

# A.I. and Disinformation – Who is Responsible When Chatbots Lie?

Emily M. Crawford  
Elysium Digital

What can you do when artificial intelligence (A.I.) lies about you? This exact question was asked recently in a [New York Times article](#) of the same name. In the piece, a Dutch politician describes how an A.I. chatbot, BlenderBot, had correctly described her political background but had also incorrectly described her as a terrorist. Her story is not unique. [An ongoing lawsuit for defamation against OpenAI alleges that OpenAI's ChatGPT generated a legal complaint accusing radio host Mark Walters of defrauding and embezzling funds from a gun rights group, Second Amendment Federation.](#) Though there is an actual ongoing lawsuit against the group, Walters has no ties to the group and is not involved in that lawsuit in any way.[1]

While A.I. chatbots come with the disclaimer that they may produce incorrect information, some individuals may see the service as “just another search tool.” However, [A.I. is prone to hallucinations](#) – sections of text output that may appear syntactically correct but contain false information. After all, generative A.I. is not forming full thoughts that it subsequently articulates; it is simply repeatedly trying to find the next best word to fit in its output given the prompt and the words it has already generated. [Who is liable for false content that may be produced by A.I.? The owner of offending content often has some liability, but who is the owner of A.I.-generated content?](#) [Courts have ruled](#) that because an A.I. system is not a human entity, it cannot claim copyright or be named as an inventor on a patent. Courts have further ruled that the content generated by A.I. should be in the public domain rather than assigned to either the creators of an A.I. system or the individual(s) who created the prompts that created the content by using the A.I. system. However, it's not terribly satisfying to blame the public for A.I. hallucinations even if those hallucinations are in the public domain, and legislation has not yet caught up to this issue.

Currently, there are few laws regulating A.I. or machine learning, though there have been many bills proposed in Congress regarding the subject, which build upon earlier laws that have already helped to define the modern Internet.[2] One particularly important earlier law is 47 U.S.C. § 230, also known as Section 230 of the Communications Decency Act, and it was part of a larger reform to the 1934 Communications Act that passed in 1996. Since then, Section 230 has been applied to protect online providers both from liability over the content posted by third parties on their sites, and from recourse based on good-faith content moderation decisions. Recognizing that “the Internet and other interactive computer services have flourished, to the benefit of all Americans, with a minimum of government regulation,”[3] Section 230 was designed to allow free speech and a platform for all, without companies fearing legal repercussions. [Section 230\(c\)\(1\) states that “No provider or user of an interactive computer](#)

service shall be treated as the publisher or speaker of any information provided by another information content provider.”[3] In essence, social media and other content providers cannot be held legally liable for content posted on their websites by third parties. This means if someone posts a defamatory statement about you on X, X Corp. does not have liability. Instead, the individual user who posted that statement is liable for their actions.

Historically, U.S. legal precedent has held that publishers, such as those producing magazines, books, and newspapers, are strictly liable for all the content they display. Traditionally, publishers of books or magazines curated and scrutinized their content, meaning that any defamatory statements should have been caught and either taken out or kept in purposefully, thus making the publisher liable for the statement. X, YouTube, and other online services, on the other hand, operate on a scale so vast that it is infeasible for such a company to keep up with everything posted its platform by its users. The comments sections of these websites essentially function more like a town square than a curated list of ideas. It was under this consensus that web platforms were deemed to have no liability regarding user-posted content. The owner of a park in the city would not be liable if someone stood in the park and yelled defamatory statements, and the same principle applies online.

Along with an exemption from liability for third-party (user) content on online platforms, Section 230 provides protections for content moderation, allowing websites to perform good-faith moderation of content posted by third parties on their site. Facebook, Instagram, and other social media sites all have some sort of community guidelines for what can be posted, and the platforms may take down posts that violate these rules. Under the previous American laws, this could make them liable as distributors since they are “curating” what is on their platforms, the same way bookstores or magazine stands do. Under U.S. law, if a shop owner learns they are selling a magazine with a defamatory statement and they fail to remove it within a reasonable period, they can also be held liable as a publisher.

If these sites perform some sort of curation, could they be held liable? This question was posed in the case *Stratton Oakmont, Inc. v. Prodigy Services Co.*, a precursor to Section 230. At the time, the ruling treated online services as distributors, but when Congress was debating Section 230, they disagreed with the ruling, arguing it was “backward” and that online services should not be punished for “good-faith” attempts at moderation.[4] This argument became a part of Section 230(c)(2), which says that a provider or user of an interactive computer service should not be held liable for actions taken in good faith to restrict access to content which might be obscene, lewd, or “otherwise objectionable.”[3] The early case *Zeran v. America Online, Inc.* exempted sites from both types of liability and solidified the distinctions between distributor and publisher. Since Section 230 became law, social media, news, and many other websites have enjoyed substantial legal protection on their services – free to moderate as they see fit and protected from any lawsuits regarding defamatory statements made by users. This protection makes the Internet the place we know it today, where there are few restrictions on content that individuals may create for all to see, for better or for worse.

As the Internet has evolved, search and recommendation algorithms have come into the spotlight, and it has become apparent that algorithmic recommendations that target users may be a form of curation or even “development” of third-party content. Platforms’ potential liability due to the development of third-party content has its roots in the case [Fair Housing Council v. Roommates.com, LLC](#), a decision that has been referenced in cases such as [Facebook v. Force](#), and the recent Supreme Court case [Gonzalez v. Google](#). In 2008, the Fair Housing Council sued Roommates.com for violating the Fair Housing Act and specifically highlighted a questionnaire on the site that required users to state their preferences for a roommate’s age, gender, sexual orientation, and number of children. This information was then displayed on user profiles. The Fair Housing Council argued that this violated the [Fair Housing Act](#) by displaying advertisements that indicated illegal preferences concerning protected classes. Roommates.com argued that the Fair Housing Council was trying to hold them accountable for third-party content on their site and that they were protected from liability by Section 230. In the final ruling, the Ninth Circuit sided with the Fair Housing Council, explaining that the questionnaire “induce[d] third parties to express illegal preferences,”[4] since completing the questionnaire was required to be an active user of the site. The court determined that Roommates.com may have helped “develop” this information in the questionnaire by “materially contributing to its unlawfulness.” In its decision, the court also cautioned that this same logic should not be extended to “passive conduits,” or “neutral tools” used by sites, such as search engines. The cases of [Facebook v. Force](#) and [Gonzalez v. Google](#) affirm that recommendation algorithms do not constitute “development” by online sites, finding that the algorithmic promotion of extremist content and the failure to take down such content in a reasonable manner is protected by Section 230.

While search algorithms still may be protected under Section 230, it is unclear whether this protection extends to generative A.I. Currently it depends on how one would categorize these types of algorithms. Using the language of Roommates, A.I. may be a “neutral tool” and thus still protected. Some sites have been using A.I. as an augmented search tool, fine-tuning the A.I. model and providing it with the pages on their site. The result is a chatbot that summarizes results and has links to pages within the site to verify its sources. In this usage, it could be considered a passive conduit for information and could be covered under Section 230. On the other hand, one may consider content generated by machine learning as content created by a third party. As discussed above, U.S. courts have determined that the rights to intellectual property created by A.I. do not rest with the A.I. system, the company that created it, or the user who used the A.I. system to create the IP. It remains unclear who would be legally liable for defamatory content online.

In the end, it is not certain how a Section 230 carve-out for generative A.I. would function, though A.I. as a topic has been on the minds of many representatives. Last summer, two senators proposed a bipartisan bill waiving Section 230 immunity for A.I. The bill would add an additional addendum to [Section 230 of the Communications Act of 1934](#) (47 U.S.C. § 230) in section E, such that Section 230 could not be used as protection for interactive computer service providers if the underlying suit or criminal charge involves the “use or provision of generative artificial intelligence by the interactive computer service.” Senator Josh Hawley, one sponsor of the bill, stated: “We can’t make the same

mistakes with generative AI as we did with Big Tech on Section 230,” indicating the desire for more hardline protections surrounding A.I. The other legislator co-sponsoring the bill, Senator Josh Blumenthal, “called the bill a ‘first step’ in establishing safeguards around artificial intelligence.”[5] For now, the debate will continue as lawmakers and courts determine the best way forward, and as artificial intelligence continues to advance and provide new benefits and risks.

[1] <https://news.bloomberglaw.com/ip-law/first-chatgpt-defamation-lawsuit-to-test-ais-legal-liability>

[2] <https://www.congress.gov/search?q=%7B%22source%22%3A%22legislation%22%2C%22congress%22%3A%22all%22%2C%22search%22%3A%22artificial%20intelligence%22%7D&pageSort=dateOfIntroduction%3Adesc>

[3] [https://uscode.house.gov/view.xhtml?req=\(title:47%20section:230%20edition:prelim\)](https://uscode.house.gov/view.xhtml?req=(title:47%20section:230%20edition:prelim))

[4] <https://crsreports.congress.gov/product/pdf/R/R46751>

[5] <https://www.reuters.com/technology/bipartisan-us-bill-would-end-section-230-immunity-generative-ai-2023-06-14/>