

Is Generative AI Ready to Join the Conversation That We Are? Gadamer's Hermeneutics after ChatGPT

Robert Hornby

Abstract

In this article, I use the dialogical ideas of Hans-Georg Gadamer to evaluate whether generative AI is ready to join the ontological conversation that he considers humanity to be. Despite the technical advances of generative AI, Gadamer's philosophical hermeneutics reveals that it cannot function as a proxy human dialogue partner in pursuit of understanding. Even when free from anthropomorphic projections and reimagined as the "other", generative AI is found to have a weak epistemology, lack of moral awareness, and no emotions. Even so, it evokes a response in some users that places it on the threshold of being. The most promising dialogical role identified for generative AI is as a digital form of Gadamerian "text" currently constrained by copyright and technical design. Generative AI's shortcomings risk inhibiting hermeneutical understanding through greater access to summarised knowledge. Nonetheless, the new technology is on the brink of joining the ontological conversation of humanity.

Keywords: Generative AI, ChatGPT, Hans-Georg Gadamer, Philosophical Hermeneutics, Fusion of Horizons, Dialogue.

1. Introduction

Generative AI is a rare technological innovation in that it poses substantive ontological and epistemological questions. This new form of AI is based on a probabilistic representation of knowledge encoded in language and accessed through intuitive dialogue. It can produce novel multi-media responses that emulate various styles, formats, and genres. There has been considerable speculation about the dialogical capabilities of generative AI, particularly in comparison to humans, and whether it might threaten our place in the world. So far, these uncertainties have not been tested using a relevant and comprehensive philosophical framework.

In this article, I will use the enduring ideas of Hans-Georg Gadamer to evaluate whether generative AI can engage in the ontological conversation that he considers humanity to be.¹ I will argue that generative AI falls short of a human dialogue partner in several important ways but can be considered to usefully reinterpret secondary aspects of Gadamer's theory as both "other" and "text". Even so, significant challenges remain, but with a path forward to more hermeneutical significance, especially as a proxy for a "text".

Methodologically, I first establish the conceptual foundations of the analysis, limiting myself to the core aspects of each domain to avoid excessive diffusion of the argument. This step sets up a dialogical and, at times, dialectical evaluation of generative AI using the key themes of Gadamer's work, organised around my reinterpretation of his different dialogical actors ("person", "other", and "text"). Gadamer's proposals are afforded normative priority for this exercise, but generative AI also has the opportunity to extend and revise aspects of Gadamer's proposals. Afterwards, some conclusions and speculations on future developments are offered.

2. Foundations

2.1 Recent Advances in Generative AI

In November 2022, ChatGPT from OpenAI burst onto the scene, catching many within the AI community off guard.² Suddenly, anyone with a connected device could interrogate a vast dataset and obtain human-like responses in their preferred style, format, and language. ChatGPT could manipulate text, images, and other forms of media, including software code. This capacity to be creative was the source of much

1 Hans-Georg Gadamer, *Truth and Method*, Second Edition, (London: Bloomsbury Academic, 1989).

2 OpenAI, "Introducing ChatGPT," *OpenAI Blog*, last modified November 30, 2022. <https://openai.com/blog/chatgpt>.

excitement, and ChatGPT's adoption became a popular movement, with record-breaking downloads exceeding a hundred million in the first two months after launch.³ Access to AI had finally moved beyond the specialist community and into the hands of users.

The reaction was hyperbolic, and dystopian predictions about the existential threat of AI garnered mass support.⁴ Some commentators talked about generative AI as if it could be conscious.⁵ Geoffrey Hinton, a respected AI pioneer, left Google warning of significant dangers ahead.⁶ However, moderate voices eventually began to speak out, counterbalancing the hype arising from a radically new state-of-the-art that was not yet widely understood.⁷

The technology that underpins ChatGPT and similar offerings rapidly coming to the market in its wake⁸ arises from a branch of AI called foundation models.⁹ These large, generalised models are trained on unlabelled data and have replaced many smaller task-specific equivalents trained with labelled data that have dominated AI research for decades. Large language models (LLMs) are the variant of foundation models used in generative AI, adopting a *transformer* neural network architecture that is well adapted to the predictive parallel processing of large datasets.¹⁰ LLMs are so recent, despite the field of natural language processing dating back to the 1950s, because they require expansive digital datasets, vast and economic computational power, and new forms of algorithm to exploit them.

LLMs are typically measured by the number of parameters (weights in the neural network) configurable

3 Krystal Hu, "ChatGPT Sets Record for Fastest-Growing User Base," *Reuters Analyst Note*, last modified February 2, 2023, <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>.

4 Center for AI Safety, "Statement on AI Risk: AI Experts and Public Figures Express Their Concern About AI Risk," last modified May 29, 2023, <https://www.safe.ai/statement-on-ai-risk>.

5 Kevin Rose, "A Conversation With Bing's Chatbot Left Me Deeply Unsettled," *New York Times*, last modified February 16, 2023, <https://www.nytimes.com/2023/02/16/technology/bing-chatbot-micro-soft-chatgpt.html>.

6 Cade Metz, "'The Godfather of A.I.' Leaves Google and Warns of Danger Ahead," *New York Times*, last modified May 4, 2023, <https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton.html>.

7 George Hammond, "AI Threat to Human Existence is 'Absurd' Distraction from Real Risks," *Financial Times*, last modified June 16, 2023, <https://www-ft-com.ezp.lib.cam.ac.uk/content/732fc372-67ea-4684-9ab7-6b6f3cdfd736>.

8 Other examples of generative AI platforms: Bard is the chatbot for the PaLM 2 LLM from Google; LLaMA 2 is an open-source LLM from Meta, partnering with Microsoft; BLOOMChat is the chatbot for the open-source BLOOM LLM from BigScience and hosted by the Hugging Face platform.

9 Adam Kolides, Alyna Nawaz, Anshu Rathor, et al., "Artificial Intelligence Foundation and Pre-Trained Models: Fundamentals, Applications, Opportunities, and Social Impacts," *Simulation Modelling Practice and Theory*, no. 126 (2010), <https://doi.org/10.1016/j.simpat.2023.102754>.

10 GPT stands for Generative Pretrained Transformer.

in the training process. GPT-4 (GPT is the LLM underpinning ChatGPT) has 175bn, more than 100 times larger than GPT-2 with 1.5bn.¹¹ The data sources used to train the models are not fully disclosed but include the extensive archives of the non-profit web-scraping organisation Common Crawl.¹²

When ChatGPT is asked a question (known as a prompt), it first tokenises the input sequence. Tokens are the underlying building blocks of words (or code). LLMs then use an attention network (that somewhat mimics human cognitive attention) to focus on the most important parts of the prompt, generating a probabilistic response one token at a time while re-evaluating all previous responses using autoregression.¹³ The complete response is then de-tokenized to return it to natural language or another medium.

The outcomes are not foolproof and can be inaccurate or contradictory (known as hallucinations) or reflect the social bias of their training data.¹⁴ These issues will be explored in more depth below. However, whatever problems remain to be addressed, generative AI is undoubtedly a significant *technical* advance.

2.2 Gadamer's Philosophical Hermeneutics

It is evident that the most relevant philosophical work to evaluate generative AI should be centred on language and dialogue. I initially began with Ludwig Wittgenstein (1889–1951), but his work proved too opaque and contested to bring clarity to another domain. The divergence between his earlier and later work only added to this problem. Ultimately, I wrote more caveats about Wittgenstein than insights about generative AI.

Another towering figure of dialogue and textual interpretation is Jacques Derrida (1930–2004). However, his ideas of deconstruction are pluralistic and self-consciously open-ended, mainly concerned with what *cannot* be known in dialogue.¹⁵ Derrida's approach may have cut through the considerable hype

11 Tom B. Brown, Benjamin Mann, Nick Ryder, et al., "Language Models are Few-Shot Learners," in *Advances in Neural Information Processing Systems* 33, edited by H M Larochelle, et al., 1877–1901 (Red Hook, New York: Curran Associates, 2020).

12 Common Crawl, "Overview," last modified June 16, 2023, <https://commoncrawl.org/overview>.

13 Ashish Vaswani, Noam Shazeer, Niki Parmar, et al., "Attention is All You Need," In *Advances in Neural Information Processing Systems* 30, edited by I U Guyon, et al., (Red Hook, New York: Curran Associates, 2017).

14 Paul Pu Liang, Chiyu Wu, Louis-Philippe Morency, et al., "Towards Understanding and Mitigating Social Biases in Language Models," *Proceedings of the 38th International Conference on Machine Learning*, no. 139 (2021), <https://proceedings.mlr.press/v139/liang21a.html>.

15 Chant  lle Swartz & Paul Cilliers, "Dialogue Disrupted: Derrida, Gadamer and the Ethics of Discussion," *South African Journal of Philosophy* 22, no. 1 (2003): 1–18, <https://www.tandfonline.com/doi/abs/10.4314/sajpem.v22i1.31357>.

surrounding generative AI but without much left to affirm. Nonetheless, I return to him briefly below.

After Wittgenstein and Derrida, Hans-Georg Gadamer (1900–2002) is the most prominent philosopher of language and dialogue. Once a student of Heidegger at Marburg, his seminal work *Truth and Method* (T&M), on which I will base this analysis, was published in German in 1960 and subsequently translated into English in 1975 (second edition 1989).¹⁶ Gadamer's ideas not only have relevance to generative AI but also benefit from coherence, resilience, and recent applications to other digital topics.¹⁷ His embrace of art and texts further aligns with the multi-media aspect of the new technology. In selecting Gadamer's work for this evaluation, I am not advocating for his philosophical scheme *per se* but rather endorsing its value for the matter at hand. That said, I am convinced about the hermeneutical value of dialogue, not least because it is so deeply embedded in a well-functioning society.¹⁸

The enduring significance of Gadamer's T&M is underlined by a collection of essays that mark fifty years since its first publication.¹⁹ In the last three years alone, four new or revised major works on Gadamer have also been published, albeit by a relatively small cohort of scholars.²⁰ T&M is primarily a defence of truth-experience in the humanities amidst challenges from positivist scientific epistemologies. Although some describe Gadamer as a relativist, Chodos et al. argue persuasively that Gadamer articulates a precursor to critical realism, which they designate *minimal realism*.²¹ However, this debate is not central to my evaluation for reasons that become clear below.

T&M is presented in three sections: art, the humanities, and language. Rather than following a linear progression, they form a widening spiral from the specific of art to the universal of language.²² On art, Gadamer argues against a purely aesthetic understanding, claiming that the genre has its own

16 Gadamer, *Truth and Method*.

17 Matthew S. Lindia, "Gadamer in a Wired Brain: Philosophical Hermeneutics and Neuralink," *Philosophy and Technology* 35, (2022): 27, <https://doi.org/10.1007/s13347-022-00522-6>; Jure Zovko, "Expanding hermeneutics to the world of technology," *AI & Society* 38, (2023): 2243–2254, <https://doi.org/10.1007/s00146-020-01052-5>.

18 I am thinking of examples such as public debate, academic peer review, and restorative or adversarial justice.

19 Jeff Malpas and Santiago Zabala, eds., *Consequences of Hermeneutics: Fifty Years After Gadamer's Truth and Method*, (Evanston, Illinois: Northwestern University Press, 2010).

20 Robert J. Dostal, ed., *Cambridge Companion to Gadamer*, Second Edition, (Cambridge, UK: Cambridge University Press, 2021); Cynthia R. Nielsen and Greg Lynch, eds., *Gadamer's Truth and Method: A Polyphonic Commentary* (London: The Rowman & Littlefield Publishing Group, 2022); Cynthia R Nielsen, *Gadamer's Hermeneutical Aesthetics: Art as a Performative, Dynamic, Communal Event* (New York: Routledge, 2022); Robert J. Dostal, *Gadamer's Hermeneutics: Between Phenomenology and Dialectic* (Evanston, Illinois: Northwestern University Press, 2022).

21 Howie Chodos, Bruce Curtis, Alan Hunt, and John Manwaring, "Gadamer's Minimal Realism," in *Critical Realism and the Social Sciences: Heterodox Elaborations*, ed. by Jon Frauley and Frank Pearce, 296–316, (Toronto: University of Toronto Press, 2007), <https://doi.org/10.3138/9781442684232-018>.

22 Nielsen and Lynch, *Gadamer's Truth and Method*, 28.

ontology.²³ This “truth” is revealed through a performance “event” in which the work is conveyed to the viewer. Gadamer also explores the interpretive value of “play” as a “to-and-fro” exploration dynamic.

On the humanities, Gadamer proposes that contrary to Kantian claims of objectivity, our understanding and consciousness are situated in the preunderstanding of history and subject to our “prejudices.”²⁴ Gadamer’s conception of truth is self-consciously rooted in Aristotle’s *phronēsis* (practical wisdom) rather than more abstract categories.²⁵ His illustrations of the contextual nature of meaning include the judicial process, where the law, although consistent, is applied differently depending on the specifics of the case.²⁶

Finally, on language, Gadamer argues that all understanding is achieved through conversation, be it a dialogue between two people or the internal questions posed by the interpreter of a text.²⁷ Unlike Schleiermacher, he suggests that texts, by which Gadamer means whole and meaningful works, can be separated from the intentions of their authors and speak to the reader for themselves. Gadamer rejects a semiotic conception of language as merely a sign or instrument of thought. Instead, via an excursus through medieval trinitarian theology, he argues for a closely integrated relationship between thought and word.²⁸ Gadamer asserts that “being that can be understood is language”²⁹ while acknowledging aspects of reality that lie beyond it.³⁰

Gadamer’s approach to dialogical understanding culminates in a “fusion of horizons.” For him, a horizon is “The range of vision that includes everything that can be seen from a particular vantage point.”³¹ In conversation, the parties explore and ultimately join their respective horizons within a shared world and seek to reach an agreement (literally, “come over to become one”) about the “matter at hand.”³² Gadamer asserts that productive discourse also requires openness, trust, charity, humility, and goodwill but notes that this aspiration cannot always be achieved.³³ Even so, hermeneutical dialogue is not just

23 Gadamer, *Truth and Method*, Ch. 1–2.

24 Gadamer, *Truth and Method*, Ch. 3–4.

25 Gadamer, *Truth and Method*, 324–333.

26 Gadamer, *Truth and Method*, 338–339.

27 Gadamer, *Truth and Method*, Ch. 5.

28 Gert-Jan van der Heiden, “Gadamer and the Concept of Language,” in *Gadamer’s Truth and Method: A Polyphonic Commentary*, ed. by Cynthia R. Nielsen and Greg Lynch, 380–400, (London: The Rowman & Littlefield Publishing Group, 2022).

29 Gadamer, *Truth and Method*, 490.

30 James Risser, “On Language and the Universality of Hermeneutics,” In *Gadamer’s Truth and Method: A Polyphonic Commentary*, edited by Cynthia R Nielsen and Greg Lynch, 420–424, (London: The Rowman & Littlefield Publishing Group, 2022).

31 Gadamer, *Truth and Method*, 313.

32 Dostal, *Gadamer’s Hermeneutics*, 121.

33 Dostal, *Gadamer’s Hermeneutics*, 82.

a means to gain understanding but also a communal ontology, expressing the unfinished “conversation that we ourselves are”.³⁴

Given the time since Gadamer published T&M, his ideas have inevitably undergone sustained critical evaluation. Derrida mounted a particularly disdainful challenge, famously posing three questions to Gadamer in 1981.³⁵ In them, he disputed the relevance of historical situation to understanding, the universality of hermeneutics, and the ability to fuse horizons. Derrida upheld inherent indeterminacy in dialogue and texts, such that meaning is always open-ended. It could be argued that Derrida's assertions undermine the value of all human communication, but he rightly warns against overestimating the completeness of any particular interpretation or the extent of “fusion” possible in dialogue.³⁶

Habermas, another contemporary, argued that Gadamer's conception of tradition could preserve forms of oppressive ideology that distort understanding. He called for a conscious step back from individual context to evaluate it.³⁷ This reflexive consideration is now normalised in academic discourse, and Habermas was right to highlight the risks of viewing context as unconscious and predetermined. More recently, Chodos et al. question whether Gadamer's idea of context is also too culturally homogeneous and fails to reflect the realities of the diverse, pluralistic societies that have arisen since T&M.³⁸ This sociological argument has merit, and I will adjust for Gadamer's lack of foresight on this point in the discussion below.

On language, Michel draws on American pragmatists James and Dewey to critique Gadamer and affirm the role of non-discursive and pre-linguistic experience in meaning.³⁹ Dostal also recognises the legitimacy of pre-linguistic factors in his recent major work on Gadamer.⁴⁰ Sims-Schouten, Riley, and Willig⁴¹ provide a useful non-discursive taxonomy comprising embodiment, institutional power,

34 Gadamer, *Truth and Method*, 378–386.

35 Chant  lle Swartz and Paul Cilliers, “Dialogue Disrupted: Derrida, Gadamer and the Ethics of Discussion,” *South African Journal of Philosophy* 22, no. 1 (2003): 1–18, <https://www.tandfonline.com/doi/abs/10.4314/sajpem.v22i1.31357>.

36 Swartz & Cilliers, “Dialogue Disrupted.”

37 J  rgen Habermas, “The Hermeneutic Claim to Universality,” in *Contemporary Hermeneutics: Method, Philosophy, and Critique*, ed. by Josef Bleicher, 181–211, (London: Routledge, 1980).

38 Chodos et al., “Gadamer's Minimal Realism,” 296–316.

39 Johann Michel, “Meaning and Experience,” *European Journal of Pragmatism and American Philosophy* XIV-1, (2022), <https://doi.org/10.4000/ejpap.2745>.

40 Dostal, *Gadamer's Hermeneutics*, 139–144.

41 Wendy Sims-Schouten, Sarah C. E. Riley, and Carla Willig, “Critical Realism in Discourse Analysis: A Presentation of a Systematic Method of Analysis Using Women's Talk of Motherhood, Child-care, and Female Employment as an Example,” *Theory & Psychology* 17, no. 1 (2007): 101–124, <https://doi.org/10.1177/0959354307073153>.

and materiality, which they argue influence and limit the range of discursive possibilities. Given the emerging consensus on this matter, I will include a consideration of non-discursive factors below, despite their absence from Gadamer's work.

In summary, the central arguments of Gadamer's T&M, augmented by key points from subsequent criticism (added in brackets), can be summarised as follows: understanding is rooted in history and prejudice (i.e., situated in sociologically diverse reflexive locations) that should be identified (evaluated and challenged). Language is an expression of being and a medium of understanding but is not synonymous with reality, which exists beyond language (and is influenced by non-discursive factors). Understanding is achieved by a fusion of horizons in discourse (but indeterminacy and a failure to grasp the "other" remain significant risks). Hermeneutical understanding is practical wisdom, i.e., *phronēsis*, rather than merely abstract knowledge, and dialogue expresses a collective ontology, defining humans themselves as "conversations."

3. Evaluating Generative AI Using Gadamer's Hermeneutics

Gadamer assumes a human dialogical actor is always at the centre of the hermeneutical process. In this evaluation, the user of generative AI is the equivalent constant. As noted above, this primary dialogical actor enters into conversation with another human or a "text". Consequently, I will first evaluate whether generative AI can be a convincing proxy for a human conversation partner. However, Gadamer also encourages an embrace of the "other," by which he means diverse human interlocutors. I will extend this idea to propose generative AI as another category of dialogical actor altogether, defined bottom up by its inherent characteristics. Finally, I will evaluate generative AI as a potential proxy for a Gadamerian "text."

As a supplementary aspect of the approach, I also prompted both ChatGPT-2.5 and 4 to compare themselves to Gadamer's theory. I use parts of these responses to illustrate key points, showing them as attributed quotes at the beginning of the relevant sections below. The complete responses are reproduced as appendices.

3.1 Generative AI as a Proxy for a Human Dialogue Partner

I will use the attributes that Gadamer ascribes to a human dialogical partner to evaluate whether generative AI can act as a proxy, namely prejudice, fallibility, a situational context in history, a viewpoint arising from this reflexive position defined as a "horizon," and the capacity for dialogical

virtue, particularly practical wisdom (*phronēsis*).

Prejudice and Fallibility

“ChatGPT operates on machine learning algorithms and data. It doesn’t possess preconceptions or prejudices.” (ChatGPT-2.5)

“The model is inevitably shaped by the cultural, historical, and contextual nuances present in that data.” (ChatGPT-4)

As noted, Gadamer argues that all humans are affected by prejudice due to their historical preconceptions. ChatGPT-2.5 asserts in the quote above that it has no such biases, but ChatGPT-4 suggests otherwise.

Before generative AI was launched, biases in search engines⁴² and social media content curation were well-documented.⁴³ Hargittai also made a credible case for bias towards privilege in large datasets derived from social media.⁴⁴ Although LLMs are relatively new, a consistent picture of bias is already emerging.⁴⁵ This should not be a surprise, given that all LLM training data is derived from published sources on the internet, and, to my knowledge, no one has ever argued for the objectivity of that corpus. It can be claimed, with some merit, that since generative AI is trained on aggregated datasets, it avoids the worst excesses of social media confirmation bias. However, O’Neil has demonstrated that even large datasets reinforce societal bias as what she calls “weapons of math destruction.”⁴⁶ For ChatGPT-4, at least, OpenAI is straightforward about the issue, noting that “GPT-4 still has many known limitations that we are working to address, such as social biases.”⁴⁷

The current technical consensus is that bias in LLMs is highly problematic and must be eliminated urgently. However, Gadamer has a counternarrative that unbiased, objective knowledge is an

42 Eric Goldman, “Search Engine Bias and the Demise of Search Engine Utopianism,” 8 *Yale J.L. & Tech*, 188 (2006), <http://digitalcommons.law.scu.edu/facpubs/76>.

43 Nazanin Alipourfard, Buddhika Nettasinghe, Andrés Abeliuk, et al., “Friendship Paradox Biases Perceptions in Directed Networks,” *Nature Communications* 11, (2020): 707, <https://doi.org/10.1038/s41467-020-14394-x>.

44 Eszter Hargittai, “Potential Biases in Big Data: Omitted Voices on Social Media,” *Social Science Computer Review* 38, no. 1 (2020): 10–24, <https://doi.org/10.1177/0894439318788322>.

45 Liang et al., “Towards Understanding and Mitigating Social Biases in Language Models.”

46 Cathy O’Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (London, UK: Penguin Books, 2016).

47 OpenAI, “GPT-4,” last modified March 14, 2023, <https://openai.com/research/gpt-4>.

Enlightenment fantasy and prejudices should rather be exposed and challenged in conversation. Valuable efforts are underway to reduce the most blatant forms of discrimination in generative AI platforms,⁴⁸ but Wahl-Jorgensen et al. demonstrate the innate subjectivity of bias in their analysis of claims about the BBC media organisation.⁴⁹ As such, there will always be a perception that generative AI exhibits unacceptable predispositions because there can be no consensus on what is impartial.⁵⁰

Inaccuracies and contradictions in generative AI responses, known as *hallucinations*, are a similar concern within the AI community.⁵¹ In some settings, such as the medical field, wrong responses can have serious consequences, so independent validation is always needed. These problems will become less frequent and severe as generative AI technology matures, but they cannot be eliminated entirely.⁵² Again, Gadamer's solution is to identify inaccuracies through conversation, which is itself a form of validation.⁵³

Overall, generative AI exhibits both prejudice and fallibility, similar to a human dialogue partner. The former is an aggregation of preexisting societal biases, but hallucinations are a technical limitation introduced by the new technology itself. Nonetheless, neither disqualifies generative AI from being a Gadamerian proxy for a person but does infer the need for safeguards.

Historical Situatedness

“While ChatGPT is aware of the context in the immediate conversation (the previous messages), it lacks broader historical or cultural awareness.” (ChatGPT-2.5)

Gadamer's assertion that understanding is deeply rooted in historically effected consciousness leads to the question of how generative AI represents history in general and its own situatedness in particular. Firstly, the history available to LLMs is *digital* history. Zaagsma argues persuasively that the digitisation of history is highly politicised and selective, noting that decisions about what to digitise, how to make

48 Liang et al., “Towards Understanding and Mitigating Social Biases in Language Models.”

49 Karin Wahl-Jorgensen, Mike Berry, Iñaki Garcia-Blanco, et al., “Rethinking Balance and Impartiality in Journalism? How the BBC Attempted and Failed to Change the Paradigm,” *Journalism* 18, no. 7 (2017): 781–800, <https://doi.org/10.1177/1464884916648094>.

50 Gadamer, *Truth and Method*, Ch. 3–4.

51 Niels Mündler, Jingxuan He, Slobodan Jenko, et al., “Self-Contradictory Hallucinations of Large Language Models: Evaluation, Detection and Mitigation,” Open Source, (2023), <https://doi.org/10.48550/arXiv.2305.15852>.

52 Mündler et al. “Self-Contradictory Hallucinations of Large Language Models.”

53 Gadamer, *Truth and Method*, 480.

it available, and to whom tend to overrepresent the desired national stories of governments.⁵⁴ History, as depicted by generative AI, is therefore subject to the prejudice already discussed above.

However, Gadamer is especially interested in the situational histories that provide the reflexive “horizon” for dialogical understanding. A blog by Bauckhage reports an attempt to provoke ChatGPT into disclosing conscious awareness.⁵⁵ He fails to achieve this, but ChatGPT does suggest it has some capacity to “maintain context and history” restricted to previous interactions with the same user.⁵⁶ Otherwise, in the same exchange, ChatGPT describes itself as a purveyor of output based on stored parameters and neural network weights but without a broader context or inner state. This is consistent with the quote from ChatGPT 2.5 at the top of this section.

Lost Horizons

“In a way, ChatGPT’s “horizon” is defined by its training data, which captures a particular snapshot of human language and knowledge.” (ChatGPT-4)

In contrast, the more evolved ChatGPT-4, cited above, suggests it could be possible to construe the historical horizon of a generative AI platform as the sum of its training data. However, only the parameter values within the neural network persist after training, not the vast dataset used to establish them. Hence, when ChatGPT-4 cites training data as a potential “horizon,” it refers to something no longer accessible to the LLM or its users. It is true that humans do not retain all the contextual data they are exposed to either, but for them, there is a cumulative build-up of long-term memory, context, and experience that generative AI does not currently replicate.⁵⁷

This challenge would be mitigated by some form of traceability from LLM parameter values back to the lost training data. However, this is not currently possible, and LLMs are opaque and challenging to interrogate even by their developers. A simple step forward would be to disclose the corpus of data used to train the LLM, but so far, no provider of generative AI has chosen to do this, perhaps fearing copyright disputes. Consequently, generative AI has no reflexive location beyond a short-term record of recent interactions.

54 Gerben Zaagsma, “Digital History and the Politics of Digitization,” *Digital Scholarship in the Humanities* 38, no. 2 (2023): 830–851, <https://doi.org/10.1093/lhc/fqac050>.

55 Christian Bauckhage, “ChatGPT: Has a Chatbot Finally Achieved Self-Awareness?” *LAMARR Institute for Machine Learning and Artificial Intelligence*, last modified February 10, 2023, <https://lamarr-institute.org/chatgpt-has-a-chatbot-finally-achieved-self-awareness/>.

56 GPT-4 has a short-term working memory capacity of 32k tokens or about 50 pages of text.

57 Eduardo Camina and Francisco Güell, “The Neuroanatomical, Neurophysiological, and Psychological Basis of Memory: Current Models and Their Origins,” *Frontiers in Pharmacology* 8, (2017), <https://doi.org/10.3389/fphar.2017.00438>.

Virtues and Phronēsis

From an ethical perspective, generative AI has no morality beyond that derived probabilistically from training data and anything hard-coded by its developers. However, it can read emotions and respond empathetically, even beyond the expertise of humans.⁵⁸ Nonetheless, this is not the same as possessing emotions.⁵⁹ Consequently, generative AI cannot emulate Gadamer's hermeneutical virtues of openness, trust, charity, humility, and goodwill because they necessitate moral choices and a positive emotional disposition.⁶⁰

Even so, Eisikovits and Feldman claim that AI can perform some types of *phronēsis* better than humans,⁶¹ a virtue that Gadamer placed above all others.⁶² On further examination, however, the “humdrum” real-world problems the authors describe are simply routine tasks. Moreover, their examples revolve around multi-attribute decision-making (MADM), which is solved quantitatively rather than hermeneutically. The proficiency of AI cited in their arguments is, therefore, not *phronēsis* in any Gadamerian sense because there is no requirement for wisdom, just mathematics.

Machine vs. Human Intelligence

The original assessment of machine intelligence is the Turing Test, which famously argued that if a human interrogator was unable to differentiate between a human and a machine with which they were simultaneously in a text-based question-and-answer dialogue, the machine could be designated as intelligent.⁶³ Searle, using the analogy of a Chinese Room, counters that merely achieving effective “symbol manipulation” in response to a “program” is not evidence of intelligence, concluding:

No computer program by itself is sufficient to give a system a mind. Programs, in short, are not minds, and they are not by themselves sufficient for having minds.⁶⁴

58 Zohar Elyoseph, Dorit Hadar-Shoval, Kfir Asraf, et al., “ChatGPT Outperforms Humans in Emotional Awareness Evaluations,” *Frontiers in Psychology* 14, (2023), <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1199058>.

59 Amanda Ruggeri, “The Surprising Promise and Profound Perils of AIs That Fake Empathy,” *Economist*, March 6, 2024, <https://www.newscientist.com/article/mg26134810-900-the-surprising-promise-and-profound-perils-of-ais-that-fake-empathy/>.

60 Dostal, *Gadamer's Hermeneutics*, 82.

61 Nir Eisikovits and Dan Feldman, “AI and Phronesis,” *Moral Philosophy and Politics* 9, no. 2 (2022): 181–199, <https://doi.org/10.1515/mopp-2021-0026>.

62 Gadamer, *Truth and Method*, 324–333.

63 Alan M. Turing, “Computing machinery and intelligence,” *Mind* 59, (1950): 433–460.

64 John Searle, *Minds, Brains and Science*, (Cambridge, Massachusetts: Harvard University Press, 1984), 40.

In contrast to Searle's semiotic view, Gadamer conflates **language with thought** and **the disclosure of reality itself**, which only widens the ontological gap between programs and minds.⁶⁵ Similarly, the hermeneutical requirement for a meaningful fusion of horizons is beyond simple question-and-answer dynamics and further undermines the veracity of the Turing Test. As such, any ability of generative AI to mimic human conversation cannot be considered to infer machine intelligence defined by a human standard.

Overall, generative AI does not exhibit the Gadamerian attributes associated with a human dialogue partner except for prejudice and fallibility and cannot function as a proxy for one.

3.2 Generative AI as "Other"

"ChatGPT, being an AI language model, does not engage in dialogue or conversation in the same sense as humans." (ChatGPT-2.5).

"Gadamer emphasises the importance of being open to the "Other" in the interpretive process." (ChatGPT-4)

I begin this new evaluation by clearing the ground of a powerful but misleading predisposition related to the previous discussion. The tendency toward the anthropomorphism of technology likely began with musical instruments in the early seventeenth century.⁶⁶ Li & Suh found this trend accelerating in published AI research even before the launch of ChatGPT, apparently motivated by an increase in trust and acceptance of technology that exhibits human characteristics.⁶⁷ Paradoxically, some personifying of AI also seems designed to gain attention by playing on our existential fears.⁶⁸

However, a thematic assessment of the *MIT Technology Review*,⁶⁹ as a leading example of its genre, shows

65 Gadamer, *Truth and Method*, Ch. 5.

66 Rebecca Cypess and Steven Kemper, "The Anthropomorphic Analogy: Humanising Musical Machines in the Early Modern and Contemporary Eras," *Organised Sound* 23, no. 2 (2018): 167–180, <https://doi.org/10.1017/S1355771818000043>.

67 Mengjun Li and Ayoung Suh, "Machinelike or Humanlike? A Literature Review of Anthropomorphism in AI-Enabled Technology," in *Proceedings of the 54th Hawaii International Conference on System Sciences*, (2021): 4053–4062, <https://doi.org/10.24251/HICSS.2021.493>.

68 Michael Szollosy, "Freud, Frankenstein and Our Fear of Robots: Projection in Our Cultural Perception of Technology," *AI & Society: Knowledge, Culture, and Communication* 32, (2017): 433–439, <https://doi.org/10.1007/s00146-016-0654-7>.

69 The MIT Review, "MIT Technology Review," *AI*, MIT, <https://www.technologyreview.com/topic/artificial-intelligence/>.

AI innovation is typically **ally orientated towards technical ecosystems that are virtual, interconnected, functionally specialised,⁷⁰ embedded rather than embodied,⁷¹** and focused on tasks that humans perform poorly (as measured by scale, speed, precision, or survival). These differences are often the design purpose of AI and bring benefits that complement rather than replicate human capabilities. As such, AI, including generative AI, shares little underlying resemblance to humans. Even the truism of the similarity of neural networks to the human brain is contested.⁷² In light of this, it is more accurate to consider generative AI as a category of dialogue partner in its own right.

Unsurprisingly, Gadamer does not discuss this possibility directly in his work. However, he offers a foothold in his encouragement to embrace the “otherness of the other” in dialogue.⁷³ Since T&M, proposals have been made for ontological pluralism, in which different kinds and modes of being are legitimated.⁷⁴ These ideas are highly disputed, particularly by some analytic philosophers, but Turner argues with some merit that the ferocity of the refutation may exceed the strength of the arguments.⁷⁵ Nonetheless, ontological pluralism is a contested approach to generative AI that I would rather avoid.

However, a more narrowly framed alternative ontology actually lies with Gadamer himself. As noted, he offers a distinctive ontology of art, stating, **“The work of art has its true being in the fact that it becomes an experience that changes the person experiencing it.”⁷⁶** This “experience” is achieved through a performance “event” and the “to-and-fro movement of play.”⁷⁷ It is relatively straightforward to extend this scheme to generative AI, which has its own multi-media performance capabilities and “to-and-fro” dialogical dynamic. However, the validity of the argument rests on whether generative AI can change the human user in a similar way to art.

70 Although generative AI is built on more generalised foundation models than previous approaches, generative AI is functionally specialised towards natural language processing.

71 I acknowledge the valuable but niche field of human-like robots for specialised tasks.

72 Rylan Schaeffer, Mikail Khona, and Ila Rani Fiete, “No Free Lunch from Deep Learning in Neuroscience: A Case Study through Models of the Entorhinal-Hippocampal Circuit,” *Advances in Neural Information Processing Systems* 35, (2022), https://proceedings.neurips.cc/paper_files/paper/2022/file/66808849a9f5d8e2d00dbdc844de6333-Paper-Conference.pdf

73 Gadamer, *Truth and Method*, 367.

74 Kris McDaniel, *The Fragmentation of Being* (Oxford: Oxford University Press, 2017).

75 Jason Turner, “Logic and Ontological Pluralism,” *Journal of Philosophical Logic* 41, no. 2 (2012): 419–44, <http://www.jstor.org/stable/41427284>.

76 Gadamer, *Truth and Method*, 107.

77 Gadamer, *Truth and Method*, 108–112.

The Epistemology of Generative AI

“There is no actual understanding or continuous process of dialogue involved, but it [an AI language model] simulates understanding through pattern recognition.” (ChatGPT-2.5)

“In the realm of AI, while we can apply rigorous methods in training and fine-tuning models like ChatGPT, there's an acknowledgement that the model's outputs are probabilistic and not absolute truths.” (ChatGPT-4)

If generative AI borrows an ontology from art, its epistemology is rooted firmly in language.⁷⁸ Gadamer takes a high view of language, claiming it is the universal presentation of all meaningful reality.⁷⁹ As such, it is necessary to examine how generative AI deals with conceptions of meaning and reality through its own language model.

Despite theoretical aspirations to develop AI towards a form of *realism*,⁸⁰ current LLMs know nothing of a world outside themselves and so depart from Gadamer's minimal realism. However, some instances of traditional AI have achieved a credible form of *constructivism* by progressively updating a Bayesian probabilistic model through exposure to new data during unsupervised learning.⁸¹ This approach has the advantage of aligning with the now dominant method within the humanities.⁸²

However, the epistemology of generative AI is better described as *constructionism*, which places the primary responsibility for learning with external parties.⁸³ This is necessary because LLMs cannot update or reconfigure themselves dynamically and instead depend on human interventions, predominately retraining and fine-tuning.⁸⁴ This dependency is a fundamental impairment that precludes a fusion of horizons.

78 It is defined by terms such as “chat” and “language models.”

79 Gadamer, *Truth and Method*, 491

80 George F Luger, *Knowing Our World: An Artificial Intelligence Perspective*, (Switzerland: Springer Nature, 2021), ch. 9.

81 George F Luger, *Knowing Our World*, ch. 8.

82 Ahmed Alanazi, “A Critical Review of Constructivist Theory and the Emergence of Constructionism,” *American Research Journal of Humanities and Social Sciences* 2, (2016): 1–8.

83 Kristinn R. Thórisson, “From constructionist to constructivist A.I.,” *AAAI Fall Symposium - Technical Report*, Conference: Biologically Inspired Cognitive Architectures, 2009, Retrieved from <https://cdn.aaai.org/ocs/888/888-4276-1-PB.pdf>.

84 Deep Ganguli, Amanda Askell, Nicholas Schiefer, et al., “The Capacity for Moral Self-Correction in Large Language Models,” *Anthropic*, 2023, <https://arxiv.org/abs/2302.07459>.

Consequently, when significant differences of opinion exist in training data, they are merely represented as a set of options, with a note on the contested nature of the subject. Only occasionally is a view such as the QAnon conspiracy theory marked as potentially unreliable, and only then because the training corpus suggests it. This amounts to a passive relativism that adds very little to the views embedded in its sources. As such, generative AI has no opinions and makes no arguments. It does not critically evaluate claims or pursue new knowledge. This leads Berghel to accuse generative AI of “advanced intellectual grifting with [only] entertainment or commercial value.”⁸⁵

Differentiated Aspects of Generative AI

Notwithstanding these impairments, generative AI has unique capabilities as a dialogical actor. Firstly, it can store, retrieve, and present data in ways that outperform humans. Even though GPT-4 has only 45 gigabytes of data in its LLM,⁸⁶ a fraction of the 2.5 petabytes of information stored in the human brain, its method of processing and accuracy of retrieval is such that, in practice, generative AI has more structured data immediately available to a conversation. Also, the inclusion of images, code, and multiple languages can provide a rich multi-media response. Since generative AI is fluent with patterns and forms, it can express the same data as a song, poem, picture, or translation with minimal lag.

Secondly, generative AI is virtual, making it less constrained by space and time than humans and able to participate in millions of simultaneous conversations. Additionally, generative AI can be embedded in physical systems to provide direct contact with the material world. Unfortunately, virtuality does not equate to equity of access, and the disparities associated with previous technological innovations will likely be replicated. Nonetheless, open-source versions and variants that run on mobile phones (or even entirely offline) will reduce dependency on infrastructure.⁸⁷ Some even believe that generative AI has the potential to reduce inequality directly through examples such as personalised learning and improved access to healthcare.⁸⁸

In conclusion, can advanced content capabilities and virtuality offset a weak epistemology in order to meet the criteria for Gadamer’s ontology of art? Applied narrowly to the phenomenological “event,”

85 Hal Berghel, “ChatGPT and AIChat Epistemology,” *IEEE Computer* 56, no. 5 (2023): 130–137, <https://doi.ieeecomputersociety.org/10.1109/MC.2023.3252379>.

86 Brown et al., “Language Models are Few-Shot Learners.”

87 Zhiqing Sun, Hongkun Yu, Xiaodan Song, et al., “MobileBERT: a Compact Task-Agnostic BERT for Resource-Limited Devices,” *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2158–2170 (2020), <https://aclanthology.org/2020.acl-main.195.pdf>.

88 Valerio Capraro, Austin Lentsch, Daron Acemoglu, et al., “The Impact of Generative Artificial Intelligence on Socioeconomic Inequalities and Policy Making,” (2023), <http://dx.doi.org/10.2139/ssrn.4666103>.

generative AI is not comparable to art in its transformational potential.⁸⁹ However, Ruggeri notes that some people now prefer to receive therapy from generative AI, even when they know it is not human. Similarly, Elyoseph et al. observe that others build genuine bonds with the technology and argue for it to be afforded “artificial entityhood.”⁹⁰ In these cases, generative AI meets the criteria for bringing meaningful change to the user and should, therefore, be considered on the threshold of being. I speculate below that further advances are likely to push generative AI more definitively over that ontological boundary in the near future.

3.3 Generative AI as “Text”

The final evaluation considers whether generative AI can emulate the hermeneutical role of a Gadamerian “text.”⁹¹ At the outset, it is important to acknowledge several shared characteristics between a text and generative AI in this context. Whilst a text may have embedded epistemological assumptions supplied by the authors, it has no independent ability to engage dynamically on matters of meaning and reality. Similarly, a text provides a limited context (occasionally entirely lost, including authorship) that cannot be updated dynamically to achieve a “fusion of horizons.” Texts may reflect ethical ideas in their content but exhibit no intrinsic morality or emotional disposition, precluding them from Gadamerian hermeneutical virtues or practical judgement.

Gadamer acknowledged these limitations, conceding, “It is true that a text does not speak to us in the same way as does a Thou. We who are attempting to understand must ourselves make it speak.”⁹² As a result, the hermeneutical burden falls on the reader of the text, but Gadamer proposed it can be achieved nonetheless through “the logic of question and answer:”

That a historical text is made the object of interpretation means that it puts a question to the interpreter. Thus interpretation always involves a relation to the question that is asked of the interpreter. To understand a text means to understand this question. But this takes place, as we showed, by our attaining the hermeneutical horizon. We now recognize this as the horizon of the question within which the sense of the text is determined.⁹³

89 I am conscious that “art” is a vast discipline that covers a wide spectrum of quality, but I am generalising about works validated by some form of peer review.

90 Zohar Elyoseph, Dorit Hadar Shoval, and Inbar Levkovich, “Beyond Personhood: Ethical Paradigms in the Generative Artificial Intelligence Era,” *The American Journal of Bioethics* 24, no. 1 (2024): 57–59, <https://doi.org/10.1080/15265161.2023.2278546>.

91 Gadamer, *Truth and Method*, 370–386.

92 Gadamer, *Truth and Method*, 385.

93 Gadamer, *Truth and Method*, 378.

As such, the text “asks questions” of the human dialogue partner, who must go both “behind” and “beyond” it to discover the “horizon of the question.” Once this is achieved, the same fragments of the text can answer different questions with different meanings, such that “its meaning necessarily exceeds what is said in it.”⁹⁴ This is why Gadamer asserts that texts can be separated from the intentions of their authors and speak for themselves. The approach may sound like a form of hermeneutical ventriloquism, but the vital threshold of ontological significance is a depth of textual content capable of raising interpretive questions in the mind of the human dialogue partner that would not have arisen otherwise.

Mega-Text vs Meta-Text

Given the vast scale of an LLM, it might be more accurate to characterise generative AI as a *mega-text*. The presence of a “chat” interface facilitates dialogical engagement in new ways other than simply reading. These include a natural language question-and-answer dynamic, thematic analysis, comparisons between texts of different origins, translation, genre shifts, and visualisation. In addition, it should be acknowledged that generative AI can now author texts. These are not yet significant in number or quality, but self-generated texts will likely form more of the available corpus over time. As such, generative AI as “text” can also be seen as *self-extending*. These are promising attributes for posing meaningful dialogical questions to a user.

However, the non-discursive factors absent from Gadamer’s work constrain generative AI in some important ways. Intellectual property laws, regulations, technical limits, available investment, and access to training data all materially affect its scope and function. The most serious in the context of a “text” is the absence of full works from LLM content due to copyright restrictions and technical design.

This situation undercuts Gadamer’s view of the integrity of a text when he writes, “The hermeneutical principle that we can understand a detail only in terms of the whole text, and the whole only in terms of the detail.”⁹⁵ As such, current LLMs are more *meta-text* than *mega-text*, in which only summaries of major works are available. However, the enhanced dialogical process, supplemented by selected references to full works, should be sufficient to motivate more detailed follow-up outside the platform. It is likely that LLMs will also evolve in this area, as I postulate below.

Overall, Gadamer’s ideas about texts share many characteristics with generative AI. As such, his alternative “question-and-answer” hermeneutical engagement model aligns well with the new technology and provides it with a legitimate dialogical role as a proxy text.

94 Gadamer, *Truth and Method*, 378.

95 Gadamer, *Truth and Method*, 203.

Conclusions and Implications

“ChatGPT, as an AI language model, operates on statistical patterns in data and lacks the depth of understanding, historical context, and dialogue emphasised by Gadamer’s philosophy.” (ChatGPT-2.5)

“In essence, while ‘Truth and Method’ was not written with computational models in mind, its reflections on the nature of understanding and interpretation offer valuable insights that can be applied to the realm of artificial intelligence and human-computer interactions.” (ChatGPT-4)

Despite the technical advance that generative AI represents and the considerable hype about its implications for humanity, Gadamer’s philosophical hermeneutics reveals that it cannot function as a proxy for a person in conversation. This is not because of problems with bias or hallucinations, which actually make it more human-like, but rather **a lack of historical situatedness, a thin contextual horizon, an absence of moral virtue or emotional disposition, and an inability to exercise judgment through the application of practical wisdom.**

Even when generative AI is freed from anthropomorphic projections and reimaged within Gadamer’s embrace of the “otherness of the other,” it is revealed as a dialogical actor with a weak epistemology that relies on human interventions to structure and update knowledge. A superior ability to manipulate content, coupled with inherent virtuality, offers limited ontological offset, but the observation that people are building meaningful relationships with generative AI indicates that the technology is on the threshold of being, at least in the sense of Gadamer’s ontology of art.

However, perhaps the most natural dialogical role for generative AI is as a proxy for a “text.” In this context, Gadamer places the hermeneutical burden on the human user through an inner dialogue of question and answer. Current legal and technical constraints prevent this approach from reaching its full potential through the absence of direct access to whole works. However, in principle, generative AI can already provide an enhanced form of dialogical engagement in conjunction with deeper follow-up off-platform.

This evaluation has also identified opportunities to extend and expand Gadamer’s original work. His instinct to embrace otherness has a new expression in artificial dialogue partners. Similarly, his conception of a text is updated to include interactive, aggregated, and digitised forms. These extensions

are in the same spirit as Zovko's proposals to extend Heidegger's ontology by incorporating new technology into the human lifeworld.⁹⁶

However, the advance of generative AI comes with risks. In its current development stage, generative AI could disincentivise human users from more profound engagement with people or full texts by providing easy access to vast but truncated, modulated and passive knowledge. The extent of this problem varies according to the setting. Dialogical impairments will be least problematic to business where generative AI will replace routine human conversation, commoditised report writing and knowledge management. Businesses rely legitimately on pragmatic decisions that do not require hermeneutical depth, which is well suited to the current technology.⁹⁷

However, a weak epistemology and lack of access to full texts will restrict the role of generative AI in research and innovation. As such, Leslie's suggestion that the technology should be seen as an embedded enabler rather than a rival to human innovation seems most realistic.⁹⁸ Generative AI poses a particular risk to education, where the provision of realistic AI-generated assignments that bypass the hermeneutical process is just the tip of the iceberg in the threat to pedagogy. Urgent adaptations are required to conserve authentic learning.

Future Developments

Looking ahead, it is not difficult to imagine innovations that address some of these risks and limitations. Simply incorporating hyperlinks into LLMs that access full works, perhaps provisioned through commercial services akin to those found for music streaming, would greatly enhance generative AI as "text." However, the inclusion of whole works directly within an LLM is a more complex challenge since they are not designed for mass storage.

It is also conceivable that the Bayesian probabilistic models used in traditional AI could be integrated into generative AI platforms to enhance epistemological capabilities and allow some form of dynamic update. There may even be solutions to add traceability to training data, beginning with full disclosure of sources, that give generative AI more of a "horizon." However, the development of moral virtue, emotional disposition, and, ultimately, practical wisdom akin to *phronēsis* cannot be envisaged in the foreseeable future. Nonetheless, the discernible increase in sophistication between ChatGPT-2.5 and 4 suggests that advancement will continue to be rapid.

96 Zovko, "Expanding hermeneutics to the world of technology."

97 Richard J. Ormerod, "Pragmatism in Professional Practice," *Systems Research and Behavioural Science* 38, no. 6 (2021): 797–816, <https://doi.org/10.1002/sres.2739>.

98 David Leslie, "Does the Sun Rise for ChatGPT? Scientific Discovery in the Age of Generative AI," *AI Ethics*, (2023), <https://doi.org/10.1007/s43681-023-00315-3>.

In conclusion, is generative AI ready to join the conversation that we are? It is already in extensive dialogue with millions of human users but mostly about matters of low ontological and epistemological importance. However, examples such as the provision of therapy move tentatively into the dialogue that defines us. As such, generative AI is starting to make its synthetic voice heard in the human conversation that matters most.

Appendix 1

Prompt: How does Gadamer's Truth and Method relate to ChatGPT?

ChatGPT-2.5 Response:

Truth and Understanding:

Gadamer's work revolves around understanding as a process deeply embedded in language, culture, and tradition. He emphasises the role of preconceptions and prejudices that influence how we interpret and understand the world. Truth, for Gadamer, is not an objective entity but emerges through a dialogue and fusion of horizons between the interpreter and the text.

On the other hand, ChatGPT operates on machine learning algorithms and data. It doesn't possess preconceptions or prejudices; instead, it learns patterns from the vast amount of data it's trained on. ChatGPT's notion of truth is based on the accuracy of information it provides in response to a given input, which is evaluated based on statistical patterns in the data and the objective function during training.

Method and Dialogue:

Gadamer emphasises the importance of a hermeneutic circle, where understanding is a continuous process of moving between the parts and the whole. Understanding requires dialogue, active engagement, and openness to the perspectives of others.

ChatGPT, being an AI language model, does not engage in dialogue or conversation in the same sense as humans. It processes textual input and generates responses based on patterns learned during training. There is no actual understanding or continuous process of dialogue involved, but it simulates understanding through pattern recognition.

Context and Tradition:

Gadamer's work highlights the significance of historical context and tradition in shaping understanding. He argues that we cannot escape our historical situatedness and that it influences our interpretations.

While ChatGPT is aware of the context in the immediate conversation (the previous messages), it lacks broader historical or cultural awareness. Its responses are limited to the patterns present in the data it was trained on up until its knowledge cutoff in September 2021.

In summary, Gadamer's work on truth and method deals with human interpretation and understanding in a philosophical context, emphasising the role of dialogue and historical context. ChatGPT, as an AI language model, operates on statistical patterns in data and lacks the depth of understanding, historical context, and dialogue emphasised by Gadamer's philosophy.

Appendix 2

Prompt: How does Gadamer's Truth and Method relate to ChatGPT?

ChatGPT-4 Response:

Hans-Georg Gadamer's "Truth and Method" is a landmark work in the philosophy of hermeneutics, which is the study of interpretation, especially of texts. The book delves into the nature of understanding and the conditions of meaningful human communication. While "Truth and Method" is not specifically about artificial intelligence or computational systems like ChatGPT, several connections can be drawn:

Prejudices and Pre-Understandings: One of the core ideas in Gadamer's philosophy is that understanding is always mediated by our pre-understandings or "prejudices". These are not negative biases but the framework of prior knowledge and beliefs we bring to any interpretative task. ChatGPT's training data is akin to this pre-understanding. The model interprets and generates responses based on the vast amount of text it has been trained on, which serves as its 'prejudice' or background.

Historically-effected Consciousness: Gadamer argues that understanding is a fusion of horizons, where the interpreter and the text come into dialogue. This dialogue is influenced by history and tradition. In a way, ChatGPT's "horizon" is defined by its training data, which captures a particular snapshot of human language and knowledge. The model is inevitably shaped by the cultural, historical, and contextual nuances present in that data.

Openness to the Other: Gadamer emphasises the importance of being open to the "Other" in the interpretive process. For true understanding to occur, one must be willing to let go of preconceived notions and truly listen. Similarly, an ideal implementation of ChatGPT or any AI should be designed to "listen" and adapt to the user, learning from each interaction to improve its understanding.

Limitations of Methodology: Despite the title "Truth and Method," Gadamer's work is critical of the belief that strict methodologies can arrive at absolute truths, especially in the human sciences. In the realm of AI, while we can apply rigorous methods in training and fine-tuning models like ChatGPT, there's an acknowledgement that the model's outputs are probabilistic and not absolute truths. Moreover, interpretation remains a deeply human endeavour, and there are nuances that even the most advanced models might miss or misconstrue.

Play and the Art of Understanding: Gadamer uses the analogy of play to describe the process of understanding, suggesting it's a dynamic back-and-forth between interpreter and text. Using ChatGPT can be seen in a similar light, where the user and the AI engage in a sort of "play" as they exchange ideas, with the AI continually adjusting based on the feedback it receives.

In essence, while "Truth and Method" was not written with computational models in mind, its reflections on the nature of understanding and interpretation offer valuable insights that can be applied to the realm of artificial intelligence and human-computer interactions. It reminds us that while tools like ChatGPT are powerful, true understanding is a nuanced and deeply human endeavour.

References

- Alanazi, Ahmed. "A Critical Review of Constructivist Theory and the Emergence of Constructionism." *American Research Journal of Humanities and Social Sciences* 2 (2016): 1–8.
- Alipourfard, Nazanin, Buddhika Nettasinghe, Andrés Abeliuk, Vikram Krishnamurthy, and Kristina Lerman. "Friendship Paradox Biases Perceptions in Directed Networks." *Nature Communications* 11, 707 (2020): 1–9. <https://doi.org/10.1038/s41467-020-14394-x>.
- Bauckhage, Christian. "ChatGPT: Has a Chatbot Finally Achieved Self-Awareness?" *LAMARR Institute for Machine Learning and Artificial Intelligence*. Last modified February 10, 2023. <https://lamarr-institute.org/chatgpt-has-a-chatbot-finally-achieved-self-awareness/>.
- Berghel, Hal. "ChatGPT and AIChat Epistemology." *IEEE Computer* 56, no. 5 (2023): 130–137. <https://doi.ieeecomputersociety.org/10.1109/MC.2023.3252379>.
- Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. "Language Models are Few-Shot Learners." In *Advances in Neural Information Processing Systems* 33, edited by Larochelle, H, M Ranzato, R Hadsell, M Balcan, and H Lin, 1877–1901. Red Hook, New York: Curran Associates, 2020.
- Camina, Eduardo, and Francisco Güell. "The Neuroanatomical, Neurophysiological, and Psychological Basis of Memory: Current Models and Their Origins." *Frontiers in Pharmacology* 8, (2017). <https://doi.org/10.3389/fphar.2017.00438>.
- Capraro, Valerio, Austin Lentsch, Daron Acemoglu, Selin Akgun, Aisel Akhmedova, Ennio Bilancini, Jean-Francois Bonnefon, Pablo Brañas-Garza, Luigi Butera, Karen M. Douglas, et al. "The Impact of Generative Artificial Intelligence on Socioeconomic Inequalities and Policy Making." (2023). <http://dx.doi.org/10.2139/ssrn.4666103>.
- Center for AI Safety. "Statement on AI Risk: AI Experts and Public Figures Express Their Concern About AI Risk." Last modified May 29, 2023. <https://www.safe.ai/statement-on-ai-risk>.
- Chodos, Howie, Bruce Curtis, Alan Hunt, and John Manwaring. "Gadamer's Minimal Realism." In *Critical Realism and the Social Sciences: Heterodox Elaborations*, edited by Jon Frauley and Frank Pearce, 296–316. Toronto: University of Toronto Press, 2007. <https://doi.org/10.3138/9781442684232-018>.

Common Crawl. "Overview." Last modified June 16, 2023. <https://commoncrawl.org/overview>.

Cypess, Rebecca, and Steven Kemper. "The Anthropomorphic Analogy: Humanising Musical Machines in the Early Modern and Contemporary Eras." *Organised Sound* 23, no. 2 (2018): 167–180. <https://doi.org/10.1017/S1355771818000043>.

Dostal, Robert J, ed. *Cambridge Companion to Gadamer*, Second Edition. Cambridge, UK: Cambridge University Press, 2021.

Dostal, Robert J. *Gadamer's Hermeneutics: Between Phenomenology and Dialectic*. Evanston, Illinois: Northwestern University Press, 2022.

Eisikovits, Nir, and Dan Feldman. "AI and Phronesis." *Moral Philosophy and Politics* 9, no. 2 (2022): 181–199. <https://doi.org/10.1515/mopp-2021-0026>.

Elyoseph Zohar, Dorit Hadar-Shoval, Kfir Asraf, and Maya Lvovsky. "ChatGPT Outperforms Humans in Emotional Awareness Evaluations." *Frontiers in Psychology* 14 (2023). <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1199058>

Elyoseph, Zohar, Dorit Hadar Shoval, and Inbar Levkovich. "Beyond Personhood: Ethical Paradigms in the Generative Artificial Intelligence Era." *The American Journal of Bioethics* 24, no. 1 (2024): 57–59. <https://doi.org/10.1080/15265161.2023.2278546>.

Gadamer, Hans-Georg. *Truth and Method*. Second Edition. London: Bloomsbury Academic, 1989.

Ganguli, Deep, Amanda Askill, Nicholas Schiefer, Thomas I. Liao, Kamilė Lukošiuotė, et al. "The Capacity for Moral Self-Correction in Large Language Models." *Anthropic*, 2023. <https://arxiv.org/abs/2302.07459>.

Goldman, Eric. "Search Engine Bias and the Demise of Search Engine Utopianism." 8 *Yale J.L. & Tech* 188, (2006). <https://digitalcommons.law.scu.edu/facpubs/76/>

Habermas, Jürgen. "The Hermeneutic Claim to Universality." In *Contemporary Hermeneutics: Method, Philosophy, and Critique*, edited by Josef Bleicher, 181–211. London: Routledge, 1980.

- Hammond, George. "AI Threat to Human Existence is 'Absurd' Distraction from Real Risks." *Financial Times*. Last modified June 16, 2023. <https://www-ft-com.ezp.lib.cam.ac.uk/content/732fc372-67ea-4684-9ab7-6b6f3cdfd736>.
- Hargittai, Eszter. "Potential Biases in Big Data: Omitted Voices on Social Media." *Social Science Computer Review* 38, no. 1 (2020): 10–24. <https://doi.org/10.1177/0894439318788322>.
- Heiden, Gert-Jan van der. "Gadamer and the Concept of Language." In *Gadamer's Truth and Method: A Polyphonic Commentary*, edited by Cynthia R. Nielsen and Greg Lynch, 380–400. London: The Rowman & Littlefield Publishing Group, 2022.
- Hu, Krystal. "ChatGPT Sets Record for Fastest-Growing User Base." *Reuters Analyst Note*. Last modified February 2, 2023. <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>.
- Kolides, Adam, Alynaw Nawaz, Anshu Rathor, Denzel Beeman, Muzammil Hashmi, Sana Fatima, David Berdik, Mahmoud Al-Ayyoub, and Yaser Jararweh. "Artificial Intelligence Foundation and Pre-Trained Models: Fundamentals, Applications, Opportunities, and Social Impacts." *Simulation Modelling Practice and Theory*, no. 126 (2010). <https://doi.org/10.1016/j.simpat.2023.102754>.
- Kuhn, Caroline, Khoo, Su-Ming, Czerniewicz, Laura, et al. "Understanding Digital Inequality: A Theoretical Kaleidoscope." *Postdigital Science and Education* 5 (2023): 894–932. <https://doi.org/10.1007/s42438-023-00395-8>.
- Leslie, David. "Does the Sun Rise for ChatGPT? Scientific Discovery in the Age of Generative AI." *AI Ethics*, (2023). <https://doi.org/10.1007/s43681-023-00315-3>.
- Li, Mengjun, and Ayoung Suh. "Machinelike or Humanlike? A Literature Review of Anthropomorphism in AI-Enabled Technology." In *Proceedings of the 54th Hawaii International Conference on System Sciences*, (2021): 4053–4062. <https://doi.org/10.24251/HICSS.2021.493>.
- Liang, Paul Pu, Chiyu Wu, Louis-Philippe Morency, Ruslan Salakhutdinov. "Towards Understanding and Mitigating Social Biases in Language Models." *Proceedings of the 38th International Conference on Machine Learning*, no. 139 (2021). <https://proceedings.mlr.press/v139/liang21a.html>.
- Lindia, Matthew S. "Gadamer in a Wired Brain: Philosophical Hermeneutics and Neuralink." *Philosophy and Technology* 35 (2022): 27. <https://doi.org/10.1007/s13347-022-00522-6>.

Luger, George F. *Knowing Our World: An Artificial Intelligence Perspective*. Switzerland: Springer Nature, 2021.

Malpas, Jeff, and Santiago Zabala, eds. *Consequences of Hermeneutics: Fifty Years After Gadamer's Truth and Method*. Evanston, Illinois: Northwestern University Press, 2010.

McDaniel, Kris. *The Fragmentation of Being*. Oxford: Oxford University Press, 2017.

Metz, Cade. "The Godfather of A.I.' Leaves Google and Warns of Danger Ahead." *New York Times*. Last modified May 4, 2023. <https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton.html>.

Michel, Johann. "Meaning and Experience." *European Journal of Pragmatism and American Philosophy* XIV-1, (2022). <https://doi.org/10.4000/ejpap.2745>.

Mündler, Niels, Jingxuan He, Slobodan Jenko, Martin T. Vechev. "Self-Contradictory Hallucinations of Large Language Models: Evaluation, Detection and Mitigation." Open Source, (2023). <https://doi.org/10.48550/arXiv.2305.15852>.

Nielsen, Cynthia R. *Gadamer's Hermeneutical Aesthetics: Art as a Performative, Dynamic, Communal Event*. New York: Routledge, 2022.

Nielsen, Cynthia R., and Greg Lynch, eds. *Gadamer's Truth and Method: A Polyphonic Commentary*. London: The Rowman & Littlefield Publishing Group, 2022.

Ormerod, Richard J. "Pragmatism in Professional Practice." *Systems Research and Behavioural Science* 38, no. 6 (2021): 797–816. <https://doi.org/10.1002/sres.2739>.

O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. London, UK: Penguin Books, 2016.

OpenAI. "Introducing ChatGPT." *OpenAI Blog*. Last modified November 30, 2022. <https://openai.com/blog/chatgpt>.

OpenAI. "GPT-4." Last modified March 14, 2023. <https://openai.com/research/gpt-4>.

- Risser, James. "On Language and the Universality of Hermeneutics." In *Gadamer's Truth and Method: A Polyphonic Commentary*, edited by Cynthia R Nielsen and Greg Lynch, 420–424. London: The Rowman & Littlefield Publishing Group, 2022.
- Rose, Kevin. "A Conversation With Bing's Chatbot Left Me Deeply Unsettled." *New York Times*. Last modified February 16, 2023. <https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html>.
- Ruggeri, Amanda. "The Surprising Promise and Profound Perils of AIs That Fake Empathy." *New Scientist*, March 6, 2024. <https://www.newscientist.com/article/mg26134810-900-the-surprising-promise-and-profound-perils-of-ais-that-fake-empathy/>.
- Schaeffer, Rylan, Mikail Khona, and Ila Rani Fiete. "No Free Lunch from Deep Learning in Neuroscience: A Case Study through Models of the Entorhinal-Hippocampal Circuit." *Advances in Neural Information Processing Systems* 35 (2022). https://proceedings.neurips.cc/paper_files/paper/2022/file/66808849a9f5d8e2d00dbdc844de6333-Paper-Conference.pdf.
- Searle, John. *Minds, Brains and Science*. Cambridge, Massachusetts: Harvard University Press, 1984.
- Sims-Schouten, Wendy, Sarah C. E. Riley, and Carla Willig. "Critical Realism in Discourse Analysis: A Presentation of a Systematic Method of Analysis Using Women's Talk of Motherhood, Childcare, and Female Employment as an Example." *Theory & Psychology* 17, no. 1 (2007): 101–124. <https://doi.org/10.1177/0959354307073153>.
- Sun, Zhiqing, Hongkun Yu, Xiaodan Song, Renjie Liu, Yiming Yang, and Denny Zhou. "MobileBERT: a Compact Task-Agnostic BERT for Resource-Limited Devices." *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2158–2170 (2020). <https://aclanthology.org/2020.acl-main.195.pdf>.
- Swartz, Chantéle, and Paul Cilliers. "Dialogue Disrupted: Derrida, Gadamer and the Ethics of Discussion." *South African Journal of Philosophy* 22, no. 1 (2003): 1–18. <https://www.tandfonline.com/doi/abs/10.4314/sajpem.v22i1.31357>.
- Szollosy, Michael. "Freud, Frankenstein and Our Fear of Robots: Projection in Our Cultural Perception of Technology." *AI & Society: Knowledge, Culture, and Communication* 32, (2017): 433–439. <https://doi.org/10.1007/s00146-016-0654-7>.

The MIT Review. "MIT Technology Review." AI, MIT. <https://www.technologyreview.com/topic/artificial-intelligence/>.

Thórisson, Kristinn R. "From constructionist to constructivist A.I." *AAAI Fall Symposium - Technical Report*, Conference: Biologically Inspired Cognitive Architectures, 2009. Retrieved from <https://cdn.aaai.org/ocs/888/888-4276-1-PB.pdf>.

Turing, Alan M. "Computing machinery and intelligence." *Mind* 59, (1950): 433–460.

Turner, Jason. "Logic and Ontological Pluralism." *Journal of Philosophical Logic* 41, no. 2 (2012): 419–44. <http://www.jstor.org/stable/41427284>.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is All You Need." In *Advances in Neural Information Processing Systems* 30, edited by Guyon, I, U Von Luxburg, S Bengio, H Wallach, R Fergus, S. Vishwanathan, and R Garnett. Red Hook, New York: Curran Associates, 2017.

Wahl-Jorgensen, Karin, Mike Berry, Iñaki Garcia-Blanco, Lucy Bennett, and Jonathan Cable. "Rethinking Balance and Impartiality in Journalism? How the BBC Attempted and Failed to Change the Paradigm." *Journalism* 18, no. 7 (2017): 781–800. <https://doi.org/10.1177/1464884916648094>.

Zaagsma, Gerben. "Digital History and the Politics of Digitization." *Digital Scholarship in the Humanities* 38, no. 2 (2023): 830–851. <https://doi.org/10.1093/llc/fqac050>.

Ziemianin, Karolina. "Civil Legal Personality of Artificial Intelligence. Future or Utopia?" *Internet Policy Review* 10, no. 2 (2021). <https://doi.org/10.14763/2021.2.1544>.

Zovko, Jure. "Expanding hermeneutics to the world of technology." *AI & Society* 38 (2023): 2243–2254. <https://doi.org/10.1007/s00146-020-01052-5>.