A Legal Risk Taxonomy for Generative Artificial Intelligence

David Atkinson   **Jacob Morrison**

Allen Institute for Artificial Intelligence

Abstract

For the first time, this paper presents a taxonomy of legal risks associated with generative AI (GenAI) by breaking down complex legal concepts to provide a common understanding of potential legal challenges for developing and deploying GenAI models. The methodology is based on (1) examining the legal claims that have been filed in existing lawsuits and (2) evaluating the reasonably foreseeable legal claims that may be filed in future lawsuits. First, we identified 22 lawsuits against prominent GenAI entities and tallied the claims of each lawsuit. From there, we identified seven claims that are cited at least four times across these lawsuits as the most likely claims for future GenAI lawsuits. For each of these seven claims, we describe the elements of the claim (what the plaintiff must prove to prevail) and provide an example of how it may apply to GenAI. Next, we identified 30 other potential claims that we consider to be more speculative, because they have been included in fewer than four lawsuits or have yet to be filed. We further separated those 30 claims into 19 that are most likely to be made in relation to pre-deployment of GenAI models and 11 that are more likely to be made in connection with post-deployment of GenAI models since the legal risks will vary between entities that create versus deploy them. For each of these claims, we describe the elements of the claim and the potential remedies that plaintiffs may seek to help entities determine their legal risks in developing or deploying GenAI. Lastly, we close the paper by noting the novelty of GenAI technology and propose some applications for the paper's taxonomy in driving further research.

1 Introduction

Generative artificial intelligence (GenAI)[1] is a form of artificial intelligence (AI) that can create new outputs, such as new text, images, or audio. GenAI models have expanded from single-modality (e.g., text or image-based)[2] to multi-modal (e.g., generating both images and text, and possibly audio, code, and video)[3]. More importantly, GenAI models are being used in many different contexts spanning medical application (Singhal et al., 2023) to companionship[4], legal review[5], and virtually every form of creative work[6].

The views and opinions expressed in this paper are those of the authors and do not necessarily reflect the official policy or position of their employers.

[1]GenAI is not to be confused with artificial general intelligence (AGI), which is not-yet-existent broad human-level intelligence.

[2]e.g., ChatGPT, Stable Diffusion

[3]e.g., Gemini

[4]e.g., Replika, PI

[5]e.g., Harvey

[6]e.g., Firefly, MusicLM

The benefits of GenAI are not without their risks, however. Researchers have already identified a slew of potential risks and categorized them into groups and hierarchies called taxonomies, such as ethical (Weidinger et al., 2022), sociotechnical (Weidinger et al., 2023), code generating (Khlaaf et al., 2022), privacy [7], and security (Gupta et al., 2023; Derner et al., 2023) risks. There are also specific taxonomies of risks for text-to-image (Bird et al., 2023), audio generation (Barnett, 2023), and other modality-specific use cases.

Most, if not all, of the taxonomies cite actual harms, as opposed to merely theorizing about potential harms. There are also databases that keep a running catalog of real-world harms caused by GenAI[8] [9].

Notably missing, however, is a taxonomy focused on the legal risks associated with developing and deploying GenAI despite nearly two dozen lawsuits against the likes of Meta, Anthropic, OpenAI, and Alphabet with claims that vary widely in legal theory, such as copyright, torts, contracts, unfair competition, trademark dilution, and more.

A legal taxonomy is important to distinguish from others given the particular risks and harms that apply in the legal context. For example, GenAI that hypersexualizes women(Wolfe et al., 2023), hallucinates(Xu et al., 2024), or spouts harmful disinformation(Chen & Shu, 2023) may have tremendous ethical implications that may cause significant reputational damage to the developers and deployers of those systems. However, it's unlikely such ethical issues will cause the type of existential harm that losing a class action lawsuit might. Indeed, many prominent GenAI models have already produced most ethical harms known to be practically possible (not merely theoretical) and yet the models and their creators still exist and, in many cases, still generate more revenue or achieve a higher market cap each month than most businesses.

This paper provides the first comprehensive overview of GenAI legal risks[10]. Definitions for legal terminology used throughout this paper are included in Table 1.

2 Scope

There are (unfortunately) dozens of potential legal claims that could conceivably be brought against GenAI developers and deployers depending on the particular facts of the situation, and how adventurous and creative the plaintiffs would like to be in bringing lawsuits. Indeed, because GenAI is so new, the body of case law applied to it is vanishingly small. For the most part, it remains unsettled as to how existing statutes and common law will apply to what may be a sufficiently unique form of technology to merit review de novo. Some of the claims below may ultimately be dismissed in most instances, but we include them in this paper for completeness. In short, the actual legal risk for certain common claims is unknown.

Though the nature of novel legal theories is no doubt riveting, this paper only aspires to discuss at some length those claims that have already been brought in federal court against prominent GenAI entities, and are therefore probably the most likely claims to be brought in future litigation against other GenAI entities over at least the next few years. We provide a more cursory overview of less likely or less common claims.

3 The Most Litigated Claims

This section will discuss the claims that are most often cited in existing GenAI cases based on a simple methodology: we tallied the frequency of each claim that appears in 22 lawsuits against prominent GenAI entities[11]. Note that frequency does not necessarily equate to strength. Just because a certain legal claim occurs more frequently than another does not mean the claim is more likely to succeed. However, frequency is a good proxy for the types of arguments that plaintiffs believe have at least a reasonable chance of success. We identified seven claims that are cited most frequently

---

[7]Shaping the future: A dynamic taxonomy for AI privacy risks: https://iapp.org/news/a/shaping-the-future-a-dynamic-taxonomy-for-ai-privacy-risks/

[8]AI Incident Database

[9]AIM: The OECD AI Incidents Monitor, an evidence base for trustworthy AI

[10]To be clear, this paper is not legal advice. It is an overview from a legal perspective. If you need legal advice, you should hire a lawyer.

[11]These are the only lawsuits based on GenAI claims the authors are aware of. There may be more.

Term	Definition
Case law	Case law, also referred to as common law, is a law that is based on precedents set by courts, rather than law as written in constitutions, statutes, or regulations.
Civil damages	Civil damages (or damages) are monetary awards that the plaintiff can seek from the defendant in a civil lawsuit. The amount and type of damages depend on the nature and extent of the harm, and the applicable law.
Civil law	Civil law includes actions in which the parties bringing the suits (the plaintiffs) are seeking to enforce private obligations or duties against the other parties (the defendants).
Criminal law	Criminal law is based on written statutes by which a state or the federal government prohibits specific kinds of conduct. The statutes allow the government to impose fines or imprisonment on persons convicted of violating them. Criminal cases are always brought by the government whose law has allegedly been violated.
Criminal penalties	These are fines or imprisonment that the government can impose on the defendant for violating a criminal statute. The penalties under these laws can vary depending on the intent, harm, and prior convictions of the defendant.
Common law	Unlike criminal law, which is based on statutes, common law is the body of law created by judges in written opinions (the rulings of the court). Common law courts look to the past decisions of courts to synthesize the legal principles of past cases.
Compensatory damages	These damages aim to restore the plaintiff to the same situation he or she was in before the injury occurred. Compensatory damages can cover both tangible and intangible losses, such as medical expenses, lost wages, property damage, pain and suffering, emotional distress, and loss of enjoyment of life.
Defendant	An individual, company, or institution sued or accused in a court of law.
de novo	*De novo is a Latin term that means "anew," "from the beginning," or "afresh." When a court hears a case "de novo," it is deciding the issues without reference to any legal conclusion or assumption made by a previous court.*
Element/Factor	An element (sometimes called a factor) is an essential requirement to make a claim or defense in court.
Injunction	Injunctions require the person to stop doing something or to perform a specific action to remedy the harm. Injunctions may be temporary or permanent, depending on the circumstances of the case.
Liability	The state of being responsible for something
Nominal damages	These damages recognize that the plaintiff was legally wronged, but did not suffer any actual harm.
Remedy	A form of court enforcement of a legal right resulting from a successful civil lawsuit that includes awarding damages or ordering an injunction.
Statute	A written law passed by a legislative body. Often just called a law.
Plaintiff	A person who brings a case against another in a court of law.
Private right of action	Allows an individual or organization (rather than just the government) to bring a lawsuit in court based on an alleged violation of a law and to seek a remedy.
Prosecution / Prosecutor	The legal party responsible for presenting the case in a criminal trial against the defendant, an individual accused of breaking the law.
Punitive damages	Aims to punish the defendant for his or her malicious, reckless, or fraudulent conduct, and to deter others from engaging in similar behavior.

Table 1: Glossary of terms and definitions related to GenAI legal risks.

across these lawsuits as further discussed below: direct copyright infringement, vicarious copyright infringement, contributory copyright infringement, DMCA 1202(b), unjust enrichment, negligence, and unfair competition.

3.1 Copyright Infringement

3.1.1 Direct Copyright Infringement

Direct copyright infringement is when a person, without authorization (e.g., in the absence of a license or other agreement between the parties), reproduces, distributes, displays, or performs a copyrighted work, or prepares a derivative work based on a copyrighted work as these are exclusive rights of the copyright owner[12]. Notably, copyright infringement applies strict liability, meaning it does not require knowledge of infringement or intent to infringe by the defendant. In other words, the plaintiff does not need to prove that the defendant knowingly or intentionally made unauthorized use of the plaintiff's copyrighted work in order to prevail.

How does one know if there was infringement? Courts use a standard called "substantial similarity," which is a way of comparing the alleged infringing material to the original to decide whether they are too similar[13]. The mechanics of how courts conduct the assessment varies by jurisdiction, but the underlying principle of comparing one version to another remains unchanged. One might wonder why such a test is required, and the reason is because if only exact copies of the original counted as infringement, people would need to only make the slightest alteration to escape a copyright claim.

Infringement is not the end of the story, however. In some circumstances, the unauthorized use of copyrighted works is considered fair use. Fair use is a legal doctrine that permits the use of copyrighted works without authorization from the copyright owner and exempts one from liability for copyright infringement using a four-part test to determine whether something qualifies as fair use, as summarized below.

Section 107 of the Copyright Act[14] identifies four factors to consider when assessing whether the use of copyrighted material is infringement or fair use:

1. The purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;

2. The nature of the copyrighted work;

3. The amount and substantiality of the portion used in relation to the copyrighted work as a whole; and

4. The effect of the use upon the potential market for or value of the copyrighted work.

The fair use test requires an assessment of all the factors together. The courts have repeatedly emphasized that there are no bright line rules, and that each case must be decided on its own facts[15].

Direct copyright infringement could apply to GenAI at several steps of the GenAI development and deployment pipeline, including during data collection, model training, when the model is released, and at the point when outputs are generated based on a prompt. Most legal scholars seem to believe that inputs to the model (i.e., the training materials and training process) are probably fair use, while outputs tend to be less defensible[16].

The penalties for direct infringement can include statutory damages ranging from $750 to $30,000 per work, or up to $150,000 per work for willful infringement, as well as actual damages and profits, injunctions, and attorney's fees[17].

---

[12]17 U.S. Code § 106

[13]https://plus.lexis.com/api/permalink/7ace1710-76cc-4b8a-b375-62681d1d78c9/?context=1530671

[14]17 U.S. Code § 107 - Limitations on exclusive rights: Fair use

[15]As of the publication of this paper there is no case law on the application of fair use to GenAI specifically.

[16]See, e.g., https://texaslawreview.org/fair-learning/

[17]https://www.copyright.gov/title17/chapter5.pdf

3.1.2 Vicarious Copyright Infringement

Vicarious infringement is a form of secondary liability for copyright infringement created by common law. It allows for Party A to be found liable for the infringing acts of Party B if (i) Party A had the the right and ability to control the infringing activity and (ii) Party A had a direct financial interest in the infringement. Importantly there must have been direct infringement by Party B for Party A to possibly face secondary infringement.

This form of liability may apply to GenAI entities if they, for example, host a model that is producing infringing content and they make people pay to use the model. By hosting the model, the GenAI entity probably has the ability to apply guardrails that would prevent infringing outputs, possibly satisfying the first prong, while the payment they receive may satisfy the second prong.

The penalties for vicarious infringement can include statutory damages ranging from $750 to $30,000 per work, or up to $150,000 per work for willful infringement, as well as actual damages and profits, injunctions, and attorney's fees[18].

3.1.3 Contributory Copyright Infringement

Contributory copyright infringement is the second type of secondary liability commonly claimed in ongoing litigation. A successful claim requires that (i) Party A makes a material contribution to the infringing activity, while (ii) having knowledge or a reason to know of the direct infringement by Party B. Making a material contribution could include knowingly inducing Party B to directly infringe, for example[19].

Contributory infringement may apply to GenAI if someone hosts a model that they know can produce infringing content (e.g., images of Nintendo characters) and someone tells the model host that such infringing content is being produced by their model, but the host takes no action to mitigate the infringement.

The penalties for contributory infringement can include statutory damages ranging from $750 to $30,000 per work, or up to $150,000 per work for willful infringement, as well as actual damages and profits, injunctions, and attorney's fees[20].

3.1.4 Digital Millennium Copyright Act 1202(b)

Section 1202(b) of the Digital Millennium Copyright Act (DMCA) was meant to prohibit people from removing or altering copyright management information (sometimes called CMI, which includes information like the creator's name and title of the work). It requires the intent to remove the CMI, distributing CMI knowing that the CMI has been removed or altered without authorization, or distributing copies of works knowing the CMI has been removed or altered, provided that the person knew or had reason to know, that such removal or alteration of CMI would "induce, enable, facilitate, or conceal an infringement..."[21] Because this law requires knowledge or intent of both the removal or alteration of CMI and that such removal or alteration would lead to infringement, it, like most laws that require knowledge or intent, is difficult to prove.

Section 1202(b) could apply where an entity collects copyrighted works, then deletes the CMI prior to training a model, and then generates infringing outputs that don't contain the CMI. However, the case law on this matter seems to require that the output be an exact copy of the original work, which is another high bar to meet[22].

The penalties for violating the section 1202 of the DMCA can include the actual damages and any additional profits of the violator, or at any time before final judgment is entered, a complaining party may elect to recover an award of statutory damages for each violation of section 1202 in the sum of not less than $2,500 or more than $25,000[23].

---

[18]https://www.copyright.gov/title17/chapter5.pdf

[19]https://plus.lexis.com/api/permalink/80c8265b-c026-4196-8dbb-9091b8646790/?context=1530671

[20]https://www.copyright.gov/title17/chapter5.pdf

[21]17 U.S. Code § 1202 - Integrity of copyright management information

[22]https://storage.courtlistener.com/recap/gov.uscourts.cand.403220/gov.uscourts.cand.403220.195.0_1.pdf

[23]17 U.S. Code § 1203 - Civil remedies

3.2 Unjust Enrichment

Unjust enrichment is common law and as such is subject to different requirements in different states. Generally, it occurs when Party A is in some agreement with Party B. Party A confers a benefit on Party B, but Party B does not comply with the terms of the agreement yet still retains the benefit conferred by Party A[24]. Some plaintiffs try to broaden the scope of this claim, though, to situations where there is no agreement, such that Party B receives a benefit from Party A without Party A's assent or consent.

The second scenario above is the one that will most likely apply to GenAI. For example, if a GenAI developer uses Party A's books as training material to create the GenAI without Party A's consent, and then the GenAI developer makes money from hosting the model, Party A may claim that revenue was the ill-gotten unjust enrichment.

There are two types of remedies for unjust enrichment: personal remedies and proprietary remedies.

1. A *personal* remedy is when the court orders the unjustly enriched party to pay the monetary value of the benefit they received to the other party. This is also called restitution, and it is the most common remedy for unjust enrichment[25].

2. A *proprietary* remedy is when the court orders the unjustly enriched party to return the specific property or asset that they gained to the other party. This is also called disgorgement, and it is usually granted when the property or asset is unique or irreplaceable[26].

3.3 Negligence

Negligence is a tort, which is a body of law that fills in the void left by contract law and criminal law where someone is harmed and someone else must be liable even if the action that caused the harm wasn't in breach of a contract or in violation of a criminal law. Negligence claims must satisfy four elements[27]:

1. the liable party must have owed a duty to act as a reasonably prudent person would under the circumstances,

2. the liable party breached that duty by acting unreasonably,

3. the breach of duty was the cause of the harm to the injured party, and

4. the injured party was in fact injured[28].

Application of negligence to GenAI can vary widely. Almost anytime someone feels injured, they can work their way back to blaming someone for the harm. One negligence claim, for example, could hinge on whether the GenAI developers acted reasonably when curating its training dataset to not collect and use third-party trade secrets. Another claim could center on whether reasonable precautions and guardrails were put in place to prevent or mitigate the model from providing harmful information to someone who then uses that information to cause harm.

The prevailing party can receive civil damages. Those damages can fall into one of the following:

1. *Nominal damages*: These are symbolic awards that recognize that the plaintiff was legally wronged, but did not suffer any actual harm. Nominal damages are rare in negligence cases, as negligence usually requires proof of injury.

2. *Compensatory damages*: These are awards that aim to restore the plaintiff to the same situation he or she was in before the negligence occurred. Compensatory damages can cover both tangible and intangible losses, such as medical expenses, lost wages, property damage, pain and suffering, emotional distress, and loss of enjoyment of life.

3. *Punitive damages*: These are awards that aim to punish the defendant for his or her malicious, reckless, or fraudulent conduct, and to deter others from engaging in similar behavior.

---

[24] https://www.law.cornell.edu/wex/unjust_enrichment
[25] https://content.next.westlaw.com/practical-law/document/Ib9aa1b161c9a11e38578f7ccc38dcbee/Restitution
[26] https://www.law.cornell.edu/wex/disgorgement
[27] https://www.law.cornell.edu/wex/damages
[28] https://www.law.cornell.edu/wex/negligence

Punitive damages are not available in all jurisdictions, and are subject to certain limits and standards.

3.4 Unfair Competition

Most claims of unfair competition are based on California law, which defines it as "any unlawful, unfair or fraudulent business act or practice and unfair, deceptive, untrue or misleading advertising..."[29] As one might expect, the statute is not precise enough on its own to make definitive conclusions about claims, so litigants must rely on case law to either bolster or distinguish their claim(s).

This claim may apply to GenAI if, for instance, the GenAI developer promised a content creator that any data scraped from the content creator's site would only be used for internal research purposes, but then subsequently used the material to offer a revenue-generating product to the public.

Generally, the remedies for unfair competition can be classified into two categories: monetary relief and injunctive relief[30].

1. *Monetary* relief is when the court orders the violator to pay compensation to the injured party or to the government. This may include actual damages, lost profits, statutory damages, punitive damages, and attorney's fees and costs.

2. *Injunctive* relief is when the court orders the violator to stop or refrain from the unlawful conduct, or to take certain actions to correct or prevent the harm. This may include cease and desist orders, corrective advertising, product recall, disgorgement of profits, and compliance monitoring.

3.5 Mitigation Strategies

The following non-comprehensive table identifies several ways GenAI developers and deployers can mitigate their legal risks pertaining to most of the possible claims discussed above.

Pre-Deployment/Release

1. Use training material that:
   (a) Was collected by others
   (b) Was collected with consent
   (c) Is not from sites you should know harbor illegally-obtained datasets
   (d) Doesn't include "unpublished" works[31]
   (e) Is public domain
   (f) Is properly-licensed[32]
   (g) Was gathered by respecting the instructions of robots.txt, and/or
   (h) Doesn't use the deep web or dark web for data

2. Filter the training data to remove material that might encode harmful behaviors.

3. Remove duplicates from the training set to reduce the likelihood of the model memorizing any given portion of training material

4. When releasing datasets, release in non-human-readable formats (preferably in formats that are irreversible to human-readable without great effort and technical know-how).

5. Remove PII/personal data.

6. Train the model(s) with strong levels of differential privacy, which not only protects PII, but also makes it so virtually nobody can distinguish, with high probability, between a model trained with a particular training example and one trained without it (Neel & Chang, 2024).

7. Conduct extensive red-teaming to try to coax potential illegal outputs from the model prior to publicly releasing it, then develop ways to minimize such occurrences in the future.

---

[29]Cal. Bus. & Prof. Code §§ 17200, et seq. (the "UCL")
[30]https://www.findlaw.com/smallbusiness/business-laws-and-regulations/unfair-competition-.html
[31]https://www.oyez.org/cases/1984/83-1632
[32]Beware: https://spectrum.ieee.org/data-ai

8. Use reinforcement learning from human feedback (RLHF) to reduce the probability of copyright infringement, right of publicity violations, torts, and privacy-violating outputs.

9. Invest in new mechanisms that make it difficult to fine-tune models for harmful uses, (e.g., "self destructing models" (Henderson et al., 2023))

10. Don't release or deploy models that pose a significant risk of violating the law.

11. For text-to-image, remove non-public domain artist names associated with works so model users can't generate outputs "in the style of" those artists.

12. For text-to-image, replace copyrighted character names with more generic descriptors.

13. Alternatively, change user prompts via prompt transformation (also called shadow prompting; covertly modifying the prompt on the back end) from specific artist or character names to more generic ones.

Post-Deployment/Release

1. Release model(s) with responsible AI license (e.g., ImpACT or RAIL)

2. Provide instance attribution if feasible.

3. Compare model output to training set data to filter and block generations that implicate copyright or privacy.

4. For live, deployed systems, create an easy way for users to report potentially harmful prompts or generations, and then continually update the prompt- and output-filtering tools

5. Detect and filter out harmful and sensitive data that users provide.

6. Implement watermarking or metadata tags on outputs.

7. Limit access to research purposes if applicable.

8. Allow a third-party audit and act on the recommendations (or record why action was not taken).

3.6 Open Questions

For the sake of brevity, this paper will not discuss open questions at length. However, we do intend to host a webpage where we will explore unsettled legal questions in greater detail. What follows is merely a sampling of the types of legal questions around GenAI that we believe deserve greater attention.

1. Does the provenance of data matter (e.g., should data intended to be publicly available be treated the same as data that is publicly available but wasn't intended to be by the data creator?

2. How important is following robots.txt?

3. How important, if at all, is allowing people to opt-out of datasets? Should this be balanced against the cost or technical burden of complying with an opt-out request?

4. Must opt-out requests always be respected? If so, is it retroactive, or only for current datasets and future models?

5. Do research organizations have to respect opt-out requests if doing so will make the research less reproducible?

6. How should licenses such as MIT or Apache, designed for code, apply to model weights?

7. What is the threshold for sufficient information security to protect datasets and user inputs?

8. Does the frequency of infringing outputs matter, or is a single output sufficient for infringement?

9. Does it matter how many times a model must be prompted before it makes an infringing output?

10. Should there be regulatory safe harbors or other legal incentives to promote transparency and safety?

11. Is there a "reasonable effort" threshold/safe harbor for entities making good faith efforts to prevent offense, illicit, and infringing outputs?

4 Other Risks: Pre-deployment

These claims either haven't been made in GenAI litigation, so they're more speculative, or they occur infrequently. Given their less prominent nature, this section will give them a more cursory overview. Here, we only provide a brief description of what the claim entails and the potential penalties if found liable or guilty so researchers can weigh the risks.

4.1 Copyright

4.1.1 DMCA 1201-1205

Section 1202 of the Digital Millennium Copyright Act (DMCA) makes it unlawful to provide or distribute false copyright management information (CMI) with the intent to induce or conceal infringement. CMI is certain information, including the title, name of the author and copyright owner, and terms for use of the work, conveyed in connection with copies, phonorecords, performances, or displays of a work[33].

The actual damages and any additional profits of the violator, or at any time before final judgment is entered, a complaining party may elect to recover an award of statutory damages for each violation of section 1202 in the sum of not less than $2,500 or more than $25,000[34].

4.2 Contract

4.2.1 Breach of Licenses

Most licenses, including permissive ones such as Apache 2.0, include requirements to: provide a copy of the license with any copies or derivatives; cause any modified files to carry prominent notices stating that the user changed the files; and the user must retain, in the source form of any derivative works that they distribute, all copyright, patent, trademark, and attribution notices from the source form of the work, excluding those notices that do not pertain to any part of the derivative works. If creators don't adhere to license terms yet use the code to train their model and provide code outputs, then they may be in breach of the license, which means both breach of contract and copyright infringement, to the extent copyright applies.

Potential penalties could include:

- If an end user breaches a material provision of the End User License Agreement (EULA), the software publisher may have the right to terminate the agreement.
- If the end user's infringement causes financial harm to the software publisher or other parties, it could result in monetary penalties.
- In specific situations, EULA breaches may be regarded as criminal offenses and subject to criminal penalties such as fines or imprisonment.
- For willful infringement, statutory damages can be awarded up to $150,000 per work infringed in copyright cases.

4.2.2 Breach of Terms of Service

Many websites have a Terms of Service or Terms of Use agreement and those terms may include prohibitions on how the site data can be accessed and/or used. There are several arguments for and against allowing liability for web scrapers based on only browsewrap and the law is unsettled when it comes to using scraped data to train LLMs.

The potential penalties are similar to breach of terms:

- If you breach the ToS, the website operator may terminate your access to their services.
- Violating the ToS might lead to civil lawsuits. For instance, if your actions cause financial harm to the website or other users, you could be held liable for damages.
- There have been cases where violating a website's ToS was treated as a criminal offense.

---

[33] 17 U.S. Code Chapter 12 - Copyright Protection and Management Systems
[34] 17 U.S. Code § 1203 - Civil Remedies

4.3 Torts

4.3.1 Intrusion Upon Seclusion

The laws regarding intrusion upon seclusion were established to protect any aspect of the plaintiff's life that he can reasonably expect will not be intruded upon. It's one of the four privacy torts created under U.S. common law.

The elements are generally: One who intentionally intrudes, physically or otherwise, upon the solitude or seclusion of another or his private affairs or concerns, is subject to liability to the other for invasion of his privacy, if the intrusion would be highly offensive to a reasonable person[35].

Civil penalties for intrusion upon seclusion can include compensatory and punitive damages. Nominal damages are rare in intrusion cases, as they usually require proof of injury.

In some jurisdictions intrusion upon seclusion can also be considered a criminal offense under the laws of trespassing, voyeurism, stalking, harassment, or cybercrime.

4.3.2 Tortious Interference with a Prospective Economic Advantage

Tortious interference with a prospective economic relationship occurs when a party improperly prevents two others from entering a contract or otherwise doing business together. Interference with prospective advantage does not require proof of a legally binding contract.

The elements are generally:

1. economic relationship between the plaintiff and some third party, with the probability of future economic benefit to the plaintiff,
2. the defendant's knowledge of the relationship,
3. intentional acts on the part of the defendant designed to disrupt the relationship,
4. actual disruption of the relationship, and
5. economic harm to the plaintiff proximately caused by the acts of the defendant.

Penalties can include civil damages. In some jurisdictions, such as Texas, tortious interference with a prospective economic relationship can additionally be considered a criminal offense under the laws of theft, fraud, or misappropriation.

4.3.3 Tortious Interference with a Contract

This is similar the the claim above, but the elements are slightly different, requiring:

1. the existence of a valid and enforceable contract between plaintiff and another,
2. defendant's awareness of the contractual relationship,
3. defendant's intentional and unjustified inducement of a breach of the contract,
4. a subsequent breach by the other caused by defendant's wrongful conduct, and
5. damages[36].

Penalties can include civil damages.

4.3.4 Civil Conspiracy

A civil conspiracy is a form of conspiracy involving an agreement between two or more parties to deprive a third party of legal rights or deceive a third party to obtain an illegal objective. A form of collusion, a conspiracy may also refer to a group of people who make an agreement to form a partnership in which each member becomes the agent or partner of every other member and engage in planning or agreeing to commit some act. It is not necessary that the conspirators be involved in all stages of planning or be aware of all details. Any voluntary agreement and some overt act by

---

[35]https://cyber.harvard.edu/privacy/Privacy_R2d_Torts_Sections.htm

[36]See, e.g., https://www.law.cornell.edu/wex/intentional_interference_with_contractual_relations

one conspirator in furtherance of the plan are the main elements necessary to prove a conspiracy. A conspiracy may exist whether legal means are used to accomplish illegal results, or illegal means used to accomplish something legal.

The elements are:

1. a combination of two or more persons,

2. the persons seek to accomplish an object or course of action,

3. the persons reach a meeting of the minds on the object or course of action,

4. one or more unlawful, overt acts are taken in pursuance of the object or course of action, and

5. damages occur as a proximate result[37].

Penalties can include civil damages.

4.3.5 Conversion

If personal data is property (and some courts think it is), then by scraping it without permission the creator may be committing some form of conversion. For the purposes of conversion, "intent" merely means the objective to possess the property or exert property rights over it. A party is liable for conversion regardless of their knowledge of property's ownership status.

The elements are:

1. the plaintiff's ownership or right to possession of the property,

2. the defendant's conversion by wrongful act inconsistent with the property rights of the plaintiff, and

3. damages[38].

Penalties can include civil damages.

4.4 Property

4.4.1 Trespass to Chattels

A plaintiff alleging trespass to chattels in the web scraping context generally must allege that a defendant accessed its computer system without authorization and caused damage. These claims often, but not always, turn on whether a plaintiff can allege that a scraper damaged its computer systems. Courts have differed on what constitutes such "damage." In some cases, a plaintiff's allegation of an increased burden on server capacity caused by a scraper's activity may constitute damage for purposes of the claim.

The basic elements of a claim of trespass to chattels are:

1. the lack of the plaintiff's consent to the trespass,

2. interference or intermeddling with possessory interest, and

3. the intentionality of the defendant's actions[39].

In a trespass to chattel claim, you can typically recover actual damages (as opposed to nominal damages). Courts measure actual damages by assessing the diminished total value of the chattel resulting from the defendant's actions. For instance, if someone damages your bicycle, the compensation would be based on the bike's reduced value due to the damage.

---

[37]https://www.findlaw.com/smallbusiness/business-laws-and-regulations/civil-conspiracy.html

[38]See, e.g., https://www.law.cornell.edu/wex/conversion

[39]See, e.g., https://www.findlaw.com/injury/torts-and-personal-injuries/trespass-to-chattels.html

4.5 Criminal

4.5.1 Computer Fraud and Abuse Act (CFAA)

The Computer Fraud and Abuse Act of 1986 (CFAA) is a United States cybersecurity bill. A claim under subsection 1030(a)(c) of the CFAA, for example, has three elements:

1. a defendant has accessed a protected computer;
2. has done so without authorization or by exceeding such authorization as was granted; and
3. he obtains information from the protected computer[40].

Penalties include fines per crime, imprisonment from 1 to 20 years, or a combination of both[41].

4.5.2 Aid and Abet

Aiding and abetting is a legal doctrine related to the guilt of someone who aids or abets (encourages, incites) another person in the commission of a crime (or in another's suicide). It applies where it cannot be shown the party personally carried out the criminal offense, but where another person may have carried out the illegal act(s) as an agent of the charged, working together with or under the direction of the charged, who is an accessory to the crime.

The elements are:

1. That the accused had specific intent to facilitate the commission of a crime by another;
2. the accused had the requisite intent of the underlying substantive offense;
3. the accused assisted or participated in the commission of the underlying substantive offense; and
4. someone committed the underlying offense[42].

Penalties vary by state. For example, in Texas a first-degree felony, like theft of more than $200,000 worth of property is punishable by five to 99 or life in prison, along with a $10,000 fine[43].

4.5.3 Larceny/Theft

If personal data is property (and some courts think it is) then by scraping it without permission the creator may be committing some form of property theft.

The elements typically include:

1. the unlawful taking,
2. of the property of another,
3. without their consent, and
4. with the specific intent to deprive the owner of that property.

As a more specific example, the California Penal Code Sec 496(a) applies to (i) every person who, (ii) buys or receives any property that has been stolen or that has been obtained in any manner constituting theft or extortion, (iii) knowing the property to be so stolen or obtained, or (iv) who conceals, sells, withholds, or aids in concealing, selling, or withholding any property from the owner, knowing the property to be so stolen or obtained.

Penalties will vary by state. For example, under California Penal Code Sec. 496 penalties could include: 1) imprisonment, 2) Any person who has been injured by a violation of subdivision (a) or (b) may bring an action for three times the amount of actual damages, if any, sustained by the plaintiff, costs of suit, and reasonable attorney's fees[44].

---

[40] 18 U.S. Code § 1030 - Fraud and related activity in connection with computers
[41] 18 U.S. Code § 1030 - Fraud and related activity in connection with computers
[42] https://www.justice.gov/archives/jm/criminal-resource-manual-2474-elements-aiding-and-abetting
[43] https://statutes.capitol.texas.gov/Docs/PE/htm/PE.12.htm
[44] https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?lawCode=PEN&sectionNum=496

4.6 Privacy

4.6.1 General Data Protection Regulation

The European Union's General Data Protection Regulation (GDPR)[45] applies to personal data related to EU residents, and covers data controllers, processors, and data subjects.

The GDPR (General Data Protection Regulation) is defined by its scope, rights of data subjects, and requirements imposed upon data processors:

- Scope:
  - Applies to personal data related to EU residents.
  - Covers data controllers, processors, and data subjects.
- Rights of Data Subjects:
  - Right to Access: Individuals can request information about their data processing.
  - Right to Rectification: Correct inaccuracies in personal data.
  - Right to Erasure (Right to Be Forgotten): Request deletion of personal data.
  - Right to Data Portability: Obtain and reuse personal data for other services.
  - Right to Object: Object to processing based on legitimate interests.
- Legal Basis for Processing:
  - Requires a lawful basis (e.g., consent, contract, legal obligation).
- Data Protection Officer (DPO):
  - Appoint a DPO if processing is regular or systematic.
- Breach Notification:
  - Notify authorities within 72 hours of a data breach.
- Privacy by Design and Default:
  - Integrate privacy into systems and processes.
- Automated Decision-Making:
  - Transparency and the right to human intervention.

Fines for Non-Compliance: Up to 4% of global annual turnover or €20 million, whichever is higher.

4.6.2 California Consumer Privacy Act

The California Consumer Privacy Act (CCPA)[46] is similarly defined by its scope, consumer rights, and requirements imposed:

- Scope:
  - Applies to personal information of California residents.
  - Covers businesses that meet specific criteria.
- Consumer Rights:
  - Right to Know: Request information about data collection.
  - Right to Delete: Request deletion of personal information.
  - Right to Opt-Out: Opt out of the sale of personal information.
  - Right to Non-Discrimination: Businesses cannot discriminate based on exercising rights.
- Transparency and Disclosure:
  - Inform consumers about data collection practices.

The California Attorney General enforces most of the CCPA. Penalties could include fines for violations, but private right of action is limited to data breaches. The private right of action only covers a data breach involving certain types of personal information.

---

[45]https://gdpr-info.eu/
[46]https://oag.ca.gov/privacy/ccpa

4.7 Miscellaneous

4.7.1 Data Broker Registration

Two states currently have data broker registration laws.

California's data broker law requiring registration[47].

- "Data broker" means a business that knowingly collects and sells to third parties the personal information of a consumer with whom the business does not have a direct relationship.
- California's law likely doesn't apply to nonprofits because they're not a "business."

Vermont's data broker law[48]:

- A Data Broker is a business, or unit or units of a business, separately or together, that knowingly collects and sells or licenses to third parties the brokered personal information of a consumer with whom the business does not have a direct relationship.
- Vermont's law specifically says it doesn't matter if the entity was organized to operate at a profit.

Penalties could include:

- California: $100 per day
- Vermont: $50 per day, not to exceed $10,000 per year for each year the entity fails to register

4.7.2 Unfair or Deceptive Trade Practices

According to the Federal Trade Commission (FTC)[49]:

An act or practice is "unfair" if it "causes or is likely to cause substantial injury to consumers which is not reasonably avoidable by consumers themselves and not outweighed by countervailing benefits to consumers or to competition."

"Deceptive" practices are defined in the Commission's Policy Statement on Deception as involving a material representation, omission or practice that is likely to mislead a consumer acting reasonably in the circumstances.

Civil Penalties[50]:

- The FTC can seek civil penalties against companies that engage in unfair or deceptive conduct. These penalties can exceed the profits earned through misconduct.
- Companies that receive a "Notice of Penalty Offenses" from the FTC and still engage in prohibited practices can face civil penalties of up to $50,120 per violation[51].

Private Enforcement:

- In some states, statutes allow private enforcement.
- Individuals may be entitled to recover punitive damages and/or statutory fines.

State-Specific Penalties:

- Each state has its own laws regarding deceptive trade practices. Penalties vary based on state statutes.

Injunctions and Restraint Orders:

- Courts may issue injunctions or restraining orders to prevent continued deceptive trade practices.

---

[47]https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB362
[48]https://sos.vermont.gov/corporations/other-services/data-brokers/
[49]https://www.ftc.gov/about-ftc/mission/enforcement-authority
[50]15 U.S. Code § 45 - Unfair methods of competition unlawful; prevention by Commission
[51]https://www.ftc.gov/enforcement/penalty-offenses

4.7.3 State Computer Access Laws

State laws generally cover two areas:

Unauthorized Access and Computer Trespass:

- Unauthorized access involves approaching, communicating with, or altering computer resources without consent.
- Laws in all 50 states, Puerto Rico, and the Virgin Islands address unauthorized access or computer trespass.

Malware and Viruses:

- Malware and viruses modify, damage, destroy, or transmit information within computer systems without permission.
- Laws address actions that disrupt normal operation of computer systems.

Penalties vary by state but may include fines, imprisonment, or both.

Examples:

- In California, unlawful access, alteration, or damage to computer systems can result in fines, imprisonment, or both[52].
- In New Mexico, causing damage exceeding $250 can lead to felony charges[53].

4.7.4 False Designation of Origin

False designation of origin occurs when the manufacturer or seller lies about the country of origin or maker of its products.

Elements[54]: Any person who, on or in connection with any goods or services, or any container for goods, uses in commerce any word, term, name, symbol, or device, or any combination thereof, or any false designation of origin, false or misleading description of fact, or false or misleading representation of fact, which:

- is likely to cause confusion, or to cause mistake, or to deceive as to the affiliation, connection, or association of such person with another person, or as to the origin, sponsorship, or approval of his or her goods, services, or commercial activities by another person, or
- in commercial advertising or promotion, misrepresents the nature, characteristics, qualities, or geographic origin of his or her or another person's goods, services, or commercial activities,

shall be liable in a civil action by any person who believes that he or she is or is likely to be damaged by such act.

Penalties could include[55]:

- Injunctions
- Any damages sustained by the plaintiff, defendant's profits, and the costs of the action.
- In exceptional cases, reasonable attorney fees
- The court may order that any infringing articles bearing the word, term, name, symbol, or device be destroyed

---

[52]https://www.calpers.ca.gov/docs/ca-penal-code-502.pdf
[53]https://law.justia.com/codes/new-mexico/2021/chapter-30/article-45/section-30-45-5/
[54]15 U.S. Code § 1125 - False designations of origin, false descriptions, and dilution forbidden
[55]15 U.S.C. §1116(a)-§1118

4.7.5 Trademark Dilution

Use in commerce of a mark or trade name if such use causes dilution by blurringe or tarnishment of the distinctive quality of a famous trademark[56].

Notes: "Dilution" is statutorily defined in 15 U.S.C. §1127 to mean "the lessening of the capacity of a famous mark to identify and distinguish goods or services, regardless of the presence or absence of ... (1) competition between the owner of the famous mark and other parties, or (2) likelihood of confusion, mistake, or deception."

"Blurring" occurs when the famous mark's ability to identify its product has been impaired due to an association in the minds of consumers arising from similarity between another mark and the famous mark.

Tarnishment occurs when the reputation of a famous mark has been harmed by negative associations arising from the similarity between another mark and the famous mark. Situations in which tarnishment could result are when a famous trademark is "linked to products of shoddy quality, or is portrayed in an unwholesome or unsavory context, with the result that the public will associate the lack of quality or lack of prestige in the defendant's goods with the plaintiff's unrelated goods."

Penalties could include[57]:

- Injunctions.

- Owners of famous marks may also be entitled to the following additional remedies listed below if the mark or trade name that is likely to cause dilution by blurring or dilution by tarnishment was first used in commerce by the alleged infringer after Oct. 6, 2006; and (A) in a dilution by blurring action, the person willfully intended to trade on the recognition of the famous mark; or (B) in a dilution by tarnishment action, the person willfully intended to harm the reputation of the famous mark.

- For a willful violation, any damages sustained by the plaintiff, defendant's profits, and the costs of the action.

- In exceptional cases, reasonable attorney fees.

- For a willful violation, the court may order that any infringing articles bearing the word, term, name, symbol, or device be destroyed.

4.7.6 Trade Secrets (State and Federal)

The Uniform Trade Secrets Act (UTSA; which is broadly adopted by the states) defines a "trade secret" as:

information, including a formula, pattern, compilation, program, device, method, technique, or process that:

- Derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable by proper means by, other persons who can obtain economic value from its disclosure or use; and

- Is the subject of efforts that are reasonable under the circumstances to maintain its secrecy[58].

Remedies include both injunctive relief and monetary damages[59]. In some instances, penalties could include imprisonment[60].

---

[56]https://www.everycrsreport.com/reports/RL34109.html#_Toc452631097
[57]Under 15 U.S.C. §1125(c)(5)
[58]https://www.law.cornell.edu/wex/trade_secret
[59]https://www.reuters.com/legal/legalindustry/remedies-trade-secret-misappropriation-2023-05-08/
[60]18 U.S. Code § 1832 - Theft of trade secrets

5 Other Risks: Post-Deployment

These claims either haven't been made in GenAI litigation, so they're more speculative, or they occur infrequently. Given their less prominent nature, this section, like the previous section, will give them a more cursory overview.

5.1 Contract

5.1.1 Clickwrap and Enforceability of Artifact Terms

Many websites have a Terms of Service or Terms of Use agreement and those terms may include prohibitions on how the site data can be accessed and/or used. There are several arguments for and against allowing liability for web scrapers based on only browsewrap and the law is unsettled when it comes to using scraped data to train LLMs. However, the consensus appears to be that only clickwrap (requiring an affirmative assent to terms) is enforceable[61].

The potential penalties are similar to breach of terms:

- If you breach the ToS, the website operator may terminate your access to their services.
- Violating the ToS might lead to civil lawsuits. For instance, if your actions cause financial harm to the website or other users, you could be held liable for damages.
- There have been cases where violating a website's ToS was treated as a criminal offense.

5.2 Torts

5.2.1 Gross Negligence

Gross negligence occurs in both common and statutory law.

According to the Legal Information Institute, gross negligence is a lack of care that demonstrates reckless disregard for the safety or lives of others, which is so great it appears to be a conscious violation of other people's rights to safety. Gross negligence is a heightened degree of negligence representing an extreme departure from the ordinary standard of care. Falling between intent to do wrongful harm and ordinary negligence, gross negligence is defined as willful, wanton, and reckless conduct affecting the life or property or another.

Potential Penalties

- Civil penalties that vary by state
- A misdemeanor conviction for gross negligence can get you up to a year in jail. A felony conviction can result in several years of imprisonment.
- A felony conviction for gross negligence may also result in the loss of certain civil rights, such as the right to vote, serve on a jury, or possess a firearm.

5.2.2 Defamation

Defamation is a statement that injures a third party's reputation[62].

Elements:

1. The defendant made a statement about the plaintiff to another party.
2. The statement was intended to cause harm.
3. The statement caused injury to the plaintiff's reputation.
4. The statement was false.

---

[61]https://blog.ericgoldman.org/archives/2022/12/hello-youve-been-referred-here-because-youre-wrong-about-web-scraping-laws-guest-blog-post-part-2-of-2.htm

[62]https://www.law.cornell.edu/wex/defamation

5. If the plaintiff is a public figure, then the defendant must have made the false statement intentionally or with reckless disregard. This is known as the actual malice standard. The same can be true if the plaintiff was not necessarily a public figure. They could instead be involved in some newsworthy event or a matter of public concern.

6. The statement was not privileged information, such as any information shared between an attorney and a client or a doctor and a patient.

Potential Penalties

- Compensatory damages intended to compensate the victim of the defamation for the actual harm or loss they suffered as a result of the false statement.

- Punitive damages intended to punish the person who made the false statement for their malicious or reckless conduct and deter them from engaging in similar behavior in the future.

- Injunctions

- The court may order a retraction

5.2.3 Public Disclosure of Private Facts

Elements of the Tort:

1. The Defendant gives publicity to a matter concerning the private life of another.

2. The Plaintiff did not consent to the publication.

3. The matter publicized or the act of publication would be highly offensive to a reasonable person.

4. The publication was not of legitimate concern to the public.

Penalties could include compensatory damages, punitive damages, and or injunctions.

5.2.4 False Light

False light differs from defamation primarily in being intended "to protect the plaintiff's mental or emotional well-being," rather than to protect a plaintiff's reputation as is the case with the tort of defamation and in being about the impression created rather than being about veracity. If a publication of information is false, then a tort of defamation might have occurred. If that communication is not technically false but is still misleading, then a tort of false light might have occurred[63].

Elements of the Tort:

1. The defendant published some information about the plaintiff.

2. The information must portray the plaintiff in a false or misleading light.

3. The information is highly offensive or embarrassing to a reasonable person.

4. The defendant must have published the information with reckless disregard for its offensiveness.

Penalties could include compensatory damages, punitive damages, and or injunctions.

5.2.5 Right to Publicity

Personality rights are generally considered to consist of two types of rights: the right of publicity, or the right to keep one's image and likeness from being commercially exploited without permission or contractual compensation, which is similar (but not identical) to the use of a trademark; and the right to privacy, or the right to be left alone and not have one's personality represented publicly without permission.

---

[63]https://en.wikipedia.org/wiki/False_light

A commonly cited justification for this doctrine, from a policy standpoint, is the notion of natural rights and the idea that every individual should have a right to control how their right of publicity is commercialized by a third party, if at all[64].

Elements:

1. Use of an individual's name or likeness
2. for commercial purposes
3. without Plaintiff's consent

Penalties could include compensatory damages, punitive damages, and or injunctions.

5.2.6 Appropriation of Name and Likeness

Appropriation occurs when a defendant uses a plaintiff's name, likeness, or image without their permission for commercial purposes. The right of publicity applies to individuals (for example, a celebrity) with a proven commercial value to their image or identity, while appropriation pertains to everyone.

Elements:

1. The defendant used the plaintiff's name or likeness,
2. the plaintiff did not consent,
3. the defendant gained a commercial benefit (or some other advantage),
4. the plaintiff was harmed, and
5. that the defendant's conduct was a substantial factor in causing the plaintiff's harm.

Penalties could include compensatory damages, punitive damages, and or injunctions.

5.2.7 Failure to Warn

The duty to warn arises in product liability cases, as manufacturers can be held liable for injuries caused by their products if the product causes an injury to a consumer and the manufacturer fails to supply adequate warnings about the risks of using the product or if they fail to supply adequate instructions for the proper use of the product. If the manufacturer fails to supply these warnings, the law will consider the product itself to be defective[65].

Elements:

1. That the manufacturer knew of the danger posed by the product;
2. That the manufacturer had a duty to warn consumers of the danger related to the product;
3. The manufacturer was negligent in relation to their duty to warn; and
4. The manufacturer's failure to warn resulted in the plaintiff's injury.

Penalties could include compensatory damages, punitive damages, and or injunctions.

5.2.8 Libel

Libel is a method of defamation expressed by print, writing, pictures, signs, effigies, or any communication embodied in physical form that is injurious to a person's reputation, exposes a person to public hatred, contempt or ridicule, or injures a person in his/her business or profession. State courts generally follow the common law of libel, which allows recovery of damages without proof of actual harm. Under the traditional rules of libel, injury is presumed from the fact of publication. However, the U.S. Supreme Court has held that the First Amendment's protection of freedom of expression limits a State's ability to award damages in actions for libel[66].

---

[64]https://en.wikipedia.org/wiki/Personality_rights
[65]https://en.wikipedia.org/wiki/Duty_to_warn
[66]https://www.law.cornell.edu/wex/libel

While it is sometimes said that the person making the libelous statement must have been intentional and malicious, actually it need only be obvious that the statement would do harm and is untrue.

Elements:

1. A defendant made a written factual and defamatory statement purporting to be fact;
2. Regarding the plaintiff;
3. That was published without privilege or authorization to others by the defendant; and
4. There was resultant injury, or some harm caused to the reputation of the person or entity who is the subject of the statement, unless the statement falls within a category of "per se" harm.

Penalties could include compensatory damages, punitive damages, and or injunctions.

5.3 Privacy

5.3.1 Children's Online Privacy Protection Act (COPPA)

COPPA applies to the online collection of personal information by persons or entities under U.S. jurisdiction about children under 13 years of age, including children outside the U.S. if the website or service is U.S.-based. It details what a website operator must include in a privacy policy, when and how to seek verifiable consent from a parent or guardian, and what responsibilities an operator has to protect children's privacy and safety online, including restrictions on the marketing of those under 13.

Although children under 13 can legally give out personal information with their parents' permission, many websites—particularly social media sites, but also other sites that collect most personal info—disallow children under 13 from using their services altogether due to the cost and work involved in complying with the law.

Penalties could include fines up to $51,744 per violation[67].

5.3.2 GDPR: Right to Erasure

The right to be forgotten (RTBF) is the right to have private information about a person be removed from Internet searches and other directories under some circumstances. The issue has arisen from desires of individuals to "determine the development of their life in an autonomous way, without being perpetually or periodically stigmatized as a consequence of a specific action performed in the past." It's distinct from the right to privacy, which constitutes information that is not publicly known, whereas the right to be forgotten involves removing information that was publicly known at a certain time and not allowing third parties to access the information.

Penalties could be up to 4% of global revenue.

6 Conclusion

This taxonomy is the first paper to organize the many potential legal challenges generative AI creators may face. The stakes are high and the uncertainty is higher. While many laws may apply to the cutting edge of AI, it is unclear under what circumstances and to what extent the AI creator should be legally liable.

Now that the claims have been identified, entities can better anticipate possible litigation claims and build stronger legal protections to ensure they are able to continue conducting the type of work they have decided to pursue. The taxonomy also helps to better identify some of the many open questions in law so researchers, policymakers, and legal scholars have a framework with which to prioritize and examine whether and how laws implemented decades ago can and should be applied to a type of artificial intelligence that sounded like science fiction just a few years ago.

In addition to clarifying the bounds of the law, researchers, legal scholars, and practitioners could also consider possible solutions that are better suited for probabilistic tools like AI rather than the

---

[67]https://www.ftc.gov/business-guidance/resources/complying-coppa-frequently-asked-questions

deterministic nature of tools in years past. Such solutions could include safe harbors so long as certain conditions are met to reduce harm or when conducting research. Another possible solution may be to broaden the application of the Learned Hand formula[68] from torts to apply to most of the issues surrounding AI. This would be a nod to the fact that it's impossible to anticipate all possible harms of AI, but it may be reasonable to require entities to implement best practices for their type of model given its capabilities.

For more information, Appendix A contains a table of GenAI lawsuits, showing the parties involved and the claims made by the plaintiffs.

References

Julia Barnett. The ethical implications of generative audio models: A systematic literature review. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '23. ACM, August 2023. doi: 10.1145/3600211.3604686. URL http://dx.doi.org/10.1145/3600211.3604686.

Charlotte Bird, Eddie L. Ungless, and Atoosa Kasirzadeh. Typology of risks of generative text-to-image models, 2023.

Canyu Chen and Kai Shu. Combating misinformation in the age of llms: Opportunities and challenges, 2023.

Erik Derner, Kristina Batistic, Jan Zah'alka, and Robert Babuška. A security risk taxonomy for large language models. *ArXiv*, abs/2311.11415, 2023. URL https://api.semanticscholar.org/CorpusID:265294893.

Maanak Gupta, CharanKumar Akiri, Kshitiz Aryal, Eli Parker, and Lopamudra Praharaj. From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy, 2023.

Peter Henderson, Eric Mitchell, Christopher D. Manning, Dan Jurafsky, and Chelsea Finn. Self-destructing models: Increasing the costs of harmful dual uses of foundation models, 2023.

Heidy Khlaaf, Pamela Mishkin, Joshua Achiam, Gretchen Krueger, and Miles Brundage. A hazard analysis framework for code synthesis large language models, 2022.

Seth Neel and Peter Chang. Privacy issues in large language models: A survey, 2024.

Karan Singhal, Tao Tu, Juraj Gottweis, Rory Sayres, Ellery Wulczyn, Le Hou, Kevin Clark, Stephen Pfohl, Heather Cole-Lewis, Darlene Neal, Mike Schaekermann, Amy Wang, Mohamed Amin, Sami Lachgar, Philip Mansfield, Sushant Prakash, Bradley Green, Ewa Dominowska, Blaise Aguera y Arcas, Nenad Tomasev, Yun Liu, Renee Wong, Christopher Semturs, S. Sara Mahdavi, Joelle Barral, Dale Webster, Greg S. Corrado, Yossi Matias, Shekoofeh Azizi, Alan Karthikesalingam, and Vivek Natarajan. Towards expert-level medical question answering with large language models, 2023.

Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, Courtney Biles, Sasha Brown, Zac Kenton, Will Hawkins, Tom Stepleton, Abeba Birhane, Lisa Anne Hendricks, Laura Rimell, William Isaac, Julia Haas, Sean Legassick, Geoffrey Irving, and Iason Gabriel. Taxonomy of risks posed by language models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, pp. 214–229, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450393522. doi: 10.1145/3531146.3533088. URL https://doi.org/10.1145/3531146.3533088.

Laura Weidinger, Maribeth Rauh, Nahema Marchal, Arianna Manzini, Lisa Anne Hendricks, Juan Mateos-Garcia, Stevie Bergman, Jackie Kay, Conor Griffin, Ben Bariach, Iason Gabriel, Verena Rieser, and William Isaac. Sociotechnical safety evaluation of generative ai systems, 2023.

---

[68]No liability if the burden of adequate precautions is greater than likelihood of the harm times the magnitude of the harm if it occurs.

Robert Wolfe, Yiwei Yang, Bill Howe, and Aylin Caliskan. Contrastive language-vision ai models pretrained on web-scraped multimodal data exhibit sexual objectification bias. In *2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23. ACM, June 2023. doi: 10.1145/3593013.3594072. URL `http://dx.doi.org/10.1145/3593013.3594072`.

Ziwei Xu, Sanjay Jain, and Mohan Kankanhalli. Hallucination is inevitable: An innate limitation of large language models, 2024.

A Lawsuits Against GenAI Companies

Plaintiff	Defendant	Claims	Date filed
J.L.	Alphabet, Deepmind, Google	1. Violation Of California Unfair Competition Law, Business And Professions Code §§ 17200, Et Seq. 2. Negligence 3. Invasion Of Privacy Under California Constitution 4. Intrusion Upon Seclusion5. Larceny/Receipt Of Stolen Property 5. Conversion 6. Unjust Enrichment 7. Direct Copyright Infringement 8. Vicarious Copyright Infringement 9. Violation Of Digital Millennium Copyright Act, 17 U.S.C. § 1202(B)	7/11/2023
Concord Music Group, Inc. et. al	Anthropic	1. Direct Copyright Infringement 2. Contributory Infringement 3. Vicarious Infringement 4. Removal Or Alteration Of Copyright Management Information	10/18/2023
Michael Chabon	Meta	1. Direct Copyright Infringement, 17 U.S.C. § 106, Et Seq. 2. Vicarious Copyright Infringement 17 U.S.C. § 106 3. Digital Millennium Copyright Act – Removal Of Copyright Management Information 17 U.S.C. § 1202(B) 4. Violations Of The California Unfair Competition Law Cal. Bus. & Prof. Code §§ 17200, Et Seq 5. Negligence 6. Unjust Enrichment	9/12/2023

Mike Huck-abee	Meta, Bloomberg, Microsoft, The EleutherAI Institute	1. Direct Copyright Infringement, 17 U.S.C. § 106, Et Seq. 2. Vicarious Copyright Infringement, 17 U.S.C. § 106 3. Digital Millennium Copyright Act – Removal Of Copyright Management Information, 17 U.S.C. § 1202(B) 4. Conversion 5. Negligence 6. Unjust Enrichment	10/17/2023
Richard Kadrey	Meta	1. Direct Copyright Infringement 17 U.S.C. § 106 2. Vicarious Copyright Infringement 17 U.S.C. § 106 3. Removal Of Copyright-Management Information And False Assertion Of Copyright 17 U.S.C. § 1202(B) 4. Unfair Competition Cal. Bus. & Prof. Code §§ 17200, Et Seq. 5. Unjust Enrichment California Common Law 6. Negligence California Common Law	7/7/2023
Abdi Nazemian	NVIDIA	1. Direct Copyright Infringement (17 U.S.C. § 501)	3/8/2024
Authors Guild	OpenAI	1. Direct Copyright Infringement (17 U.S.C. § 501) 2. Vicarious Copyright Infringement 3. Contributory Copyright Infringement	9/19/2023
The New York Times	Microsoft, OpenAI	1. Copyright Infringement (17 U.S.C. § 501) 2. Vicarious Copyright Infringement 3. Contributory Copyright Infringement 4. Digital Millennium Copyright Act – Removal Of Copyright Management Information (17 U.S.C. § 1202) 5. Common Law Unfair Competition By Misappropriation 6. Trademark Dilution (15 U.S.C. § 1125(C))	12/27/2023

Julian Sancton/Alter	OpenAI, Microsoft	1. Copyright Infringement (17 U.S.C. § 501) 2. Contributory Infringement	11/21/2023
A.T., J.H.	OpenAI, Microsoft	1. Violation Of Electronic Communications Privacy Act, 18 U.S.C. §§ 2510, Et Seq. 2. Violation Of The Computer Fraud And Abuse Act, 18 U.S.C. § 1030 3. Violation Of The California Invasion Of Privacy Act ("Cipa"), Cal. Penal Code § 631 4. Violation Of California Unfair Competition Law, Business And Professions Code §§ 17200, Et Seq. 5. Negligence 6. Invasion Of Privacy 7. Intrusion Upon Seclusion 8. Larceny/Receipt Of Stolen Property 9. Conversion 10. Unjust Enrichment 11. New York General Business Law §§ 349, Et Seq.	9/5/2023
Basbanes	Microsoft, OpenAI	1. Direct Copyright Infringement (17 U.S.C. § 501) 2. Vicarious Copyright Infringement 3. Contributory Copyright Infringement	1/5/2023
The Intercept Media	OpenAI, Microsoft	1. Violation of 17 U.S.C. § 1202(b)(1) 2. Violation of 17 U.S.C. § 1202(b)(3)	2/28/2024
Raw Story Media	OpenAI	1. Violation of 17 U.S.C. § 1202(b)(1)	2/28/2024

A.S.	OpenAI, Microsoft	1. Violation Of Electronic Communications Privacy Act, 18 U.S.C. §§ 2510, Et Seq. 2. Violation Of The Computer Fraud And Abuse Act, 18 U.S.C. § 1030 3. Violation Of The California Invasion Of Privacy Act ("Cipa"), Cal. Penal Code § 631 4. Violation Of California Unfair Competition Law, Business And Professions Code §§ 17200, Et Seq. 5. Negligence 6. Invasion Of Privacy 7. Intrusion Upon Seclusion 8. Larceny/Receipt Of Stolen Property 9. Conversion 10. Unjust Enrichment	2/27/2024
Michael Chabon	OpenAI	1. Direct Copyright Infringement, 17 U.S.C. § 106, Et Seq. 2. Vicarious Copyright Infringement 17 U.S.C. § 106 3. Digital Millennium Copyright Act – Removal Of Copyright Management Information 17 U.S.C. § 1202(B) 4. Violations Of The California Unfair Competition Law Cal. Bus. & Prof. Code §§ 17200, Et Seq. 5. Negligence 6. Unjust Enrichment	9/8/2023
Sarah Silverman	OpenAI	1. Direct Copyright Infringement 17 U.S.C. § 106 2. Vicarious Copyright Infringement 17 U.S.C. § 106 3. Digital Millennium Copyright Act— Removal Of Copyright Management Information 17 U.S.C. § 1202(B) 4. Unfair Competition Cal. Bus. & Prof. Code §§ 17200, Et Seq 5. Negligence Under California Common Law 6. Unjust Enrichment Under California Common Law	7/7/2023

Paul Tremblay	OpenAI	1. Direct Copyright Infringement 17 U.S.C. § 106 2. Vicarious Copyright Infringement 17 U.S.C. § 106 3. Digital Millennium Copyright Act—Removal Of Copyright Management Information 17 U.S.C. § 1202(B) 4. Unfair Competition Cal. Bus. & Prof. Code §§ 17200, Et Seq. 5. Negligence 6. Unjust Enrichment	6/28/2023
Mark Walters	OpenAI	1. Negligence 2. Libel	6/5/2023
Doe 1	OpenAI, Github, Microsoft	1. Violation Of The Digital Millennium Copyright Act 17 U.S.C. §§ 1201–1205 2. Breach Of Contract—Open-Source License Violations Common Law 3. Tortious Interference In A Contractual Relationship Common Law 4. Fraud Common Law 5. False Designation Of Origin—Reverse Passing Off 15 U.S.C. § 1125 6. Unjust Enrichment Cal. Bus. & Prof. Code §§ 17200, Et Seq. And Common Law 7. Unfair Competition 15 U.S.C. § 1125; Cal. Bus. & Prof. Code §§ 17200, Et Seq.; And Common Law 8. Breach Of Contract Violation Of Github Privacy Policy And Terms Of Service Cal. Bus. & Prof. Code § 22575–22579; Cal. Civ. Code § 1798.150; And Common Law 9. Violation Of The California Consumer Privacy Act Cal. Civ. Code § 1798.150 10. Negligence—Negligent Handling Of Personal Data Common Law 11. Civil Conspiracy Common Law 12. Declaratory Relief 28 U.S.C. § 2201(A) And Cal. Code Civ. Proc. § 1060	11/3/2022
Thomson Reuters	ROSS Intelligence	1. Copyright Infringement (17 U.S.C. § 101 et seq.) 2. Tortious Interference with Contract	9/25/2023

Sarah Andersen	Stability AI, Midjourney, DeviantArt	1. Direct Copyright Infringement 17 U.S.C. §§ 106, Et Seq. 2. Vicarious Copyright Infringement 17 U.S.C. §§ 106, Et Seq. 3. Violation Of The Digital Millennium Copyright Act 17 U.S.C. §§ 1201–1205 4. Violation Of The Statutory Right Of Publicity Cal. Civ. Code § 3344 5. Violation Of The Common Law Right Of Publicity Common Law 6. Unfair Competition 15 U.S.C. § 1125; Cal. Bus. & Prof. Code §§ 17200, Et Seq.; And Common Law 7. Breach Of Contract Violation Of Deviantart Policies Cal. Bus. & Prof. Code § 22575–22579; Cal. Civ. Code § 1798.150; And Common Law 8. Declaratory Relief 28 U.S.C. § 2201(A) And Cal. Code Civ. Proc. § 1060	1/13/2023
Getty Images	Stability AI	1. Copyright Infringement (17 U.S.C. § 101 Et Seq.) 2. Providing False Copyright Management Information In Violation Of 17 U.S.C. § 1202(A) 3. Removal Or Alteration Of Copyright Management Information In Violation Of Section 1202(B) 4. Trademark Infringement In Violation Of Section 32 Of The Lanham Act, 15 U.S.C. § 1114(1) 5. Unfair Competition In Violation Of Section 43(A) Of The Lanham Act, 15 U.S.C. § 1125(A) 6. Trademark Dilution In Violation Of Section 43(C) Of The Lanham Act, 15 U.S.C. § 1125(C) 7. Deceptive Trade Practices In Violation Of Delaware's Uniform Deceptive Trade Practices Act 8. Trademark Dilution In Violation Of Section 3313 Of The Delaware Trademark Act	2/3/2023

Table 2: Lawsuits against GenAI companies.