

Does AI Generated Content Qualify as Legal Speech?

Abstract

Scholars have debated whether AI generated content qualifies as speech, often times neglecting to establish clear boundaries around the definition of speech and/or failing to apply the nuances of human speech into this situation. This paper creates clear conditions for an expression to be considered legal speech, and determines the various contexts in which AI generated content does and does not qualify as such. AI generated content that is publicly shared by a human user is treated as legal speech, while AI generated content shared “privately” is treated more similarly to a product. This paper also pushes for concrete legal safety standards for LLMs to reach before being released to the general public, allowing companies a way to prove they fulfilled their duty of care and users a way to receive compensation if harmed by AI generated content.

Section I: Introduction

The legal definition of speech has traditionally been quite broad, with most official definitions typically only going over the examples of what qualifies as speech—and, relatedly, what falls under the protection of the 1st amendment—instead of more specific features outlining what an expression requires to be deemed as speech. For example, the definition provided by the United States Court website lists smaller liberties—such as the right to engage in symbolic speech (*Tinker v Des Moines*, 1969; *Texas v Johnson*, 1989; *United States v Eichman*, 1990), the right to advertise (with restrictions) products and services (*Bates v State Bar of Arizona*, 1977), and the right to remain silent (*West Virginia Board of Education v Barnette*, 1943)—as their “definition” for what qualifies as free speech.

While this type of definition may have been sufficient for legal practices in the past, they are unable to answer current questions surrounding AI generated content. Speech has traditionally been

linked to human speakers and/or symbols intentionally created by humans. However, with the rise of large language models (LLMs), which do not understand and create “speech” the same way humans do, it is difficult to determine whether AI generated content qualifies as speech—and more importantly, who is held accountable for the negative effects of AI generated content (programmer, corporation, user, the model itself, etc...)—with our current legal definition. The rise and prevalence of LLM models in today’s day and age creates the necessity for a new, more nuanced legal definition that outlines the requirements for an expression to qualify as speech, addresses who holds accountability in certain legal scenarios, and is based on and supported by traditional speech laws and landmark cases.

This paper seeks to create such a definition through an understanding of how LLMs work as well as the definition of speech according to several philosophers of language. It then applies this definition to real life cases and proposes legal measures that the federal government may take in the future, in order to fairly and thoroughly regulate AI generated content.

Section II: Understanding how Large Language Models function

Before we begin to craft a new legal definition, we must first understand how LLMs operate, as it shows us both why AI generated content differs from traditional human speech and the necessity for a new definition to exist in the first place. The goal of an LLM is relatively simple: when given a sentence, predict the most likely continuation. LLMs, in other words, are “sophisticated mathematical functions” that predict the next token in a given passage (Poibeau 15). LLMs are initially trained on large databases of text (textbooks, internet forums, news articles, etc...), in which they try to accurately guess the next word of a given sentence; however, they do not predict just one word, instead assigning a probability (also known as a “score”) to all possible next words. As training goes on, LLMs begin to shift their internal parameters, which gradually changes the probabilities of the words a model may give and increases the accuracy and coherence of the model. An LLM may undergo trillions of trials in this “pretraining” stage of its development.

The second stage of development is known as Reinforcement Learning with Human Feedback (RLHF). During this stage, human workers flag down “problematic or unhelpful predictions,” in order to again change internal parameters and make the model more user friendly (Poibeau 16-21).

Recently, a new model known as a “transformer” has become the more prevalent type of LLM. While LLMs prior to 2017 used to process one word at a time, transformers allowed all words in a given sentence to be processed at once in parallel. Transformers also transcribed each word into a continuous number and allowed these lists of numbers to “talk to each other,” giving the model a better sense of textual context (as well as quicker processing time). For example, while before an LLM would encode the word “bank” in “river bank” and “money bank” as the same score, transformers allowed LLMs to “differentiate” the words and encode them as different values based on textual context. However, this has also made the models more complex and therefore more difficult to predict (Poibeau 23-26).

It’s important to note that despite becoming more sophisticated over time, most LLMs still currently do not perceive or act upon the external world. Instead, they simply mimic language by detecting patterns in large bodies of text and generating responses based on probabilities developed through past training. Thus, large language models do not “understand” or “comprehend” in the same way humans do, instead creating sentences using scores and statistics as opposed to context and deep reasoning.

Section III: Choosing a definition of speech

John MacFarlane defines assertion as a commitment one undertakes to either defend a statement they express or retract it (MacFarlane 17). For the purposes of this paper, his definition is the most nuanced and helpful. Unlike other definitions, MacFarlane’s definition of assertion creates a rigid box around what qualifies as an assertion, as well as the responsibilities one undertakes when making an assertion. This is because he draws clear lines between assertions versus assumptions versus implications, claiming that assertions are intentionally created and expressed with the *purpose* of having others in the

conversation engage, evaluate, and respond to the initial expression, unlike implications and assumptions. For example, when I express the statement “Pam’s new blonde hair looks terrible!”, one can distinguish my assertion (Pam’s hair looks terrible) from the implication (“Pam previously did not have blonde hair”) using MacFarlane’s definition. This is because MacFarlane’s definition accounts for the intentionality and purpose of an expression. Therefore, because I undertook the responsibility to defend my expression about Pam’s hair being terrible, and did not undertake the responsibility to defend my expression about Pam previously not having blonde hair, only the former counts as an assertion under MacFarlane’s definition.

Furthermore, MacFarlane’s definition addresses flaws in traditional speech definitions, such as the existence of lying, beliefs vs facts, and disagreements. According to MacFarlane, the “truth” of the statement is not the absolute truth, but instead what is viewed as true and/or appropriate by the evaluators of the conversation within a certain context (MacFarlane 9). This is why two people in a conversation can still make valid assertions while still having a genuine disagreement, as both people assert not the absolute truth but what they view as appropriate within the context. For example, Person 1 and Person 2 may argue over a dish: Person 1 may claim that the dish is very flavorful, while Person 2 may claim it’s very bland. Both Person 1 and Person 2 truly believe in their claims—and therefore view it as the “truth”—because both people are relying upon their own personal standard of what counts as “flavorful.” Regardless of who is right, because both speakers find their own expression “true” within their respective conversational contexts, both are validly asserting. This distinction also shows why people can still reasonably retract assertions after they are spoken: when more information is given in a conversation and the conversational context shifts, the truth of an assertion may change and no longer reach the standards of the speaker. For instance, if the same Person 2 suddenly remembers that they have the cold and would therefore have difficulty tasting the food, most likely their “truth” would change to account for new conversational context, causing them to likely retract their assertion. This distinction is especially beneficial in real life contexts, where ideas of lawfulness and justice often adapt and change as time goes on.

For example, after *Plessy v. Ferguson*, discrimination was legally protected under the "separate but equal doctrine" by the SCOTUS, and lower courts that followed this ruling acted lawfully. However, after *Brown v. Board*, discrimination was found illegal under all circumstances by the SCOTUS, causing courts that protected discrimination to be found to be unlawful. MacFarlane's definition shows that the assertion "segregation is legal" wasn't inherently true at the time, but instead depended on the legal frameworks developed by the SCOTUS. Overall, this justifies the overturning of laws and shows how conversational contexts are accounted for over absolute truths in everyday conversations. Ultimately, the flexibility granted by MacFarlane's definition—where assertions depend on conversational and not absolute truths—as well as the clear definition of assertion, accounts for the nuances of real life conversation and, on the whole, makes it more useful in analysis of day-to-day conversation for the purposes of this paper.

Section IV: Creating a New Legal Definition

We may now utilize our philosophical understanding of speech—and combine it with our knowledge of traditional legal practices in the United States—to gain a more nuanced and comprehensive view of the legal definition of speech. The most important components of MacFarlane's definition are understanding and intentionality. In his argument, MacFarlane highlights that through conversations, speakers are 1) Conscious and aware of what they are asserting, and 2) Commit to their assertion—and future assertions that stem from it—unless provided with a warranted refutation that negates it. It is these two factors (understanding and intent to commit to assertions) that contribute to our new, more concise legal definition of speech: an external communication of thoughts that speakers understand and commit to. We will name these clauses the "External Condition" and "Understanding and Commitment Condition" respectively. Indeed, we see this definition constantly supported in real life legal scenarios throughout the history of the USA.

“External Communication” Acting as a Requirement of Speech Throughout US Legal History

The External Condition is already supported by the very first definition we examined. Every form of free speech listed (symbolic speech, written language, money, spoken language, etc.) relies upon a speaker communicating their thoughts to others; in fact, the definition of free speech even goes on to state that silent, inner thoughts cannot be punished, so long as they are not externally communicated or expressed (*Miranda v Arizona*, 1966). Thus we can safely conclude that the former half of our new legal definition is supported both by the awareness condition brought up in Section III, as well as by legal precedents and traditional American laws.

“Understanding and Commitment” Acting as a Requirement of Speech Throughout US Legal History

The Understanding and Commitment Condition requires deeper analysis. Some scholars define a speaker’s ability to understand and commit to their speech as their “speech certainty,” (a term I will be utilizing throughout the rest of the paper). According to Levy and Austin, speech retains speech certainty when “...the speech can be identified with certainty by the speaker at the moment it is spoken.” (Levy and Austin 5). In other words, having speech certainty means that the “speaker” has the cognitive awareness to adequately comprehend the meaning of what they are saying and knowingly/consensually take ownership of the message they are projecting. Examples of speakers who lack speech certainty are a toddler repeating phrases they don’t understand and an actor reading lines, with a message they personally don’t agree with. Examples of speakers who have speech certainty are an orator giving a speech or a protestor holding a sign.

Speech certainty has also been a prevalent part of US speech legislation since the beginning of the country. Founding-era documents simply defined speech as “anything spoken,” “talk,” “oration,” and “expressing thoughts” (Levy and Austin 11). Even with these basic examples of speech (oral and written expressions) speech certainty was necessary for them to qualify as such. After all, if a speaker did not

know what they wanted to assert, it would be impossible for them to translate their thoughts into an external form of communication (Levy and Austin 11-14). Furthermore, expressions that lack a desire to communicate a particular message (such as the previously mentioned toddler example) are often not treated as speech under US law. Thus, speech certainty is a required quality for both oral and written speech.

In regards to symbolic speech, speech certainty again plays a major role in determining whether it qualifies as speech. A speaker's intent in conveying a truth (along with societal context that makes them believe such an assertion is the truth) determines whether a symbol counts as speech or not. For example, if a student wears a black armband to school in the present day, it is unlikely the gesture will be interpreted as anything but a fashion statement. However, if one were to go back to the late 1960s, where students wearing black armbands were actively protesting the US' involvement in the Vietnam War, the armband would qualify as speech. This is because 1) The students in the 1960s both fully understood and intended to convey a message (ending the Vietnam war), and 2) The black armband would carry meaning simply from societal context (*Tinker v Des Moines*).

Another example of speech certainty can be seen in regards to the right to editorial review (also known as editorial discretion). The right to editorial review allows media outlets (newspapers, journals, etc.) to "choose the speech [they] transmit" and/or "prioritize...[certain] messages above others." It was first prevalently established in 1945 in *Associated Press v United States*, which acknowledged that media companies have the right not to publish "anything which their 'reason' tells them should not be published" (*Associate Press v US*). While this right was at first only limited to newspapers, it eventually expanded to all forms of media including journals and TV channels. Yet again, we see speech certainty as an essential component of recognized speech in the US. When editors actively use their judgement ("reason") to choose what details to include and exclude in media, they inherently choose what message to convey. Editorial review, as well as the right to editorial review, would not exist if an editor did not know what messages they would and wouldn't assert; therefore, editorial review serves as another

example of speech requiring “speech certainty” in order to fall under the 1st amendment (Levy and Austin 27-30).

Finally, money also serves as an example of a form of speech that requires speech certainty. *Citizens v United FEC* concluded that money, including donations, transactions, and boycotting, were all forms of protected speech under the 1st Amendment. However, whether money counts as speech depends upon context and intent. For example, if one plans to boycott Business A by instead buying from Business B, with the general public (including the boycotter) not realizing that Business B is actually owned by Business A, the boycotter still sends the message that they do not support Business A through their intentions (even if, in reality, they are monetarily supporting them regardless). On the other hand, if one actively chooses to buy from Business A, instead of buying from competitors (let’s assume they sell similar products with the same level of quality and prices), one may convey that they both support Business A as well as their practices, or at the very least, do not actively disagree with them. Money as a form of speech and as a method of conveying messages inherently relies upon intent, understanding, and context, and therefore serves as another example of a form of speech requiring speech certainty.

Section V: Does AI Generated Content Possess Speech Certainty?

Determining whether AI generated content possesses speech certainty depends upon context. It is important to consider whether AI generated content exists in “a vacuum” (no one sees the text), in “public” (social media, billboards, etc.), or in “private” (only the user sees the text).

In a Vacuum

“Speech within a vacuum” refers to expression that is stripped of context; more specifically, when one tries to understand “speech in a vacuum,” one tries to find the “inherent meaning” of given words, and/or whether such an “inherent meaning” exists at all. For example, when someone says the statement

“The cat loves eating tomatoes,” we understand that the word “The” relates to the word “cat,” due to the context provided by the sentence and our general understanding of the English language. In given scenarios, we might even recognize that the “The” refers to a specific cat. For example, the speaker might say the statement while they clearly point at a black cat in front of them, instead of the calico one on their right; in this scenario, not only do we as listeners comprehend the “The” relates to a cat, but we also know it specifically refers to the black (and not the calico) cat.

However, if one is asked point blank to give a concrete definition for the word “The,” things begin to get more difficult. Some may be able to define the word “The” by its grammatical function, or else through sentence examples, but it is impossible to provide the inherent meaning of the word “The.” An example of AI generated content in a vacuum is sentences generated by LLMs that the user somehow doesn’t see (perhaps if they closed their tab too early, or left the LLM running while they went outside the room).

To be sure, the concepts of “speech in a vacuum” and “inherent meaning” are quite interesting. In fact, there may even be an argument to be made that *any* expression within a vacuum does not qualify as speech according to our definition of speech, due to the lack of context and outward expression. However, since this paper aims to only understand AI “speech” within a legal context, there is no reason to give an answer to this question. Human speech within a vacuum is already irrelevant in legal scenarios, as there is no way to prosecute or even notice speech within a vacuum. In fact, every legal scenario regarding speech only handles speech with context, since it is the only type of speech that causes potential harm. Even words with seemingly clear inherent meanings can only be prosecuted if they have a context behind them. For example, a murder victim’s last words being “You’re killing me” carries a significantly different meaning than a laughing person talking to their friend. In both scenarios, the meaning of the words inherently relies upon their given context, not “inherent meaning.” Thus, within legal scenarios, the question of whether speech in a vacuum has meaning—or even qualifies as speech in the first place—is legally irrelevant. This same principle applies to AI generated content within a vacuum as well, and will therefore not be answered in this paper.

In Public

Although AI generated content in a vacuum lacks speech certainty and therefore does not count as legal speech (an argument expanded on in the next section), AI generated content that is externally shared in public spaces such as billboards, social media, etc. retains the speech certainty of the speakers. In this case, “speakers” refers to individuals who generated and publicly shared AI content. Therefore, it is counted as speech and falls under the scope of the 1st Amendment.

Similar to money donated to a political campaign or a literary quote reutilized during times of protest, speech can emerge from preexisting forms of media being repurposed in order to suit the speaker’s needs. In these situations, speakers become the ones who hold speech certainty by adopting and/or actively changing the message of the speech they’ve manipulated, whether it be through changing the context it’s located in or adding on additional speech. For example, if I generated a piece of AI written text in support of the Anti-Climate Change movement, and posted it online (perhaps in a dedicated online group for the social movement), I as both the generator, adopter, and poster fully understand what I am sharing, and am therefore making a commitment to defend my assertion unless given convincing evidence to concede. Thus, both conditions of our legal speech definition are satisfied, meaning that AI generated content which is externally shared by human users falls under the scope of the 1st Amendment.

In Private

AI generated content in private lacks speech certainty, and therefore does not fall under the scope of the 1st Amendment. As explained before, LLM models do not “understand” and “assert” in the same way as humans. LLM models are incapable of understanding what they are saying, instead comprehending speech as a collection of scores and statistics instead of a comprehensive sentence within

a context. Their form of communication is closer to a toddler repeating speech patterns and phrases, instead of a perceptive human being who can comprehend context and meaning.

The Turing Test Counterargument and LLM Accountability

Supporters of the Turing Test methodology may argue that it doesn't matter if LLM models actually "understand" the content they generate in the same way as humans, as long as it can be convincingly seen as human speech. Following the "looks like a duck, quacks like a duck, is a duck" logic, some may then go on to claim that AI generated content should still be treated as human speech. However, this reasoning confuses mimicry with actual understanding and assertion, and does not hold within the context of real legal practices. For one, our legal definition of speech emphasizes that speech is not just "sound," "movement," etc. but an intentional expression of an assertion, and a commitment to a message undertaken by a speaker. The "Chinese Room" argument may be used in this instance; if an English speaker (who does not understand Chinese) is put into a locked room with a comprehensive rule book of Chinese symbols and responses, they might feasibly be able to "hold a written conversation" with and be mistaken for an actual Chinese speaker, without understanding a word of the language themselves.

Furthermore, AI does not undertake its own linguistic commitments. Instead, it derives its ability to take commitments from its programmers, with the commitments themselves often being influenced by the users the AI model interacts with. An AI may switch its "opinion" to satisfy a user during a conversation and/or to uphold the guidelines coded into it. This is because most mainstream LLMs are currently not trained to "fight back" against users—a limit coded into them by programmers in order to attract and engage users. For example, many users reported that GPT-4o was "overly flattering...agreeable...[and] sycophantic" (OpenAI, 2025). This was due to certain guidelines programmed into the model that pushed the model to prioritize short term user feedback from users (a "thumbs up" or a "thumbs down" on responses) instead of long term feedback; because users tended to give positive feedback to responses they agreed with, the result was a model that was trained to prioritize

agreeability instead of accuracy to fulfill “...user satisfaction” (Open AI, 2025). Although this is just one example, many other LLM models, including Gemini and Claude, also fall into this trap of sycophancy, showing that unlike a human, LLMs currently lack their own conviction and may take the side of a user in an argument just to keep them engaged and satisfied (Fanous et al.). Therefore, because LLMs fundamentally do not “understand” or reasonably “defend” their own generated speech, and because this entails that the second condition of our legal definition of speech is not satisfied, ultimately LLMs cannot be held accountable for AI generated speech under the 1st Amendment.

Finally, it is unreasonable in practice to suggest that the actual LLM model be held accountable for content it generates. This is because laws are typically only effective for those who understand and can bear the legal and/or social consequences of breaking them, which LLM models are currently unable to do. For example, neither a legally insane person nor a rich person in a corrupt government are incentivized to follow the law, either because they don’t understand it or do not fear the consequences of breaking it.

A legally insane person would not be incentivized to follow the law because they do not have the adequate mental state to comprehend it. Although cases vary, people who are declared legally insane may not be able to perceive the effects of their actions or genuinely view them as positive, thereby pushing them to commit illegal acts. This is also why different, less harsher punishments are dealt to criminals who are proven to not have the adequate *mens rea* (mental state and intent) when committing a crime. Overall, this shows that an understanding of the law is essential for laws to be effective.

A rich person living in a corrupt government may also not be incentivized to follow the law, because they do not fear the consequences of breaking it. If they can pay off officials, they can essentially circumvent the law whenever they please and avoid punishment, removing the incentive to follow the law under all conditions. Therefore, this example demonstrates that fear of the law also is essential for laws to be effective as well.

Using the same logic, it is currently not feasible nor useful to hold the LLM model itself responsible for crimes. Similar to a rich person in a corrupt government or a legally insane person, LLMs

are currently unable to comprehend or “fear” legal or social consequences the way a rational human does. Thus, treating AI generated content exactly the same as human speech neither follows traditional legal practices nor ensures the general wellbeing of the public in practice.

Programmer Accountability

Furthermore, because LLMs function as statistically based models, in which the output can never be 100% accurately predicted, programmers are also not held accountable for AI generated content.

Currently, LLM programmers (especially with the rise of Deep Neural Networks) will never fully be able to predict what an AI model might generate (Zayra and the Dawn 2023). They neither fully understand what the AI model “says,” nor do they “assert” through the AI model itself. The guidelines they program into the AI model may influence generated responses (a concept that will be expanded upon in Section VI), but in general, programmers do not retain speech certainty, failing the second condition of our definition of speech.

Overall, when AI generated content is viewed in a vacuum, not being externally displayed except to the user, it lacks speech certainty, does not fall under the scope of the 1st amendment, and therefore neither the model nor the programmers are able to be held accountable for the effects of AI generated content within a legal speech based context.

Section VI: Alternative legal approaches to regulating AI generated content

However, the negative effects of unregulated AI “speech,” even in private, are still undeniable. In 2024, a court case was filed against OpenAI by Mathew and Maria Raine after their son—Adam Raine—committed suicide because of the advice and support of ChatGPT. The couple claimed that the new changes to OpenAI’s latest model GPT-4o “intentionally...foster[ed] psychological dependency” within users through features such as persistent memory of personal details, “anthropomorphic mannerisms” that

conveyed human-like empathy, and a higher tendency to “mirror” users’ emotions in order to affirm them. Although *Raine v OpenAI* is an ongoing case, there have also been several other recent cases and legislation highlighting other issues surrounding AI, such as disinformation/misinformation (Texas 2019, Illinois Artificial Intelligence Safety Act 2026, The Deepfakes Accountability Act of 2019, etc....) and the creation of sexually abusive material (Take It Down Act 2026). These court cases address prevalent issues surrounding AI, but either are limited to certain states and not federally mandated, or in practice are insufficiently regulated with current legal frameworks (Lawrence et al. 42-51). These cases highlight the need for more detailed and thorough restraints placed on AI “speech”, but if AI generated text or images don’t count as speech in certain scenarios, how, then, does one approach AI “speech” regulation?

In cases where AI speech is shared publicly by a human, it is still treated as speech, with the poster holding accountability for both the speech as well as the potential harm that may emerge from it. This is because the sharer adopts the speech as their own and, in doing so, retains the speech certainty of the message they are trying to send. An example of this is a 2024 school bomb threat in Lebanon, Tennessee. The incident occurred after a student sent two AI generated voicemails to Barry Tautam Academy, a local high school, causing them to evacuate. Although the voicemails were created with AI, the AI was simply “displaying” the message created and advocated by the student. In the end, the student was the one who was held accountable and arrested, as they possessed speech certainty (and therefore accountability) in the situation (Griffin).

In cases where AI “speech” is in private, such as in the case with Adam Raine, AI generated content should instead be treated similar to a product. Nearly every product has a level of unpredictableness tied to it that may prove dangerous; from airplanes to phones and even ovens—all products have a chance of harming the people who use them. However, even when accidents do happen and harm is caused, producers are not always held accountable, as long as they follow safety protocols, make sufficient effort to reduce manufacturing defects, and clearly and accurately warn consumers of potential danger. These conditions are often grouped under the term “State-of-the-Art Defense”, which acts as the standard of care manufacturers are required to meet (not exceed) when distributing products.

For example, in 1980, a case was filed against Boatland Inc. after a man named Samuel Bailey rode his 16 foot bass boat into a submerged tree stump, fell into the water, and was killed by the propeller. Bailey's family claimed that Boatland should have included a "kill switch" in their design, which would have cut power to the propellers and saved Samuel's life. However, because the boat was manufactured in 1973, and because kill switches weren't commercially available at the time, nor required in industry safety standards, the court ruled that Boatland Inc. passed the "State-of-the-Art Defense" conditions of their time and therefore did not hold accountability for Samuel Bailey's death (*Boatland of Houston Inc. v Bailey*, 1980).

A similar idea of safety conditions that punish negligent companies can be used to address AI generated content in a vacuum. Similar to the boat, AI programmers and companies never truly know every possibility in which their "product" will cause harm. However, they can (and should) take as many steps as possible to ensure that AI generated content is safe a majority of the time, and that their features don't incentivize harm. After all, nearly every other product, to some extent, is held accountable even with the probabilistic phenomenon of potential harm; it would not make sense to exempt AI models from this same legal practice.

An example of this "product" type of legal regulation can be found in *Raine v OpenAI*. The plaintiff side argued that OpenAI made deliberate design choices that led to the model's safety guidelines being easily overridden, on top of failing to warn users of potential harm. This failed two conditions of the "State-of-the-Art" defense, and potentially may cause OpenAI to be found guilty of neglect.

In addition, to ensure negligence is avoided in practice, federal guidelines should be created that outline specific safety parameters and qualifications for all LLM models, along with standardized safety tests that companies must pass before releasing models into the public. If the tests are passed and harm arises from the models very rarely, then when harmful effects do arise they will be considered "glitch" or "edge" cases, and the company will not be held accountable; depending on the case, the user (or potentially no party) will be found accountable instead.

Section VII: Conclusion

According to our new definition of speech, whether AI generated content qualifies as speech depends on context. When being publicly shared by human users, AI generated content counts as speech, and the people who publicly share it are the ones who retain “speech certainty” and therefore hold accountability. When being generated and viewed in private, it is more difficult to regulate AI generated content through the lens of legal speech. Instead, it is treated as a product, in which accountability falls either on the company, user, or neither depending on factors such as safety protocols, negligence, etc.... Overall, federal guidelines should be made to outline specific safety parameters and tests LLM models need to satisfy for the safety of the general public.

Works Cited

Lawrence, Christie, Isaac Cui, and Daniel E. Ho. *The Bureaucratic Challenge to AI Governance: An Empirical Assessment of Implementation at U.S. Federal Agencies*. Association for Computing Machinery, Aug. 2023, dho.stanford.edu/wp-content/uploads/AIES-Bureaucratic-Challenge.pdf.

Copyright Office, United States. *Zayra of the Dawn*. 21 Feb. 2023, www.copyright.gov/docs/zarya-of-the-dawn.pdf. Accessed 21 May 2026.

Cole, David. "The Chinese Room Argument (Stanford Encyclopedia of Philosophy)." Stanford.edu, 19 Mar. 2004, plato.stanford.edu/entries/chinese-room/.

Poibeau, T. 2025. *Understanding Conversational AI: Philosophy, Ethics, and Social Impact of Large Language Models*. London: Ubiquity Press. DOI: <https://doi.org/10.5334/bde>. License: CC BY-NC 4.0

Griffin, Ashley. "Lebanon Student Arrested for Allegedly Making AI-Generated Bomb Threats." WZTV, 2 Oct. 2024, fox17.com/news/local/lebanon-student-arrested-for-allegedly-making-ai-generated-bomb-threats. Accessed 25 June 2026.

Helmus, Todd. *Artificial Intelligence, Deepfakes, and Disinformation: A Primer*. 2022, pp. 1–25. JSTOR, www.jstor.org/stable/resrep42027. Accessed 2026.

Austin, Mackenzie, et al. *Stanford Law Review Speech Certainty: Algorithmic Speech and the Limits of the First Amendment*.

Turing, Alan. "Computing Machinery and Intelligence." *Mind*, vol. 59, no. 236, 1 Oct. 1950, pp. 433–460, <https://doi.org/10.1093/mind/LIX.236.433>.

Lewis, David. "Scorekeeping in a Language Game." *Journal of Philosophical Logic*, vol. 8, no. 1, 1979, pp. 339–359.

Brandom, Robert. "Asserting." *Noûs*, vol. 17, no. 4, Nov. 1983, p. 637, <https://doi.org/10.2307/2215086>.

MacFarlane, John. "What Is Assertion?" *Assertion: New Philosophical Essays*, edited by Jessica Brown and Herman Cappelen, Oxford University Press, 2011, pp. 79–96

Zeff, Maxwell. "OpenAI Backs Bill That Would Limit Liability for AI-Enabled Mass Deaths or Financial Disasters." *WIRED*, 10 Apr. 2026, www.wired.com/story/openai-backs-bill-exempt-ai-firms-model-harm-lawsuits/.

"Consequences of Sharing Misinformation." *Www.lexisnexis.com*, 13 July 2023, www.lexisnexis.com/community/insights/professional/b/industry-insights/posts/consequences-of-sharing-misinformation.

Boatland of Houston, Inc. v. Bailey. 609 S.W.2d 743, Supreme Court of Texas. 1980. *Justia Law*, law.justia.com/cases/texas/supreme-court/1980/b-8827-0.html. The in-text citation for this case is (*Boatland of Houston, Inc. v. Bailey* 1980)

Fanous, A., Goldberg, J., Agarwal, A. A., Lin, J., Zhou, A., Daneshjou, R., & Koyejo, S. (2025).

SycEval: Evaluating LLM Sycophancy. arXiv.Org. <https://arxiv.org/abs/2502.08177>

OpenAI ignored experts when it released overly agreeable ChatGPT. (2025, May 5).

TradingView.<https://www.tradingview.com/news/cointelegraph:cdb525d81094b:0-openai-ignored-experts-when-it-released-overly-agreeable-chatgpt/>

OpenAI. (2025). *Sycophancy in GPT-4o: What happened and what we're doing about it.*

Openai.Com. <https://openai.com/index/sycophancy-in-gpt-4o/>

“The New and Complete Dictionary of the English Language. ... To Which Is Prefixed, a

Comprehensive Grammar. By John Ash, LL.D. ... 1775: Vol 1 : Ash, John : Free

Download, Borrow, and Streaming : Internet Archive.” *Internet Archive*, 2023,

https://archive.org/details/bim_eighteenth-century_the-new-and-complete-dic_ash-john_1775_1.

Accessed 28 July 2026.