Daniel Chen

danielchen082208@gmail.com

Explain whether existing defamation and libel laws are or are not sufficient to prohibit baseless AI generated and distributed content.

The introduction of ChatGPT changed the world forever. This advanced generative AI system demonstrated just a glimpse of the immense potential this technology could bring. It would create a lasting impact in the lives of students, workers, and anyone with access to the internet. Though people could tell when text was written by ChatGPT at the time, newer models' responses became more humanlike. Today, fake text generated by ChatGPT is just one of many forms of malicious content generated to defame individuals. Although libel laws were designed to handle cases of defamation, they are not sufficient to prohibit baseless AI generated content from damaging people's reputations because there is no clear culprit and our technological age makes AI generated defamation spread too quickly for defamation policy to take action.

Libel laws were designed mainly to alleviate the effects of defamation by providing victims with legal recourse and compensation. These laws cover content that cause damage regardless if it was spoken, typed, or posted online. However, it becomes difficult for defamation and libel laws to keep up with the pace of AI generated content; they become increasingly irrelevant. For example, one idea is that platforms can use AI detection of false information and flag/disprove content before it is posted to the internet and avert any harm caused to the victim. This aligns with the principle of "notice and takedown" previously established in many jurisdictions where defamatory content is removed promptly once identified. However, this policy was not made with the internet's powers in mind, it was meant to deal with print newspapers, rumors, and forms of spreading that were able to be contained timely. Today, with billions connected to the internet, any form of digital defamation spreads like wildfire, too quick for takedown to be relevant. These falsehoods spread even faster than truthful information: an MIT study found that false news stories are 70% more likely to be retweeted than true stories.[1] Fake content can be posted and spread so quickly that the victim may not even know that their reputation was attacked until the damage is already done, rendering the protection offered by libel and defamation laws useless.

Certain jurisdictions and legal systems – such as the United States – may allow for "punitive damages" and/or criminal penalties in certain libel cases to the defendant for particularly severe attacks. These cases require a defined perpetrator, which becomes hard to pin down in the field of spreading AI generated content. Without a defined culprit, current libel and defamation laws simply cannot be applied to these types of cases. Should the person who posted the content take the blame? Or the developer for letting the AI generate this content, or the platform that hosts the AI? There are equal arguments to all sides, and therefore there is no clear entity to be punished. One might argue that because of Free Speech of the First Amendment, people shouldn't be punished for what they say. However, this just gives malicious attackers a sense of security that they can post whatever they want without any backlash: the perpetrator cannot be clearly defined, and they can use free speech to justify how they haven't broken any laws. These reckless individuals will only cause more harm as AI expands to become more realistic, lifelike, and accessible. Just like how there are laws for assaulting someone physically,

[1] Dizikes, Peter. "Study: On Twitter, false news travels faster than true stories." MIT News, Massachusetts Institute of Technology, 8 Mar. 2018, news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308. Accessed 1 Dec. 2024.

there must be laws specifically targeting baseless AI generated content designed to assault someone’s reputation.

To ensure beneficial relationships between humanity and AI, model providers must create mechanisms like watermarks to create traceability and ensure responsible creation of AI content. This would allow for easier identification of AI, similar to how parental guidelines help identify age-appropriate media. When it comes to distributions, platforms like social media must work proactively to swiftly detect and take down defamatory content, and perhaps require verification of content before it gets posted online. Finally, raising public awareness about AI is essential. Being informed on the potential risks of AI can empower individuals to better navigate the digital landscape and avoid being misled.

Historically, we’ve been able to tackle issues presented by newer technology by altering associated laws. The Clean Air Act of 1970 tackled the pollution problem caused by the then new technology of cars; as a result new passenger vehicles were 98-99% cleaner for most tailpipe pollutants compared to the 1960s.[2] Therefore, amends to control baseless AI generated content are only possible by strengthening libel and defamation laws.

[2] "Accomplishments and Successes of Reducing Air Pollution from Transportation in the United States." EPA, United States Environmental Protection Agency, www.epa.gov/transportation-air-pollution-and-climate-change/accomplishments-and-successes-reducing-air. Accessed 1 Dec. 2024.

Daniel Chen

danielchen082208@gmail.com

List of Sources:

Dizikes, Peter. "Study: On Twitter, false news travels faster than true stories." MIT News, Massachusetts Institute of Technology, 8 Mar. 2018, news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308. Accessed 1 Dec. 2024.

"Accomplishments and Successes of Reducing Air Pollution from Transportation in the United States." EPA, United States Environmental Protection Agency, https://www.epa.gov/transportation-air-pollution-and-climate-change/accomplishments-and-successes-reducing-air Accessed 1 Dec. 2024.