Center for AI Policy

AI, Liability, and Copyright

Key Takeaways:

- Section 230 blocks liability for ***hosting*** content, not for ***generating*** content. If generative AI like ChatGPT generates libel, then OpenAI is liable under existing law.
- It's not clear whether courts will include "training an AI" as a "fair use" of a book or a movie under copyright law. In the most similar prior case, for Google Books, the court seemed to base its ruling on the value that Google Books would add to society.
- Strictly enforcing copyright can help some artists get compensated, but it doesn't offer a long-term solution for unemployment caused by ongoing automation of the workforce.
- Generative AI hallucinates and discriminates roughly as often as humans. NIST's AI Risk Management Framework, if properly and vigorously applied, would go a long way toward constructively addressing these problems.
- Existing law doesn't provide a solution for the catastrophic risks posed by advanced AI, like bioweapons and automated hacking. CAIP's bill would help address these risks with a strict liability regime and mandatory safety measures.

What are the barriers to holding companies liable for harms committed by their AIs?

- Identifying the bad actor can be a challenge. For example, if a large corporate developer makes an AI with a sloppy third-party interface that doesn't screen out security threats, and then a freelance designer adds a potentially dangerous extension for the AI, and then the end user installs that plug-in and uses it to develop a bioweapon, who caused the harm? The corporate developer, the freelance designer, or the end user?
    - The answer is that all three parties are probably liable to various degrees, but proving this can be difficult.
    - CAIP's current bill would solve this problem by making the developer, the designer, and the user all jointly and severally liable.
- **Section 230** exempts the owner of an "interactive computer service" that publishes third-party content. For example, if Howard Dean claims on Reddit that Senator Romney murders clowns for fun, then Senator Romney can't sue Reddit for defamation – he could

sue Mr. Dean, but Reddit itself is safe, even if Reddit was sloppy about responding to requests to delete Mr. Dean's accusation of clown-murder.

  - This protection is valued by the Electronic Frontier Foundation and the tech companies who rely on it for their advertising revenue.
  - The Supreme Court has interpreted exceptions to Section 230 narrowly. For example, it approved rulings finding that Twitter did not violate the Anti-Terrorism Act by passively hosting conversations in which terrorist leaders recruited new agents, and that Reddit did not knowingly benefit from child pornography by passively hosting bulletin boards on which third parties posted illegal videos.
- Senators Hawley (R-MO) and Blumenthal (D-CT) are sponsoring a bill that would deny Section 230 immunity for charges and claims related to generative AI.
  - Our understanding is that existing law already makes it such that there is no immunity for generative AI (because generating your own text is unlikely to qualify as passively hosting third-party content) but the Hawley-Blumenthal bill attempts to eliminate any possible ambiguity.

Do foundational AI models violate copyright?

- Most texts and images are copyrighted. They may come with licenses that allow certain types of users to use them in certain ways, but these licenses almost never grant permission for the content to be scraped and used in training runs by AI developers. A work that is valuable enough for its owners to protect usually remains under copyright for literally 95 years – this means that books published in the 1920s are only just now entering the public domain.
- Making use of a copyrighted work without permission carries fines of up to $30,000 per work, or $150,000 per work if the infringement was "willful." If the owners of the work lost profits that are larger than these fines, they can recover their profits, instead.
- A "fair use" of copyrighted material does not violate copyright law. Literary critics, journalists, teachers, comedians, and researchers are usually allowed to cite medium-sized portions of copyrighted works for the purpose of commenting on them without having to get permission or pay royalties.
  - A court is more likely to find that a work is a "fair use" if it is a "transformative use," i.e., if the work adds something new, of a different purpose or character, and thus does not directly substitute or compete with the original work.
  - The question of whether, e.g., Hugging Face is making fair use of the images it was trained on is currently being litigated; there is genuine uncertainty about how courts will rule.
  - The comedian Sarah Silverman is spearheading a similar lawsuit against Meta and OpenAI for training their LLMs on illegally downloaded copies of her copyrighted books. She argues that these models would not be able to summarize her books if

they had not 'read' them, and that they did not pay her (or other authors) for the right to read them.
    - The most important similar lawsuit was Author's Guild vs. Google Books; a consortium of authors and publishers sued Google for violating their copyrights by scanning books in order to put them into a searchable database and offer long snippets from each book. The parties tentatively agreed on a settlement that would have had Google paying modest per-book fees and allowed authors to affirmatively opt-out of the database, but then abandoned the settlement; Google ultimately won a complete victory in court. It appears that the Second Circuit (which heard the case) was more interested in the public utility provided by having a searchable database of nearly all books than in the literal application of existing copyright law.

When should AI companies be sued for copyright violations?

- Weak enforcement of copyright will make it easier to develop profitable AI businesses, and easier for consumers to access low-priced entertainment, but it will also contribute to putting authors, actors, and musicians out of work and exacerbate economic inequality.
    - For example, several actors are alleging that they were paid $100 to work as an extra for the day on a movie set – the usual rate – and then a director took detailed digital scans of them that would enable movie producers to keep on using the actor's likeness in future movies, all without any additional pay.
    - Meanwhile, studios are offering up to $900,000 to hire AI technicians.
    - Entertainers already have very low bargaining power – for example, Spotify offers musicians only a fraction of a cent each time a song is played; to earn even **minimum wage** from Spotify, a typical artist would need to have 400,000 streams per month. This is roughly the equivalent of releasing a new gold record album every single year.
- Stronger copyright enforcement will make it somewhat harder for AI entertainment to succeed.
    - Transaction costs will slow down the production of AI-assisted artwork; it's challenging to negotiate with each person whose art you want to use, and you might not be able to find them or might not know who owns their catalog.
    - Even if most artists have the right to say 'no' to AI scanning and scraping (by enforcing their copyright), if a few artists say 'yes', then there could still be mass unemployment in the field. It's plausible that given a sample of a few hundred scanned actors, AI will be able to extrapolate enough digital images to cover any situations that might come up in TV and movies. The most passionate fans might pay a premium to watch likenesses of their favorite actors, but most consumers will probably be happy enough with high-quality (but generic) AI actors, plus a few familiar faces from the actors who agreed to be scanned.

When should AI companies be liable for hallucinations?

- The current generation of chatbots often "hallucinate," i.e., invent detailed and confident-sounding claims that have no basis in reality. The FTC is investigating OpenAI for such hallucinations right now.
- This is mostly a problem if people don't realize that it's a possibility. At some point, blindly trusting something because you heard it from a chatbot is like trusting something because you read it in the *National Enquirer* or *The Onion*.
  - If you open ChatGPT, you get a screen that prominently warns you that one of its "limitations" is that it "may occasionally generate incorrect information." At the bottom of the screen is a disclaimer that "ChatGPT may produce inaccurate information about people, places, or facts." OpenAI's advertising campaigns have mostly been transparent about what ChatGPT can and can't do and what its limitations are.
  - Other chatbots may be less scrupulous about their disclaimers.
- Banning any chatbot that has any tendency to hallucinate would destroy significant economic value – ChatGPT is "right" or at least "useful" often enough that if you double-check its suggestions, you'll usually still do your work faster and better than if you tried to work without hearing those suggestions at all.
- To balance the usefulness of chatbots with their tendency to hallucinate, it makes sense to require chatbots to offer strong disclaimers, and then hold them harmless for hallucinations if and only if the disclaimers are adequate.

When should AI companies be liable for discrimination and bias?

- Discrimination based on protected categories is an ongoing problem that's mostly highlighted by AI rather than caused by AI. As a result, efforts to address underlying causes of discrimination will also help reduce discrimination caused by AI. For the most part, AIs that make biased decisions are doing so because they were trained on data that reflects the biases of average Americans.
- Some AI-based discrimination reflects poor design choices on the part of engineers. For example, default camera settings tend to be suboptimal for photographing darker skin tones. This can lead to racial bias in AI-powered facial recognition technology.
  - Companies should be legally required to seek out and address these types of problems.
  - The NIST AI Risk Management Framework ("AI RMF") offers tools and resources for companies that voluntarily choose to pursue these internal investigations.
- Algorithms are often evaluated for bias based on the "four-fifths rule," meaning that if an algorithm is selecting only four-fifths as many people from a protected group, then it is biased. This is a crude rule of thumb; as algorithms become increasingly important in daily life, we should look for ways to make more precise assessments. For example, we

can ask whether the difference in selection rates is statistically significant using a chi-squared test, and then if there is a statistically significant difference, we can ask whether there is a legitimate reason for this difference that is not due to discrimination and that adequately explains the difference.

- Companies who develop AI have the same legal obligations to avoid discrimination as any landlord under the Fair Housing Act, any employer under the Civil Rights Act of 1964, and so on. These laws can adequately protect against most types of discrimination from AI as long as we apply them thoughtfully.
- AI can sometimes hide or 'launder' an unlawful or immoral bias by making a decision appear to be objective.
  - For example, an AI-generated "sentencing guideline" that's calculated based on isolated facts about a criminal defendant may seem more impartial than a sentence ordered by a particular judge to that defendant's face after seeing and talking with the defendant.
  - It is important to educate people to assess the actual bias present in both human and AI-assisted decision-making systems. People should be taught not to assume that any decision generated by a computer must be impartial. This could involve mandatory training for the end users of AI recommendation engines, such as judges and parole officers.

What about liability for catastrophic risks?

- AI is becoming increasingly powerful. Very soon, it could gain capabilities that could cause destruction on a mass scale, such as designing new bioweapons, automatically seizing control of millions of computers, or manufacturing its own drones and missiles.
- In the past, most software was only dangerous to people who voluntarily chose to use it. Even a computer virus typically cannot affect people who do not install malware or click on a suspicious link.
  - As a result, we have almost no laws that regulate dangerous software.
  - It would be very difficult to hold people properly accountable for creating or distributing dangerous AI.
- The Center for AI Policy's current bill would address this gap by establishing strict liability for catastrophic risks. The bill also aims to reduce these extreme risks by monitoring concentrations of the special chips needed to power advanced AI, requiring safety precautions for advanced AI developers, and allowing for swift intervention in the case of an emergency.