

Free Expression

Section 230 and its Applicability to Generative AI: A Legal Analysis

September 4, 2024

By: CDT Intern Noor Waheed

Introduction

The rise of Generative AI in the world of technology, especially within the last year, has unsurprisingly led to myriad questions and concerns regarding its responsible governance in policy spaces across the political spectrum. One such hotly debated question is whether [Section 230 of the Communications Decency Act 1996](#) applies to the outputs created by Generative AI systems. In 2023, Senator Josh Hawley introduced legislation intended to exclude Generative AI systems from the [purview of Section 230](#). In the same year, Senator Ron Wyden and former Representative Chris Cox, the authors of Section 230, wrote that [Section 230 does not protect generative AI outputs](#). The answer, however, may

be more complicated.

Section 230 is the [federal safe-harbor](#) law that protects online intermediaries from liability for the third-party content posted and disseminated on their platforms. At the heart of Section 230 is Congress' realization that the free expression rights of everyday users of online services [depend on online intermediaries' ability](#) to host a variety of content with their own content guidelines and removal standards. At the same time, Section 230 incentivizes online speech intermediaries to curate spaces by moderating or removing "objectionable" content from their services, without fear of liability. That being said, Section 230 immunity is not unlimited. It only protects online intermediaries from liability for third-party generated content. These online intermediaries [remain legally responsible](#) for the content they generate themselves, whether in whole or in part, and for their own conduct. This means that when illegal content created by a user or by a third party is uploaded on X or Facebook, the platform would be protected by Section 230 from liability for that content. However, if either company contributed to the creation of the illegal content, in whole or in part, then Section 230 would provide no protection.

The recent exponential growth and development in Generative AI technologies such as OpenAI's ChatGPT and Dall-E, and Google's Bard may blur the line between what content is considered user-generated versus created in whole or in part by generative AI systems themselves. Generative AI systems use machine learning to generate new content, which can take the form of text, audio, video, or images. Massive amounts of data is required to train the Generative AI model, which can then create new outputs, based upon, but not always identical to, the

training data. Additionally, the output generated by the model can be influenced by text prompts made by the user. Generative AI may also in some cases produce “hallucinated outputs” – i.e., incorrect or misleading results – that exceed the scope of the training data or can be the result of incorrectly decoded data. The outputs of generative AI systems are a product of complex interactions between many moving parts – generative AI models learn patterns from underlying data, which interact with user-generated prompts and additional content filters applied on the back end by companies to limit harmful or policy-violating inputs and outputs. This can understandably complicate the determination of who is responsible for “creating” the end product. The fact that generative AI models can produce novel content in response to user prompts creates interesting legal questions surrounding whether the output of Generative AI systems receive Section 230 protection.

The Criteria for Section 230 Immunity

Section 230 subsection (c)(1), in relevant part, states that no provider or user of an interactive computer service may be treated as the publisher or speaker of content provided by another information content provider. In order to qualify for this immunity, an online service provider must be an “interactive computer service,” i.e., “any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server.” The [courts have found](#) that companies providing such services include broadband internet access services, social media services, and basically any other service that transmits information over the internet. An interactive computer service also notably includes in the definition “access software provider” which is a provider of software or enabling tools that can “filter, screen, allow, or

disallow content”; “pick, choose, analyze, or digest content”; or “transmit, receive, display, forward, cache, search, subset, organize, reorganize, or translate content.” This criteria is important because it captures many of the functions performed by Generative AI systems and makes clear that Generative AI systems are, indeed, a type of interactive computer service.

In the most [familiar application of Section 230](#), information content providers are most commonly users themselves, posting or reposting content to a platform. Section 230 explicitly protects providers and users of an interactive computer service from being treated as the publisher or speaker of any content provided by *another* information content provider (e.g., a user). The interactive computer service, however, [is not exempt from liability](#) if it acts as an “information content provider” with respect to the content at issue in a given case. Courts have found that interactive computer services are acting as information content providers when they conform to the statutory definition, i.e., they are “responsible, in whole or in part, for the creation or development of information provided through the Internet.” The critical question for the availability of Section 230’s shield, therefore, often relates to whether an online intermediary is also an information content provider with respect to the content at issue. To answer this question, [courts look to](#) whether the service provider developed the content, in whole or in part.

The Case (or Lack Thereof) for Section 230’s Application to Generative AI System Outputs

The question of whether Section 230 applies to Generative AI system outputs will probably most often turn on whether the model developed

the content at issue in whole or in part. Given that Generative AI systems engage in a wide breadth of functions, some of which may not involve the creation of original content or might be significantly shaped by user prompts, determining whether the system is an “information content provider” with respect to particular content, and thus [outside the scope](#) of Section 230 immunity, would likely vary on a case by case basis.

In cases involving tools like ChatGPT or Dall-E, where Generative AI is involved in the creation, “in whole or in part” of the offending material, it appears likely that courts will find that Section 230 immunity does not apply. Though the Supreme Court has yet to rule on this, the precedent thus far spanning across various federal appeals courts generally involves applying the “material contribution test” to determine if an entity has contributed significantly enough to the creation of content to qualify as an information content provider. In the *Fair Housing Council of San Francisco Valley v. Roommates.com LLC*, the case that announced the material contribution test, the Ninth Circuit distinguished between mere dissemination of user created input (in this case discriminatory criteria for a housing opportunity included in a blank text input box) and systems that by design limited housing listings based on discriminatory criteria (i.e., a drop-down menu that required users to make discriminatory choices). In the former instance, the Court held that the website did not materially contribute to the offending content (but rather acted like a “neutral tool”) and in the latter case, where the site design required potentially discriminatory inputs, it did. Applying this reasoning to the output of Generative AI systems, if, for example, the system returned a verbatim prompt created by a user, there may not be a material contribution to the creation of the content. On the other hand, if the model were to return a wholly new original output that contained

potentially defamatory or illegally discriminatory content in response to an otherwise legal prompt, the output would likely be a material contribution to the allegedly illegal nature of the content.

Recently, two cases have emerged regarding the legal accountability of Generative AI systems for their outputs that potentially raise the question of Section 230 applicability and the threshold for the material contribution test in the case of Generative AI. In *Walters v. OpenAI*, Mark Walters, a radio host, was falsely described as having embezzled funds from a non-profit organization when ChatGPT generated the information in response to a request [from a third party journalist](#) asking to summarize a real federal court case by linking to an online PDF. ChatGPT created a false summary with factually incorrect information in addition to the false allegations against Walters. In *Battle v. Microsoft*, Bing using GPT-4 technology [merged the results for technology expert Jeffery Battle](#) with the convicted terrorist Jeffrey Battle (different spelling) to produce an inaccurate and potentially defamatory text result. As of now, it appears that neither OpenAI nor Microsoft have raised Section 230 as a defense, [limiting the degree to which](#) these cases will serve as useful indicators of Section 230's scope in this context. Even so, the companies' decision not to assert Section 230 in these cases may signal how they anticipate the analysis would have played out, at least with regard to these facts.

It's further worth noting that, in the [oral arguments for Gonzalez v. Google](#), the US Supreme Court indicated that Generative AI system outputs may simply always be an information content provider with respect to model outputs. If this reasoning continues to hold sway with the Court, Generative AI outputs may eventually be definitively ruled [out](#)

of the sphere of Section 230 immunity.

That could mean, for example, that when an X user posts or shares an AI-generated image containing potentially illegal content on X, courts may find that X is protected from liability by Section 230, but that the Generative AI model that created the image would not be protected by Section 230.

However, it should be noted that in cases where Generative AI outputs are so close to the source material that the Generative AI model does not materially contribute to the creation of the illegal content and is acting solely in the capacity of an interactive computer service, [currently applicable case law](#) indicates those outputs could still receive Section 230 immunity. To this end, there is existing precedent in which courts have held that algorithmic recommendations or use of algorithms even in the proliferation of potentially illegal content was protected under Section 230 immunity. According to the courts this immunity also extends over instances of “automated editorial acts” or minor alterations that did not contribute materially to the offensive nature of the content.

Conclusion

As it stands, legal precedent indicates that the more “creative” outputs of Generative AI will likely fall outside the parameters of Section 230 immunity. This is because, while Generative AI does constitute an interactive computer service, when involved in the creation of new content, even in response to a user prompt, the system is more likely acting as an information content provider. However, Generative AI companies might still attempt to avail themselves of Section 230

immunity in cases where they could claim that the Generative AI system did not materially contribute in whole or in part to the creation of the illegal content. In other words, when it comes to whether Section 230 protects Generative AI systems from liability for their outputs, at least under current precedent, the answer is: it depends.

Share

Related Reading



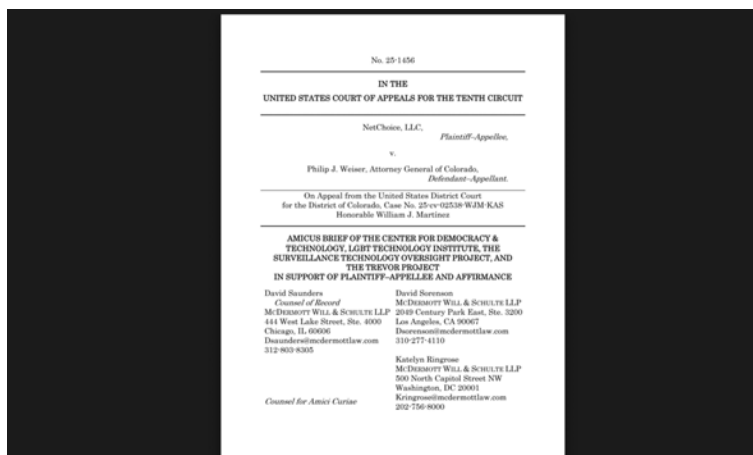
Tech Talks: Regulating AI in the States — From the Legislative Process to Implementation

July 24, 2026



Senate AI and Kids Safety Package Won't Justify Preempting State AI Laws

July 21, 2026



CDT Files Joint Amicus Brief In Opposition to Misleading Social Media Warning Labels

July 17, 2026

CDT works to strengthen individual rights and freedoms by defining, promoting, and influencing technology policy and the architecture of the internet that impacts our daily lives.

The content throughout this website that originates with CDT can be freely copied and used as long as you make no substantive changes and clearly give us credit. More on CDT's content reuse policy is available [here](#). CDT’s generative AI usage policy is available [here](#).

People	Organization	Areas of Focus
CDT Europe	2025 Annual Report	AI Policy & Governance
CDT AI Governance Lab	Financials	Cybersecurity & Standards
CDT Research	Press	Elections & Democracy
Staff	Events	Equity in Civic Tech
Board		

- CDT Advisory Council

CDT's Non-Resident Fellows

Collaborating Attorney Networks

CDT Alumni Network

Working Groups
- Donate

Careers

CDT's Tech Talk
- Free Expression

Government Surveillance

Open Internet

Privacy & Data

Privacy Policy

Contact Us

Keep in Touch



CenDemTech

f

X

in

m

🦋

CDT Europe

X

in

m

🦋

Copyright © 2026 by Center for Democracy and Technology. Created by nclud.