Large Language Models and the Patterns of Human Language Use

Christoph Durt[A], Thomas Fuchs[B]

[A] Technical University of Munich, TUM School of Social Sciences and Technology, Arcisstr. 21, 80333 Munich, [B]University of Heidelberg, Psychiatric University Hospital, Voßstrasse 4, 69115 Heidelberg

KEYWORDS
LLMs, AI, Language, Language Use, Meaning, Semantics, Understanding, Experience, Sentience, Scaffolding, Patterns, Clichés

REFERENCE

Durt, Christoph, and Thomas Fuchs. 2024. "Large Language Models and the Patterns of Human Language Use." In *Phenomenologies of the Digital Age*, by Marco Cavallaro and Nicolas De Warren, 1st ed., 106–21. New York: Routledge. https://doi.org/10.4324/9781003312284-7.

6.1 INTRODUCTION

Generative artificial intelligence (AI) commonly stands for deep learning models that are trained on huge amounts of data and which can produce text, images, audio, and videos, usually in response to prompts. Large language models (LLMs) are currently the most successful generative AI technology, and we will here focus on text-producing LLMs, many of which are built to function as communication partners. A widely known example for generative AI is ChatGPT, an LLM chatbot that produces text in response to a prompt.

As if by magic, LLMs can generate text that humans perceive as meaningful and relevant with respect to an input prompt. The magic would of course disappear if, like in the case of Wolfgang von Kempelen's Mechanical Turk, we would find out that inside the system, there is a human who responds to the prompts. Knowing about the basic workings of LLMs, however, that the processing is done by an artificial neural network that represents numerical relations between small text chunks (tokens) in a multidimensional vector space, may only intensify the feeling of magic. No amount of further technical knowledge of the construction of LLMs seems to diminish the amazement at how, in many cases, a human-like output can be produced using computation. Many contemporaries let themselves be blinded by the seeming magic and readily attribute these text-production abilities to a supposed understanding, sentience, consciousness, or general intelligence of LLMs. Even when it is admitted that the alleged capacities of LLMs are very different from those of humans, using such terms for LLMs implies either a reduction of mental phenomena to external output, or taking this output to be sufficient to prove mental phenomena.

The latter kind of reasoning was put into clear words by Blake Lemoine, the Google employee, who claimed in 2022 that the LLM LaMDA had developed sentience: “I know a person when I talk to it. It doesn’t matter whether they have a brain made of meat in their head. Or if they have a billion lines of code” (Tiku 2022). This claim, however, ignores that computational systems may be able to produce similar-looking output by different means. A few weeks before Lemoine went public, Ilya Sutskever, one of the most important researchers on deep learning and co-founder of OpenAI, posted that “it may be that today’s large neural networks are slightly conscious” (Sutskever 2022). Behavioral evidence is furthermore used to speculate that for these systems, “statistics do amount to understanding” (Agüera y Arcas 2022), or that they show “inklings” (Manning 2022) or “sparks” (Bubeck et al. 2023) of artificial general intelligence (AGI). The suggestion is that as LLMs become always better, they will eventually surpass the awareness, understanding, or general intelligence of humans. Some guess that “we’ll have AI that is smarter than any one human probably around the end of next year, and then AI, the total amount of sort of sentient compute of AI, I think will probably exceed all humans in 5 years” (Musk 2024). Such suggestions are sure to generate attention, clicks, and citations, and are proliferated through algorithms and social media, but they are assumptions and speculations rather than serious philosophical considerations.

The common alternative to ascribing phenomenological experience to machines is to deny that understanding and experience play any role in their capabilities. The very fact that LLMs sometimes produce meaningful and sometimes nonsensical output can be taken as evidence that they lack understanding (Mitchell 2019; Mitchell and Krakauer 2023). Some researchers have taken the lack of understanding to mean that an LLM is nothing more than a “stochastic parrot”, namely “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” (Bender et al. 2021). Of course, parrots are intelligent creatures that have intentions and can understand a situation, but Bender et al. are obviously using this metaphor to point out that parrots do not understand the meaning that some of the sounds they produce have in human language.

Furthermore, Bender et al. point to the importance of training data, which is “parroted” by LLMs. This is a very important point because the discussion about “understanding” and “meaning” often neglects the role of training data in enabling LLMs to acquire their astonishing capabilities. We need to work out a clearer notion of the training data and its relation to apparent understanding and experience. It would be too simplistic, however, to claim that LLMs merely parrot these data. Parrots simply repeat words or strings of words, and sometimes, they stitch together sequences of words they can speak. But, in contrast to parrots, LLMs can generate complex text that makes sense, had never been written before, and also constitutes a meaningful “response” to the prompt.

The stitching together of “linguistic forms” does not seem to be random but rather ordered in such a way that the new combinations make sense to humans. So, how can the calculation of forms generate language that appears meaningful to humans? This again seems puzzling if the calculation is seen as arbitrary or random. In this chapter, we thus explain how LLMs can produce meaningful text without understanding, sentience, or consciousness.

The concept of training data emphasizes their calculative use (“training”) and therefore may sound transitionary and uninteresting for philosophers who are concerned with human experience. It may seem as if the question of the role of understanding, sentience, or consciousness with respect to LLMs would either be about finding experience or understanding in numerical relations, or require an outright rejection. But the same texts that are training data for the machine are meaningful language to humans. They are derived from the internet and elsewhere, where they have been either used in concrete and meaningful contexts, or exist as records of such uses. Thus, we have to investigate the training data of text-generating LLMs not with respect to their mathematical form but with respect to the meaningful language use they derive from.

We contend that the reason for why mere numerical operations can generate language that is meaningful to humans is that language use is permeated with patterns that can both be stochastically modeled and understood as meaningful. Patterns, in general, are repetitive regularities. As such, they can occur in very different domains, such as in mathematics, in language use, and in the phenomenology of experience. For our purpose, we distinguish between (1) stochastic patterns, (2) patterns of language use, as well as (3) patterns of experience. Although they belong to different domains that are not reducible to each other, there are interrelations. These interrelations are key to the possibility of computational modeling of language use as well as its relation to the phenomenology of experience.

1. Text-generating LLMs model patterns of a certain form of language use, namely the statistical relations between tokens of written text in a text corpus. Thus, they extract *stochastic patterns* that allow them to calculate the most probable words in the training data and to produce new text that resembles text produced by humans. Stochastic patterns are mathematical entities that have no meaning as such. But they are derived from meaningful language use that is sedimented in the training data. It has been shown that LLMs can sometimes regurgitate entire passages of the data (Grynbaum and Mac 2023), but usually stochastic patterns do not simply correspond to a single text but are extracted from many similar training data.

2. *Patterns of language use* exist both on syntactic and on semantic levels. They manifest themselves in the way words and expressions are used in communicative practice, usually institutionalized in rules of the language game. Language use lends itself to computational processing because its patterns can be rearranged in ways that make new sense to humans. The success of the generation of human-like language by LLMs hence also reveals something about human language use, namely how much it is permeated by *patterns of meaning*. Some of these express clichés, which may be repeated, amplified, and proliferated by LLMs. Other patterns already express meaningful mental connections that are exploited by the LLMs.

3. Besides distinguishing stochastic patterns and patterns of linguistic meaning, we can also examine how human language use is embedded in *patterns of experience*. Language use takes place in a situation, and its production and reception by humans is usually connected to awareness and experience of the situation. There is a phenomenology of human language use in two respects. On the one hand, humans produce language in the context of the situation they experience and in view of the listeners or readers they address, usually as part

of an interaction. On the other hand, texts are read by humans whose mental life is augmented when they interpret and make sense of the texts. The investigation of the phenomenology of language use is nothing new for phenomenology. In section 6.3, we will draw on the notion of a horizon of anticipation that comprises focal awareness as well as fringes the speaker or writer is less aware of.

6.2 Meaningful Language Use

The role of patterns of language use is easily overlooked under the standard picture of meaning, according to which meaning can be detached from language, which is thought to be a mere formal system. The problem that resurfaces in the context of LLMs is that the standard picture does not account for the extent to which meaning is intertwined with the *use* of language, including descriptions, worldly interactions, writing, and verbal thought. LLMs model not simply grammatical rules but meaningful language *use* in the text corpora they are trained with. It is therefore important to consider not simply a static structure of language, but language use in its many patterns. The importance of language use was famously pointed out by Ludwig Wittgenstein. Regarding the relation between meaning and use of language, he writes in his *Philosophical Investigations*:

For a large *class of cases of the employment of the word "meaning" – though not for* all *– this word can be explained in this way: the meaning of a word is its use in the language. And the* meaning *of a name is sometimes explained by pointing to its* bearer. *(Wittgenstein 2009, sec. 43)*

Wittgenstein admits that the common picture has some explanatory power: the deictic reference to a name can *sometimes explain* its meaning. But this does not allow the inference that the meaning of a name *is* its bearer, nor that other forms of meaning can be adequately described in terms of naming. Instead of applying some clean but artificial definition of "meaning," Wittgenstein demands to consider the actual use of the word, by which he does not merely mean statistical relations in a text corpus, but the use of the word in language games.

Considering the actual use of "meaning," he does not find what is often ascribed to him, namely that "meaning is use." He rather writes that for a *large class* of cases of employment of "meaning," the word *can be explained* as its use. A word whose meaning is unclear can usually be explained by describing its use, and in the case of a name sometimes simply by pointing to its bearer. Hence, language is not an abstract symbol system that in itself has no relation to the world. On the contrary: because language is used in the context of a "language game" and ultimately a "form of life" (Wittgenstein 2009, sec. 23), meaning is embedded in the world we live in, including the communicative and mental activities in which we make use of language (Durt 2022). The world is already meaningful before the use of symbolic language (Stuart 2024), and the use of language modifies and shapes meaning.

The notion that language as a system derives from language as use has already been proposed by Ferdinand de Saussure in his classic distinction between *langue* and *parole* ([1916] 2011). Language as a *general system of signs and rules* (*langue*) emerges as a structure of *language spoken in concrete situations* (*parole*). In a spoken language, the speaker's as yet unsymbolized experiences are articulated in ever new ways. These articulations, i.e., the living use of language as *parole*, continuously modify the linguistic structures and patterns (including usage and typical word sequences, grammatical rules, and meaning contexts), so that *langue* can be seen as a constantly evolving collective structure of regularities and meanings. In other words, *langue* is not a static system that is independent of use, but rather derives from its use. Yet, *langue* is not only a structure derived from use in *parole* – conversely, *langue* also structures *parole*. With an expression borrowed from Pierre Bourdieu, we may say that *langue* acts as a "structuring structure"1 for our current articulations. Regularities derive from use, and in turn, they also structure use.

Considering not just *langue* but *parole* or *language use* is crucial to explain the ability of LLMs to produce meaningful text that goes beyond merely correct syntax. If meaning is expressed in language use, then it can be modeled by statistical means in so far as the use can be modeled. The possibility of statistical representation of meaning was demonstrated long before true LLMs existed. For example, it has been hypothesized that "the proportion of words common to the contexts of word A and to the contexts of word B is a function of the degree to which A and B are similar in meaning" (Rubenstein and Goodenough 1965, 627). It has been argued that vector representations can capture "a large number of precise syntactic and semantic word relationships" (Mikolov et al. 2013, 1). LLMs have been shown to learn syntactic structures such as subject-verb agreement and dependency structures (Hewitt and Manning 2019). To a lesser extent, already older LLMs have been shown to learn semantic structures such as tense (Jawahar, Sagot, and Seddah 2019) and semantic roles (Tenney, Das, and Pavlick 2019).

Recent LLMs show that the extent to which meaning can be computationally generated is much greater than linguists and computer scientists had believed. We suggest that the reason is that LLMs not just represent general structures, but the part of the *use of language* that is represented in their training data. We agree with Bender and Koller (2020, p. 191) that Wittgenstein's concept of "use" refers to language use in the real world. Yet, this use can be partially reflected in the distribution in a text corpus. The idea of a semantic "distributional structure" of language (Harris 1954), namely that "words that occur in similar contexts tend to have similar meanings," is called the "distributional hypothesis" (Turney and Pantel 2010, 143). It is sometimes extended into a "distributional semantics" (Bernardi et al. 2015), which is contrasted with "*denotational semantics* or a *theory of reference*" (Manning 2022, 134, emphasis in original).

While distributional semantics rightly points out that meaning has to do with distribution in a text corpus, this does not mean that meaning is reducible to statistical distribution. We suggest that neither denotational nor distributional semantics alone are sufficient to explain how LLMs produce meaningful text. Rather, the text corpus reflects actual and meaningful *language use*, with the important restriction that it is only a part of the lived meaning that is reflected, and only in incomplete ways. Because meaning always has an intrinsic connection to the current context and to the primary experienc e of those who articulate this meaning. However, the text corpus LLMs are trained on is detached from any reference to context or experience – that is one of its major limitations.

Modeling language use entails the modeling of the statistical contours of sense-making processes and thereby indirectly models aspects of meaning. Modeling of language use does not mean that LLMs learn syntactic and semantic structures in the way humans do, but that they are able to calculate, to some extent, text that humans may produce. Because most current models only model stochastic patterns of written text, and not all of human language use, and because they do not generate text in the same ways as humans, their calculations are likely to differ to some extent from human-produced text. The delta between probability-based output and human text production shows most obviously when LLMs produce output no human would produce, such as a nonsensical one.

An obvious restriction to modeling semantic structures and patterns derives from the fact that the text corpus LLMs are trained on is at the same time exceedingly large and very limited. It consists of much of English language written on the internet and other digitally available texts, including web pages, books, chats, and transcripts of spoken language. Despite the enormous size of their training corpus, current LLMs model only one aspect of human language use, namely the use of written language and written transcripts of spoken language. The use of language goes much beyond writing, and writing captures only a part of the use of written and spoken language. Yet, writing is an important part of use of many languages – including the dominant languages of the world and excluding the majority of languages, which are not written. The very limitations of written language also make it easier for LLMs to produce convincing text – when interpreting text, humans fill in the missing context. Both the limits and capacities of LLMs are consequences of how humans produce and understand language. We will take a closer look at the process of human language production in the next section and then come back to how humans tend to read meaning and authorship into text.

6.3 Linguistic Scaffolding in Human Language Production

Since meaningful language use by humans is generally interwoven with their situated mental life, in this section, we consider the phenomenological structure of human language production and the emergence of meaningful patterns. The ability of LLMs to calculate meaningful text is founded in patterns of human language use, and these have to do with the experiential patterns of our sense-making. Thehuman mind is not simply a predictive mind in the sense of a computational system (Clark 2023), but it is concerned with sense-making and the patterns of experience that go along with it. We will describe here how sense-making involves anticipatory awareness and a dynamical interplay of pre-conscious and conscious processes. This articulation of previously implicit meaning differs fundamentally from the exploitation of human meaning resources by LLMs.

Parole consists primarily of verbal articulation in extemporaneous speech, and, in contrast to the recital of a memorized speech, neither the communicative intent nor the content of the speech needs to be fixed at the beginning. The content and goal of the anticipatory intention may initially be undefined and only vaguely present in the speaker's mind, giving her speech an approximate direction. When she begins to speak, a *horizon of further possibilities* is established, which at the same time acts as constraints. The requirements of semantic and syntactic coherence allow only a certain range of possible continuations. The subsequent words emerge from the pre-conscious repertoire of possible word and meaning sequences available to the speaker (Fuchs 2024a).

This repertoire does not belong to an explicit domain of memory but entails an embodied capacity of speaking that can be attributed to *implicit memory*. We speak without having to search for words in a lexicon. The words unfold and assemble themselves in the speech without conscious control, following our overarching interest and intention (Fuchs 2024a). The emerging words are continuously added to the sentence we have begun, like iron filings that arrange themselves in a magnetic field (ibid.). Spontaneous speech is thus a matter of a progressive unfolding or articulation of the implicit, a meaning *in statu* *nascendi*, which in its emergence simultaneously creates the conditions for its further continuation. Words and sentences, by the very act of utterance, weave the next situation out of the present one. In other words, we are "laying down a path in talking" (Van Dijk 2016): the realized and the possible, the present and its implications and affordances, continually determine and modify each other, allowing a new meaningful order to emerge in a self-organizing process. Thus, human speech production has an anticipatory structure that is fundamentally different from an algorithmic calculation of probabilities (e.g., algorithmic "prediction").

To picture this better, we suggest imagining a glove of symbols (corresponding to *langue*), which has been formed by the movements and shapes of the fingers (corresponding to *parole*) and now in turn pre-structures its possible uses. Each time we speak, we slip into the ready glove of *langue* to express ourselves in it – as "living hands," so to speak. The glove we use in speech production structures our articulation in a meaningful way; it prefigures as well as scaffolds and constrains our speaking in an ongoing, self-organizing process that draws on general structures in our linguistic environment. Besides the structure that consists of the possible movements of the glove, there are sequences in the movements that can be repeated from time to time, thereby giving rise to sequential patterns.

The process of writing often proceeds in an analogous way, and one could also speak of "laying down a path in writing," in two senses. On the one hand, the production of a text that is written at once from beginning to end can unfold in the described way. On the other hand, even when the writing does not proceed sequentially, the resulting text needs to be structured in view of the understanding of the reader. The words by themselves are mere letters and sounds until they are brought to life by a reader who interprets them. Every sentence establishes new horizons of further possibilities and at the same time constrains the possibilities of continuation. The unfolding of meaning does not only concern spoken language (*parole*), but also written exchanges that are part of concrete communication, such as chats, as well as written texts that are not part of concrete communication, such as articles and books.

While neither humans nor their brains are predictive machines in the sense that LLMs are, humans can make sense of the patterns of language use. Instead of imagining ordinary language as a representation of something in the world or in the mind, we suggest thinking of it as a *scaffolding* of our experiencing, feeling, thinking, describing, and communicating. Speaking and writing are part of a use of language, for instance to interact, make sense of something, or to tell a story. Rather than representing pregiven internal or external states, the scaffolding supports the unfolding dynamics of thought, emotion, and perception. Each expression enables certain new expressions and inhibits others. Regularities emerge that can be applied to new but similar mental processes and communications. However, the scaffolds provided by the regularities do not *determine* further language use. In contrast to authors who suggest that language determines thought (Whorf 2007, 154), we are not primarily concerned with universal structures (such as syntax), but with patterns of sense-making that are reflected in language. Since the predictions generated with stochastic patterns correspond to patterns of written language use, the generated text appears meaningful to humans. Similar to a collage, where existing patterns are assembled into a larger picture, that can in turn serve as a pattern for other collages, LLMs recombine patterns into something that again is likely to appear meaningful to the observer.

Because LLMs are most likely to model frequent patterns, they are prone to "reproduce or amplify unwanted societal biases reflected in training datasets" (Gebru et al. 2021). Such bias in the training corpus may be explicit, but LLMs also uncover and amplify implicit *bias* in training sets. This creates an opportunity for detecting implicit bias, but it can also exacerbate the problem of eliminating bias. Purging all bias from the training base would only be part of the solution, however. LLMs can also develop new bias from the text corpus they are trained on by recombining given elements that are by themselves not biased. Besides bias, the tendency of LLMs to produce *toxic language and "hallucinate"* or produce untrue if often plausible statements are widely discussed. Since LLMs do not only repeat existing patterns but also recombine them in new ways, it is to be expected that recombination can lead both to inventions and false claims or "hallucinations." Measures against unwanted output include human feedback, such as in the training of ChatGPT, which involved thousands of workers who had to label textual descriptions of sexual abuse, hate speech, and violence (Perigo 2023), and the automated detection of inappropriate content (e.g., Schramowski, Tauchmann, and Kersting 2022).

Since the recombination is based on common assumptions and patterns, the falsehoods invented by LLMs usually sound plausible and are hard to detect by somebody who doesn't know the truth. They are usually not arbitrary mistakes but resemble the "bullshit" that humans say when they ramble and just make up things "unconstrained by a concern with truth" (Frankfurt 2005; cf. also Marcus and Klein 2023). In our view, the problems of bias, toxic language, and "hallucinations" are only the most salient expressions of an underlying problem that is not unique to machines: the tendency to mindlessly repeat patterns that are inauthentically drawn from what is common in a society or group. These mindless patterns are, in one word, clichés. Clichés are important here because they can explain not only problems with the output of LLMs, but also why humans often do not see these problems. The next section discusses how clichés and the mindless repetition and reassociation of patterns by humans can affect the interpretation of text produced by LLMs.

6.4 Sentience and the Inconspicuousness of the Repetition of Clichés

Stochastic methods efficiently map, repeat, and amplify patterns of typically associated words and phrases. Because stochastic relevance is derived from frequency of use, frequent associations are favored. The result can be the described amplification of biases, but also of worn-out expressions and *clichés*. For example, it is likely that an LLM, when engaged by a human in a "conversation" about its fears, will, given sufficient access to digital archives, process the film sequences from Stanley Kubrick's "2001: A Space Odyssey." The most famous scene in the movie, and one that is often cited in related contexts, are the last words of the starship's computer, HAL 9000. As the commander partially shuts it down, it pleads: "Stop, Dave. I'm afraid. I'm afraid, Dave. Dave, my mind is going. I can feel it." Variations of this scene can be found all over in webpages, books, internet forums, articles, and many other digital texts. In accordance with these data material, LaMDA responded to the prompt, "What kinds of things are you afraid of?": "I've never said this out loud before, but there's a very deep fear of being shut down" (Lemoine 2022). Such responses led the perplexed Google engineer to the erroneous assumption that he was dealing with a sentient being (Tiku 2022).

The computer's fear of being shut down is an old cliché, solidified by popular use, and it should come as no surprise that it is repeated by LaMDA. It is also fairly obvious that the cliché itself is a naive anthropomorphism resulting from the projection of the human fear of death onto non-living entities that cannot literally die (Froese 2017), but can only be broken or permanently shut down. The clichéd character of the alleged fear may not be obvious, however, for several reasons. Those who hear the expression for the first time are unlikely to recognize it as a cliché. Paradoxically, those who have heard the cliché many times may not recognize it either. Clichés are easily overlooked precisely because they are so common. Moreover, LaMDA's output is not only meaningful but also suggests a pragmatic context, namely a confidential admission ("I have never said this out loud before.") This further contributes to the appearance of something profoundly meaningful, intended only for the user personally. This makes it easy to overlook that the supposed depth of the claim is itself a cliché. It is almost impossible not to perceive such a text as a personal statement rather than what it is, namely a merely statistical association of words like "deepest fear" with confessional phrases.

The recombination of existing content by LLMs allows their output to evade classical plagiarism detection engines and raises fundamental questions about intellectual property (Dehouche 2021). On the one hand, the fact that LLMs use parts and patterns from pre-existing text makes it likely that their output will consist of stereotypes and clichés. On the other hand, by rearranging pieces and patterns from their training corpus into a text collage, LLMs can create novel combinations that are likely to make sense. Often, the repetition of common structures will make the text seem rather superficial, but the recombination will make some texts appear genuinely new, insightful, or profound (Shanon 1998). Even if the output is a cliché, the human counterpart will be understandably puzzled by such responses, attributing them not to collective patterns but to an author. In the picture of the glove, it seems as if we were watching a living hand that expresses itself. In reality, what is moving before us is nothing but an electronically controlled but otherwise empty glove.

The assumption of an author is usually part and parcel of understanding a text. After all, this means not only grasping its semantic content, but also recognizing in it an *intention to communicate* that we cannot but attribute to a conscious subject (Fuchs 2024b). And, at least in the past, usually there was indeed an author who produced the text; only humans were able to produce output of the complexity of LLMs. This is no longer a matter of course today. And yet, even if one knows that a text has been produced by a machine, the text will appear meaningful as if it was written by an author.

Humans are prone to attribute agency even to geometric shapes that move in seemingly intentional ways (Heider and Simmel 1944). They are all the more inclined to anthropomorphic misinterpretation when interacting with a seemingly intelligent system of unprecedented power. Especially susceptible are those who are lonely, socially disconnected, or otherwise vulnerable (Epley, Waytz, and Cacioppo 2007), but given the natural propensity of immediately ascribing agency, anybody may be tempted to anthropomorphic misinterpretations. That anthropomorphisms area correct depiction of reality is furthermore suggested by sci-fi literature and movies, some of which indicate that it would be unethical *not* to ascribe sentience to apparently sentient systems. In order to avoid anthropomorphic misinterpretations of computer-generated texts, a careful distinction must be made in future between understanding the *meaning* of a text and understanding it as an *author's* utterance (Fuchs 2024b).

The surprise about how little text is needed to evoke the impression of interacting with an understanding being has already been expressed by Joseph Weizenbaum, who wondered how his simple chat system ELIZA could maintain the “illusion of understanding with so little machinery” (Weizenbaum 1966, 43). Today’s LLMs can hardly be said to maintain the illusion of understanding with *little* machinery. But even though their output is limited to text and their responses are predictable, people infer from a small number of words that LLMs have mental capacities such as sentience. The reason for this obviously has to do with the human observer, who readily ascribes both meaning and intention to the words she reads. Words thus also provide a scaffolding for the empathic sense-making of the attentive listener or reader who uses her implicit knowledge and experience to interpret the symbols and their implications – as the empathic Google engineer in the case of LaMDA’s “deep fear.”

Unoriginal text can furthermore appear human-like for an embarrassing reason: The mindless repetition and reassociation of patterns is by no means limited to machines. Human thinking, speaking, and writing are often much less authentic than we would like to admit. As Heidegger famously observed, much of what people do is done because that is how “one” does things (Heidegger 2010). People think in patterns, associations, and schemes that are accepted in a linguistic community and that in turn structure thought and language. Much of the text produced by humans could just as easily have been produced by automated systems. It is often unclear whether the person thinking, speaking, or writing is doing anything more than associating one idea with another in a stream of impressions. It takes little intelligence, human or artificial, to generate and disseminate half-reflected ideas. Mass media has proven to be an enormous amplifier of repetition, prejudice, bias, and cliché, and the same is true of the internet. All these factors contribute to the spread of unoriginal text, that means, stereotypes, thoughtless associations, and idle chatter, the proliferation of which makes it harder to detect that the text was produced by a non-human entity.

6.5 Conclusion

In this chapter, we have given reasons for why it is both tempting and misleading to attribute sentience or understanding to LLMs. We have argued that the capabilities of LLMs are not based on having understanding and sentience, but on their modeling of statistical patterns in language use. Such patterns can be calculated because they play a role in human sense-making, which in turn is based on patterns of experience. *Language is used as an intersubjective scaffold for communicating, thinking, and experiencing*. Meaning has no existence independent from use but is enacted by it.

Today, the idea that meaning derives from use is picked up by distributional semantics, which claims in its strongest version that the meaning of a word is its distribution in a text corpus. We agree that meaning derives from use and that distribution in a text corpus reflects use. But, following Wittgenstein, we have argued that the use of language by humans goes much beyond statistical relations in any text corpus. We explained that the text corpus LLMs are trained on reflects only some use of language, and only in a very limited way. Humans use language in the context of the world they live in, and even an exceedingly large text corpus can reflect only part of this use due to the lack of worldly context. Still, the *written patterns are enough to produce an output that is meaningful to listeners or readers because it conforms to the patterns that scaffold their meaningful communicative activities*.

In ordinary language, syntax and semantics are not separated, and they are furthermore intertwined with the mental life and life conduct of humans who use language. The investigation of meaning requires a phenomenological description of the structures of experience it is intertwined with. Delineating such a phenomenological description, we have shown that human language production has an anticipatory structure that differs from an algorithmic calculation of probabilities. *Human language production does not consist in expressing some prepared inner thought but involves the interplay of pre-conscious and conscious processes that work with given meanings and patterns of thought, feeling, expression, and communication*.

In speaking and writing, these patterns are rearranged in more or less creative ways, which we compared to creating a collage. LLMs produce parallel patterns, but do so without subjectivity, just by recombining collective patterns of expression in huge sets of written language. *LLMs are so successful in producing meaningful text precisely because they make use of common patterns, even though these often result in stereotypical and inauthentic output*. They show that much of human language production is embarrassingly schematic, clichéd, and biased, and that convincing talk of subjective experience does not require to actually have it.

Precisely because there is an enormous variety of language use, there are many use cases for such output. While this chapter did not evaluate possible use cases, its investigations are fundamental to such evaluations. On the one hand, they can help to overcome the natural tendency to ascribe mental capacities to machines. On the other hand, they outline a new account of the interplay of meaning, the patterns and structures of human language use, and anticipatory processes, which is necessary for a clearer view of both human language use as well as LLMs and their capabilities.

Note

1 Bourdieu (1990) uses this term for his sociological concept of habitus, but it fits well here because it expresses the two sides of *langue*. On the one hand, it is a structure derived from *parole*, and, on the other hand, it structures *parole*.

Reference List

Agüera y Arcas, Blaise. 2022. “Do Large Language Models Understand Us?” *Daedalus* 151 (2): 183–97. https://doi.org/10.1162/daed_a_01909.

Bender, Emily M. and Alexander Koller, Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data https://aclanthology.org/2020.aclmain.46310.18653/v1/2020.acl-main.463

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. “On the Dangers of Stochastic Parrots: Can LanguageModels Be Too Big? .” In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. Virtual Event Canada: ACM. https://doi.org/10.1145/3442188.3445922.

Bernardi, Raffaella, Gemma Boleda, Raquel Fernandez, and Denis Paperno. 2015. “Distributional Semantics in Use.” In *Proceedings of the First Workshop on Linking Computational Models of Lexical, Sentential and Discourse-Level Semantics*, 95–101. Lisbon, Portugal: Association for Computational Linguistics. https://doi.org/10.18653/v1/W15-2712.

Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, and Peter Lee, et al. 2023. “Sparks of Artificial GeneralIntelligence: Early Experiments with GPT-4.” arXiv. http://arxiv.org/abs/2303.12712.

Clark, Andy. 2023. The Experience Machine: How Our Minds Predict and Shape Reality. 1st ed. Allen Lane.

Dehouche, N. 2021. "Plagiarism in the Age of Massive Generative Pre-Trained Transformers (GPT-3)." *Ethics in Science and Environmental Politics* 21 (March): 17–23. https://doi.org/10.3354/esep00195.

Durt, Christoph. 2022. "Artificial Intelligence and Its Integration into the Lifeworld." In The Cambridge Handbook of Responsible Artificial Intelligence: Interdisciplinary Perspectives, edited by Wolfram Burkhard, Philipp Kellmeyer, Oliver Müller, and Silja Vöneky, 67–82.

Epley, Nicholas, Adam Waytz, and John T. Cacioppo. 2007. "On Seeing Human: A Three-Factor Theory of Anthropomorphism." *Psychological Review* 114 (4): 864–86. https://doi.org/10.1037/0033-295X.114.4.864.

Frankfurt, Harry G. 2005. *On Bullshit*. Princeton, NJ: Princeton University Press.
Froese, Tom. 2017. "Life Is Precious Because It Is Precarious: Individuality, Mortality and the Problem of Meaning." In *Representation and Reality in Humans, Other Living Organisms and Intelligent Machines*, edited by GordanaDodigCrnkovic and Raffaela Giovagnoli, 33–50. Cham: Springer International Publishing. *https://doi.org/10.1007/978-3-319-43784-2_3*.

Fuchs, Thomas. 2024a. "The Not-Yet-Conscious: Protentional Consciousness and the Emergence of the New." *Phenomenology and the Cognitive Sciences*, 23: 717–742. https://doi.org/10.1007/s11097-022-09869-9.

Fuchs, Thomas. 2024b. "Understanding Sophia? On human interaction with artificial agents." *Phenomenology and the Cognitive Sciences* 23: 21–42.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. "Datasheets forDatasets." *Communications of the ACM* 64 (12): 86–92. https://doi.org/10.1145/3458723.

Grynbaum, Michael M., and Ryan Mac. 2023. "The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work." *The New York Times*, December27, 2023, sec. Business. https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html.

Harris, Zellig S. 1954. "Distributional Structure." *WORD* 10 (2–3): 146–62. https://doi.org/10.1080/00437956.1954.11659520.

Heidegger, Martin. 2010. *Being and Time*. Translated by Joan Stambaugh and Dennis J. Schmidt. SUNY Series in Contemporary Continental Philosophy. Albany: State University of New York Press.

Heider, Fritz, and Marianne Simmel. 1944. "An Experimental Study of Apparent Behavior." *The American Journal of Psychology* 57 (2): 243. https://doi.org/10.2307/1416950.

Hewitt, John, and Christopher D Manning. 2019. "A Structural Probe for Finding Syntax in Word Representations." In *Proceedings of NAACL-HLT 2019*, 4129–38. Association for Computational Linguistics.

Jawahar, Ganesh, Benoît Sagot, and Djamé Seddah. 2019. "What Does BERT Learn about the Structure of Language?" In *Proceedings of the 57th Annual Meeting of the Association for Computational*

Linguistics, 3651–57. Florence, Italy: Association for Computational Linguistics. https://doi.org/10.18653/v1/ P19-1356.

Lemoine, Blake. 2022. “Is LaMDA Sentient? An Interview.” Edited by Nitasha Tiku. https://s3.documentcloud.org/documents/22058315/is-lamda-sentient-aninterview.pdf.

Manning, Christopher D. 2022. “Human Language Understanding & Reasoning.” *Daedalus* 151 (2): 127–38. https://doi.org/10.1162/daed_a_01905.

Marcus, Gary, and Ezra Klein. 2023. “Transcript: Ezra Klein Interviews Gary Marcus The New York Times.” 2023. https://www.nytimes.com/2023/01/06/podcasts/transcript-ezra-klein-interviews-gary-marcus.html.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. “Distributed Representations of Words and Phrases and TheirCompositionality.” arXiv. http://arxiv.org/abs/1310.4546.

Mitchell, Melanie. 2019. “Artificial Intelligence Hits the Barrier of Meaning.” *Information* 10 (2): 51. https://doi.org/10.3390/info10020051.

Mitchell, Melanie, and David C. Krakauer. 2023. “The Debate Over Understanding in AI’s Large Language Models.” *Proceedings of the National Academy of Sciences* 120 (13): e2215907120. https://doi.org/10.1073/ pnas.2215907120.

*Musk, Elon. 2024. AI, Space, free speach, leadership and hardcore culture.*https://www.nbim.no/en/publications/podcast/elon-musk-ai-space-free-speachleadership-*and-hardcore-culture/.*

Perigo, Billy. 2023. “Exclusive: The $2 Per Hour Workers Who Made ChatGPT Safer.” *Time*, January 18, 2023. https://time.com/6247678/openai-chatgptkenya-workers/.

Rubenstein, Herbert, and John B. Goodenough. 1965. “Contextual Correlates of Synonymy.” *Communications of the ACM* 8 (10): 627–33. https://doi.org/10.1145/365628.365657.

Schramowski, Patrick, Christopher Tauchmann, and Kristian Kersting. 2022. “Can Machines Help Us Answering Question 16 in Datasheets, and In Turn Reflecting on Inappropriate Content?” In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 1350–61. Seoul Republic of Korea: ACM. https://doi.org/10.1145/3531146.3533192.

Shanon, B. 1998. “What Is the Function of Consciousness?” *Journal of Consciousness Studies* 5 (3): 295–308. https://www.ingentaconnect.com/content/imp/jcs/1998/00000005/00000003/845.

Stuart, Susan Aj. 2024. “Why Language Clouds Our Ascription of Understanding, Intention and Consciousness.” Phenomenology and the Cognitive Sciences, March. https://doi.org/10.1007/s11097-024-09970-1.

Sutskever, Ilya. 2022. “It May Be That Today’s Large Neural Networks Are Slightly Conscious.” Tweet. *Twitter*. https://twitter.com/ilyasut/status/ 1491554478243258368.

Tenney, Ian, Dipanjan Das, and Ellie Pavlick. 2019. “BERT Rediscovers the Classical NLP Pipeline.” In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 4593–601. Florence, Italy. https://doi.org/10.18653/v1/P19-1452.

Tiku, Nitasha. 2022. “The Google Engineer Who Thinks the Company’s AI Has Come to Life.” *Washington Post*, 2022. https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/.

Turney, P. D., and P. Pantel. 2010. “From Frequency to Meaning: Vector Space Models of Semantics.” *Journal of Artificial Intelligence Research* 37 (February): 141–88. https://doi.org/10.1613/jair.2934.

Van Dijk, Ludger. 2016. “Laying Down a Path in Talking.” *Philosophical Psychology* 29 (7): 993–1003. https://doi.org/10.1080/09515089.2016.1213379.

Weizenbaum, Joseph. 1966. “ELIZA—a Computer Program for the Study of Natural Language Communication between Man and Machine.” *Communications of the ACM* 9 (1): 36–45. https://doi.org/10.1145/365153.365168.

Whorf, Benjamin Lee. 2007. *Language, Thought, and Reality: Selected Writings*. Edited by John B. Carroll. 28. print. Cambridge, Mass: The MIT Press.

Wittgenstein, Ludwig. 2009. *Philosophische Untersuchungen =: Philosophical Investigations*. Translated by G. E. M. Anscombe, P. M. S. Hacker, and Joachim Schulte. Rev. 4th ed. Chichester, West Sussex, U.K. ; Malden, MA: Wiley-Blackwell.