

Bridging the Legal Void Surrounding Generative AI Liability: A Proposed Path Forward

Lucas D. Jacob*

Abstract

The impressive growth of Generative AI has exposed numerous gaps within the United States' liability laws that current frameworks were never designed to address. Section 230 of the Communications Decency Act shields Internet platforms from liability for user-posted content, but generative AI actually creates content rather than just hosting it. This Note examines why courts have avoided ruling on Section 230's applicability to AI-generated content and why traditional tort claims like product liability and negligence face serious obstacles when applied to AI. This Note then compares how California, the EU, and China have approached AI regulation with each attempting to protect consumers without stifling innovation. Drawing on these efforts, this Note argues that Congress should amend Section 230 to clarify that generative AI falls outside its scope of protection and establish a federal agency to oversee a risk-based regulatory scheme. This framework would offer safe harbours for compliant developers while shifting the burden of proof through a rebuttable presumption of causation, giving harmed individuals a realistic path to relief.

Keywords: Generative Artificial Intelligence, Section 230, liability, risk-based regulation, Safe Harbour

A Introduction

Artificial intelligence (AI) has surged in popularity within the past few years. Businesses are exploring potential applications for this continually evolving technology to increase efficiency and productivity in every part of a company's framework.¹ Individuals can now readily generate new text, videos, and images within moments, and many hope that this technology will revolutionize industries worldwide.² However, novel questions have arisen about the extent to

* Lucas D. Jacob, Indiana University Robert H. McKinney School of Law - Indianapolis, Indiana, USA.

1 *Generative AI Could Raise Global GDP by 7%*, Goldman Sachs (5 April 2023), <https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent> [<https://perma.cc/T494-E2DS>].

2 *Artificial Intelligence's Use and Rapid Growth Highlight Its Possibilities and Perils*, U.S. Government Accountability Office (6 September 2023), <https://www.gao.gov/blog/artificial-intelligences-use-and-rapid-growth-highlight-its-possibilities-and-perils> [<https://perma.cc/FYH2-K25Z>].

which this technology should be used and what regulations should be put in place to reduce abuse.

Historically, Section 230 of the Communications Decency Act (CDA) aimed to shield Internet content platforms from liability for harmful content posted by a user, provided that the platform did not materially contribute to the content's creation.³ Generative artificial intelligence (Gen AI) presents a unique shift in determining Section 230's applicability to content providers.⁴ Upon a user's prompt, Gen AI platforms formulate a unique response by pulling on incredible amounts of data to predict the correct or most likely desired answer.⁵ This content generation makes it distinct from traditional Internet platforms in that the content generated is not solely of human origin.⁶ Instead, Gen AI materially contributes to the generation of the output, raising the question of whether or not the Gen AI service fills the role of an information content provider. Under that interpretation, the application to Section 230 regarding AI companies is suspect.⁷

To illustrate, imagine that one day, as you browse social media, you come across a news article featuring an image that resembles you, stating that a person with your name has committed a violent bank robbery. You eventually discover that the news agency, which uses an AI tool to skim articles and produce content they can then post to their website, acknowledges that their AI tool has made a horrible mistake. It has combined two articles: one about an actual bank robbery and another detailing your philanthropic efforts around the city that featured a picture of you, creating a false story by their combination. The damage is done, though, as you get calls from friends and family asking if you were the assailant. When you file a lawsuit against the company for defamation, you also find that the AI developer is a national company facing multiple similar claims in other states. However, you see that federal safe-harbour law, Section 230 of the CDA, has been cited and effectively prohibits this type of lawsuit because AI developers have historically been found not to be information content providers. Further, the outcomes of other cases have yielded results indicating that the complexity of AI systems makes it nearly impossible to prove causation. Although this story is purely hypothetical, it highlights the risk posed by Gen AI and the potential consequences that may come to fruition without more structured regulations on AI.

Absent the hypothetical prohibition in the example, claims arising from Gen AI-related harm could be solved through traditional tort schemes. However, under these different schemes, one event bringing about multiple claims in several states could foreseeably lead to judgments being rendered differently in each state. This approach not only poses issues for potential plaintiffs and defendants, as there is a

3 Pete J. Benson & Valerie C. Brannon, Cong. Rsch. Serv., LSB11097, Section 230 Immunity and Generative Artificial Intelligence 2 (28 December 2023), <https://crsreports.congress.gov/product/pdf/LSB/LSB11097> [https://perma.cc/W2YS-67E6].

4 *Ibid.*

5 Cole Stryker & Mark Scapicchio, *What Is Generative AI?*, IBM (22 March 2024), <https://www.ibm.com/topics/generative-ai> [https://perma.cc/6TW3-EX5M].

6 *Ibid.*

7 Benson & Brannon, *supra* note 3, at 3.

lack of clarity on how the issues will be decided, but it also threatens to disrupt the global benefit that AI brings. As such, the complexity of this technology requires a consistent federal approach that allows AI technology to grow and enables individuals harmed by AI services to find protection under the law. However, no consistent federal approach exists to address this problem.⁸

A Gen AI system's potential outputs range from simple images and videos to enhanced medical imagery or complex weather forecasts.⁹ The only true limitation is defined by the creativity of the prompt and the amount of data available to that platform to develop the content.¹⁰ Internally, many of these Gen AI companies have policies regulating the platform's content.¹¹ There is a growing push for the AI industry to be self-regulated due to technology's complexity and fast-moving nature.¹² However, federal courts lack guiding statutes or regulations that can clearly and succinctly define liability regarding content created by a Gen AI platform.¹³

This Note argues that a clear and consistent federal approach that clarifies Section 230's reach and sets clear obligations for AI developers is necessary to provide more legal clarity to consumers and AI developers. The increasing ubiquity of AI underscores the need for a unified regulatory framework that respects the delicate balance between proper regulation and over-regulation to avoid chilling the growth of AI. Part I of this Note explains generally what Gen AI is and the history of its development. It then discusses the benefits of Gen AI and some associated risks that have already begun to surface around the country. Part II of this Note details the federal government's attempt to regulate the Internet in its early days, which parallels the explosive growth of AI today. Further, Part II discusses Section 230 of the CDA and explains why courts have hesitated to develop case law surrounding this statute, which was created to regulate the open Internet and could potentially be used to immunize AI companies from harm that their systems arguably contribute to producing. Part III then performs a comparative analysis of the regulatory framework for AI in California, the EU and China. Finally, Part IV of the Note concludes with a proposal for a statutory framework that addresses the concerns presented by Section 230 and other unique legal issues of Gen AI.

B Background of Generative AI

I Creation of Generative AI

The details surrounding how AI and, more specifically, Gen AI work are complex; however, only a basic understanding of these systems is needed to understand the

⁸ *Ibid.*, at 3-5.

⁹ *What Is Generative AI?* McKinsey & Company (2 April 2024), <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai> [<https://perma.cc/LWFF9-E824>].

¹⁰ *Ibid.*

¹¹ Melissa Heikkilä, *AI Companies Promised to Self-Regulate One Year Ago. What's Changed?*, MIT Technology Review (22 July 2024), <https://www.technologyreview.com/2024/07/22/1095193/ai-companies-promised-the-white-house-to-self-regulate-one-year-ago-whats-changed/> [<https://perma.cc/YL23-SNKG>].

¹² *Ibid.*

¹³ Benson & Brannon, *supra* note 3, at 3-5.

possible benefits and risks. Machine learning is a subset of AI that allows a computer to process a foundational model on which it was built, find patterns by engaging in an iterative process and predict desired outputs.¹⁴ Most machine learning platforms analyse past historical data to predict future outcomes of inputs by essentially using statistical models for predictions.¹⁵ The system looks at the foundation on which it was built, tests the proposal or assertion against similar examples within its foundation, and presents the most likely desired outcome.¹⁶ Deep learning models are a highly advanced subset of machine learning that simulates human interactions using immense amounts of data to produce results.¹⁷ These deep learning models are regularly used to develop outputs in various mediums, and the foundation from which a result is derived depends on the output content's format.¹⁸ For example, outputs can range from text to computer code to photos to videos or audio.¹⁹ Each output type relies on potentially billions of data points to properly educate the model.²⁰

The primary distinction between basic machine learning models, which have been used for decades for predictive analysis, and today's Gen AI, which more closely resembles a deep learning model, involves the apparent originality of the output data.²¹ Gen AI's distinguishing feature is that its content remains similar to the training data but is novel content tailored to the user's request.²²

II Potential Benefits of Adopting Gen AI

Some sources anticipate Gen AI adding \$2.6 to \$4.4 trillion annually to the global economy throughout all industry sectors.²³ Within those sectors, "customer operations, marketing and sales, software engineering, and research and development" are anticipated to experience the most increase in value.²⁴ Analysts predict that Gen AI could replace 60% to 70% of an employee's daily tasks because of its ability to process and understand natural language or non-technical and regular human-speak.²⁵ While reading through the potential impacts that Gen AI could have, it is important to consider the potential liability issues arising from these uses.

An AI system's ability to analyse vast amounts of data in conjunction with a medical professional's guidance promises to be a significant benefit in the medical

14 *What Is Machine Learning (ML)?*, Berkely: Ischoolonline (26 June 2020), <https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/> [https://perma.cc/6U7P-Q3CB].

15 *Ibid.*

16 *Ibid.*

17 Stryker & Scapicchio, *supra* note 5.

18 *Ibid.*

19 *Ibid.*

20 *Ibid.*

21 Adam Zewe, *Explained: Generative AI*, MIT: MIT News (9 November 2023), <https://news.mit.edu/2023/explained-generative-ai-1109> [https://perma.cc/H4UX-BMG9].

22 *Ibid.*

23 Michael Chui et al., *The Economic Potential of Generative AI: The Next Productivity Frontier*, McKinsey: McKinsey Digital (14 June 2023), <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#key-insights> [https://perma.cc/N2P7-U784].

24 *Ibid.*

25 *Ibid.*

field.²⁶ Traditional decision-making tools are being outmatched by ever-evolving AI systems.²⁷ For example, the Modified Early Warning Score is a conventional tool used to determine a patient's risk of rapid deterioration while in hospital care.²⁸ AI has already shown itself to be more accurate than this traditional and time-tested tool.²⁹ Hospitals capture 3.6 billion images through different diagnostic testing procedures, around 97% of which go unused.³⁰ The profound amount of data captured, though collectively too vast for human analysis, presents a perfect opportunity for a Gen AI system to create new insights by analysing patterns.³¹ Healthcare professionals have a prime opportunity to "structure, index and leverage this information for more accurate diagnostics".³² AI's integration into the healthcare sector aims to improve patient safety, error detection, patient data organization and drug delivery.³³ Many healthcare leaders are betting on AI and believe it will soon play a pivotal role in clinical decision-making.³⁴

Pharmaceutical companies and the medical products field are another industry that will yield significant upticks in revenue from using Gen AI, estimated by some experts to be \$60 to \$110 billion annually.³⁵ Many of these gains can be seen within research and development departments, which, on average, consume up to twenty percent of a pharmaceutical company's revenue.³⁶

Banking has shown to be another prime industry in which Gen AI could have a significant impact.³⁷ The banking industry has been an early adopter of Gen AI, mainly in the marketing and customer-facing operations sectors.³⁸ In the future, however, analysts predict that the banking industry will find improvements in many other areas, such as customer credit analysis, internal chatbots acting as databases aiding in employee efficiency, and even automated systems that scan the jurisdiction to detect regulatory changes in the places that the banks are doing business.³⁹

III Risks Associated with Generative AI

The few examples of benefits listed previously merely hint at the potential benefits that AI is expected to deliver. However, as the following section portrays, there are risks that underscore the need for more intentional oversight.

26 *How AI Is Improving Diagnostics, Decision-Making and Care*, American Hospital Association, <https://www.aha.org/aha-center-health-innovation-market-scan/2023-05-09-how-ai-improving-diagnostics-decision-making-and-care> (accessed 30 September 2024) [<https://perma.cc/RZ8P-9GPK>].

27 *Ibid.*

28 *Ibid.*

29 *Ibid.*

30 *Ibid.*

31 *Ibid.*

32 *Ibid.*

33 *Ibid.*

34 *Ibid.*

35 Chui et al., *supra* note 23.

36 *Ibid.*

37 *Ibid.*

38 *Ibid.*

39 *Ibid.*

In March 2022, a startlingly realistic video was posted to X that seemed to portray Ukrainian President Volodymyr Zelenskyy asking his soldiers to lay down their arms and surrender to Russian forces.⁴⁰ Deepfakes, such as the one portraying President Zelenskyy, are digital creations that use sophisticated Gen AI platforms to create realistic-looking videos or photos of a subject that can mimic facial expressions, voice and overall appearance of an individual.⁴¹ In a separate situation involving deepfakes, San Francisco filed a first-of-its-kind lawsuit against two U.S. and several international AI companies, accused of aiding in the proliferation of AI-generated child porn.⁴² The lawsuit claims that those companies broke numerous “state and federal laws prohibiting deepfake pornography, revenge pornography, and child pornography, as well as violations of [California’s] Unfair Competition Law”.⁴³ The city seeks to shut down the platforms that created this content and pursue civil penalties.⁴⁴

Gen AI has been shown to amplify gender and racial stereotypes even worse than those found in society.⁴⁵ Bloomberg ran a test in which they prompted the AI system to create hundreds of images of people representing different jobs.⁴⁶ The results showed that lighter-skinned subjects were more represented in AI-generated images depicting higher paying jobs,⁴⁷ while darker skin tones were mainly depicted in jobs like “fast-food worker” and “social worker”.⁴⁸ This bias was also observed when assigning specific roles by gender.⁴⁹ Referring back to the benefits suggested within the banking industry, precisely the credit analysis suggestion, a biased Gen AI system could foreseeably cause significant harm to the discriminated-against group if these biases were to go uncorrected.

Image-based Gen AI outputs have created alarming results; however, text-based Gen AI outputs introduce unique challenges.⁵⁰ Text-based Gen AI systems can present information that appears authentic and accurate but is actually an untruthful fabrication contrived from the data sets used as a foundation for the

40 Bobby Allyn, *Deepfake Video of Zelenskyy Could Be ‘Tip of the Iceberg’ in Info War, Experts Warn*, NPR (16 March 2022), <https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia> [https://perma.cc/F8SG-42RV].

41 Alakananda Mitra, Saraju P. Mohanty & Elias Kougianos, *The World of Generative AI: Deepfakes and Large Language Models*, arxiv.org (6 February 2024), <https://arxiv.org/html/2402.04373v1#S2> [https://perma.cc/NKC3-U8FC].

42 Tim Fang, *San Francisco City Attorney Sues Websites Creating AI-Generated Deepfake Pornography*, CBS News (15 August 2024), <https://www.cbsnews.com/sanfrancisco/news/sf-city-attorney-sues-websites-creating-ai-generated-deepfake-pornography/> [https://perma.cc/S3NF-Z5F8].

43 *Ibid.*

44 *Ibid.*

45 Leonardo Nicoletti & Dina Bass, *Humans Are Biased. Generative AI Is Even Worse*, Bloomberg: Bloomberg Tech. & Equal. (9 June 2023), <https://www.bloomberg.com/graphics/2023-generative-ai-bias/> [https://perma.cc/NMB4-82G8].

46 *Ibid.*

47 *Ibid.*

48 *Ibid.*

49 *Ibid.*

50 Allyn, *supra* note 40.

system.⁵¹ These fabrications are typically called “hallucinations” within the AI industry.⁵² These hallucinations can manifest from an inaccurate context of the initial prompt or an untruthful foundation.⁵³ There are obvious risks associated with this hallucinated information. For example, companies planning to use Gen AI to synthesize large bodies of text into concise summaries, who trust that the text is a derivation of the source text, could find unrelated or novel conclusions that are not based on any of the source data. In a medical or pharmaceutical context, fabricated results incorrectly summarizing a patient’s medical records or suggesting inferior treatment options highlight some risks that hallucinated information could present.⁵⁴

The global accessibility of Gen AI and the potential for spreading wrong or misleading information could create defamation claims. A Georgia case filed in 2023 details an individual’s attempt to hold an AI company liable for its defamatory content.⁵⁵ *Walters v. OpenAI, LLC* was filed after ChatGPT generated false and damaging information about an individual.⁵⁶ Though several factual issues remain to be determined in this case, the court dismissed OpenAI’s motion to dismiss and allowed the case to proceed to trial.⁵⁷

Another case involved Microsoft’s Gen AI chatbot confusing two individuals with the same name, one of whom was convicted of seditious conspiracy after attempting to join the Taliban following 9/11.⁵⁸ In *Battle v. Microsoft Corporation*, the plaintiff accuses Microsoft of defamation for “propagating false information about [Mr. Battle]...”⁵⁹ Since the initial writing of this Note, the parties have elected to resolve the matter in a binding arbitration outside of court.⁶⁰

The rapid development and widespread adoption of Gen AI present both incredible opportunities and significant challenges across many sectors. While these Gen AI systems offer promising increases in efficiency and productivity, they also introduce complex risks. The tensions between Gen AI’s incredible benefits and its apparent risks create an urgent need for appropriate and updated legal frameworks to address these emerging issues. The current regulatory landscape in the United States, primarily centred around outdated Internet legislation, reveals significant gaps when applied to these new technologies. The following section explores the historical context of these

51 Matt O’Brien, *Chatbots Sometimes Make Things Up. Is AI’s Hallucination Problem Fixable?*, AP News: Business (1 August 2023), <https://apnews.com/article/artificial-intelligence-hallucination-chatbots-chatgpt-falsehoods-ac4672c5b06e6f91050aa46ee731bcf4> [https://perma.cc/47GJ-CVQM].

52 *Ibid.*

53 *Ibid.*

54 Chui et al., *supra* note 23.

55 Complaint at 3, *Walters v. OpenAI, LLC* (N.D. Ga. June 5, 2023) (No. 23-A-04860-2).

56 *Ibid.*

57 *Walters v. OpenAI, LLC* (N.D. Ga. June 5, 2023) (No. 23-A-04860-2).

58 Motion for Injunction at 3, *Battle v. Microsoft Corp.* (D. Md. May 2, 2024) (1:23-cv-01822).

59 *Ibid.*

60 Memorandum Opinion at 1, *Battle v. Microsoft Corp.* (D. Md. May 2, 2024) (1:23-cv-01822).

regulations and analyses the limitations of the existing legal mechanisms potentially available to address the unique challenges that Gen AI poses.

C Overview of the AI Regulation Landscape in the United States

I History of Section 230 of the CDA

Section 230 is a potential regulatory tool that could be applied to the Gen AI industry if courts decide that the system's output implicates the defences available within the federal statute.⁶¹ However, the two previously mentioned cases, *Battle v. Microsoft Corporation* and *Walters v. OpenAI, LLC*, are examples that have created questions as to the applicability of Section 230 against the outputs of a Gen AI system.⁶²

The development of Section 230 began in the early days of the Internet, and it dealt primarily with two cases that sought to resolve the issue of liability for illicit content posted to a website.⁶³ The first case, *Cubby v. CompuServe, Inc.*, involved an online service, CompuServe, Inc., which operated as a gateway to multiple other online forums.⁶⁴ One of those forums, Rumorville, was accused of creating false and defamatory statements about the plaintiff.⁶⁵ There, the District Court analogized CompuServe, Inc. to a newsstand and stated, “[I]t would be no more feasible for CompuServe to examine every publication it carries for potentially defamatory statements than it would be for any other distributor to do so”.⁶⁶ Accordingly, the court found that because CompuServe, Inc., serving as a distributor, “neither knew nor had reason to know of the allegedly defamatory Rumorville statements”, the company could not, as a matter of law, be held liable for the defamatory statements.⁶⁷

Next was a New York Supreme Court case called *Stratton Oakmont, Inc. v. Prodigy Services Co.*, which involved a dispute surrounding a defamatory comment posted to an online forum.⁶⁸ This forum, Prodigy, was eventually found liable for the comment posted because, as the court noted, “PRODIGY’s conscious choice, to gain the benefits of editorial control, has opened it up to a greater liability than CompuServe and other computer networks that make no such choice”.⁶⁹ Essentially, the court found that Prodigy was similar to a publisher because of its editorial functions and, therefore, was responsible for the content and thus liable.⁷⁰

Two House Representatives highlighted the dichotomy presented in these two high-profile cases.⁷¹ If a company like the one in *Cubby* were found to have hosted

61 Benson & Brannon, *supra* note 3, at 3.

62 *Ibid.*

63 Ash Johnson & Daniel Castro, *Overview of Section 230: What It Is, Why It Was Created, and What It Has Achieved*, ITIF (22 February 2021), <https://itif.org/publications/2021/02/22/overview-section-230-what-it-why-it-was-created-and-what-it-has-achieved/> [<https://perma.cc/5S64-P7PR>].

64 *Cubby, Inc. v. CompuServe, Inc.*, 776 F. Supp. 135, 138 (S.D.N.Y. 1991).

65 *Ibid.*

66 *Ibid.*, at 140.

67 *Ibid.*, at 141.

68 *Stratton Oakmont, Inc. v. Prodigy Servs. Co.*, 1995 WL 323710, at *1 (N.Y. Sup. Ct. May 24, 1995).

69 *Ibid.*, at *5.

70 *Ibid.*, at *3.

71 Johnson & Castro, *supra* note 63.

illegal content without any editorial involvement, it would be immune from any legal actions stemming from the content, provided the company remained unaware of the allegedly unlawful content.⁷² Conversely, the New York Supreme Court found that if a company, acting responsibly by editing out harmful content, inadvertently hosted a defamatory comment on a website, it could still be held liable because it has undertaken the role of an editor.⁷³ Section 230 was created in response to discrepancies in liability enforcement and a desire to nurture a budding and rapidly growing industry.⁷⁴

When *Stratton Oakmont* was decided in the mid-1990s, approximately 45 million people were using the Internet; just five years later, that number had skyrocketed to 150 million people worldwide.⁷⁵ In response to this tremendous growth, Congress sought to implement safeguards to protect those who might be harmed by illegal content and those who might inadvertently host that illegal content.⁷⁶ As the Internet changed people's lives worldwide at its advent, so does AI.⁷⁷

Although impactful in the simpler days of the Internet, Section 230 has been criticized today for being a simplistic regulatory tool within an incredibly complicated industry.⁷⁸ The Justice Department's 2020 proposal to update Section 230 expresses concern that the effective tool of the 1990s has outgrown its usefulness and requires an update.⁷⁹ Succinctly put, the Justice Department explains why Section 230 must be amended and stated:

[Websites] no longer function as simple forums for posting third-party content, but use sophisticated algorithms to suggest and promote content and connect users. Platforms also now offer an ever-expanding array of services, playing an increasingly essential role in how Americans communicate, access media, engage in commerce, and generally carry on their everyday lives.⁸⁰

II The Uncertainty Surrounding Liability Involving AI-Generated Content

Section 230 of the CDA immunizes providers and users of "interactive computer services" from liability stemming from content posted by another individual because those providers and users shall not be considered publishers.⁸¹ An "interactive computer service", commonly thought to include Facebook, Google or Amazon, is defined as any service that "provides or enables computer access by

⁷² *Ibid.*

⁷³ *Ibid.*

⁷⁴ Department of Justice's Review of Section 230 of the Communications Decency Act of 1996, U.S. Dep't of Just (16 September 2024), <https://www.justice.gov/archives/ag/departments-justice-s-review-section-230-communications-decency-act-1996> [https://perma.cc/C9ND-ZR2M].

⁷⁵ 1960s–1990s: Internet, Elon University: Imagining the Internet, <https://www.elon.edu/u/imagining/time-capsule/150-years/back-1960-1990/#:-:text=In%201996%2C%20there%20were%20approximately,were%20407%20million%20users%20worldwide> (accessed 30 September 2024) [https://perma.cc/6HRD-6LQJ].

⁷⁶ U.S. Dep't of Just, *supra* note 74.

⁷⁷ *Ibid.*

⁷⁸ *Ibid.*

⁷⁹ *Ibid.*

⁸⁰ *Ibid.*

⁸¹ Benson & Brannon, *supra* note 3, at 1.

multiple users to a computer server”.⁸² Section 230 confirms that only those who “[are] responsible, in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service” may be considered information content providers and, in effect, open themselves up to liability for illegal content.⁸³

Applying this definition to a Gen AI system creates a question. If a Gen AI system is responsible, in whole or in part, for the development of information provided through the Internet, would they be considered information content providers and, therefore, open themselves up to liability? Conversely, would this statute shield them from essentially all harmful content developed if a court alternatively finds that they are not information content providers under Section 230?

Several courts have weighed in on interpreting this statute, and some have utilized the “material contribution test” in determining the meaning of development in the context of a service’s influence on content posted by an external user.⁸⁴ For example, in the case *Fair Housing Council of San Fernando Valley v. Roommates.Com, LLC*, the Ninth Circuit court analysed a website’s involvement in a discriminatory feature inadvertently exploited on the website.⁸⁵ The court found that Roommates.com was not protected by Section 230 when they were “sufficiently involved with the design and operation of the search and email systems”, which limited access to housing based on several protected characteristics.⁸⁶ In addition, the court clarified that only a material contribution to a material illegality would satisfy the “development” criteria outlined in Section 230.⁸⁷

Courts have been surprisingly silent when analysing claims from algorithm-generated or, more specifically, Gen AI-related content and Section 230’s place within the analysis.⁸⁸ In the Ninth Circuit case *Gonzalez v. Google*, the court reviewed a district court’s application of Section 230 to dismiss the majority of the plaintiff’s claims.⁸⁹

On appeal, the court reviewed several terrorist attacks in Paris, Istanbul and San Bernardino, California, and the involvement that certain social media platforms had in supporting those attacks.⁹⁰ The plaintiffs claim that Facebook, Twitter and Google aided and abetted ISIS through the use of algorithmic content suggestions, which were generated in response to a user’s viewing habits.⁹¹ Of the three appeals addressed in this case, Google was the only

82 *Ibid.*

83 47 U.S.C. § 230.

84 Benson & Brannon, *supra* note 3, at 1.

85 *Fair Hous, Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1170 (9th Cir. 2008).

86 *Ibid.*

87 *Ibid.*, at 1168.

88 Benson & Brannon, *supra* note 3, at 2.

89 *Gonzalez v. Google LLC*, 2 F.4th 871, 885-886 (9th Cir. 2021), vacated and remanded, 598 U.S. 617 (2023), and rev’d sub nom. *Twitter, Inc. v. Taamneh*, 598 U.S. 471 (2023).

90 *Ibid.*, at 880.

91 *Ibid.*

company to assert that Section 230 immunized them from liability.⁹² In response, the plaintiff argued that

the immunity afforded by § 230 does not bar their claims because § 230 immunizes only those who publish content created by third parties, and their claims are directed to content created by Google.⁹³

The court analysed the “material contribution” test through which the role of a website in the “creation or development” of specific content can be determined.⁹⁴ They highlighted that

making a material contribution does not mean ‘merely taking action that is necessary to the display of the allegedly illegal content,’ but rather, ‘being responsible for what makes the displayed content allegedly unlawful.’⁹⁵

The Ninth Circuit affirmed the District Court’s ruling regarding the application of Section 230; however, the Supreme Court in *Twitter v. Taamneh*, reversed these cases on non-Section 230 grounds.⁹⁶ However, during the oral arguments presented to the Court, Justice Gorsuch implied that Gen AI should not be protected by Section 230.⁹⁷ He said, “I mean, artificial intelligence generates poetry, it generates polemics today. That – that would be content that goes beyond picking, choosing, analysing, or digesting content. And that is not protected”.⁹⁸ Though the Court did not directly analyse the use of Section 230 applied against a machine-based algorithm, the Ninth Circuit and Supreme Court’s distance from it has raised questions on its applicability to Gen AI.⁹⁹

III Using Traditional Legal Principles to Solve AI Liability Claims

If a court were to find that Section 230 does not shield an AI company from liability, the claimant would still need to detail a legal theory through which they could bring their claim. Tort law is one pathway, and although many states generally follow the same high-level principles, states could potentially have different tort laws. One potential tort that plaintiffs could use for relief is through a product liability claim.¹⁰⁰ According to the Second Restatement of Torts, for a product liability claim to succeed, a plaintiff must successfully show that a product existed with a defective condition, putting the consumer in an unreasonably dangerous position, and

92 *Ibid.*

93 *Ibid.*, at 890.

94 *Ibid.*, at 892.

95 *Ibid.*

96 *Twitter, Inc. v. Taamneh*, 598 U.S. 471 (2023).

97 Transcript of Oral Argument at 49-50, *Gonzalez v. Google, LLC*, 143 S.Ct. 1191 (2023) (No. 21-1333).

98 *Ibid.*

99 Lauren Willard Zehmer et al., *The U.S. Supreme Court Punts on Section 230 in Gonzalez v. Google LLC*, Covington (18 May 2023), <https://www.globalpolicywatch.com/2023/05/the-u-s-supreme-court-punts-on-section-230-in-gonzalez-v-google-llc/> [<https://perma.cc/4VVP-DQCP>].

100 Restatement (Second) of Torts § 402A (1965).

that the product eventually caused physical harm.¹⁰¹ The defendant must have sold the product, and the product must be in substantially the same condition as when it was originally sold.¹⁰² Lastly, even if a seller has exercised all possible care in producing the product and even if there was no contractual relationship with the injured party, the claim could still be brought.¹⁰³

From a surface-level reading of the elements required for a product liability claim, some questions immediately arise regarding its applicability to Gen AI. The first is the distinction between a product and a service. Modern courts have attempted to tackle this distinction in cases concerning social media platforms, applications and other online service providers by using the Cheapest Cost Avoider test, which looks at who is in the best position to control the potential harm caused by the service.¹⁰⁴ This test is a distinct offshoot from the traditional approach, which was developed in cases involving professional service providers like engineers, physicians or accountants, where the tangibility of the item produced was assessed to determine if the disputed item was a service or a product.¹⁰⁵ The less an output from a service provider looks like a professional service itself, the more likely a court would be to accept the claim under product liability.¹⁰⁶ Unfortunately, the case law surrounding product liability claims has not provided a clear precedent on the correct approach if a plaintiff brought a Gen AI claim under product liability.¹⁰⁷ Further, adapting this approach for something as global and far-reaching as Gen AI could create inconsistent adjudications. This inconsistency would harm consumers and producers of Gen AI alike by muddying the legal landscape with a patchwork liability system.

Negligence is another potential approach to remedying harm caused by AI.¹⁰⁸ The Second Restatement of Torts defines negligence as “conduct which falls below the standard established by law for the protection of others against unreasonable risk of harm”.¹⁰⁹ The Restatement goes on to say that negligent conduct is “an act which the actor as a reasonable man should recognize as involving an unreasonable risk of causing an invasion of an interest of another”.¹¹⁰ The Restatement’s reasonability assessment is generally focused on the tortfeasor’s actions.¹¹¹

A significant roadblock for bringing claims against an AI system under negligence claims involves one of the fundamental components of negligence: foreseeability. This consists of evaluating and determining whether the effect is

101 *Ibid.*

102 *Ibid.*

103 *Ibid.*

104 Catherine Sharkey, *Products Liability for Artificial Intelligence*, Lawfare (25 September 2024), <https://www.lawfaremedia.org/article/products-liability-for-artificial-intelligence> [https://perma.cc/J9JU-VERE].

105 *Ibid.*

106 *Ibid.*

107 *Ibid.*

108 Restatement (Second) of Torts § 282 (1965).

109 *Ibid.*

110 Restatement (Second) of Torts § 284 (1965).

111 Restatement (Second) of Torts § 282 cmt. c (1965).

too attenuated for fault to be found from the cause.¹¹² A different approach that could, albeit inadvertently, shield AI-system developers is an assessment of foreseeability applied against the user. One of the main problems with negligence law applied to AI can be seen when trying to understand what is foreseeable in light of an AI's output. This problem is more significant in consideration of AI claims due to the rapid evolution of the technology and the apparent magic box from where the outputs are generated.¹¹³ Some scholars argue that these AI-specific characteristics make it especially difficult for a plaintiff to show that the harm caused was a foreseeable result of the AI systems act, potentially preventing a plaintiff from showing that the AI system owed a duty of care.¹¹⁴ As AI becomes more integrated into society, questions arise over what a reasonable person would view as a foreseeable outcome to the use of certain AI systems and then whether courts would view that as eliminating the duty aspect necessary to bring a negligence claim.¹¹⁵

D Analysis of Regulations Around the World

I AI Regulations in California

Jurisdictions worldwide have been racing to implement relevant and adaptable AI regulations, aiming to create regulations that remain pertinent as the technology quickly evolves.¹¹⁶ Although some of these jurisdictions' approaches differ, efforts are being made to create unified frameworks that are applicable worldwide.¹¹⁷ California is an example of an American jurisdiction taking the lead in its attempt to implement AI regulation.¹¹⁸ They have a good reason, too – 35 of the top 50 AI companies worldwide are in California.¹¹⁹ The state aims to be the centre of this technological boom, but it recognizes that an unregulated growth of this incredibly complex technology could lead to unforeseen risks.¹²⁰ California has proposed numerous bills that cover a large amount of AI-focused regulation assigning responsibilities to AI companies, with some being vetoed, accepted or still pending.¹²¹

112 *Palsgraf v. Long Island R. Co.*, 162 N.E. 99, 104 (1928).

113 Andrew D. Selbst, *Negligence and AI's Human Users*, 100 B. U. L. Rev. 1315, 1315-76 (2020) (analysing the potential pitfalls of bringing a negligence claim against an AI system and suggesting that new negligence doctrines should be implemented to alleviate the issues associated with this complex technology).

114 See *ibid.*, at 1360-65.

115 *Ibid.*

116 AI Watch: *Global Regulatory Tracker*, White & Case LLP, <https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker#home> (accessed 19 January 2024) [https://perma.cc/WGG6-9846].

117 *Ibid.*

118 Cal. Exec. Order No. N-12-23 (30 September 2023), www.gov.ca.gov/wp-content/uploads/2023/09/AI-EO-No.12-_GGN-Signed.pdf [https://perma.cc/6V6C-4UWQ].

119 *Ibid.*

120 *Ibid.*

121 Chanley T. Howell & Leighton B. R. Allen, *Decoding California's Recent Flurry of AI Laws*, Foley: Innovative Technology Insights (4 October 2024), <https://www.foley.com/insights/publications/2024/10/decoding-california-recent-ai-laws/> [https://perma.cc/D9E2-ZMLS].

California's now vetoed *Safe and Secure Innovation for Frontier Artificial Intelligence Models Act*, S.B. 1047, was a significant bill that aimed to curb potentially harmful content earlier in the process stream by focusing partly on the development phase of an AI system.¹²² Some industry experts were concerned by the proposed statute's ambiguity and complexity.¹²³ Andrew Ng is an industry leader in AI who has spoken out about the collective industry concerns surrounding S.B. 1047 and similar bills.¹²⁴ In a blog post concerning the bill, Ng expressed concern over how the government would designate an AI model as "hazardous".¹²⁵ California suggested an approach that required an analysis of whether the AI system could inflict damages surpassing five hundred million dollars.¹²⁶ In his article, he also urges the governments to consider the application of the model and not the technology itself.¹²⁷ He explains that "[t]echnology refers to tools that can be applied in many ways to solve various problems. Applications are specific implementations of technologies designed to meet particular customer needs".¹²⁸ He encourages regulators to consider a more fact-specific and narrowly tailored approach to regulating AI instead of a broad-sweeping, technology-based regulation that he claims would stifle innovation.¹²⁹

The bill states, among other things, that a model cannot be held liable for a mass casualty event involving an AI if "information [that caused the harm] is otherwise reasonably publicly accessible by an ordinary person from sources other than a covered model...".¹³⁰ Additionally, the vetoed bill stipulates that the model would be implicated for "[a]cts with limited human oversight, intervention, or supervision".¹³¹ An additional major proposal in the bill required AI developers to include a shutdown procedure and the procedures and conditions under which this shutdown would occur.¹³²

Under this bill, an AI company must submit a disclosure detailing the types of critical harms and the magnitude of those harms for which the model may reasonably cause or enable before it allows public use of the model.¹³³ As before, some industry leaders chastise the ambiguity of such a requirement.¹³⁴ They ask, "[H]ow is a team of developers supposed to figure this out and declare – under

122 S.B. 1047, 2023-24 Leg. Reg. Sess. (Cal. 2024) (bill vetoed by the Governor of California).

123 Andrew Ng, *Issue 252, The Batch: Weekly Issues* (5 June 2024), <https://www.deeplearning.ai/the-batch/issue-252/> [<https://perma.cc/95E8-SGB9>].

124 *About*, Andrew Ng, <https://www.andrewng.org/about/> (accessed 1 October 2024) [<https://perma.cc/THE4-ZFRF>].

125 See Ng, *supra* note 123; see S.B. 1047, *supra* note 122.

126 Ng, *supra* note 123.

127 *Ibid.*

128 *Ibid.*

129 *Ibid.*

130 S.B. 1047, § 22602, 2023-24 Leg. Reg. Sess. (Cal. 2024).

131 *Ibid.*

132 *Ibid.*, at § 22603.

133 *Ibid.*

134 Andrew Ng, *Issue 257, The Batch: Weekly Issues* (10 July 2024), <https://www.deeplearning.ai/the-batch/issue-257/> [<https://perma.cc/4L9F-M985>].

penalty of perjury – that they meet this requirement?”¹³⁵ Under this requirement, any specific tests used in this “critical harm” assessment would be publicly available for five years to ensure a third party could replicate the evaluation.¹³⁶ Additionally, all covered AI model providers would be required to “retain a third-party auditor that conducts audits” to ensure compliance with the proposed bill’s various provisions.¹³⁷ This auditor would be allowed access to the AI company’s internal workings and must publish a redacted report detailing the company’s compliance with the bill.¹³⁸

A bill designed in part to prevent the spread of non-consensual sexual imagery is SB 926.¹³⁹ California’s Governor approved this bill in September 2024 to curb the distribution of intimate computer-generated images.¹⁴⁰ It outright prohibits the non-consensual distribution of images that appear to represent the victim.¹⁴¹ This bill expressly includes deepfake imagery in its ban by including in the description any image that “was created in a manner that would cause a reasonable person to believe the image is an authentic image of the person depicted”.¹⁴²

Senate Bill No. 942, California’s AI Transparency Act, was signed into law by Governor Newsom of California on 19 September 2024.¹⁴³ The bill’s primary focus is transparency for users of Gen AI services. Covered providers are now required to indicate that users are interacting with content created by a Gen AI platform.¹⁴⁴ This is achieved through providing a manifest disclosure or adding a latent disclosure within the content’s metadata.¹⁴⁵ Further, the bill requires that covered providers make available, at no cost, an AI detection tool that allows users to analyse content and determine if it was AI-created.¹⁴⁶

II Discussion of AI Laws in the EU

The EU has concentrated its efforts on a unified package that aims to foster AI innovation while ensuring that sectors utilizing AI are regulated with the appropriate level of care specific to each industry within reach of the Artificial Intelligence Act (AI Act).¹⁴⁷ In fact, this regulation is credited with being the world’s first major, binding law regarding the use and development of AI.¹⁴⁸ The Parliament hopes that

135 *Ibid.*

136 S.B. 1047, § 22603, 2023-24 Leg. Reg. Sess. (Cal. 2024).

137 *Ibid.*

138 *Ibid.*

139 S.B. 926, 2023-24 Leg. Reg. Sess. (Cal. 2024).

140 *Ibid.*

141 S.B. 926, § 647(j)(4)(A)(ii), 2023-24 Leg. Reg. Sess. (Cal. 2024).

142 *Ibid.*

143 S.B. 942, § 22757.3, 2023-24 Leg. Reg. Sess. (Cal. 2024).

144 *Ibid.*

145 *Ibid.*

146 *Ibid.*, at § 22757.2.

147 *EU AI Act: First Regulation on Artificial Intelligence*, Eur. Parl.: A.I., <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#ai-act-different-rules-for-different-risk-levels-0> [<https://perma.cc/XV72-R4TM>].

148 *Artificial Intelligence Act: MEPs Adopt Landmark Law*, Eur. Parl.: A.I., <https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law> [<https://perma.cc/7LXQ-GA66>].

this regulation sets Europe up to lead the path forward regarding AI and the regulations to follow.¹⁴⁹

A significant aspect of the AI Act involves separating specific AI systems into risk categories, each with a different regulatory scheme.¹⁵⁰ Systems that present unacceptable risks are grouped as “unacceptable” and are banned outright from use in jurisdictions adopting the AI Act.¹⁵¹ One example includes systems that employ subliminal techniques to manipulate individuals by distorting their decision-making ability.¹⁵² An example of this system would be a voice-activated toy that manipulates a child’s thinking into believing that a certain dangerous behaviour is safe.¹⁵³ Further, any system that autonomously exploits vulnerabilities found within a group of people, including “age, disabilities, or specific social or economic situations”, with the goal of materially distorting the victim’s behaviour would be prohibited.¹⁵⁴

Another example is the use of an AI system to create a social scoring system based on known or inferred knowledge of the individual, which can result in detrimental treatment to that person.¹⁵⁵ Specifically, the EU is concerned with discrimination and potential bias creation that would impair a number of “fundamental rights”.¹⁵⁶ Related to the concern of “fundamental rights” being impaired, using “real-time” biometric identification systems that could otherwise be utilized in public places would be prohibited except for a few very narrow exceptions for law enforcement.¹⁵⁷

The next category, “High-Risk”, involves, for example, AI systems implemented as a safety component within a product.¹⁵⁸ AI systems are not considered high-risk if they do not pose a risk to the “health, safety, or fundamental rights of natural persons, including by not materially influencing the outcome of decision making”.¹⁵⁹ AI systems that can manipulate, supersede or adversely influence human decision-making are to be considered high-risk.¹⁶⁰ Additionally, systems involving critical infrastructure like road traffic management or utility supply, as well as AI systems that categorize and make decisions regarding education, vocation, employee recruitment, employee advancement and anything similar would be considered high-risk.¹⁶¹ The AI Act goes on to explain that systems that generate

149 *Ibid.*

150 Eur. Parl.: A.I., *supra* note 147.

151 *Ibid.*

152 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonized Rules on Artificial Intelligence and Amending Regulations (EC) No. 300/2008, (EU) No. 167/2013, (EU) No. 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828, 2024 O.J. (L) 1 [hereinafter AI Act].

153 *Ibid.*

154 *Ibid.*

155 *Ibid.*, at 9.

156 *Ibid.*

157 *Ibid.*

158 *Ibid.*, at 53.

159 *Ibid.*, at 54.

160 *Ibid.*

161 *Ibid.*, at 14.

medical data, creditworthiness ratings or insurance decisions would be regarded as high-risk.¹⁶²

The next category of defined AI systems would be considered not-high-risk or general-purpose AI systems, and these systems are overseen with less regulation because of the relatively minimal impact they could have on a person.¹⁶³ AI systems designed to perform specific or narrowly tailored tasks, such as data summarization, would be considered general purpose.¹⁶⁴ Another example would be an AI system used to improve, but not materially modify, previously completed human activity.¹⁶⁵ Next, the AI Act says that if an

AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment, without proper human review, it would be considered general purpose.¹⁶⁶

Another critical component of the AI Act is that content created from a Gen AI system must be labelled to indicate that the product was machine-created or manipulated.¹⁶⁷ More generally, the law requires that AI systems notify individuals when interacting with a Gen AI derivative, even if a reasonable person may be unaware otherwise.¹⁶⁸ Interestingly, the AI Act also contemplates the extra notice required when dealing with disabled or elderly individuals who are especially at risk.¹⁶⁹

A 2022 proposal, commonly referred to as the AI Liability Directive (Directive), offers that no redress has been developed for individuals whom AI has specifically harmed.¹⁷⁰ Regarding the AI Act, the proposal states, “[t]hey do not provide for compensation of the injured person for damage caused by an output or the failure to produce an output by an AI system”.¹⁷¹ The Directive discusses the liability approach that the EU would adopt when an AI system causes actual harm to an individual.¹⁷² Like the liability question in the United States, the Directive discusses the traditional approach that may be taken when a cause of action arises.¹⁷³ When a victim seeks compensation, traditional legal mechanisms can be utilized to assign fault and seek compensation.¹⁷⁴ However, the proposal explains that “the specific characteristics of certain AI systems, such as opacity, autonomous behavior, and complexity”, may make the link between the harm and the cause a complicated link

¹⁶² *Ibid.*

¹⁶³ *Ibid.*

¹⁶⁴ *Ibid.*

¹⁶⁵ *Ibid.*, at 15.

¹⁶⁶ *Ibid.*, at 54.

¹⁶⁷ *Ibid.*, at 33.

¹⁶⁸ *Ibid.*

¹⁶⁹ *Ibid.*

¹⁷⁰ Commission Proposal for a Directive of the European Parliament and of the Council on Adapting Non-contractual Civil Liability Rules to Artificial Intelligence (AI Liability Directive), at 2, COM (2022) 496 Final (28 September 2022) [hereinafter The Directive].

¹⁷¹ *Ibid.*, at 16.

¹⁷² *Ibid.*, at 1.

¹⁷³ *Ibid.*, at 2.

¹⁷⁴ *Ibid.*, at 16.

to connect.¹⁷⁵ The Directive reinforces that the complexity of AI technology should not be a barrier to harmed individuals seeking accountability for damages.¹⁷⁶

The Directive suggests a “rebuttable presumption of causality” that requires an AI company to prove that the damage was not caused by the AI system’s act or omission.¹⁷⁷ The claimant must demonstrate several conditions for the “causal link between the fault of the defendant and the output produced by the AI system...” to apply.¹⁷⁸ First, the claimant must show that the AI system was non-compliant “with a duty of care laid down in Union or national law directly intended to protect against the damage that occurred”.¹⁷⁹ Second, the claimant must show that it is “reasonably likely” that the non-compliance with the statute was responsible for the harmful output or omission of the AI system.¹⁸⁰ Lastly, there must also be evidence that the damage was caused by the AI system’s act or omission.¹⁸¹

III Discussion of AI Laws in China

China has been at the forefront of AI research and is interested in setting regulatory precedents in the future, considering the global nature of AI.¹⁸² The Interim Measures for the Management of Generative Artificial Intelligence Service (Measures for Generative AI Services) is a regulation created by the Cyberspace Administration of China that aims to regulate most forms of Gen AI and provides obligations to the developers of those AI systems.¹⁸³ Similarly to the regulations in the aforementioned jurisdictions, there is an underlying strong desire to support AI development in China and prevent chilling of the industry’s growth.¹⁸⁴ This desire was so strong that a provision was written into the law itself:

The state is to adhere to the principle of placing equal emphasis on development and security, merging the promotion of innovation with governance in accordance with law; employing effective measures to encourage innovation and development in generative AI, and carrying out tolerant and cautious graded management by category of generative AI services.¹⁸⁵

¹⁷⁵ *Ibid.*

¹⁷⁶ *Ibid.*

¹⁷⁷ *Ibid.*, at 13.

¹⁷⁸ *Ibid.*, at 26.

¹⁷⁹ *Ibid.*

¹⁸⁰ *Ibid.*

¹⁸¹ *Ibid.*

¹⁸² Matt Sheehan, *China’s AI Regulations and How They Get Made*, Carnegie Endowment for International Peace: Asia (10 July 2023), <https://carnegieendowment.org/research/2023/07/chinas-ai-regulations-and-how-they-get-made?lang=en> [https://perma.cc/EZ62-2JD6].

¹⁸³ Interim Measures for the Management of Generative Artificial Intelligence Services (promulgated by the Cyberspace Admin. of China, 10 July 2023, effective 15 August 2023), <https://www.chinalawtranslate.com/en/generative-ai-interim/> [https://perma.cc/8WR2-DPAT] [hereinafter Interim Measures].

¹⁸⁴ *Ibid.*

¹⁸⁵ *Ibid.*, at Art. 3.

The regulation is surprisingly direct in its approach. Among other things, it says that when designing a Gen AI system, it must not discriminate against any protected class, religion or profession.¹⁸⁶ It includes broad, sweeping bans for creating “content that is prohibited by laws and administrative regulations”.¹⁸⁷ Further, AI systems must “[r]espect the lawful rights and interests of others [and] the physical and psychological well-being of others must not be endangered...”¹⁸⁸ It also says that Gen AI systems “must employ effective measures to increase transparency in generative AI services and to increase the accuracy and reliability of generated content”.¹⁸⁹ Importantly, this regulatory scheme does not apply to any organization during the development phase but focuses on the Gen AI system’s interaction with the general public.¹⁹⁰

AI system developers must only use “lawful sources” when building the foundation models driving the AI system.¹⁹¹ This law also pins liability on AI providers “as the producers of online information content”.¹⁹² Additionally, the law requires that the AI provider sign a service agreement with any individual who registers for their service.¹⁹³ It appears that if Gen AI produced a harmful output, China would take a more contractual approach to the remedies.¹⁹⁴

The bill also requires labelling and identifying content generated through the service.¹⁹⁵ It obliges providers to track illegal activity, report to the government the use of their systems for that illegal activity, and employ immediate measures to ensure compliance with the laws.¹⁹⁶ Lastly, the bill stipulates that any publicly accessible service must regularly perform security assessments to ensure law compliance.¹⁹⁷

E Proposed Solutions

I A Basic Regulatory Framework

California, the EU and China appear to agree that over-regulation would create a chilling effect across the industry. As previously discussed in this Note, the possibility of such incredible economic benefits should be a massive consideration for what regulations to impose.¹⁹⁸ To that end, creating a harmonized legal framework with well-defined requirements ensures that the AI systems it applies to can advance confidently, free from the uncertainties of an unequipped legal system.

186 *Ibid.*, at Art. 4.

187 *Ibid.*

188 *Ibid.*

189 *Ibid.*

190 *Ibid.*, at Art. 2.

191 *Ibid.*, at Art. 7.

192 *Ibid.*, at Art. 9.

193 *Ibid.*

194 *Ibid.*

195 *Ibid.*, at Art. 12.

196 *Ibid.*, at Art. 14.

197 *Ibid.*, at Art. 17.

198 Chui et al., *supra* note 23.

A common trend among the different jurisdictions discussed previously involves the potential for over-regulation of this new technology. Creating a regulatory scheme that is quickly adaptable, flexible to the wide span that AI covers, clear to stakeholders and intelligent in its approach is a complicated balance; however, creating a new Agency or Office solely tasked with making such a scheme is vital. Following a clear and thorough delegation of authority from Congress, this administrative body would first need to hire industry experts and legal minds to draft and propose guidelines that aim to protect this new and promising industry and the end users alike. To that end, synthesizing previously discussed regulations sheds light on what might succeed.

Like the AI Act and California's variety of statutes, these agency-developed rules would apply to different AI systems based on the risk levels attributed to those systems. Instead of creating a threshold based on computing power for whom these rules apply, similar to California's vetoed S.B. 1047, these rules would apply to any system that categorized itself as an algorithmic tool or, essentially, an AI system.¹⁹⁹ A tiered risk framework should be created parallel to the frameworks established in the EU's AI Act.²⁰⁰ These categories must be general enough to grow along with the AI industry properly but precise enough for courts to have a substantial basis when interpreting the rule.

Just as Section 230 was created in response to the explosive growth of the Internet to help its content providers avoid certain liabilities, AI may face the same hurdles that the Internet avoided if the AI industry does not receive comparable protections.²⁰¹ An essential component of this new regulatory package would be creating legal safe harbours for Gen AI developers who develop systems that abide by legal criteria. As we have seen with Section 230, this safe-harbour law allows developers to create innovative products with more legal security than would otherwise exist. Modifying Section 230 is vital to erasing the existing ambiguity surrounding the statute, and language should be inserted that clearly excludes Gen AI systems from the statute's reach. In its place, potentially included with the enabling act creating the agency or enabling a current agency, should be language that clearly stipulates under what conditions a new safe-harbour law could be utilized to the benefit of AI developers.

The EU's AI Act prohibits the public use of specific AI systems based on their inherent risks.²⁰² These include AI systems designed to manipulate or deceive individuals by exploiting known vulnerabilities like age, disability and nationality.²⁰³ Congressional action facially banning a particular type of content must, at minimum, show that the government has a substantial interest in prohibiting this material and that the regulation's purpose is one other than banning the speech

199 S.B. 1047, *supra* note 122.

200 Eur. Parl.: A.I., *supra* note 147.

201 Johnson & Castro, *supra* note 63.

202 AI Act, *supra* note 152, at 8.

203 *Ibid.*

but more geared towards protecting a class of people, as discussed in *Moody*.²⁰⁴ This Note suggests that the United States should take a similar approach in banning Gen AI systems that create the material defined previous. Additionally, Gen AI systems that create sexual imagery like revenge porn or any other sexual images created without the consent of a party should be prohibited, similarly to the proposed California bill, S.B. 926.²⁰⁵

Next would be a high-risk category, which includes safety-related AI systems that could adversely affect a person's health, safety or well-being or systems that can manipulate, supersede or adversely influence human decision-making. Critical infrastructure systems like road traffic management or utility supply would also be considered high-risk. Also included would be Gen AI systems that generate summaries of medical data, create creditworthiness ratings or make insurance decisions due to the risk of discrimination, bias and harmful potential outcomes. Consistent auditing of these systems would be a vital responsibility of the regulatory agency. Similarly to the concerns expressed in the AI Act, auditing these systems would help ensure that risks to health, safety and fundamental rights are mitigated.²⁰⁶ Further, agency audits and mandatory reporting would be required to help ensure that the technical opacity and mystery behind the technology is broken down. For systems placed within this category, users must be made aware that they are interacting with a Gen AI system through clear and apparent notifications and warnings. Notice and transparency are important features because of the potential for exploitation inherent in the systems within this category.

Following that would be low-risk AI systems. These would involve AI systems that could cause damage on failure; however, the risks are minimal compared to a high-risk or prohibited AI system. These AI systems would be subjected to the lowest regulatory scrutiny. This would include chatbots (ChatGPT, Gemini, CoPilot, etc.) and other assistive technology. Similarly to the EU's AI Act, the output of these systems would have significant human oversight. Additionally, AI systems that act to summarize data but do not substantially alter that data would be categorized as low-risk.

In each of these suggested categories there is the implicit concern of the suppression of free speech. The implications of the First Amendment's free speech provision are important considerations when enacting content regulations in the United States.²⁰⁷ As the Court in *Moody v. Netchoice, LLC* affirmed, a government's regulation of a private individual's speech implicates First Amendment concerns, which courts must analyse through either strict or intermediate scrutiny.²⁰⁸ Even at the lesser level of scrutiny, the government

204 *Moody v. NetChoice, LLC*, 603 U.S. 707, 740 (2024).

205 S.B. 926, *supra* note 139.

206 AI Act, *supra* note 152, at 54.

207 See Generally *Moody*, 603 U.S., at 743 (discussing Texas' attempt to regulate content posted on social media platforms through the guise of balancing ideological differences).

208 *Moody*, 603 U.S., at 740.

must show that the regulation serves “substantial governmental interest” and is “unrelated to the suppression of free expression”.²⁰⁹

II *Enforcement of Applicable Statutes or Rules*

One of the critical issues discussed in this Note centres around the difficulty in assigning fault due to AI’s complexity. To address those concerns, the EU proposes a rebuttable presumption of causality that presumes a causal link between the defendant’s fault and the AI system’s act or omission when certain conditions are met.²¹⁰ A scheme similar to that proposed in the EU would work in the United States because the same complexity and opacity cited as the general concerns surrounding claims brought against a Gen AI are universal.

Agencies across the country employ court-like adjudications to enforce federal rules.²¹¹ In many of these hearings, an Administrative Law Judge oversees court-like proceedings to ensure that the requirements of the Administrative Procedures Act and any other applicable statute are followed.²¹² An administrative agency explicitly created for AI regulation should be responsible for promulgating the previously discussed statutory scheme. Deference to an agency’s interpretation and expertise of a statute has been questioned with recent high-profile Supreme Court cases.²¹³ The reduced deference given to administrative agencies underscores the importance of a thorough and clearly defined delegation from Congress.

This Note proposes that the applicable administrative agency employ the rebuttable presumption of causation suggested in the EU’s AI Directive. This presumption of causation is similar to the negligence per se doctrine used by several state courts, where some courts accept a presumption of causation.²¹⁴ The Restatement Third of Torts explains that although this scheme is not found in all jurisdictions throughout the United States, it does reflect “some courts’ willingness to adapt the burden of proof depending on the type of tortious conduct and the difficulties of proof the plaintiff faces”.²¹⁵ This reasoning mirrors the AI Directives’ reasoning for the existence of that scheme.

According to the Third Restatement of Torts, negligence per se is defined as follows:

[a]n actor is negligent if, without excuse, the actor violates a statute that is designed to protect against the type of accident the actor’s conduct causes, and if the accident victim is within the class of persons the statute is designed to protect.²¹⁶

209 *Ibid.*

210 The Directive, *supra* note 170, at 13.

211 Ben Harrington & Daniel J. Sheffner, Cong. Rsch. Serv., R46930, Informal Administrative Adjudication: An Overview 2 (28 December 2023), <https://crsreports.congress.gov/product/pdf/R/R46930> [<https://perma.cc/6C72-KN8T>].

212 *Ibid.*

213 *See Loper Bright Enterprises v. Raimondo*, 603 U.S. 369 (2024).

214 Restatement (Third) of Torts: Liability for Physical Harm § 28 cmt. b (2005).

215 *Ibid.*

216 Restatement (Third) of Torts: Phys. & Emot. Harm § 14 (2010).

Hypothetically, in a jurisdiction that recognizes a presumption of causation, once the plaintiff sufficiently demonstrates that the AI provider's actions violated the applicable statute and that violation likely led to the alleged harm, the burden to prove otherwise could shift to an AI developer. This approach is similar to the EU's Directive, which says that a claimant must show that it is "reasonably likely" that the non-compliance with the statute was responsible for the harmful output or omission of the AI system.²¹⁷ The claimant must also present evidence that the damage was caused by the AI system's act or omission.²¹⁸

The complexities surrounding the precise language to be included in an enabling statute, including the agency's rules and sanctions, would need to be developed through intensive stakeholder studies and discussions. However, this Note suggests that the federal government should take the lessons learned from Section 230's success in affecting the growth of the Internet and implement similar safe-harbour laws in conjunction with a clear statutory framework that sets the basis for administrative agencies to enforce the law effectively.

F Conclusion

Section 230 of the CDA allowed the Internet to flourish by safe-harbouring content providers from litigation stemming from content that the provider itself did not create.²¹⁹ However, a regulatory void has surfaced as technology has outgrown this relatively simple statute. If it is left to the courts to decide whether Section 230 applies to Gen AI providers, the innovation this technology promises risks being stifled by a messy regulatory landscape.

Understanding the incredible benefits and potential risks this technology could create in society is vital when crafting a regulatory scheme. This Note discussed Section 230 of the CDA because of the legal questions its potential application to Gen AI brings and the analogous nature that the rise of the Internet has to the rise of Gen AI.

Numerous schemes worldwide have been created to regulate Gen AI, each reflecting unique attributes of the jurisdiction from where it came. Still, within each jurisdiction, there is a common goal: support the development of AI and protect people from the risks. Open and flexible regulations of Gen AI are required in the United States because this technology's potential is too high to regulate it into non-existence. Further, a sufficiently open and flexible regulatory scheme is necessary to prevent AI companies, small and large, from chilling their development efforts for fear of litigation. By creating a new federal agency or delegating that authority to an existing agency, this government body could engage industry experts and consumers alike to make effective regulations that adequately protect all stakeholders and allow this promising technology to flourish.

²¹⁷ The Directive, *supra* note 170, at 26.

²¹⁸ *Ibid.*

²¹⁹ Benson & Brannon, *supra* note 3, at 1.