What Does It Mean When We Say That The Meaning of a Word Is Its Use?

Yufa Zhou
Duke University
yufa.zhou@duke.edu

Abstract

This essay revisits Wittgenstein's claim that meaning is not grounded in inner representations but in use, and argues—drawing on Stanley Cavell—that such use is essentially normative and bound up with judgment and mutual accountability. Meaning, on this view, does not derive from a hidden framework of rules or a conceptual scheme, but is immanent in our talk and action within a shared form of life. Against this background, I examine contemporary large language models (LLMs), which acquire linguistic competence by learning statistical regularities across vast corpora. While these systems display striking fluency across many language games, their successes and failures reveal a gap between reproducing patterns of use and participating in the practices that sustain meaning. This distinction helps clarify both the power and the limits of current AI systems, and to rethink important questions about understanding modern AI.

1 The Old Picture: The Meaning–Word Gap

Classical picture of language, exemplified by Augustine's picture at the opening of *Philosophical Investigations* [Wit09], treats meaning as something that exists prior to language. **Words are taken to name inner ideas, and sentences simply combine these names.** This view implies a very specific model of communication: each speaker privately associates words with mental images or concepts, and successful communication occurs when the listener reconstructs the same inner contents. On this picture, **learning a language looks like mapping sounds to pre-existing ideas**—like attaching labels to items in a mental inventory. If I utter "tree," I supposedly point you to the idea of *tree* already stored in your mind; if you lack that idea, the word simply has no meaning for you.

Wittgenstein illustrates this with the "five red apples" example. On the classical naming model, understanding this request would mean linking *"five," "red,"* and *"apples"* to corresponding inner notions and then combining them. The model works tolerably well for concrete nouns like *apple*, where one can imagine a stable signifier–signified pair: a word paired with a mental picture of an object. But Wittgenstein stresses that **this is only *one* kind of language-game**. The model breaks down for most of our vocabulary—verbs, adjectives, adverbs, number-words, words for abstract concepts, expressions like "now," "tomorrow," "pain," "game," "good," or "understand." These do not neatly correspond to pre-formed inner pictures, and we in fact use them in a wide range of activities with no single mental image tying them together.

Recognizing this exposes the deeper problem: **if meaning is grounded in private inner representations, then the signifier–signified relation becomes inherently unstable.** There is always a gap between the word I utter and the inner image you attach to it. Nothing guarantees that your concept of *apple*, *five*, or *good* matches mine. The classical picture thus *cannot* explain how shared meaning or public criteria for correct use are even possible—a tension Wittgenstein uses to motivate his shift from "inner ideas" to meaning as use in practice.

2 Wittgenstein's Critique: Meaning Is Use

Wittgenstein's later philosophy rejects the old representational model. He argues that meaning does not sit behind words like a mental object waiting to be named. **Meaning *emerges from use*—**from how people actually deploy words in the flow of everyday activities. To understand a word is not to look inward, but to know how to use it correctly within a community. **Meaning, then, is public, social, and practical, anchored in what Wittgenstein calls our "forms of life."**

To make this vivid, imagine a small group of people trying to create an entirely new language from scratch. At first, they establish a few basic words by pointing and acting: someone holds up a stone and says "dor," touches water and says "lin." These ostensive moves get the game started, but they are only the scaffolding. As they work together—building a shelter, navigating a river, teaching a child—they invent new expressions on the fly. One person says "dor-lin!" to warn that the stones are slippery; another repurposes "lin" to mean "cool down." Over time, the group acquires ways of giving orders, making promises, joking, arguing, and correcting one another. The vocabulary thickens, old words branch into new uses, and new terms arise to handle new situations. **Meaning here is dynamic, negotiated, and multiply layered. No inner picture guarantees it; *practice* does.**

Seen this way, the old "word–meaning gap" evaporates. Language is not a mirror of private mental states, but part of the public practices through which such states are expressed, recognized, and responded to. The representational picture held us captive because it made us stuck in the wrong place—inside the mind instead of at our shared forms of life—so we must simply let that picture go.

3 Language Games: How Words Gain Meaning from Context

Wittgenstein's notion of *language games* marks a decisive break from the classical picture of meaning. A language game is not a formal grammar or a rigid code. It is a **way of acting with words** within a shared form of life. Each game involves characteristic activities—asking a question, giving an order, joking, warning, comforting—and each activity comes with its own gestures, expectations, and tacit rules.

A simple example shows why meaning cannot be fixed inside the word itself. Consider the word *"check."* Shouted across a chessboard, it warns the opponent. At a restaurant, it requests the bill. At a bank, it designates a financial document. The sound remains the same, but its meaning shifts with the **practical situation** in which it is used. Meaning is not stored as an inner mental content; it is enacted through use. This is the core of Wittgenstein's insight: **words gain significance from the roles they play in human practices, not from any underlying representation.**

3.1 Family Resemblances and the Limits of Models

Because meaning depends on use, Wittgenstein argues that most concepts lack sharp, essential boundaries. They resemble families—sets of overlapping similarities where no single feature is shared by all members. This is why models so easily mislead us. We expect a theory to capture the "essence" of things, as if language must match reality like a blueprint. Wittgenstein instead treats models as **yardsticks**: helpful comparisons that orient us, not reveal hidden structures.

This connects directly to language games. Our uses of words overlap in messy, evolving ways. Consider "game": some games involve luck, some skill, some competition, some none of these. A theory attempting to define "the essence of gamehood" will inevitably distort the actual diversity of our practices. A better philosophical method is comparative—to illuminate similarities and differences without forcing a single mold onto all cases.

This shift also alters the metaphysical picture. Philosophy does not uncover metaphysical objects lurking behind our words. It clarifies how we speak, where confusion creeps in, and which misleading pictures tempt us. Rules and definitions cannot guarantee absolute precision because rules themselves require interpretation. The perceived "gap" between words and meanings dissolves once we stop expecting language to operate by rigid internal correspondences.

3.2 The Private Language Argument: Meaning Without Inner Objects

Wittgenstein's private-language argument offers a vivid demonstration of this idea. In the "beetle in the box" thought experiment, each person has a box no one else can inspect. Everyone calls whatever is inside a "beetle." The contents might differ radically—or the boxes might be empty. Yet the word "beetle" functions perfectly well in ordinary speech.

The lesson is clear: **the meaning of the word does not depend on a private inner object**. Meaning comes from public use—how people talk, correct each other, coordinate actions, or respond within a community. A purely private reference point can vanish without affecting communication at all.

A modern analogy makes this intuitive. Online communities constantly generate slang, memes, and coded expressions. Outsiders hear only noise; insiders instantly grasp the meaning because they share the practices in which the terms are used. Over time, even the original reference of the slang may disappear, yet the word continues to function. The "thing in the box" drops out, and the shared practice remains. For Wittgenstein, this is the key: **language is possible because of public criteria, not private objects**.

3.3 Pain: Neither a Hidden Object nor a Mere Nothing

This framework helps clarify Wittgenstein's remarks on sensations such as pain. Grammar tempts us to treat nouns as naming things. So we ask: *What is pain? Where is it located? Is it a Something or a Nothing?* Wittgenstein argues that this entire way of posing the question is shaped by misleading linguistic habits.

When someone says "I'm in pain," they are not identifying a mysterious inner entity. They are participating in a recognizable practice—crying out, asking for help, explaining their behavior. Pain has no private, object-like essence hidden inside a mental chamber; yet it is not unreal either. Its significance lies in the **patterns of feeling, expression, response, and acknowledgment** that structure human life. A "Nothing" would fill the same linguistic role as a metaphysical "Something" about which nothing further can be said. Once we let go of the idea that words must name objects, the paradox evaporates.

Pain, then, is neither an inner object nor a philosophical illusion. It is a phenomenon whose meaning emerges from our shared practices of life.

3.4 What Language Games Reveal About Meaning

Taken together, these arguments show that meaning is neither a private possession in our heads nor a metaphysical entity hovering behind our words. We **inherit** it from our community, **perform** it in our actions, and **reshape** it as contexts shift. Language functions across many overlapping language games, and significance emerges only within the forms of life in which those uses are learned and sustained. This is why definitions alone cannot secure meaning, and why rules cannot eliminate ambiguity. What holds language together is not an internal mental code, but the ways we live, act, train, and respond together. Instead of looking behind words for hidden metaphysical items, Wittgenstein invites us to observe how we actually use language—how our linguistic activities form a rich, interwoven fabric. Meaning is not behind the word; it lives in its use within a shared form of life.

4 Beyond Use Alone: Criteria and Normativity

However, it is not enough to say that meaning arises from use. Use must itself be understood as a constant act of **judgment**. This is where Wittgenstein's notion of *criteria*, as developed by **Cavell** [Cav99], becomes essential. Criteria are the shared standards that allow us to say a word was used rightly or wrongly, appropriately or irresponsibly. Criteria are the shared agreements in our form of life that are **revealed within** everything we say. Criteria are not abstract rules but the shared ways in which we recognize whether someone's words make sense, whether they fit the situation, whether they count as an answer, a promise, a refusal, or an excuse.

For example, when someone says "I want to go shopping," or asks "How could she say that to me?", we rely on criteria to understand what counts as wanting, what counts as saying something hurtful, what counts as speaking sincerely or evasively. These criteria belong to our ordinary practices; they show themselves in our reactions and expectations. In these moments, intelligibility depends on our attunement to the context and the specific "point" of the utterance. Thus, **use reveals criteria**. To speak is to project our shared criteria, and our agreement in criteria is nothing deeper than our agreement in judgment itself.

Cavell captures the essence of the skeptical impulse when he describes the temptation to "empty out my contribution to words, so that language itself, as if beyond me, exclusively takes over the responsibility for meaning" [Cav69]. The skeptic dreams of an "ideal" language—a fixed framework—that works automatically, without the risk of human error. Because ordinary speech relies on our fragile, moment-to-moment judgment, the skeptic finds it questionable, arbitrary, or "mediocre". We can liken this to a book sitting on a shelf: we might assume the story exists perfectly within the binding, independent of us. But without a reader to engage with it, the book is merely physical paper and ink; it possesses no intelligibility on its own. Language does not guarantee its own meaning any more than a book reads itself. Just as a reader must actively bring the text to life, we must "project" our own grounds in every act of speech. Meaning is not a pre-existing object we find; it is a responsibility we must continuously shoulder.

Crucially, Affeldt argues that this normativity—the "force" that makes language binding rather than arbitrary—does not derive from a rigid "framework of rules" or a "conceptual scheme" [Aff98]. To think so is to imagine that grammar dictates meaning to us from the outside. Instead, the normativity of language is "simply in our talk and action". It is grounded in our mutual accountability: the fact that we hold each other responsible for what we say. When we employ a concept, we are not merely triggering a rule; we are entering a space of shared expectations where we can be challenged, corrected, or found unintelligible by others. The "must" of grammar—why we *must* mean this if we say that—is not a logical compulsion from a hidden rulebook, but a necessity born of our shared form of life. To be a speaker is to accept the burden of this responsibility, a burden the skeptic seeks to evade by fantasizing that the system can do the judging for them.

5 LLMs and Modern AI: How Do They Learn Meaning From Data?

A large language model (LLM) can be formally characterized as a conditional probability distribution over sequences of tokens. Given a prefix $x_{<t} = (x_1, x_2, \ldots, x_{t-1})$, the model assigns a probability distribution to the next token x_t,

$$x_t \sim \Pr(x_t \mid x_{<t}; \theta),$$

where θ denotes parameters learned during training. Training proceeds by maximizing the likelihood of observed sequences in a large corpus, enabling the model to capture the statistical regularities that govern how sentences, paragraphs, and longer texts typically continue. In this sense, an LLM operates by predicting the next word conditioned on prior context, gradually refining its internal parameters to model patterns of linguistic use at scale.

A crucial feature of such a model is its sensitivity to context. Any change in the preceding words reshapes the model's expectations about what is likely to come next. As a result, the model does not store fixed dictionary-like meanings. Instead, it continually recalibrates its output based on the evolving linguistic environment. This explains both its remarkable strengths—such as its ability to produce coherent arguments, instructions, or stories—and its weaknesses. Whenever genuine understanding depends on factors not present in the prompt, such as user intentions, emotions, embodied cues, or background knowledge that humans typically share, the predictive machinery can misfire.

Viewed through a Wittgensteinian lens, LLMs learn meaning by immersing themselves in patterns of use. With no explicit semantic grounding or predefined rules, they become fluent participants in many language games: answering questions, offering advice, debating, joking, translating, and narrating. Their success appears to support the idea that meaning arises not from abstract definitions or mental representations, but from the ways words function within activity.

Yet their failures are just as revealing. If meaning were nothing more than use, and LLMs are masters of patterned linguistic use, then why do they still hallucinate, misapply norms, or give inconsistent moral or epistemic judgments? These errors point to an important distinction: reproducing the surface patterns of a practice is not the same as being embedded in the **form of life** that gives those patterns their significance. LLMs can mimic participation in a practice, but they do not inhabit the social, moral, or experiential contexts in which human understanding is rooted.

Their competence is real, but it is a competence of prediction rather than of lived engagement. And this gap forces us to refine what "use" itself means in Wittgenstein's framework—use is not mere statistical patterning but use within the lived human activities that give language its point.

6 Language Games and Artificial Intelligence

One way to think about intelligence—both human and artificial—is through the **range and complexity of the language games** an agent can competently participate in. Animals like dogs or cats may not possess full-fledged linguistic systems, yet they can engage in very simple language games: recognizing an object when named, responding to commands, or reacting to human gestures. Human linguistic ability expands dramatically through education and experience. Children begin with elementary games such as naming objects or following simple instructions; adults gradually acquire the capacity to participate in far more abstract games involving mathematics, physics, ethics, humor, or emotional nuance. The expansion of one's linguistic repertoire reflects the expansion of one's forms of life.

A similar gradient appears in the development of large language models. Early models, such as GPT-3, struggled even with basic instruction-following. Later systems—GPT-4, Claude, GPT-5—display sophisticated capacities that resemble graduate- or doctoral-level proficiency in many domains. Their

"intelligence," however artificial, seems to scale with the **variety of language games** they can imitate, from technical explanation to legal reasoning to emotional support.

Wittgenstein's distinction between *primitive languages* and full-fledged *language games* helps illuminate this. Primitive languages—like the builder's calls for bricks or the symbolic manipulations of formal logic—are narrow systems with explicit rules and clear goals. These are precisely the domains where LLMs excel: programming, mathematics, classification, translation. The structure is rigid and the standards of success are well defined. Prediction-based models thrive here because the relevant patterns are stable and the rules are explicit and relatively stable.

But language games in the broader Wittgensteinian sense are far more demanding. They involve shifting context, tacit background knowledge, cultural norms, emotional expectation, and subtle cues—humor, sarcasm, irony, metaphor, consolation. These games do not have fixed rules; their "correctness" depends on how well one's words resonate with shared human practices. For this reason, LLMs often falter. They can reproduce the surface patterns, but the social, embodied, and cultural grounding that gives those patterns their meaning is missing.

This asymmetry also explains an intriguing contrast: **AI often excels at mathematics, while many humans find mathematics difficult**, yet humans grasp humor, tone, and social nuance much more naturally. Mathematics is a tightly constrained practice with explicit rules and minimal dependence on context, as it is self-contained by construction. Most of the relevant information is present in the written symbols themselves; inference proceeds within a closed system. By contrast, human social interaction relies on vast, shifting, implicitly learned contexts—childhood experience, cultural memory, embodied perception, emotional intelligence. LLMs, lacking these forms of life, find such games harder to play.

This reveals something important about working with models: providing the **right context** is often the decisive factor in achieving good performance. Humans can bring unspoken background knowledge to a conversation automatically; LLMs cannot. To make an AI system behave reliably, users must carefully supply the contextual scaffolding that humans ordinarily take for granted.

In this sense, part of the art of using an LLM is learning how to "teach" it the language game you want it to play—how to specify the background, constraints, roles, and purposes that allow the model's predictive machinery to function effectively. Much of what appears as intelligence in these systems is really competence within the boundaries of the language game we construct for them.

7 Can LLMs Judge? And the Alignment Question

What LLMs seem to lack currently is what Cavell calls *criteria*. As Affeldt emphasizes, we must not confuse the "grammatical framework" of language with the human ground of intelligibility. What is missing in LLMs is not more data or more patterns of use, but participation in the shared forms of life in which criteria are lived: the ability to recognize sincerity, evasiveness, reassurance, insult, comfort, or responsibility in another's words.

Cavell and Affeldt clarify that criteria are not rules. They are our shared judgments of the *right use* of words—judgments rooted in our lives as embodied, responsive, mutually accountable human beings. LLMs can reproduce linguistic uses with remarkable fluency, but they do not yet *judge*. They do not respond as one who has been trained within a practice, corrected by others, or held responsible for what they say.

This leads to a deeper question: **Can mere exposure to linguistic patterns ever amount to genuine participation in shared meaning, or do shared meanings depend on specifically human capacities that make criteria possible?**

From the perspective of AI alignment, this becomes the central problem. The challenge is not simply to supply models with the "correct" standards. The philosophical worry is whether contemporary AI

systems are capable of anything that could count as *judging*, and how such a capacity—if it emerged—could be recognized or tested.

At present (2025), AI systems like GPT-1 can behave humorously or pick up emotional cues, yet they remain passive participants in language. They do not initiate actions or questions; they speak only when spoken to. This sharpens rather than resolves the issue. Do they genuinely grasp humor, comfort, or tone, or do they merely reproduce patterns that mimic human responses? Can they act, play, and actively *judge*? Their apparent sensitivity to emotion only deepens the puzzle.

These questions lead directly to practical concerns for alignment. Some researchers hope that methods such as reinforcement learning—where an AI acts, receives feedback, and attempts to improve—might offer something closer to participation in a practice rather than mere imitation of linguistic patterns. Could such training bring AI systems nearer to the human capacities that underlie criteria? Or is there something irreducibly human in those capacities that no amount of reinforcement can provide?

Thus, I argue that for LLMs to transcend the status of a static "framework of rules" and achieve genuine intelligibility, they must move beyond passive response. They require **active learning and self-improvement** mechanisms that allow them not just to mimic patterns, but to "project" a position within a dialogue. By actively communicating and continuously refining their own internal states (their weights) in response to others, they might begin to move away from letting "language itself… exclusively take over the responsibility for meaning" [Cav69] and start assuming that responsibility themselves. A system that merely reacts perfectly remains a superintelligent instrument—a "tool man" operating within a fixed Mulhallian framework. Whether a system could become a "vivid" participant in our forms of life without embodiment, or without inclusion in a genuinely human form of life, remains an open question—one this essay does not attempt to resolve.

8 What Counts as Human and AI?

This brings us to a final question: if LLMs can mimic so many aspects of human linguistic life, what counts as intelligence or as human? *Blade Runner* reminds us that this boundary is not merely a matter of inner qualities. The replicants speak like us, feel like us, and even suffer like us, yet they are treated as "just machines." Their exclusion is not based on what they are, but on how society chooses to **use** them—much like the way human beings have historically reduced other humans to slave labor. The film shows that the boundary between human and non-human is not discovered but **drawn**, shaped by our practices, interests, and refusals.

The same question feels sharper today. Contemporary AI systems can deploy the language of warmth, empathy, and emotional support—often with greater consistency than humans—raising the question of whether such uses amount to meaning, or merely to successful imitation. Their designers now intentionally limit such human-likeness, precisely to prevent people from forming deep emotional attachments.

Yet a paradox emerges. AI is always available, never angry, never manipulative, and reliably helpful. In the short term, this can feel ideal, like the perfect companion. But its limitations soon become clear: it lacks initiative, desire, and genuine disagreement. It does not surprise us, challenge us, or insist on its own point of view. Over time, this begins to resemble a kind, steady friend who never truly grows, never pushes back, and never moves unless prompted.

What this reveals is not simply a technological limitation but a conceptual one. To count as human—or even as a partner in meaning—may require more than linguistic fluency. It may require the capacity to judge, to act, to refuse, to correct, to be responsible: all the things that Cavell, Wittgenstein, and Affeldt gather under the idea of *criteria*.

9 Conclusion

Wittgenstein shows that use provides the fundamental basis of meaning, and Cavell reminds us that such use is always already a projection of **criteria**—the normative force that makes meaning shared and accountable. Modern LLMs vividly illustrate the distinction between **mere pattern-matching and genuine participation**: they master patterns of use with extraordinary fluency, yet falter when meaning demands judgment, justification, and value-sensitive interpretation. Their successes reveal how far mere use can take us; their failures reveal how much of our linguistic life depends on capacities rooted in human forms of life.

Meaning, therefore, begins in use, but its depth—its accountability, seriousness, and ethical significance—emerges only within the shared practices in which criteria become visible. This essay has aimed to explore this question by connecting Wittgenstein's insights with contemporary AI, and by offering my own interpretation of how use, criteria, and shared practices jointly constitute meaning.

References

[Aff98] Steven G Affeldt. The ground of mutuality: criteria, judgment and intelligibility in stephen mulhall and stanley cavell. *European Journal of Philosophy*, 6(1):1–31, 1998.

[Cav69] Stanley Cavell. *Must We Mean What We Say?: A Book of Essays*. Cambridge University Press, 1969.

[Cav99] Stanley Cavell. *The Claim of Reason: Wittgenstein, Skepticism, Morality, and Tragedy*. Oxford University Press USA, 1999.

[Wit09] Ludwig Wittgenstein. *Philosophical Investigations*. Wiley-Blackwell, 4th edition, 2009.