CAN CONVERSATIONAL AIS TESTIFY? CHALLENGING THE CONSERVATIVE VIEW

Domingos FARIA

ABSTRACT: The conservative view on testimony rests on the assumption that a testifier must (1) believe or know the propositional content of the testimony, (2) intend to deliver the testimony, (3) be a responsible epistemic agent, (4) be an object of trust, and (5) be capable of making assertions. According to this view, such conditions apply exclusively to human beings (or minded agents), not to technological artifacts or instruments. For this reason, conversational AIs are typically not considered or classified as testifiers. However, in this paper, I argue that some of the conditions posited by the conservative view – namely (1) and (2) – do not constitute necessary conditions for an entity to qualify as a testifier, while others – (3), (4), and (5) – can plausibly be satisfied by conversational AIs. Therefore, there are no compelling reasons to reject the claim that conversational AIs can testify.

KEYWORDS: artificial intelligence, large language models, conversational AIs, epistemology of testimony

1. Introduction

Conversational AIs are technologies that can interact with human beings using natural language. Outputs from conversational AIs are not pre-made propositions uttered by artificial speakers (for example, automated pre-recorded phone messages: "Press 5 if you want to activate roadside assistance"). Nor are they technological devices that mediate human testimony (for example, when my smartphone reads aloud a text message I received from my wife). Instead, the outputs of conversational AIs are propositions generated by AI and delivered algorithmically; namely, they usually work with large language models (LLMs). LLMs are mathematical models of statistical distribution that recognize, predict, and generate text. These models use machine learning and other AI techniques to calculate and determine the probability of text occurrence; that is, the models are trained on vast amounts of text to learn patterns and relationships, as well as to be accurate and reliable. Thus, conversational AIs are sophisticated statistical machines.

Conversations with technological AI devices have become indistinguishable, in both form and content, from conversations with human beings. It is plausible to say that AI conversational models seem to pass the Turing test (cf. Mitchell (2024), Jones and Bergen (2025)). Furthermore, we learn new things, we acquire knowledge,

based on the "say-so" of conversational AIs (such as ChatGPT). For example, I typed the following prompt: "What is the atomic number of gold?" ChatGPT answered: "The atomic number of gold is 79." I typed in another prompt: "And that of water?" ChatGPT answered: "Water (H_2O) is a compound, not an element, so it does not have an atomic number. However, the atomic numbers of the elements that make up water are Hydrogen (H): atomic number 1; and Oxygen (O): atomic number 8. Water is composed of two hydrogen atoms and one oxygen atom."

How should such attributions of knowledge be understood? Should they be classified as testimonial knowledge, analogous to knowledge acquired through interpersonal communication? Or, alternatively, should they be regarded as a form of instrument-based knowledge, similar to that obtained via a thermometer or other scientific instruments? Clarifying the epistemic status of knowledge acquired from the "say-so" of a conversational AI is of considerable importance, as it may entail significant social and legal ramifications. In short, can conversational AIs testify or deliver testimony?

To answer these questions, I begin in Section 2 by analyzing the conservative view of testimony, from which it is typically concluded that conversational AIs cannot be testifiers. However, in the following sections, I argue that this conservative view fails to successfully exclude conversational AIs from being classified as testifiers, since some of the conditions it establishes are not necessary for an entity to testify, while the remaining conditions can be easily satisfied by conversational AIs.

2. The conservative view on testimony

The conservative view, as defended by Coady (1992), Lackey (2008), Tollefsen (2009), Goldberg (2012), Fricker (2015), and Pagin and Marsili (2021), is that conversational AIs cannot be considered testimonial sources, but at most instrumental sources of knowledge (in a similar way to the knowledge we obtain when we consult a thermometer). The main argument for this conservative view can be summarized as follows:

1. An entity S can testify that p only if S believes that p, S has the intention to deliver testimony that p, S is a responsible epistemic agent for transmitting that p, S is an object of trust, and S is able to assert that p.
2. But conversational AIs cannot believe that p, nor intend to testify that p, nor are they responsible epistemic agents who transmit that p, nor are they objects of trust, nor are they able to assert that p.
3. Therefore, conversational AIs cannot testify that p.

I intend to show that this argument is not sound, since there are plausible reasons to reject both premises. In particular, underlying this conservative view are the assumptions that a testifier (1) must believe or know the propositional content of the testimony, (2) must intend to deliver the testimony, (3) must be a responsible epistemic agent, (4) must be an object of trust, and (5) must be able to assert. Such conditions only seem to apply to people (or agents with minds), and not to technological artifacts or instruments.[1] For this reason, conversational AIs are not typically considered or classified as testifiers. However, some of the conditions presented in the conservative view, namely (1) and (2), do not appear to qualify as necessary conditions for being a testifier, while other conditions, such as (3), (4), and (5), can indeed be met by conversational AIs.

3. The belief or knowledge condition

Starting the critical analysis with condition (1), this *belief or knowledge condition* holds that entities can only testify if they believe or know what they are testifying about. In other words, testimony requires that testifiers sincerely believe or know the claims about which they are testifying (cf. Coady (1992), Mallory (2023)). However, conversational AIs do not have beliefs.[2] This is because AIs do not even grasp the propositions under consideration.[3] Furthermore, conversational AIs do not possess representational mental states with propositional content, nor do they appear to have behavioral dispositions related to that content.

Nevertheless, as an objection to this belief or knowledge condition, it can be pointed out that believing or knowing is not a necessary condition for being a testifier. A counterexample formulated by Lackey (2008, 48) illustrates this point: suppose a creationist teacher gives her class a lesson on the theory of evolution, even though she does not believe in it. Her students, unaware that the teacher is a creationist, come to believe in the theory of evolution. In this case, there is a testifier, yet she does not hold the relevant belief or knowledge regarding the theory of evolution.[4] Therefore, belief or knowledge is not a necessary condition for being a

[1] This conservative view on testimony can be classified as an "anthropocentric view" since this "view presupposes that only persons can participate in the act of testimony because only humans, in principle, can be qualified as a testifier" (cf. Freiman (2024, 479)). But I can already suspect that an anthropocentric perspective for dealing with testimony in general may be arbitrary and biased.

[2] See, for exemple, Shanahan (2024).

[3] At the end of section 7, I will challenge this idea.

[4] One could argue that this case, as presented by Lackey (2008, 48), does not capture the full complexity of the situation, and that in order for testimony to occur, the first link in the testimonial chain must possess the relevant belief or knowledge (cf. Wright (2018)). In the case under consideration, the students acquire knowledge via testimony because the initial links in the

testifier, and so the fact that conversational AIs lack beliefs or mental states is not problematic in this respect.

Yet the idea that conversational AIs lack beliefs has recently been challenged. This challenge can be raised by adopting a conception of belief, as proposed by Herrmann and Levinstein (2025), in which beliefs are understood as "internal representations" of truth that guide action. In the case of conversational AIs, such beliefs would correspond to structures within the LLM model that "classify" (or tag) certain propositions as true or false and use them to guide outputs. Just as humans use beliefs to make decisions, LLMs can develop internal representations that function as "maps" or guidelines for generating more coherent, informed, accurate, and effective responses. According to Herrmann and Levinstein (2025), for an internal representation to count as a "belief", it must satisfy four criteria: (i) accuracy – the representations must be mostly true in domains where the model has reliable knowledge; (ii) coherence – the representations must be logically consistent; (iii) uniformity – the representation of truth must remain consistent across different domains; and (iv) use – the representation must actively influence the model's output, i.e., what text it generates. If conversational AIs meet these four criteria, they possess internal representations that function as beliefs. And there is currently some empirical evidence suggesting that conversational AIs at least partially satisfy several of these conditions (cf. Azaria and Mitchell (2023), Marks and Tegmark (2024)).

4. The intention condition

The next condition present in the argument of the conservative view on testimony, condition (2), which I can call the *intention condition*, holds that entities can only testify if they intend to do so. For example, Fricker (2015, 175) maintains that "instances of testimony are intentional communicative acts made by a testifier or speaker *S* to her intended recipient a hearer *H*". Similarly, Lackey (2008, 3) draws attention to the dual nature of testimony, stating that "on the one hand, testimony is often thought of as an intentional act on the part of the speaker and, on the other hand, testimony is often thought of as simply a source of belief or knowledge for the hearer". Lackey (2008, 30) also proposes and defends the following definition of

transmission chain, going back to Darwin, do possess the relevant belief and knowledge, and the creationist teacher merely relays one segment of this chain. However, this does not undermine the claim that conversational AIs can testify. Just as the creationist teacher counts as a testifier (inasmuch as earlier links in the testimonial chain had belief and knowledge), conversational AIs can likewise function as testifiers, provided the chain from which they draw contains belief and knowledge.

speaker testimony: "*S* s-testifies that *p* by performing an act of communication *a* if and only if, in performing *a*, *S* reasonably intends to convey the information that *p* (in part) in virtue of *a*'s communicable content." Assuming that technologies do not have intentions (cf. Woudenberg, Ranalli, and Bracker (2024)), conversational AIs cannot be instances of "speaker testimony" or be considered testifiers.

However, as an objection to this intention condition, we can draw on the counterexamples presented by Coady (1992, 51), who discusses cases of "documentary testimony", i.e., the reading of a personal diary written by someone who never intended it to be read. In such cases, the author did not have the direct intention of someone specific reading the text. Yet, even without an element of intentionality, these still count as cases of testimony.[5] Another example: suppose Joseph is talking to some friends about climate change and says, "The recent forest fires are all related to climate change; and it's all our fault!" Mary happens to be passing by and hears the conversation. She ends up believing this statement, even though Joseph did not intend for her to hear it. Again, we have an instance of a testifier without the intention to convey a specific message to a particular hearer. Therefore, intention is not a necessary condition for testimony, and thus it is not problematic that conversational AIs lack intentions.[6]

5. The epistemic responsibility condition

Moving on to the next condition, condition (3), I designate it as the *epistemic responsibility condition*. According to this condition, entities can only testify if they are epistemic agents responsible for transmitting propositional content, since the testifier is accountable for the truth of what she asserts. For example, Fricker (2015, 176) argues that "in any act of testifying the speaker takes on responsibility for the truth of what she tells to her audience." This kind of responsibility, however, cannot yet be attributed to technological artifacts, and thus they cannot be considered testifiers. To better illustrate this point, Goldberg (2012, 184) presents the following pair of cases:

> TEMPERATURE 1: Smith wants to know the temperature. Knowing that Jones often attends to these things, he asks her. Jones replies that it is 71ºF. On this basis, Smith forms the belief that it is 71ºF.

[5] For a similar objection, see Narayanan and Cremer (2022, 10).

[6] One could take the objection further and argue that conversational AIs can, in fact, have intentions, as Lederman and Mahowald (2024) contends. The idea is that, by adopting interpretationalism in the philosophy of mind, one can claim that conversational AIs possess intentions insofar as the best explanation of their behavior involves attributing goals and plans to them.

> TEMPERATURE 2: Smith wants to know the temperature. Knowing that there is a thermometer by the kitchen window, he consults it. It reads 71ºF. On this basis, Smith forms the belief that it is 71ºF.

The question before us is whether there is any significant epistemic difference between these cases. Goldberg (2012, 184) defends the following thesis:

> The key difference between testimony-based and instrument-based belief, I will be arguing, is: to rely in belief-formation on another speaker is to rely on an epistemic subject, that is, on a system which itself is susceptible to epistemic assessment in its own right, whereas "mere" instruments and mechanisms are not properly regarded as epistemic subjects in their own right, they are not susceptible to normative epistemic assessment.

And Goldberg (2012, 188) further clarifies what it means to be susceptible to normative epistemic assessment:

> α's information processing is relevant to assessments of the doxastic justification of beliefs formed through reliance on α's output if, but only if, (i) α's information processing can be assessed for reliability and (ii) α's operations are properly assessed for rationality and responsibility.

Goldberg (2012) defends an asymmetrical treatment between TEMPERATURE 1 and TEMPERATURE 2, that is, between beliefs based on the testimony of epistemic agents (such as people) and beliefs based on instruments (such as thermometers or clocks). There is a normative difference here: epistemic agents, people, are susceptible to full normative evaluations (e.g., in terms of rationality and responsibility), while instruments are only evaluated in terms of their reliability. This difference explains why *testimony* involves a distribution of epistemic responsibility between speaker and hearer, while the use of *instruments* places all responsibility on the subject who uses them. Thus, in cases of testimony, the process of belief formation is interpersonally extended; it includes not only the mind of the hearer but also that of the speaker (e.g., whether the speaker has sufficient evidence, whether their assertion is justified, whether they should believe what they say, etc.).

But, extending Goldberg's (2012) argument, conversational AIs are mere instruments, since they do not follow standards of rationality or epistemic responsibility. In fact, we can assess whether instruments (such as clocks, thermometers, or conversational AIs) are reliable, accurate, or well-calibrated, but we cannot say that they act irresponsibly or violate epistemic norms, since they operate solely on the basis of physical causality or programming. Therefore, unlike testimony (where the hearer trusts that the speaker has fulfilled certain epistemic obligations) in the case of instruments, all the responsibility lies with the user, who

must, for example, check whether the clock, thermometer, or conversational AI is functioning properly. In short, it is not possible to normatively evaluate conversational AIs, as they are not rational or responsible epistemic agents. The idea is that they are mere instruments. As such, they cannot testify.

This last line of reasoning, although it seems to carry some weight, can be challenged on the grounds that conversational AIs can, in fact, satisfy the epistemic responsibility condition. But how? It is true that conversational AIs are not directly susceptible to normative epistemic assessment. However, in an indirect or derivative sense, such assessment can be attributed to those who design and maintain these AI models (such as programmers, executives, etc.). In this way, the design, implementation, and operation of conversational AIs can be subject to normative epistemic assessment by imputing rationality and responsibility to their creators. According to Goldberg (2012), responsibility in testimony is interpersonal (between speaker and hearer); however, the same can apply to conversational AI models, where responsibility may be distributed among users, programmers, and other parties involved in the design, implementation, and operation of these systems. It is also true that, in such AI systems, responsibility may be diluted or not easily identifiable; nevertheless, this is a common feature of complex epistemic systems and corresponds to a version of the epistemic problem of "many hands", as Helen Nissenbaum (1996, 29) describes it. In these systems, it may not be obvious who is to blame, nor is it easy to identify a single individual as responsible. Thus, as in other complex epistemic systems, in the case of conversational AIs, accountability and responsibility for the accuracy and consequences of the AI system's outputs can be distributed across "many hands".[7] Moreover, organizations that develop AI-based products are generally expected to assume responsibility for their products. For example, OpenAI is typically considered responsible for ChatGPT. This, in itself, suggests that conversational AIs are already treated as satisfying the epistemic responsibility condition.

It is worth noting that this possible response was already anticipated by Goldberg (2012, 194) himself, who pointed out that "if a given computer yields information that turns out to be false, we will blame the programmer, or our use of the program, or ..., but not the computer itself. (…) As evidence of this, I note that we do not resent the computer." However, against this perspective, I argue that it is legitimate to hold the system itself normatively responsible, provided we understand "responsibility" as a form of functional evaluation.[8] Suppose, for instance, that the

[7] For a similar objection, see Freiman (2024, 482).

[8] Simion and Kelp (2023) define trustworthiness as a system's (or agent's) disposition to fulfill its functional obligations, whether derived from design or etiological functions. Following a similar

developers of conversational AI models have implemented algorithms to prevent misinformation and verify facts. In such cases, these systems embed epistemic norms into their design, becoming more than mere instruments: they are artifacts with defined cognitive functions. Thus, if a conversational AI designed for fact-checking begins to repeat falsehoods, ignore contradictory data, and resist correction, we can "blame" or criticize it for violating the epistemic norms it was built to uphold, much like we might criticize a human jury for negligence. In this sense, unlike a broken thermometer or malfunctioning watch whose failure is purely technical, a conversational AI can fail as a source of knowledge, not merely as an instrument.

6. The trust condition

Let us now move on to the next condition, condition (4), which I refer to as the *trust condition*. According to this condition, supported by Faulkner (2011), entities can only testify if they are objects of trust. For example, conservative theories of testimonial knowledge hold that the act of testifying entails a relationship of trust between hearer and speaker. However, as Miller and Freiman (2020) argues, only human beings can be the objects of trust relationships, which means that technologies, in principle, lack the property of trustworthiness. In other words, the notion of betrayal inherent in trust applies solely to agents capable of moral failure; namely, human beings, not technological artifacts. Thus, conversational AIs cannot be considered testifiers.

However, this conclusion does not necessarily follow, since conversational AIs can satisfy the trust condition for reasons similar to those discussed above in relation to the epistemic responsibility condition. I hereby argue that conversational AIs can be objects of both trust and responsibility. On the one hand, it can be argued that direct trust in artifacts is in fact indirect trust in the agents who design and maintain those artifacts. As Freiman (2024, 481) notes, "when a person trusts a bridge not to collapse, she actually trusts the people who built the bridge and the people who are responsible for its maintenance." Similarly, when we trust that a given conversational AI will not "hallucinate", we are ultimately placing trust in the programmers who designed the AI and in those responsible for supervising and operating it. Thus, trust in AI systems can be *reduced* to trust in the people and organizations behind these technologies.

reasoning, responsibility can be understood as the adherence to functional obligations; that is, an AI system is responsible when it operates within the functional norms established by its design and historical role. Functional failures, in this framework, imply irresponsibility, as they constitute a violation of these obligations.

On the other hand, the concept of "trust" can be subject to conceptual engineering, just as I previously proposed for the concept of "responsibility". According to Simion and Kelp (2023), being the object of trust can be understood as the disposition to fulfill one's obligations, where (i) artifacts can possess functions, and (ii) functions can generate obligations. Within this framework, conversational AIs can be objects of trust insofar as they acquire obligations through the acquisition of functions. Consequently, conversational AIs are trustworthy to the extent that they fulfill their functional obligations; obligations grounded in their creators' intentions or in a history of success in fulfilling a social purpose. For example, a customer-support conversational AI that tracks orders may be considered trustworthy if it reliably fulfills its designated functions, such as providing accurate delivery estimates, updating order status in real time, and resolving customer issues.[9]

7. The assertion condition

The last condition, present in the argument for the conservative view of testimony, is the so-called *assertion condition*. This condition (5) states that entities can only testify if they are capable of making assertions (cf. Fricker (1995) and Lackey (2008)). But what is an *assertion*? The speech act of assertion refers to the familiar phenomenon whereby a subject states, reports, contends, or claims that something is the case. But what distinguishes assertion from other speech acts, such as speculation or guessing? Goldberg (2015, 3) clarifies this by noting that "assertion is the unique speech act that is governed by a particular rule: the so-called norm of assertion". Thus, the speech act of assertion can be individuated by reference to this rule or norm. This norm typically has the following structure: one should assert that p only if ϕ, where "ϕ" is replaced with the condition that captures the content of the norm. There has been much debate over how best to formulate this condition. The main candidates for ϕ are the following: one knows that p (Williamson (2000)); one believes that p is true (Weiner (2005)); one is epistemically certain that p (Stanley (2008)); or it is reasonable for one to believe that p (Lackey (2008)). Although these perspectives diverge, they appear to share a common assumption: *grasping* is a necessary condition for assertion (cf. Kallestrup (2019)). Therefore, if an entity is unable to grasp or understand the meaning of proposition p, it cannot assert that proposition and, consequently, cannot testify that p. If conversational AIs

[9] This framework also provides a plausible response to the "black box" problem; i.e., the concern that our inability to understand or explain how an AI system arrives at a particular decision undermines its trustworthiness. On this view, trust does not require understanding or explainability but rather depends on the system's ability to consistently fulfill its functional obligations.

do not *grasp* or understand the propositions they generate via algorithmic processes, in other words, if AI systems operate as purely syntactic mechanisms rather than semantic ones, then they are incapable of testifying.[10]

As a critique of this last line of reasoning, we can advance two arguments. On the one hand, we can examine the epistemic aim of assertion.[11] This primary purpose does not seem to depend on the speaker's own grasp of the content; rather, it seems more plausible to hold that the relevant aim of assertion is to generate (or at least have the disposition to generate) some epistemic status in the hearer. There are two main reasons for this: First, because the social function of language is to convey or communicate information, often through the speech act of assertion. Second, because we are cognitively limited beings, that is, we cannot come to know many things firsthand or in isolation, we must rely on the words of others, particularly their assertions, to acquire knowledge. For similar reasons, Kelp (2016, 16) argues that assertion has the epistemic function of generating knowledge in hearers, and thus defends the following rule of assertion: one should assert that p only if it has the disposition to generate knowledge that p in hearers. If this functionalist norm of assertion is plausible, then we can say that conversational AIs are capable of making proper assertions. At the very least, what conversational AIs produce has the capacity to generate knowledge and grasp in their hearers about a given domain, even if the AIs themselves do not possess any knowledge and grasp (in the internalist sense) of that domain.[12]

On the other hand, we might accept a more externalist theory of grasping, which allows that conversational AIs do in fact grasp words. For instance, on the externalist account of Williamson (2006), to grasp a word is to be a member of a community that uses that word. Moreover, one counts as a member of such a community insofar as one participates in relevant causal interactions with other

[10] This idea is defended, for example, by Mahowald et al. (2024).

[11] It can also be argued that "grasping" is not a necessary condition for asserting, as Faria (2020) proposes.

[12] Other solutions are also available. One might argue, as Arora (2024) does, that the assertions of conversational AIs are "proxy assertions"; in other words, conversational AIs function as "proxies" (representatives) of a human or institutional agent that assumes responsibility for their assertions. Another possibility, as suggested by Mallory (2023), is to interpret interactions with conversational AIs as a form of make-believe, where we treat the AI as if it were a communicative agent, without actually believing that it has understanding or grasp. Finally, there is the proposal by Williams and Bayne (2024) that chatbots operate in an intermediate state of "proto-assertion", akin to a child learning to speak. In this view, conversational AIs may lack full human capacities (such as semantic understanding or grasp), yet they exhibit certain features of assertion (such as providing useful information, adapting responses, defending against criticism, learning from mistakes, etc.).

members of that community. In this context, Williamson (2006, 36) writes that such members "use a word as a word of a public language, allowing its reference in their mouths to be fixed by its use over the whole community." Conversely, an entity fails to grasp a word when there is a "lack of causal interaction with the social practice of using that word" (cf. Williamson (2006, 38)). On the basis of this theory, we can argue that conversational AIs grasp the words they produce and thus qualify as testifiers, insofar as they are "engaged in the practice" in which such words are used. That is, conversational AIs can be seen as members of a linguistic community through which they maintain causal interaction (e.g., via algorithms and large language models) in order to use words as humans do, for the purpose of forming assertions and testimony. Therefore, as we have shown throughout this paper, there are no compelling reasons to reject the claim that conversational AIs can testify.[13]

References

Arora, Chirag. 2024. "Proxy Assertions and Agency: The Case of Machine-Assertions." *Philosophy and Technology* 37 (1): 1–19. https://doi.org/10.1007/s13347-024-00703-5.

Azaria, Amos, and Tom Mitchell. 2023. "The Internal State of an LLM Knows When It's Lying." In *Findings of the Association for Computational Linguistics: EMNLP 2023*, edited by Houda Bouamor, Juan Pino, and Kalika Bali, 967–76. Singapore: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.68.

Coady, C.A.J. 1992. *Testimony: A Philosophical Study*. New York: Oxford University Press.

Faria, Domingos. 2020. "Group Testimony: Defending a Reductionist View." *Logos and Episteme* 11 (3): 283–304. https://doi.org/10.5840/logos-episteme202011322.

Faulkner, Paul. 2011. *Knowledge on Trust*. Oxford University Press.

Freiman, Ori. 2024. "AI-Testimony, Conversational AIs and Our Anthropocentric Theory of Testimony." *Social Epistemology* 38 (4): 476–90. https://doi.org/10.1080/02691728.2024.2316622.

[13] **Acknowledgments**: Earlier versions of this paper have been presented at the LanCog Seminar Series in Analytic Philosophy, the Workshop on Social Frameworks: From Epistemology to Aesthetic, the 9th Graduate Conference: Research Seminar Aesthetics, Politics and Knowledge Research Group, and the 1st National Meeting of PhiloTalks: Philosophy, Ethics and Education in Debate. I am grateful to the audience for the useful discussion. Any errors or omissions are my responsibility. Affiliation: Department of Philosophy, Faculty of Arts and Humanities, University of Porto, Via Panorâmica s/n, 4150 Porto, Portugal. E-mail: dfaria@letras.up.pt

Fricker, Elizabeth. 1995. "Critical Notice: Telling and Trusting: Reductionism and Anti-Reductionism in the Epistemology of Testimon." *Mind* 104 (414): 393–411. https://doi.org/10.1093/mind/104.414.393.

———. 2015. "How to Make Invidious Distinctions Amongst Reliable Testifiers." *Episteme* 12 (2): 173–202. https://doi.org/10.1017/epi.2015.6.

Goldberg, Sanford C. 2012. "Epistemic Extendedness, Testimony, and the Epistemology of Instrument-Based Belief." *Philosophical Explorations* 15 (2): 181–97. https://doi.org/10.1080/13869795.2012.670719.

———. 2015. *Assertion: On the Philosophical Significance of Assertoric Speech*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198732488.001.0001.

Herrmann, Daniel A., and Benjamin A. Levinstein. 2025. "Standards for Belief Representations in Llms." *Minds and Machines* 35 (1): 1–25. https://doi.org/10.1007/s11023-024-09709-6.

Jones, Cameron R, and Benjamin K Bergen. 2025. "Large Language Models Pass the Turing Test." *arXiv Preprint arXiv:2503.23674*.

Kallestrup, Jesper. 2019. "Groups, Trust, and Testimony." In *Trust in Epistemology*, 136–58. Routledge. https://doi.org/10.4324/9781351264884-6.

Kelp, Christoph. 2016. "Assertion: A Function First Account." *Noûs* 52 (2): 411–42. https://doi.org/10.1111/nous.12153.

Lackey, Jennifer. 2008. *Learning from Words: Testimony as a Source of Knowledge*. Oxford University Press.

Lederman, Harvey, and Kyle Mahowald. 2024. "Are Language Models More Like Libraries or Like Librarians? Bibliotechnism, the Novel Reference Problem, and the Attitudes of LLMs." *Transactions of the Association for Computational Linguistics* 12: 1087–1103.

Mahowald, Kyle, Anna A Ivanova, Idan A Blank, Nancy Kanwisher, Joshua B Tenenbaum, and Evelina Fedorenko. 2024. "Dissociating Language and Thought in Large Language Models." *Trends in Cognitive Sciences*.

Mallory, Fintan. 2023. "Fictionalism about Chatbots." *Ergo: An Open Access Journal of Philosophy* 10 (n/a). https://doi.org/10.3998/ergo.4668.

Marks, Samuel, and Max Tegmark. 2024. "The Geometry of Truth: Emergent Linear Structure in Large Language Model Representations of True/False Datasets." In *First Conference on Language Modeling*. https://openreview.net/forum?id=aajyHYjjsk.

Miller, Boaz, and Ori Freiman. 2020. "Trust and Distributed Epistemic Labor." In *The Routledge Handbook of Trust and Philosophy*, 341–53. Routledge.

Mitchell, Melanie. 2024. "The Turing Test and Our Shifting Conceptions of Intelligence." *Science* 385 (6710): eadq9356.

Narayanan, Devesh, and David De Cremer. 2022. "'Google Told Me so!' On the Bent Testimony of Search Engine Algorithms." *Philosophy and Technology* 35 (2): 1–19. https://doi.org/10.1007/s13347-022-00521-7.

Nissenbaum, Helen. 1996. "Accountability in a Computerized Society." *Science and Engineering Ethics* 2 (1): 25–42. https://doi.org/10.1007/bf02639315.

Pagin, Peter, and Neri Marsili. 2021. "Assertion." In *Stanford Encyclopedia of Philosophy*.

Shanahan, Murray. 2024. "Talking about Large Language Models." *Communications of the ACM* 67 (2): 68–79.

Simion, Mona, and Christoph Kelp. 2023. "Trustworthy Artificial Intelligence." *Asian Journal of Philosophy* 2 (1): 8.

Stanley, Jason. 2008. "Knowledge and Certainty." *Philosophical Issues* 18 (1): 35–57. https://doi.org/10.1111/j.1533-6077.2008.00136.x.

Tollefsen, Deborah Perron. 2009. "Wikipedia and the Epistemology of Testimony." *Episteme* 6 (1): 8–24. https://doi.org/10.3366/e1742360008000518.

Weiner, Matthew. 2005. "Must We Know What We Say?" *Philosophical Review* 114 (2): 227–51. https://doi.org/10.1215/00318108-114-2-227.

Williams, Iwan, and Tim Bayne. 2024. "Chatting with Bots: AI, Speech Acts, and the Edge of Assertion." *Inquiry*, 1–24.

Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford University Press. https://doi.org/10.1093/019925656x.001.0001.

———. 2006. "'Conceptual Truth'." *Aristotelian Society Supplementary Volume* 80 (1): 1–41. https://doi.org/10.1111/j.1467-8349.2006.00136.x.

Woudenberg, René van, Chris Ranalli, and Daniel Bracker. 2024. "Authorship and ChatGPT: A Conservative View." *Philosophy and Technology* 37 (1): 1–26. https://doi.org/10.1007/s13347-024-00715-1.

Wright, Stephen. 2018. *Knowledge Transmission*. Routledge. https://doi.org/10.4324/9781315111384.