Beyond the Imitation Game: The Failure of LLMs to Participate in the Social Practice of Giving and Asking for Reasons

Bahareh Izadi
Graduate Student in Department of Philosophy, Concordia University
Montreal, Quebec, Canada

Paper presented at the International Conference on ***Reasons, Rationality, and Rationalism***
University of Zurich, September 12th, 2025

ABSTRACT

Since ELIZA in 1966, the performance of chatbots has evolved considerably. Chatbots such as PARRY, Jabberwacky, GPT, and others were designed to generate suitable responses to user prompts. Recent advances in Large Language Models (LLMs), such as the signs of primitive "metacognitive" circuits in Claude 3.5 Haiku that allows the model to know the extent of its own knowledge, have demonstrated unprecedented sophistication in text generation, prompting some to suggest LLMs will even exceed the limits of Artificial General Intelligence (AGI) (Altman, 2025).

In this ongoing research of my thesis, I argue that despite their capabilities, LLMs fundamentally cannot participate in the practice of "giving and asking for reasons" (Brandom 1994, 5)—the normative practice that constitutes rationality.

Drawing on Inferentialism, I argue that this incapacity stems from two interrelated limitations: (1) current technical constraints inherent in the transformer architecture's approach to language as probabilistic prediction rather than normative practice, and (2) LLMs constitutive inability to achieve an autonomous intentional membership in normative communities where participants mutually recognize each other's authority to make claims and be held responsible for them. I argue that the normative dimension of language is irreducibly social, emerging through practices of mutual recognition that transcend statistical pattern recognition.

I presuppose and do not defend an inferentialist approach to meaning and rationality. I do not address the question whether human consciousness is the only form of consciousness, and whether we can call LLMs conscious. Rather, the importance of my research lies in proving that these magnificent tools which cannot be held *responsible* to provide reasons for their outputs will always need the help of a being that can be held responsible.

This interpretation reveals a categorical distinction between statistical simulation and genuine discursive practice—a distinction that establishes limitations on Artificial Intelligence (AI). The implications extend beyond technical questions about AI to inquiries concerning the nature of rationality, normativity, and the societal aspect of meaning.

I. CONTEXT OF THE ARGUMENT

There has been a discussion going on regarding the assessment of LLMs. This gives rise to a dialectic between mainly two competing views—one that privileges outputs of LLMs and another that insists on the mechanistic interpretations concerning internal processes of LLMs (Chalmers, 2025). Central to this tension is a question about the nature of having the ability to utilize the language, as a participant in the social practice of giving and asking for reasons: How can this participation be granted to LLMs? Is it fundamentally behavioral, emerging from the pattern recognition behind the responses to stimuli, or is it irreducibly normative, emerging from social practices of commitment-undertaking and responsibility-bearing (Turing 1950; Searle 1980; Brandom 1994)? This question takes on an urgency as we confront technologies that can simulate human linguistic performance while potentially lacking the intentionality thought necessary for any genuine participation; mainly in the social practice of giving and asking for reasons (Gubelmann, 2024).

A. BEHAVIORAL INTERPRETABILITY

At the extreme end of this spectrum, strong behavioralism adjudicates a LLM's participation in discursive practice by assessing its capacity to perform intelligent linguistic tasks—a methodological commitment that has proven enormously productive in AI research. This approach suggests that a system capable of producing appropriate linguistic outputs might reasonably be described as a discursive creature. Research on LLMs often implicitly adopts this framework, focusing on performance metrics such as contextual appropriateness rather than ontological questions about what these systems *really* are (Arai and Tsugawa, 2023).

B. MECHANISTIC INTERPRETABILITY

Language, intrinsically, is not simply about producing strings of symbols in statistically appropriate patterns but about undertaking commitments and acknowledging responsibilities within a community of reason-givers and reason-takers (Brandom, 1994). To assert a proposition is not merely to utter a declarative sentence but to undertake a commitment to its truth and to take responsibility for the inferences it licenses. This normative dimension of language use points toward the essentially social character of linguistics (Wittgenstein, 1953). Some believe that if a being lacks this dimension, it fails in participating in giving and asking for reasons.

C. INFERENTIALIST APPROACH

MASTERY OF INFERENTIAL ROLE

Inferentialism elaborates this by understanding conceptual content in terms of inference rather than reference—what matters is not how words hook onto the world but how assertions fit within a network of inferential relations. This approach opens possibilities for AI, since mastery of inferential roles might be achievable through training on linguistic data alone, without the need

for direct sensory engagement with the world (Gubelmann, 2023). If "mastery of inferential role is sufficient for possession of all concepts" (Simonelli 2025, 7':37"), then systems trained exclusively on linguistic data might in principle be the participants in the social practice of giving and asking for reasons. Thus, the implications of inferentialism for LLMs are not obvious and call for the kind of investigation that my thesis will provide.

SAPIENCE WITHOUT SENTIENCE

So, is it possible for a system to possess sapience (conceptual understanding) without sentience (conscious awareness)? Traditional accounts have generally assumed these two are necessarily connected—that conceptual understanding requires some form of conscious awareness (Searle, 1980). But LLMs challenge this assumption by demonstrating capabilities without anything resembling sentience. They appear capable of reading without giving a damn—of processing and generating meaningful text without caring about or being aware of what they're doing (Haugeland 1985, Simonelli 2025).

D. NORMATIVITY

There are reasons to doubt whether LLMs genuinely participate in language games in the full normative sense. The ability to predict what word is likely to come next in a sequence is not the same as grasping the commitments and responsibilities involved in assertion. While LLMs can simulate participation in the game of giving and asking for reasons, they lack *testimonial standing*—the capacity to stand behind their assertions and take responsibility for their correctness (Redaelli, 2024). This *testimony gap* reflects a limitation: LLMs do not and cannot care about the truth of what they say because they lack the intentional architecture necessary for caring (Sparrow and Flenady, 2025). This precludes the idea of being held responsible and demanding such.

Even those who acknowledge the remarkable capabilities of current LLMs and project their continued improvement, often recognize that something essential to human understanding may be missing from these systems (Müller and Bostrom, 2016). The question is whether this missing element—be it conscious awareness, intentionality, or normative commitment—is merely a contingent limitation of current technology or a necessary consequence of the fundamentally different architecture of AI. Some argue that future advances in machine learning will eventually bridge this gap, while others maintain that language exploitation requires a form of embodied, world-engaged consciousness that computational systems inherently lack (Bender and Koller, 2020).

What emerges from this analysis is a recognition that the question whether machines can genuinely use and express language cannot be answered without first clarifying what we mean by this usage. If it is merely a matter of producing appropriate linguistic responses to stimuli, then LLMs already demonstrate efficient language usage (Turing 1950; Arai and Tsugawa 2023). But if it involves normative commitments undertaken within a community of reason-givers and

reason-takers, then LLMs remain fundamentally limited. Their performance reflects pattern recognition rather than genuine participation in the normative practices that constitute rational discourse.

II. THE CATEGORIAL DISTINCTION AND THE NORMATIVE COMMUNITY

A. Establishing the Categorial Distinction

I want to establish why Large Language Models fail and will continue to fail, regardless of all future magnificent progresses, to participate in the social practice of giving and asking for reasons. This failure stems from a categorial distinction between mechanistic interpretability (explaining AI systems in terms of their internal mechanisms) and behavioral interpretability (explaining AI in terms of their performance and generated texts).

LLMs' algorithms, though able to follow the probabilities of linguistic propositions, miss the most important part of rational discourse: intentional membership in the normative community.

B. The Three Constitutive Ties of Normative Community

The normative community consists of participants in the social practice of giving and asking for reasons, bound together by three constitutive ties that cannot be simulated or approximated:

1. ***Mutual Recognition of Authority and Responsibility***

The first tie is mutual recognition of members as possessing both authority and responsibility in the game of giving and asking for reasons. This creates a dialectical relationship between norms and attitudes. As Brandom explains, norms as socially instituted by attitudes of reciprocal recognition (Brandom, 2019).

This mutual recognition creates normative statuses through our mutual acknowledgment of each other as beings capable of undertaking commitments and being held responsible for them. When I recognize you as a discursive practitioner, I treat your assertions as potentially binding on me through testimonial inheritance, as subject to my challenges through the default and challenge structure, and as capable of altering the deontic score we jointly maintain.

Being responsible means having the ability to come up with a response, the ability to give reasons. We hold participants responsible for their assertions not primarily by evaluating if they are free intentional agents, but by first evaluating if excusing conditions apply to them. I cannot participate in the practice of giving and asking for reasons with a participant that has no control over their assertions—just as I cannot engage with a drunk person, I cannot engage in discursive practice with a tool that has no intentions of setting goals for itself (Strawson, 1963).

2. ***Commitment to the Inferential Rule***

The second tie is a general commitment to the inferential rule—the pragmatic significance of making claims explicit through logical vocabulary, which requires practical knowledge. LLMs cannot commit to the inferential rule because they lack intentionality and desires of their own (as tools) and have no control over their generated texts.

Committing to the inferential rule is not just following patterns of "if-then" statements. It is understanding that conditionals make explicit the endorsement of material inferences, that negation makes explicit incompatibility commitments, and that quantifiers make explicit substitutional commitments. This means understanding not just the meaning but also the implications and consequences.

Understanding propositional content involves practical knowledge—grasping a propositional content means being able to accord practical significance to assertions of it, knowing what one is doing in asserting it (Brandom, 1994). This practical knowledge transcends mere statistical regularities and requires active participation in a socially articulated space of reasons.

3. *Intentional Autonomous Commitment*

The third tie is intentional autonomous commitment to the normative community rules or the inferential rule. This internal decision to see the urge to give and ask for reasons cannot be induced by external forces but emerges from recognition of oneself as a rational agent capable of both authority and responsibility.

III. CENTRAL ARGUMENT

LLMs fundamentally cannot participate in the game of giving and asking for reasons. This inability comes from two restrictions: (1) technical constraints, which may or may not be overcome and (2) their inability to achieve intentional autonomous membership in normative communities, which they will never overcome.

A. INFERENTIALISM AND DISCURSIVE PRACTICE

In Inferentialism, semantic content derives from the inferential roles within the game of giving and asking for reasons. This game constitutes the foundation of rationality and serves as the basis of discursive practice. Discursive practice is fundamentally normative, governed by implicit rules that participants both follow and enforce.

The failure in being a member of this normative community results in a form of alienation from norms which demarcates LLMs from participating in the social practice of giving and asking for reasons. According to Brandom's interpretation of Hegel, this alienation represents not only a contingent exclusion but a constitutive impossibility, for in the absence of the reciprocal recognition that establishes normative statuses with normative attitudes, an entity remains outside the structure (of Geist)—the socially constituted space where conceptual content and

commitments become determinate through the mutual allocation of authority and responsibility. Furthermore, lacking membership in a normative community entails the absence of recollective rationality—the capacity to reconstruct and thereby rationally vindicate the inferential commitments one has undertaken, a capacity that requires not just the tracking of patterns but genuine participation in the historical process through which norms are instituted, recognized, and transformed.

Within this framework, claims emerge as the primary units of discursive exchange. Subsequently, giving reasons presupposes the possibility of asking for them, or at least the possibility that claims often stand in need of reasons. Claims are therefore inherently dual-functional: they simultaneously offer reasons and can have reasons demanded of them.

B. DEONTIC SCOREKEEPING

The practice of giving and asking for reasons depends on *deontic scorekeeping*—the social tracking of commitments and entitlements. As participants engage in discourse, they implicitly maintain a record of what others are committed and entitled to claim.

...It is by reference to the attitudes of others toward the deontic status (attributing a commitment) that the attitude of the one whose status is in question (acknowledging or undertaking a commitment) is to be understood. (Brandom 1994, 162)

This scorekeeping practice is social and normative, requiring participants to adopt stances toward one another as beings capable of undertaking commitments and being held responsible for them. Members must master multi-perspectival scorekeeping—tracking not just what commitments and entitlements I attribute to others, but also what I take them to attribute to me and to third parties. This creates the social articulation of content that makes objectivity possible.

C. THE INTRINSICALLY SOCIAL CHARACTER OF DISCURSIVE PRACTICE

The practice of giving and asking for reasons is irreducibly social and therefore normative. When someone makes an assertion, they undertake a commitment with three dimensions of normative significance:

- The commissive dimension—what the assertion commits them to
- The permissive dimension—what the assertion entitles them to
- The preclusive dimension—what the assertion makes them incompatible with claiming

These "downstream" consequences are matched by "upstream" dimensions: the circumstances under which one is entitled to make an assertion, the defeasors that would undermine that entitlement, and the challenges that might demand vindication. This six-fold structure cannot be captured by any system that merely tracks patterns, because it requires genuine normative attitudes that institute these statuses through social recognition.

D. TECHNICAL LIMITATIONS OF LLMs AS DISCURSIVE AGENTS

LLMs currently operate through statistical pattern recognition, utilizing architectures that predict token sequences based on their training data. Despite their impressive capabilities in text generation, their mechanism remains probabilistic predictors rather than normative followers.

The back-propagation process in LLMs optimizes prediction accuracy but lacks any mechanism for tracking normative statuses or understanding the consequences of assertions. This architectural limitation means that LLMs do not and cannot grasp the normative significance of the linguistic patterns they reproduce.

LLMs simulate linguistic behavior through pattern matching but cannot genuinely participate in the normative practices that constitute reasoning. Their outputs may superficially resemble claims, but they lack the essential normative dimension that defines genuine assertion. There lies a difference between merely conforming to norms and genuinely following them—LLMs exhibit the former but cannot achieve the latter.

E. THE INTENTIONALITY GAP

The requirement for intentional participation reveals a fundamental gap that cannot be bridged by engineering. To gain membership in the normative community, one must come to an internal decision to give reasons and to ask for them.

"One who succeeds in making a promise still authorizes others to rely on one's future performance, to hold one responsible for a failure to perform according to one's commitment. These are both authority and responsibility, adding up to commitment" (Brandom 1994, 165).

The participant must be able to foresee and grasp the urgency of providing reasons for their assertions. They must recognize when their commitments have been challenged and understand what would count as adequate vindication. Most crucially, they must individually choose to answer the call as a responsible agent.

The source of provided reasons must be essentially intrinsic rather than extrinsic. LLMs are purposeful in an extrinsic sense—their engineered purpose derives from user needs. Without stimuli, there are no responses. Without prompts, there will be no analysis. The absence of spontaneous engagement with normative statuses reveals the absence of genuine intentionality.

F. THE FAILURE OF EXTENSIONAL INTENTIONALITY

When developers create LLMs with certain intentions, those intentions do not transfer to the tools themselves. An elevator doesn't choose to move between floors; a car doesn't decide to transport you. LLMs too do not choose to generate well-structured responses, with no matter how many signs of autonomous learning through RLHF.

This differs even from "Reliable Differential Responsive Dispositions" (RDRD) found in simple organisms. A parrot that mimics speech can choose when to vocalize for rewards. Despite

superior performance, LLMs lack even this minimal intentionality—they execute probabilistic functions without genuine choice or recognition of normative significance.

IV. THE NECESSITY OF NORMATIVE COMMUNITY MEMBERSHIP

A. THE SOCIAL CONSTITUTION OF NORMS

Inferentialism conceives of norms as socially constituted and maintained regularities. This dialectical relationship between norms and attitudes requires a community of practitioners who mutually recognize one another as participants in the normative practice.

B. AUTHORITY AND RESPONSIBILITY IN DISCURSIVE PRACTICE

Participation in the game of giving and asking for reasons requires both the exercise of authority and the assumption of responsibility. To make a claim is simultaneously to exercise authority by putting forward a content as worthy of endorsement and to take responsibility for defending that content when challenged.

C. WHY LLMs CANNOT ACHIEVE MEMBERSHIP

LLMs cannot become discursive creatures, which also implies these models cannot achieve membership in normative communities for:

1. *Absence of Genuine Recognition*: LLMs cannot recognize others as authorities capable of holding them responsible, nor can they recognize themselves as beings who can be legitimately held to account.
2. *Inability to Undertake Commitments*: Without the capacity for genuine normative attitudes, LLMs cannot undertake the commitments that constitute assertion.
3. *Lack of Normative Standing*: LLMs have no normative standing within discursive communities—they cannot be praised or blamed, held responsible or credited in the way that genuine discursive participants can.

The game of giving and asking for reasons depends on normative statuses (commitments and entitlements) that are instituted by social practices of mutual recognition of each participant.

As Brandom explains, the normative statuses characteristic of concept use is domesticated into a form of normative attitudes, just what practitioners take or treat each other as being committed or entitled to (Brandom, 1994). This irreducibly social dimension of normativity places a principled limitation on what LLMs can achieve.

D. FORMAL ARGUMENT

The main argument can be formalized as follows:

- **P1:** Participation in the practice of giving and asking for reasons requires an intentional autonomous membership in a normative community where participants mutually recognize each other's authority to make claims and be held responsible for them.
- **P2:** LLMs cannot achieve an intentional autonomous membership in a normative community where participants mutually recognize each other's authority to make claims and be held responsible for them.
- **C:** Therefore, LLMs cannot participate in the practice of giving and asking for reasons.

V. OBJECTIONS AND RESPONSES

A. THE SIMONELLI CHALLENGE: INFERENTIALISM CRITIQUE

Ryan Simonelli suggests that LLMs might indeed qualify as genuine discursive creatures through their Mastery of Inferential Roles. Simonelli contends that the requirement for concept possession is the mastery of inferential roles, which implies that training on linguistic data [might be] sufficient for possessing of all concepts. On this account, LLMs trained exclusively on data could potentially possess conceptual capacities sufficient for discursive practice.

- A_x = Being trained on linguistic datasets
- B_x = Achieves mastery of inferential role
- C_x = Possesses all concepts
- a = LLMs

$A_x \rightarrow B_x$
$B_x \rightarrow C_x$

$A_x \rightarrow C_x$ $\quad\quad$ $A_a \rightarrow C_a$

This argument overlooks two prerequisite conditions:

- D_x = Intentionally engages with and grasps concepts
- E_x = Autonomously sets the intentions

The proper inference requires: $E_x \wedge D_x \wedge A_x \rightarrow B_x$, and $B_x \rightarrow C_x$. But we cannot establish D_a and E_a for LLMs.

In Jackson's Mary's room thought experiment, when Mary comes out of the room and sees red, she doesn't just process information—she chooses to grapple with whether she is seeing red. This intentional autonomous engagement with conceptual content cannot be programmed through datasets alone. The difference between Mary being forced to identify colors versus willingly

examining them illustrates why intentionality matters. In the voluntary case, Mary can go beyond the set goal and provide reasons even when valid inferences might not be accessible through the datasets being embedded in its algorithms.

LLMs can only give reasons based on datasets embedded in their models and connections between them. They cannot go beyond the limit of the goal set for them because they cannot choose otherwise. This makes them unable to be intentional autonomous members of normative community.

B. THE NON-MONOTONIC MULTI-SUCCEDENT CHALLENGE

One might think we could close this gap by integrating normative rules directly into AI systems. The non-monotonic multi-succedent (NMMS) logical system that Brandom and Hlobil develop provides formal tools for representing defeasible reasoning and multiple simultaneous conclusions. Could we not simply implement such a system in LLMs?

Response

The NMMS system captures important features of reasoning—the non-monotonic character where adding premises can defeat previous conclusions, the multi-succedent structure where we can entertain multiple incompatible conclusions simultaneously, and the defeasible nature of material inferences. But implementing such a formal system would only enable LLMs to simulate following the patterns of normative reasoning, not to engage in genuine normative practices.

The distinction is between conforming to norms and following them—between exhibiting behaviors that accord with rules and being bound by those rules. An LLM implementing NMMS logic would still be executing probabilistic functions, not undertaking commitments or recognizing responsibilities.

C. BEHAVIORALISM CRITIQUE

Another objection argues that actual participation in normative communities may be unnecessary for discursivity. According to this view, an agent does not need to belong to a normative community to function within discursive practices.

Response

This objection fails to recognize three key dimensions of normativity:

***SOCIETAL ASPECT*:** The norms in language are not merely regularities that can be statistically approximated but constitutively social phenomena that emerge through practices of mutual recognition.

PRAGMATISM: Concept possession transcends pattern recognition, requiring instead the capacity to understand and respond to normative statuses within a community of recognizers.

This involves a practical knowledge—knowing what one is doing in deploying concepts—that remains inaccessible to LLMs.

LLMs DEPENDENCY

AI remains dependent on human discursive practices. The tools may exhibit convincing simulations of reasoning, but they necessarily derive their normative status from the communities that create and interpret them.

VI. IMPLICATIONS AND CONCLUSION

This analysis does not diminish the achievements of AI. LLMs are magnificent tools that can process and generate text with sophistication. They can help us explore patterns, generate hypotheses, and reveal implicit connections. But we must be clear about what they are and are not capable of. As one can play dumb without actually being dumb; one can also play intelligent without actually being intelligent. How? Only through extended participation in the social practice of giving and asking for reasons.

LLMs are not rational agents. They are not members of our normative community. They cannot genuinely give or ask for reasons. They will always rely on genuine discursive agents-us-to provide the normative framework that makes their outputs meaningful.

But are we sure they cannot ever obtain intentions? What if tomorrow an AI engineer shares that they succeeded in adding intentionality to machine learning algorithms? The test would be examining whether LLMs can participate in discursive practices outside of engineer's defined goals. Without prompts, would LLMs initiate discursive practice on their own, without satisfying user purposes? Would they give and ask for reasons beyond programmed limits? They will not.

Regardless of how sophisticated LLMs become in mimicking linguistic behaviors, they will remain categorically distinct from genuine discursive agents. The game of giving and asking for reasons—the practice that defines us as rational beings—will remain beyond their reach not merely due to technical limitations but due to their fundamental inability to enter the normative communities that constitute and sustain discursive practice.

The social practice of giving and asking for reasons is constitutively tied to membership in normative communities. Such membership requires capacities that cannot be engineered into systems lacking genuine intentionality, unable to undertake real commitments, and incapable of reciprocal recognition. LLMs, regardless of sophistication, remain outside this space of reasons.

To advance AI development, we must accept this categorial distinction and work within it. LLMs are extraordinary extensions of human discursive practice but cannot become participants in that practice. They require genuine discursive agents to provide the normative framework making their outputs meaningful. This is not a limitation to overcome but a fundamental feature

distinguishing tools from rational agents, pattern matching from genuine understanding, simulation from participation in the social practice of giving and asking for reasons.

REFERENCES

Alammar, Jay, and Maarten Grootendorst. *Hands-On Large Language Models*. O'Reilly Media, Inc., 2024.

Arai, Koji, and Satoshi Tsugawa. 2023. [Reference to be completed]

Bender, Emily M., and Alexander Koller. "Climbing Towards NLU: On Meaning, Form, and Understanding in the Age of Data." *Proceedings of the Annual Meeting of the Association for Computational Linguistics* 58 (2020): 5185-98.

Boisseau, É. "Imitation and Large Language Models." *Minds and Machines* 34 (2024): 42.

Brandom, Robert. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press, 1994.

Brandom, Robert B. *A Spirit of Trust: A Reading of Hegel's Phenomenology*. Cambridge, MA: Harvard University Press, 2019.

Brandom, Robert B. *Between Saying and Doing: Towards an Analytic Pragmatism*. Oxford: Oxford University Press, 2008.

Butlin, Patrick, and Emanuel Viebahn. "AI Assertion." *Ergo: An Open Access Journal of Philosophy* (2023).

Chalmers, David. “Propositional Interpretability in Artificial Intelligence.” (2025).

Chalmers, David J. "Could a Large Language Model Be Conscious?" *Boston Review* 1 (2023).

Culbertson, Carolyn. "Shared Understanding Before Semantic Agreement: Gadamer on the Hidden Ground of Linguistic Community." *Journal of the British Society for Phenomenology* 56, no. 1 (2024): 57-69.

Gubelmann, Reto. "A Loosely Wittgensteinian Conception of the Linguistic Understanding of Large Language Models like BERT, GPT-3, and ChatGPT." *Grazer Philosophische Studien* 99, no. 4 (2023): 485-523.

Gubelmann, Reto. "Large Language Models, Agency, and Why Speech Acts Are Beyond Them (For Now) — A Kantian-Cum-Pragmatist Case." *Philosophy and Technology* 37, no. 1 (2024): 1-24.

Gubelmann, Reto, Ioannis Katis, Christina Niklaus, and Siegfried Handschuh. "Capturing the Varieties of Natural Language Inference: A Systematic Survey of Existing Datasets and Two Novel Benchmarks." *Journal of Logic, Language and Information* 33, no. 1 (2023): 21-48.

Gubelmann, Reto, and Siegfried Handschuh. "Context Matters: A Pragmatic Study of PLMs' Negation Understanding." In *Proceedings of the ACL*, 2022.

Haugeland, John. *Artificial Intelligence: The Very Idea*. Cambridge, MA: MIT Press, 1985.

Havlík, Vladimír. "Meaning and Understanding in Large Language Models." *Synthese* 205, no. 1 (2024): 1-21.

Heinrichs, Bert, and Sebastian Knell. "Aliens in the Space of Reasons? On the Interaction Between Humans and Artificial Intelligent Agents." *Philosophy and Technology* 34, no. 4 (2021): 1569-1580.

Hlobil, Ulf, and Robert Brandom. [Reference to NMMS system to be completed]

Holliday, Wesley H., Matthew Mandelkern, and Cedegao Zhang. "Conditional and Modal Reasoning in Large Language Models." Manuscript. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024)*.

Jackson, Frank. [Reference to Mary's room thought experiment to be completed]

Jack Lindsey et al., "On the Biology of a Large Language Model," *Transformer Circuits*, March 27, 2025, https://transformer-circuits.pub/2025/attribution-graphs/biology.html

Legg, Catherine. "Peirce and Generative AI." Forthcoming in Robert Lane, *Pragmatism Revisited*. Cambridge University Press.

Miller, Ryan. "Does Artificial Intelligence Use Private Language?" In Ines Skelac and Ante Belić, eds., *What Cannot Be Shown Cannot Be Said: Proceedings of the International Ludwig Wittgenstein Symposium, Zagreb, Croatia, 2021*, 113-24. Lit Verlag, 2023.

Monti, Paolo. "AI Enters Public Discourse: A Habermasian Assessment of the Moral Status of Large Language Models." *Ethics and Politics* 61, no. 1 (2024): 61-80.

Müller, Vincent C., and Nick Bostrom. "Future Progress in Artificial Intelligence: A Survey of Expert Opinion." In Vincent C. Müller, ed., *Fundamental Issues of Artificial Intelligence*, 553-571. Cham: Springer, 2016.

Redaelli, R. "Intentionality Gap and Preter-Intentionality in Generative Artificial Intelligence." *AI & Society* (2024).

Searle, John. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3, no. 3 (1980): 417-57.

Simonelli, Ryan. "Sapience without Sentience: An Inferentialist Approach to LLMs." YouTube video, 1:21:33. Posted by "Ryan Simonelli," August 4, 2023. https://www.youtube.com/watch?v=nocCJAUencw

Simonelli, Ryan. "The Normative/Agentive Correspondence." *Journal of Transcendental Philosophy* 3, no. 1 (2022): 71-101.

Sparrow, Robert, and Gene Flenady. "The Testimony Gap: Machines and Reasons." *Minds and Machines* 35, no. 1 (2025): 1-16.

Stoljar, Daniel, and Vincent Zhihe Zhang. "Why ChatGPT Doesn't Think: An Argument from Rationality." *Inquiry: An Interdisciplinary Journal of Philosophy*. Forthcoming 2024.

Strawson, Peter (1963). Freedom and Resentment. Proceedings of the British Academy 48:187-211.

TED. *OpenAI's Sam Altman Talks ChatGPT, AI Agents and Superintelligence — Live at TED2025*. YouTube video, 1:08:17. Posted by "TED," April 14, 2024. https://www.youtube.com/watch?v=5MWT_doo68k

Turing, Alan. "Computing Machinery and Intelligence." *Mind* 59, no. 236 (1950): 433-60.

Wittgenstein, Ludwig. *On Certainty*. Edited by G. E. M. Anscombe, G. H. von Wright, and Mel Bochner. San Francisco: Harper Torchbooks, 1969.

Wittgenstein, Ludwig. *Philosophical Investigations*. Edited by G. E. M. Anscombe. New York, NY: Wiley-Blackwell, 1953.