New Research in Islamic Humanities Studies
Online ISSN : 2980 - 8944
Lorestan University
https://api.lu.ac.ir/

The Differance Engine

Large Language Models and Poststructuralism

David. J. Gunkel

Professor, Department of Philosophy and Ethics, School of Information and Communication Technology, Northern Illinois University, United States of America. Email: dgunkel@niu.edu

Article Info

Article type:
Research Article

Article history:
Received 28 September 2025
Received in revised form 28 November 2025
Accepted 02 January 2026
Available online 12 January 2026

Keywords:
Artificial Intelligence,
Jacques Derrida,
Différance,
Large Language Models,
Philosophy of Technology,
Poststructuralism,
Semiotics

ABSTRACT

Objective: This paper argues that large language models (LLMs), such as transformer-based architectures like GPT, represent a practical actualization of Jacques Derrida's concept of différance.

Method: Drawing on the framework of poststructuralist theory and semiotics, the study examines how LLMs generate meaningful content by computing statistical differences across vast textual corpora. This analysis foregrounds the computational enactment of processes analogous to Derridean spacing, temporalization, and trace.

Results: The paper demonstrates that LLMs can be understood as "différance engines," computationally enacting the very mechanisms theorized by Derrida. It further elaborates on the philosophical consequences of this alignment, particularly the challenges it poses to logocentrism, authorship, and the metaphysics of presence. In addressing three potential criticisms, the paper contends that applying Derrida's work in this context constitutes not a misappropriation, but a continuation and reiteration of its inherent logic.

Conclusions: The paper concludes by identifying three systemic limitations inherent to this framework and outlines opportunities for future research in this domain. Ultimately, it illustrates how LLMs can be interpreted through the lens of poststructuralist theory, while simultaneously demonstrating how that theory can be clarified and rendered more accessible through the technical operations of contemporary AI.

Cite this article: J. Gunkel, D. (2026). The Differance Engine (Large Language Models and Poststructuralism). *New Research in Islamic Humanities Studies*, 4 (Special Issue), 1-34. https://doi.org/10.22034/api.2026.735002

Publisher: Lorestan University.
DOI: https://doi.org/10.22034/api.2026.735002

Introduction

Jacques Derrida's neologism *différance* (Derrida 1982) is one of the most influential yet notoriously elusive concepts in contemporary critical theory. Derived from the French verb *différer* which, depending on contexts of use, can mean either "to differ" or "to defer" *différance* designates the way that words and other kinds of signs or signifiers acquire meaning not by referring to a stable referent but through its difference from and deferral to other signs within that same system.

Though *différance* was originally introduced and developed in the context of semiotics, poststructuralism, and the phenomenological tradition, its relevance extends well beyond the scope of literary theory and continental philosophy. One unexpected domain in which the concept which Derrida (1982, 3) insists is "literally neither a word nor a concept" finds an uncanny application is large language models (LLMs), such as GPT-type models and other transformer-based architectures. These systems actualize, at a technical level, the very mechanisms of *difference* and *deferral* that Derrida describes and theorizes. That is, they generate seemingly meaningful linguistic content not by referring to or having access to real-world referents but through calculating statistical differences that are discoverable in large sets of textual data.

This essay examines how LLM artificial intelligence (AI) can be seen as mechanisms of *différance* or what the title to this essay calls, in a deliberate repurposing of the work of Charles Babbage, *Différance Engines*. 1) We will begin by identifying and explicating key characteristics of *différance* as developed by Derrida in his titular essay on the subject. 2) We will then show how LLMs, especially transformer-based models, operationalize these concepts through their reliance on statistical difference and contextual deferrals. 3) In the third part we describe and discuss the philosophical consequences of understanding LLM AI and LLM-generated textual content through this deconstructive lens. 4) This is followed by the fourth part where we address three possible criticisms of this effort and provide responses to these potential objections. 5) Finally, the essay conclude by identifying three systemic limitations of the current investigation and by proposing directions for future research that would expand, refine, or elaborate upon the claims advanced here.

Difference

Introduced in a 1968 lecture and essay by the same name, Derrida fabricated *différance* as a deliberate and calculated misspelling of the common French noun "différence," replacing the "e" with an "a." Interestingly this seemingly simple substitution is something that is only available to us in and by writing. It is, as Derrida (1982, 3) explains "purely graphic: it is read, or it is written, but it cannot be heard. It cannot be apprehended in speech."[1] More importantly, however, this

1. This did, as one might anticipate, create some trouble for Derrida in the course of delivering the *Différence* lecture to the French Society of Philosophy. Since the "discrete graphical intervention" that distinguishes *différence* from *différance* cannot be spoken or made audible, Derrida provided his audience with the following warning: "In effect, I cannot let you know through my discourse, through the speech being addressed at this moment to the French Society of Philosophy, what difference I am talking about when I talk about it. I can speak

silent difference a silence, that Derrida (1982, 4) emphasizes, can only function within alphabetic or "so called phonetic writing" is not just a typographic pun or simple play on words; it is a deliberate critical intervention in "the classically determined structure of the sign" (Derrida 1982, 9).Beginning with (at least) Aristotle (1938, 16a), language has typically been understood to consist of signs that refer and defer to things. When we write the words "large language model," for instance, it is assumed that those linguistic tokens stand for and refer to some real thing out there in the world, like the ChatGPT application developed by OpenAI. "The signification 'sign,'" Derrida (1978, 281) explains, "has always been understood and determined, in its meaning, as sign-of, a signifier referring to a signified, a signifier different from its signified." *Différance* is Derrida's attempt to intervene in this logic by extending and elaborating upon the structural linguistics of Ferdinand de Saussure[1], which characterizes language and meaning-making as a matter of difference situated within language itself. "What is key here," as Peter Salmon (2021, 106) explains, "is that each signifier gets its validity not from some quality of itself; rather, it gets it from how it differs from other signifiers. Cat is cat, and not cut, because it differs in one of its phonemes. *Language is a system of differences*."

Understood in this way, signs do not at least not principally or exclusively come to have meaning by reference to things that exist outside the system of signs; signs differ from and defer to other signs in the movement of *differance*. Obviously much more could be said about this pivotal (non)concept and the famous (but notoriously difficult) essay in which it is developed and presented. For our purposes, it will be sufficient to note the following three items:

Différance as Spacing

The sign whether a word, written mark, or linguistic token does not derive meaning from any intrinsic property or direct relation to a referent. Rather, as Derrida (1982, 13) emphasizes, it acquires meaning through its differential positioning within a system of signs. This difference is necessarily spatial: it is the *spacing* the structural intervals and gaps between signs that allows them to be recognized, identified, and distinguished from one another. Without such spacing, there could be no differentiation, and without differentiation, no signification. Meaning, then, emerges not from self-presence or identity but from positional difference what something is depends on what it is *not*. Though Derrida does not use this example, it is the dictionary that

---

of this graphic difference only through a very indirect discourse on writing, and on the condition that I specify, each time, whether I am referring to difference with an "e" or *différance* with an "a." Which will not simplify things today, and will give us all, you and me, a great deal of trouble, if, at least, we wish to understand each other" (Derrida 1982, 4). Peter Salmon (2021, 78), in his well-received biograph of Derrida, adds that subsequent generations of scholars, especially in the English-speaking world, have, for better or worse, often tried to make this difference audible through a deliberate (and unfortunately often clumsy) affectation that seeks to "sound hyper French when pronouncing it."

1.What is crucial in this context is a quotation taken from Saussure's posthumously published *Course on General Linguistics* and mobilized by Derrida in a number of different texts: "Everything that has been said up to this point boils down to this: in language there are only differences. Even more important: a difference generally implies positive terms between which the difference is set up; but in language there are only differences without positive terms" (Saussure 1959, 117).

provides an easily accessible illustration. In a dictionary, words come to have meaning through their differential relationship with other words. In pursuing definitions of words in the dictionary, one remains within the space and spacing of linguistic signifiers and never gets outside the space language, to the referent or to what semioticians call "the transcendental signified."

Différance as Temporalization

Différance is not only a matter of spatial differentiation but also of temporal deferral what Derrida (1982, 8) calls *temporalization*. Meaning is never fully present in any given moment; it is always postponed, always *to come*. A sign's significance, rather than being grasped in a self-contained instant, is constituted across time through its relation to what precedes and what follows. Each sign refers not to a fixed meaning but to other signs, and since those too are deferred, the chain of signification is never complete. Meaning, then, is not something that arrives *in the present* but something that is continually delayed constituted through a future that never quite arrives and a past that is never fully recoverable. *Différance*, therefore, names both spacing the structural difference between signs and deferral the temporal postponement of presence. Meaning is never present *as such*; it is always in process.

Différance as Trace and Iterability

In the movement of *différance* that is, the co-constitutive operations of differing and deferring no sign (whether a word, mark, or linguistic unit) can be understood as autonomous or fully self-present. Each sign is necessarily inscribed within a network of relations, bearing the residual traces of other terms, contexts, and discursive formations that condition its significance (Derrida 1982, 24). Significance, therefore, can never be completely localized within a single sign; it is always and already haunted by the trace of prior iterations and the possibility of future reuse and recontextualization. Moreover, insofar as signs are *iterable* i.e. capable of being reused, remixed, and repeated in different temporal and contextual settings every iteration alters its range of possible significances, further destablizing any pretense to there being a final or secure meaning.

LLMs as Mechanisms of Différance

Large language models, such as OpenAI's GPT series of algorithms, Google's Gemini, Anthropic's Claude, and Deepseek , rely on massive datasets of text and sophisticated neural network architectures known as transformers. These models as many critics have pointed out do not understand language. Instead, they arrange legible sequences of linguistic tokens based on calculating the probable distribution of signs discoverable in their training data. Consequently, because these models depend on *differences* and *deferrals* situated in the materiality of language, they can be read as algorithmic simulations of *différance*.[1] Here's how:

1. The term *simulation* admits of a number of different and not necessarily compatible definitions. Etymologically, the verb *simulate*, from the Latin *simulare*, indicates "to copy," "to imitate," or "to feign." In computer science and related disciplines, the nominal form of the word *simulation* refers to using computer software to mimic the behavior of a real-world system or process. It involves creating a computer model that represents a system and then running the model to observe its behavior under different conditions. There is also a specific denotation of

Tokenization and Word Embeddings

LLMs generate word sequences. But technically speaking they do not work with words nor do they understand what words are or are not. They work with *tokens*, which consist of words (the, cat, in, hat), fragments of words (un believ able) or even individual characters and punctuation marks. Transforming words into tokens is called tokenization, and the point of tokenization is to converts textual data words, sentences, entire paragraphs, etc. into a sequence of tokens the model can utilize and process. And each LLM, as Jerry Kaplan (2025, 52) explains "uses its own scheme for converting words into tokens" though most "appear to favor using subword tokenization, because this offers a mix of efficiency and flexibility."

Once tokenized each token is represented by an array of numbers that describes a vector in an imaginary high-dimensional space called an embedding. In one model (the English Wikipedia model), "dog" can be represented by 300 individual numbers: "-0.03301828354597092, 0.05134638026356697, 0.0036009703762829304, -0.04066073149442673, 0.103614 30048942566. . ." (Fares et al. 2017). Thus, embeddings represent tokens as vectors in high-dimensional space, where the meaning of each token is determined not by any intrinsic property of the token, but by its proximity to and distance from other tokens. The significance of a linguistic token such as *dog*, for example, is determined by how its vector (again an ordered array of numbers) differs from and relates to the vectors of other linguistic tokens, like *puppy*, *pet*, *cat*, and so on.

These embeddings are learned during pre-training and are designed to capture semantic and syntactical relationships in terms of spatial difference. Thus, word embeddings do not just resemble *différance* they operationalize it in computational form. Meaning in these systems is never a matter of presence anchored in a transcendental signified; it is a matter of differences within a spatial field of embedded relations. LLMs therefore do not represent meaning as fixed content. They do not adhere to or play by the rules of the classically determined structure of the sign. They participate in the deconstruction of this metaphysical tradition by actualizing *différance* at scale.

Next Token Prediction

LLMs generate seemingly significant sequences of text through a computational process called next token prediction. When we prompt ChatGPT with a statement like "explain next token prediction," the model attempts to predict, i.e. make a guess about, the next most likely word that would follow from this pattern of words. It is this predictive mechanism that explains and

---

the word that is developed in poststructuralist theory, especially (but not exclusively) Jean Baudrillard. "Simulation," Baudrillard (1983, 1) wrote in his now famous eponymous essay, "is no longer the that of a territory, a referential being or substance. It is the generation by models of a real without origin or reality." This version of *simulation* is not simply the opposite of what is operationalized in computer science. It is its deconstruction. In the context of this sentence, *simulation* is used in the computer science sense of the word. Though that use is accompanied or, perhaps better stated, *haunted* by both the word's etymology and its deconstruction in Baudrillard's work.

underlies all of the model's seemingly intelligent behaviors, from composing short essays, answering questions, engaging in dialogue, and even generating code.

This process unfolds as follows: The input words are tokenized and mapped to these high-dimensional vectors that (as described above) represents their relative positions with respect to other tokens in this imaginary high-dimensional space of differences. These vectors are then processed through a stack of transformer layers a kind of neural network comprised of both a next token prediction network and attention network where the output of one layer becomes the input for the next. Current LLMs consist of many layers of transformers. OpenAI's GPT-3, for instance, has 96 layers, and its successor GPT-4 boasts 1.8 trillion parameters (i.e. adjustable weights in the model's neural network) across 120 layers. It is this density that makes these language models "large."

At each step in this process, the model produces a probability distribution over possible continuations of a given sequence and selects the most likely next token. Take for instance the sequence "To be or not to be that is the." There are a number possible words that could come next in this sequence: "question" "dilemma" "problem" "issue" and so on. The model selects one of these words often the most likely one (i.e. the one that has the highest probability of coming next as determined by the relative proximities and differences in the embedded representation) but sometimes one that is selected with an element of randomness and adds it to the sequence. This new, longer sequence is then fed back into the model, and the process repeats. It does this again and again through many iterations.

Thus next word prediction is a sequential operation that dynamically unfolds in time. The model generates one token after another, without having access to some fully formed and complete utterance in advance. As such, the significance or semantic content of any given token is always in process and remains provisional, awaiting the specification or transformation that will be brought about by the generation of subsequent tokens. The model, therefore, does not have access to some completely formed meaning and then expresses this content in a sequence of linguistic tokens. Rather, the meaning of the generated content emerges through an iterative process in which meaning is continually deferred in its dependence on future tokens that have yet to be generated.

This procedure not only looks and sounds like *différance* but actualizes it in computational form. In the movement of *différance*, as Derrida argues, the sign never arrives at a final, self-present meaning; instead, each signifier is dependent upon another in an endless chain of difference and deferral, where significance is dynamic and always *to come*. In the case of LLM AI, each token is selected not as some definitive semantic content but as a contingent placeholder a statistical best guess which acquires significance only in differential relation to deferred future tokens that are themselves caught in this movement of *différance*.

Nothing Outside the Text

When one of these transformer-based LLMs generates a word sequence, like "LLMs are a kind of generative AI," it does not know (nor does it not know) what it is saying, because it does

not have access to nor does it understand what these sequences of linguistic tokens refer to. This important difference has been illustrated in what is one of the defining thought experiments in the philosophy of AI John Searle's Chinese Room (1999, 115). Like the man inside Searle's imaginary room, LLMs do not know what they are doing. They do not understand language in the way that we allegedly understand and use language. They are simply and superficially playing with different signs. In the case of the Chinese Room this is done by following a predefined set of transformations analogous to how GOFAI symbolic reasoning systems operate. In the case of an LLM, by contrast, this is accomplished by measuring and manipulating the spatial differences represented by and encoded in word embeddings. Thus, for the LLM, there is no transcendental signified that can arrest and anchor the chain of signification. There is nothing beyond or outside the different relations between tokens that are endlessly deferred in the process of next token prediction.

Thus, LLMs actualize what is perhaps the most famous (or notorious) statement that has come to be associated with Derrida (1976, 158): *Il n'y a pas de hors-texte* "there is nothing outside the text" or "there is no outside- the- text." This is not some anti-realist statement, and it does not mean as many critics have mistakenly assumed that nothing is real or objectively true and everything is just a socially constructed artifact or effect of discourse. And Derrida had explained as much in the course of a debate with John Searle that has been recorded for us in the book *Limited, Inc.*: "'There is nothing outside the text.' That does not mean that all referents are suspended, denied, or enclosed in a book, as people have claimed, or have been naïve enough to believe and to have accused me of believing. But it does mean that every referent, all reality has the structure of a differential trace, and that one cannot refer to this 'real' except in an interpretive experience" (Derrida 1993, 148).

What this means is that a text whether it is written by a human writer or artificially generated by an LLM like ChatGPT (with the nudge of a human prompt) comes to have meaning not by referring and deferring to some external signified (what Aristotle in *De Interpretatione* calls thoughts or the things to which thoughts ultimately refer). It comes to enact and perform meaning by way of interdependent relationships to other words with which it is already associated and differentiated from. It is for this reason that we can say, following Ludwig Wittgenstein (1995, 5.6) that for LLMs the limits of their language (model) mean the limits of their world. Consequently, what has often been offered as a criticism of LLM technology namely, that these algorithms only circulate different signs without access to the signified might not be the indictment critics think it is. LLMs are *différance engines* that deconstruct the defining logic of classical semiotics.

Philosophical Consequences

Différance provide a conceptual mechanism and vocabulary by which to make sense of and explain the philosophical exigencies and consequences of LLMs. But simply connecting the dots between the operational features of LLM AI and Derrida's *différance* is not, not in and of itself, sufficient to close the deal. We still need to ask, what difference this make in our understanding

and critique of LLMs and, perhaps more importantly, our concept and understanding of language. Here we can identify at least three important philosophical consequences:

Deconstruction of Logocentrism

Central to Jacques Derrida's entire philosophical project of *deconstruction* (Gunkel 2021) is the critique of logocentrism. The term *logocentrism* was coined in the early part of the 20th century by German philosopher Ludwig Klages (Josephson-Storm 2017, 221), who used it to identify the assumed primacy and importance of the spoken word as a direct sign of things and thus relegating writing to being a sign of speech or the sign of a sign. Derrida (1976, 11) locates the original formulation of this concept in Aristotle's *De Interpretatione*: "If for Aristotle spoken words (*ta en te phone*) are the symbols of mental experience (*pathemata tes psyches*) and written words are the symbols of spoken words, it is because the voice, producer of the first symbols, has a relationship of essential and immediate proximity with the mind. Producer of the first signifier, it is not just a simple signifier among others. It signifies 'mental experiences' which themselves reflect or mirror things by natural resemblance." If we ask the question "What is it Derrida is trying to say, here?" that very question a mode of inquiry which seeks to discover what an author is *saying* in and by the writing is logocentrism par excellence.

Derrida does not just challenge this way of thinking about words and things. He advocates deconstruction of the ruling conceptual opposition that is its organizing principle the binary distinction that differentiates the full presence of speech from its derivative, deceptive, and deficient other: writing. In *Of Grammatology*, Derrida (1976) exposes how Western philosophy has historically valorized spoken language as the immediate, authentic carrier of thought, while relegating writing to a secondary, derivative status. Logocentrism thus rests on a metaphysics of presence: the belief that meaning is grounded in a fully present, self-conscious subject who speaks and therefore has something to say in and by their words.

LLMs participate in and operationalize the deconstruction of logocentric metaphysics. They are trained almost exclusively on written text[1] and yet produce outputs that look like speech, thought, or narrative coherence without ever invoking or being anchored in the intensions of a speaking subject. They therefore foreground the materiality of the signifier the traces of language as it is written and circulated detaching it from any living, breathing voice or origin. In this way then, they overturn and disrupt which, as Derrida (1981a, 41) explains, is the constitutive "double gesture of deconstruction" the logocentric privilege. With the LLM, writing is no longer a secondary after effect but is foundational. In order to identify this subtitle difference, Derrida (1976) renames this primary sense of writing "arche-writing." And the effect of this intervention challenges formative assumptions across the entire history of Western philosophy.

Consequently, LLMs do not just mimic human language use; they expose the very logics that underlie its functioning. They reveal that language operates through a systemic play of differences not self-presence, where significance is not derived from intention or origin, but emerges in the

1. Even when these models are trained on other kinds of media content as in contemporary multimodal systems this content is a kind of writing insofar as it is recorded and preserved in the sign system of digital data.

movement of *différence*. In short, LLMs are mechanisms in which the metaphysics of presence unravel and the operations of deconstruction are made computable.

Death of the Author

Derrida's deconstruction of logocentrism destabilizes the assumption that meaning is anchored in and guaranteed by a singular, originary intention. Rather than presuming an authoritative and authorial source from which truth is transmitted, *différance* names the process by which meaning is endlessly deferred and constituted through relational difference. In parallel, Roland Barthes's (1978, 148) declaration of "the death of the author" rejects the idea that the author is the ultimate arbiter of meaning, insisting instead that the text is a site of multiplicity, "a tissue of quotations drawn from the innumerable centers of culture." LLMs operationalize these theoretical insights in a radically literal form. Trained on vast archives of textual traces, these models generate responses through algorithmic pattern recognition, without reference to a sovereign subject or authorial intention. The resulting text is not authored in any traditional sense but assembled through iterative processes that track and operationalize the movement of *différance*.

When a user prompts an LLM and receives a coherent output, the question of authorship becomes radically undecidable. Who, or what, speaks? The LLM is not a speaking subject, yet it produces legible linguistic content in the voice of many; it recombines and remixes fragments without origin; and proliferates meanings without closure. Thus, the texts generated by LLM AI are "quite literally *unauthorized"* (Coeckelbergh and Gunkel 2025, 7). Once the written text is cut-loose from the controlling interests and intentions of an author, the question concerning significance gets turned around. Specifically, the meaning of a piece of writing is not something that can be guaranteed *a priori* by the authentic character or *ethos* of the one who is assumed to be speaking through the medium of the text. Instead, meaning transpires in and from the experience of reading. And if it is the case that this significance had been customarily attributed to the original intensions of an author, that attribution is (and has always actually and only been) projected backwards from the reader onto a supposed and oftentimes absent author. In effect, an effect of reading gets turned around to become its own cause.

This flipping of the script on literary theory alters the location of meaning-making from the original intentions of the author/writer who (it is assumed) has "something to say" to the interpretive activity of the reader who makes meaning in or generates it out of the materiality of the written content. As Barthes (1978, 148) explains: "text is made of multiple writings, drawn from many cultures and entering into mutual relations of dialogue, parody, contestation, but there is one place where this multiplicity is focused and that place is the reader…A text's unity lies not in its origin but in its destination."

This also explains how LLM generated content comes come to have meaning. The critics are correct when they point out, for instance, that LLMs manipulate words or linguistic tokens but do not "truly comprehend the meaning behind the words" (Bogost 2022) because they "have no access to real-world, embodied referents" (Bender quoted in Weil 2023). But it would be impetuous to conclude from this fact that what an LLM generates is pure non-sense, meaningless,

or bullshit (Hicks et al. 2024). These writings are and can be meaningful, and what they mean is something that happens in the process of our reading, interpretation, and evaluation of the generated content. And this fact is not something that is specific to LLMs but is, as Barthes had already proposed and demonstrated, a defining characteristic of all writing. LLMs just happen to make it legible.

Critical AI

LLMs and other forms of generative AI are undeniably powerful technologies, and it is essential to approach them critically attuned not only to their potential benefits but also to the conceptual frameworks they both rely on and disturb. Many current responses from linguists, philosophers, and AI experts tend to reaffirm logocentric assumptions about meaning, authorship, and intelligence assumptions that were already challenged by twentieth-century developments in literary theory and continental philosophy. And the problem is not that these ways of thinking have somehow failed to work in the face of these new technologies of writing. It's quite the opposite. The problem is they work all too well, exerting their influence over our thinking about writing and writing about thinking in ways that go by largely without notice.

Ultimately, the root cause of the problem might already be contained in the way *artificial intelligence* both the term and the concept that it identifies has been formulated and defined. Because of its nominal focus on "intelligence," the output of these mechanisms is taken to be either external signs of the actual presence of intelligent thought or, in those situations where the device seems to spit out nonsense or hallucinate, the lack thereof. This procedure taking the generation of written content (ostensibly external signifiers) as a sign or symptom of intelligence (an internal cognitive capability) has been the defining condition of/for machine intelligence since the time of Alan Turing's (1950) agenda setting paper on the subject. Though Turing does not identify it as such, his imitation game plays by the rules of this metaphysical game. And the current criticisms of LLM AI namely, that they know nothing of what they speak or that they are stochastic parrots (Bender et al. 2021) simply repeats, or "parrots," this formulation.

LLMs are philosophically significant precisely because they displace these assumptions. They generate texts without authors, statements without intention, writing without speech. They mark a break in the chain that links linguistic expression to the voice of a knowing subject. From within the prevailing logocentric tradition, this appears as a profound crisis. If language no longer guarantees truth through the voice of an authorial presence, what becomes of meaning, literature, or thought itself?

But from another perspective one shaped by poststructuralist critique this is not a crisis but a philosophical opportunity. Rather than mourning the disruption of presence, what we see in LLMs is an actualization of what deconstruction has long argued. Namely, that meaning is never self-present, never guaranteed by intention, and never fully anchored in an origin. LLMs do not undermine language, authorship, or truth as such; they challenge a historically specific conception of these terms one grounded in metaphysical assumptions that deserve interrogation.

To understand LLMs as *différance engines* is to see them not as defective thinkers or broken communicators, but as machines that expose the already-deferred, already-differential logic of signification. They do not confirm the absence of intelligence so much as they call into question what we have taken intelligence to mean. In doing so, they offer us the opportunity to reexamine the very foundations of linguistic meaning, authorship, and truth.

Criticisms and Responses

We began by proposing that LLMs are *différance* engines and the foregoing has not only explained how and why that is the case but also identified the important philosophical consequences of this shift in perspective. That said, there are potential criticisms that one could venture in response. We will conclude, therefore, by considering three of these, each one formulated in terms of a critical question:

Isn’t This Just Word games?

A common objection to theoretical essays of this kind particularly those that bring Derridean concepts into dialogue with contemporary technology is that what they amount to seems to be little more than conceptual substitutions or an elaborate kind of wordplay. The concern is that by reinterpreting LLM AI through the lens of *différance*, we are not getting any new or genuine insights, but merely replacing one set of terms (e.g. token, embeddings, statistical models) with another (e.g. signifier, trace, deferral). From this vantage point, the essay is not really philosophical but rhetorical. It may be a sophisticated and even persuasive performance, but it is not a true contribution to understanding.

This mode of criticism, however, can be shown to enact, and thus be deconstructed by, the logic of *différance*. To accuse the essay of "just playing with words" is to presuppose that there is a more stable, less playful, and somehow more *real* discourse elsewhere a pure and immediate language that transparently says what it means and means what it says. Yet Derrida’s central claim is precisely that *all* discourse, including the language of critique, is already entangled in *différance*. There is no final metalanguage, no pure realm of concepts or Platonic ideas that are untouched by the trace of *differance*.

Thus, what the critic calls "just words" is already a metaphysical commitment to the logocentric idea that meaning can or should be grounded outside of language, in some pre-given presence or empirical certainty. But LLMs offer a kind of technological demonstration and proof that such meaning is never so fixed and assured, due to the fact that their operations depend on statistical difference and deferral rather than by reference to a stable referent. Consequently, the model’s output is in fact "just words" and precisely in the way Derrida identifies *not* because it is empty, but because it foregrounds the very instability that haunts all linguistic systems, including those used in the process of formulating this criticism.

In other words, this objection assumes a metaphysical center it cannot secure, and thus falls into the very word play it seeks to exclude. The accusation of "word games" becomes a performative instance of the trace, as it relies on the distinction between “serious” and “playful”

language, a binary Derrida not only demonstrates is untenable but which is itself subject to deconstruction. So rather than dismiss the essay's approach as mere linguistic substitutions, we might see it as exposing the deeper logic already operative in both natural and artificial languages. To think LLMs through *différance* is not to indulge in abstraction for its own sake, but to illuminate how these systems themselves simulate the structural conditions of meaning that Derrida identifies.

If this is so, then couldn't an LLM have written the entire essay?

The second criticism has a performative dimension to it. In order to be internally consistent and avoid the charge of performative contradiction, everything that has been written here about the author, writing, and meaning-making would need to apply to and hold for this text. It is therefore reasonable for the skeptical reader to inquire whether the argument they have just read is the product of a human author, output that has been generated by an LLM, or the result of some form of human machine collaboration? To put it in its most direct and accusative form: Couldn't an LLM have written the entire essay?

This objection does not refute the essay's thesis; it confirms it in its most radical form. The very possibility that an LLM could produce this text does not disprove *différance*. It is an enactment of it. If language functions not through authorial presence but emerges through differential repetition, if meaning arises from spacing, deferral, and iterability, then the source of a given utterance whether human written or machine generated is no longer foundational and determinative. In fact, it is this undecidability of origin an undecidability that is asked about by way of this particular criticism that *différance* seeks to name.

Additionally, this criticism is pre-emptively answered by Derrida in the pages of *Différance*: "This implies," Derrida (1982, 15) writes, "that the subject (in its identity with itself, or eventually in its consciousness of its identity with itself, its self-consciousness) is inscribed in language, is a 'function' of language, becomes a speaking subject only by making speech conform to the system of the rules of language as a system of differences, or at very least by conforming to the general law of *différance*." In other words, what LLMs seem to lack i.e. a conscious speaking subject who uses language to say something intelligible about the world is already missing from human language use insofar as the very concept of a "speaking subject" is an effect of *différance*. Consequently, even if we offer the customary tokens of authenticity the proper name of the author, a written declaration that this is "100 percent genuine human-generated content," a watermark, or some other official sign they are all, always and already, inescapably involved in the movement of difference and deferral that is *différance*.

Thus, questioning whether an LLM could have written the essay as an objection to what is argued in the essay is to fall back on and reinscribe the logocentric metaphysics of presence. But this is precisely what deconstruction calls into question and, in a word, deconstructs. If the model can produce writing that participates in the logic of *différance*, the fact that it was machine generated (or not) does not make a difference, or at least not the difference one might think it

does. In other words, the suspicion that a machine could have written the essay is not the point at which the argument breakdowns. It is the point at which it becomes real.

What would Derrida say about this use (or misuse) of his work?

Obviously, Derrida did not write about contemporary artificial intelligence, large language models, or the computational infrastructure of contemporary machine learning systems. To apply his thinking to these domains may seem, to some, to be an overextension perhaps even a distortion. Thus, doesn't such an application risk instrumentalizing or misappropriating his work for purposes it was never meant to serve?

This objection stages a familiar anxiety: that of fidelity to the original meaning intended by an author. And yet, Derrida's work persistently undermines the very idea of hermeneutic closure or interpretive orthodoxy guarded by authorial intent. From *Of Grammatology* (1976) to *Limited Inc*. (1993), Derrida emphasized that texts, once written, exceed their origin. They become iterable, open to contexts unforeseeable by their progenitor. There is no meaning that is not always and already contextual, and no context that is ever fully saturated such that it can ever fully delimit significance. And in pointing this out, Derrida is channeling an ancient problem, one that was already identified by Socrates in Plato's *Phaedrus* (1982, 275e) "And every word, when once it is written, is bandied about, alike among those who understand and those have no interest in it, and it knows not to whom to speak or not speak."[1]

So to ask "What would Derrida say?" is, in a sense, to re-inscribe the very metaphysics of presence that the deconstruction of logocentrism seeks to overturn and displace. We cannot consult the (dead) author to confirm our usage; nor should we pretend that any text including Derrida's—can be stabilized once and for all. To engage Derrida in the context of LLMs is not to violate or to do violence to the integrity of his work but to participate in its performative logic: to extend the play of *différance* into a new domain, one whose conditions of signification make that extension not only possible, but necessary.

Indeed, to refuse such a movement out of fear of misinterpretation or misrepresentation would be to betray the very force of Derrida's innovations, which calls us to interrogate boundaries between philosophy and technology, human beings and machines, speech and writing. The risk of misreading is not an accident of deconstruction; it is its very condition. As Derrida might put it: there is no responsible reading that does not risk this irresponsibility.

Limitations and Future Direction

The objective of this essay was to initiate an interdisciplinary dialogue between recent innovations in generative AI and late-20th-century poststructuralist theory. Such an approach has the potential to upend existing ways of thinking, generate new insight, and reveal novel opportunities. However, as this is only a first step, several issues identified in the course of the

1. Derrida directly addresses this scene of writing and the written dialogue in which it appears in the essay "Plato's Pharmacy" (Derrida 1981b).

investigation could not be fully addressed and will therefore require further attention. In this final section, I highlight three in particular:[1]

Structuralism vs. Poststructuralism

One limitation of the present essay is that it cannot fully address the complex and ongoing debate in structuralist and poststructuralist semiotics over Derrida's "revisionist reading" of Saussure, especially the *Course in General Linguistics* (Harris 2003). This issue, which already has quite a storied history, has recently been revisited by Leif Weatherby in *Language Machines*, which argues that Saussure's structuralism offers a more robust theoretical framework for explaining and interpreting contemporary language generators. As Weatherby (2025, 73) writes, "it is simply not clear that we need Derrida's revision of structuralism to proceed with a concrete analysis of computational language," and his book aims "to make the case that the implementation of contemporary language generators matches the theory of language that European structuralism advanced nearly a century ago." No Derrida, *différance*, or deconstruction required.

Addressing and doing justice to this debate would require more sustained engagement with both Saussurean and Derridean semiology than is possible here. The terms of that debate, however, are now clearly drawn. Weatherby (2025, 43) sides with structuralism, concluding that "poststructuralism, which often has useful diagnostic things to say about contemporary culture, has not produced a concrete enough analytical tool set to deal with automated language generation." This essay offers the counterargument, demonstrating how Derrida's remix of Saussure provides a more substantive critical lens for understanding the technical operations of the large language model. That debate cannot be resolved here. This is where it begins.

The Stakes of Deconstruction

A second potential limitation concerns the philosophical consequences of the investigation. Because this essay has focused on situating the (non)concept of *différance* as a framework for understanding the technical operations of LLMs, one might come away with an impression of Derrida's work that could potentially obscure what is truly at stake in his philosophical enterprise. In other words, the effort to demonstrate the relevance of poststructuralist theory for AI systems, might inadvertently obscure the deeper philosophical consequences of deconstruction. It risks, as the saying goes, missing the philosophical forest for the trees.

Though Derrida did not have the occasion to address LLMs directly, his engagement with the "animal question" (Derrida 2008) provides a useful parallel. In deconstructing the human/animal binary, Derrida's aim is not to extend some exclusive human privilege to animals but to interrogate whether humans ever legitimately possessed such a privilege in the first place. Applied to LLMs, this reframes the debate entirely. What is ultimately at stake, then, is not merely tracking

1. These three items come directly from the anonymous reviews of the initial submission. I note this for two reasons: 1) To express my gratitude to the reviewers for their insight and for the work they put into reading and responding to the initial draft. And 2) to recognize the continued importance and viability of the peer review process in academic publishing.

how machines challenge human capacities but recognizing how these applications reveal the fact that these capacities language, creativity, and authorship have never been as autonomous, self-present, or secure as we might have assumed or assured ourselves.

Future research could build on and extend this insight this by engaging more directly with Derrida's writings on "originary technicity" (Bradley 2011) and his deconstruction of entrenched metaphysical binaries human/machine, natural/artificial, speech/writing, etc. This would not only offer a more nuanced account of the philosophical consequences of aligning deconstruction with LLM technologies but would also sharpen our understanding of how these systems expose the structural conditions of language that deconstruction has long theorized and, in a word, deconstructed.

LLMs and the Question of Meaning

A third limitation related to and directly proceeding from the previous one concerns the consequences of a central claim advanced in this essay: that LLMs do not merely imitate human language use, but expose the very logics that underlie its functioning. Taken seriously, this could suggest that human language use itself might be explicable, at least in part, in terms of the operations of these generative models.

This opens an important line of inquiry: What, if anything, do LLMs reveal about the structures of human thought and language use? And how might this affect our understanding not only of meaning but of ourselves? One possible response, consistent with the strategy of deconstruction, would be to resist the binary opposition that has been situated between the "we are all just stochastic parrots now" and "we are not parrots, because we know what words mean." Rather than choosing sides in this dispute, Derrida's work can be read as charting a third path one that deconstructs the metaphysical terms of the debate itself.

While the present essay gestures toward this possibility, a full elaboration of it exceeds the scope of the current inquiry. To pursue it seriously would require a separate and sustained engagement not only with Derrida's corpus but also with contemporary debates in philosophy of language, cognitive science, and computational linguistics. In short, this essay simply set out to connect-the-dots between *différance* and LLMs, but what this ultimately means for human thought and language use is a different, albeit related, undertaking, and one that must, for now at least, be deferred.

Conclusions

If *différance* is, as Derrida (1982, 3) insists, "literally neither a word nor a concept," then its resonance with large language models is not a metaphor but a structuring principle. What LLMs actualize through their predictive architectures, their dependence on prior textual contexts, and their inability to access and refer to "real world embodied referents" (Bender quoted in Weil 2023) is not a departure from meaning but an actualization of its constitutive conditions. These models do not "understand" in the classical philosophical sense; they operate through difference

and deferral, generating linguistic outputs that are intelligible precisely because they are situated and operate within a system of difference.

What this essay has called "Différance Engines" do not merely imitate human language use; they demonstrate the difference and deferral that is signification. This is not to suggest that LLMs are or have even been designed to be deconstructive agents, nor that Derrida foresaw this technological development, but rather that the appearance of such systems reveals anew the play of signification that deconstruction theorized decades ago. If anything, the emergence of LLMs makes the stakes of *différance* legible, rending visible, and perhaps uncomfortably tangible, the unsettling decentering of meaning that deconstruction has always articulated.

To recognize this is not to reduce philosophy to technology nor to subordinate poststructuralist theory to computation. Rather, it is to allow theory to follow the movement of *différance* itself beyond its original context, across disciplinary boundaries, and into other kinds of texts and contexts. In doing so, we are not misusing Derrida but responding to the iterable logic he situated in language the fact that meaning is never fully present, always already deferred, and that no discourse human or machinic is absolved of the play of differences.

Author Contributions

The author contributed equally to the conceptualization of the article and writing of the original and subsequent drafts.

Data Availability Statement

Data available on request from the authors.

Acknowledgements

Thanks and appreciation have not been expressed.

Ethical Considerations

The author strictly adhered to the highest standards of research integrity. The authors avoided data fabrication, falsification, plagiarism, and any other form of scientific misconduct.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The author declare no conflict of interest.

References

Aristotle. (1938). *Categories. On Interpretation. Prior Analytics*. Translated by H. P. Cooke. Cambridge, MA: Harvard University Press.

Barthes, Roland. (1978). Death of the Author. In *Image, Music, Text*. Translated by Stephen Heath, 142–148. New York: Hill & Wang.

Baudrillard, Jean. (1983). *Simulations*. Translated by Paul Foss, Paul Patton, and Philip Beitchman. New York: Semiotext(e).

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 Conference on Fairness, Accountability, and Transparency*, March 3–10, 2021 (FAccT '21), Virtual Event, Canada, ACM. https://doi .org/10.1145/3442188.3445922.

Bogost, Ian. (2022). ChatGPT Is Dumber Than You Think. *The Atlantic.* https://www.theatlantic.com/technology/archive/2022/12/ chatgpt -openai-artificial-intelligence-writing-ethics/672386.

Bradley, Arthur. (2011). *Originary Technicity: The Theory of Technology from Marx to Derrida.* New York: Palgrave Macmillan

Coeckelbergh, Mark and David J. Gunkel. (2025). *Communicative AI: A Critical Introduction to Large Language Models.* Cambridge: Polity.

Derrida, Jacques. (1976). *Of Grammatology*. Translated by Gayatri Chakravorty Spivak. Baltimore,MD: The Johns Hopkins University Press.

Derrida, Jacques. (1978). *Writing and Difference*. Translated by Alan Bass. Chicago: University of Chicago Press.

Derrida, Jacques. (1981a). *Positions*. Translated by Alan Bass. Chicago: University of Chicago Press, 1981.

Derrida, Jacques. (1981b). *Dissemination*. Translated by Barbara Johnson. Chicago: University of Chicago Press.

Derrida, Jacques. (1982). *Margins of Philosophy*. Translated by Alan Bass. Chicago: University of Chicago Press.

Derrida, Jacques. (1993). *Limited Inc.* Evanston, IL: Northwestern University Press.

Derrida, Jacques. (2008). *The Animal That Therefore I Am*. Translated by David Wills. Edited by

Marie-Louise Mallet. New York: Fordham University Press.

Fares, Murhaf, Andrei Kutuzov, Stephan Oepen, and Erik Velldal. (2017). Word vectors, reuse, and replicability: Towards a community repository of large-text resources. In Jörg Tiedemann (ed.), *Proceedings of the 21st Nordic Conference on Computational Linguistics* (NoDaLiDa), 22–24 May 2017. Linköping University Electronic Press. https://doi.org/978-91-7685-601-7. (See also http://vectors.nlpl.eu/explore/embeddings/en.)

Gunkel, David J. (2021). *Deconstruction*. Cambridge, MA: MIT Press.

Harris, Roy. (2003). *Saussure and His Interpreters*. Edinburgh: University of Edinburgh Press.

Hicks, Michael Townsen, James Humphries, and Joe Slater. (2024). ChatGPT Is Bullshit. *Ethics and Information Technology*. https://doi .org/10.1007/s10676-024-09775-5.

Josephson-Storm, Jason A. (2017). *The Myth of Disenchantment: Magic, Modernity, and the* Birth of the Human Sciences. Chicago, IL: University of Chicago Press.

Kaplan, Jerry. (2025). *Generative Artificial Intelligence: What Everyone Needs to Know*. New York: Oxford University Press.

Plato. (1982). *Euthyphro, Apology, Crito, Phaedo, Phaedrus*. Translated by Harold North Fowler. Cambridge, MA: Harvard University Press.

Salmon, Peter. (2021). *An Event, Perhaps: A Biography of Jacques Derrida*. London: Verso.

Saussure, Ferdinand de. (1959). *Course in General Linguistics*. Translated by Wade Baskin. London: Peter Owen.

Searle, John. (1999). The Chinese Room. In R. A. Wilson and F. Keil (eds.), *The MIT Encyclopedia of the Cognitive Sciences*, 115–116. Cambridge, MA: MIT Press.

Turing, Alan. (1950). Computing Machinery and Intelligence. *Mind* 59 (236): 433–460. https://doi.org/10.1093/mind/LIX.236.433.

Weatherby, Leif. (2025). *Language Machines: Cultural AI and the End of Remainder Humanism*. Minneapolis, MN: University of Minnesota Press.

Weil, Elizabeth. (2023). You Are Not a Parrot: And a Chatbot Is Not a Human: And a Linguist Named Emily M. Bender Is Very Worried What Will Happen When We Forget This. *New York Magazine*, March 1. https://nymag.com/intelligencer/article/ai-artificial-intelli gence-chatbots-emily-m-bender.html.

Wittgenstein, Ludwig. (1995). *Tractatus Logico-Philosophicus*. Translated by D. F. Pears and B. F. McGuinness. New York: Routledge.

پژوهش های نوین در مطالعات علوم انسانی اسلامی
شاپا الکترونیکی: ۲۹۸۰-۸۹۴۴
https://api.lu.ac.ir/
دانشگاه لرستان

موتور دیفرانس

مدل‌های زبانی بزرگ و پساساختارگرایی

دیوید جی گانکل

استاد، گروه فلسفه و اخلاق، دانشکده فناوری اطلاعات و ارتباطات، دانشگاه ایلینوی شمالی، ایالات متحده آمریکا. رایانامه: dgunkel@niu.edu

اطلاعات مقاله

نوع مقاله:
مقاله پژوهشی،

تاریخچه مقاله:
تاریخ دریافت: ۱۴۰۴/۰۷/۰۶
تاریخ بازنگری: ۱۴۰۴/۰۹/۰۷
تاریخ پذیرش: ۱۴۰۴/۱۰/۱۲
تاریخ انتشار: ۱۴۰۴/۱۰/۲۲

کلیدواژه‌ها:
هوش مصنوعی،
ژاک دریدا،
دیفرانس،
مدل‌های زبان بزرگ،
فلسفه فناوری،
پساساختارگرایی،
نشانه‌شناسی

چکیده

هدف: این مقاله استدلال می‌کند که مدل‌های زبانی بزرگ، مانند معماری‌های ترنسفورمر از نوع GPT، مفهوم «دیفرانس» ژاک دریدا را به شکل عملی تحقق می‌بخشند.

روش پژوهش: با تکیه بر چارچوب نظریه پساساختارگرایی و نشانه‌شناسی، مقاله بررسی می‌کند که مدل‌های زبانی بزرگ چگونه با محاسبه تفاوت‌های آماری در میان مجموعه‌های متنی عظیم، محتوای معنادار تولید کرده و بدین ترتیب فرآیندهای فاصله‌گذاری، زمان‌مندی و اثر را در کانون توجه قرار می‌دهند.

یافته‌ها: مقاله نشان می‌دهد که مدل‌های زبانی بزرگ را می‌توان به عنوان «موتورهای دیفرانس» در نظر گرفت که سازوکارهای نظریه‌پردازی‌شده توسط دریدا را به صورت محاسباتی به اجرا درمی‌آورند. همچنین مقاله به پیامدهای فلسفی این هم‌ترازی از جمله چالش‌های پیش روی نشانه‌محوری، مؤلف‌محوری و متافیزیک حضور می‌پردازد و به سه نقد احتمالی پاسخ می‌گوید که استفاده از آثار دریدا در این زمینه ادامه و بازبیان منطق آن است.

نتیجه‌گیری: مقاله با شناسایی سه محدودیت سیستمی و ترسیم فرصت‌هایی برای پژوهش‌های آتی در این حوزه نتیجه‌گیری می‌کند و نشان می‌دهد که از یک سو چگونه می‌توان مدل‌های زبانی بزرگ را از منظر نظریه پساساختارگرایی خوانش کرد، و از سوی دیگر چگونه نظریه پساساختارگرایی می‌تواند از طریق عملیات فنی هوش مصنوعی معاصر تبیین و قابل دسترس شود.

استناد: جی گانکل، دیوید. (۱۴۰۴). موتور دیفرانس (مدل‌های زبانی بزرگ و پساساختارگرایی). *پژوهش های نوین در مطالعات علوم انسانی اسلامی*، (ویژه نامه) ۴، ۱-۳۴. https://doi.org/10.22034/api.2026.735002

ناشر: دانشگاه لرستان.

DOI: https://doi.org/10.22034/api.2026.735002

مقدمه

واژه جدید «différance» ژاک دریدا (Derrida ۱۹۸۲) یکی از تأثیرگذارترین و در عین حال به طرز بدنامی مبهم‌ترین مفاهیم در نظریه انتقادی معاصر است. این واژه از فعل فرانسوی différer گرفته شده است. که بسته به زمینه‌ی کاربرد، می‌تواند به معنای «متفاوت بودن» یا «به تعویق انداختن» باشد. *تفاوت*این اصطلاح، شیوه‌ای را مشخص می‌کند که کلمات و سایر انواع نشانه‌ها یا دال‌ها، نه با ارجاع به یک مرجع پایدار، بلکه از طریق تفاوت و ارجاع آن به سایر نشانه‌ها در همان نظام، معنا کسب می‌کنند.

اگرچه مفهوم «دیفرانس» در اصل در بستر نشانه‌شناسی، پساساختارگرایی و سنت پدیدارشناسی معرفی و توسعه یافت، اما اهمیت آن بسیار فراتر از محدوده نظریه ادبی و فلسفه قاره‌ای است. یکی از حوزه‌های غیرمنتظره‌ای که این مفهوم در آن قرار می‌گیرد که دریدا (۱۹۸۲، ۳) اصرار دارد که «به معنای واقعی کلمه نه یک کلمه است و نه یک مفهوم» کاربرد غیرعادی آن در مدل‌های زبانی بزرگ (LLM) مانند مدل‌های نوع GPT و سایر معماری‌های مبتنی بر ترانسفورماتور است. این سیستم‌ها، در سطح فنی، همان مکانیسم‌های تفاوت و تعویقی را که دریدا توصیف و نظریه‌پردازی می‌کند، محقق می‌کنند. یعنی، آنها محتوای زبانی ظاهراً معناداری را نه با ارجاع یا دسترسی به ارجاعات دنیای واقعی، بلکه از طریق محاسبه تفاوت‌های آماری که در مجموعه‌های بزرگی از داده‌های متنی قابل کشف هستند، تولید می‌کنند.

این مقاله بررسی می‌کند که چگونه هوش مصنوعی (AI) در سطح LLM می‌تواند به عنوان مکانیسم‌های تفرق یا آنچه عنوان این مقاله آن را می‌نامد، در بازآفرینی عمدی اثر چارلز بابیج، موتورهای تفرق، دیده شود. ۱) ما با شناسایی و توضیح ویژگی‌های کلیدی تفرق، همانطور که دریدا در مقاله عنوانی خود در مورد این موضوع توسعه داده است، شروع خواهیم کرد. ۲) سپس نشان خواهیم داد که چگونه LLMها، به ویژه مدل‌های مبتنی بر ترانسفورماتور، این مفاهیم را از طریق اتکای خود به تفاوت آماری و تعویق‌های زمینه‌ای عملیاتی می‌کنند. ۳) در بخش سوم، پیامدهای فلسفی درک هوش مصنوعی LLM و محتوای متنی تولید شده توسط LLM را از طریق این دیدگاه ساختارشکنانه توصیف و مورد بحث قرار می‌دهیم. ۴) پس از آن، بخش چهارم آمده است که در آن به سه انتقاد احتمالی از این تلاش می‌پردازیم و به این اعتراضات بالقوه پاسخ می‌دهیم. ۵) در نهایت، مقاله با شناسایی سه محدودیت سیستماتیک تحقیق فعلی و با پیشنهاد مسیرهایی برای تحقیقات آینده که ادعاهای مطرح شده در اینجا را گسترش، اصلاح یا شرح می‌دهند، نتیجه‌گیری می‌کند.

تفاضل

دریدا در سال ۱۹۶۸ در یک سخنرانی و مقاله با همین نام، واژه‌ی «différance» را به عنوان یک غلط املایی عمدی و حساب‌شده از اسم رایج فرانسوی «différence» ابداع کرد و «e» را با «a» جایگزین کرد. جالب اینجاست که این جایگزینی به ظاهر ساده چیزی است که فقط در نوشتار و از طریق نوشتار برای ما در دسترس است. همانطور که دریدا (۱۹۸۲، ۳) توضیح می‌دهد، «کاملاً گرافیکی است: خوانده می‌شود، یا نوشته می‌شود، اما نمی‌توان آن را شنید. نمی‌توان آن را در گفتار درک کرد.»[۱] با این حال، مهم‌تر از آن، این سکوتِ تفاوت‌آمیز، سکوتی که دریدا (۱۹۸۲، ۴) بر آن تأکید می‌کند، تنها می‌تواند در چارچوب الفبایی

---

۱. همانطور که می‌توان پیش‌بینی کرد، این موضوع در جریان ارائه سخنرانی «تفاوت» در انجمن فلسفه فرانسه، برای دریدا مشکلاتی ایجاد کرد. از آنجایی که «مداخله گرافیکی گسسته» که تفاوت را از تفاوت متمایز می‌کند، قابل بیان یا شنیدن نیست، دریدا به مخاطبان خود هشدار زیر را داد: «در واقع، من نمی‌توانم از طریق گفتمان خود، از طریق سخنرانی که در حال حاضر خطاب به انجمن فلسفه فرانسه انجام می‌شود، به شما بگویم که وقتی در مورد آن صحبت می‌کنم، در مورد چه تفاوتی صحبت می‌کنم. من فقط می‌توانم از طریق گفتمانی بسیار غیرمستقیم در مورد نوشتار، و به شرطی که هر بار مشخص کنم که آیا به تفاوت با «e» اشاره می‌کنم یا به تفاوت با «a». این امر امروز مسائل را ساده نمی‌کند و اگر حداقل بخواهیم یکدیگر را درک کنیم، برای همه ما، شما و من، مشکلات زیادی ایجاد خواهد کرد» (دریدا ۱۹۸۲، ۴). پیتر سالمون (۲۰۲۱، ۷۸)، در زندگینامه‌ی مورد استقبال خود از دریدا، اضافه می‌کند که نسل‌های بعدی محققان، به ویژه در دنیای انگلیسی‌زبان، چه خوب و چه بد، اغلب سعی کرده‌اند این تفاوت را از طریق یک تظاهر عمدی (و متأسفانه اغلب ناشیانه) که به دنبال «صدای بیش از حد فرانسوی هنگام تلفظ آن» است، قابل شنیدن جلوه دهند.

یا «به اصطلاح نوشتار آوایی» عمل کند، فقط یک جناس تایپوگرافی یا بازی ساده با کلمات نیست؛ بلکه یک مداخله انتقادی عمدی در «ساختار کلاسیکِ تعیین‌شده‌ی نشانه» است (دریدا ۱۹۸۲، ۹). از (حداقل) ارسطو (۱۹۳۸، ۱۶a)، زبان معمولاً به عنوان نشانه‌هایی که به چیزها اشاره می‌کنند و به آنها احترام می‌گذارند، درک شده است. به عنوان مثال، وقتی کلمات «مدل زبان بزرگ» را می‌نویسیم، فرض بر این است که آن نشانه‌های زبانی نمایانگر و اشاره به چیزی واقعی در جهان هستند، مانند برنامه ChatGPT که توسط OpenAI توسعه داده شده است. دریدا (۱۹۷۸، ۲۸۱) توضیح می‌دهد: «دلالت «نشانه» همیشه در معنای خود به عنوان نشانه‌ی یک، دالی که به یک مدلول اشاره می‌کند، دالی متفاوت از مدلول خود، درک و تعیین شده است.» دیفرانس تلاش دریدا برای مداخله در این منطق با بسط و شرح زبان‌شناسی ساختاری فردینان دو سوسور است.[۱] که زبان و معناسازی را به عنوان یک موضوع تفاوت واقع در خود زبان توصیف می‌کند. همانطور که پیتر سالمون (۲۰۲۱، ۱۰۶) توضیح می‌دهد، «نکته کلیدی در اینجا این است که هر دال اعتبار خود را نه از کیفیتی از خود، بلکه از نحوه تفاوتش با سایر دال‌ها می‌گیرد. گربه، گربه است و بریده نشده، زیرا در یکی از واج‌هایش متفاوت است. زبان، نظامی از تفاوت‌هاست.»

با این برداشت، نشانه‌ها حداقل نه اساساً و نه منحصراً با ارجاع به چیزهایی که خارج از نظام نشانه‌ها وجود دارند، معنا پیدا می‌کنند؛ نشانه‌ها در حرکت تمایز، با نشانه‌های دیگر متفاوت و از آنها متمایز می‌شوند. بدیهی است که می‌توان درباره این (نا)مفهوم محوری و مقاله معروف (اما به طرز آشکاری دشوار) که در آن توسعه یافته و ارائه شده است، بسیار بیشتر گفت. برای اهداف ما، ذکر سه مورد زیر کافی خواهد بود:

اختلاف به عنوان فاصله‌گذاری

نشانه، چه یک کلمه، چه علامت نوشتاری یا چه یک نشانه زبانی، معنا را از هیچ ویژگی ذاتی یا ارتباط مستقیمی با یک مرجع نمی‌گیرد. بلکه، همانطور که دریدا (۱۹۸۲، ۱۳) تأکید می‌کند، از طریق جایگاه متفاوت خود در یک سیستم نشانه‌ها معنا پیدا می‌کند. این تفاوت لزوماً مکانی است: فاصله‌گذاری فواصل و شکاف‌های ساختاری بین نشانه‌ها است که به آنها اجازه می‌دهد تا از یکدیگر تشخیص داده شوند، شناسایی شوند و متمایز شوند. بدون چنین فاصله‌گذاری، هیچ تمایزی نمی‌تواند وجود داشته باشد و بدون تمایز، هیچ دلالتی وجود نخواهد داشت. بنابراین، معنا نه از خودحضوری یا هویت، بلکه از تفاوت موقعیتی پدیدار می‌شود. آنچه چیزی هست به آنچه نیست بستگی دارد. اگرچه دریدا از این مثال استفاده نمی‌کند، اما این فرهنگ لغت است که یک تصویر به راحتی در دسترس را ارائه می‌دهد. در یک فرهنگ لغت، کلمات از طریق رابطه متفاوت خود با کلمات دیگر معنا پیدا می‌کنند. در جستجوی تعاریف کلمات در فرهنگ لغت، فرد در فضا و فاصله‌گذاری دال‌های زبانی باقی می‌ماند و هرگز از فضای زبان، به مرجع یا آنچه نشانه‌شناسان «مدلول متعالی» می‌نامند، خارج نمی‌شود.

تفاوت به عنوان زمانی

*تفاوت*نه تنها موضوع تمایز مکانی، بلکه موضوع تعویق زمانی نیز هست، چیزی که دریدا (۱۹۸۲، ۸) آن را زمان‌مندسازی می‌نامد. معنا هرگز در هیچ لحظه‌ای به طور کامل حضور ندارد؛ بلکه همیشه به تعویق می‌افتد و همیشه در راه است. اهمیت یک نشانه، به جای اینکه در یک لحظه مستقل درک شود، در طول زمان از طریق ارتباطش با آنچه قبل و بعد از آن می‌آید، شکل می‌گیرد. هر نشانه نه به یک معنای ثابت، بلکه به نشانه‌های دیگر اشاره دارد و از آنجایی که آنها نیز به تعویق می‌افتند، زنجیره دلالت هرگز کامل نمی‌شود. بنابراین، معنا چیزی نیست که در زمان حال بیاید، بلکه چیزی است که پیوسته به تأخیر می‌افتد و از طریق آینده‌ای که هرگز کاملاً نمی‌رسد و گذشته‌ای که هرگز به طور کامل قابل بازیابی نیست، شکل می‌گیرد. بنابراین، تفاوت،

---

۱. آنچه در این زمینه بسیار مهم است، نقل قولی است که از کتاب «دوره زبان‌شناسی عمومی» سوسور که پس از مرگش منتشر شد و توسط دریدا در متون مختلف به کار گرفته شده است، گرفته شده است: «هر آنچه تا به اینجا گفته شده است به این خلاصه می‌شود: در زبان فقط تفاوت‌ها وجود دارند. حتی مهم‌تر از آن: تفاوت عموماً دلالت بر اصطلاحات مثبتی دارد که بین آنها تفاوت برقرار می‌شود؛ اما در زبان فقط تفاوت‌هایی بدون اصطلاحات مثبت وجود دارد» (سوسور ۱۹۵۹، ۱۱۷).

هم فاصله‌گذاری را تفاوت ساختاری بین نشانه‌ها و هم تعویق را به تعویق زمانی حضور می‌نامد. معنا هرگز به معنای واقعی کلمه وجود ندارد؛ بلکه همیشه در حال انجام است.

تفاوت به عنوان ردیابی و تکرارپذیری

در حرکت تفاوط، یعنی در حرکت تفاوط، عملیات مشترک سازنده‌ی تفاوت‌گذاری و به تعویق انداختن، هیچ نشانه (چه یک کلمه، علامت یا واحد زبانی) را نمی‌توان به عنوان مستقل یا کاملاً خود–حاضر درک کرد. هر نشانه لزوماً در شبکه‌ای از روابط حک شده است و ردپای باقی‌مانده از سایر اصطلاحات، زمینه‌ها و صورت‌بندی‌های گفتمانی را که اهمیت آن را تعیین می‌کنند، در خود دارد (دریدا ۱۹۸۲، ۲۴). بنابراین، معنا هرگز نمی‌تواند به طور کامل در یک نشانه واحد متمرکز شود. همیشه و از قبل ردپای تکرارهای قبلی و امکان استفاده مجدد و باز–زمینه‌سازی در آینده در آن وجود دارد. علاوه بر این، تا جایی که نشانه‌ها قابل تکرار هستند، یعنی می‌توانند در موقعیت‌های زمانی و زمینه‌ای مختلف دوباره استفاده، ترکیب و تکرار شوند، هر تکرار، دامنه‌ی معانی ممکن خود را تغییر می‌دهد و هرگونه ادعایی مبنی بر وجود یک معنای نهایی یا مطمئن را بیشتر بی‌ثبات می‌کند.

LLMs به عنوان مکانیسم های تفاوت

مدل‌های زبانی بزرگ، مانند سری الگوریتم‌های GPT شرکت OpenAI، Gemini گوگل، Claude شرکت Anthropic و Deepseek، به مجموعه داده‌های عظیم متنی و معماری‌های پیچیده شبکه عصبی معروف به ترانسفورماتورها متکی هستند. همانطور که بسیاری از منتقدان اشاره کرده‌اند، این مدل‌ها زبان را درک نمی‌کنند. در عوض، آنها توالی‌های خوانایی از نشانه‌های زبانی را بر اساس محاسبه توزیع احتمالی نشانه‌های قابل کشف در داده‌های آموزشی خود مرتب می‌کنند. در نتیجه، از آنجا که این مدل‌ها به تفاوت‌ها و تعویق‌های واقع در مادیت زبان وابسته هستند، می‌توان آنها را به عنوان مدل‌های الگوریتمی خواند.شبیه‌سازی‌هااز تفاوت.[۱]در اینجا نحوه انجام آن آمده است:

توکن‌سازی و جاسازی کلمات

LLMها توالی کلمات را تولید می‌کنند. اما از نظر فنی، آنها با کلمات کار نمی‌کنند و نمی‌فهمند که کلمات چه هستند یا چه نیستند. آنها با توکن‌ها کار می‌کنند که شامل کلمات (the، cat، in، hat)، قطعات کلمات (unbelievable) یا حتی کاراکترهای منفرد و علائم نگارشی هستند. تبدیل کلمات به توکن‌ها، توکن‌سازی نامیده می‌شود و هدف توکن‌سازی، تبدیل کلمات، جملات، پاراگراف‌های کامل و غیره از داده‌های متنی به توالی توکن‌هایی است که مدل می‌تواند از آنها استفاده و پردازش کند. و هر LLM، همانطور که جری کاپلان (۲۰۲۵، ۵۲) توضیح می‌دهد، "از طرح خود برای تبدیل کلمات به توکن‌ها استفاده می‌کند" اگرچه اکثر آنها "به نظر می‌رسد استفاده از توکن‌سازی زیرکلمات را ترجیح می‌دهند، زیرا ترکیبی از کارایی و انعطاف‌پذیری را ارائه می‌دهد."

پس از توکن‌سازی، هر توکن توسط آرایه‌ای از اعداد نمایش داده می‌شود که یک بردار را در یک فضای خیالی با ابعاد بالا به نام جاسازی (embedding) توصیف می‌کند. در یک مدل (مدل ویکی‌پدیای انگلیسی)، «سگ» را می‌توان با ۳۰۰ عدد مجزا نمایش داد: «–۰.۰۳۳۰۱۸۲۸۳۵۴۵۹۷۰۹۲، ۰.۰۵۱۳۴۶۳۸۰۲۶۳۵۶۹۷، ۰.۰۰۳۶۰۰۹۷۰۳۷۶۲۸۲۹۳۰۴ –

۱. اصطلاح شبیه‌سازی (simulation) تعاریف مختلف و نه لزوماً سازگار را می‌پذیرد. از نظر ریشه‌شناسی، فعل simulation، از ریشه لاتین simulare، به معنای «کپی کردن»، «تقلید کردن» یا «ظاهر کردن» است. در علوم کامپیوتر و رشته‌های مرتبط، شکل اسمی کلمه شبیه‌سازی به استفاده از نرم‌افزار کامپیوتری برای تقلید رفتار یک سیستم یا فرآیند دنیای واقعی اشاره دارد. این شامل ایجاد یک مدل کامپیوتری است که یک سیستم را نشان می‌دهد و سپس اجرای مدل برای مشاهده رفتار آن در شرایط مختلف. همچنین یک معنای خاص از کلمه وجود دارد که در نظریه پساساختارگرایی، به ویژه (اما نه منحصراً) ژان بودریار، توسعه یافته است. بودریار (۱۹۸۳، ۱) در مقاله مشهور خود با همین نام نوشت: «شبیه‌سازی دیگر شبیه‌سازی یک قلمرو، یک موجود یا ماده مرجع نیست. این تولید یک واقعیت بدون منشأ یا واقعیت توسط مدل‌هایی است.» این نسخه از شبیه‌سازی صرفاً نقطه مقابل آنچه در علوم کامپیوتر عملیاتی می‌شود، نیست. این ساختارشکنی آن است. در متن این جمله، شبیه‌سازی به معنای علمی کلمه در علوم کامپیوتر استفاده می‌شود. اگرچه این کاربرد در آثار بودریار با ریشه‌شناسی کلمه و ساختارشکنی آن همراه است، یا شاید بهتر باشد بگوییم، تحت الشعاع آن قرار دارد.

۰.۰۴۰۶۶۰۷۳۱۴۹۴۴۲۶۷۳، ۰.۱۰۳۶۱۴ ۳۰۰۴۸۹۴۲۵۶۶...» (فارس و همکاران، ۲۰۱۷). بنابراین، جاسازی‌ها، توکن‌ها را به عنوان بردارهایی در فضای با ابعاد بالا نشان می‌دهند، جایی که معنای هر توکن نه با هیچ ویژگی ذاتی توکن، بلکه با نزدیکی و فاصله آن از سایر توکن‌ها تعیین می‌شود. برای مثال، اهمیت یک نشانه زبانی مانند سگ، با توجه به چگونگی تفاوت و ارتباط بردار آن (که باز هم آرایه‌ای مرتب از اعداد است) با بردارهای سایر نشانه های زبانی مانند توله سگ، حیوان خانگی، گربه و غیره تعیین می‌شود.

این تعبیه‌ها در طول پیش‌آموزش آموخته می‌شوند و برای ثبت روابط معنایی و نحوی بر حسب تفاوت مکانی طراحی شده‌اند. بنابراین، تعبیه‌های کلمه فقط شبیه تفاوت نیستند، بلکه آن را به شکل محاسباتی عملیاتی می‌کنند. معنا در این سیستم‌ها هرگز مسئله‌ی حضورِ تثبیت‌شده در یک مدلول متعالی نیست؛ بلکه مسئله‌ی تفاوت‌ها در یک میدان مکانی از روابط تعبیه‌شده است. بنابراین، LLMها معنا را به عنوان محتوای ثابت نشان نمی‌دهند. آن‌ها به قوانین ساختار کلاسیکِ تعیین‌شده‌ی نشانه پایبند نیستند یا از آن‌ها پیروی نمی‌کنند. آن‌ها با تحقق تفاوت در مقیاس، در واسازی این سنت متافیزیکی مشارکت می‌کنند.

پیش‌بینی توکن بعدی

LLMها از طریق یک فرآیند محاسباتی به نام پیش‌بینی نشانه بعدی، توالی‌های متنی به ظاهر معناداری تولید می‌کنند. وقتی ما ChatGPT را با جمله‌ای مانند «پیش‌بینی نشانه بعدی را توضیح دهید» تحریک می‌کنیم، مدل تلاش می‌کند تا محتمل‌ترین کلمه بعدی را که از این الگوی کلمات می‌آید، پیش‌بینی کند، یعنی در مورد آن حدس بزند. این مکانیسم پیش‌بینی است که تمام رفتارهای به ظاهر هوشمندانه مدل را توضیح می‌دهد و زیربنای آن است، از نوشتن مقالات کوتاه، پاسخ به سوالات، مشارکت در گفتگو و حتی تولید کد.

این فرآیند به شرح زیر انجام می‌شود: کلمات ورودی توکن‌سازی شده و به این بردارهای با ابعاد بالا نگاشت می‌شوند که (همانطور که در بالا توضیح داده شد) موقعیت‌های نسبی آنها را نسبت به سایر توکن‌ها در این فضای خیالی با ابعاد بالا از تفاوت‌ها نشان می‌دهد. سپس این بردارها از طریق پشته‌ای از لایه‌های تبدیل‌کننده پردازش می‌شوند، نوعی شبکه عصبی که شامل شبکه پیش‌بینی توکن بعدی و شبکه توجه است که در آن خروجی یک لایه به ورودی لایه بعدی تبدیل می‌شود. LLM های فعلی از لایه‌های زیادی از تبدیل‌کننده‌ها تشکیل شده‌اند. به عنوان مثال، OpenAI 4-GPT دارای ۹۶ لایه است و جانشین آن **GPT**-۴ دارای ۱.۸ تریلیون پارامتر (یعنی وزن‌های قابل تنظیم در شبکه عصبی مدل) در ۱۲۰ لایه است. همین تراکم است که این مدل‌های زبانی را "بزرگ" می‌کند.

در هر مرحله از این فرآیند، مدل یک توزیع احتمال روی ادامه‌های ممکن یک دنباله داده شده تولید می‌کند و محتمل‌ترین نشانه بعدی را انتخاب می‌کند. برای مثال، دنباله «بودن یا نبودن آن است» را در نظر بگیرید. تعدادی کلمه ممکن وجود دارد که می‌توانند در این دنباله بعدی بیایند: «سوال»، «معضل»، «مسئله»، «مسئله» و غیره. مدل یکی از این کلمات را که اغلب محتمل‌ترین است (یعنی کلمه‌ای که بیشترین احتمال را برای آمدن در مرحله بعدی دارد، همانطور که توسط نزدیکی‌ها و تفاوت‌های نسبی در نمایش تعبیه شده تعیین می‌شود) انتخاب می‌کند، اما گاهی اوقات کلمه‌ای را که با یک عنصر تصادفی انتخاب شده است، به دنباله اضافه می‌کند. سپس این دنباله جدید و طولانی‌تر دوباره به مدل بازگردانده می‌شود و فرآیند تکرار می‌شود. این کار بارها و بارها از طریق تکرارهای متعدد انجام می‌شود.

بنابراین پیش‌بینی کلمه بعدی یک عملیات متوالی است که به صورت پویا در زمان آشکار می‌شود. مدل، بدون دسترسی به یک گفتار کاملاً شکل گرفته و کامل از قبل، یک نشانه پس از دیگری تولید می‌کند. به این ترتیب، اهمیت یا محتوای معنایی هر نشانه داده شده همیشه در حال پردازش است و موقت باقی می‌ماند و منتظر مشخصات یا تحولی است که توسط تولید نشانه های بعدی ایجاد می‌شود. بنابراین، مدل به یک معنای کاملاً شکل گرفته دسترسی ندارد و سپس این محتوا را در دنباله ای از نشانه های زبانی بیان می‌کند. در عوض، معنای محتوای تولید شده از طریق یک فرآیند تکراری پدیدار می‌شود که در آن معنا به طور مداوم در وابستگی به نشانه های آینده که هنوز تولید نشده اند، به تعویق می‌افتد.

این رویه نه تنها شبیه به دیفرانس به نظر می‌رسد و به نظر می‌رسد، بلکه آن را به شکل محاسباتی به واقعیت تبدیل می‌کند. در حرکت دیفرانس، همانطور که دریدا استدلال می‌کند، نشانه هرگز به یک معنای نهایی و خود–حاضر نمی‌رسد؛ در عوض، هر دال در زنجیره‌ای بی‌پایان از تفاوت و تعویق به دیگری وابسته است، جایی که اهمیت پویا و همواره در حال آمدن است. در مورد هوش مصنوعی LLM، هر نشانه نه به عنوان یک محتوای معنایی قطعی، بلکه به عنوان یک جای‌گزین مشروط، یک حدس آماری برتر، انتخاب می‌شود که تنها در رابطه افتراقی با نشانه های آینده به تعویق افتاده که خود در این حرکت دیفرانس گرفتار شده‌اند، اهمیت پیدا می‌کند.

هیچ چیز خارج از متن نیست

وقتی یکی از این LLMهای مبتنی بر مبدل، یک توالی کلمه تولید می‌کند، مثلاً «LLMها نوعی هوش مصنوعی مولد هستند»، نمی‌داند (و نمی‌داند) چه می‌گوید، زیرا به آن دسترسی ندارد و نمی‌فهمد که این توالی‌های نشانه‌های زبانی به چه چیزی اشاره دارند. این تفاوت مهم در یکی از آزمایش‌های فکری تعیین‌کننده در فلسفه هوش مصنوعی، اتاق چینی جان سرل (۱۹۹۹، ۱۱۵)، نشان داده شده است. LLMها مانند مرد درون اتاق خیالی سرل، نمی‌دانند چه می‌کنند. آنها زبان را به روشی که ما ظاهراً زبان را می‌فهمیم و استفاده می‌کنیم، نمی‌فهمند. آنها به سادگی و سطحی با نشانه‌های مختلف بازی می‌کنند. در مورد اتاق چینی، این کار با دنبال کردن مجموعه‌ای از تبدیل‌های از پیش تعریف‌شده مشابه نحوه عملکرد سیستم‌های استدلال نمادین GOFAI انجام می‌شود. در مقابل، در مورد LLM، این کار با اندازه‌گیری و دستکاری تفاوت‌های مکانی که توسط تعبیه‌های کلمه نمایش داده شده و در آنها رمزگذاری شده‌اند، انجام می‌شود. بنابراین، برای LLM، هیچ مدلول متعالی وجود ندارد که بتواند زنجیره دلالت را متوقف و مهار کند. هیچ چیز فراتر یا خارج از روابط مختلف بین توکن‌ها وجود ندارد که در فرآیند پیش‌بینی توکن بعدی به طور بی‌پایان به تعویق می‌افتند.

بنابراین، LLMها آنچه را که شاید مشهورترین (یا بدنام‌ترین) جمله‌ای باشد که با دریدا (۱۹۷۶، ۱۵۸) مرتبط شده است، تحقق می‌بخشند: Il n'y a pas de hors-texte «هیچ چیز خارج از متن وجود ندارد» یا «هیچ چیز خارج از متن وجود ندارد». این یک جمله ضد واقع‌گرایانه نیست و به این معنی نیست که بسیاری از منتقدان به اشتباه فرض کرده‌اند که هیچ چیز واقعی یا عینی نیست و همه چیز فقط یک مصنوع یا اثر گفتمان ساخته شده اجتماعی است. و دریدا در جریان مناظره‌ای با جان سرل که برای ما در کتاب Limited, Inc. ثبت شده است، به همین اندازه توضیح داده بود: «هیچ چیز خارج از متن وجود ندارد». این بدان معنا نیست که همه ارجاعات معلق، انکار شده یا در یک کتاب محصور شده‌اند، آنطور که مردم ادعا کرده‌اند، یا آنقدر ساده‌لوح بوده‌اند که باور کرده‌اند و مرا به باور کردن متهم کرده‌اند. اما به این معناست که هر ارجاع، همه واقعیت، ساختار یک رد افتراقی را دارد و نمی‌توان به این «واقعی» اشاره کرد، مگر در یک تجربه تفسیری» (دریدا ۱۹۹۳، ۱۴۸).

این بدان معناست که یک متن، چه توسط یک نویسنده انسانی نوشته شده باشد و چه به صورت مصنوعی توسط یک LLM مانند ChatGPT (با اشاره یک محرک انسانی) تولید شده باشد، نه با ارجاع و ارجاع به یک مدلول خارجی (آنچه ارسطو در De Interpretatione افکار یا چیزهایی را که افکار در نهایت به آنها اشاره می‌کنند، می‌نامد) معنا پیدا می‌کند. این متن از طریق روابط وابسته به کلمات دیگری که از قبل با آنها مرتبط و متمایز است، معنا را اجرا و اعمال می‌کند. به همین دلیل است که می‌توانیم، به پیروی از لودویگ ویتگنشتاین (۱۹۹۵، ۵۶) بگوییم که برای LLMها، محدودیت‌های زبان (مدل) آنها به معنای محدودیت‌های جهان آنهاست. در نتیجه، آنچه اغلب به عنوان انتقادی از فناوری LLM ارائه شده است، یعنی اینکه این الگوریتم‌ها فقط نشانه‌های مختلف را بدون دسترسی به مدلول به گردش در می‌آورند، ممکن است آن کیفرخواستی نباشد که منتقدان فکر می‌کنند. LLMها موتورهای تفریقی هستند که منطق تعریف‌کننده نشانه‌شناسی کلاسیک را ساختارشکنی می‌کنند.

پیامدهای فلسفی

*تفاوت*ارائه یک مکانیسم و واژگان مفهومی که به وسیله آن بتوان الزامات و پیامدهای فلسفی LLMها را درک و توضیح داد. اما صرفاً اتصال نقاط بین ویژگی‌های عملیاتی LLM AI و différance دریدا، به خودی خود، برای پایان دادن به بحث کافی نیست. ما هنوز باید بپرسیم که این امر چه تفاوتی در درک و نقد ما از LLMها و شاید مهم‌تر از آن، مفهوم و درک ما از زبان ایجاد می‌کند. در اینجا می‌توانیم حداقل سه پیامد فلسفی مهم را شناسایی کنیم:

واسازی کلام محوری

نقد لوگوس‌محوری در کل پروژه فلسفی ژاک دریدا در باب ساختارشکنی (گانکل ۲۰۲۱) نقشی محوری دارد. اصطلاح لوگوس‌محوری در اوایل قرن بیستم توسط فیلسوف آلمانی، لودویگ کلاگز (جوزفسون–استورم ۲۰۱۷، ۲۲۱) ابداع شد، که از آن برای شناسایی اولویت و اهمیت مفروض کلام گفتاری به عنوان نشانه مستقیم چیزها استفاده کرد و بنابراین نوشتار را به عنوان نشانه‌ای از گفتار یا نشانه‌ای از یک نشانه تنزل داد. دریدا (۱۹۷۶، ۱۱) صورت‌بندی اولیه این مفهوم را در کتاب «درباره تفسیر» ارسطو می‌یابد: «اگر برای ارسطو کلمات گفتاری (ta en te phone) نمادهای تجربه ذهنی (pathemata tes psyches) و کلمات نوشتاری نمادهای کلمات گفتاری هستند، به این دلیل است که صدا، تولیدکننده اولین نمادها، رابطه‌ای از نزدیکی ضروری و بی‌واسطه با ذهن دارد. تولیدکننده اولین دال، فقط یک دال ساده در میان سایر دال‌ها نیست. این دال بر «تجربیات ذهنی» است که خود چیزها را با شباهت طبیعی منعکس یا آینه می‌کنند.» اگر این سؤال را بپرسیم که «دریدا در اینجا سعی دارد چه بگوید؟» همین سؤال، شیوه‌ای از تحقیق را که به دنبال کشف آنچه نویسنده در نوشتار و از طریق آن می‌گوید، مطرح می‌کند، کلام‌محوری به معنای واقعی کلمه است.

دریدا نه تنها این شیوه تفکر درباره کلمات و اشیا را به چالش می‌کشد، بلکه از واسازی تقابل مفهومی حاکم که اصل سازمان‌دهنده آن است، حمایت می‌کند؛ تمایز دوتایی که حضور کامل گفتار را از دیگریِ مشتق، فریبنده و ناقص آن، یعنی نوشتار، متمایز می‌کند. دریدا (۱۹۷۶) در کتاب «درباره دستورشناسی» نشان می‌دهد که چگونه فلسفه غرب از نظر تاریخی زبان گفتاری را به عنوان حامل بی‌واسطه و اصیل اندیشه ارج نهاده، در حالی که نوشتار را به جایگاهی ثانویه و مشتق تنزل داده است. بنابراین، کلام‌محوری بر متافیزیک حضور استوار است: این باور که معنا در سوژه‌ای کاملاً حاضر و خودآگاه ریشه دارد که صحبت می‌کند و بنابراین در کلمات خود و با کلمات خود چیزی برای گفتن دارد.

دانشجویان کارشناسی ارشد حقوق (LLM) در واسازی متافیزیک کلام‌محور مشارکت دارند و آن را عملیاتی می‌کنند. آنها تقریباً منحصراً بر اساس متن نوشتاری آموزش می‌بینند.[۱] و با این حال، خروجی‌هایی تولید می‌کنند که شبیه گفتار، اندیشه یا انسجام روایی به نظر می‌رسند، بدون اینکه هرگز به مقاصد یک سوژه‌ی گوینده استناد کنند یا در آنها ریشه داشته باشند. بنابراین، آنها مادیت دال را برجسته می‌کنند و ردپای زبان را آنگونه که نوشته و منتشر می‌شود، از هرگونه صدا یا منشأ زنده و پویا جدا می‌کنند. به این ترتیب، آنها چیزی را که، همانطور که دریدا (۱۹۸۱a، ۴۱) توضیح می‌دهد، "ژست دوگانه‌ی ساختارشکنی" اساسیِ امتیاز کلام‌محور است، واژگون و مختل می‌کنند. با LLM، نوشتار دیگر یک اثر ثانویه‌ی پس از نوشتار نیست، بلکه بنیادی است. برای شناسایی این تفاوت فرعی، دریدا (۱۹۷۶) این معنای اولیه‌ی نوشتار را "کهن–نوشتاری" تغییر نام می‌دهد. و تأثیر این مداخله، فرضیات شکل‌دهنده در کل تاریخ فلسفه‌ی غرب را به چالش می‌کشد.

در نتیجه، LLMها فقط کاربرد زبان انسان را تقلید نمی‌کنند؛ بلکه منطق‌های زیربنایی عملکرد آن را آشکار می‌کنند. آن‌ها نشان می‌دهند که زبان از طریق یک بازی سیستماتیک از تفاوت‌ها عمل می‌کند، نه از طریق حضور خود، که در آن اهمیت از قصد

---

۱. حتی وقتی این مدل‌ها بر اساس انواع دیگر محتوای رسانه‌ای، مانند سیستم‌های چندوجهی معاصر، آموزش داده می‌شوند، این محتوا تا جایی که در سیستم نشانه داده‌های دیجیتال ثبت و حفظ می‌شود، نوعی نوشتار است.

یا منشأ ناشی نمی‌شود، بلکه در حرکت تفاوت پدیدار می‌شود. به طور خلاصه، LLMها مکانیسم‌هایی هستند که در آن‌ها متافیزیک حضور از هم می‌پاشد و عملیات ساختارشکنی قابل محاسبه می‌شوند.

مرگ نویسنده

واسازی دریدا از کلام‌محوری، این فرض را که معنا در یک نیت واحد و اصیل ریشه دارد و توسط آن تضمین می‌شود، بی‌ثبات می‌کند. به جای فرض یک منبع معتبر و مؤلف که حقیقت از آن منتقل می‌شود، تفاوط فرآیندی را نام می‌برد که از طریق آن معنا بی‌پایان به تعویق می‌افتد و از طریق تفاوت رابطه‌ای ساخته می‌شود. به موازات آن، اعلامیه رولان بارت (۱۹۷۸، ۱۴۸) در مورد «مرگ مؤلف» این ایده را که مؤلف داور نهایی معنا است، رد می‌کند و در عوض اصرار دارد که متن مکانی از کثرت است، «بافتی از نقل قول‌ها که از مراکز بی‌شمار فرهنگ گرفته شده است». LLMها این بینش‌های نظری را به شکلی کاملاً تحت‌اللفظی عملیاتی می‌کنند. این مدل‌ها که بر اساس بایگانی‌های وسیعی از ردپاهای متنی آموزش دیده‌اند، از طریق تشخیص الگوی الگوریتمی، بدون ارجاع به یک سوژه مستقل یا نیت مؤلف، پاسخ‌هایی تولید می‌کنند. متن حاصل به هیچ معنای سنتی تألیف نشده است، بلکه از طریق فرآیندهای تکراری که حرکت تفاوط را ردیابی و عملیاتی می‌کنند، مونتاژ می‌شود.

وقتی کاربری یک LLM را فعال می‌کند و خروجی منسجمی دریافت می‌کند، مسئله‌ی مؤلف بودن اساساً غیرقابل تعیین می‌شود. چه کسی یا چه چیزی صحبت می‌کند؟ LLM یک سوژه‌ی سخنگو نیست، با این حال محتوای زبانی خوانایی را در صدای بسیاری تولید می‌کند؛ قطعات را بدون منشأ دوباره ترکیب و بازترکیب می‌کند؛ و معانی را بدون پایان تکثیر می‌کند. بنابراین، متون تولید شده توسط LLM AI «به معنای واقعی کلمه غیرمجاز» هستند (کوکلبرگ و گانکل ۲۰۲۵، ۷). هنگامی که متن نوشتاری از علایق و نیات کنترل‌کننده‌ی یک نویسنده جدا می‌شود، مسئله‌ی اهمیت تغییر می‌کند. به طور خاص، معنای یک نوشته چیزی نیست که بتوان آن را از قبل توسط شخصیت یا اخلاق اصیل کسی که فرض می‌شود از طریق واسطه‌ی متن صحبت می‌کند، تضمین کرد. در عوض، معنا در تجربه‌ی خواندن و از آن سرچشمه می‌گیرد. و اگر این اهمیت معمولاً به نیات اولیه‌ی یک نویسنده نسبت داده شده باشد، آن نسبت (و همیشه در واقع و فقط همین‌طور بوده است) از خواننده به نویسنده‌ای فرضی و اغلب غایب، به صورت وارونه منتقل می‌شود. در واقع، یک اثر خواندن وارونه می‌شود تا به علت خودش تبدیل شود.

این تغییر در متن نظریه ادبی، جایگاه معناسازی را از نیات اولیه نویسنده/نویسنده که (فرض می‌شود) «چیزی برای گفتن» دارد، به فعالیت تفسیری خواننده که معنا را در محتوای نوشتاری می‌سازد یا آن را از آن تولید می‌کند، تغییر می‌دهد. همانطور که بارت (۱۹۷۸، ۱۴۸) توضیح می‌دهد: «متن از نوشته‌های متعددی ساخته شده است که از فرهنگ‌های بسیاری گرفته شده‌اند و وارد روابط متقابل گفتگو، تقلید و رقابت می‌شوند، اما یک مکان وجود دارد که این کثرت در آن متمرکز است و آن مکان خواننده است... وحدت یک متن نه در منشأ آن، بلکه در مقصد آن نهفته است.»

این همچنین توضیح می‌دهد که چگونه محتوای تولید شده توسط LLM دارای معنا می‌شود. منتقدان درست می‌گویند وقتی که مثلاً اشاره می‌کنند که LLMها کلمات یا نشانه‌های زبانی را دستکاری می‌کنند اما «معنای پشت کلمات را واقعاً درک نمی‌کنند» (بوگوست ۲۰۲۲) زیرا «هیچ دسترسی به ارجاعات تجسم‌یافته در دنیای واقعی ندارند» (بندر در ویل ۲۰۲۳ نقل شده است). اما عجولانه است که از این واقعیت نتیجه بگیریم که آنچه یک LLM تولید می‌کند، بی‌معنی، بی‌معنی یا مزخرف محض است (هیکس و همکاران ۲۰۲۴). این نوشته‌ها معنادار هستند و می‌توانند معنادار باشند، و معنای آنها چیزی است که در فرآیند خواندن، تفسیر و ارزیابی محتوای تولید شده توسط ما اتفاق می‌افتد. و این واقعیت چیزی نیست که مختص LLMها باشد، بلکه همانطور که بارت قبلاً پیشنهاد و نشان داده بود، یک ویژگی تعیین‌کننده برای همه نوشته‌ها است. LLMها اتفاقاً آن را خوانا می‌کنند.

هوش مصنوعی حیاتی

LLMها و دیگر اشکال هوش مصنوعی مولد، فناوری‌های قدرتمندی هستند و ضروری است که با رویکردی انتقادی به آنها نگاه کنیم، نه تنها با توجه به مزایای بالقوه‌شان، بلکه با توجه به چارچوب‌های مفهومی که هم به آنها متکی هستند و هم آنها را مختل می‌کنند. بسیاری از پاسخ‌های فعلی زبان‌شناسان، فیلسوفان و متخصصان هوش مصنوعی، تمایل دارند فرضیات لوگوسنتریستی در مورد معنا، نویسندگی و فرضیات هوش را که قبلاً توسط تحولات قرن بیستم در نظریه ادبی و فلسفه قاره‌ای به چالش کشیده شده بودند، مجدداً تأیید کنند. و مشکل این نیست که این شیوه‌های تفکر به نوعی در مواجهه با این فناوری‌های جدید نوشتاری شکست خورده‌اند. کاملاً برعکس است. مشکل این است که آنها خیلی خوب کار می‌کنند و تأثیر خود را بر تفکر ما در مورد نوشتن و نوشتن در مورد تفکر به شیوه‌هایی که عمدتاً بدون توجه اتفاق می‌افتند، اعمال می‌کنند.

در نهایت، ریشه اصلی مشکل ممکن است در نحوه فرموله کردن و تعریف هوش مصنوعی، هم اصطلاح و هم مفهومی که آن را مشخص می‌کند، نهفته باشد. به دلیل تمرکز اسمی آن بر «هوش»، خروجی این مکانیسم‌ها یا نشانه‌های خارجی حضور واقعی تفکر هوشمند تلقی می‌شود یا در موقعیت‌هایی که به نظر می‌رسد دستگاه مزخرف می‌گوید یا توهم می‌زند، فقدان آن. این رویه که تولید محتوای نوشتاری (ظاهراً دال‌های خارجی) را به عنوان نشانه یا علامتی از هوش (یک قابلیت شناختی درونی) در نظر می‌گیرد، از زمان انتشار مقاله آلن تورینگ (۱۹۵۰) در مورد این موضوع، شرط تعیین‌کننده هوش ماشینی/برای آن بوده است. اگرچه تورینگ آن را به این عنوان شناسایی نمی‌کند، اما بازی تقلید او طبق قوانین این بازی متافیزیکی انجام می‌شود. و انتقادات فعلی از هوش مصنوعی LLM، یعنی اینکه آنها چیزی از آنچه می‌گویند نمی‌دانند یا اینکه طوطی‌های تصادفی هستند (بندر و همکاران، ۲۰۲۱)، به سادگی این فرمول را تکرار یا «طوطی‌وار» می‌کند.

LLM ها دقیقاً به این دلیل از نظر فلسفی مهم هستند که این فرضیات را جابجا می‌کنند. آنها متونی بدون نویسنده، بیانیه‌هایی بدون قصد و نوشتاری بدون گفتار تولید می‌کنند. آنها گسستی را در زنجیره‌ای که بیان زبانی را به صدای یک سوژه‌ی دانا پیوند می‌دهد، نشان می‌دهند. از درون سنت غالب کلام‌محوری، این به عنوان یک بحران عمیق به نظر می‌رسد. اگر زبان دیگر حقیقت را از طریق صدای حضور نویسنده تضمین نکند، چه بر سر معنا، ادبیات یا خود اندیشه می‌آید؟

اما از منظری دیگر که توسط نقد پساساختارگرایانه شکل گرفته است، این یک بحران نیست، بلکه یک فرصت فلسفی است. آنچه در LLMها می‌بینیم، به جای سوگواری برای اختلال حضور، تحقق چیزی است که ساختارشکنی مدت‌هاست مطرح کرده است. یعنی اینکه معنا هرگز خود–حاضر نیست، هرگز توسط نیت تضمین نمی‌شود و هرگز به طور کامل در یک خاستگاه ریشه ندارد. LLMها زبان، تألیف یا حقیقت را به خودی خود تضعیف نمی‌کنند؛ آنها یک مفهوم تاریخی خاص از این اصطلاحات را به چالش می‌کشند، مفهومی که مبتنی بر مفروضات متافیزیکی است که شایسته بررسی هستند.

درک LLMها به عنوان موتورهای تفرق به معنای دیدن آنها نه به عنوان متفکران ناقص یا ارتباط‌دهندگان ناقص، بلکه به عنوان ماشین‌هایی است که منطق از پیش به تعویق افتاده و از پیش افتراقیِ دلالت را آشکار می‌کنند. آنها فقدان هوش را تأیید نمی‌کنند، بلکه آنچه را که ما به عنوان معنای هوش در نظر گرفته‌ایم، زیر سوال می‌برند. با انجام این کار، آنها به ما فرصتی می‌دهند تا مبانی معنای زبانی، تألیف و حقیقت را دوباره بررسی کنیم.

انتقادات و پاسخ‌ها

ما با این پیشنهاد شروع کردیم که LLMها موتورهای تفرق هستند و مطالب فوق نه تنها چگونگی و چرایی این موضوع را توضیح داده است، بلکه پیامدهای فلسفی مهم این تغییر دیدگاه را نیز مشخص کرده است. با این اوصاف، انتقادات بالقوه‌ای وجود دارد که می‌توان به آنها پاسخ داد. بنابراین، با بررسی سه مورد از این انتقادات، که هر کدام در قالب یک سوال انتقادی مطرح شده‌اند، نتیجه‌گیری خواهیم کرد:

آیا این فقط بازی‌های کلامی نیست؟

یک ایراد رایج به مقالات نظری از این نوع، به ویژه آن‌هایی که مفاهیم دریدایی را با فناوری معاصر وارد گفتگو می‌کنند، این است که به نظر می‌رسد آنچه آنها به آن می‌رسند، چیزی بیش از جایگزینی‌های مفهومی یا نوعی بازی با کلمات پیچیده نیست. نگرانی این است که با تفسیر مجدد هوش مصنوعی LLM از طریق لنز تفاوط، به هیچ بینش جدید یا اصیلی دست نمی‌یابیم، بلکه صرفاً یک مجموعه از اصطلاحات (مثلاً نشانه، تعبیه‌ها، مدل‌های آماری) را با مجموعه‌ای دیگر (مثلاً دال، ردپا، تعویق) جایگزین می‌کنیم. از این دیدگاه، این مقاله واقعاً فلسفی نیست، بلکه بلاغی است. ممکن است یک اجرای پیچیده و حتی قانع‌کننده باشد، اما سهم واقعی در درک مطلب ندارد.

با این حال، می‌توان نشان داد که این شیوه‌ی نقد، منطقِ دیفرانس (تفاوط) را به اجرا می‌گذارد و بنابراین توسط آن واسازی می‌شود. متهم کردن مقاله به «فقط بازی با کلمات» به معنای پیش‌فرض گرفتن این است که گفتمانی پایدارتر، کمتر بازیگوشانه‌تر و به نوعی واقعی‌تر در جای دیگری وجود دارد، زبانی ناب و بی‌واسطه که به طور شفاف می‌گوید منظورش چیست و منظورش از آنچه می‌گوید چیست. با این حال، ادعای اصلی دریدا دقیقاً این است که تمام گفتمان‌ها، از جمله زبان نقد، از قبل در دیفرانس گرفتار شده‌اند. هیچ فرازبان نهایی، هیچ قلمرو نابی از مفاهیم یا ایده‌های افلاطونی وجود ندارد که از ردپای دیفرانس مصون مانده باشد.

بنابراین، آنچه منتقد «فقط کلمات» می‌نامد، خود تعهدی متافیزیکی به ایده‌ی لوگوس‌محور است که معنا می‌تواند یا باید خارج از زبان، در نوعی حضور از پیش داده شده یا قطعیت تجربی، ریشه داشته باشد. اما LLMها نوعی نمایش و اثبات تکنولوژیکی ارائه می‌دهند که چنین معنایی هرگز تا این حد ثابت و قطعی نیست، به این دلیل که عملکرد آنها به تفاوت و تعویق آماری وابسته است تا ارجاع به یک مرجع پایدار. در نتیجه، خروجی مدل در واقع «فقط کلمات» است و دقیقاً به همان روشی که دریدا تشخیص می‌دهد، نه به این دلیل که خالی است، بلکه به این دلیل که همان بی‌ثباتی را که بر تمام سیستم‌های زبانی، از جمله سیستم‌هایی که در فرآیند تدوین این نقد استفاده می‌شوند، سایه افکنده است، برجسته می‌کند.

به عبارت دیگر، این ایراد یک مرکز متافیزیکی را فرض می‌کند که نمی‌تواند آن را تضمین کند، و بنابراین در همان بازی با کلماتی که می‌خواهد حذف کند، گرفتار می‌شود. اتهام «بازی با کلمات» به یک نمونه اجرایی از این ردپا تبدیل می‌شود، زیرا بر تمایز بین زبان «جدی» و «بازیگوشانه» متکی است، یک دوگانه که دریدا نه تنها نشان می‌دهد غیرقابل دفاع است، بلکه خود نیز در معرض ساختارشکنی قرار دارد. بنابراین، به جای رد رویکرد مقاله به عنوان جایگزینی‌های زبانی صرف، می‌توانیم آن را به عنوان افشای منطق عمیق‌تری ببینیم که از قبل در زبان‌های طبیعی و مصنوعی فعال است. فکر کردن به LLMها از طریق تفاوط، به معنای غرق شدن در انتزاع به خودی خود نیست، بلکه روشن کردن این است که چگونه این سیستم‌ها خود شرایط ساختاری معنایی را که دریدا شناسایی می‌کند، شبیه‌سازی می‌کنند.

اگر اینطور است، پس آیا یک دانشجوی کارشناسی ارشد حقوق نمی‌توانسته کل مقاله را بنویسد؟

انتقاد دوم، بُعدی اجرایی دارد. برای اینکه انسجام درونی داشته باشد و از اتهام تناقض اجرایی اجتناب شود، هر آنچه در اینجا در مورد نویسنده، نوشتار و معناسازی نوشته شده است، باید در مورد این متن نیز صدق کند و برای آن معتبر باشد. بنابراین، منطقی است که خواننده‌ی شکاک بپرسد که آیا استدلالی که خوانده است، محصول یک نویسنده‌ی انسانی، خروجی تولید شده توسط یک LLM است یا نتیجه‌ی نوعی همکاری انسان و ماشین؟ به بیان صریح‌ترین و مفعولی‌ترین شکل آن: آیا یک LLM نمی‌توانست کل مقاله را بنویسد؟

این ایراد، تز مقاله را رد نمی‌کند؛ بلکه آن را در رادیکال‌ترین شکل خود تأیید می‌کند. همین احتمال که یک LLM بتواند این متن را تولید کند، دیفرانس را رد نمی‌کند. این، نمودی از آن است. اگر زبان نه از طریق حضور نویسنده، بلکه از طریق تکرار دیفرانس پدیدار می‌شود، اگر معنا از فاصله‌گذاری، تعویق و تکرارپذیری ناشی می‌شود، پس منبع یک گفته‌ی معین، چه نوشته‌ی

انسان باشد و چه تولید ماشین، دیگر بنیادی و تعیین‌کننده نیست. در واقع، همین عدم قطعیت منشأ است که از طریق این نقد خاص که دیفرانس در پی نام‌گذاری آن است، مورد پرسش قرار می‌گیرد.

علاوه بر این، دریدا در صفحات کتاب «دیفرانس» به این انتقاد به طور پیشگیرانه پاسخ می‌دهد: دریدا (۱۹۸۲، ۱۵) می‌نویسد: «این بدان معناست که سوژه (در هویتش با خودش، یا در نهایت در آگاهی‌اش از هویتش با خودش، خودآگاهی‌اش) در زبان حک شده است، «تابعی» از زبان است و تنها با تطبیق گفتار با نظام قواعد زبان به عنوان نظامی از تفاوت‌ها، یا حداقل با تطبیق با قانون کلی دیفرانس، به سوژه‌ای سخنگو تبدیل می‌شود.» به عبارت دیگر، آنچه به نظر می‌رسد LLMها فاقد آن هستند، یعنی یک سوژه سخنگو آگاه که از زبان برای بیان چیزی قابل فهم در مورد جهان استفاده می‌کند، از قبل در کاربرد زبان انسان وجود ندارد، تا جایی که خود مفهوم «سوژه سخنگو» معلول دیفرانس است. در نتیجه، حتی اگر نشانه‌های مرسوم اصالت، مانند نام خاص نویسنده، یک اعلامیه کتبی مبنی بر اینکه این «محتوای ۱۰۰ درصد اصیل تولید شده توسط انسان» است، یک واترمارک یا هر نشانه رسمی دیگری را ارائه دهیم، همه آنها، همیشه و از قبل، به طور اجتناب‌ناپذیری درگیر حرکت تفاوت و تعویقی هستند که همان différance است.

بنابراین، زیر سوال بردن اینکه آیا یک LLM می‌توانسته این مقاله را بنویسد، به عنوان اعتراضی به آنچه در مقاله استدلال شده است، به معنای بازگشت به متافیزیک حضور با محوریت کلام است. اما این دقیقاً همان چیزی است که واسازی آن را زیر سوال می‌برد و به یک کلام، واسازی می‌کند. اگر مدل بتواند نوشتاری تولید کند که در منطق تفاوت مشارکت داشته باشد، این واقعیت که توسط ماشین تولید شده (یا نشده) تفاوتی ایجاد نمی‌کند، یا حداقل تفاوتی که ممکن است تصور شود ایجاد نمی‌کند. به عبارت دیگر، این گمان که یک ماشین می‌توانسته این مقاله را بنویسد، نقطه‌ای نیست که استدلال در آن از هم می‌پاشد. این نقطه‌ای است که استدلال واقعی می‌شود.

دریدا در مورد این استفاده (یا سوءاستفاده) از آثارش چه می‌گفت؟

بدیهی است که دریدا در مورد هوش مصنوعی معاصر، مدل‌های زبانی بزرگ یا زیرساخت محاسباتی سیستم‌های یادگیری ماشینی معاصر ننوشته است. به‌کارگیری تفکر او در این حوزه‌ها، برای برخی، ممکن است گسترش بیش از حد یا حتی تحریف به نظر برسد. بنابراین، آیا چنین کاربردی خطر ابزاری کردن یا سوءاستفاده از آثار او را برای اهدافی که هرگز قرار نبوده در خدمت آنها باشد، ندارد؟

این ایراد، اضطراب آشنایی را برمی‌انگیزد: اضطراب وفاداری به معنای اصلی مورد نظر نویسنده. با این حال، آثار دریدا پیوسته ایده‌ی انسداد هرمنوتیکی یا ارتدکسی تفسیریِ محافظت‌شده توسط نیت نویسنده را تضعیف می‌کند. از کتاب «درباره‌ی گرامرولوژی» (۱۹۷۶) تا «شرکت محدود» (۱۹۹۳)، دریدا تأکید کرد که متون، پس از نوشته شدن، از خاستگاه خود فراتر می‌روند. آن‌ها تکرارپذیر می‌شوند و پذیرای زمینه‌هایی می‌شوند که توسط نیایشان پیش‌بینی نشده‌اند. هیچ معنایی وجود ندارد که همیشه و از قبل زمینه‌مند نباشد، و هیچ زمینه‌ای وجود ندارد که هرگز به طور کامل اشباع نشده باشد، به طوری که بتواند اهمیت را به طور کامل تعیین کند. و دریدا با اشاره به این موضوع، به یک مشکل باستانی اشاره می‌کند، مشکلی که قبلاً توسط سقراط در فایدروس افلاطون (۱۹۸۲، ۲۷۵e) شناسایی شده بود: «و هر کلمه، هنگامی که نوشته می‌شود، بین کسانی که آن را می‌فهمند و کسانی که به آن علاقه‌ای ندارند، دست به دست می‌شود و نمی‌داند با چه کسی صحبت کند یا نکند.»[۱]

بنابراین پرسیدن «دریدا چه می‌گفت؟» به یک معنا، بازنویسی همان متافیزیک حضور است که واسازی کلام‌محوری در پی واژگون کردن و جابه‌جا کردن آن است. ما نمی‌توانیم برای تأیید کاربرد خود با نویسنده (مرده) مشورت کنیم؛ و همچنین نباید وانمود کنیم که هر متنی، از جمله متن دریدا، می‌تواند یک بار برای همیشه تثبیت شود. مشارکت در دریدا در زمینه LLMها به

۱. دریدا مستقیماً به این صحنه از نوشتار و گفتگوی نوشتاری که در آن ظاهر می‌شود، در مقاله «داروخانه افلاطون» (دریدا ۱۹۸۱ب) می‌پردازد.

معنای نقض یا خدشه به تمامیت اثر او نیست، بلکه مشارکت در منطق اجرایی آن است: گسترش بازی تفاوت به حوزه‌ای جدید، حوزه‌ای که شرایط دلالت آن، این گسترش را نه تنها ممکن، بلکه ضروری می‌سازد.

در واقع، امتناع از چنین جنبشی به دلیل ترس از تفسیر نادرست یا بازنمایی نادرست، خیانت به نیروی نوآوری‌های دریدا خواهد بود، نوآوری‌هایی که ما را به بررسی مرزهای بین فلسفه و فناوری، انسان و ماشین، گفتار و نوشتار فرا می‌خواند. خطر سوء تعبیر، حادثه‌ای ناشی از ساختارشکنی نیست؛ بلکه شرط آن است. همانطور که دریدا ممکن است بگوید: هیچ خوانش مسئولانه‌ای وجود ندارد که این بی‌مسئولیتی را به خطر نیندازد.

محدودیت‌ها و جهت‌گیری‌های آینده

هدف این مقاله، آغاز گفتگویی میان‌رشته‌ای بین نوآوری‌های اخیر در هوش مصنوعی مولد و نظریه پساساختارگرایی اواخر قرن بیستم بود. چنین رویکردی پتانسیل آن را دارد که شیوه‌های تفکر موجود را دگرگون کند، بینش جدیدی ایجاد کند و فرصت‌های بدیعی را آشکار سازد. با این حال، از آنجایی که این تنها گام اول است، چندین مسئله شناسایی شده در جریان تحقیقات به طور کامل قابل بررسی نیستند و بنابراین نیاز به توجه بیشتر دارند. در این بخش پایانی، به طور خاص به سه مورد اشاره می‌کنم:[۱]

ساختارگرایی در مقابل پساساختارگرایی

یکی از محدودیت‌های مقاله حاضر این است که نمی‌تواند به طور کامل به بحث پیچیده و مداوم در نشانه‌شناسی ساختارگرا و پساساختارگرا در مورد «قرائت تجدیدنظرطلبانه» دریدا از سوسور، به ویژه در دوره زبان‌شناسی عمومی (هریس ۲۰۰۳) بپردازد. این موضوع که از قبل تاریخچه مفصلی دارد، اخیراً توسط لیف وِدربی در کتاب «ماشین‌های زبان» مورد بررسی مجدد قرار گرفته است، که استدلال می‌کند ساختارگرایی سوسور چارچوب نظری قوی‌تری برای توضیح و تفسیر مولدهای زبان معاصر ارائه می‌دهد. همانطور که وِدربی (۲۰۲۵، ۷۳) می‌نویسد، «به سادگی مشخص نیست که ما برای پیشبرد تحلیل ملموس زبان محاسباتی به بازنگری دریدا از ساختارگرایی نیاز داشته باشیم» و هدف کتاب او «بیان این موضوع است که پیاده‌سازی مولدهای زبان معاصر با نظریه زبانی که ساختارگرایی اروپایی تقریباً یک قرن پیش مطرح کرد، مطابقت دارد». نیازی به دریدا، دیفرانس یا واسازی نیست.

پرداختن به این بحث و ادای حق مطلب در مورد آن، مستلزم تعامل پایدارتری با نشانه‌شناسی سوسوری و دریدایی است، بیش از آنچه در اینجا امکان‌پذیر است. با این حال، اصطلاحات آن بحث اکنون به روشنی ترسیم شده‌اند. وِدربی (۲۰۲۵، ۴۳) با ساختارگرایی موافق است و نتیجه می‌گیرد که «پساساختارگرایی، که اغلب حرف‌های تشخیصی مفیدی در مورد فرهنگ معاصر دارد، ابزار تحلیلی به اندازه کافی مشخصی برای مقابله با تولید خودکار زبان ارائه نکرده است.» این مقاله، استدلال متقابل را ارائه می‌دهد و نشان می‌دهد که چگونه ترکیب دریدا از سوسور، دریچه انتقادی اساسی‌تری برای درک عملکردهای فنی مدل بزرگ زبان فراهم می‌کند. این بحث در اینجا قابل حل نیست. اینجاست که آغاز می‌شود.

مخاطرات واسازی

دومین محدودیت بالقوه مربوط به پیامدهای فلسفی این تحقیق است. از آنجا که این مقاله بر قرار دادن (عدم)مفهوم تفاوط به عنوان چارچوبی برای درک عملیات فنی LLMها تمرکز کرده است، ممکن است برداشتی از کار دریدا حاصل شود که می‌تواند به طور بالقوه آنچه را که واقعاً در کار فلسفی او مطرح است، مبهم کند. به عبارت دیگر، تلاش برای نشان دادن ارتباط نظریه پساساختارگرایی با سیستم‌های هوش مصنوعی، ممکن است ناخواسته پیامدهای فلسفی عمیق‌تر ساختارشکنی را مبهم کند. این کار، همانطور که می‌گویند، خطر از دست دادن جنگل فلسفی به خاطر درختان را به همراه دارد.

---

۱. این سه مورد مستقیماً از بررسی‌های ناشناسِ ارسال اولیه گرفته شده‌اند. من این را به دو دلیل ذکر می‌کنم: ۱) برای ابراز قدردانی از داوران به خاطر بینش و کاری که برای خواندن و پاسخ به پیش‌نویس اولیه انجام داده‌اند. و ۲) برای به رسمیت شناختن اهمیت و قابلیت دوام فرآیند بررسی همتا در انتشارات دانشگاهی.

اگرچه دریدا فرصت پرداختن مستقیم به LLMها را نداشت، اما پرداختن او به «مسئله حیوان» (Derrida ۲۰۰۸) یک الگوی مفید را ارائه می‌دهد. هدف دریدا در تجزیه و تحلیل دوگانه انسان/حیوان، تعمیم امتیاز انحصاری انسانی به حیوانات نیست، بلکه بررسی این موضوع است که آیا انسان‌ها از ابتدا به طور مشروع چنین امتیازی را داشته‌اند یا خیر. با اعمال این موضوع به LLMها، بحث به طور کامل تغییر شکل می‌دهد. بنابراین، آنچه در نهایت مورد توجه است، صرفاً ردیابی چگونگی به چالش کشیدن ظرفیت‌های انسانی توسط ماشین‌ها نیست، بلکه تشخیص این است که چگونه این کاربردها این واقعیت را آشکار می‌کنند که این ظرفیت‌ها، زبان، خلاقیت و نویسندگی، هرگز به آن اندازه که ما فرض می‌کردیم یا به خودمان اطمینان می‌دادیم، مستقل، خودحاضر یا ایمن نبوده‌اند.

تحقیقات آینده می‌تواند با پرداختن مستقیم‌تر به نوشته‌های دریدا در مورد «فناوری اصیل» (بردلی ۲۰۱۱) و واسازی او از دوگانه‌های متافیزیکی ریشه‌دار انسان/ماشین، طبیعی/مصنوعی، گفتار/نوشتار و غیره، بر این بینش بنا نهاده و آن را گسترش دهد. این امر نه تنها شرح ظریف‌تری از پیامدهای فلسفی همسوسازی واسازی با فناوری‌های LLM ارائه می‌دهد، بلکه درک ما را از چگونگی افشای شرایط ساختاری زبان توسط این سیستم‌ها که واسازی مدت‌هاست آن را نظریه‌پردازی و به یک کلام، واسازی کرده است، نیز افزایش می‌دهد.

LLMها و مسئله‌ی معنا

محدودیت سوم که به محدودیت قبلی مرتبط است و مستقیماً از آن ناشی می‌شود، مربوط به پیامدهای ادعای اصلی مطرح شده در این مقاله است: اینکه LLMها صرفاً از کاربرد زبان انسان تقلید نمی‌کنند، بلکه منطق‌های زیربنایی عملکرد آن را آشکار می‌کنند. اگر این موضوع را جدی بگیریم، می‌توان گفت که خود کاربرد زبان انسان، حداقل تا حدی، می‌تواند بر اساس عملکرد این مدل‌های تولیدی قابل توضیح باشد.

این موضوع، مسیر مهمی را برای تحقیق باز می‌کند: LLMها چه چیزی را در مورد ساختارهای تفکر انسان و کاربرد زبان آشکار می‌کنند، اگر اصلاً چنین چیزی وجود داشته باشد؟ و این چگونه می‌تواند بر درک ما نه تنها از معنا، بلکه از خودمان نیز تأثیر بگذارد؟ یک پاسخ احتمالی، مطابق با استراتژی ساختارشکنی، مقاومت در برابر تقابل دوتایی است که بین «همه ما اکنون فقط طوطی‌های تصادفی هستیم» و «ما طوطی نیستیم، زیرا می‌دانیم کلمات چه معنایی دارند» قرار گرفته است. به جای جانبداری در این اختلاف، می‌توان اثر دریدا را به عنوان ترسیم مسیر سومی خواند که اصطلاحات متافیزیکی خود بحث را ساختارشکنی می‌کند.

اگرچه مقاله حاضر به این احتمال اشاره می‌کند، اما شرح کامل آن از محدوده پژوهش فعلی فراتر می‌رود. پیگیری جدی آن مستلزم تعامل جداگانه و مداوم نه تنها با آثار دریدا، بلکه با مباحث معاصر در فلسفه زبان، علوم شناختی و زبان‌شناسی محاسباتی است. به طور خلاصه، این مقاله صرفاً به دنبال پیوند دادن بین تفاوط و LLMها بود، اما معنای نهایی این امر برای تفکر انسان و کاربرد زبان، کاری متفاوت، هرچند مرتبط، است و کاری که حداقل فعلاً باید به تعویق انداخته شود.

نتیجه‌گیری

اگر différance، همانطور که دریدا (۱۹۸۲، ۳) اصرار دارد، «به معنای واقعی کلمه نه یک کلمه است و نه یک مفهوم»، پس طنین آن با مدل‌های بزرگ زبانی یک استعاره نیست، بلکه یک اصل ساختاردهنده است. آنچه LLMها از طریق معماری‌های پیش‌بینی‌کننده خود، وابستگی‌شان به بافت‌های متنی قبلی و ناتوانی‌شان در دسترسی و ارجاع به «مراجع تجسم‌یافته دنیای واقعی» (به نقل از بندر در ویل ۲۰۲۳) تحقق می‌بخشند، انحراف از معنا نیست، بلکه تحقق شرایط سازنده آن است. این مدل‌ها به معنای فلسفی کلاسیک «درک» نمی‌کنند؛ آنها از طریق تفاوت و تعویق عمل می‌کنند و خروجی‌های زبانی تولید می‌کنند که دقیقاً به این دلیل قابل فهم هستند که در یک سیستم تفاوت قرار دارند و عمل می‌کنند.

آنچه در این مقاله «موتورهای تفاضل» نامیده شده است، صرفاً از کاربرد زبان انسان تقلید نمی‌کنند؛ بلکه تفاوت و تعویقی را که همان دلالت است، نشان می‌دهند. این به آن معنا نیست که موتورهای تفاضل محدود (LLM) به عنوان عوامل واسازنده طراحی شده‌اند یا حتی طراحی شده‌اند، و همچنین به این معنا نیست که دریدا این توسعه تکنولوژیکی را پیش‌بینی کرده باشد، بلکه به این معناست که ظهور چنین سیستم‌هایی، بازی دلالتی را که واسازنده دهه‌ها پیش نظریه‌پردازی کرده بود، از نو آشکار می‌کند. اگر بخواهیم دقیق‌تر بگوییم، ظهور موتورهای تفاضل محدود، اهمیت تفاضل را خوانا، و آن مرکززدایی نگران‌کننده از معنایی را که واسازنده همواره بیان کرده است، مرئی و شاید به طرز ناخوشایندی ملموس می‌کند.

تشخیص این موضوع به معنای تقلیل فلسفه به فناوری یا تابع کردن نظریه پساساختارگرا به محاسبات نیست. بلکه به معنای اجازه دادن به نظریه است تا حرکت خودِ تفاوت را فراتر از بستر اصلی‌اش، در سراسر مرزهای رشته‌ای، و به درون انواع دیگر متون و زمینه‌ها دنبال کند. با انجام این کار، ما از دریدا سوءاستفاده نمی‌کنیم، بلکه به منطق تکرارپذیری که او در زبان قرار داد، پاسخ می‌دهیم؛ این واقعیت که معنا هرگز به طور کامل وجود ندارد، همیشه از قبل به تعویق افتاده است، و هیچ گفتمان انسانی یا ماشینی از بازی تفاوت‌ها مبرا نیست.

ملاحظات اخلاقی

مشارکت نویسندگان

مشارکت نویسنده در این مقاله به شکل زیر است:

نویسنده به تنهایی مسئول مفهوم‌پردازی، نگارش و بازبینی مقاله است.

تعارض منافع

بر اساس اظهارات نویسنده، این مقاله تعارض منافعی ندارد.

حامی مالی

بنابر اظهارات نویسنده این پژوهش هیچگونه حامی مالی ندارد.

سپاسگزاری

از تمامی مشارکت کنندگان در این پژوهش سپاسگزاری می شود.

References

Aristotle. (1938). *Categories. On Interpretation. Prior Analytics*. Translated by H. P. Cooke. Cambridge, MA: Harvard University Press.

Barthes, Roland. (1978). Death of the Author. In *Image, Music, Text*. Translated by Stephen Heath, 142–148. New York: Hill & Wang.

Baudrillard, Jean. (1983). *Simulations*. Translated by Paul Foss, Paul Patton, and Philip Beitchman. New York: Semiotext(e).

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 Conference on Fairness, Accountability, and Transparency*, March 3–10, 2021 (FAccT '21), Virtual Event, Canada, ACM. https://doi .org/10.1145/3442188.3445922.

Bogost, Ian. (2022). ChatGPT Is Dumber Than You Think. *The Atlantic.* https://www.theatlantic.com/technology/archive/2022/12/ chatgpt -openai-artificial-intelligence-writing-ethics/672386.

Bradley, Arthur. (2011). *Originary Technicity: The Theory of Technology from Marx to Derrida.* New York: Palgrave Macmillan

Coeckelbergh, Mark and David J. Gunkel. (2025). *Communicative AI: A Critical Introduction to Large Language Models.* Cambridge: Polity.

Derrida, Jacques. (1976). *Of Grammatology*. Translated by Gayatri Chakravorty Spivak. Baltimore,MD: The Johns Hopkins University Press.

Derrida, Jacques. (1978). *Writing and Difference*. Translated by Alan Bass. Chicago: University of Chicago Press.

Derrida, Jacques. (1981a). *Positions*. Translated by Alan Bass. Chicago: University of Chicago Press, 1981.

Derrida, Jacques. (1981b). *Dissemination*. Translated by Barbara Johnson. Chicago: University of Chicago Press.

Derrida, Jacques. (1982). *Margins of Philosophy*. Translated by Alan Bass. Chicago: University of Chicago Press.

Derrida, Jacques. (1993). *Limited Inc.* Evanston, IL: Northwestern University Press.

Derrida, Jacques. (2008). *The Animal That Therefore I Am*. Translated by David Wills. Edited by

Marie-Louise Mallet. New York: Fordham University Press.

Fares, Murhaf, Andrei Kutuzov, Stephan Oepen, and Erik Velldal. (2017). Word vectors, reuse, and replicability: Towards a community repository of large-text resources. In Jörg Tiedemann (ed.), *Proceedings of the 21st Nordic Conference on Computational Linguistics* (NoDaLiDa), 22–24 May 2017. Linköping University Electronic Press. https://doi.org/978-91-7685-601-7. (See also http://vectors.nlpl.eu/explore/embeddings/en.)

Gunkel, David J. (2021). *Deconstruction*. Cambridge, MA: MIT Press.

Harris, Roy. (2003). *Saussure and His Interpreters*. Edinburgh: University of Edinburgh Press.

Hicks, Michael Townsen, James Humphries, and Joe Slater. (2024). ChatGPT Is Bullshit. *Ethics and Information Technology*. https://doi .org/10.1007/s10676-024-09775-5.

Josephson-Storm, Jason A. (2017). *The Myth of Disenchantment: Magic, Modernity, and the* Birth of the Human Sciences*.* Chicago, IL: University of Chicago Press.

Kaplan, Jerry. (2025). *Generative Artificial Intelligence: What Everyone Needs to Know*. New York: Oxford University Press.

Plato. (1982). *Euthyphro, Apology, Crito, Phaedo, Phaedrus*. Translated by Harold North Fowler. Cambridge, MA: Harvard University Press.

Salmon, Peter. (2021). *An Event, Perhaps: A Biography of Jacques Derrida*. London: Verso.

Saussure, Ferdinand de. (1959). *Course in General Linguistics*. Translated by Wade Baskin. London: Peter Owen.

Searle, John. (1999). The Chinese Room. In R. A. Wilson and F. Keil (eds.), *The MIT Encyclopedia of the Cognitive Sciences*, 115–116. Cambridge, MA: MIT Press.

Turing, Alan. (1950). Computing Machinery and Intelligence. *Mind* 59 (236): 433–460. https://doi.org/10.1093/mind/LIX.236.433.

Weatherby, Leif. (2025). *Language Machines: Cultural AI and the End of Remainder Humanism*. Minneapolis, MN: University of Minnesota Press.

Weil, Elizabeth. (2023). You Are Not a Parrot: And a Chatbot Is Not a Human: And a Linguist Named Emily M. Bender Is Very Worried What Will Happen When We Forget This. *New York Magazine*, March 1. https://nymag.com/intelligencer/article/ai-artificial-intelli gence-chatbots-emily-m-bender.html.

Wittgenstein, Ludwig. (1995). *Tractatus Logico-Philosophicus*. Translated by D. F. Pears and B. F. McGuinness. New York: Routledge.