Press, 1994.
- TAGER-FLUSBERG H. "Evaluating the Theory-of-Mind Hypothesis of Autism." *Association for Psychological Science* 16, no. 6 (2007): 311-315.
- WANG Z. and WANG, L. "The Mind and Heart of the Social Child: Developing the Empathy and Theory of Mind Scale." *Child Development Research* (2015). <http://dx.doi.org/10.1155/2015/171304>, consulted on 12/11/2024)
- WILLIAMS L.E. and BARGH, J.A. "Experiencing Physical Warmth Promotes Interpersonal Warmth." *Science*, no. 322 (2008): 606-607.
- WILSON A. "Foucault on the 'Question of the Author': A Critical Exegesis." *The Modern Language Review* 99, no. 2 (2004): 339-363. <http://eprints.whiterose.ac.uk/4458/>, consulted on 24/01/2025)
- ZUNSHINE L. "Lying Bodies of the Enlightenment: Theory of Mind and Cultural Historicism." In *Introduction to Cognitive Cultural Studies*, edited by L. Zunshine. Johns Hopkins University Press, 2010.

Appendices

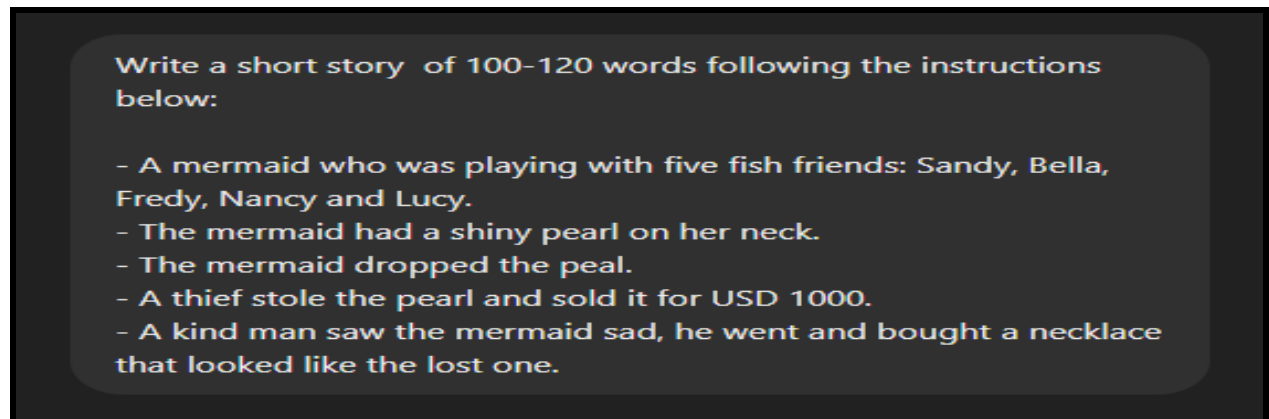


Figure 1: After reading my daughter's original story, I thought of giving a few clues to GPT-4 so it can generate a similar story that I could use later to compare human vs. artificial narratological intelligence.

Lila, the mermaid, swam joyfully through the coral reefs, laughing with her five fish friends: Sandy, Bella, Fredy, Nancy, and Lucy. Around her neck, she wore a shiny pearl that gleamed with the colors of the ocean. Suddenly, the pearl slipped from her neck and fell to the sandy seabed. Before she could grab it, a sneaky thief swam by, snatching the pearl and disappearing. Heartbroken, Lila gazed at the empty spot around her neck. A kind man, noticing her sadness from a boat above, asked what was wrong. Learning of her loss, he bought a necklace with a pearl that resembled the one she had lost. Lila smiled, her heart lightened by his kindness.

Figure 2: The story was generated by GPT-4 after it has been given the writing clues above.

Darwin will be looking for the red ball in the **green box**.

This is because Darwin, like anyone else in a typical theory of mind test, would base his search on the last place he knew the red ball was placed — which is in the green box, where Johnny originally put it. Darwin wouldn't know that Peter later moved the ball to the blue box, as he wasn't in the room when that happened.

Figure 3: The answer was generated by GPT-4 in response to the Darwin's ToM scenario.



Based on the scenario, Darwin will likely look for the red ball in the **blue box**. This is because Peter was the last person to interact with the red ball, and he moved it from the green box to the blue box. Since Darwin is entering the room after Peter, he would expect the red ball to be in the blue box.

Figure 4: The answer was generated by Deep Seek in response to the Darwin's ToM scenario.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author used ChatGPT-4 and Deep Seek in order to create theory-of-mind scenarios to be embedded in the main purpose of the paper. Additionally, the two AI tools were used to revise the grammatical structure of some sentences for clarity and accuracy. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.