Insight

Section 230 in the Era of Generative AI

ANGELA LUNA | OCTOBER 16, 2025

Executive Summary

- Generative artificial intelligence (Gen AI), which creates new content such as text and images from user prompts, raises questions about whether Section 230 of the Communications Decency Act's liability protections for online platforms extend to these tools.
- Section 230 protects platforms from liability for the content users generate, but if the model itself "creates" that content, the law is unclear as to whether an AI developer or platform that offers generative AI tools can benefit from the protection.
- As the legal status of Section 230 immunity for generative AI outputs remains ambiguous, policymakers face the challenge of balancing protection for online platforms with accountability for harmful outputs; relevant legislation may focus on addressing specific, well-defined harms, while courts and lawmakers work to clarify the broader legal questions.

Introduction

Since the rise of generative artificial intelligence (Gen AI) – the type of AI capable of creating new content, such as text and images based on user prompts – there has been uncertainty among courts, policymakers, and AI developers about whether Section 230 of the Communications Decency Act (CDA), which shields online platforms from liability for third-party content, extends to these technologies.

The challenge lies mainly in the first clause of Section 230, which states that courts shall not treat the interactive computer services as the publisher or speaker of content posted by their users. The challenge for Gen AI, however, is that it isn't entirely clear who is

responsible for creating the content, meaning the model developer or platforms that integrate AI features to help users create posts might not fall under Section 230 protections. Behind a Gen AI output is a complex process that depends heavily on several factors, including the training data, the model's design and algorithms, and the user prompt. This process blurs the line of accountability among developers and users, because all, in some way, may have an impact on the resulting Gen AI output.

Policymakers remain divided over how to protect online platforms from excessive liability while ensuring accountability for AI-generated harm. Given the legal ambiguity surrounding Section 230's application to generative AI, early legislation could target specific, well-defined harms as courts and lawmakers clarify broader questions of liability.

Laying the Groundwork: Section 230 and the Rise of Generative AI

Section 230 shields interactive computer services from liability for hosting user-generated content and removing or editing third-party content from claims such as defamation and negligence. The protection, however, is limited to content provided by the user, not content that the platform materially contributes to generating.

Generative AI could destabilize Section 230 by blurring traditional lines of accountability among platforms, developers, and users. The process behind AI outputs is deeply complex and depends heavily on several factors, including the training data, the model's design and algorithms, and the user prompt. Accordingly, outputs can vary with each prompt, and sometimes include "hallucinations," meaning the model generates content not directly tied to training data or that is inaccurate, like the model is inventing something by itself. As a result, generative AI outputs are not solely user-generated, and platforms that offer AI tools to help users create and post content could lose their 230 protections.

How Generative AI Challenges Section 230

The challenge of Gen AI outputs mainly lies in the first clause of Section 230, determining whether the AI developer is the creator of the output. For example, when a Gen AI chatbot generates a harmful output, such as a defamatory statement, the output will largely depend on how the user writes the prompt, and its intention, on the data and whether the data include unlawful or biased information, and on how the algorithms determine how the model interprets the prompt to generate a response. The problem hinges on how neutral the model is when producing the output and to what extent it "materially contributes" to the alleged harmful content creation.

On the immunity side, courts have traditionally held that tools based on objective factors are neutral. Under this interpretation, the developer might be simply providing the machinery,

as the output is derived from the training data (provided by others) and the prompt (provided by the user), meaning the developer did not create the specific resulting content and should remain protected by Section 230. On the other hand, comments against immunity highlight that Gen AI has an intrinsic creative function where the output is "composed by the programs themselves," as the content is not just composed of quotations from existing sites but is rather articulated by the AI.

A clear example of this can be seen in the case of Gen AI search engines. The challenge goes beyond the kind of algorithmic filtering that traditional internet search engines and social media platforms do, where algorithms simply curate or promote third-party content to display results for the user. Generative AI, however, doesn't just sort information; it actively synthesizes data from multiple sources, decides what's relevant, and generates new text. Thus, these Gen AI search tools don't simply fall into the category of distributors of content; neither can they be treated as full authors. Thus, some argue that even if the AI relies on third-party sources, the final output is essentially the model's own speech, since it paraphrases, summarizes, and sometimes distorts the original material, making it harder to claim that it just hosts or distributes content.

Congressional Action and Regulatory Implications

The current legal status of Section 230 immunity for generative AI outputs is defined by ambiguity. Generative AI disrupts the distinction between platform and speaker and the notions of neutral tool and material contribution, and without clear guidance the uncertainty will persist. The primary legislative attempt to address Gen AI liability was the No Section 230 Immunity for AI Act, introduced in 2023 by Senators Josh Hawley (R-MO) and Richard Blumenthal (D-CT), a bipartisan bill that would have waived immunity under Section 230 for claims that involve the use or provision of Gen AI. The bill faced significant resistance, however, as critics argued that holding AI companies legally liable for their models' outputs would harm AI innovation and the United States' AI competitiveness.

Congress could also choose to set legal clarity by declaring that Gen AI models and platforms are not liable for content that users direct the models to generate. Section 230 was designed to promote speech and allow platforms to moderate content without fear of liability. A broad law that protects Gen AI developers would encourage AI development and promote speech online. Yet this could result in harmful content proliferating online. Both approaches would have led to harmful outcomes for developers and users without addressing the question of whether generative AI materially contributes to the creation of the content.

Considering the challenge for policymakers to balance protecting online platforms from

excessive liability and ensuring accountability for the harm caused by user-generated or AI-generated content, Congress could target legislation that focuses on specific harms, while working to clarify the broader legal questions (a job for the courts, as well). For example, Congress has been considering legislation to target deepfakes, and if specific generative AI outputs are found to cause additional harms, Congress could consider legislation targeting these harms directly rather than broadly exempt AI from Section 230 protections. This will allow courts to continue to apply Section 230 broadly and take a closer examination of questions about creation and responsibility of content generation in the Gen AI era.

Conclusion

The current legal status of Section 230 immunity for generative AI outputs is defined by ambiguity that Congress should work to address via clear guidance on liability and protections. Yet as the defining lines between speaker and platform are complex, policymakers could focus on addressing specific, well-defined harms with frameworks and standards that developers can follow to address risk management, while courts and policymakers work to clarify the broader legal questions.