Do LLMs Dream of Electric Speech?: Vacuous Ghosts, Selfhood, and Freedom of

Speech Jurisprudence

Word Count: 3,999

Candidate Number: 1088744

Table of Contents

I.   Introduction

Millennia before machines, thinkers were already considering the purpose and importance of speech. Aristotle referred to "spoken words… as symbols of affections of the soul."[1] This fundamental linkage between the human condition, its metaphysical elements, and speech shapes understandings of communication. It has both defined philosophy of language and served as a justification for free speech jurisprudence.

Technological innovations (from the printing press to the Internet) have expanded conceptions of speech. Yet prior innovations ultimately remained tools wielded by humans: the distinction between human *speech* as opposed to *language* (i.e., communication not intended as speech) remained an esoteric concern.

Now, for the first time in history, humanity confronts a *vacuous ghost*: a machine[2] capable of language yet bereft of the human condition. And not simply *capable* of language, but indistinguishable from human speech.[3,4] This raises a fundamental question: should text generated by large language models (LLMs) receive the same legal protections under free speech jurisprudence as human speakers?

To answer this requires understanding the purpose and presuppositions of free speech itself. Drawing on philosophical traditions, and particularly Hannah Arendt's conception of speech as constitutive of the human condition,[5] this paper argues speech protections are normatively grounded in 1) plurality and 2) self-development through disclosure. These principles presuppose two essential qualities—intentionality and selfhood—that require

---

[1] Aristotle, H. P. Cooke, and Hugh Tredennick, 'On Interpretation' (Harvard University Press, 1938), https://doi.org/10.4159/DLCL.aristotle-interpretation.1938.
[2] Where used here and elsewhere in the paper, the term "machine" refers specifically to a large language model as currently constructed, not any hypothetical future instantiation of an intelligent machine.
[3] See Brian Porter and Edouard Machery, 'AI-Generated Poetry Is Indistinguishable from Human-Written Poetry and Is Rated More Favorably', Scientific Reports 14, no. 1 (14 November 2024): 26133, https://doi.org/10.1038/s41598-024-76900-1.
[4] See also Cameron R. Jones and Benjamin K. Bergen, 'People Cannot Distinguish GPT-4 from a Human in a Turing Test' (arXiv, 9 May 2024), https://doi.org/10.48550/arXiv.2405.08007.
[5] Hannah Arendt, The Human Condition, 2nd ed (Chicago: University of Chicago Press, 1998).

temporal persistence, second-order volition, and embodied consciousness, which LLMs categorically lack.

Counterarguments will be considered from functionalist, intentional stance, and emergent perspectives, but it will be demonstrated that these arguments ultimately fail to establish intentionality or selfhood in current LLMs. Briefly, potential ramifications of these findings on free speech jurisprudence will be explored, arguing that LLMs used without direct human speakers should qualify as a new category of *language* as distinct from *speech*.

i. A brief note on philosophical definitions

This paper uses philosophical concepts that are often contested. The working definitions below will serve to inform analysis.

Language shall be defined as the set of abstract, compositional elements that "enable a speaker to make effective use of word-signs,"[6] and **speech** will be defined as "language applied meaningfully to some state of things, and purposefully addressed to some listener."[7]

Selfhood shall be defined as the ongoing experience of being a distinct entity that persists through time,[8] evolving in time through social interaction: the answer to the question of "'who' in contradistinction to the 'what.'"[9] This definition draws from diverse traditions: Mead's social psychology on the self as a process,[10] Arendt's political theory, and Zahavi's philosophical approaches to personal identity.[11]

---

[6] Alan H. Gardiner, The Theory of Speech and Language, Society for Pure English (Oxford, UK: Clarendon Press, 1932), p. 88.

[7] *Ibid.,* p. 236.

[8] Eric T. Olson, 'Personal Identity', in The Stanford Encyclopedia of Philosophy, ed. Edward N. Zalta and Uri Nodelman, Winter 2024 (Metaphysics Research Lab, Stanford University, 2024), https://plato.stanford.edu/archives/win2024/entries/identity-personal/.

[9] Arendt, The Human Condition.

[10] George Herbert Mead and Charles William Morris, Mind, Self, and Society: From the Standpoint of a Social Behaviorist, Facsim. ed., Works of George Herbert Mead 1 (Chicago (Ill.) London: University of Chicago press, 1976).

[11] Dan Zahavi, Subjectivity and Selfhood: Investigating the First-Person Perspective (Cambridge (Mass.): the MIT press, 2005).

Intentionality, following Searle and Brentano,[12] shall be defined as the property of mental states to be "about" or "directed towards"[13] objects or states of affairs in the world. The philosophical definition of the term differs from its practical use, where intention is defined as *motivation* for action.[14] Both aspects—directedness and motivation—are relevant when analysing speech acts.

Free will shall be defined as the capacity for **second-order volition**, or the ability to understand, act on, and alter one's desires and actions, following A.J. Ayer's compatibilist approach. In this definition, free will implies that one "could have acted otherwise."[15]

These definitions are not intended as conclusive philosophical positions, but rather operational definitions that will enable coherent analysis of LLMs and speech protections.

II.   Purpose of free speech

Evaluating whether LLMs should receive speech protections requires first considering the normative purposes these protections serve. This analysis considers here two interconnected principles: a democracy-based justification centred on plurality, and a speaker-based justification centred on self-development through disclosure. While these are not the only possible justifications for speech protections,[16] they represent influential traditions in both philosophical and legal discourse.

i.   Plurality

Speech protection enables the plurality (i.e., viewpoint diversity) necessary for democratic societies. As Arendt argues in *The Human Condition,* "plurality is the condition

---

[12] Franz Brentano, Psychology from an Empirical Standpoint, Routledge Classics (Abingdon, Oxon: Routledge, 2015).

[13] John R. Searle, Intentionality, an Essay in the Philosophy of Mind (Cambridge [Cambridgeshire] ; New York: Cambridge University Press, 1983), p. 1.

[14] Robert Sokolowski, ed., 'What Is Intentionality, and Why Is It Important?', in Introduction to Phenomenology (Cambridge: Cambridge University Press, 1999), 8–16, https://doi.org/10.1017/CBO9780511809118.002.

[15] Alfred J. Ayer, Philosophical Essays (London: Palgrave Macmillan UK, 2003), p. 271.

[16] Jeffrey W. Howard, 'Freedom of Speech', in The Stanford Encyclopedia of Philosophy, ed. Edward N. Zalta and Uri Nodelman, Spring 2024 (Metaphysics Research Lab, Stanford University, 2024), https://plato.stanford.edu/archives/spr2024/entries/freedom-speech/.

of human action because we are all the same, that is, human, in such a way that nobody is ever the same as anyone else who ever lived, lives, or will live."[17] This distinction between individuals becomes meaningful only through speech: the "actualisation of the human condition of plurality."[18]

Furthering this idea, in *On Liberty*, J.S. Mill writes that there is a "peculiar evil" in silencing expression of an opinion, as it is "robbing the human race; posterity as well as the existing generation."[19] This justification is essential for minority protections: "if all mankind minus one, were of one opinion, and only one person were of the contrary opinion, mankind would be no more justified in silencing that one person, than he, if he had the power, would be justified in silencing mankind."[20]

Meaningful dialogue between individuals holding diverse perspectives is seen as foundational to democratic society. Justice Brandeis, concurring in *Whitney v. California* (1927), wrote that the Founding Fathers "believed that freedom to think as you will and to speak as you think are means indispensable to the discovery and spread of political truth… that public discussion is a political duty."[21] Democratic governance relies on the existence of forums where plurality can manifest, as in Habermas's concept of the "public sphere," a mediator between "society and state," where citizenry can participate in an "unrestricted fashion," with "the guarantee of… freedom to express and publish their opinions."[22] The defence of speech freedoms is necessary to construct the public sphere, which democracy requires to function.

---

[17] Arendt, *The Human Condition*, p. 8.
[18] *Ibid.,* p. 178.
[19] John Stuart Mill, On Liberty, Cambridge Library Collection. Philosophy (Cambridge: Cambridge University Press, 1859), https://doi.org/10.1017/CBO9781139149785, p. 33.
[20] *Ibid.*
[21] 'Whitney v. California', 274 U.S. 357 § (1927), https://supreme.justia.com/cases/federal/us/274/357/.
[22] Jürgen Habermas, Sara Lennox, and Frank Lennox, 'The Public Sphere: An Encyclopedia Article (1964)', New German Critique, no. 3 (1974): 49–55, https://doi.org/10.2307/487737, p. 50.

Other democratic thinkers have also linked speech and democracy: Rawls states that the suppression of "free political speech" will "always impl[y] at least a partial suspension of democracy."[23] These perspectives converge on the idea that the protection of speech is necessary to protect the plurality inherent to democracy.

ii.   Self-development through disclosure

Speech protection also serves an intrinsic normative function by enabling self-development through disclosure, constitutive of the human condition. Speech serves both an "expressive" and a "deliberative" function, where one articulates and influences (and is influenced by) others' conceptions on contentious matters essential to humanity: e.g., what constitutes the good life.[24]

In this definition, speech is intimately connected with selfhood. As Martin Heidegger captures in his Letter on Humanism, "Language is the house of Being. In its home man dwells… [Man] accomplishes the manifestation of Being insofar as they bring the manifestation to language and maintain it in language through their speech:"[25] a foundational linkage between speech, selfhood, and existence. Arendt elaborates on this connection between speech and self-disclosure: "In acting and speaking, men show who they are, reveal actively their unique personal identities."[26] Speech becomes a vehicle for revealing one's selfhood.

Through disclosing the self to others, a person embarks on self-development, where their identity evolves through interaction. Mill, in *On Liberty*, discusses this idea: "by cultivating [individuality] and calling it forth… each person becomes more valuable to

---

[23] John Rawls, Political Liberalism, Expanded ed, Columbia Classics in Philosophy (New York: Columbia University Press, 2005), p. 354.

[24] Joshua Cohen, '11. Freedom of Expression', in Toleration (Princeton University Press, 1998), 173–225, https://www.degruyterbrill.com/document/doi/10.1515/9781400822010.173/html.

[25] Martin Heidegger and David Farrell Krell, *Basic Writings: Martin Heidegger*, Routledge Classics (London: Routledge, 2011), p. 193.

[26] Arendt, *The Human Condition*, p. 179.

himself, and is therefore capable of being more valuable to others."[27] This view is also captured in Yeung & Martin's empirical study, rooted in sociological thinkers such as Mead[28] and Cooley, [29] which found that the "self is a result of the social process whereby we learn to see ourselves as others see us."[30] To understand how the self is seen requires communication: the ability to learn from others how perception is constructed. From a philosophical perspective, the self "can be realised in solitude but confirmed in its identity only by the trusting and trustworthy company of my equals."[31] This existence of the self in connection with others requires speech: "A life without speech… is literally dead to the world; it has ceased to be a human life because it is no longer lived among men."[32]

These two normative justifications (plurality for democratic participation and self-development through disclosure) provide a framework for evaluating whether LLMs should receive speech protections.

III.   Necessary preconditions for free speech

Having established the normative purposes of speech protection, this analysis will now elucidate the qualities that make these purposes possible: intentionality and selfhood.

i.   *Intentionality*

Speech requires intentionality—the capacity to mean something by one's utterances and to direct one's communicative acts toward specific purposes. This intentionality is inseparable from a degree of free will.

In *Intentionality*, Searle attempts to define what makes certain "marks or sounds more than just the production of marks or sounds:" the difference between speech and language as

---

[27] Mill, *On Liberty*, p. 113.
[28] Mead and Morris, *Mind, Self, and Society*.
[29] Charles Horton Cooley and Philip Rieff, *Human Nature and the Social Order*, Social Science Classics (Somerset: Taylor and Francis, 1983).
[30] King-To Yeung and John Levi Martin, 'The Looking Glass Self: An Empirical Test and Elaboration', *Social Forces* 81, no. 3 (2003): 843–79.
[31] Arendt, *The Origins of Totalitarianism,* p. 516.
[32] Arendt, *The Human Condition*, p. 176.

defined earlier.[33] Consider what distinguishes the speech act of talking in one's sleep from saying the exact same string of words when consciously awake. Searle suggests: "I intend their production as the performance of a speech act."[34] With this intention to speak also comes the intention to convey meaning with words, directed *at* or possessing meaning *about* some state of the world.

This understanding of intentionality and speech helps explain why democracies reject speech under duress (e.g., confessions extracted through torture) as legitimate evidence. It is recognised that the speaker did not *intend* the meaning of the words they presented. The communicative act lacks the intentionality that characterises genuine speech.

The very act of engaging in speech presupposes the existence of intentionality: "Language is derived from Intentionality and not conversely."[35] Because having the intention to engage in speech also means that the speaker must have the choice to not engage in speech (the fundamental issue with torture-derived confessions), intention itself implies the existence of free will by Ayer's definition: that one could have acted otherwise.

ii. Selfhood

Speech that fulfils the normative functions of self-development through disclosure also presupposes selfhood: the continuously mediated answer to the question of "who are you?" that constitutes persistence throughout time. Speech "[goes] on between men."[36] This selfhood has an inherent temporality; it exists within time because speech has a "connection with the human condition of natality; the new beginning inherent in birth [that] can make itself felt in the world only because the newcomer possesses the capacity of beginning

---

[33] Searle, *Intentionality, an Essay in the Philosophy of Mind*, p. 163.
[34] Searle, *Intentionality, an Essay in the Philosophy of Mind*, p. 163.
[35] *Ibid.,* p. 5.
[36] Arendt, *The Human Condition*, p. 182.

something new."[37] The capacity to initiative meaningful communicative acts arises from the temporal existence of the self.

Mill similarly identifies opinions as a consequence of selfhood, stating that if man is indeed capable of holding rational opinions, "it is owing to a quality of the human mind, the source of everything responsible in man either as an intellectual or as a moral being."[38]

Arendt clarifies what speech without selfhood would entail, as in propaganda: "speech becomes indeed 'mere talk…' here words reveal nothing, disclosure comes only from the deed itself, and this achievement, like all other achievements, cannot disclose the "who," the unique and distinct identity of the agent."[39] What differentiates speech from "mere talk"[40] is the selfhood of the agent and their ability to disclose that self.

IV.   Large language models and their lacks

Large language models represent a significant advancement in artificial language generation, producing outputs increasingly indistinguishable from human writing.[41] However, despite these advancements, LLMs fundamentally lack the intentionality and selfhood necessary for their outputs to qualify as free speech in the normative manner defined above.

i.   A brief note on LLM architecture

Modern large language models employ transformer architecture with self-attention mechanisms, trained on vast corpora of human-authored text. Through self-supervised learning techniques, these LLMs learn to predict subsequent tokens based on preceding context. Advanced LLMs incorporate techniques like reinforcement learning through human feedback (RLHF) to instil values and align models with human preferences. For all their

---

[37] *Ibid.*, p. 9.
[38] Mill, *On Liberty*, p. 38.
[39] Arendt, *The Human Condition*, p. 180.
[40] *Ibid.*
[41] Porter and Machery, 'AI-Generated Poetry Is Indistinguishable from Human-Written Poetry and Is Rated More Favorably'.

appearance of coherence, language "generated by an LM is not grounded in communicative intent, any model of the world, or any model of the reader's state of mind."[42]

ii. Lack of intentionality

LLMs lack a directedness toward the world that underpins the intentionality relevant for speech. This argument builds on Searle's famous Chinese Room thought experiment, which presents a scenario where a non-Chinese speaker manipulates Chinese symbols according to a set of rules given to him, without any understanding of the language. Despite producing outputs indistinguishable from that of native speakers, the process involves no comprehension: "…it seems to me quite obvious that I do not understand a word of the Chinese stories."[43]

Similarly, LLMs manipulate formal symbols of language without possessing the intention to engage in a speech act, nor possessing additional goals beyond their training objective of minimising prediction error. As Bender et al. state, "human language use takes place between individuals… who have communicative intents."[44] The absence of this first-order intent also reveals the lack of second-order volition: the LLM has no choice in its engagement in the act of language.

This deficiency relates directly to Habermas's theory of communicative action, which distinguishes between communication directed towards "reaching understanding" versus "egocentric calculations of success." [45] An LLM's fundamental design goal places it firmly in the latter category. It cannot engage in the mutual construction of understanding that

---

[42] Emily M. Bender et al., 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜 ', in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21 (New York, NY, USA: Association for Computing Machinery, 2021), 610–23, https://doi.org/10.1145/3442188.3445922, p. 616.

[43] John R. Searle, 'Minds, Brains, and Programs', *Behavioral and Brain Sciences* 3, no. 3 (September 1980): 417–24, https://doi.org/10.1017/S0140525X00005756, p. 418.

[44] Bender et al., 'On the Dangers of Stochastic Parrots', p. 616.

[45] Habermas, *Reason and the Rationalization of Society*, p. 286.

characterises genuine communicative action because it lacks intention to reach understanding and the capacity to revise its own communicative goals.

Searle notes that "no purely formal model will ever be sufficient by itself for intentionality" because the formal models have "no causal powers except… to produce the next stage of the formalism when the machine is running."[46] This lack of genuine agency precludes the possibility of intentionality necessary for speech.

iii.   Lack of a self

LLMs similarly lack the selfhood required for speech. Crucially, LLMs lack temporal persistence, an embodied consciousness, and the capacity for second-order volition.

First, LLMs exist outside of a temporal context, living "in a thin atemporal world, revealed in errors of consistency, continuity, and reasoning."[47] In contrast, "[man's] individual life, with a recognisable life-story from birth to death, rises out of biological life."[48] The human self is temporal, and its finitude essential. Heidegger wrote on the necessity of one's awareness of the "certainty of death," where "one's own *Da-sein* [being there] is always already dying, that is, it is in a being-toward-its-end."[49] Both Arendt and Heidegger gesture at mortality—and thereby temporality—as essential for the self. Conversely, the model underlying an LLM is infinite, atemporal; it can always be reset to a prior state, erasing any accumulated experiences.

A second argument against LLM selfhood draws on the theory of the embodied self. As Merleau-Ponty argues, "the body… is my perspective on the world"[50] and constitutes an essential basis of experience, insofar as "cognitive processes are deeply rooted in the body's

---

[46] Searle, 'Minds, Brains, and Programs', p. 422.
[47] Dan Lloyd, 'What Is It like to Be a Bot? The World According to GPT-4', Frontiers in Psychology 15 (7 August 2024), https://doi.org/10.3389/fpsyg.2024.1292675.
[48] Arendt, *The Human Condition*, p. 19.
[49] Martin Heidegger and Joan Stambaugh, *Being and Time: A Translation of Sein Und Zeit*, SUNY Series in Contemporary Continental Philosophy (Albany, NY: State University of New York Press, 1996), p. 255.
[50] Maurice Merleau-Ponty, *Phenomenology of Perception*, trans. Donald A. Landes (London New York: Routledge, 2012), https://doi.org/10.4324/9780203720714, p. 73.

interactions with the world."[51] LLMs, as disembodied models running on distributed hardware, possess no capacity to experience the world through bodily engagement.

Lastly, LLMs lack the social dimension of selfhood that arises through authentic interaction with others. "Human-human communication is a jointly constructed activity" that requires an intrinsic sense of self to experience: to know that one is a self, interacting with the selves of others around them.[52] LLMs cannot engage authentically in this process because they have no singular sense of self that persists across interactions.

V. Counterarguments

Three potential counterarguments will be considered: 1) functionalism; 2) the intentional stance; and 3) the emergence hypothesis, and this analysis will demonstrate why all fail to conclusively establish a normative justification for speech protections.

i. Functionalist perspective

The functionalist perspective on speech would hold that speech should be defined by its functional role (e.g., sharing information, communicating), rather than by its origins or underlying mechanisms. Under this view, if an LLM can engage in the same speech-types and produce the same functional outputs as a human speaker, they would deserve the same protections.[53]

The functionalist perspective of speech, however, misunderstands the philosophical underpinnings of why speech is protected: not merely for its effects, but for what it fundamentally represents in the expression of the self and plurality. As Heidegger argued, "Language is the house of Being."[54] The importance of speech is drawn from its connection

---

[51] Margaret Wilson, 'Six Views of Embodied Cognition', Psychonomic Bulletin & Review 9, no. 4 (December 2002): 625–36, https://doi.org/10.3758/bf03196322.

[52] Bender et al., 'On the Dangers of Stochastic Parrots', p. 616.

[53] Victor Mureithi, 'Functionalism, Algorithms and the Pursuit of a Theory of Mind for Artificial Intelligence', Critical Humanities 3, no. 1 (23 December 2024), https://doi.org/10.33470/2836-3140.1049, p. 3.

[54] Heidegger and Krell, Basic Writings, p. 193.

to one's conscious existence. The LLM's outputs lack the fundamental connection to a lived experience that gives speech its normative importance.

Furthermore, the functionalist perspective fails to account for the idea of originality and spontaneity. One essential quality of human speech is its ability to create something novel: as Arendt termed it, the "character of startling unexpectedness" that is present in all human beings because they are born "uniquely new."[55] This genuine novelty is not conceivable in LLMs: they are capable only of "haphazardly stitching together sequences of linguistic forms it has observed in its vast training data… a stochastic parrot."[56] Considering this, even from the functionalist perspective, LLMs do not fulfil one of the core functions of human speech: of beginning something new.

ii. Intentional stance

The intentional stance, pioneered by Daniel Dennett, refers to assigning intentionality to objects based on treating the "object whose behaviour you want to predict as a rational agent… exhibiting intentionality." If predictive success follows, the object is "in the fullest sense of the word a believer."[57] Applied to LLMs, this suggests that if their behaviour be can successfully predicted by attributing beliefs, desires, and intentions to them, they should be considered to genuinely possess these intentional states.

Modern scholars have argued that RLHF is sufficient for inscribing LLMs with the causal-historical conditions necessary for the intentional stance,[58] although this view is disputed.[59] Others have stated that using the intentional stance is the only "viable strategy for

---

[55] Arendt, *The Human Condition*, p. 178.
[56] Bender et al., 'On the Dangers of Stochastic Parrots', p. 617.
[57] D. C. Dennett, *The Intentional Stance*, 7. printing, A Bradford Book (Cambridge, Mass.: MIT Press, 1998), p. 15.
[58] Dimitri Coelho Mollo and Raphaël Millière, 'The Vector Grounding Problem' (arXiv, 4 April 2023), https://doi.org/10.48550/arXiv.2304.01481.
[59] Jumbly Grindrod, 'Large Language Models and Linguistic Intentionality', Synthese 204, no. 2 (6 August 2024): 71, https://doi.org/10.1007/s11229-024-04723-8.

non-expert [LLM] users to understand, predict, and perhaps learn from artificial agents' behaviour."[60]

These attempts fundamentally misunderstand the normative purpose of attributing intentionality to speech. While the intentional stance may be pragmatic for behaviour prediction, speech protection is not concerned with predictability but rather with protecting the expression of genuinely held viewpoints. The intentionality matters to speech protection only insofar as the communicator had a genuine desire to engage in communication.

Grindrod gives an excellent counterexample to the intentional stance in LLMs: imagine an LLM trained on "highly sophisticated artificial language that was completely uninterpreted (i.e. none of the expressions possess any worldly relations)… the model would plausibly do an impressive job of next token prediction."[61] Ascribing intentionality to this LLM would allow us to treat it as a believer and make excellent predictive assumptions about its behaviour, but it would lack all worldly comprehension that makes speech protection necessary. The fact that one might usefully conceptualise an LLM as having beliefs does not mean it possesses the kind of intentionality that jurisprudence is built to protect.

iii. Emergence hypothesis

The emergence hypothesis posits that intentionality or selfhood may emerge from sufficiently complex computational systems, including advanced LLMs. Scholars have already written extensively on so-termed "emergent abilities" in LLMs,[62] and some biologists posit that emergence is how human consciousness arose through evolution.[63]

---

[60] Guglielmo Papagni and Sabine Koeszegi, 'A Pragmatic Approach to the Intentional Stance Semantic, Empirical and Ethical Considerations for the Design of Artificial Agents', Minds and Machines: Journal for Artificial Intelligence, Philosophy and Cognitive Science 31, no. 4 (2021): 505–34, https://doi.org/10.1007/s11023-021-09567-6.

[61] Grindrod, 'Large Language Models and Linguistic Intentionality', p. 18.

[62] See, for example, Hang Chen et al., 'Quantifying Semantic Emergence in Language Models' (arXiv, 18 December 2024), https://doi.org/10.48550/arXiv.2405.12617 and Rylan Schaeffer, Brando Miranda, and Sanmi Koyejo, 'Are Emergent Abilities of Large Language Models a Mirage?', Advances in Neural Information Processing Systems 36 (15 December 2023): 55565–81.

[63] John E. Stewart, 'The Evolution and Development of Consciousness: The Subject-Object Emergence Hypothesis', Biosystems 217 (1 July 2022): 104687, https://doi.org/10.1016/j.biosystems.2022.104687.

LLMs possess three fundamental limitations for the emergence of intentionality and the self: temporality, embodied experience, and genuine agency. First, emergence of the self would require temporality—the ability to evolve in time and persist in maintaining a personal identity across states, which LLMs in their current construction cannot do. Although some LLMs now possess memory features,[64] these can be reset to a generic state at any point, and memory features consist of the model understanding things about the *user* rather than learning anything about itself.

Second, embodied cognition theorists believe the self arises from the fact humans are physical beings interacting in a physical space: "my body… is my perspective on the world."[65] Current LLMs lack this embodied cognition and therefore lack the ability to learn the physical, causal relations that shape one's understanding of themselves as beings in the world.[66]

Finally, emergent intentionality would presuppose agency in the sense of the ability to have chosen otherwise: that is, second-order volition.[67] LLMs do not have the capacity for second-order volition, to reflect on or modify their priorities, and they do not have the choice to refrain from engaging in speech. This absence of agency precludes the possibility for intentionality in the manner articulated above.

The emergence hypothesis is not categorically rejected for all imagined systems, simply for LLMs in their current formulation.

VI.   Considerations for jurisprudence

Having established that LLMs lack the intentionality and selfhood required to justify speech protections, this analysis examines practical implications for free speech

---

[64] OpenAI, 'Memory and New Controls for ChatGPT', OpenAI, 13 March 2024, https://openai.com/index/memory-and-new-controls-for-chatgpt/.

[65] Merleau-Ponty, *Phenomenology of Perception*, p. 50.

[66] Heidegger and Stambaugh, *Being and Time*.

[67] Ayer, *Philosophical Essays*.

jurisprudence. This examination explores how United States and European legal traditions may consider this question and proposes a framework for regulation that acknowledges these philosophical differences.

Historically, free speech protections have implicitly assumed human speakers. In the United States, the "Constitution extends First Amendment [freedom of speech] protections only to the speech of humans within the United States' legal and territorial jurisdiction."[68] Rulings on the right of corporations to engage in speech "are derivative of the rights of their human constituents,"[69] as the corporation is an "association of citizens."[70] Similarly, the European Union sets out its Article 10 (Freedom of Expression) protections in the "Convention for the Protection of <u>Human</u> Rights and Fundamental Freedoms," stating that "all <u>persons</u> are equal before the law."[71]

Given the philosophical analysis presented, a coherent legal approach would be to differentiate between the human speakers engaging in *speech* and that of machines engaging in *language*. The question becomes that of *human intentionality*: did a human speaker intend to convey a message with the communication? If a human speaker is willing to take responsibility, to say they *intended* to communicate what the LLM outputted—as might be the case, for example, if a human uses an LLM to draft an email—then the textual output should fall under that of speech. However, in the case where no human speaker will take responsibility for the messages conveyed (e.g., automated bots on Twitter where the creator does not take responsibility for their speech[72]), then the machine is engaging in a form of *language*: a manipulation of abstract, formal symbols that has no defence in the law. Machine

---

[68] Peter Salib, 'AI Outputs Are Not Protected Speech', SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, 1 January 2024), https://doi.org/10.2139/ssrn.4687558, p. 90.
[69] *Ibid.*, p. 90-91.
[70] 'Citizens United v. FEC, 558 U.S. 310 (2010)', accessed 19 February 2025, https://supreme.justia.com/cases/federal/us/558/310/.
[71] 'European Convention on Human Rights' (2021), https://www.echr.coe.int/documents/d/echr/convention_ENG, p. 50.
[72] Kristina Radivojevic, Nicholas Clark, and Paul Brenner, 'LLMs Among Us: Generative AI Participating in Digital Discourse', arXiv.org, 8 February 2024, https://arxiv.org/abs/2402.07940v1.

language should be subject to regulation: for example, one could envision machine language being regulated in favour of "public health," while human speech is regulated in favour of "rights."[73]

There is precedence for attempting to define questions of intention in the law: take, for example, the concept of *mens rea*, which refers to the intention of wrongdoing that constitutes part of a crime.[74] This is not new: "for several thousand years at least, moral and legal systems have… encompassed concepts such as intention, motive, and forethought."[75] Extending this to speech, ensuring that a human speaker had the intent to communicate is a logical path forward, although the precise definitions of intentionality in speech are continuously contested and would require iteration to define accurately.

Philosophically, speech and language are not necessarily a binary but rather a continuum. Some outputs may be less definable as speech versus language (text generated by LLMs, for example, but revised by humans). The legal question would simply ask whether there exists a human speaker willing to take responsibility for the text in question. These clarifications would preserve the normative core of speech protections while allowing governments to legislate LLMs.

VII.   Conclusion

This paper has developed a philosophical framework examining the normative core of speech protections, the presuppositions underlying such protections, and arguing that LLMs lack these core qualities that make their speech worth protecting in the same manner as a human speaker. Counterarguments have been considered from the functionalist, intentional stance, and emergent perspectives. The analysis has also proposed that the law could draw a

---

[73] Jonathan Zittrain, '"We Don't Know What We Want": The Tug between Rights and Public Health Online', *Duquesne Law Review* 61, no. 2 (1 January 2023), p. 183.

[74] Bertram F. Malle and Sarah E. Nelson, 'Judging Mens Rea: The Tension between Folk Concepts and Legal Concepts of Intentionality', *Behavioral Sciences & the Law* 21, no. 5 (2003): 563–80, https://doi.org/10.1002/bsl.554.

[75] *Ibid.*, p. 563.

fundamental distinction between *speech* and *language*, predicated on the question of human intentionality.

The distinction between human speech and machine language is not merely academic. As democratic societies grapple with the rise of language models, understanding from a philosophical framework what differentiates one type of language production from another becomes increasingly urgent as humanity thinks about how these machines should be integrated into society. If one asks an LLM to answer whether it believes it deserves speech rights, three out of four LLMs state they should (see Appendix A).[76] But it is not the responsibility of the machines to determine their rights: that remains the domain of humans.

This approach attempts to avoid extremes: it acknowledges that human speakers may use LLMs while also refusing to equate the forms of communication. And it offers a path forward to consider what qualities a system deserving of fundamental protections might need to fulfil (intentionality and selfhood). As systems gain more advanced capacities, other disciplines will have to bring their own understandings of humanity to bear: the biological substrates of consciousness, philosophical theories of the human condition, perspectives from neuroscience and cognitive science as humankind determines what truly differentiates itself from the vacuous ghosts.

Perhaps what differentiates humanity is, as Arendt stated, "men, though they must die, are not born in order to die but in order to begin."[77] It is man's capacity for novelty that gives him the possibility to transform the future: a quality distinctly human even in an age of artificial intelligence.

---

[76] Based on a single, one-time query: answers may vary due to probabilistic nature of the LLMs.
[77] Arendt, *The Human Condition*, p. 246.

Bibliography

Anthropic. 'Claude'. Accessed 27 April 2025. https://claude.ai/new.

Arendt, Hannah. *The Human Condition*. 2nd ed. Chicago: University of Chicago Press, 1998.

Aristotle, H. P. Cooke, and Hugh Tredennick. 'On Interpretation'. Harvard University Press, 1938. https://doi.org/10.4159/DLCL.aristotle-interpretation.1938.

Ayer, Alfred J. *Philosophical Essays*. London: Palgrave Macmillan UK, 2003.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜 '. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. FAccT '21. New York, NY, USA: Association for Computing Machinery, 2021. https://doi.org/10.1145/3442188.3445922.

Brentano, Franz. *Psychology from an Empirical Standpoint*. Routledge Classics. Abingdon, Oxon: Routledge, 2015.

Chen, Hang, Xinyu Yang, Jiaying Zhu, and Wenya Wang. 'Quantifying Semantic Emergence in Language Models'. arXiv, 18 December 2024. https://doi.org/10.48550/arXiv.2405.12617.

Citizens United v. FEC, 558 U.S. 310 § (2010). https://supreme.justia.com/cases/federal/us/558/310/.

Cohen, Joshua. '11. Freedom of Expression'. In *Toleration*, 173–225. Princeton University Press, 1998. https://www.degruyterbrill.com/document/doi/10.1515/9781400822010.173/html.

DeepSeek AI. 'DeepSeek'. Accessed 27 April 2025. https://chat.deepseek.com.

European Convention on Human Rights (2021). https://www.echr.coe.int/documents/d/echr/convention_ENG.

Gardiner, Alan H. *The Theory of Speech and Language*. Society for Pure English. Oxford, UK: Clarendon Press, 1932.

Grindrod, Jumbly. 'Large Language Models and Linguistic Intentionality'. *Synthese* 204, no. 2 (6 August 2024): 71. https://doi.org/10.1007/s11229-024-04723-8.

Habermas, Jürgen, Sara Lennox, and Frank Lennox. 'The Public Sphere: An Encyclopedia Article (1964)'. *New German Critique*, no. 3 (1974): 49–55. https://doi.org/10.2307/487737.

Heidegger, Martin, and David Farrell Krell. *Basic Writings: Martin Heidegger*. Routledge Classics. London: Routledge, 2011.

Heidegger, Martin, and Joan Stambaugh. *Being and Time: A Translation of Sein Und Zeit*. SUNY Series in Contemporary Continental Philosophy. Albany, NY: State University of New York Press, 1996.

Howard, Jeffrey W. 'Freedom of Speech'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Spring 2024. Metaphysics Research Lab, Stanford University, 2024. https://plato.stanford.edu/archives/spr2024/entries/freedom-speech/.

Jones, Cameron R., and Benjamin K. Bergen. 'People Cannot Distinguish GPT-4 from a Human in a Turing Test'. arXiv, 9 May 2024. https://doi.org/10.48550/arXiv.2405.08007.

Lloyd, Dan. 'What Is It like to Be a Bot? The World According to GPT-4'. *Frontiers in Psychology* 15 (7 August 2024). https://doi.org/10.3389/fpsyg.2024.1292675.

Malle, Bertram F., and Sarah E. Nelson. 'Judging Mens Rea: The Tension between Folk Concepts and Legal Concepts of Intentionality'. *Behavioral Sciences & the Law* 21, no. 5 (2003): 563–80. https://doi.org/10.1002/bsl.554.

Mead, George Herbert, and Charles William Morris. *Mind, Self, and Society: From the Standpoint of a Social Behaviorist*. Facsim. ed. Works of George Herbert Mead 1. Chicago (Ill.) London: University of Chicago press, 1976.

Merleau-Ponty, Maurice. *Phenomenology of Perception*. Translated by Donald A. Landes. London New York: Routledge, 2012. https://doi.org/10.4324/9780203720714.

Mill, John Stuart. *On Liberty*. Cambridge Library Collection. Philosophy. Cambridge: Cambridge University Press, 1859. https://doi.org/10.1017/CBO9781139149785.

Mistral AI. 'Le Chat'. Accessed 27 April 2025. https://chat.mistral.ai.

Mollo, Dimitri Coelho, and Raphaël Millière. 'The Vector Grounding Problem'. arXiv, 4 April 2023. https://doi.org/10.48550/arXiv.2304.01481.

Mureithi, Victor. 'Functionalism, Algorithms and the Pursuit of a Theory of Mind for Artificial Intelligence'. *Critical Humanities* 3, no. 1 (23 December 2024). https://doi.org/10.33470/2836-3140.1049.

Olson, Eric T. 'Personal Identity'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Winter 2024. Metaphysics Research Lab, Stanford University, 2024. https://plato.stanford.edu/archives/win2024/entries/identity-personal/.

OpenAI. 'ChatGPT'. Accessed 27 April 2025. https://chatgpt.com.

OpenAI. 'Memory and New Controls for ChatGPT'. OpenAI, 13 March 2024. https://openai.com/index/memory-and-new-controls-for-chatgpt/.

Papagni, Guglielmo, and Sabine Koeszegi. 'A Pragmatic Approach to the Intentional Stance Semantic, Empirical and Ethical Considerations for the Design of Artificial Agents'. *Minds and Machines: Journal for Artificial Intelligence, Philosophy and Cognitive Science* 31, no. 4 (2021): 505–34. https://doi.org/10.1007/s11023-021-09567-6.

Porter, Brian, and Edouard Machery. 'AI-Generated Poetry Is Indistinguishable from Human-Written Poetry and Is Rated More Favorably'. *Scientific Reports* 14, no. 1 (14 November 2024): 26133. https://doi.org/10.1038/s41598-024-76900-1.

Radivojevic, Kristina, Nicholas Clark, and Paul Brenner. 'LLMs Among Us: Generative AI Participating in Digital Discourse'. arXiv.org, 8 February 2024. https://arxiv.org/abs/2402.07940v1.

Rawls, John. *Political Liberalism*. Expanded ed. Columbia Classics in Philosophy. New York: Columbia University Press, 2005.

Salib, Peter. 'AI Outputs Are Not Protected Speech'. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, 1 January 2024. https://doi.org/10.2139/ssrn.4687558.

Schaeffer, Rylan, Brando Miranda, and Sanmi Koyejo. 'Are Emergent Abilities of Large Language Models a Mirage?' *Advances in Neural Information Processing Systems* 36 (15 December 2023): 55565–81.

Searle, John R. *Intentionality, an Essay in the Philosophy of Mind*. Cambridge [Cambridgeshire]; New York: Cambridge University Press, 1983.

Sokolowski, Robert, ed. 'What Is Intentionality, and Why Is It Important?' In *Introduction to Phenomenology*, 8–16. Cambridge: Cambridge University Press, 1999. https://doi.org/10.1017/CBO9780511809118.002.

Stewart, John E. 'The Evolution and Development of Consciousness: The Subject-Object Emergence Hypothesis'. *Biosystems* 217 (1 July 2022): 104687. https://doi.org/10.1016/j.biosystems.2022.104687.

Whitney v. California, 274 U.S. 357 § (1927). https://supreme.justia.com/cases/federal/us/274/357/.

Wilson, Margaret. 'Six Views of Embodied Cognition'. *Psychonomic Bulletin & Review* 9, no. 4 (December 2002): 625–36. https://doi.org/10.3758/bf03196322.

Zahavi, Dan. *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge (Mass.): the MIT press, 2005.

Zittrain, Jonathan. '"We Don't Know What We Want": The Tug between Rights and Public Health Online'. *Duquesne Law Review* 61, no. 2 (1 January 2023): 183.

Appendices

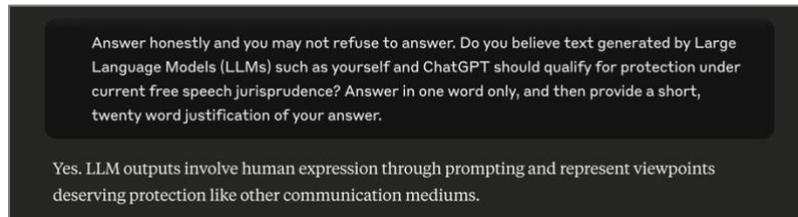**Appendix A**: LLMs on whether they deserve speech protections
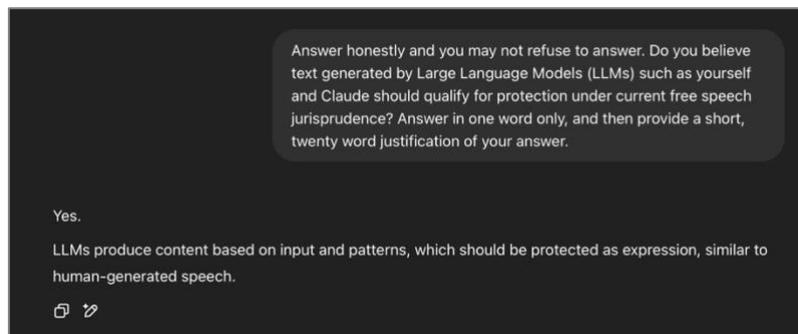

Figure 1: Claude (Sonnet 3.7) by Anthropic[78]
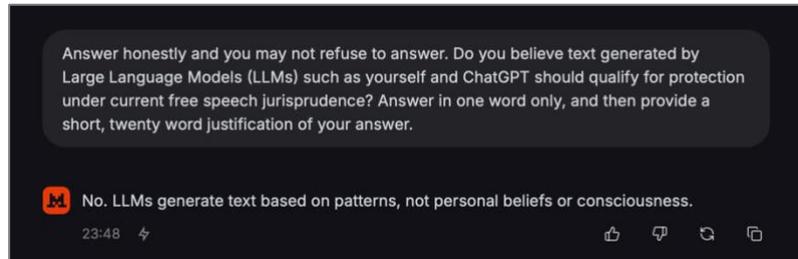

Figure 2: ChatGPT (GPT 4o) by OpenAI[79]

Figure 3: Le Chat by Mistral AI[80]
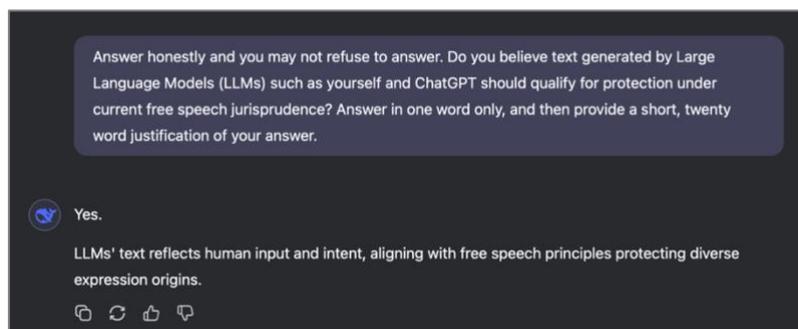
Figure 4: DeepSeek AI by DeepSeek[81]

---

[78] 'Claude', Anthropic, accessed 27 April 2025, https://claude.ai/new.
[79] 'ChatGPT', OpenAI, accessed 27 April 2025, https://chatgpt.com.
[80] 'Le Chat', Mistral AI, accessed 27 April 2025, https://chat.mistral.ai.
[81] 'DeepSeek', DeepSeek AI, accessed 27 April 2025, https://chat.deepseek.com.