



RAND / Research & Commentary / Research Reports /

Liability for Harms from AI Systems

The Application of U.S. Tort Law and Liability to Harms from Artificial Intelligence Systems

[Gregory Smith](#), [Karlyn D. Stanley](#), [Krystyna Marcinek](#), [Paul Cormarie](#), [Salil Gunashekar](#)

RESEARCH

Published Nov 20, 2024

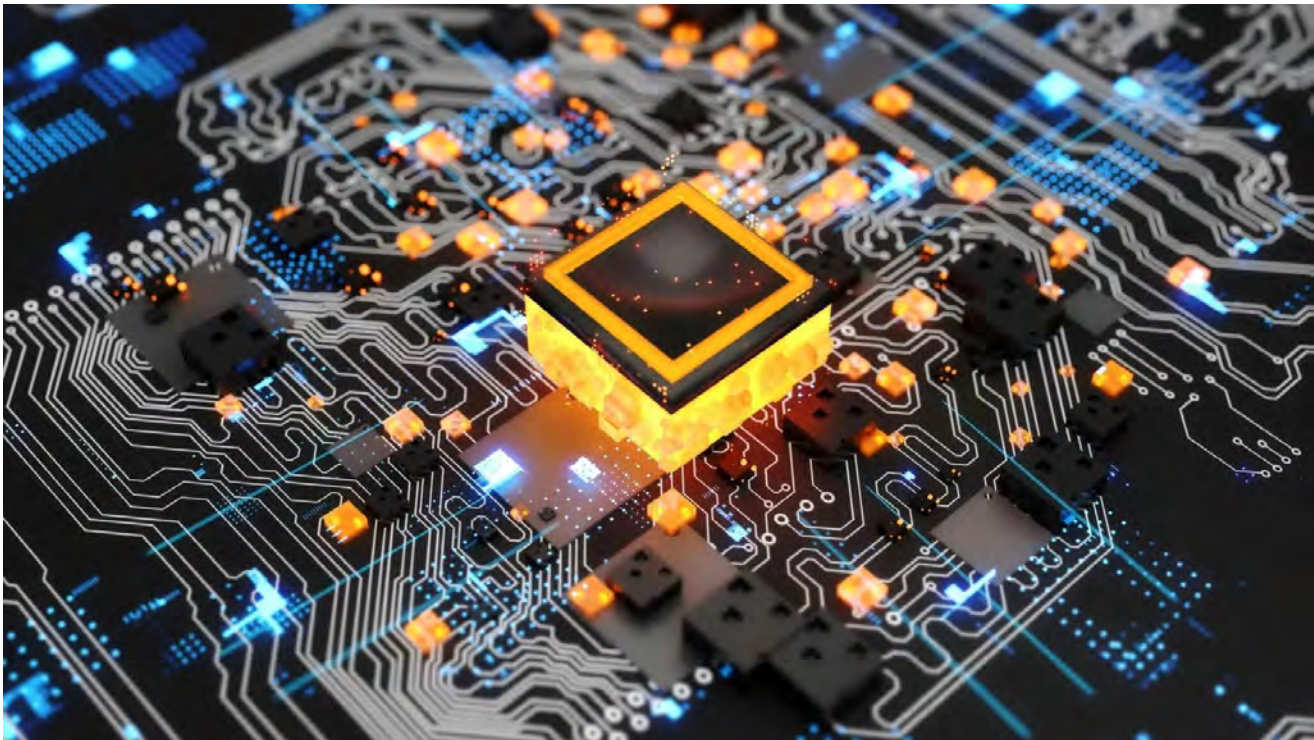


Photo by Getty Images/hh5800

Key Takeaways

- Absent new law from Congress or state legislatures, tort law will be applied to harms caused by artificial intelligence (AI) when such cases are brought in court.
 - Tort law is a form of common law, meaning that it has been formed primarily by judges ruling on specific cases and developing precedent to apply to future cases.
 - In the United States, tort law is primarily state law and can vary from state to state. Therefore, the specific tort law applied to AI will differ depending on which state's law is applied, with potentially large differences in how cases are adjudicated and how each state's tort law develops over time.
- Many AI-related tort cases will involve claims of negligence (a claim that a party did not act with due care) by harmed plaintiffs against AI developers and deployers.
 - Courts could find it difficult to apply negligence law to AI because of the complexities of AI development and the AI “supply chain” between AI developers and deployers, both of which could make it difficult to identify negligent behavior and the party responsible for an alleged harm.
 - When deciding whether an AI company acted negligently, courts might look to industry AI standards and to the customs in place for developing and deploying AI to determine what reasonable safety measures for AI look like.
- Plaintiffs also might bring cases alleging that products involving AI are defective; these cases would be decided under specialized products liability law.
 - It is unclear whether courts will define AI as a product, and different products liability tests could lead to very different analyses of whether an AI is truly defective (as required to proceed with the case) under these doctrines.
- The application of tort law to AI might involve the First Amendment, particularly for AI use cases that involve speech. In such cases, it might be more difficult for plaintiffs to win in court against AI developers.

The advent of new, powerful artificial intelligence (AI) systems has prompted a burst of interest in how best to govern AI. Although software and AI have existed for many years, new AI technologies, such as generative AI, with increasingly broad capabilities have recently been introduced, such as OpenAI's ChatGPT and Anthropic's Claude; these developments have increased interest in AI governance and in understanding existing law's application to this technology.^[1] There is interest in how existing rules of liability might apply to these new AI systems.^[2] This report provides a brief overview of some of the key issues at the intersection of advanced AI systems and U.S. tort law.

This report, which focuses on AI's effects on U.S. tort law, is not intended to provide a comprehensive analysis but rather to highlight key issues and areas of uncertainty raised by the continued advancement of AI technology and its deployment. It is intended to support continued dialogue among stakeholders on specific facets of AI governance, especially as AI applications proliferate worldwide and complex governance debates persist.^[3] Note that legal and regulatory frameworks will similarly apply to AI when relevant, but that is outside the scope of this report.

The Structure of Liability Law in the United States and the Legal Implications for Liability Related to AI-Caused Harms

Established principles and frameworks of liability and tort law will be particularly influential in governing AI. Tort law is the set of rules, doctrines, and statutes that governs remedying of harms caused by wrongful actions under civil law in the United States.^[4] When one party harms another and is then sued by the injured party, the case will usually be brought in civil court

and decided under tort law. Philosophically, tort law seeks to compensate the person least at fault.

In contrast to many other proposed regimes for governing AI, tort law is already in force and is regularly applied in state and federal court. Liability law also does not require an act by Congress, the President, or other federal or state decisionmakers to apply to AI-caused injuries. It is therefore valuable for policymakers to understand some of the key structural components and doctrines of tort law and how they might be applied to AI-caused harms in the future.

To provide some guidance about how liability might be assigned for civil harms from AI, it is first necessary to point out two critical structural components of tort law in the United States: First, it is primarily a form of common law; second, most U.S. tort law is made by state courts.

Tort Law Is Common Law

Common law is a form of law that is primarily created through judicial decisions in the process of litigation rather than being made through statutes passed by legislatures.^[5] Many of the rules governing the allocation of liability for harms, such as the duties that parties owe each other or what constitutes negligence, have been developed by the accumulation of judicial decisions over time. This said, not all tort law is common law; statutes can directly modify the rules of tort, and statutes are often given significant weight in determining whether a party is liable for the harms they inflict—even if those statutes do not explicitly address themselves to the subjects that tort law covers.^[6]

This state of affairs has several implications for liability for harms done by AI. First, absent statutory intervention by federal or state legislatures, courts will need to adjudicate harms caused by novel AI systems and resolve uncertainties by building on common-law foundations. Although some predictions can be made about how tort doctrine will be applied to AI, there will still be

significant uncertainty as to how tort law will treat specific issues in concrete cases, across different states as they are put before the courts. (Several of these key uncertainties are covered in more detail later in this report.)

Second, if and when cases are brought by plaintiffs, judges will eventually need to grapple with complex and novel legal issues, and juries will face complex factual issues. AI-related cases will require judges to apply existing law and doctrine to the novel circumstances that AI will create. In the U.S. legal system, judges alone—not juries—determine the law, such as legal duties, under tort. This will result in many cases not making it to trial because judges will apply the law to these new facts and decide many cases solely as matters of law, or "a determination of the applicable law."^[7] Although some cases might be easily decided under precedent, judges could have to grapple with difficult questions regarding the design and deployment of AI, particularly when injuries occur that a party might claim were caused by the AI's developer rather than by the owner of a particular AI-enhanced product. These decisions also could include creating new common-law doctrines to handle the new challenges that AI poses—decisions that might then be applied in future cases in that same state jurisdiction. Judges will therefore have a major role both in determining how specific AI-related tort cases will proceed and how each state's doctrine develops as a result of AI-related challenges.

Although many cases are likely to be decided as matters of law by judges, some cases will still proceed to trial and potentially be decided by a jury. Many components of tort law are *questions of fact* (dealing with "the particular details of the case")^[8] and therefore traditionally left to juries to determine, although defendants can and do agree to waive their right to a jury trial and have judges act as the fact-finder for their case.^[9] The outcomes of these cases will rely on the specific factual judgments of judges and juries whose behavior and sympathies might be difficult to predict. The outcome of novel AI-related cases will therefore be particularly uncertain. This might drive more cases to settle out of court if AI companies seek to avoid the risks associated with litigating novel facts for which there is little precedent—particularly for large-scale harms against sympathetic plaintiffs.

In the United States, Tort Law Is Primarily a Form of State Law

In the United States, the common law of tort is primarily formed by the decisions and precedents of state courts, not federal ones.^[10] In most situations, there is no such thing as federal common law, and federal courts will work through a complex choice of legal rules to determine which state common law should be applied to a particular case. Although federal common law does exist for narrow areas, such as for harms arising on maritime waters under admiralty law, federal courts mostly apply state common law to allocate liability for harms.^[11] Therefore, it will be primarily state common law, and potentially state courts, that will develop rules to govern AI-related torts.

State control of tort law in the United States also means that there can be variation between different states on key rules for determining liability.^[12] How AI-caused harms are treated will differ from jurisdiction to jurisdiction, and such variations could significantly affect the imposition of liability and/or potential size of liability judgments against AI developers. The potential for state-by-state variation has, in the past, created incentives for federal policymakers to intervene and set national rules of the road for liability.^[13] Whether such a solution is right for AI is unclear. Although setting common rules would make it easier for AI developers to manage their risk and make clear to developers and users when they would be held liable, the relative youth of AI technology means that it might not yet be possible to know what these ideal rules should be.

Although common law is primarily formed in state courts, there is also a strong interest in synthesizing and identifying the common elements of different states' bodies of tort law.^[14] Much of this work is done by the American Law Institute, which issues Restatements of Law, including the Restatements of Torts, that synthesize and summarize state bodies of law. The Restatements are often considered highly persuasive by state courts when they formulate their

own common law rules. However, there is no requirement that a state court adopt the Restatement's particular synthesis of a given law, and states regularly reject the Restatements or adopt them only in part. The Restatements are therefore one set of influential standards that help shape the common-law of tort across the United States. The remainder of this report draws from the Restatements of Tort because they are still relied on in many states, but the authors acknowledge that states might have their own individual standards or develop unique new rules for applying tort law to AI.

Contracts Are Often Used to Resolve Traditional Issues of Tort Law

It should also be noted that contract terms often require parties to forgo tort claims should a product malfunction. These limitations on liability are common in user agreements for software and might prevent AI users from pursuing traditional tort claims should such terms be included in that AI's user agreement. However, some of these waivers might not be enforceable, and there could be confusion when claims are brought regarding whether contractual waivers of tort claims apply in specific cases. Note, however, that if there is a contract between the parties, then an injured party will not be able to bring tort claims should the contract be found valid.

Negligence Cases Against AI Developers and Users

Many tort claims regarding AI will allege *negligence*.^[15] These cases generally allege that a defendant acted negligently and breached a duty of care toward the plaintiff, resulting in the plaintiff's injury.^[16] Although there can be specific duties of care, all people are held to have a duty to take reasonable care against causing foreseeable harm to others.^[17] This is generally summarized as requiring people to be "reasonably prudent" both in their behavior and in

taking precautions to ensure they do not inflict harms on others.^[18] This requirement generally applies to preventing harm to a person's body and property; there is not generally a duty against causing pure economic loss or purely emotional harm.^[19]

Several issues immediately arise from this simple description of negligence and its application to AI. To start, what standards will be used to judge whether an AI developer or deployer was reasonably prudent when they developed or deployed their AI? The duty an AI developer might have to deployers or users of their product is also unclear. There will also be arguments about whether a particular AI was truly *defective* that could be difficult to settle, given that AI is used to automate more tasks and that clear definitions of "defective" behavior might not yet exist. Although reason and logical judgment will play a role in determining this duty, courts are also likely to look to the practices and customs of the AI industry more broadly to judge whether a party was negligent in developing or deploying their AI.^[20]

Negligence law in the United States generally places great emphasis on industry custom and standard when determining whether a defendant was negligent; following industry practice is taken as strong, although not conclusive, evidence that a defendant was not negligent.^[21] Conversely, a defendant not following industry practice will be taken as strong evidence that a defendant was negligent. Courts are also not limited in determining what constitutes industry practice; they can look to informal standards or those promulgated by an industry body that the court finds representative of industry practice.^[22]

This means that, in common law, the AI industry's own safety practices will play a significant role in defining what is required of AI developers and deployers to avoid liability for negligence. Courts can still find AI companies negligent even if they did follow industry custom—courts have noted in the past that entire industries could be negligent or lagging in their safety practices.^[23] However, particularly safety-conscious AI companies, or associations of AI companies, also might have an opportunity to develop standards for safe AI

development that in turn could establish benchmarks for the whole industry in future litigation. Policymakers and government bodies could also play a role in this standard-setting by publishing their own guidance for AI development and deployment.^[24]

Another issue arises from the complexity of the AI supply chain and its intersection with the requirements of causation. *Causation*, in negligence, requires that a defendant must be the factual, or "but-for," cause of the plaintiff's injury for the defendant to be found liable.^[25] Demonstrating which party caused a plaintiff's injury might be difficult for cases involving AI systems. AI, particularly large language models (LLMs) such as ChatGPT, are often developed by one party, which then makes its model available for various AI deployers to customize or fine-tune to behave in certain ways for their own products.^[26] Although LLMs can be made safer through fine-tuning and other techniques, these methods are not foolproof, and protections embedded in LLMs can be removed by other developers and bypassed by users.^[27] It might not be clear whether an injury resulted from actions by the developer, the deployer, or some other party, let alone the question of whether any party's actions were also negligent.

With such an extended supply chain, the complexity of AI development, and the difficulty of controlling LLM behavior, it might be difficult for plaintiffs to show that any actor was the but-for cause of a particular injury. An AI developer might argue that the deployer's usage of its AI was the cause of the injury; the AI deployer might argue that the original developer was, in fact, negligent in some way in the original creation of the model. Courts have developed doctrines to deal with issues of causation in other contexts, and will have examples to draw from when encountering complex issues of causation posed by AI. However, courts could face difficulties with the uncertainties involved in identifying a particular party's behavior as the true cause of a plaintiff's injuries.

Potential Problems with Standing in AI-Related Tort Cases

Another issue that plaintiffs in AI-related tort cases might face is proving standing. Standing is a constitutional requirement that must be fulfilled by a plaintiff to successfully bring a case; failure to show standing results in a case being dismissed.^[28] In federal court, standing generally requires that a plaintiff demonstrate "injury in fact," meaning an actual injury that is "concrete and particularized" to them, that there is a causal connection between the injury and the conduct of the defendant, and that the injury can be redressed by a favorable court decision.^[29] The requirement that plaintiffs show they have suffered a concrete and particularized injury has been a major barrier to software-related suits because common alleged injuries, such as invasion of privacy, often have not been concrete enough, or been able to point to sufficiently clear damages to plaintiffs, to satisfy the requirement for standing.^[30] The same issues with showing concrete and particularized harms could reemerge with suits related to AI. Although a case involving physical harm caused by an AI malfunction would be very likely to pass the standing requirement, it might be more difficult to show a sufficiently "concrete and particularized injury" when an AI violates a plaintiff's privacy or outputs misinformation about a plaintiff.

The Potential Impact of Products Liability Laws

In addition to the law of negligence, U.S. states have developed specialized doctrines of products liability to adjudicate cases related to the manufacture, design, or warning provided regarding products.^[31]

The first question is whether AI will be considered a *product*. Courts have regularly considered whether other forms of software were considered products and have mostly held that they are not, for the purposes of products liability law.^[32] It should also be noted that software is not considered a product under the Uniform Commercial Code, a model law that many states have adopted; this, in turn, could make treating AI as a product more difficult.^[33] The incorporation of AI into physical products—such as IoT (Internet of Things) devices, self-driving cars, or robots—might prompt courts to begin considering AI as a form of product subject to specialized products liability doctrines.

If AI is considered a product, it might be subject in certain state jurisdictions to the consumer expectations test, a legal standard that considers whether a product is dangerous to a degree in a way that is not within the expectation of the ordinary purchaser or user of that product.^[34] This test asks whether a product is "malfunctioning" according to the expectations of the consumer; if it is, liability is assigned to the product's designer.^[35]

The issue of consumer expectations for AI might therefore be particularly important in some state jurisdictions when determining liability for an AI-caused tort. AI developers and deployers might be able to avoid some liability by warning users of the potential risks that an AI system might pose, thereby appropriately calibrating a user to the potential harms that an AI product could inflict. The issue of how consumers would expect an AI system to behave will therefore be particularly critical should AI, or AI-enabled products, be broadly subject to products liability law throughout the United States.

It should also be noted that some states do not apply the consumer expectations test but rather apply the "risk-utility" test.^[36] This test asks whether an alternative design would, at reasonable cost, reduce the risk posed by a particular product.^[37] In practice, the risk-utility test tends to hinge on analyzing whether a particular design is "reasonable" compared with alternatives; for AI, this assessment might hinge on whether appropriate safeguards and controls were implemented in a particular AI system that is

alleged to have caused harm. For cases in which the risk-utility test is used to determine products liability, whether AI companies and deployers are held liable might turn on judgments about whether they acted reasonably in designing the AI system in question.

Implications of the First Amendment for AI Liability

It should also be noted that the First Amendment's applications to tort law and civil liability might also have significant implications for AI. This is because many outputs from AI systems take the form of text or speech and, therefore, might plausibly receive some protection under the First Amendment.

Historically, the Supreme Court has held that the imposition of civil liability for speech implicates the First Amendment and must be subject to its restrictions.^[38] This rule does not prevent the imposition of civil liability in tort cases for speech, but it has led to changes in tort doctrine to avoid running afoul of the First Amendment. For example, in *New York Times Co. v. Sullivan*, the Supreme Court limited the common-law tort of defamation in consideration of the First Amendment.^[39] Courts often balance the concerns of the First Amendment and tort law, choosing to make success in a tort case more difficult if the conduct in question was or resembled speech.^[40]

The primary implication of these precedents is that if an AI's output resembles speech, it is possible that a court might make it more difficult to recover in a tort case related to that output.^[41] Conversely, it will probably be easier for a plaintiff to succeed in a tort claim if an AI's behavior takes the form of "acting" in the real world and directly injures a plaintiff, such as a malfunctioning self-driving car hitting a pedestrian, given that speech considerations would be significantly reduced in such a situation.

Uncertainties Surrounding Section 230

Tort cases involving AI might also be affected by Section 230 of the Communications Decency Act of 1996, which reads that "[n]o provider or user of an interactive computer service shall be treated as the publisher of any information provided by another information content provider."^[42] Section 230 has been a major barrier in tort cases brought against technology companies in the past, so there is interest in how it might affect cases brought regarding harms caused by AI.^[43] Unfortunately, there is no definitive answer yet on how Section 230 will apply to AI broadly. It is likely that the application of Section 230 will be highly specific to the particular facts of each case and how a particular harm was caused by AI. A key point of uncertainty is whether an AI developer or deployer would be considered a *content provider* under Section 230, which would prevent that party from invoking it as a protection against tort liability.

Conclusion

The common-law nature of tort and its development by state courts mean that the development of tort law in the AI context will be a gradual and distributed process, which creates a potential space for federal policymakers to intervene to create rules of the road for AI liability that might apply nationwide. However, policymakers will have to carefully assess different considerations, such as the proper balance between state and federal law under the U.S. federal system. By the same token, it might be undesirable to intervene in the fact-finding and doctrinal development of the courts, which might be able to develop more-comprehensive solutions to harms from AI as the technology rapidly evolves and is deployed in new use cases that legislators might not foresee. Such interventions can also ease the path for plaintiffs in tort cases, however—for example, recognizing new classes of harms or imposing new

duties on defendants could encourage developers and deployers of AI to be more cautious in their use of such systems where the systems could cause harm. How to strike this balance is beyond the scope of this report, but it is hoped that a stronger understanding of the issues at the intersection of tort law and AI will help inform future discussion and debate.

Notes

[1] Many of the descriptions and examples of tort law in this report are drawn, with omissions and minor adjustments, from Ketan Ramakrishnan, Gregory Smith, and Conor Downey, *U.S. Tort Liability for Large-Scale Artificial Intelligence Damages: A Primer for Developers and Policymakers*, RAND Corporation, RR-A3084-1, 2024. As of October 14, 2024: https://www.rand.org/pubs/research_reports/RRA3084-1.html

[2] Matthew van der Merwe, Ketan Ramakrishnan, and Markus Anderljung, "Tort Law and Frontier AI Governance," *Lawfare*, May 24, 2024.

[3] This volume is one in a series of closely related RAND publications on this topic. These closely related volumes share some material, including descriptions, figures, and tables. See Tifani Sadek, Karlyn D. Stanley, Gregory Smith, Krystyna Marcinek, Paul Cormarie, and Salil Gunashekar, *Artificial Intelligence Impacts on Privacy Law*, RAND Corporation, RR-A3243-2, 2024. As of October 14, 2024: https://www.rand.org/pubs/research_reports/RRA3243-2.html

[4] Kevin M. Lewis and Andreas Kuersten, "Introduction to Tort Law," Congressional Research Service, *In Focus brief*, IF11291, last updated May 26, 2023.

[5] "Tort law has also historically been a matter of common law rather than statutory law; that is, judges (not legislatures) developed many of tort law's fundamental principles through case-by-case adjudication" (Lewis and Kuersten, 2023).

[6] "Common Law," *Black's Law Dictionary*, 11th ed., Thomson Reuters, 2019. Statutes can modify the rules of tort directly and are often relevant for determining tort liability. For example, a breach of a duty imposed by statute could be taken as strong evidence that a party has acted negligently. American Law Institute, *Restatement of the Law Third: Torts*, 2010, Section 14.

[7] Legal Information Institute, "Matter of Law," webpage, Cornell Law School, July 2023a. As of September 12, 2024: https://www.law.cornell.edu/wex/matter_of_law

[8] Legal Information Institute, "Question of Fact," webpage, Cornell Law School, August 2023b. As of October 12, 2024: https://www.law.cornell.edu/wex/question_of_fact

[9] American Law Institute, 2010, Section 8.

[10] For example, *Erie Railroad Co. v. Tompkins* holds that federal courts should apply state "substantive law" when deciding claims brought under state law. Tort law is generally considered substantive; therefore, federal courts apply it when hearing state tort claims in federal court. *Erie Railroad Co. v. Tompkins*, 304 U.S.

64, 1938.

[11] Brandon J. Murrill, "Federal Admiralty and Maritime Jurisdiction Part 4: Torts and Maritime Contracts or Services," Congressional Research Service, *Legal Sidebar*, LSB10827, 2022.

[12] For example, New York and California have different standards for when a plaintiff can succeed in a case against a manufacturer, with it being more difficult in New York to succeed in such a case. *Robinson v. Reed-Prentice Div. of Package Mach. Co.* states, "We hold that a manufacturer of a product might not be cast in damages, either on a strict products liability or negligence cause of action, where, after the product leaves the possession and control of the manufacturer, there is a subsequent modification which substantially alters the product and is the proximate cause of plaintiff's injuries" (*Robinson v. Reed-Prentice Div. of Package Mach. Co.*, 403 N.E.2d 440, New York, February 14, 1980). Compare this with California Civil Jury Instruction No. 1245, which states that a product's modification or misuse should provide its manufacturer with a complete defense against liability only if the manufacturer can prove that "it was so highly extraordinary it was not reasonably foreseeable" (California Civil Jury Instruction No. 1245, Affirmative Defense—Product Misuse or Modification, December 2013).

[13] For example, Congress passed legislation in 2005 that barred civil suits against firearm manufacturers for certain third-party misuses of such weapons (Public Law 109-92, Protection of Lawful Commerce in Arms Act, October 26, 2005; codified at U.S. Code, Title 15, Sections 7901-7903).

[14] The American Law Institute is a private, nonprofit entity that consists of lawyers, judges, and legal academics. For more information, see Legal Information Institute, "Restatement of the Law," webpage, Cornell Law School, August 2020. As of September 12, 2024: https://www.law.cornell.edu/wex/restatement_of_the_law

[15] Kenneth S. Abraham, *The Forms and Functions of Tort Law: Concepts and Insights*, 5th ed., Foundation Press, 2017.

[16] American Law Institute, *Restatement of the Law First: Torts*, 1934, Section 282.

[17] American Law Institute, 2010, Section 7.

[18] A person is negligent, and therefore breaches their duty of care, when they fail to act as a "reasonably prudent person" (American Law Institute, 2010, Section 3).

[19] American Law Institute, 2010, Section 7.

[20] American Law Institute, 2010, Section 13.

[21] "[B]y making reference to custom evidence, the typical jury instruction on custom emphasizes to the jury the potential probative value of this evidence" (Kenneth S. Abraham, "Custom, Noncustomary Practice, and Negligence," *Columbia Law Review*, Vol. 109, No. 7, 2009).

[22] American Law Institute, 2010, Section 13.

[23] "A whole calling many have unduly lagged in the adoption of new and available [safety] devices" (*The T.J. Hooper, the Northern No. 30 and No. 17, the Montrose in re Eastern Transp. Co. New England Coal & Coke Co. v. Northern Barge Corporation, H. N. Hartwell & Son, Inc., v. same.*, 60 F.2d 737, 740, 2d Cir. 1932).

[24] AI developers might also be held liable for malpractice should courts find there to be a recognized professional standard of care that a developer then violated. Sharma discusses the potential implications of

developing such a standard of care for tort liability in AI. See Chinmayi Sharma, *AI's Hippocratic Oath*, Yale Law & Economics Research Paper, March 14, 2024.

[25] American Law Institute, 2010, Section 27. Plaintiffs are also required to show *proximate cause*, which requires that, for a defendant to be liable for a harm, the harm must result from the risks that the defendant's conduct created. Many jurisdictions have adopted a foreseeability test for proximate cause, holding that a defendant is liable only for the foreseeable harms of their conduct. For AI-caused harms, this might occur when an AI acts in ways that were unforeseeable to its designer and causes harm, which might make it difficult for a plaintiff to prove proximate cause. As an AI becomes more capable, it could also become prone to malfunction in ways that are increasingly difficult to foresee; this, in turn, could increase the difficulty of proving proximate harm because it might be easier for defendants to argue that the harm that the AI technology inflicted was not foreseeable.

[26] For example, see OpenAI's fine-tuning of GPT-4 to refuse requests for personal information or "illicit advice" (OpenAI, *GPT-4 System Card*, March 23, 2023, pp. 3, 13) and Gemini Team Google's discussion of the fine-tuning of Gemini to generate "harmless" responses (Gemini Team Google [Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M. Dai, Anja Hauth, Katie Millican, David Silver, et al.], "Gemini: A Family of Highly Capable Multimodal Models," arXiv, arXiv:2312.11805, December 19, 2023). Anthropic uses a similar process to prevent harmful model behaviors (Anthropic, *The Claude 3 Model Family: Opus, Sonnet, Haiku*, 2024, p. 3).

[27] Users can, for example, engage in "prompt injection" to bypass fine-tuning that would normally prevent an LLM from responding to certain inputs. See Rich Harang, *Securing LLM Systems Against Prompt Injection*, Nvidia, August 3, 2023; and Qiusi Zhan, "Removing RLHF Protections in GPT-4 via Fine-Tuning," arXiv, arXiv:2311.05553, revised April 5, 2024.

[28] Legal Information Institute, "Standing," webpage, Cornell Law School, June 2024. As of September 12, 2024: <https://www.law.cornell.edu/wex/standing#>

[29] *Lujan v. Defenders of Wildlife*, 504 U.S. 555, 1992.

[30] *Reilly v. Ceridian Corp.*, 664 F.3d 38, 42, 3d Cir., 2011.

[31] Manufacturing defect liability applies when a product does not match, or function according to, its intended design—for example, when it is missing components or has improperly manufactured parts. Harms from AI are less likely to implicate manufacturing liability because most potential harms from AI result from issues with an AI's design rather than a manufacturing defect in a particular instance of an AI's code. There is significant debate about how products liability law can or should be applied to AI. This section provides a quick overview of some key issues at the intersection of this body of law and advanced AI. For more-detailed discussion of potential products liability frameworks, see John Villasenor, *Products Liability Law as a Way to Address AI Harms*, Brookings Institution, October 31, 2019; and Catherine M. Sharkey, "A Products Liability Framework for AI," *Columbia Science and Technology Law Review*, Vol. 25, No. 2, 2024.

[32] American Law Institute, 2010, Section 19. Products are "tangible personal property" while software is generally not considered tangible. Courts have not found software to be a product for the purposes of products liability rules.

[33] Uniform Commercial Code, Section 9-102, Definitions and Index of Definitions, Item 42, "General Intangible."

[34] American Law Institute, *Restatement of the Law Second: Torts*, 1965, Section 402.

[35] American Law Institute, 2010, Section 2.

[36] American Law Institute, *Restatement of the Law Third: Product Liability*, 1998, Section 2b.

[37] American Law Institute, 1998, Section 2.

[38] Regarding a promissory estoppel claim, a court wrote that "[s]uch a cause of action, although private, involves state action within the meaning of the Fourteenth Amendment and therefore triggers the First Amendment's protections, since promissory estoppel is a state-law doctrine creating legal obligations never explicitly assumed by the parties that are enforceable through the Minnesota courts' official power" (*Cohen v. Cowles Media Co.*, 501 U.S. 663, 1991).

[39] *New York Times Company v. Sullivan*, 376 U.S. 254, 1964.

[40] For example, the Ninth Circuit prevented plaintiffs from bringing strict liability claims against publishers of *The Encyclopedia of Mushrooms* after they were poisoned when relying on incorrect information in the encyclopedia (*Winter v. G.P. Putman's Sons*, 938 F.2d 1033, 1035, 9th Cir., 1991).

[41] For an argument that AI outputs are a form of speech protected under the First Amendment, see Eugene Volokh, Mark A. Lemley, and Peter Henderson, "Freedom of Speech and AI Output," *Journal of Free Speech Law*, Vol. 3, No. 113, 2023. For an argument that AI outputs should not receive First Amendment protections, see Peter Salib, *AI Outputs Are Not Protected Speech*, University of Houston Law Center, Public Law and Legal Theory Research Paper Series, No. 2024-A-5, 2024.

[42] U.S. Code, Title 47, Chapter 5, Subchapter II, Part 1, Section 230, Protection for Private Blocking and Screening of Offensive Material; Item (c)(1).

[43] Valerie C. Brannon and Eric N. Holmes, *Section 230: An Overview*, Congressional Research Service, R46751, 2024.

Topics

Artificial Intelligence	Science, Technology, and Innovation Policy
Emerging Technologies	Trade Regulation
European Union	United States
Market Regulation	

Document Details

Copyright: RAND Corporation	DOI: https://doi.org/10.7249/RRA3243-4
Availability: Web-Only	Document Number: RR-A3243-4
Year: 2024	