INVITED ARTICLE

The Concept of a Large Language Model: Misconceptions About and Beneficial Conceptualizations of LLM Technology

Matthew Peterson [a], **Colleen Patton** [b], **Michael Green** [c], **and Brayan Díaz** [d]

[a] Department of Graphic Design and Industrial Design, North Carolina State University, Raleigh, NC, USA; [b] Department of Psychology, North Carolina State University, Raleigh, NC, USA; [c] Laboratory for Analytic Sciences, North Carolina State University, Raleigh, NC, USA; [d] School of Teacher Education and Leadership, Utah State University, Logan, Utah, USA

Corresponding author: Matthew Peterson (mopeters[at]ncsu.edu)

Abstract: Large language models (LLMs) are transforming writing and planning practices across the breadth of knowledge work, where misconceptions about the technology abound. We contend that these misconceptions are rooted in the dispersal of disciplinary knowledge—that multiple disciplines are needed to understand LLMs. We bring together disciplinary knowledge to address four constructs commonly applied to LLMs: writing, learning, hallucination, and thinking. Misconceptions about LLM technology are embodied in and transmitted through conceptual metaphors and analogies. We thus provide a basic description of LLM technology with a careful accounting for metaphor. This description and the discussion of the four constructs leads to corrected statements for understanding LLMs. For instance, the assertion that "LLMs think" is replaced with: "LLMs produce outputs that resemble the products of thinking." Finally, we consider common and potential metaphors for more accurately conceptualizing LLMs, and favor LLMs as answer simulators, remixers, fuzzy thesauruses, and plagiarism traps. The statements and metaphors provided here are a starting point for well-grounded usage patterns with LLMs for professionals, researchers, and students alike.

Implications for practice: The descriptions of writing, learning, hallucination, and thinking constructs (Section 3) combine with the basic description of LLM technology (Section 2) to give practitioners a reflective space for reconsidering LLMs. Practitioners can then refer to the provided statements (Section 3, summarized in Table 1) and metaphors (Section 4) to positively impact internal corporate and institutional communications about LLMs. A more careful and consistent

@: ISSUE〉 ARTICLE〉

Cite this article:

Peterson, M., Patton, C., Green, M., & Díaz, B. (2026). The concept of a large language model: Misconceptions about and beneficial conceptualizations of LLM technology. *Visible Language, 60*(2), 210–239.

First published online April 26, 2026.

https://visible-language.org/journal/

Visible Language Consortium:
University of Leeds (UK)
University of Cincinnati (USA)
North Carolina State University (USA)

use of language can serve to informally educate collaborators, managers, and decision makers on the LLM technology that they are actively employing.

Implications for research: Researcher-provided analogies have been shown to impact how people reason through concepts and models (see Section 1). The statements and metaphors about LLMs provided here (Sections 3.6 and 4) could likewise be tested for their impact on how people plan to use LLMs, what people expect to happen with such use, and how people interpret resultant LLM outputs. It would also be beneficial to systematically study how (and how frequently) constructs such as writing, learning, hallucination, and thinking are invoked in communication about LLMs—by companies, users, and researchers.

Keywords: analogical reasoning; conceptual metaphor theory; large language models; LLM hallucination; metaphor; multidisciplinarity; semiotics; writing

1. Introduction

Large language models (LLMs) such as ChatGPT, Claude, and Gemini have pervaded knowledge workers' work and students' studies. Knowledge workers are being pressured into incorporating artificial intelligence (AI) into their workflows, if they are not already enthusiastically doing so of their own volition. Students are being prodded by software to utilize AI (e.g., a Gemini-powered "help me write" prompt that appears on the fourth blink of the cursor in every new paragraph in Google Docs). School teachers and college instructors are receiving LLM-written papers for their assignments, and academic journal editors are fielding LLM-written submissions and LLM-written peer reviews.

Whether prodded or not, people are making *decisions* about when and how to use LLM technology. These decisions are informed by *something,* even when they are not entirely conscious decisions.

We contend that a major factor behind decisions on when and how to use LLMs is the user's fundamental conceptualization of LLM technology, and furthermore that LLM use is in fact too commonly informed by misconceptions of the technology's nature. LLMs have at least superficially passed the Turing test (Gibney, 2025), meaning they can imitate text-based human communication well enough to appear human. The influence of this imitation on users is compounded by the language humans utilize when describing the technology. For instance, according to IBM's knowledge base on LLM technology (Stryker, n.d.), LLMs are "capable of understanding" and "AI agents don't just think."

If LLMs do not in fact understand or think, then these are examples of metaphor or analogy. The literature reveals that they constrain and even constitute our understanding

of reality. Conceptual metaphor theory (Lakoff & Johnson, 1980) posits that much of human thought is guided by conceptual metaphors, which utilize existing concepts like *building* and embodied experiences like a flutter in the chest to constitute more abstract concepts like *theory* and *love,* respectively. Underlying the more basic concepts and organizing embodied experiences are image schemas (Johnson, 1987; or conceptual primitives, Mandler, 1992) like the notion of a *path.* These image schemas are necessary building blocks for more structured conceptual metaphors. Thus, theories can be understood in relation to buildings—"We need to *buttress* the theory with *solid* arguments" (Lakoff & Johnson, 1980, p. 46)—and love is a journey, predicated on a *path* that partners *travel* together (Lakoff & Johnson, 1999, pp. 63–73).

Analogical reasoning similarly utilizes conceptual mapping to understand one concept in terms of another (Gentner, 1983, 2003), and analogy involves similarity of relations more than attributes (Gentner & Markman, 1997, p. 48). For instance, electronic circuitry, and thus electricity as a concept, can be conceptualized according to the structure of crowds of people moving through passageways, or by water flowing through plumbing—with electrons as people or water molecules. Gentner and Gentner (1983) showed how using one of these analogies over the other changed expectations about the nature of electricity, and being incomplete as all analogies must be, that different errors in understanding electricity result from the analogies. Ultimately, whether conceptual metaphor or analogical reasoning, conceptualizing LLMs according to constructs from other domains of understanding (e.g., thinking in humans) necessarily determines and delimits our relationship to the technology—importantly in terms of decisions on when and how to use it.

If it is necessary to comprehend LLM technology through something like conceptual metaphor or analogical reasoning, what expertise can be leveraged to derive reasonably accurate metaphors and analogies? LLM technology's nature is bound not only to the processes by which its developers created it, but also to nuances of thought, language, and reality. We assume that an ideal conceptualization of LLM technology would rely on synthesizing disparate areas of expertise such as computer science, linguistics, psychology, the learning sciences, semiotics, and philosophy, at the very least.

Imagine an LLM developer who knows the technology well, but only has a folk understanding of thinking, and they ascribe thinking to LLMs. And imagine a cognitive psychologist who has a technical understanding of thinking, but does not appreciate how LLMs work, and they too ascribe thinking to LLMs. Yet these two individuals could come together to share knowledge and collectively determine that LLMs do not think. Perhaps only coordinated multidisciplinary expertise can hope to apprehend the reality of LLMs.

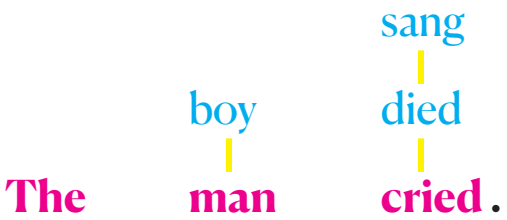
To that end, we utilize multiple areas of academic expertise—embodied in us, the authors—to describe LLM technology and interpret some constructs that are frequently used to conceptualize it. We identify misconceptions according to these areas of expertise and ultimately offer alternative conceptualizations—including conceptual metaphors—that may lead to better decision making with LLMs. The goal is to synthesize knowledge from relevant fields to provide a pragmatically more accurate or helpful conceptualization of LLMs.

2. A Basic Description of LLM Technology

The following description of LLM technology is necessarily written at a high level of abstraction, especially because we are minimizing use of metaphors. Its abstraction is a demonstration of why addressing the nature of LLMs is so frequently, even necessarily, dependent upon metaphor.

Model architecture. The large language model is a type of machine learning model called a *neural network,* with an architecture somewhat analogous to the neurons in a human brain. In the brain, each neuron is connected to multiple other neurons via synapses. Based on the strength of signals provided to it by these connections, each neuron produces its own signal, which is distributed to other neurons in turn. Already inescapable in the basic categorical designation of LLM technology is a brain–algorithm

Introduction to Boxes 1–4. In semiotic structural analysis, meaning is substantially derived from differences among signs (including words), and the paradigmatic axis (expressed in vertical chains below, and in all boxes) is a range of signs that could substitute for a chosen one; even "apparent synonymous signifiers" have significant differences (Chandler, 2021). In effect, at some level a chosen sign implies its alternatives, and the fact that those alternatives were not used impacts interpretation.



Box 1. A simple demonstration of the paradigmatic axis, as shown in Chandler (2021). The words in cyan are possible substitutions for the words in magenta.

analogy, which uses the neuron to characterize what might be neutrally described as a *node*. The correlational relationships among an LLM's nodes are produced mathematically, through several layers of interconnected mathematical functions that take inputs from previous layers of similar functions, modulate and combine them to some degree, and produce outputs to be fed to subsequent layers. The final layer of this network produces results that can be matched to human-comprehensible tasks, such as a yes/no decision or a choice among possible words in a vocabulary.

Model training. The strength of the signals in the connections among nodes in these mathematical representations is not explicitly programmed. The signal strength of node connections is determined by processing training data and through trial and error. The way an LLM incorporates information into a network of node connections is threefold, as shown in Figure 1 (we avoid characterizing this incorporation as *learning* here, though that term is commonly used).

In *pretraining,* an LLM is first trained on an enormous text corpus that may contain trillions of tokens. Tokens are the words or subword units that represent the compositional building blocks of language in current machine learning applications. These tokens are presented to a model as sequences of varying length and as they occur in the source texts. Using a process called *causal self-attention,* a model attempts to predict the next token in the given sequence. In this process, the model is prevented from seeing future tokens and conditionally determines the most significant earlier tokens in the sequence to focus on in its calculations. These next-token predictions are computed across many positions in parallel, and the predicted tokens are compared with the actual next tokens from the text with the differences contributing to the accumulated error or loss over the batch. Based on the aggregated loss, the numerical parameters or weights associated with each node in the network are adjusted a small amount at a time, proceeding one layer at a time in reverse order from model output back to input, through backpropagation. Training proceeds over many iterations as described, each cycle resulting in incremental adjustments to the model. Eventually the resulting

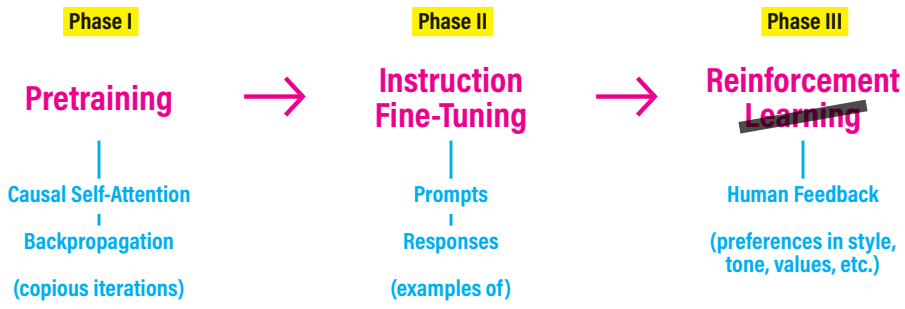


Figure 1. Training phases for large language models.

weights define a function that produces an increasingly accurate expression of the likelihood of each token in a model's vocabulary being the next token in the text. After a stopping criterion is met, the phase is complete and the model represents statistical and structural patterns of text in the training corpus. These patterns potentially generalize beyond the exact examples seen in the training data. However, we emphasize that these patterns may not match with reality or the world outside the training context.

The remaining training phases make the LLM more useful for humans. The second phase is *instruction fine-tuning,* in which examples of questions and answers—or prompts and responses—are provided. These examples are intended to approximate likely user requests. Again, node connection weights are revised through next-token prediction. The third and final phase is *reinforcement* (usually called reinforcement *learning*). This phase is a human feedback process that focuses on human preferences in style, tone, values, and other qualities (e.g., politeness, confidence). The sequence example and token counts in the latter two phases are much lower than in pretraining, but they are high nonetheless.

Output variability. Following training, an LLM outputs tokens into word sequences in a manner that, being based on probabilistic selection among token options and not deterministic pathways, is variable—results are not precisely repeatable. Several strategies have emerged to constrain this variability for situations that were not adequately represented in training. These include prompt engineering, expanded context windows, "reasoning models," and systems of AI agents, among others. In such extended articulation, more deterministic results can be achieved, pairing the LLM's pattern matching capabilities with reusable functions, verifiable data sources, or other resources.

3. Problematic Characterizations of LLM Technology

The above description of LLM technology was carefully written to appear dry and to preempt common misconceptions. Due to the terminology regularly applied to LLMs, we still had to invoke the constructs of *neuron* and *learning*. We now address the latter and other constructs that underlie common problematic conceptions of LLM technology, and we consider the entailments that necessarily follow from common but misleading descriptions. If we unpack and reject individual misconceptions, a healthier general concept of LLM technology is possible, one that can lead to better human decision making in using and choosing to use LLMs.

3.1. Writing

When a human user submits a prompt in an LLM chat interface, there is a brief pause before sequences of words are produced from start to finish, resulting in a coherent explanatory narrative. This observable production looks like writing in the folk sense,

with two obvious differences. First and most obvious, the speed at which the narrative is produced is beyond human capability. Second, at least with many LLMs, words and sentences are produced sequentially, with no regressions to revise earlier passages along the way. In contrast, an outside observer to the human writing of the previous sentence would have seen "*make edits* along the way" change to "*revise earlier passages* along the way," with the latter half retained while the former half is reconsidered. These two LLM–human differences matter because LLM speed promises extraordinary efficiency, and because the impression can be that the technology is superhuman in its capability. But to be superhuman suggests more overlap with "human" than perhaps can reasonably be said to exist, bestowing a kind of authority and even familiarity to LLMs—as personification.

There is a less obvious difference between human and LLM text production. The folk concept of writing collapses writing-to-communicate and writing-to-learn into a single act, if it does not ignore the latter entirely in favor of writing as an unproblematic form of communication. "Let me get my thoughts down on paper" sounds sensible to many, but not to experts on composition. In their cognitive process theory of writing, Flower and Hayes (1981) describe writing as an elaborate, nonlinear process of goal setting and goal satisfaction. The intentional process of writing reveals the writer's ideas as dynamic, not pre-formed:

> Writers create their own goals in two key ways: by generating both high-level goals and supporting sub-goals which embody the writer's *developing* sense of purpose, and then, at times, by *changing* major goals or even *establishing entirely new ones* based on what has been *learned* in the act of writing. (p. 366, emphasis added)

To the writer, composition "often seems a serendipitous experience," as in, "'I don't know what I mean until I see what I say'" (p. 377). But this serendipity is in fact constituted of *process goals,* or personal instructions for writing of the moment, and *content goals,* or what the writer intends to convey. Especially interesting is that Flower and Hayes note that much of what has been learned about how writers compose text could not have been ascertained by researchers through retrospective measures, because "people rapidly forget many of their own local working goals once those goals have been satisfied" (p. 377). This reality in part suggests why an inaccurate folk sense of writing may have formed—the subjective experience of writing is incomplete or forgotten.

Emig (1977) and Flower and Hayes (1981) establish that writing has an epistemic function. Because the experience of writing is personally beneficial, students and experienced writers face asymmetric consequences in LLM usage. The ability to use LLMs critically requires metacognitive awareness and writing process fluency that is not evenly distributed among writers with varying experience and expertise. Writers

who already have strong process fluency are more capable of integrating LLMs as genuine tools; those who do not are most at risk of substitution patterns. So there is an ethical dimension in the discrepancy between writing in the normal sense and the LLM production of texts. For the new writer, the more an LLM speaks for them, the less they can develop their own agency. They are subject to a form of cognitive surrender that neutralizes intuition and deliberation (Shaw & Nave, 2026).

It is thus misleading to consider the LLM as a co-author. A pair of human co-authors are engaged in their own personal and dynamic goal setting and satisfying behaviors (Flower & Hayes, 1981), which leads to individual learning that refines intentional meaning as embodied in the text. That is, each of the pair contributes the full spectrum of what writing entails. In human–machine teaming of one human author engaging an LLM through prompting, only the human can contribute the full spectrum. But in practice engaging with an LLM can sidestep the learning benefits of writing for the human operator. Thus LLMs are less co-authors than a means to bypass the personal benefits of learning through writing when producing text, which disproportionately impacts inexperienced writers. If the learning process constitutive of full-spectrum writing improves the thoughts that become text, then the speed benefits of LLMs are necessarily paired with quality limitations.

The quality limitations of LLMs extend beyond substitution of a human author's otherwise deep engagement through writing, to the probabilistic nature of next-token prediction. Because the token-to-token weights that determine an LLM's probabilities are derived from commonalities in a massive corpus of training data, its outputs are necessarily derivative in the unflattering sense that it is prone to repetition at the expense of novelty. To inexperienced writers such as students, an LLM's outputs may appear polished and novel, while an experienced writer may find the same writing to be unsophisticated, incoherent, and overly familiar. The student cannot perceive the patterns that their instructor can. And problematically, the lack of novelty—repetition derived from the dataset—can easily slide into unwitting plagiarism. This plagiarism is unwitting to the human operator who does not recognize the sources, and unwitting to the LLM simply because it has no inner life, and thus no wits. Instead of creating unique works—even if it seems this way to users—LLMs generate derivative works. This is a reversal from *creativity* to *replication*.

At a more fundamental level, it is somewhat misleading to make the simple claim that LLMs write texts. There is a scope asymmetry between what happens when a human writes on one hand, and next-token prediction on the other. Though less elegant, it is more accurate to suggest that LLMs enable the production of coherent text in the absence of writing. This is a reversal from *ignoring the limitations of an invoked sense of writing* to *emphasizing the absence of crucial facets of writing*.

Box 2. An absurdist demonstration of the paradigmatic axis (from semiotics; see Mickus, 2024, for its application to LLMs). The words in cyan are possible substitutions for the words in magenta. Boxes 2–4 are human-created, accomplished by following associations through the Oxford American Writer's Thesaurus. Vertically adjacent words have direct connections in the thesaurus. The first column (i.e., people, LLMs) is an exception; it was created in the spirit of the paradigmatic axis without thesaurus connections.

3.2. Learning

Across popular, technical, and even academic discussions, the language of learning is frequently applied to AI systems without sufficient conceptual clarification or critical examination. As a result, fundamentally different processes are often treated as equivalent. Two common uses of *learning* illustrate this problem.

First, the LLM training process is routinely described through educational metaphors. Authors refer to "teaching" a model or having a model "learn" from examples (Patil & Gudivada, 2024). Such language can be misleading because it conflates machine training with human learning.

Human learning does not occur through passive exposure to vast quantities of information. Rather, educational research has consistently demonstrated that learning emerges through active engagement, experimentation, reflection, social interaction, and meaning-making. As Díaz and Delgado (2024) argue:

> Should we educate students as we train ChatGPT, having them read all of Wikipedia plus thousands of books and magazines (Johri, 2023; Ouyang et al., 2022)? Or should we implement pedagogical practices that promote students' problem solving, critical thinking, and teamwork skills instead of memorization? We have known for decades that the latter approach is preferable. (p. 2576)

The danger of the learning metaphor is not merely semantic. If machine training becomes an implicit model for human learning, educational practice risks being reduced to information transmission, positioning students as passive recipients of content rather than active participants in knowledge construction.

A second and even more widespread source of confusion arises from the term *machine learning* itself. Within computer science, machine learning refers to computational methods that enable systems to identify patterns in data and improve performance on specific tasks without explicit reprogramming (Bell, 2022). While this definition is technically appropriate within the field, describing the process as learning raises an important conceptual question: Is statistical optimization equivalent to learning as understood in educational, psychological, and social theories? Through training, LLMs improve their ability to predict linguistic patterns and generate increasingly sophisticated outputs. This happens in the absence of cognitive characteristics such as understanding, experience, intentionality, and meaning-making.

Educational theories have proposed diverse definitions of learning. Within behaviorism, learning is typically defined as a change in observable behavior resulting from interactions between environmental stimuli and responses (Watson, 2017). Learning is demonstrated through measurable behavioral changes rather than internal cognitive

processes. Constructivist perspectives, particularly those associated with Piaget, shifted attention toward the learner's active role in constructing knowledge (Duckworth, 1965). From this perspective, learning occurs through processes of assimilation and *accommodation* (i.e., updating personal schemas when faced with conflicting information), whereby individuals reorganize existing cognitive structures in response to new experiences. Cognitive theories further emphasized internal mental processes involved in acquiring, organizing, storing, and retrieving information (Bruner, 1966). Although traditional cognitivism is often criticized for conceptualizing learners as information processors, it nevertheless assumes active mental engagement rather than information accumulation.

Moving beyond individualistic perspectives, social constructivism emphasized the fundamentally social nature of learning. Vygotsky (1978) argued that higher psychological processes emerge through social interaction and cultural mediation. His concept of the *zone of proximal development* highlighted how learning occurs through guided participation with more knowledgeable others. Through this lens, individuals learn from a scaffolded learning process. The emphasis on participation was further developed through situated learning theory (Lave & Wenger, 1991), which conceptualized learning as inseparable from the contexts in which it occurs. Wenger's (1998) communities of practice framework extended this argument by defining learning as increasing participation within social communities. Importantly, Wenger explicitly distinguished human participation from machine processes: "Participation is an active process, but I will reserve the term for actors who are members of social communities. For instance, I will not say that a computer 'participates' in a community of practice…" (Wenger, 1998, p. 56).

This distinction remains particularly relevant in contemporary discussions of AI. More recent perspectives have further expanded the concept of learning beyond cognition alone. Drawing on Maturana's work on *autopoiesis* (i.e., on self-maintaining systems; Maturana, 1975; Maturana & Varela, 1991), scholars have emphasized the biological, emotional, ecological, and relational dimensions of learning (Bekerman & Zembylas, 2018; Nussbaum et al., 2026). Under this biological theory approach, learning is understood as a property of living, self-organizing systems that continuously adapt through interactions with their environments. Learning is therefore inseparable from biological embodiment and lived experience.

Despite substantial theoretical differences, behaviorist, constructivist, cognitive, sociocultural, situated, and biological perspectives converge on several characteristics of human learning. Learning is understood as experiential, contextually situated, socially mediated, embodied, and grounded in lived participation. By contrast, the training of LLMs consists of statistical optimization across large datasets. Machine

Box 3. An absurdist demonstration of the paradigmatic axis. The words in cyan are possible substitutions for the words in magenta. Boxes 2–4 are human-created, accomplished by following associations through the Oxford American Writer's Thesaurus. Vertically adjacent words have direct connections in the thesaurus. The first column (i.e., chickens, horses, people, LLMs) is an exception; it was created in the spirit of the paradigmatic axis without thesaurus connections.

	prompt			
	egg on			
	arouse			
	animate	horse sense		research
Chickens	fortify	insights		learning
Horses	consolidate	inspiration		scholarship
People	**integrate**	**stimuli**	**as**	**knowledge.**
LLMs	assimilate	prompts		expertise
	digest			command
	dissolve			resolution
	disintegrate			intention
				design

learning systems identify patterns, adjust parameters, and improve predictive performance. This distinction does not diminish the remarkable capabilities of LLMs. Rather, it clarifies the ontological differences between computational optimization and human learning. Failing to recognize these differences can lead researchers, educators, and policymakers to overestimate what LLMs are doing and misunderstand how humans learn. Learning cannot be reduced to a process of parameter optimization or information accumulation.

LLMs internalize, in a fractured manner, a massive amount of information. But while LLMs acquire information, humans construct knowledge. We thus emphasize that LLMs are trained on information. (We would say "human-created information," but LLMs' datasets will increasingly include outputs from earlier LLMs.) Unlike learning, the term training does not to us suggest much about who or what is being trained. LLMs are trained on information while humans learn and thus construct knowledge. This is a reversal from *implying internal qualities of LLMs when referring to their development* to *emphasizing their artificiality as information stores.*

3.3. Hallucination

LLMs are frequently said to hallucinate when their outputs appear to humans as confident despite failing to reflect reality. Unlike the above construct of learning, the folk sense of hallucination is at least relatively simple, and for many people hallucinations are never experienced. Aleman and Larøi (2008) describe their model of hallucination as presuming:

> that a sensory experience (either perception due to *external input* or hallucination due to *internal input*) can be induced by activation of sensory cortical areas (either primary or secondary). In normal perception, signals emanating from the sensory organs play a decisive role, whereas in imagery and hallucinations, internally generated signals (emanating from cortical centers) play a decisive role in determining the final "percept." (unpaginated ebook version, emphasis added)

LLMs of course have no sensory system, and their "inputs" may be considered to all be internal. As internal inputs are associated with hallucination, it is tempting to simply state that everything LLMs do is a matter of hallucination, though many of these hallucinations end up adequately representing ground truths as humans accept them.

The American Psychological Association's *Dictionary of Psychology* defines visual hallucination as:

> visual perception in the absence of any external stimulus. Visual hallucinations may be unformed (e.g., shapes, colors) or complex (e.g., figures, faces, scenes). In hallucination associated with psychoses (e.g., paranoid schizophrenia, alcohol- or hallucinogen-induced psychotic disorder), *the individual is unaware of the unreality of the perception,* whereas insight is retained in other conditions (e.g., pathological states of the visual system)... (APA, 2018, emphasis added)

If LLMs have no direct connection to reality, then certainly they cannot, in any way, be *aware* of unrealities. They are thus entirely disconnected from any ground truth. Instead, they are connected to what Baudrillard (1981/1994) called *hyperreality,* a self-contained

simulation as a world in itself. Following Wittgenstein's (1921/1922) description of the proposition (in language) as "a model of the reality as we think it is" (p. 39), Baudrillard (1981/1994) stated, long before the advent of LLMs, that:

> Today abstraction is no longer that of the map, the double, the mirror, or the concept. Simulation is no longer that of a territory, a referential being, or a substance. It is the generation by models of a real without the origin or reality: a hyperreal. The territory no longer precedes the map, nor does it survive it. (p. 1)

Baudrillard's ideas about media are all the more relevant to LLMs and their outputs. For LLMs, there is no separable normal experience to compare with hallucinatory experience, and so the basic implication of hallucination as a deviation is nonsensical. What are typically called LLM hallucinations are simply probabilistic results from next-token prediction.

Stating that LLMs occasionally hallucinate carries with it the unfortunate suggestion that they regularly perceive reality, which in turn implies that they can (unless something goes wrong) report on reality. It would be far less misleading to suggest instead that LLMs probabilistically produce outputs that are frequently—that often happen to be—reflective of reality. This is a reversal from *emphasizing errors as a performance deviation* to *emphasizing accuracy as a performance probability.*

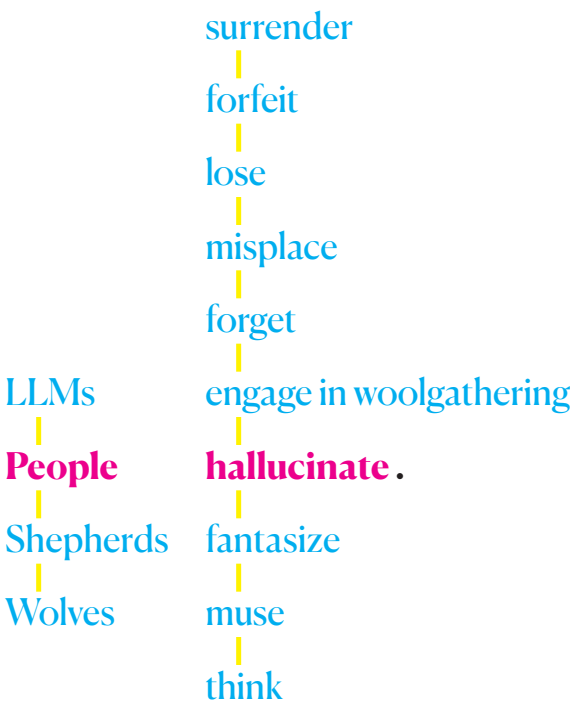
In a negative sense, LLMs do not have access to reality and thus do not connect human users to reality. They act on internal representations derived from a dataset. In a positive sense, LLMs provide humans in reality access to a fragmented hyperreality constituted of a sizable portion of the internet, though it includes conflicting statements, outright falsehoods, and fictional accounts. (This hyperreality is fragmented because it is constituted of probability distributions rather than regular form.) This is a reversal from *epistemic authority* to *epistemic instability.*

3.4. Thinking

As noted above, writing is a process that forms or crystallizes thoughts. We have experienced this in preparing this section on thinking. The result of our writing process here is tentativeness, which leads to an emphasis on pragmatism (why ascribing thinking to LLMs matters in practice) over ontology (what LLMs are, what thinking is) or epistemology (what we can know about LLMs, how we talk about thinking).

We will begin with a brief sample discussion of aspects of thinking and correspondence with LLM technology. We will then note—and not comprehensively—how thinking entails a massive scope of theoretical constructs in and beyond psychology. This will lead us, in fact, to abandon the exercise and conclude that the richness of the thinking construct not only makes our unpacking of it impractical, but that this richness is

Box 4. An absurdist demonstration of the paradigmatic axis. The words in cyan are possible substitutions for the words in magenta. Boxes 2–4 are human-created, accomplished by following associations through the Oxford American Writer's Thesaurus. Vertically adjacent words have direct connections in the thesaurus. The first column (i.e., LLMs, people, shepherds, wolves) is an exception; it was created in the spirit of the paradigmatic axis without thesaurus connections.



entirely the problem of ascribing thinking to LLMs: it cannot possibly be done responsibly. This, ultimately, is the crux of our general argument, that characterizations of LLM technology must be carefully formed, and some common characterizations cannot be vetted in the short term for a black-box technology. A conservative approach suggests favoring characterizations that can be given with confidence, that hedge bets.

Thinking is not as clearly delineated a term as is *cognition,* but cognition is at least a good starting point. One aspect of cognition is decision making. As an LLM produces word after word in sequences that cohere into an explanatory narrative, decisions are being made (i.e., next words are being "predicted"). For LLMs, this is a matter of probabilities derived from the training process. Is decision making so different in humans?

Much of the way in which people make decisions is through pattern matching. Recognition-primed decision making (Klein et al., 1989) suggests that humans decide on an action by matching cues from the current situation to past situations. Experts can be defined as such, in part, because they have such a breadth of prior patterns to match that they often make good decisions. LLMs exhibit exemplary pattern matching from and within the bounds of a huge set of patterns, frequently resulting in good decisions on what text to output. This idea can be transferred from decision making broadly into writing more specifically—often, good writers have enough experience to recognize where grammatical formatting belongs, or ways to structure an argument. LLMs do not tire of writing or typing, they can hold and compare amounts of information that we cannot fathom, and they can produce outputs—that is, continue to make decisions on what words to produce—without ever hitting writer's block.

Human pattern matching in decision making is influenced by biases such as the saliency of the information (e.g., the emotion attached to it) or recency (e.g., how recently the information was encountered). These biases can make people hold onto already written work, recall certain pieces of information that may not be closely matched to the current idea, or have emotions evoked from a piece of writing—none of which are comparable to an LLM's pattern matching. Humans acquire their patterns by being in the world, connected to it through sensory systems. They encode information in rich mental imagery and in linguistic propositions (Kosslyn et al., 2006), and process it in dedicated resources in working memory (the visuospatial sketchpad and phonological loop, respectively; Baddeley, 2018). Affect impacts much processing, with immediate feelings and personal values influencing interpretation and response. Decisions are understood through a conceptual system that uses conceptual metaphors (Lakoff & Johnson, 1980) to make the world meaningful, and these are based on formative and embodied experiences that provide basic units of meaning (Johnson, 1987; Shapiro, 2019; Wilson, 2002). Decisions are necessarily influenced by past experiences of living in a social world.

We quickly reach a point in considering decision making, itself only part of cognition, and that only part of what thinking entails, where we accumulate so many considerations that we must simply state the following:

Thinking is a rich capability of humans that is incredibly complex in its mechanisms and its source material. Most of what constitutes thinking in human beings has no parallel in LLMs, but creating a list enumerating the differences is impractical. Instead we are left with a simpler question: Do discrepancies between what humans do and what LLMs do suggest that LLMs cannot think, or that our definition of thinking is too narrow? Chen et al. (2026) address this question, or something close to it. They suggest that LLMs have general intelligence, and argue that differences between humans

and LLMs should not disqualify LLMs. It is in fact anthropocentric bias that leads to disqualification of LLMs, and "by reasonable standards, including Turing's own, we have artificial systems that are generally intelligent" (p. 36).

We the authors are overwhelmed. Instead of making a scientific statement about the nature of thought, in which we could have little confidence, we are left with a more designerly response. *Pragmatically,* whether we ultimately describe what LLMs do as thinking or not, we presently believe that invoking *thinking* for LLMs is problematic.

In their structure mapping of similarity and analogy, Gentner and Markman (1997) use simple node diagrams to represent correspondence between two conceptual domains. Figure 2 is an example of pure correspondence. The nodes could represent any number of things. For instance, in an analogy between an atom (Domain A) and the solar system (Domain B), some nodes might be: (1) a large central element, nucleus or sun; (2) a small secondary element, electron or planet; (3) the large element restricting the small element's movements; and (4) the large element being relatively fixed within the system (at which point at least the atom half of the analogy is becoming somewhat misleading).

Unlike the atom and solar system example, what LLMs do as a process is much simpler than the embodied, experiential, and social processes that underlie human thought. As shown in Figure 3, this can be visualized as a profound discrepancy in complexity of connections. The blurred magenta nodes on the LLM side of the diagram represent aspects of human thought that are not present in what we might call an LLM's "thinking," but that might result in "candidate inferences" (Gentner & Markman, 1997), which use of the term *thinking* impels people to make when they consider the nature of LLMs. That is, the invocation of *thinking* when discussing LLMs is likely to result in people making inferences from their assumptions about human thinking, itself enriched by their own experiences of their own thoughts, as well as interactions with others. The

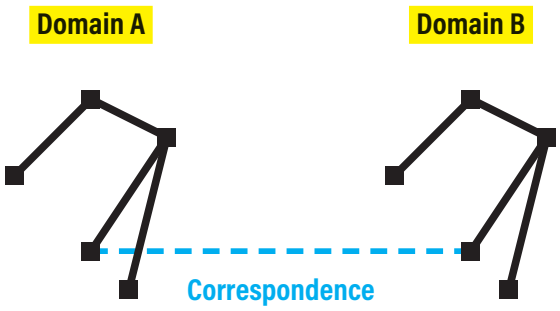


Figure 2. An abstraction of the analogy between two conceptual domains, after Gentner and Markman (1997). In this case, the two domains are in fact analogical. The dashed line indicates correspondence between nodes in the two domains. Only one line of correspondence is labeled, but each of the five nodes in Domain A corresponds to a node in Domain B, as indicated by matching network shapes.

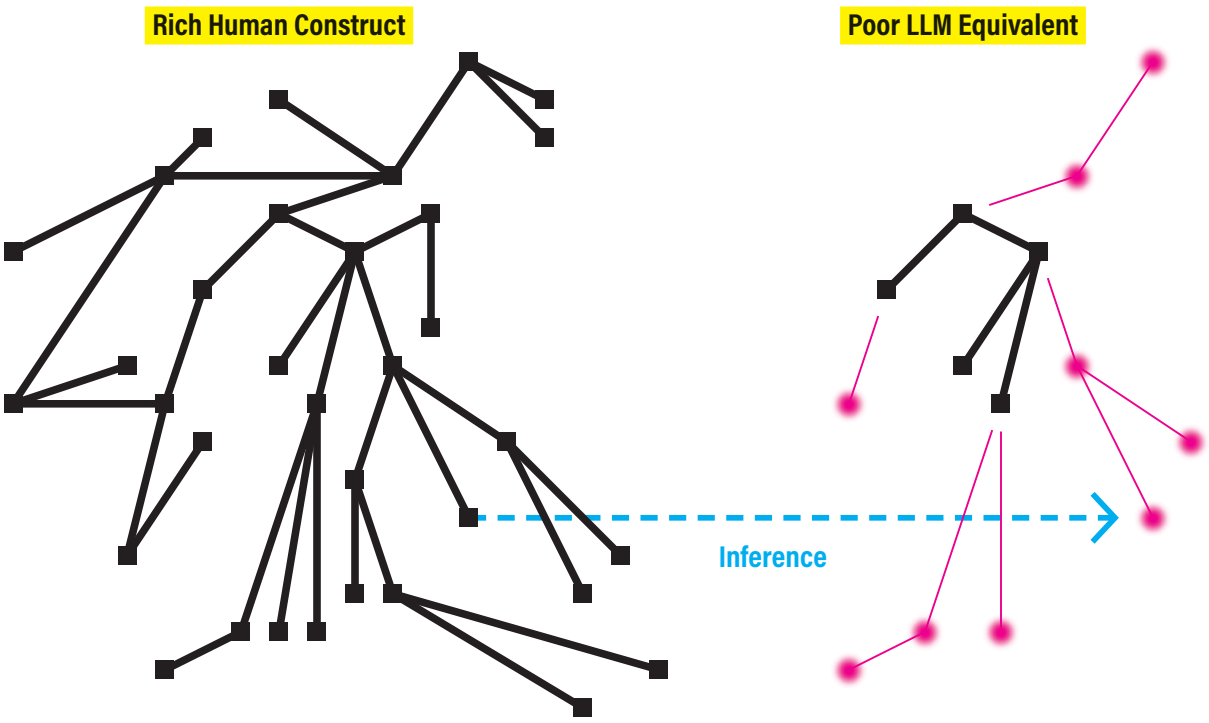


Figure 3. Complexity asymmetry between two domains in an imperfect analogy: a rich human construct such as *thinking* is paired with a comparably poor LLM equivalent. Inappropriate inferences may be made about the LLM as a target due to the richness of the thinking construct as a source, as indicated with the dashed cyan line. Inferred but inaccurate aspects of the LLM domain are indicated with blurred magenta circles. Properly corresponding aspects of the LLM domain are indicated with solid black squares.

discrepancies in LLM and human versions of "thinking" are too great for us to know all of the implications, and it is prudent to assume that many of the countless implications matter, perhaps to a great degree.

It is fair to state that there is an anthropocentric bias to selectively claiming that human-only aspects of thought are defining factors of thought, as is argued by Chen et al. (2026). However, saying that we do not have a complete enough sense of thinking to conclude that LLMs *do not* think is simply not a good argument for saying that they *do* think. In sharp contrast to Chen et al. (2026), we advocate for agnosticism on the matter, at the least. (While that is our prudent conclusion, we are individually and on a more personal level fairly comfortable stating that LLMs do not think, and do not come remotely close to doing so.)

Thus, instead of formally countering a statement that LLMs think with a definitive statement that they do not, we emphasize that the impression of thinking is based on similarity of outputs, and not on internal processes. Wittgenstein (1921/1922), in

considering the proposition (in language) to be a model of reality (p. 39), noted the discrepancy between output (language) and processing (thought):

> Language disguises the thought; so that from the external form of the clothes one cannot infer the form of the thought they clothe, because the external form of the clothes is constructed with quite another object than to let the form of the body be recognized. (p. 39)

Chen et al. (2026) suggest in their argument for LLMs' possession of general intelligence that an LLM "will correctly predict shattering in one case and not the other" for "dropping a glass or a pillow on a tile floor" (p. 39), but this is misleading. A human making this prediction will imagine a glass falling and shattering, at least partially pictorially—we think in pictures—and thus will predict based on an activated *model of reality*. They will, indeed, predict the result of a glass or a pillow falling. The LLM does not predict what the glass or the pillow will do. It predicts the words that a human is likely to use when the human is predicting the actual outcome. The outputs may be equivalent, but the underlying process is not. Bender and Koller (2020) offer insight here, warning that:

> we must be extra careful in devising evaluations for machine understanding, as Searle (1980) elaborates with his Chinese Room experiment: he develops the metaphor of a "system" in which a person who does not speak Chinese answers Chinese questions by consulting a library of Chinese books according to predefined rules. From the outside, the system seems like it "understands" Chinese, although in reality no actual understanding happens anywhere inside the system. (p. 5188)

(Searle was anticipating, in 1980, "strong AI" that "really is a mind, in the sense that computers given the right programs can be literally said to *understand* and have other cognitive states" [p. 417]. Amusingly, and reflecting the present and heated debates on the nature of LLMs, Searle's article is immediately followed in its publishing journal's issue by an open peer commentary entitled "Searle's argument is just a set of Chinese symbols" [Abelson, 1980].)

Imagine asking a person a question, which they answer. And then, as a follow-up, questioning them again, and again receiving an answer. We would not consider this person's thoughts to exist only in and as they articulate their responses. The first response, rather, would be considered an immediate utterance that is, to some degree, the result of a lifetime of thought that preceded it. And in the space between their two answers would be more relevant thought, influenced by having to consider the first answer, how the questioner responded through affect and gesture, and the second question. But for the LLM, anything remotely akin to thought occurs in the *moments of*

response alone, as responses, not thoughts on how to respond, and not in the context of any internal life.

LLMs produce outputs that resemble the products of thinking—this need not imply that they think. This is a reversal from *a definitive assessment of thinking* (one way or another) to *an acknowledgment that there are likely important production discrepancies despite similarities in resultant products.*

3.5. Learning, Thinking, Hallucination, and Writing in Synthesis

The learning, thinking, and hallucination constructs, when considered together in sequence, offer another conceptual reversal. In an LLM's initial training phases—during development—holistic source material is not "learned" and internalized as understood by humans. It rather disassembles mass amounts of information from original gestalts into weights in its probability distribution. The LLM does not retain the integrated wholes from which it derived isolated parts. This is fundamental to how LLMs work, even though additional technologies can mitigate this to varying degrees.

When the LLM is subsequently prompted by a human—during use—it works from these isolated "parts" to consolidate some of them in its node network into something that the human operator will perceive as a whole. But this is a new whole, not a recovered whole. If this new whole corresponds with something that the LLM was originally trained on, we would call it a reconstruction. But the "re-" prefix is a human reflection, an assessment; the LLM was nevertheless merely constructing something from raw material.

(This may appear reminiscent of the fact that when humans access long term memory, there is a creative process that can alter the memories stored there. Elizabeth Loftus's research on "repressed" memories, eyewitness testimony, and leading questions demonstrated this [Mook, 2004], but *unstable* gestalts are far removed from the *absence of gestalts,* so we do not consider this evidence of human–machine thinking correspondence.)

The same human assessment of an LLM's outputs that may deem one a construction instead of a reconstruction could result in the label of *hallucination.* But as previously discussed, hallucination implies a connection to ground truth through a violation of that ground truth. In viewing LLMs as disassembling wholes into parts and subsequently assembling something else with a selection of those parts, hallucination again rings false.

Returning to writing as a practice, consider a common occurrence: an LLM is asked to substitute for a human user's literature search, and among its outputs are multiple formal references to articles and books. We would typically say that if some of these references are not real—i.e., the publications they list in fact do not exist—then *these*

references in particular are hallucinations. But all of the references were produced in the same manner. The LLM was originally trained on a massive number of references, all of which were decomposed into weights in a probability distribution, and subsequently during use, it fabricated references. *Fabricated* here is a synonym for construction. So though we might be inclined to ultimately assess *particular* references as "fabricated references," in fact even the references we accept have been fabricated. As an LLM assembles a reference token by token, it is following pathways through its probability distribution, with the potential for it to make all of what we would call "right turns," and output an authentic reference; but also with the potential to veer away.

What we typically call hallucination is inseparable from the very nature of LLMs. Any innovations that completely eliminate the problem of hallucination will result in something, by definition, other than an LLM.

3.6. Implications of Alternative Statements

Statements. The preceding description of four constructs affiliated with LLM technology led to misleading statements about LLMs. We suggested alternative statements for these misconceptions, which are collected in Table 1. Our premise is that these alternative statements support better human decision making with LLMs. Though this premise could hold even if it were impossible to observe directly, we next consider some overt ways in which these statements might express themselves in decision making.

Decisions. Consider a college student in an introductory course in their field of study who believes that *LLMs write texts.* When given a writing assignment, this student is likely to use an LLM to write their paper. They may have heard that LLMs are useful tools, and thus they "use" this tool to "write their paper." But if they instead believe that *LLMs produce coherent text without writing,* they may instead see using the LLM as sidestepping the writing task ("without" writing), and they may be slightly more likely to conclude, even unconsciously, that using an LLM to produce their paper is improper. Their instructor is likely to agree.

Now consider an experienced writer who is an expert in their discipline, whose potential use of AI is not governed by a syllabus policy that makes it "improper." If this expert believes that *LLMs write texts,* they may be comfortable offloading parts of their writing process to an LLM without questioning its impact on their writing process overall. But if they believe that *LLMs produce coherent text without writing,* they may use their sophisticated understanding of the writing process—informed by their own metacognition over thousands of hours of writing—to selectively compartmentalize aspects of their writing process, engaging in a preservation of their writing's integrity.

Next consider the college student and the disciplinary expert together. If instead of viewing LLMs as creative in the normal sense, they think that they *generate text patterns*

Table 1. Misleading statements about LLMs compared with corrected versions.

Misleading statement	Corrected statement
LLMs think.	LLMs produce outputs that resemble products of thinking.
	LLMs recombine semantic parts that are separate from their original wholes.
LLMs learn from accumulated human knowledge.	LLMs are trained on patterns in a collective information store.
	LLMs disassemble meaningful gestalts during training.
LLMs occasionally hallucinate.	LLM outputs frequently reflect reality.
	LLMs fabricate texts.
LLMs model reality.	LLMs provide humans in the real world access to a fragmented hyperreality.
LLMs write texts.	LLMs produce coherent text without writing.
	LLMs assemble gestalts in which people find meaning.
LLMs are creative.	LLMs generate text patterns through partial replication of other text patterns.

through partial replication of other text patterns, they may be more careful in employing an LLM in writing. The student may fear cheating while the expert fears plagiarism. Replication of existing language suggests a high level of responsibility for interrogating outputs and investigating provenance.

Finally, consider the disciplinary expert when they have agreed to peer review a submission for an academic journal. If they believe that *LLMs think,* then they may consider it adequate to prompt an LLM with their intentions for review, and then have the LLM "review" the manuscript. But if they believe that *LLMs produce outputs that resemble the products of thinking* and thus that LLMs cannot truly reason or evaluate, they may conclude that the LLM cannot substitute for their expert evaluation. The manuscript's authors and the journal editor are likely to agree.

4. Alternative Characterizations of LLM Technology

Our assumption is that because LLM technology is unfamiliar and mysterious as a black-box technology, that metaphors consistently used to characterize it will function as conceptual metaphors (Lakoff & Johnson, 1980), constraining thoughts about LLMs and thereby influencing usage patterns with them.

In a remarkable paper that raised alarms about LLM technology, Emily Bender, Timnit Gebru, and colleagues (2021) called the language model a *stochastic parrot. Stochastic* in this case refers to the underlying mechanism of probabilistic weights used in next-token prediction, and it is generally defined as "randomly determined; having a random probability distribution or pattern that may be analyzed statistically but may not be predicted precisely" (New Oxford American Dictionary, June 2, 2026). The *parrot* half of the metaphor emphasizes the absence of meaning, in that the LLM has only its patterns of relationships among tokens. Considering the relationship of language model (LM) to human user in text-based communication, Bender et al. (2021) state that:

> The problem is, if one side of the communication does not have meaning, then the comprehension of the implicit meaning is an illusion arising from our singular human understanding of language (independent of the model). [...] Contrary to how it may seem when we observe its output, an LM is a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning. (pp. 616–617)

This illusory meaning is predicated on an understanding of language that distinguishes between form and meaning, as described by Bender and Koller (2020):

> We take *form* to be any observable realization of language: marks on a page, pixels or bytes in a digital representation of text, or movements of the articulators. [...] We take *meaning* to be the relation between the form and something external to language... (pp. 5186–5187)

The "something external" that the LLM's language is disconnected from is anything in reality itself. Bender and Koller also emphasize intentionality, an important facet of the stochastic parrot, which "parrots" language that it does not understand: "When humans use language, we do so for a purpose: We do not talk for the joy of moving our articulators, but in order to achieve some *communicative intent*" (p. 5187).

The stochastic parrot metaphor is popular (Bender et al., 2025)—as of June 2, 2026, only roughly five years after publication, Bender et al. (2021) had accumulated over 14 thousand citations on Google Scholar. However, it may over-correct in its move away from unjustified positivity in favor of LLMs—a positivity to which companies that make money on LLMs are prone. The stochastic parrot argument largely holds up in its assertion that models represent patterns determined by statistical prevalence in training data and not deeper conceptual understanding or embodied experience. But the term *parrot* minimizes the current capabilities of LLMs. Models generalize and display emergent behaviors not explicit in the training data, creating connections between seemingly dissimilar patterns. They effectively translate languages, write

code, solve problems, use tools, and follow complex instructions where they are not just repeating strings of previously seen text. While in the act of predicting next tokens, LLMs form abstract internal representations of possible future continuations, latent goals, and structural constraints that suggest that there is something more going on than simple mimicry (Lindsey et al., 2025). Our search for metaphors continues.

Fanciful metaphors about LLM technology abound. Ethan Mollick offers two noteworthy metaphors. In lamenting LLM-generated posts online, Mollick (2026) describes them as "just *meaning-shaped attention vampires* that take mental effort to decode and give you no equivalent understanding in return" (emphasis added). He describes the LLM-based chatbot (ChatGPT specifically) as "an infinitely sort of *patient intern that sometimes lies to you*" (Mollick & Nickisch, 2023, emphasis added). We can invent our own fanciful metaphors that eschew AI hype, for instance, by emphasizing the LLM's limitation of its connection to a dataset (i.e., internal representations alone) as a hyperreality:

- LLMs are *hyperreality ambassadors,* giving us partial access not to reality, but to the hyperreality of the internet, with its biases, contradictions, and falsehoods. While this metaphor draws a hard line between the LLM's disconnected and fractured dataset and the world, ambassadors are people with intentionality, emotion, interior lives, and personal histories. There are too many problematic entailments in such personification, and so we reject this metaphor.
- LLMs are *gaslighters,* emphasizing that the human user interacting from a position in reality with an LLM that is in hyperreality can confuse what is reality. Though we offer this up as an example, we immediately reject it on the grounds that gaslighting is done systematically and intentionally, and this is too extreme a characterization, and too negative.
- LLMs are *hallucinators.* We reject this metaphor also, because it implies a connection to reality that is bypassed. LLMs have no such connection to bypass—again, they frequently reflect reality rather than occasionally hallucinate.

Other available metaphors emphasize some aspect of LLM technology without drawing human parallels, and we believe work better by characterizing LLMs as artificial:

- LLMs are *meaning emulators.* This metaphor is terminologically reminiscent of *stochastic parrot* but offers another verb (*emulation*) to describe the work of LLMs.
- LLMs are *answer simulators* that appear to answer questions in the thoughtful way that people do, but in fact are predicting (token by token) what a person would answer. This is not the same as answering.
- LLMs are *averaging machines* or *algorithmic homogenizers,* because their weighted probabilities will have a central tendency within the distribution of their training data. This is in one sense a strength of LLMs: they give us access to something approaching a consensus. But the metaphor is useful when an inexperienced

writer is trying to use an LLM creatively. The result is not likely to be special due to this central tendency—one person's prompted result is likely to be overly similar to another person's.

- LLMs are *remixers,* collecting knowledge contributions from humans and combining them—sometimes plagiarizing them.
- LLMs are *plagiarism traps* that regularly obscure sourcing with which their human operators are unfamiliar. This means that the central tendency that leads to homogenized results also leads to inappropriately copied results.

The above characterizations may appear to be rather negative. They do in fact emphasize limitations. (Far more negative machine metaphors are available, such as *bullshit generator,* a pre-existing term that is frequently applied to LLMs.) The following metaphor emphasizes how LLMs can be used as tools when only limited function is required of them:

- LLMs are *fuzzy thesauruses.* A conventional thesaurus enables writers to find synonyms and antonyms for specific words. But the chatbot interface for an LLM permits writers to explore a looser terminological range of possibility even when they cannot call to mind a particular word as a starting point, which a conventional thesaurus requires. With a chatbot, a writer can circle around a concept instead of needing to first name the concept. The term *fuzzy,* as in fuzzy set theory, suggests a gradient of semantic inclusion for similarity rather than a binary of either is-similar or is-not-similar. So the LLM can reach further out than a standard set of synonyms.

We used ChatGPT as a fuzzy thesaurus to generate alternatives to *fuzzy thesaurus,* beginning with prompts that loaded the "nearest neighbor" and "paradigmatic axis" concepts, and it suggested *probability harvester, paradigmatic displacement apparatus,* and *semantic neighborhood watch,* among other metaphors (ChatGPT, June 8, 2026). (A *nearest neighbor* is a point in a set that is most similar to another point, and the *paradigmatic axis* in semiotics contends that a word's usage is meaningful in part as a substitution, or in contrast to all the alternatives that could have been but were not used [which we explore in Boxes 1–4].) Here conceptual material that is meaningful to a human prompter *before an interaction with a chatbot* (e.g., nearest neighbor, paradigmatic axis) ensures that suggestions are imbued with meaning (i.e., the LLM is not a standalone stochastic parrot). But we consider these ChatGPT metaphors to be novelties rather than good options.

We find four of the above metaphors for LLMs to be the most balanced and reflective of their nature. The first three are the LLM as an answer simulator, as a remixer, and as a fuzzy thesaurus. Together they characterize the LLM as a non-thinking tool (an answer simulator; neither respondent nor consultant), which presents information in

a changed form (a remixer), and functions through flexible word associations (a fuzzy thesaurus). An additional metaphor, the LLM as a plagiarism trap, may appear negative, but it is actually balanced because people can avoid traps when they know how the traps work and where they can be found. The metaphor presents a warning while implying that humans can engage safely, if carefully. These more balanced metaphors leave room for principled use by human operators while avoiding major misconceptions about LLM technology.

5. Conclusion

Large language models do similar things to what people have traditionally done. This challenges our assumptions or blind spots about those things. For instance, writing is a complex endeavor, but our shorthand notion of writing is simply producing coherent language in written form. Once an LLM can produce coherent language in written form, we arrive at some contradictions. An instructor would not consider it acceptable for a student to "write" a paper with a single prompt and then take credit as the author. And no serious academic journal would accept a paper with ChatGPT as an author with a human prompter credited in the acknowledgments. So what is writing, actually? The point is that this situation begs for more explicit descriptions of constructs like writing, learning, hallucination, and thinking, and all that we can compare these constructs with is humans—with our own versions of them. With a given construct, what is the deviation from the human version, and when is the construct too far removed to even be invoked? LLMs may be intelligent in some way, but that intelligence is probably quite a bit different from human intelligence. This results in difficulties around assumptions or expectations of truth, and whether (or when) we can trust the output of AI models and systems.

We provided a "basic" description of LLM technology (i.e., nontechnical and with minimal use of metaphor), and compared it to four constructs as understood separately from LLMs: writing, learning, hallucination, and thinking. In doing so, we highlighted numerous alternative statements about the technology:

1. LLMs produce outputs that resemble products of thinking.
2. LLMs recombine semantic parts that are separate from their original wholes.
3. LLMs are trained on patterns in a collective information store.
4. LLMs disassemble meaningful gestalts during training.
5. LLM outputs frequently reflect reality.
6. LLMs fabricate texts.
7. LLMs provide humans in the real world access to a fragmented hyperreality.
8. LLMs produce coherent text without writing.

9. LLMs assemble gestalts in which people find meaning.
10. LLMs generate text patterns through partial replication of other text patterns.

An additional statement includes a warning to inexperienced writers—no, all writers—about LLM use:

11. LLMs can be used to bypass the personal benefits of learning through writing when producing text.

We then offered a selection of conceptual metaphors, and rejected many as overly negative or misleading. We ultimately favor characterizing LLMs as:

- Answer simulators
- Remixers
- Fuzzy thesauruses
- Plagiarism traps

Language constrains thought, and thought leads to action. The use of LLMs in writing and in other endeavors should be done thoughtfully and with reasonable conceptions of LLM technology. It is undeniable that the large language model is an amazing innovation, especially difficult for humans to understand. Accurately identifying a technology's inherent limitations does not deny that technology's importance. LLMs are not what they are not, but they certainly are what they are.

6. Acknowledgments

Kirsti Cole—a professor of English and co-director of the Campus Writing and Speaking Program at NC State University—provided extensive notes on an early version of the manuscript. Many of her comments were processed and are embodied in the final version, but this does not imply that she agrees with all assertions presented here.

The third author is a United States governmental employee, and thus we provide the following acknowledgment. This material is based upon work done, in whole or in part, in coordination with the Department of Defense (DoD). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the DoD and/or any agency or entity of the United States Government.

7. References

Abelson, R. P. (1980). Searle's argument is just a set of Chinese symbols [Open peer commentary]. *Behavioral and Brain Sciences, 3*(3), 424–425. https://doi.org/10.1017/S0140525X00005756

Aleman, A., & Larøi, F. (2008). Hallucinations: The science of idiosyncratic perception. American Psychological Association. https://doi.org/10.1037/11751-000

APA (2018). Visual hallucination. In *APA dictionary of psychology*. https://dictionary.apa.org/visual-hallucination

Baddeley, A. D. (2018). Working memory: Theories, models, and controversies. In *Exploring working memory: Selected works of Alan Baddeley* (pp. 332–369). Routledge.

Baudrillard, J. (1981/1994). *Simulacra and simulation* (Trans. S. F. Glaser). The University of Michigan Press.

Bekerman, Z., & Zembylas, M. (2018). *Psychologized language in education: Denaturalizing a regime of truth.* Palgrave Macmillan.

Bell, J. (2022). What is machine learning?. In S. Carta (Ed.), *Machine learning and the city: Applications in architecture and urban design* (pp. 207–216). John Wiley & Sons Ltd.

Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th annual meeting of the Association for Computational Linguistics* (pp. 5185–5198). Association for Computational Linguistics. https://aclanthology.org/2020.acl-main.463.pdf

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? 🦜. In *Conference on Fairness, Accountability, and Transparency (FAccT '21),* March 3–10, 2021, Virtual Event, Canada. ACM. https://doi.org/10.1145/3442188.3445922

Bender, E. M., Costello, E., Lee, K., Farrow, R., & Ferreira, G. (2025). Unsafe AI for education: A conversation on stochastic parrots and other learning metaphors ⚠️. *Journal of Interactive Media in Education, 2025*(1), article no. 10. https://doi.org/10.5334/jime.1079

Bruner, J. S. (1966). *Toward a theory of instruction.* Harvard University Press.

Chandler, D. (2021, November 23). Paradigms and syntagms [Webpage]. *Semiotics for beginners.* http://visual-memory.co.uk/daniel/Documents/S4B/sem03.html

Chen, E. K., Belkin, M., Bergen, L., & Danks, D. (2026, February 2). Does AI already have human-level intelligence? The evidence is clear. *Nature, 650,* 36–40. https://www.nature.com/articles/d41586-026-00285-6

Díaz, B., & Delgado, C. (2024). Artificial intelligence: Tool or teammate?. *Journal of Research in Science Teaching, 61*(10), 2575–2584. https://doi.org/10.1002/tea.21993

Duckworth, E. (1964). Piaget rediscovered. *The Arithmetic Teacher, 11*(7), 496–499. https://www.jstor.org/stable/41186862

Emig, J. (1977). Writing as a mode of learning. *College Composition and Communication, 28*(2), 122–128. https://doi.org/10.2307/356095

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365–387. https://www.jstor.org/stable/356600

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155–170.

Gentner, D. (2003). Analogical reasoning, psychology of. In L. Nadel, R. French, P. Culicover, D. Chalmers, R. Goldstone, M. Gasser, D. Noelle, K. McCabe, J. Levin, & J. Odling-Smee (Eds.), *Encyclopedia of cognitive science* (pp. 106–112). https://doi.org/10.1002/0470018860.s00473

Gentner, D., & Gentner, D. R. (1983). Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner, & A. L. Stevens (Eds.), *Mental models* (pp. 99–129). Lawrence Erlbaum Associates, Publishers.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*(1), 45–56. https://doi.org/10.1037/0003-066X.52.1.45

Gibney, E. (2025, October 20). AI language models killed the Turing test: Do we even need a replacement? *Nature: News.* https://doi.org/10.1038/d41586-025-03386-w

Klein, G. A., Calderwood, R., & Macgregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics, 19*(3), 462–472. https://doi.org/10.1109/21.31053

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery.* Oxford University Press.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason.* University of Chicago Press, 1987.

Johri, S. (2023). The making of ChatGPT: From data to dialogue [Blog post]. *Science in the news.* https://sites.harvard.edu/sitn/2023/06/06/the-making-of-chatgpt-from-data-to-dialogue/

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought.* Basic Books.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511815355

Lindsey, J., Gurnee, W., Ameisen, E., Chen, B., Pearce, A., Turner, N. L., Citro, C. Abrahams, D., Carter, S., Hosmer, B., Marcus, J., Sklar, M., Templeton, A., Bricken, T., McDougall, C., Cunningham, H., Henighan, T., Jermyn, D., Jones, A., . . . Batson, J. (2025). On the biology of a large language model. *Transformer Circuits.* https://transformer-circuits.pub/2025/attribution-graphs/biology.html

Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review, 99*(4), 587–604. https://doi.org/10.1037/0033-295x.99.4.587

Maturana, H. R. (1975). The organization of the living: A theory of the living organization. *International Journal of Man-Machine Studies, 7*(3), 313–332. https://doi.org/10.1016/S0020-7373(75)80015-0

Maturana, H. R., & Varela, F. J. (1991). *Autopoiesis and cognition: The realization of the living* (Vol. 42). Springer Science & Business Media.

Mickus, T. (2024). Language models and the paradigmatic axis. In *Proceedings of the Society for Computation in Linguistics 2024* (pp. 63–74). Association for Computational Linguistics. https://aclanthology.org/2024.scil-1.7/

Mollick, E. (2026, May 26). Choosing to stay human… Means choosing when and how to use AI [Blog post]. *One Useful Thing.* https://www.oneusefulthing.org/p/choosing-to-stay-human

Mollick, E., & Nickisch, C. (2023, March 14). Why you (and your company) need to experiment with ChatGPT now [Podcast episode]. *HBR IdeaCast,* episode 901. https://hbr.org/podcast/2023/03/why-you-and-your-company-need-to-experiment-with-chatgpt-now

Mook, D. (2004). Elizabeth Loftus: Leading questions and false memories. In D. Mook (Ed.), *Classic experiments in psychology* (pp. 199–202). Greenwood Press.

Nussbaum, M., Bekerman, Z., Archambault, L., Ranieri, M., Yilmaz, R., van Braak, J., Zou, D., & Tsai, C. C. (2026). *Teachers' doings in the AI era: From transmission to relation. Computers & Education, 249,* 1–8. https://doi.org/10.1016/j.compedu.2026.105617

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback [Preprint]. *ArXiv.* https://doi.org/10.48550/arXiv.2203.02155

Patil, R., & Gudivada, V. (2024). A review of current trends, techniques, and challenges in large language models (LLMs). *Applied Sciences, 14*(5), 2074. https://doi.org/10.3390/app14052074

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

Shapiro, L. (2019). *Embodied cognition* (2nd ed.). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781315180380

Shaw, S. D., & Nave, G. (2026). Thinking—fast, slow, and artificial: How AI is reshaping human reasoning and the rise of cognitive surrender [Preprint]. https://doi.org/10.31234/osf.io/yk25n_v1

Stryker, M. (n.d.). What are LLMs? [Webpage]. *IBM: Think.* Accessed March 18, 2026. https://www.ibm.com/think/topics/large-language-models

Vygotsky, L. (1978). *Mind in society.* Harvard University Press.

Watson, J. B. (2017). *Behaviorism.* Routledge.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity.* Cambridge University Press.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review, 9*(4), 625–636. https://doi.org/10.3758/BF03196322

Wittgenstein, L. (1921/1922). *Tractatus logico-philosophicus* (Trans. C. K. Ogden). Kegan Paul, Trench, Trubner & Co., Ltd. https://www.gutenberg.org/files/5740/5740-pdf.pdf

Authors

Matthew Peterson is an associate professor of graphic and experience design at NC State University. His research focuses on visual representation in user interface design, including the facilitation of AI in intelligence analysis workflows through human–machine teaming, text–image integration in immersive user information systems, and the facilitation of scale cognition and numeracy through interaction design.

Colleen Patton is an assistant professor of psychology at NC State University. She received her PhD in Cognitive Psychology from Colorado State University in 2023. Her research focuses on optimizing the performance of human-automation teams in complex decision making contexts, emphasising attentional limitations and the subjective experiences of human operators.

Michael Green is a machine learning researcher at the Laboratory for Analytic Sciences at NC State University and a US Department of Defense employee. His work explores means of making AI useful to knowledge workers and on the development and evaluation of systems that augment analytic decision making.

Brayan Díaz is an assistant professor of STEM education at Utah State University and a researcher affiliated with the Chilean National Center for Artificial Intelligence. His research examines how artificial intelligence can support teaching, learning, and teacher professional development, with a particular focus on AI literacy, pedagogical decision-making, equity, and access in STEM education. He also studies collaborative learning, communities of practice, and the design of interdisciplinary STEM learning environments.