

Restraining ChatGPT

ROEE SAREL[†]

ChatGPT is a prominent example of how Artificial Intelligence (AI) has stormed into our lives. Within a matter of weeks, this new AI—which produces coherent and humanlike textual answers to questions—managed to become an object of both admiration and anxiety. Can we trust generative AI systems, such as ChatGPT, without regulatory oversight?

Designing an effective legal framework for AI requires answering three main questions: (i) is there a market failure that requires legal intervention?; (ii) should AI be governed through public regulation, tort liability, or a mixture of both?; and (iii) should liability be based on strict liability or a fault-based regime such as negligence? Law and economics literature offers clear considerations for these choices, focusing on the incentives of injurers and victims to take precautions, engage in efficient activity levels, and acquire information.

This Article is the first to comprehensively apply these considerations to ChatGPT as a leading test case. As the United States is lagging in its response to the AI revolution, I focus on the recent proposals in the European Union to restrain AI systems, which apply a risk-based approach and combine regulation and liability. The analysis reveals that this approach does not map neatly onto the relevant distinctions in law and economics, such as market failures, unilateral versus bilateral care, and known versus unknown risks. Hence, the existing proposals may lead to various incentive distortions and inefficiencies. This Article, therefore, calls upon regulators to emphasize law and economics concepts in their design of AI policy.

[†] Junior Professor of Private Law and Law & Economics, Institute of Law and Economics, University of Hamburg. Comments may be directed to: roee.sarel@uni-hamburg.de. The Author thanks Danielle D’Onforo, Eric Langlais, Christine Susanne Rabe and other participants of the following conferences and seminars: American Law & Economics Association (2023), German Law & Economics Association (2023), French Law & Economics Association (2023), and Leibniz Business School (2023) for useful comments. The Author also thanks the Oxford Business Law Blog and the Columbia Law School’s Blue Sky Blog for featuring a post about the Article.

TABLE OF CONTENTS

INTRODUCTION	118
I. HOW CAN CHATGPT CAUSE HARM?.....	123
II. RESTRAINING AI: REGULATION OR LIABILITY?	128
A. MARKET FAILURE AS AN ECONOMIC PREREQUISITE FOR INTERVENTION.....	128
B. LIABILITY OR REGULATION: FOUR DETERMINANTS.....	129
C. LIABILITY AND REGULATION: FURTHER CONSIDERATIONS	133
III. LIABILITY OF AI: STRICT LIABILITY VERSUS NEGLIGENCE.....	134
A. UNILATERAL CARE	137
1. <i>Incentives to Take Precautions</i>	137
2. <i>Incentives to Choose an Efficient Level of Activity</i>	140
B. BILATERAL CARE: VICTIM’S INCENTIVES.....	141
C. RISK ATTITUDES AND INSURANCE	143
1. <i>No Market Insurance</i>	144
2. <i>Market Insurance</i>	144
3. <i>Victim’s Activity Level</i>	145
D. UNKNOWN RISKS	146
IV. THE EUROPEAN UNION’S PROPOSALS TO RESTRAIN ARTIFICIAL INTELLIGENCE.....	148
A. THE EUROPEAN UNION’S APPROACH TO AI – EARLY MILESTONES.....	148
B. THE AI ACT	149
1. <i>AI ACT: Original Proposal</i>	149
2. <i>AI ACT: Revisions Proposed by EU Entities</i>	153
C. THE AI LIABILITY DIRECTIVE	156
1. <i>The Directive, Generally</i>	156
2. <i>Scope and Content</i>	158
D. THE REVISED PRODUCT LIABILITY DIRECTIVE.....	160
E. RECENT PROPOSALS IN THE UNITED STATES	162
V. A LAW AND ECONOMICS EVALUATION OF AI RESTRAINTS	164
A. MARKET FAILURE, WHERE ART THOU?	165
B. SHALL LIABILITY AND REGULATION GO TOGETHER?	166
C. STRICT LIABILITY OR NEGLIGENCE?.....	168
1. <i>Unilateral versus Bilateral Care</i>	169
2. <i>Levels of Activity</i>	169
3. <i>Risk Attitudes and Insurance</i>	170
4. <i>Incentives to Research Unknown Risks</i>	171

D. PRESUMING A CAUSAL LINK	171
E. LEGAL UNCERTAINTY	172
CONCLUSION	172

INTRODUCTION

Artificial intelligence (AI) has swooped into our lives in a blink of an eye.¹ What was once considered well within the realm of science fiction now seems to manifest as reality. Autonomous vehicles,² medical diagnoses,³ fraud detection,⁴ robotics,⁵ image and speech recognition,⁶ and other AI-enabled functions, have all made substantial progress in recent years. One particular category of AI, however, now dominates the news cycle: generative AI systems.⁷

1. See, e.g., Cynthia Estlund, *What Should We Do After Work? Automation and Employment Law*, 128 YALE L.J. 254, 266 (2018) (noting how fast the technology of AI is advancing); Dorothy E. Roberts, *Digitizing the Carceral State*, 132 HARV. L. REV. 1695, 1695 (2019) (reviewing VIRGINIA EUBANKS, *AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR* (2018)) (“Big data, predictive analytics, and automated decisionmaking are used in every major type of state system, including law enforcement, national security, public assistance, health care, education, and child welfare.”); Andrea Roth, *Machine Testimony*, 126 YALE L.J. 1972, 1975 (2017) (discussing the reliance on machine evidence); Jonathan Gingerich, *Is Spotify Bad for Democracy? Artificial Intelligence, Cultural Democracy, and Law*, 24 YALE J.L. & TECH. 227, 229 (2022) (“Increasingly, we rely on algorithmic artificial intelligence (AI) to tell us what music to listen to, what movies to watch, what audio books to queue up, what groceries to buy, what stories, jokes, and photos to glance at on our phones, and with whom to go on dates.”); Alexander I. Platt, *Beyond “Market Transparency”: Investor Disclosure and Corporate Governance*, 74 STAN. L. REV. 1393, 1398 (2022) (describing “today’s world” as full of “algorithmic traders, machine learning, and robo-advisors”).

2. Andrew Guthrie Ferguson, *Digital Habit Evidence*, 72 DUKE L.J. 723, 752–53 (2023) (“[T]he rise of digital smart objects has led to several global trends including . . . ‘autonomous vehicles’ . . .”); see also Mariano-Florentino Cuéllar, *A Common Law for the Age of Artificial Intelligence: Incremental Adjudication, Institutions, and Relational Non-Arbitrariness*, 119 COLUM. L. REV. 1773, 1776 (2019) (discussing how common law should deal with AI, in particularly autonomous vehicles). For discussions on the balancing of liability and regulation for autonomous vehicles, see Jack Boeglin, Note, *The Costs of Self-Driving Cars: Reconciling Freedom and Privacy with Tort Liability in Autonomous Vehicle Regulation*, 17 YALE J.L. & TECH. 171, 176 (2015) (discussing how to reconcile liability with regulation of autonomous vehicles); Steven Shavell, *On the Redesign of Accident Liability for the World of Autonomous Vehicles*, 49 J. LEGAL STUD. 243, 283 (2020).

3. Harry Surden, *Artificial Intelligence and Law: An Overview*, 35 GA. ST. U.L. REV. 1305, 1316–17 (2019) (noting that AI enables “automated, expert-level diagnoses . . . (e.g., If patient has symptoms X and Y, the expert system, using its rules, determines that it is likely medical condition Z).”).

4. *Id.* at 1311 (“In essence, most machine-learning methods work by detecting useful patterns in large amounts of data. These systems can then apply these patterns in various tasks, such as driving a car or detecting fraud . . .”).

5. *Id.* at 1310 (“AI is truly an interdisciplinary enterprise that incorporates ideas, techniques, and researchers from multiple fields, including . . . robotics . . .”).

6. Ashley Deeks, *The Judicial Demand for Explainable Artificial Intelligence*, 119 COLUM. L. REV. 1829, 1832 (2019) (“[S]cientists and scholars often use the term [Artificial Intelligence] to encompass technologies that include machine learning, speech recognition, natural language processing, and image recognition.”).

7. For an overview of ChatGPT and generative AIs, see PHILIPP HACKER, ANDREAS ENGEL & MARCO MAUER, *REGULATING CHATGPT AND OTHER LARGE GENERATIVE AI MODELS* 1112 (2023).

As the name suggests, such systems enable users to generate convincing visual artwork,⁸ audio and video files,⁹ and humanlike textual outputs,¹⁰ all in a simple click of a button. The most conspicuous example is OpenAI's ChatGPT.¹¹

At first glance, ChatGPT may seem like a simple chatbot, providing answers to human questions in a chat-like manner. However, ChatGPT is no ordinary chatbot. Rather, it is a variant of the GPT (Generative Pre-training Transformer) model—a powerful AI tool that uses so-called neural networks to extract information and provide answers that feel natural, almost as if a human wrote them.¹² Upon its introduction, ChatGPT took the world by storm, reaching over one million users within five days of its launch in November 2022.¹³ The line of events that followed can only be described as a massive and disruptive tidal wave. Google declared a “code red” and introduced a competitor chatbot (“Bard”).¹⁴ Microsoft decided to incorporate ChatGPT into its Bing search engine and possibly into Microsoft Office tools.¹⁵ New York City quickly

8. See generally Kieran M. R. Hunt, COULD ARTIFICIAL INTELLIGENCE WIN THE NEXT WEATHER PHOTOGRAPHER OF THE YEAR COMPETITION? 108 (2022). Popular examples of generative AI include Dall-E-2, Midjourney, and Stable Diffusion. *DALL-E 2*, OPENAI, <https://openai.com/dall-e-2> (last visited Jan. 31, 2023); MIDJOURNEY, <https://www.midjourney.com/home/?callbackUrl=%2Fapp%2F> (last visited Dec. 18, 2023); *Stable Diffusion*, STABILITY.AI, <https://stability.ai/> (last visited Dec. 19, 2023).

9. HACKER ET AL., *supra* note 7 (mentioning Synthesia and MusicLM as two examples).

10. See generally Tamsin Maxwell & Burkhard Schafer, *Natural Language Processing and Query Expansion in Legal Information Retrieval: Challenges and a Response*, 24 INT'L REV. L. COMPUT. & TECH. 63, 63 (2010) (discussing that Natural Language Processing (NLP) refers to tasks that entail understanding or generating natural language, such as summarizing text, translating text to different languages, or supplying humanlike answers through a virtual assistant).

11. H. Holden Thorp, *ChatGPT is Fun, But Not an Author*, 379 SCI. 313 (Jan. 26, 2023), <https://www.science.org/doi/10.1126/science.adg7879> (describing ChatGPT as a “cultural sensation”); CHATGPT, <https://chat.openai.com/auth/login> (last visited Dec. 18, 2023).

12. For a general overview of GPT models, see Roberto Gozalo-Brizuela & Eduardo C. Garrido-Merchán, *ChatGPT is Not All You Need. A State of the Art Review of Large Generative AI Models* (Jan. 11, 2023) (unpublished manuscript) (on file with Cornell University).

13. See, e.g., Mehmet Firat, *What ChatGPT Means for Universities: Perceptions of Scholars and Students*, 6 J. APPLIED LEARNING & TEACHING 57, 58 (2023). For continuously updating statistics, see <https://www.demandsage.com/chatgpt-statistics>.

14. Nico Grant & Cade Metz, *A New Chat Bot is a ‘Code Red’ for Google’s Search Business*, N.Y. TIMES (Dec. 21, 2022), <https://www.nytimes.com/2022/12/21/technology/ai-chatgpt-google-search.html>; Dan Milmo, *Google Trials Its Own AI Chatbot Bard After Success of ChatGPT*, GUARDIAN (Feb. 6, 2023, 2:47 PM EST), <https://www.theguardian.com/technology/2023/feb/06/google-releases-its-own-ai-chatbot-bard-after-success-of-chatgpt>.

15. Johana Bhuiyan et al., *Microsoft Reportedly to Add ChatGPT to Bing Search Engine*, GUARDIAN (Jan. 5, 2023, 2:52 PM EST), <https://www.theguardian.com/technology/2023/jan/05/microsoft-chatgpt-bing-search-engine>; Luke Hughes, *Microsoft Now Wants to Add ChatGPT to Word, Email*, TECHRADAR (Jan. 9, 2023), <https://www.techradar.com/news/microsoft-now-wants-to-add-chatgpt-to-word-email>; Geoffrey A. Fowler, *Trying Microsoft’s New AI Chatbot Search Engine, Some Answers are Uh-Oh*, WASH. POST, <https://www.washingtonpost.com/technology/2023/02/07/microsoft-bing-chatgpt/> (Feb. 8, 2023, 6:33 PM EST).

prohibited the use of ChatGPT in schools due to concerns of misuse.¹⁶ Several university professors declared they intend to redesign their courses¹⁷ following the discovery that ChatGPT can pass exams in law schools,¹⁸ business schools,¹⁹ and even medical schools.²⁰ ChatGPT also caused turmoil in academic research, as researchers began crediting it with authorship, leading publishers to clarify that this practice was unacceptable.²¹ In parallel, a study showing fake abstracts ChatGPT wrote could fool scientists led to serious concerns for the future of peer review.²²

However, somewhat less attention has been given to ChatGPT's degree of accuracy. In a recent study, Hadar Jabotinsky and Roe Sarel illustrated how ChatGPT could be used to generate coherent answers to the types of questions scholars tend to ask, much like those analyzed in law reviews.²³ Although ChatGPT could produce clear answers, it suffered from one major flaw: some of the "academic sources" provided in the output were starkly inaccurate or flat-out nonexistent. While an experienced researcher could probably overcome such

16. Kalhan Rosenblatt, *ChatGPT Banned from New York City Public Schools' Devices and Networks*, NBC NEWS (Jan. 5, 2023, 2:16 PM PST), <https://www.nbcnews.com/tech/tech-news/new-york-city-public-schools-ban-chatgpt-devices-networks-rcna64446>. ChatGPT was also temporarily banned in Italy, due to privacy concerns. See Adi Robertson, *ChatGPT Returns to Italy After Ban*, VERGE (Apr. 28, 2023, 12:17 PM PDT), <https://www.theverge.com/2023/4/28/23702883/chatgpt-italy-ban-lifted-gdp-data-protection-age-verification>.

17. Kalley Huang, *Alarmed by A.I. Chatbots, Universities Start Revamping How They Teach*, N.Y. TIMES (Jan. 16, 2023), <https://www.nytimes.com/2023/01/16/technology/chatgpt-artificial-intelligence-universities.html>. But see Kevin Roose, *Don't Ban ChatGPT in Schools. Teach with It*, N.Y. TIMES (Jan. 12, 2023), <https://www.nytimes.com/2023/01/12/technology/chatgpt-schools-teachers.html> (arguing that ChatGPT should be embraced and used as an educational tool).

18. See generally Jonathan H. Choi, Kristin E. Hickman, Amy B. Monahan & Daniel Schwarcz., *ChatGPT Goes to Law School* 5 (Jan. 23, 2023) (unpublished manuscript) (on file with the University of Minnesota Law School) (last revised Oct. 20, 2023).

19. Samantha Murphy Kelly, *ChatGPT Passes Exams from Law and Business Schools*, CNN, <https://www.cnn.com/2023/01/26/tech/chatgpt-passes-exams/index.html> (Jan. 26, 2023, 1:35 PM).

20. Matt Delaney, *AI Language Bot ChatGPT Can Pass Parts of Business, Law and Medical Exams*, WASH. TIMES (Jan. 29, 2023), <https://www.washingtontimes.com/news/2023/jan/29/chatgpt-ai-language-bot-can-pass-business-law-and/>.

21. See generally Chris Stokel-Walker, *ChatGPT Listed as Author on Research Papers: Many Scientists Disapprove*, NATURE (Jan. 18, 2023), <https://www.nature.com/articles/d41586-023-00107-z>; James Vincent, *ChatGPT Can't Be Credited as an Author, Says World's Largest Academic Publisher*, VERGE (Jan. 26, 2023, 3:27 AM), <https://www.theverge.com/2023/1/26/23570967/chatgpt-author-scientific-papers-springer-nature-ban>; *Tools Such as ChatGPT Threaten Transparent Science; Here Are Our Ground Rules for Their Use*, NATURE (Jan. 24, 2023), <https://www.nature.com/articles/d41586-023-00191-1>; H. Holden Thorp, *ChatGPT Is Fun, but Not an Author*, SCIENCE (Jan. 26, 2023), <https://www.science.org/doi/10.1126/science.adg7879>.

22. Holly Else, *Abstracts Written by ChatGPT Fool Scientists*, NATURE (Jan. 12, 2023), <https://www.nature.com/articles/d41586-023-00056-7>. See also Brian Bushard, *Fake Scientific Abstracts Written by ChatGPT Fooled Scientists, Study Finds*, FORBES (Jan. 10, 2023, 1:44 PM EST), <https://www.forbes.com/sites/brianbushard/2023/01/10/fake-scientific-abstracts-written-by-chatgpt-fooled-scientists-study-finds/?sh=2f0b73bd18b6>.

23. Hadar Y. Jabotinsky & Roe Sarel, *Co-authoring with an AI? Ethical Dilemmas and Artificial Intelligence* (forthcoming) (manuscript at 29) (on file with the Arizona State Law Journal) (last revised Sept. 12, 2023).

a problem by verifying whether a source exists, this requires effort and possibly some degree of expertise. As the popularity of ChatGPT grows, an obvious concern arises: how can we prevent harm due to misleading (or otherwise problematic) outputs?

The question can be broken down into three sub-questions. First, is there a need for legal intervention at all? For instance, ChatGPT is based on a machine learning process. Its quality may already improve through this adaptive process swiftly enough without intervention that it will be sufficiently accurate before there has been a market failure justifying intervention. Second, if a market failure does occur, is this the type of problem best mitigated using public regulation, tort liability, or a mixture of both? Third, if liability is imposed, should the standard be based on strict liability or a fault-based regime, such as negligence?

The United States currently does not have comprehensive legislation to address these questions at the national level.²⁴ Thus, this Article focuses instead on a more comprehensive attempt to restrain AI by turning to proposals in the European Union (EU). In particular, the EU has been promoting three pieces of proposed legislation: a general regulation on AI (“AI Act”),²⁵ a directive focused on the tort liability of AI (AI Liability Directive (“AILD”)),²⁶ and a revised Product Liability Directive (“PLD”).²⁷

This Article is the first to comprehensively evaluate whether such proposals efficiently address the challenges of ChatGPT through the lens of law and economics.²⁸ This analysis yields several key insights.

24. H. MARK LYON, FRANCES WALDMANN, SAMANTHA ABRAMS-WIDDICOMBE, TONY BEDEL, IMAN CHARANIA, KEVIN KIM, EVAN KRATZER, BRENDAN KRIMSKY, EMILY LAMM & PRACHI MISTRY, GIBSON DUNN & CRUTCHER LLP, ARTIFICIAL INTELLIGENCE AND AUTOMATED SYSTEMS 2022 LEGAL REVIEW (Jan. 25, 2023) (noting that while the EU has launched a first attempt, “[t]here is (still) no comparable governance framework on the horizon in the U.S.”). This Article discusses the recent frameworks in the U.S. See *infra* Part V.E.

25. *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, at 1, COM (2021) 206 final (Apr. 21, 2021) [hereinafter *AI Act*].

26. *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 1, COM (2022) 496 final, (Sept. 28, 2022) [hereinafter *AILD*]. In addition to the AI Act and the AILD, the EU has other proposals on digital services in place that currently do not seem to apply. See HACKER ET AL., *supra* note 7, at 1113 (“We show that the EU’s prime instrument to combat harmful speech, the [Digital Services Act] . . . does not apply to [generative AI systems], creating a dangerous regulatory loophole.”).

27. *Proposal for a Directive of the European Parliament and of the Council on Liability for Defective Products*, at 1, COM (2022) 495 final (Sept. 28, 2023) [hereinafter *PLD*].

28. See *infra* Part V. For existing work on regulating ChatGPT, which does not focus on law and economics concepts, see HACKER ET AL., *supra* note 7, at 1114 (arguing that the EU’s proposals are over-inclusive and insufficiently address content moderation); Adam Zaremba & Ender Demir, ChatGPT: Unlocking the Future of NLP in Finance (Jan. 13, 2023) (unpublished manuscript) (on file with the Montpellier Business School) (an experimental demonstration of ChatGPT’s capabilities that demonstrates the need to regulate AI technology in

First, the EU's proposal to adopt a risk-based approach does not guarantee that regulation will apply only whenever there is a market failure. This reduces the economic justification for intervention.

Second, the EU's choice to combine liability and regulation may lead to ambiguous results in terms of efficiency. On the one hand, it may create synergies and avoid gaps that standalone regulation creates. On the other hand, the proposal ties liability with the defendant's regulatory obligations, which can lead to distorted incentives (for example, defendants over-complying with the regulation just to avoid liability on top).

Third, the AILD may lead to a de facto distinction between high-risk and non-high-risk AI systems when deciding the applicable liability regime—the former being subject to strict liability, whereas the latter is subject to a fault-based (negligence-like) regime. As this distinction does not necessarily reflect the theoretical considerations for choosing between the two regimes, there is a concern that additional inefficiencies will arise.

Finally, the EU's triad of proposals artificially separates harms caused by defective AIs and other AI-related harms, which can lead to uncertainty and further irrelevant distinctions. These various problems are especially worrisome given the well-known “Brussels Effect,”²⁹ which suggests that legal actions taken by the EU can have a global effect.

Thus, the Article calls upon policymakers to pay closer attention to law and economics in order to avoid a framework that neglects incentives.

The remainder of this Article is organized as follows: Part II explains what ChatGPT is and how it may cause harm. Part III discusses the choice between public regulation, tort liability, and their combination for ChatGPT. Part IV summarizes the main considerations in law and economics in favor of strict liability versus negligence and applies them to ChatGPT. Part V surveys the three proposals at the EU and compares them to some recent attempts to deal with AI in the United States. Part VI contains the heart of the argument, evaluating whether the existing proposals are compatible with the law and economics principles covered in the previous parts. Part VII concludes.

finance). Another recent working paper lists some arguments related to law and economics, but its focus and scope are very different. See Philipp Hacker, *The European AI Liability Directives – Critique of a Half-Hearted Approach and Lessons for the Future* (Cornell Univ., Working Paper No. 2211.13960, 2023).

29. See, e.g., Roe Sarel, Hadar Y. Jabotinsky & Israel Klein, *Globalize Me: Regulating Distributed Ledger Technology*, 56 VAND. J. TRANSNAT'L L. 435, 476 (2023). For a general discussion of the Brussels Effect, see Anu Bradford, *The Brussels Effect*, 107 NW. U.L. REV. 1, 3 (2012).

I. HOW CAN CHATGPT CAUSE HARM?

ChatGPT is a Large Generative AI Model (LGAIM), a technology trained using an enormous number of parameters.³⁰ More specifically, it is a Generative Pre-trained Transformer (GPT), a machine-learning model that is able to generate text after a process of “pre-training” on some textual database.³¹ The purpose of a GPT model is simple: to convincingly generate text that a human could have written.³² At the time of ChatGPT’s launch in November 2022, the most advanced GPT model was GPT-3, which was trained on approximately 175 billion parameters.³³ In March 2023, however, OpenAI—the firm behind ChatGPT—introduced a newer version (GPT-4), which is trained on a larger, undisclosed number of parameters.³⁴ ChatGPT now runs in two versions: a free version, based on a modification of GPT-3 (“GPT 3.5”) and a paid version based on GPT-4 (“ChatGPT Plus”). Importantly, both versions were trained on a set of texts published before the end of 2021.³⁵

The interface of ChatGPT resembles that of a regular chatbot, such that humans can enter queries in natural language (input) and receive an answer from the computer (output). Users can also give feedback (for example, clicking a “thumbs up” button), reflecting the concept of reinforced learning.³⁶ The

30. See, e.g., HACKER ET AL., *supra* note 7, at 1113 (stating that large language models are “usually trained with several billion, if not hundreds of billions, parameters . . .”).

31. *Id.*

32. Ben Lutkevich & Ronald Schmelzer, *Definition: GPT-3*, TECHTARGET, <https://www.techtargget.com/searchenterpriseai/definition/GPT-3> (Aug. 2023).

33. *Id.*

34. See, e.g., Drew Harwell & Nitasha Tiku, *GPT-4 Has Arrived. It Will Blow ChatGPT Out of the Water.*, WASH. POST (Mar. 13, 2023, 6:23 PM EDT), <https://www.washingtonpost.com/technology/2023/03/14/gpt-4-has-arrived-it-will-blow-chatgpt-out-water/> (“An earlier version of this story offered an incorrect number for GPT-4’s parameters. The company has declined to give an estimate.”); Tianyu Wu, Shizhu He, Jingping Liu, Siqi Sun, Kang Liu, Qing-Long Han & Yang Tang, *A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development*, 10 IEEE/CAA J. AUTOMATICA SINICA 1112, 1123 tbl. 1 (2023) (showing that the number of parameters for GPT-4 is “unpublished”). Another advantage of GPT-4 is that it was trained on images in addition to text. See generally Katharine Sanderson, *GPT-4 is Here: What Scientists Think*, NATURE (Mar. 30, 2023), <https://www.nature.com/articles/d41586-023-00816-5> (“GPT-4 was released on 14 March, and one upgrade is that it can now handle images as well as text.”).

35. See *GPT-4*, OPENAI, <https://openai.com/research/gpt-4> (last visited Sept. 29, 2023) (“GPT-4 generally lacks knowledge of events that have occurred after the vast majority of its data cuts off (September 2021) . . .”). *Model Index for Researchers*, OPENAI, <https://platform.openai.com/docs/model-index-for-researchers> (last visited Sept. 29, 2023) (“GPT-3.5 series is a series of models that was trained on a blend of text and code from before Q4 2021.”); David Baidoo-Anu, & Leticia Owusu Ansah, *Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning*, at 14–15 (Apr. 13, 2023) (unpublished manuscript) (on file with author) (“[A] cursory look at the ChatGPT-generated responses in this study reveals that ChatGPT has no idea of the world after 2021 and hence it could not add any references or information after 2021. This is because ChatGPT was trained with information only up to 2021.”).

36. Douglas L. Mann, *Artificial Intelligence Discusses the Role of Artificial Intelligence in Translational Medicine: A JACC: Basic to Translational Science Interview with ChatGPT*, 8 JACC: BASIC TRANSLATIONAL SCI. 221, 221 (2023) (referring to GPT-3.5 as an “AI text generator that uses Reinforcement Learning from Human Feedback to inform its language model.”).

algorithm can improve over time, even without further training on parameters, simply by incorporating the human response to the output.³⁷ Responses are generated through a process of sampling and mixing existing data using probability distributions to predict which phrases most likely fit together as an answer to the user's question.³⁸ Because the answer is probabilistic, and given that the data used to train ChatGPT may be imperfect, it is unsurprising that its output may be subject to bias and inaccuracy.³⁹

To illustrate, consider the following scenario: Jack is a lawyer writing a brief for a court case involving a good-faith purchaser of a stolen iPhone. As part of his argument, Jack seeks to compare the regime in the United States and Europe. He turns to ChatGPT with the following question:

"Please compare the law regarding good-faith purchases of mobile items in the United States and in Europe."

ChatGPT then responds as follows:

In the United States, the Uniform Commercial Code (UCC) governs the sale of goods, including mobile items such as vehicles. Under the UCC, a good-faith purchaser for value acquires clear title to the item, free from any defects or liens, as long as they did not have knowledge of or participate in any fraudulent activity related to the item.

In Europe, the laws regarding good-faith purchases of mobile items can vary by country. However, many countries in Europe have laws similar to the UCC. For example, the Sale of Goods Act in the United Kingdom provides that a purchaser in good faith and for value acquires a good title to the goods, free from any third-party claims.

In summary, both the US and many European countries have laws that protect good-faith purchasers of mobile items from defects and liens, as long as they did not have knowledge of or participate in any fraudulent activity related to the item.⁴⁰

A reader familiar with the legal material will directly notice the inaccuracy of the answer. While the law in much of Europe indeed grants good-faith buyers

37. See Wu, He, Liu, Sun, Liu, Han & Tang, *supra* note 34, at 1123 ("[R]esearchers use reinforcement learning with human feedback (RLHF) to incrementally train the GPT-3 model . . ."). Note that reinforced learning from human feedback is also part of the process of the initial training of the GPT-3 model; that is, the feedback comes also from individuals asked by the developers to provide feedback (and not only from end users). *Id.*

38. See HACKER ET AL., *supra* note 7, at 1113 (explaining the technical foundations of ChatGPT).

39. *Id.*

40. ChatGPT gave this exact answer when asking the question on January 20, 2023 (using the free version). A screenshot is provided as Figure 1 in the Appendix, *infra*. As a disclaimer regarding this particular quote: "The author generated this text in part with ChatGPT, OpenAI's large-scale language-generation model. Upon generating draft language, the author reviewed, edited, and revised the language to their own liking and takes ultimate responsibility for the content of this publication."

a clean property right,⁴¹ American law makes a strong distinction between cases where the contract is *void* and *voidable*.⁴² When contracts are void, such as when a thief tries to pass on stolen goods to a buyer,⁴³ the buyer will never acquire good title, even if the purchase was in good faith.⁴⁴ Conversely, when the contract is only voidable, such as if the seller is a minor, a good-faith purchaser can acquire such legal rights.⁴⁵

If Jack is an expert on property law, he would likely catch the inaccuracy and perhaps even ask ChatGPT to verify this point (in which case, it might).⁴⁶ However, if Jack's field of expertise were different—say, if Jack's main line of work was constitutional law—he might decide to take ChatGPT “at its word” and insert an incorrect statement into his brief, arguing that US law and European law are similar on this point.

What is the harm of such a mistake? Obviously, Jack himself may suffer a reputational loss if the judge reprimands him for misleading the court.⁴⁷ Such

41. ROBERT COOTER & THOMAS ULEN, *LAW AND ECONOMICS* 152 (6th ed. 2016) (“According to the rule in America, transferors can usually convey only those property rights that they legitimately have. . . . A different rule prevails in much of Europe, where the buyer acquires title by purchasing the good ‘in good faith.’”). Note that this rule does not apply in every European country. *See, e.g.*, Bürgerliches Gesetzbuch [BGB] [Civil Code] § 929, § 935, https://www.gesetze-im-internet.de/englisch_bgb/index.html (Ger.); *see generally* Arthur F. Salomons, *Good Faith Acquisition of Movables*, in *TOWARDS A EUROPEAN CIVIL CODE 1065* (Arthur S. Hartkamp, Martijn W. Hesselink, Ewoud Hondius, Chantal Mak, & Edgar Du Perron, eds., 4th rev. and exp. ed. 2011).

42. *See* Melissa Yang, *Void Versus Voidable Contracts: The Subtle Distinction That Can Affect Good-Faith Purchasers' Title to Goods*, 19 *NYLITIGATOR* 31, 33 (2014) (“As opposed to contracts for the sale of goods procured by economic duress and fraud, contracts for the sale of stolen goods are void. . . . At common law, a thief acquires no title to stolen goods and therefore passes no title to a good-faith purchaser of those goods.”); *see generally* John Henry Merryman, *The Good Faith Acquisition of Stolen Art* (Stan. L. Sch. John M. Olin Program L. & Econ., Working Paper No. 364, 2007); Ashton Hawkins, Richard A. Rothman & David B. Goldstein, *A Tale of Two Innocents: Creating an Equitable Balance Between the Rights of Former Owners and Good Faith Purchasers of Stolen Art*, 64 *FORDHAM L. REV.* 49, 50 (1995) (“Anglo-American law is well-settled that neither a thief nor a good faith purchaser from the thief, nor even subsequent good faith purchasers, can pass good title.”).

43. *See, e.g.*, Shyamkrishna Balganes, *Copyright and Good Faith Purchasers*, 104 *CALIF. L. REV.* 269, 283 (2016) (“To this day, possession through theft remains the primary, and perhaps only, instance of void title at common law, one that is incapable of protecting a good faith purchaser.”).

44. Hawkins et al., *supra* note 42, at 50. This is known as the *nemo dat* doctrine. *See, e.g.*, Yun-Chien Chang, *247 Jurisdictions in the World Get the Good-Faith Purchase Problem Wrong: A New Economic Framework*, (N.Y.U. Ctr. for L., Econ. & Org., Working Paper No. 19-25, 2019) (“American common law adopts the *nemo dat* doctrine—which basically extinguishes the good-faith purchase doctrine—and the Uniform Commercial Code stipulates only two exceptions: the voidable title rule and the entrustment rule.”).

45. U.C.C. § 2-403(1) (AM. L. INST. & UNIF. L. COMM'N 1988). For a discussion on the good-faith provision of the UCC, *see generally* Grant Gilmore, *The Good Faith Purchase Idea and the Uniform Commercial Code: Confessions of a Repentant Draftsman*, 15 *GA. L. REV.* 605, 605 (1981).

46. When asked for a clarification, ChatGPT provided a slightly more accurate answer, but still did not make the exact relevant distinction.

47. When finalizing this Article, a case somewhat similar to the example of Jack the lawyer emerged in real life: two lawyers apparently used ChatGPT to draft a court filing for a federal case and received an output

harm is somewhat un concerning to ChatGPT because it is already governed by the contractual relationship between Jack and ChatGPT through the terms of service. As of June 26, 2023, the terms of service indeed not only state that OpenAI (the creator of ChatGPT) bears no liability but also that the users of the chatbot must indemnify OpenAI for any third-party claims.⁴⁸

Thus, the more interesting question is what precisely are the third-party claims that may arise. In the example of Jack the lawyer, his client may decide to add ChatGPT to a malpractice lawsuit, blaming the AI itself for providing inaccurate results. As Jack's client does not have a contractual relationship with OpenAI, such a lawsuit would likely be based on a tort claim that points to the client's loss as the relevant harm.⁴⁹

More generally, however, whenever the output of ChatGPT generates some sort of "fake news" by making misleading information available to the public as part of a newspaper article or an academic paper, one could imagine a lawsuit that seeks to hold the creators of the AI liable for various harms. These harms may include spreading misinformation, which then induced people to make decisions based on inaccurate information.⁵⁰ In turn, this may lead to negative

which cited nonexistent cases (much like the hallucination of fake Articles; see Jabotinsky & Sarel, *supra* note 23, at 12. When this was discovered, the lawyers blamed ChatGPT. The judge eventually imposed a \$5000 fine on the lawyers, creating a monetary (in addition to reputational) harm. See Benjamin Weiser, *ChatGPT Lawyers Are Ordered to Consider Seeking Forgiveness*, N.Y. TIMES (June 22, 2023), <https://www.nytimes.com/2023/06/22/nyregion/lawyers-chatgpt-schwartz-loduca.html>; Associated Press, *Lawyers Submitted Bogus Case Law Created by ChatGPT. A Judge Fined Them \$5,000*, MYNORTHWEST (June 22, 2023, 3:16 PM), <https://mynorthwest.com/3901722/lawyers-submitted-bogus-case-law-created-by-chatgpt-a-judge-fined-them-5000/>.

48. *Terms of Use*, OPENAI (Mar. 14, 2023), <https://openai.com/policies/terms-of-use> (last visited June 26, 2023).

49. It is also possible that such a client would pursue a claim of unjust enrichment, pointing at profits made by AI at his expense. However, to simplify things, this Article omits this type of lawsuit from consideration. For a recent discussion of the unjust enrichment doctrine, see generally Maytal Gilboa, Yotam Kaplan & Roece Sarel, *Climate Change as Unjust Enrichment*, GEO. L.J. (forthcoming 2024) (manuscript at 37) (on file with the Georgetown Law Journal).

50. See, e.g., Emma M. Savino, Note, *Fake News: No One Is Liable, and That Is a Problem*, 65 BUFF. L. REV. 1101, 1103 (2017) ("Experts opine that the dissemination of fake news is "creating confusion, punching holes in what is true, causing a kind of fun-house effect that leaves the reader doubting everything, including real news.").

effects such as loss of trust in institutions,⁵¹ harm to reputation,⁵² and even incitement to violence.⁵³

A partial list of risks that AI creates in general can also be found in the recent Artificial Intelligence Risk Management Framework (“AI RMF”), published by the National Institute of Standards and Technology (“NIST”) of the Department of Commerce.⁵⁴ The AI RMF is a voluntary framework aiming to increase trust in AI systems.⁵⁵ Among other things, it provides examples of three types of harm caused by AI. First, AI can cause harm to *people*, such as infringing on civil liberties, endangering the psychological or physical safety of groups or individuals, or distorting democratic participation.⁵⁶ Second, AI can cause harm to *organizations*, for example, by interfering with business operations, facilitating security breaches, or causing harm to reputation.⁵⁷ Third, AI can cause harm to the “ecosystem,” for instance, by damaging interconnected resources, interrupting the global financial system, or causing environmental harm.⁵⁸

ChatGPT seems relevant to only a small subset of this list. It is clear how it may indirectly lead to the infringement of human rights (for example, if dictators use it to produce manifests) or to security breaches (for example, if it is misused to generate a hacking program), but other harms seem too disconnected to be considered here. For simplicity, this Article will mostly refer back to the example of Jack the lawyer throughout to illustrate the main points.

51. See, e.g., Katherine Ognyanova, David Lazard, Ronald E. Robertson & Christo Wilson, *Misinformation in Action: Fake News Exposure Is Linked to Lower Trust in Media, Higher Trust in Government When Your Side Is in Power*, 1 HARV. KENNEDY SCH. MISINFO. REV. 1, 1 (2020) (“One major concern about fake news is that it could damage the public trust in democratic institutions.”).

52. See, e.g., Alim Al Ayub Ahmed, Ayman Aljabouh, Praveen Kumar Donepudi & Myung Suh Choi, *Detecting Fake News Using Machine Learning: A Systematic Literature Review*, PSYCH & EDUC. J. 1932, 1932 (2021) (“There are different issues in this digital world. One of them is fake news. Someone can easily spread a fake news. Fake news is spread to harm the reputation of a person or an organization.”).

53. Rohan Goswami, *ChatGPT’s ‘Jailbreak’ Tries to Make the A.I. Break Its Own Rules, or Die*, CNBC, <https://www.cnbc.com/2023/02/06/chatgpt-jailbreak-forces-it-to-break-its-own-rules.html> (Feb. 8. 2023, 3:03 PM) (explaining that some users found a way to attain outputs from ChatGPT that incite violence by asking for hypothetical answers that another AI, less moral, would provide).

54. NAT’L INST. STANDARDS & TECH., NIST AI 100-1, ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0) (2023) [hereinafter AI RMF]. The AI RMF is also associated with the National Artificial Intelligence Initiative Act of 2020, which was enacted as part of the William M. (Mac.) Thornberry National Defense Authorization Act For Fiscal Year 2021. William M. (Mac.) Thornberry National Defense Authorization Act for Fiscal Year 2021, Pub. L. No. 116-283, § 5101, 134 Stat. 3388, 3415 (2021) [hereinafter NAIIA].

55. AI RMF, *supra* note 54.

56. *Id.*

57. *Id.*

58. *Id.*

II. RESTRAINING AI: REGULATION OR LIABILITY?

This Part discusses the choice between regulation and liability to impose restraints on ChatGPT. Subpart II.A briefly iterates the economic rationale for intervention: a market failure. Closely following the seminal article by Steven Shavell on this matter, Subpart II.B reviews the determinants for choosing between regulation and liability and applies them to ChatGPT.⁵⁹ Subpart II.C discusses further considerations when combining regulation and liability.

A. MARKET FAILURE AS AN ECONOMIC PREREQUISITE FOR INTERVENTION

While harm may be undesirable, not every harm requires legal intervention, at least from an economic perspective. The economic rationale for intervening in markets is the occurrence of a market failure.⁶⁰ Market failures can take on various forms,⁶¹ but the one most closely related to harm concerns externalities—positive or negative effects on individuals who are not in a position to bargain with the person responsible for the harm. For instance, when Jack the lawyer enters into an agreement with OpenAI, the interests of Jack’s clients are not directly represented, so the transaction entails externalities with respect to the clients. However, harm caused to Jack by ChatGPT is not, by itself, an externality.

Externalities are a typical justification for tort law.⁶² Specifically, the concern is that individuals who selfishly engage in risky activities (that may cause harm to others) do not directly care about the negative effect of their actions on others.⁶³ If the victim could easily contact the injurer, the two parties could negotiate a deal in which the injurer gets paid in exchange for ceasing the harmful activity. However, if the victim cannot effectively negotiate with the injurer, then any harm incurred by the victim constitutes a negative externality. Put differently, when high transaction costs prevent negotiations between the victim and the injurer, it is not guaranteed that the injurer’s decision as to

59. Steven Shavell, *Liability for Harm Versus Regulation of Safety*, 13 J. LEGAL STUD. 357, 358 (1984).

60. See, e.g., Hadar Y. Jabotinsky & Roece Sarel, *How Crisis Affects Crypto: Coronavirus as a Test Case*, 74 HASTINGS L.J. 433, 452 (2023).

61. *Id.* (“The literature on law and economics typically divides market failures into four different categories: excessive market power, information asymmetry, externalities, and behavioral market failures.”).

62. See, e.g., Michael Faure, Louis Visscher & Franziska Weber, *Liability for Unknown Risks - A Law and Economics Perspective*, 7 J. EUR. TORT L. 198, 200–01 (2016) (“The primary market failure experienced in tort law consists of a negative externality.”).

63. The usual assumption in neoclassical economics is that individuals are both rational and selfish, what economists refer to as a “homoeconomicus.” See, e.g., Jan-Philip Elm & Roece Sarel, *No Policy is an Island: Mitigating COVID-19 in View of Interaction Effects*, 48 AM. J.L. & MED. 7, 22 (2022). Consequently, injurers fail to internalize the cost of their actions on others unless there is liability. See, e.g., COOTER & ULEN, *supra* note 41, at 189–90 (“The economic purpose of tort liability is to induce injurers and victims to internalize the costs of harm that can occur from failing to take care. Tort law internalizes these costs by making the injurer compensate the victim.”).

whether (and how much) to invest in precautions would be optimal from a social perspective.⁶⁴ Thus, to overcome the externality, legal interventions can step in and take one of several possible forms, including safety regulation and tort liability.⁶⁵ The difference between the two is summarized by Shavell as follows:

Liability in tort and the regulation of safety represent two very different approaches for controlling activities that create risks of harm to others. Tort liability is private in nature and works not by social command but rather indirectly, through the deterrent effect of damage actions that may be brought once harm occurs. Standards, prohibitions, and other forms of safety regulation, in contrast, are public in character and modify behavior in an immediate way through requirements that are imposed before, or at least independently of, the actual occurrence of harm.⁶⁶

In other words, regulation concerns *public* authority measures in the form of rules or prohibitions, whereas liability delegates the enforcement of the norm to a *private* party, namely the plaintiff who files a lawsuit against the injurer.

B. LIABILITY OR REGULATION: FOUR DETERMINANTS

Given the concern that ChatGPT may cause harm that constitutes a negative externality, should it be regulated by a public authority, or should potential victims be granted the right to file a lawsuit if harm occurs? This classical question of regulation versus liability has kept scholars of law and economics busy for years. Steven Shavell offered four determinants for analyzing whether liability is more effective than regulation: (1) differential knowledge between the regulator and individuals,⁶⁷ (2) whether the injurers are

64. Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089, 1096 (1972) (arguing that liability rules, which are the standard in tort law, are superior to property rules whenever transaction costs are high). Robert Cooter and Thomas Ulen paraphrase Calabresi and Melamed's rule. COOTER & ULEN, *supra* note 41, at 100. ("Where there are obstacles to cooperation (that is, high transaction costs), the more efficient remedy is to award compensatory money damages.").

65. Peter N. Salib, *The Pigouvian Constitution*, 88 U. CHI. L. REV. 1081, 1082–83 (2021). Another common intervention to deal with externalities concerns taxation. By imposing a tax on activities that are inefficient, that is, when their social costs exceed their social benefits, the state can discourage the activity by charging a tax. Such taxes are known as Pigouvian taxes. *Id.* at 1084 ("Pigouvian tax is a fee imposed on an actor engaging in some activity, equal in amount to the expected social costs of that activity.").

66. Shavell, *supra* note 59, at 357.

67. *Id.* at 359.

solvent (“judgment-proofness”);⁶⁸ (3) whether injurers anticipate that they will evade liability because no lawsuit will be filed,⁶⁹ and (4) administrative costs.⁷⁰

The first determinant of *differential information* asks whether the regulator holds superior information over private parties (notably, the injurers) regarding the risky activities in question.⁷¹ If private parties have superior information on the activities they engage in—including their benefits, the potential harm to others, and the costs of precaution—then it is more efficient to let them make their own decisions.⁷² This is achieved through liability; the private parties know that if they cause harm, they might be sued. Respectively, in cases where the regulator holds superior knowledge, including special access to technical expertise or data, it is preferable to let the regulator decide on the best course of action.

The determinant of differential information usually supports liability over regulation.⁷³ But is this also the case for ChatGPT? While regulators dealing with AI may have access to some technology experts, AI creators probably have a relative informational advantage over regulators due to the fast-evolving nature of the technology—a gap known as the “pacing problem.”⁷⁴ In particular, AI creators probably have superior knowledge of the type of precautions available to them and the cost of implementing them into their specific code. Thus, the first determinant supports liability over regulation in the case of ChatGPT.⁷⁵

68. *Id.* at 360.

69. *Id.* at 363. The idea behind Shavell’s argument is that the injurer cares about the *expected* damages—the product of the probability of being sued and the magnitude of the damages. Shavell notes that if victims anticipate losing the case, they will be hesitant to sue, such that injurers “escape liability.” *Id.* Note that the same effect will generally arise also if the injurers anticipate that lawsuits would be filed but then end in a favorable verdict.

70. *Id.* at 363–64.

71. *Id.* at 359.

72. *Id.* (“Where private parties have superior knowledge of these elements, it would be better for them to decide about the control of risks, indicating an advantage of liability rules, other things being equal.”).

73. *Id.* (“Under liability, however, the outcome would likely be better.”).

74. See, e.g., Barry Friedman, Farhang Heydari, Max Isaacs & Katie Kinsey, *Policing Police Tech: A Soft Law Solution*, 37 BERKELEY TECH. L.J. 701, 117 (2022). (“Technological development today is happening ‘at an unprecedented pace,’ which makes it ‘harder than ever to govern using traditional legal and regulatory means’—a phenomenon commonly referred to as the ‘pacing problem.’”); Niels van Dijk, Simone Casiraghi & Serge Gutwirth, *The ‘Ethification’ of ICT Governance. Artificial Intelligence and Data Protection in the European Union*, 43 COMPUT. L. & SEC. REV. 1, 7 (2021) (“[L]egislation is often also depicted as too weak, too ‘thin’ and too slow to protect individuals in a rapidly developing technological world. One reading of the so-called ‘law-lag’ or ‘pacing problem’ is that changes in the technological landscape are so quick and profound, that regulations become rapidly ineffectual to properly protect individuals.”).

75. Shavell, *supra* note 59, at 359. Shavell also notes two cases where regulators tend to have an informational advantage, neither of which seem relevant here. Namely: (1) there is no reason to assume that regulators would have trouble communicating their information to AI creators, and (2) as each AI creator likely builds its algorithm, the concern that some “general information” would elude private parties because of a free-rider problem seems weak. *Id.* at 360.

The second determinant asks whether the potentially regulated entities can pay for the full magnitude of the harm.⁷⁶ In other words, the question is whether the injurer⁷⁷ is insolvent and therefore *judgment-proof*.⁷⁸ Determining whether AI creators are likely to be judgment-proof clearly depends on the case at hand. Is a rich company being sued or an individual creator? Is the harm massive or minor? And does the law entail loopholes that shield the defendant's assets from creditors? Providing a general answer to these questions seems difficult, but in the example of Jack the lawyer, the harms to the client are likely to be moderate or minor. However, insofar that the harm spreads to the general public due to fake-news-like issues, the harms could grow swiftly. Thus, the second determinant would support liability over regulation only if AI creators have sufficient resources such that the victim can recover.⁷⁹

The third determinant asks whether injurers face a threat of suit for harm they cause⁸⁰ or whether injurers anticipate that they will *escape a lawsuit*. Generally speaking, it is less likely that lawsuits will be filed in cases where (1) the harm is highly dispersed so that no single victim has enough incentive to sue; (2) the harm takes a long time to manifest, thereby causing the quality of the evidence to deteriorate; and (3) it is difficult to evaluate which entity caused the harm.⁸¹ Escaping lawsuits is usually an argument for regulation rather than liability.⁸² In the example of Jack the lawyer, the harm is concentrated with his client and occurs immediately, but the direct injurer is the lawyer, not ChatGPT. Hence, it is not obvious that the AI creators owe a duty of care to the client as an indirect victim of misinformation delivered to the lawyer, especially under the explicit contractual condition that the creator is not liable.

The same holds more generally when the harm is dispersed. If ChatGPT generates misinformation, the harm may trickle down in various directions, so

76. *Id.* at 360.

77. The terms “injurer” and “tortfeasor” are used synonymously throughout this Article to describe someone who has caused harm of any kind to a victim.

78. For a general overview of the judgement-proof problem, see Steven Shavell, *The Judgment Proof Problem*, 6 INT'L REV. L. & ECON. 45, 45 (1986); Mattias K. Polborn, *Mandatory Insurance and the Judgment-Proof Problem*, 18 INT'L REV. L. & ECON. 141, 141 (1998); Steven Shavell, *Minimum Asset Requirements and Compulsory Liability Insurance as Solutions to the Judgment-Proof Problem*, 36 RAND J. ECON. 63, 64 (2005).

79. Shavell notes an additional issue with judgement-proofness: if the injurer knows his liability is limited, he will have no incentive to purchase insurance. Shavell, *supra* note 59, at 361.

80. *Id.* at 363.

81. *Id.* (“One reason that a defendant can escape tort liability is that the harm she generates are widely dispersed, making it unattractive for any victim individually to initiate legal action. A second cause of failure to sue is the passage of a long period of time before harm manifests itself. This raises the possibility that by the time suit is contemplated, the evidence necessary for a successful action will be stale or the responsible parties out of business. A third reason for failure to sue is difficulty in attributing harm to the parties who are in fact responsible for producing it. This problem could arise from simple ignorance that a given harm or disease was caused by a human agency (as opposed to being ‘natural’ in origin) or from inability to identify which one or several out of many parties was the cause of harm.”).

82. *Id.* at 365.

identifying a specific duty of care toward any victim would be challenging. This situation resembles problems arising in other cases where the harm is dispersed, such as climate litigation. Moreover, even if the harm manifests immediately, the fact that the AI keeps learning may cause the evidence on its previous versions to deteriorate. By the time a lawsuit is filed, the AI may already learn the correct answer, so that harm could only be proven by some elaborate reverse-engineering process. Thus, the third determinant would seem to support regulation rather than liability, unless liability takes on some specific form that somehow overcomes these problems and makes the threat of lawsuits credible.

Finally, the fourth determinant concerns *administrative costs* incurred by private parties and the public.⁸³ Is it more costly to run an AI regulatory agency or to facilitate tort lawsuits against the creators of an AI? Generally speaking, the administrative costs of regulation tend to be higher than those of torts because regulation applies to all the regulated entities indiscriminately, whereas tort lawsuits are filed only when harms occur and tend to be directed at those who actually engage in dangerous activities.⁸⁴ Therefore, administrative costs are usually considered an argument for liability rather than regulation.⁸⁵

For the case of ChatGPT, regulation seems particularly costly because of the technology's cross-border nature, which creates a large set of problems. First, if any country attempts to dictate what AI is allowed to do, the creators might simply move their servers to another country, a problem known as "forum shopping," or "regulatory arbitrage."⁸⁶ Second, getting countries to collaborate on regulation is challenging due to incentive problems, such as the free-rider problem (where each country avoids investing regulatory effort because they prefer that others will do so for them).⁸⁷ Third, unless regulators can easily distinguish who is causing harm and who is not, the rules would impose compliance costs also on AIs that are already complying anyway. Thus, the fourth determinant would tend to support liability over regulation in the case of ChatGPT.

To summarize, three out of four determinants (differential information, judgment-proofness, and administrative costs) tend to support liability for

83. *Id.* at 363. For a similar argument, see also Steven Shavell, *A Fundamental Enforcement Cost Advantage of the Negligence Rule Over Regulation*, 42 J. LEGAL STUD. 275, 278–80 (2013).

84. Shavell, *supra* note 59, at 364.

85. *Id.*

86. See Elizabeth Pollman, *Tech, Regulatory Arbitrage, and Limits*, 20 EUR. BUS. ORG. L. REV. 567, 569 (2019); Wolf-Georg Ringe, *Forum Shopping Under the EU Insolvency Regulation*, 9 EUR. BUS. ORG. L. REV. 579, 580 (2008).

87. See Sarel et al., *supra* note 29, at 441 ("[T]here is also under-supply of efficient regulation, meaning a regulation that protects the entire (global) public. This occurs due to the free-rider problem: local regulators only have an incentive to adopt rules to protect the local set of users but have no reason to exert effort to protect the entire globe."). Another relevant problem is the tragedy of the commons, where the fact that the AI is a shared resource across border causes "over-consumption" by regulators, in the form of too much regulation. *Id.*

ChatGPT, whereas one determinant (escaping lawsuits) only supports that conclusion under specific conditions—conditional on liability being shaped in a way that overcomes the difficulties.

C. LIABILITY AND REGULATION: FURTHER CONSIDERATIONS

In many situations, regulation and liability co-exist. Consider, for instance, defective products. It is often the case that producers must comply with certain safety standards but are also exposed to tort lawsuits for product liability in case harm occurs.⁸⁸ Prima facie, one may wonder why the two should not always be combined, as “two are better than one.” Nonetheless, several of the considerations Shavell identified cast doubt on whether regulation and liability should always go hand in hand.

First, if both regulation and liability are effective, they could offset one another and become what economists call “strategic substitutes.”⁸⁹ For instance, suppose that the regulator knows that consumers can file a tort lawsuit if a product is faulty and assumes that the threat of a lawsuit already deters the producer. In this case, the regulator might decide to shirk its responsibility and rely on the deterrence effect of tort liability. Consequently, the producer is no longer facing the threat of regulation.

Second, if regulation does function properly, there is arguably little added value of liability on top,⁹⁰ as it certainly leads to some additional costs, including litigation costs, without much benefit.

Third, if liability is intertwined with regulation so that it is triggered only if a regulatory obligation was breached, incentive distortions could occur. Specifically, if compliance with regulation would exempt injurers from liability, they would have an incentive to only just comply.⁹¹ Thus, if there are heterogeneous injurers and a uniform standard, injurers that can cheaply prevent harm may nonetheless rely on minimum compliance with the regulation to exempt themselves from civil liability. Similarly, if noncompliance with the regulation by itself gives rise to liability, then some injurers would comply to avoid liability even though their compliance is undesirable.⁹² For example, this

88. Compare A. Mitchell Polinsky & Steven Shavell, *The Uneasy Case for Product Liability*, 123 HARV. L. REV. 1437, 1440 (2010) (arguing that one of the reasons why product liability may be unnecessary is the existence of strong regulations), with John C.P. Goldberg & Benjamin C. Zipursky, *The Easy Case for Products Liability Law: A Response to Professors Polinsky and Shavell*, 123 HARV. L. REV. 1919, 1920 (2010) (arguing for the importance of product liability law as deterrence).

89. See, e.g., Elm & Sarel, *supra* note 63, at 11.

90. Polinsky & Shavell, *supra* note 88, at 1441.

91. Shavell, *supra* note 59, at 365.

92. One solution to such a problem, proposed by Steven Shavell in another article, is to simply lower the regulatory standard, such that it would be lower than it would have been without added liability. See Steven Shavell, *A Model of the Optimal use of Liability and Safety Regulation*, 15 RAND J. Econ. 271, 271–72 (1984)

is the case when they have unusually high compliance costs that outweigh the social benefit.⁹³ A similar concern arises if potential injurers suspect that the courts might make mistakes and, in response, overinvest in precautions just to make it abundantly clear to the court that they were careful.⁹⁴

Fifth, whenever victims are consumers and injurers are producers, liability on top of regulation may indirectly hurt the victims in the form of higher prices.⁹⁵ Producers must price their products to also reflect the expected cost of litigation,⁹⁶ so any increase in the likelihood of being sued would lead to an increase in price as well. Alternatively, the additional expected costs of liability may push small companies out of the market, leaving only wealthy AI creators with high market power in the market.⁹⁷

Summarizing once more, while there may be good reasons to combine liability and regulation, it is not obvious that their combination is efficient. Rather, in some cases, it could backfire.

III. LIABILITY OF AI: STRICT LIABILITY VERSUS NEGLIGENCE

Technological advancements have sparked an ongoing debate on how to design a liability regime for AI. From a purely deontological perspective, holding AI creators liable for the actions of the autonomous machines they create is morally problematic, as the creators might be unable to predict how such machines will behave in the future.⁹⁸ There are also doctrinal issues, such as determining whether an AI is a “product,” and therefore subject to product

(“. . . [E]ither regulation or liability could turn out to be preferred when considered as an alternative to the other. But . . . it is often socially advantageous for the two means of controlling risk to be jointly employed . . . in this case, parties causing other than relatively low risks are led to do more than to satisfy the regulatory standard, for their potential liability makes that worth their while. . . . [Therefore,] it is socially desirable for the regulatory standard to be lower than if regulation was used alone.”).

93. Shavell, *supra* note 59, at 365.

94. See generally Richard Craswell & John E. Calfee, *Deterrence and Uncertain Legal Standards*, 2 J.L. ECON. & ORG. 279, 300 (1986).

95. Polinsky & Shavell, *supra* note 88, at 1441.

96. *Id.* (“In addition to the direct legal expenses of the product liability system, there are indirect costs generated by the effect of legal expenses on product prices. Specifically, because firms must raise the prices of their products to cover their legal costs, consumers are undesirably discouraged from purchasing goods.”). Polinsky and Shavell also mention two related points: (1) that the increase in prices holds some benefit by signaling the fact that it is a risky product and (2) that the higher prices induces consumers to purchase insurance that they would have avoided had the price been lower. *Id.* at 1459, 1468. As these seem less important for the case of ChatGPT, this Article does not expand further on these points.

97. But see Anat Lior, *AI Strict Liability Vis-à-Vis AI Monopolization*, 22 COLUM. SCI. & TECH. L. REV. 90, 106 (2020) (providing various arguments for why the concern of a monopolization by wealthy AI creators is not always convincing).

98. Andreas Matthias, *The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata*, 6 ETHICS & INFO. TECH. 175, 175 (2004).

liability,⁹⁹ or whether AI can have a legal personality.¹⁰⁰ Some scholars have offered to adapt the liability law concerning dangerous animals to include AI,¹⁰¹ but it has also been argued that courts may then still abstain from imposing liability because the harm was unforeseeable.¹⁰² Some scholars instead have proposed to impose a strict liability regime and pay out damages from a common fund to which AI creators would be forced to donate.¹⁰³ Others have suggested viewing AI as tools and holding the person using them liable because tools do not have a mind of their own.¹⁰⁴ A few scholars have raised the possibility of imposing criminal sanctions on AIs (which, for our purpose, would fall under regulation rather than liability), but also have acknowledged the difficulties of such a step, such as proving *mens rea* of AI.¹⁰⁵ Overall, there is no consensus on how liability rules for AI should be designed.¹⁰⁶

99. See, e.g., Woodrow Barfield, *Liability for Autonomous and Artificially Intelligent Robots*, 9 PALADYN, J. BEHAV. ROBOTICS 193, 196 (2018) (noting that the classification of AI as a product is an “evolving topic under contract law”).

100. See generally Paulius Čerka, Jurgita Grigiene & Gintarė Sirbikytė, *Liability for Damages Caused by Artificial Intelligence*, 31 COMPUT. L. & SEC. REV. 376, 377 (2015) (arguing that AI is, as of 2015, not a subject of law).

101. *Id.* at 197; Richard Kelley, Enrique Schaerer, Micaela Gomez & Monica Nicolescu, *Liability in Robotics: An International Perspective on Robots as Animals*, 24 ADVANCED ROBOTICS 1861, 1863 (2010).

102. Barfield, *supra* note 99, at 202 (“[T]he issue of foreseeability of the type of harm experienced (that is, were the damages foreseeable), will be a defense raised by the defendant in any strict liability tort action.”).

103. *Id.*

104. Čerka et al., *supra* note 100, at 384 (discussing the AI-as-Tools concept of liability).

105. See Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 554 (2015) (discussing *mens rea* of robots); John K.C. Kingston, *Artificial Intelligence and Legal Liability*, in RESEARCH AND DEVELOPMENT IN INTELLIGENT SYSTEMS XXXIII: INCORPORATING APPLICATIONS AND INNOVATIONS IN INTELLIGENT SYSTEMS 270, 271–72 (Max Bramer & Miltos Petridis eds., 2016) (analyzing whether criminal liability can be applied to AI); Alessio Azzutti, *AI Trading and the Limits of EU Law Enforcement in Detering Market Manipulation*, 45 COMPUT. L. & SEC. REV. 1, 12–14 (2022) (analyzing AI market manipulation as corporate crime, but also raising the possibility of deterring AIs ex-ante through code that implements a relevant mechanism).

106. See generally Emiliano Marchisio, *In Support of “No-Fault” Civil Liability Rules for Artificial Intelligence*, 1 SN SOC. SCI. 54, 57 (2021) (arguing in favor of strict liability); Anat Lior, Note, *AI Entities as AI Agents: Artificial Intelligence Liability and the AI Respondeat Superior Analogy*, 46 MITCHELL HAMLIN L. REV. 1043, 1102 (2019) (arguing in favor of strict liability); Matjaž Perc, Mahmut Ozer & Janja Hojnik, *Social and Juristic Challenges of Artificial Intelligence*, 5 PALGRAVE COMM’NS 1, 4–5 (2019) (mentioning that some legal scholars argue that when AI becomes sufficiently intelligent, one should consider negligence rather than strict liability); Gerald Spindler, *User Liability and Strict Liability in the Internet of Things and for Robots*, in LIABILITY FOR ARTIFICIAL INTELLIGENCE AND THE INTERNET OF THINGS 125, 137 (Sebastian Lohsse, Reiner Schulze & Dirk Staudenmayer eds., 2019) (proposing strict liability with a cap on damages); David C. Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117, 128–29 (2014) (arguing that strict liability should be imposed on the manufacturer if he is in the best position to prevent the harm, but liability should be shared with others otherwise); Alessio Azzutti, Wolf-Georg Ringe & H. Siegfried Stiehl, *Machine Learning, Market Manipulation, and Collusion on Capital Markets: Why the “Black Box” Matters*, 43 U. PA. J. INT’L L. 79, 127 (2021) (arguing that strict liability of AI in financial markets would not fit well with the rationales for market regulation). See also Christiane Wendehorst, *Strict Liability for AI and Other Emerging Technologies*, 11 J. EUR. TORT L. 150, 180 (2020).

Suppose that one nonetheless is interested in adopting a liability regime for ChatGPT, with or without regulation. What goals should such a liability regime strive for? From an economic perspective, an efficient liability regime for AI requires paying attention to the various costs it imposes when an injurer (accidentally) causes harm to a victim. Following a highly influential book by Guido Calabresi, it is customary to consider three types of cost-reductions in liability regimes.¹⁰⁷ Primary reductions aim to minimize the frequency and severity of accidents through deterrence of injurers.¹⁰⁸ Secondary reductions focus on the allocation of risk.¹⁰⁹ Tertiary costs concern the administrative costs of the liability regime.¹¹⁰ An efficient liability regime for AI aims to maximize the benefits that the technology brings.¹¹¹

However, from a practical perspective, the choice usually boils down to deciding on a liability standard.¹¹² Should such a liability regime take the form of negligence so that AI creators who prove that they took efficient precautions would be exempted from liability? Or should it take the form of strict liability so that AI creators would be liable for any harm caused, with no exemptions?

The choice between the two regimes has been a major topic of discussion in the law and economics literature, entailing a variety of arguments. The consensus seems to be that the choice involves tradeoffs and that each regime is superior in addressing some issues but inferior in addressing others.¹¹³ There are four main groups of arguments: (1) incentives to take precautions, (2) incentives

107. GUIDO CALABRESI, *THE COST OF ACCIDENTS* 26 (Yale Univ. Press, 1st ed., 1970).

108. *Id.* (“This cost, or loss, reduction goal can be divided into three subgoals. The first is reduction of the number and severity of accidents. This ‘primary reduction’ of accident costs can be attempted in two basic ways.”). *See also id.* at 27 (mentioning deterrence).

109. *Id.* at 27–28 (“The second cost reduction subgoal . . . concentrates instead on reducing societal costs resulting from accidents. . . . The secondary cost reduction goal can be accomplished through . . . a shifting of accident losses: the risk (or loss) spreading method and the deep pockets method.”). However, the subsequent literature mostly restricts attention to risk allocation. *See, e.g.*, Dominique Demougin, *Accident Liability with Rank Dependent Expected Utility: A Comment*, 90 KRITISCHE VIERTELJAHRRESSCHRIFT FÜR GESETZGEBUNG & RECHTSWISSENSCHAFT [KRITV] 172, 172 (2007) (Ger.) (“Since Calabresi (1970), law and economics scholars, when analysing liability rules, divide costs associated with accidents into primary, secondary and tertiary costs. Primary costs are associated with care effort and expected accident damages, secondary costs with risk (in the light of individual risk aversion) and tertiary costs with expenses of administering the system of deterrence and spreading risk.”).

110. CALABRESI, *supra* note 107, at 28 (“The third subgoal of accident cost reduction . . . involves reducing the costs of administering our treatment of accidents. It may be termed ‘tertiary’ . . .”). Recall that lower administrative costs were also a justification for choosing liability over regulation. Shavell, *supra* note 59, at 364.

111. Richard A Posner, *Book Review*, 37 U. CHI. L. REV. 636, 638 (1979) (reviewing GUIDO CALABRESI, *THE COST OF ACCIDENTS: A LEGAL AND ECONOMIC ANALYSIS* (1970)) (“Notice that the goal is not to minimize accidents or accident costs, unless by accident costs we mean costs net of any benefits.”).

112. Secondary and tertiary costs are sometimes neglected altogether. *See* Demougin, *supra* note 109, at 172.

113. *See, e.g.*, COOTER & ULEN, *supra* note 41, at 204 tbl.6.2 (showing how different combinations affect the incentives of the tortfeasor and the victim).

to choose an efficient level of activity, (3) implications of risk attitudes and insurance, and (4) incentives to acquire information when risks are not fully known.¹¹⁴ All these incentives must be evaluated for both the injurer and the victim, paying special attention to whether the scenario in question involves *unilateral care*, when only the injurer can reduce the likelihood of an accident, or *bilateral care*, when both the injurer and the victim can reduce the likelihood of an accident.¹¹⁵ This Article discusses each of these in turn, relying heavily on the summary of the existing theory in the standard book on law and economics by Thomas Ulen and Robert Cooter,¹¹⁶ alongside the work of others and my own examples.

Subpart III.A discusses the case of unilateral care. Subpart III.B considers bilateral care. Subpart III.C extends the discussion to the relationship between risk and insurance. Subpart III.D lays out the existing theory on unknown risks, such as those entailed in new technologies.

A. UNILATERAL CARE

1. *Incentives to Take Precautions*

When looking at primary cost reductions, the main point of interest lies in the incentives to take precautions when doing so is efficient. In the spirit of the famous “Hand Rule,”¹¹⁷ the question is whether the cost of taking precaution is lower or higher than the expected harm that would occur absent the precaution. If the precaution’s cost is lower than the expected harm, it is efficient for an injurer to adopt it. Respectively, precautions whose cost exceeds the expected harm should not be taken. This reflects the idea of efficient deterrence,¹¹⁸ that is, discouraging potential injurers from causing only those harms that can be cheaply prevented.

To illustrate how this works, consider a simple case where an injurer is facing a binary choice of whether or not to take some precaution. For example,

114. See CALABRESI, *supra* note 107, at 46, 110 (In Calabresi’s terminology, points (1), (2) and (4) relate to primary cost reductions, as they concern the frequency and severity of an accident, whereas point (3) relates to the secondary cost reduction).

115. Faure et al., *supra* note 62, at 201 (“Within tort law . . . one distinguishes . . . between unilateral (where only the tortfeasor affects the accident probability) and bilateral cases (where also the victim impacts the accident probability).”).

116. COOTER & ULEN, *supra* note 41, at 199–230.

117. The Hand Rule, also known as the “Hand Formula,” was developed by Judge Learned Hand. *United States v. Carroll Towing Co.*, 159 F.2d 169, 173 (2d Cir. 1947). It states that an injurer should not be held liable if their (marginal) cost of precaution exceeds the (marginal) benefit, which is calculated as the expected decrease in harm due to the precaution: “if the probability be called P; the injury, L; and the burden, B; liability depends upon whether B is less than L multiplied by P, i.e., whether $B < PL$.” See COOTER & ULEN, *supra* note 41, at 214.

118. See, e.g., Gary T. Schwartz, *Ethics and the Economics of Tort Liability Insurance*, 75 CORNELL L. REV. 312, 337 (1989) (illustrating efficient deterrence in torts using a numerical example).

an AI creator might need to decide whether or not to review the code for an extra hour before releasing the final version to the public. Taking the extra time would reduce the likelihood of misleading output by the AI but requires the creator to forgo another project for which he could have earned some extra cash. Suppose that reviewing the code would help avert a mistake that causes a harm of \$20,000, where the extra hour spent on review would make the mistake 1% less likely to occur. The expected benefit of the precaution is, therefore, \$200 (1% of \$20,000). Thus, it would be efficient for the creator to invest the extra hour as long he forgoes less than \$200 from the other project.

In this case, the difference between negligence and strict liability lies in the scenario where the harm still ends up occurring, notwithstanding the precautions taken. Then, the injurer pays \$20,000 damages under strict liability but is exempted from negligence. Both negligence and strict liability can lead to the same result here, at least when the standard for what constitutes negligence is set efficiently. Namely, an efficient negligence standard here requires holding the AI creator liable for the harm if the cost of precaution is lower than \$200 and exempting him otherwise.

Under negligence, the AI creator would be incentivized to take the precaution because doing so would be cheaper on average than risking a negligence lawsuit. The same outcome also occurs under strict liability. As the AI creator must always pay for the harm, he has an incentive to avoid the harm only when it is cheap to do so—and this is indeed the case whenever the cost of precaution is less than \$200. In other words, if it is efficient to take precautions, the AI creator will do so irrespective of whether he is facing a negligence regime or a strict liability regime.

As a more comprehensive numerical example, Table 1 illustrates the difference for a scenario where the AI creator has to forgo only \$100 to review the code. The first row summarizes the case where the creator *does not* review the code, which then leads to a harm of \$20,000 with an overall probability of 10%, so that the expected total social cost is \$2,000 (10% of \$20,000).¹¹⁹ In this case, where the creator *does not* take precautions, he is liable for this entire cost both under negligence and strict liability, so that his expected costs are \$2,000.

The second row, conversely, reflects the case where the AI creator *does* review the code, reducing the probability of harm by 1%. Under negligence, he is exempted from liability, so his costs are only \$100. Conversely, under strict liability, he also expects to pay damages with a probability of 9%, so he is facing expected damages of \$1800 (9% of \$20,000) plus the cost of precaution of \$100, and altogether \$1,900. But \$1,900 is still less than \$2,000; so even under strict

119. “Social cost” means here only the harm to the victim and the cost of precautions.

liability, he prefers to take precautions. Thus, in this scenario, both liability regimes yield the same result.

Injurer takes precaution	Cost of precaution	Harm	Probability of harm occurring	Injurer's expected cost		Expected Total social cost
				Negligence	Strict liability	
No	0	20,000	10%	2,000	2,000	2000
Yes	100	20,000	9%	100	1,900	1900

Table 1: Numerical Example: Efficient Precautions

However, slight adjustments of the assumptions can give an advantage for one regime over the other. For instance, setting an efficient negligence standard requires that the legislator holds all the relevant information, including the cost of precaution, the types of precautions one could take, the expected harm, the probabilities, and so on.¹²⁰ Unless the legislator has a differential knowledge advantage, it seems easier to impose strict liability and count on the AI creator to make his calculations according to the information to which he has better access.¹²¹ From this perspective, strict liability would seem to dominate negligence in the unilateral care case. As a counterexample, because a strict liability regime does not require courts to evaluate the level of care, courts can never update their beliefs about how injurers behave.¹²² Consequently, negligence can dominate strict liability to induce learning by the court specifically in those cases where the courts lack perfect information. In light of these arguments and many others, there is no consensus on which standard is

120. See, e.g., Shavell, *supra* note 59, at 359 (“Consider, for instance, the situation where private parties possess perfect information about risky activities of which a regulator has poor knowledge. Then to vest in the regulator the power of control would create a great chance of error. If the regulator overestimates the potential for harm, its standard will be too stringent, and the same will be the case if it underestimates the value of the activity or the cost of reducing risk. If the regulator makes the reverse mistakes, moreover, it will announce standards that are lax.”); Steven Shavell, *The Mistaken Restriction of Strict Liability to Uncommon Activities*, 10 J. LEGAL ANALYSIS 1, 12 (2018) (emphasizing that for strict liability there is no need for the court to have information, but for negligence, imperfect information may lead to inefficient standard setting); Stephen G. Gilles, *Negligence, Strict Liability, and the Cheapest Cost-Avoider*, 78 VA. L. REV. 1291, 1293 (1992) (discussing the different arguments and arguing in favor of the cheapest-cost-avoider as the most relevant criterion for choosing a regime).

121. Strict liability also eliminates the need to prove fault ex-post and hence yields lower administrative costs (Cf. STEVEN SHAVELL, *FOUNDATIONS OF ECONOMIC ANALYSIS OF LAW* 198 (Belknap Press, 2004) (arguing that the courts need to verify the activity, which creates a “substantial burden”).

122. See generally Eberhard Feess & Ansgar Wohlschlegel, *Liability and Information Transmission: The Advantage of Negligence Based Rules*, 92 ECON. LETTERS 63, 63 (2006).

better, even in the unilateral care case.¹²³ Instead, as mentioned, these different standards should be seen as involving tradeoffs.

2. *Incentives to choose an efficient level of activity*

A second point often discussed in the literature concerns the level of activity,¹²⁴ that is, the choice of the activity's intensity.¹²⁵ For instance, suppose that a potential injurer can construct one polluting plant, two polluting plants, or even one hundred plants. The more plants are built, the more harm will occur. For the level of activity, the literature generally finds an advantage for strict liability over negligence. Specifically, strict liability means that the injurer bears all of the social cost, so the more he pollutes, the more he pays. Conversely, negligence means that the injurer only pays for the cost of precaution. This raises a concern that the injurer will not directly care about the harm he causes, because so long as he is not negligent, he never pays for the harm.¹²⁶

To illustrate using an example closer to ChatGPT, suppose that the AI creator launches one or two algorithms, each of which can independently cause a harm of \$20,000 with a probability of 10% (so that the expected harm is \$2,000). Assume further that the creator earns \$2,500 from one algorithm but

123. See generally Dominique Demougin & Claude Fluet, *A Further Justification for the Negligence Rule*, 19 INT'L REV. L. & ECON. 33, 33 (1999) (showing how negligence rules can dominate strict liability if the tortfeasor is an agent); Henrik Lando, *Two Advantages of the Negligence Rule Over Strict Liability When the Parties Are Risk Averse*, 16 REV. L. & ECON. 1, 13 (2020) (identifying two advantages of negligence: inducing victims to bring forward information, and reduced transaction costs when insurance is imperfect); James A. Henderson Jr., *Why Negligence Dominates Tort*, 50 UCLA L. REV. 377, 396 (2002) (arguing that negligence generally dominates, e.g., because strict liability would illogically hold all those whose products are somehow involved in an accident jointly liable, even if none were at fault, without clear rules on how to adjudicate the case).

124. See Shavell, *The Mistaken Restriction*, *supra* note 120, at 13; COOTER & ULEN, *supra* note 41, at 211–13; Steven Shavell, *Strict Liability Versus Negligence*, 9 J. LEGAL STUD. 1, 2 (1980).

125. The assumption is that courts do not consider activity levels in their determination of what constitutes negligence. See, e.g., Stephen G. Gilles, *Rule-Based Negligence and the Regulation of Activity Levels*, 21 J. LEGAL STUD. 319, 320 (1992) (“The leading economic analysts of tort law have suggested, as an empirical matter, that courts employing a negligence standard ordinarily will consider parties’ levels of care—for example, how carefully someone drives a car—but not their levels of activity—whether, or how often, that person should drive at all.”). Gilles argues, however, that courts do take activity levels into account. *Id.*

126. See, e.g., Shavell, *Strict Liability Versus Negligence*, *supra* note 124, at 2 (“By definition, under the negligence rule all that an injurer needs to do to avoid the possibility of liability is to make sure to exercise due care if he engages in his activity. Consequently he will not be motivated to consider the effect on accident losses of his choice of whether to engage in his activity or, more generally, of the level at which to engage in his activity; he will choose his level of activity in accordance only with the personal benefits so derived. But surely any increase in his level of activity will typically raise expected accident losses (holding constant the level of care). Thus he will be led to choose too high a level of activity . . .”).

only \$1,000 from launching a second algorithm.¹²⁷ Table 2 summarizes the example.

Nr. Algorithms	Cost of precautions	Harm (with precautions)	Total cost	Total Revenues	Injurer's payoff (with precautions)	
					Negligence	Strict liability
0	0	0	0	0	0	0
1	100	2000	2100	2500	2400	100
2	200	4000	4200	3500	3300	-700

Table 2: Numerical Example: Efficient Activity Levels

In this example, it is more efficient for the AI creator to launch one algorithm because it yields \$400 more in revenues than in costs. Conversely, choosing two algorithms is inefficient: it yields a total revenue of \$3,500 but incurs a larger total cost of \$4,200. Comparing the injurer's payoffs in the last two columns reveals that he will behave differently under negligence and strict liability. Under strict liability, he will choose to launch only one algorithm (yielding a net income of \$100, which is higher than the alternatives). In this example, this is also the most efficient option, as it maximizes the joint sum of the social cost and the injurer's revenues.¹²⁸

Conversely, under negligence, the injurer does not care directly about the harm because he only needs to pay for the precautions. Consequently, he prefers to launch two algorithms (yielding \$3300 in net profits after paying for the precautions). This illustrates that negligence tends to induce inefficiently high activity levels.¹²⁹

B. BILATERAL CARE: VICTIM'S INCENTIVES

Conversely, under negligence, the injurer does not care directly about the harm because he only needs to pay for the precautions. Consequently, he prefers to launch two algorithms (yielding \$3300 in net profits after paying for the precautions). Suppose instead that the victim can also take precautions to reduce the likelihood of an accident. For instance, in the aforementioned example of

127. The lower earnings for the second algorithm reflect the typical concept in economics of diminishing marginal benefit. In other words, every additional product yields less profit than the previous one. See, e.g., David Gamage, *Preventing State Budget Crises: Managing the Fiscal Volatility Problem*, 98 CALIF. L. REV. 759, 779 (2010) ("Individuals receive diminishing marginal utility from nearly all forms of consumption.").

128. Note that choosing no algorithms yields a joint sum of \$0, choosing one algorithm yields a positive net sum of \$400 (= \$2500 - \$2100), and choosing two algorithms yields a negative net sum of -\$700 (= \$3500 - \$4200). This can be seen also in the last column of the table.

129. But see David Gilo & Ehud Guttel, *Negligence and Insufficient Activity: The Missing Paradigm in Torts*, 108 MICH. L. REV. 277, 317-18 (2009) (arguing that negligence can also lead to inefficiently low levels of activity).

Jack the lawyer, Jack himself can double-check the output of ChatGPT by turning to some external source. If we view Jack as the victim (rather than his client), we would want him to invest in precautions when it is efficient to do so, which occurs when the cost of precaution is lower than the benefit from reducing the likelihood of an accident.¹³⁰

Here, it does matter whether the injurer's liability regime is negligence or strict liability. If the injurer (the AI creator) is strictly liable for any harm, then the victim (Jack) is implicitly insured for the harm.¹³¹ That is, if there is no accident, there is no harm; and if there is an accident, the victim gets paid in damages. Consequently, Jack implicitly bears no risk. Thus, with strict liability, Jack (being the victim for the purpose of this example) has no incentive to take costly precautions because they would not yield any additional payment beyond the damages from the AI creator if an accident occurs.¹³² Conversely, under negligence, the AI creator is exempt from liability once he spends the extra hour reviewing the code. Then, if there is an accident that causes harm to Jack, there are no damages paid. Jack would then prefer to take precautions to avoid incurring harm. In other words, to maintain efficient incentives for the victim, negligence of the tortfeasor is superior to strict liability.¹³³

However, there is another path that achieves the same result even under strict liability by adding either contributory negligence or comparative negligence. Adopting a regime with a contributory negligence defense would imply that if the victim was negligent, then the injurer is no longer liable. This implies that Jack can only receive compensation if he takes efficient precautions. A similar result is achieved through comparative negligence, which means that Jack can only claim damages for harms he could not have efficiently prevented. To illustrate, suppose that the harm was \$1,000, and Jack could have invested \$50 to reduce it to \$800. Contributory negligence means that Jack will get zero, whereas comparative negligence would mean he can still claim \$800, as this cannot be reduced further. Thus, strict liability plus contributory or comparative

130. In cases where it is efficient that only one party invests in precautions—either the injurer or the victim—then the rule is that the least-cost avoider should bear the cost. *See, e.g.*, Eric Rasmusen, *Agency Law and Contract Formation*, 369 AM. L. & ECON. REV. 369, 380 (2004) (“The least-cost avoider principle, broadly stated, asks which party has the lower cost of avoiding harm, and assigns liability to that party.”).

131. *See, e.g.*, Harry E. Frech III, *State-Dependent Utility and the Tort System as Insurance: Strict Liability Versus Negligence*, 14 INT'L REV. L. & ECON. 261, 262 (1994) (“Under strict liability, plaintiffs are more likely to prevail and receive damage payments. In this view, strict liability provides better insurance than a negligence standard. . . . Many writers have commented on the efficiency of the implicit insurance provided by tort law.”). Note, however, that Frech also points out a common criticism that this insurance may not apply to the types of risks that consumers would usually want insurance for. *Id.*

132. *See* COOTER & ULEN, *supra* note 41, at 204–05.

133. *Id.* at 5 (referring to the problem on the implicit insurance function of strict liability as the “paradox of compensation,” and explaining that “[t]he solution to the paradox of compensation lies in a negligence rule. . . . [A] negligence rule can give efficient incentives to the victim and the injurer”).

negligence can overcome the disadvantage of pure strict liability concerning the victim's incentives.

Yet this solution only works if the standard for contributory (or comparative) negligence is set efficiently. For instance, if the court incorrectly determines that Jack would have to invest \$500, instead of \$50, to prevent the harm, then Jack would not be found negligent and would not take precautions. In this case, adding contributory or comparative negligence will not solve the problem.

Steven Shavell recently proposed an alternative solution in an article concerning AI, specifically on autonomous vehicles.¹³⁴ This solution decouples the injurer's strict liability from the victim's payoff. Instead of paying damages to the victim, damages are to be paid to the state.¹³⁵ Consequently, the victim no longer has implicit insurance and, therefore, will be incentivized to take precautions.¹³⁶ Shavell speculates that such a solution has not been implemented thus far because it reflects a strong deviation from fault-based liability to strict liability.¹³⁷ However, today, there are parallels for such a solution already. First, any regulation that imposes a fine proportional to the harm can be thought of as a strict liability with payment to the state. Second, the *cy pres* doctrine,¹³⁸ which enables courts to order damages to be paid to a charity, can work along similar lines.

C. RISK ATTITUDES AND INSURANCE

In the previous examples, we assumed that the AI creator and the victim are both risk-neutral, meaning that they only care about the expected harm but are unaffected by the fact that it is probabilistic. If we relax that assumption, things change on two fronts. First, we might be concerned with the insurance question—asking which liability regime best incentivizes the risk-averse parties to buy insurance. Second, if no market insurance is available, the question becomes who should be the residual bearer of harm: the injurer or the victim. Specifically, negligence turns the victim into the residual cost-bearer, whereas strict liability turns the injurer into the residual bearer.¹³⁹

134. Shavell, *supra* note 2, at 244–45. This article is mainly concerned with car accidents, so both parties are both victims and injurers. *Id.* at 243. This seems less relevant for ChatGPT.

135. *Id.* at 244–45.

136. See also Marie Obidzinski & Yves Oytana, *Prediction, Human Decision and Liability Rules 4* (Paris Ctr. for L. & Econ., Working Paper No. 2022-06, 2022) (arguing that negligence with liability sharing would be optimal, as it does induce the victim of an AI to take precautions).

137. Shavell, *supra* note 2, at 283.

138. See generally Rob Atkinson, *Reforming Cy Pres Reform*, 44 HASTINGS L.J. 1111, 1114 (1993).

139. See, e.g., Steven Shavell, *On Liability and Insurance*, 13 BELL J. ECON. 120, 121 (1982) (“Under strict liability injurers bear risk and victims are protected against risk, whereas under the negligence rule injurers do not bear risk—if they are not negligent, they will not have to pay damages when involved in accidents—and victims do bear risk.”).

To avoid a tedious list of the possible combination of risk attitudes across victims and injurers, let us assume that the AI creator is risk-neutral because it is a large limited liability company,¹⁴⁰ but the victim may be either risk-neutral or risk-averse. If the victim is risk-neutral, we are back to the examples above, so that no further necessary discussions are needed. But if the victim is risk-averse, we must consider whether he can buy insurance and whether that insurance is perfect or imperfect.

1. *No Market Insurance*

If there is no market insurance and the victim is risk-averse, he incurs some disutility from being the residual risk-bearer. To illustrate, suppose that the harm is \$20,000 and its probability of occurring is 10% so the expected harm is \$200 (10% of \$20,000). A risk-averse victim would value such a loss as larger than \$200 in terms of disutility because of the risk (of possibly incurring \$20,000). If the victim cannot sell his risk to an insurance company because there is no available insurance, then a risk-averse victim is the residual risk-bearer. This creates an additional cost. Hence, without insurance, there is an additional argument in favor of strict liability (of the injurer) over negligence because it provides the aforementioned implicit insurance function to the victim.¹⁴¹

2. *Market Insurance*

Suppose instead that the victim can simply buy insurance. Then, we no longer face a problem in terms of risk allocation. A risk-averse victim will simply pay a fee in exchange for transferring the risk to the insurance company.¹⁴² Thus, with perfect insurance, risk attitudes do not matter much for efficiency.

However, sometimes there is market insurance available, but it is imperfect in the sense that the insurance company cannot fully observe whether the insured party takes precautions. For instance, suppose that Jack could get some malpractice insurance that reimburses him if he must pay damages to his client. If the insurance company cannot verify whether or not Jack took precautions, there are a few additional concerns. First, Jack might not take precautions but then claim that he did, hoping that the insurance will not notice.¹⁴³ This is a

140. See Byungjoon Yoo, Vidyand Choudhary & Tridas Mukhopadhyay, *A Study of Sourcing Channels for Electronic Business Transactions*, 28 J. MGMT. INFO. SYS. 145, 156 (2011) (“[I]t is often assumed that large firms are risk neutral . . .”).

141. See, e.g., Shavell, *supra* note 139, at 121.

142. *Id.*

143. See, e.g., Shavell, *supra* note 59, at 361 (“What then is [the insured person’s] incentive to take care? The answer depends on whether insurers can easily determine risk-reducing behavior—so that they can link the premium charged or the other terms or conditions of coverage to the party’s precautions. Where this linkage can be established, the party’s incentive to take care should be tolerably good. But if insurers find it too costly to verify insureds’ efforts at risk-reduction, then their incentives to take care may be insufficient . . .”).

typical moral hazard problem, which occurs when the hidden actions of an agent affect the payoff of a principal.¹⁴⁴ Second, the insurance company may anticipate this problem and therefore refuse to insure Jack, returning things to the state of no insurance. Finally, Jack may try to take extra and unnecessary precautions just so that he can convince the insurance company that he was careful. These various extra costs are all wasteful. In light of these costs, strict liability, which leaves the injurer as the residual risk bearer, might again be preferable.

3. *Victim's Activity Level*

A final consideration concerns the victim's activity level. Much like the case of an AI creator who can launch several algorithms, the victim may decide between using AI for one search or several searches. This seems particularly relevant for ChatGPT because victims may use it for various purposes and based on multiple queries. For an efficient activity level of the victim, the victim should never be fully exempt from incurring the costs of an accident. Otherwise, even if the victim takes precautions when he uses ChatGPT, the number of queries (or the number of times they are relied upon) might be inefficiently high. The general intuition for this, as put forth by Robert Cooter and Thomas Ulen, is that:

Under each rule, the ultimate bearer of harm has incentives for an efficient activity level, whereas the party who escapes bearing the cost of accidental harm has incentives for an inefficient activity level.¹⁴⁵

What should a regulator concerned with inefficient activity levels of victims do? Cooter and Ulen clarify that the regulator must either (1) choose which activity level is more important and adopt a regime that is efficient for optimizing only one party's incentives (either the victim or the injurer) or (2) add an external measure to control the level of activity.¹⁴⁶ Thus, for ChatGPT, one must either focus on the incentives problems of one party (either the AI creator or the victim) or add an external measure—such as regulation. In other words, the tradeoff between negligence and strict liability provides an additional argument for why regulation and liability might be combined, notwithstanding the abovementioned difficulties.¹⁴⁷

144. Shavell, *supra* note 139, at 121 (“[M]atters are complicated if liability insurers cannot monitor injurer behavior, since problems of ‘moral hazard’ would then result in injurers’ purchasing only partial coverage.”). See also COOTER & ULEN, *supra* note 41, at 48 (“*Moral hazard* arises when the behavior of the insured person or entity changes after the purchase of insurance so that the probability of loss or the size of the loss increases.”).

145. COOTER & ULEN, *supra* note 41, at 212.

146. *Id.* at 213.

147. See *supra* Part II.C.

D. UNKNOWN RISKS

There is some debate on whether the concepts covered above are relevant if risks are unknown, which may well be the case for a rapidly changing technological tool such as ChatGPT.¹⁴⁸ Could AI creators do anything to prevent unknown harms? The answer largely depends on whether there are precautions that can potentially be relevant for various risks. Suppose that an AI creator can invest in cybersecurity software, reducing the hacking probability. The precise risk, if a hack occurs may be unknown, but the tougher security measures may prevent all sorts of malicious hacks. Thus, one could imagine that the incentive to invest in such a measure would be affected by the risk, even if it is not precisely known.

Similar questions can be raised for the level of activity, where the unknown risks could either have no effect (as one cannot base decisions on the unknown) or lead to a “chilling effect,”¹⁴⁹ where the potential injurer reduces their level of activity in the hope of avoiding the liability for an unknown risk.¹⁵⁰ In the worst-case scenario, the AI creator’s fear of liability would be so large that their level of activity would be reduced to zero, which means that entrepreneurs would avoid innovating altogether.¹⁵¹

Finally, if the risk is unknown to the injurer, but could be revealed by investments in information acquisition, liability might make sense as a mechanism to induce potential injurers to gather information. Thus, Michael Faure et al. argue that “[I]ability for unknown risks can induce the potential injurer to do more research, in order to find out if there are more risks than are currently known.”¹⁵²

However, they also point out that the incentive to collect information may not be the same under deterrence and strict liability.¹⁵³ Specifically, under strict liability, the injurer internalizes the entire social costs and benefits—including those involved in the search for information.¹⁵⁴ Thus, strict liability induces optimal incentives to research unknown risks.¹⁵⁵ This is not necessarily the case under negligence. Suppose the negligence standard is set to reflect only whether

148. See Faure et al., *supra* note 62, at 204–05.

149. See generally Jonathon W. Penney, *Understanding Chilling Effects*, 106 MINN. L. REV. 1451, 1456 (2022); Louis Kaplow, *Optimal Proof Burdens, Deterrence, and the Chilling of Desirable Behavior*, 101 AM. ECON. REV. 277, 277 (2011); Louis Kaplow, *Burden of Proof*, 121 YALE L.J. 738, 746 n.16 (2012).

150. Faure et al., *supra* note 62, at 205.

151. See *id.* at 211 (“[I]t is . . . possible that liability for unknown risks does not stimulate innovation, but actually hinders it. Fear of liability for possible hidden risks in these new products may result in firms not developing such new products. This could hence result in less innovation.”).

152. *Id.* at 206.

153. *Id.* at 207–10.

154. *Id.* at 207 (“[W]ith strict liability there is no difference between private benefits and social benefits: the injurer bears all costs and receives all benefits of additional research.”).

155. *Id.*

the injurer acted reasonably given the set of information that they already have (rather than the set they could have had, had they invested in gathering additional information). In that case, injurers may avoid gathering information on purpose.¹⁵⁶ For example, an AI creator might abstain from investing in a bug-detection tool to reveal new information about the risks, so that no one could later complain about why he did not take precautions to repair the bugs that would have been detected. Conversely, if the negligence standard is set such that the behavior of the uninformed injurer is judged in comparison to that of a fully informed injurer, this would give a strong incentive to invest in information acquisition.¹⁵⁷ Hence, a correctly set negligence standard can also induce injurers to find out what is “state of the art” concerning technological risks.¹⁵⁸

In light of the above, Faure et al. further argue in favor of combining regulation and liability for unknown risks, aiming to achieve a “smart mix” between the two.¹⁵⁹ Their argument closely follows the four determinants discussed in Part II.B and mostly focuses on the issue of information. Specifically, they argue that the regulators’ ability to attain an informational advantage may be hindered by the private interests of public officials, along the lines of public choice theory.¹⁶⁰ In particular, when politically influential firms launch new technologies, there is a concern of regulatory capture,¹⁶¹ which would frustrate the regulators’ incentives to discover all the relevant risks. Consequently, unknown risks provide an additional argument for implementing liability together with regulation rather than sticking to only regulation.

Of course, the above analysis is simplified. Among other things, it assumes that there is only one injurer who can acquire relevant information. Yet even in the simple example of Jack the lawyer, three parties can potentially gather information: the AI creator (OpenAI), the AI’s user (Jack), and the clients. More general criteria is needed to determine who precisely should be liable as the development of AI technology may generally include a multitude of entities (for example, training data supplier, AI producers, humans involved in feedback

156. *Id.* at 208–09.

157. The analysis can also be extended to a scenario where the victim can invest in information acquisition. In this case, the principle of the least-cost-information-gatherer would apply, so that the choice of liability will depend on who can gather the information more cheaply. *See, e.g.,* Ann-Sophie Vandenberghe, *Employment Contracts*, in *ENCYCLOPEDIA OF LAW AND ECONOMICS* 61, 83 (Gerrit De Geest ed., 1st ed. 2009) (“The central insight from the law and economics literature on pre-contractual information is that efficient rules have the effect that information is produced and revealed by the least-cost information gatherer.”).

158. Faure et al., *supra* note 62, at 216.

159. *Id.* at 216–19.

160. *Id.* at 219. Public choice theory assumes that public officials are self-interested and therefore maximize their own private utility rather than social welfare. *Id.*

161. For a general discussion of regulatory capture and its various definitions, see David Freeman Engstrom, *Corralling Capture*, 36 *HARV. J.L. & PUB. POL’Y* 31, 31–33 (2013). For a discussion in the context of technology, see, for example, Eva Micheler & Anna Whaley, *Regulatory Technology: Replacing Law with Computer Code*, 21 *EUR. BUS. ORG. L. REV.* 349, 363 (2020).

loops, etc.). Here, law and economics propose the least-cost information gatherer principle: whomever can get the information more cheaply should be incentivized to invest in information acquisition.¹⁶² Thus, if the AI creator has easier access to information on how the algorithm works compared to Jack, then the creator should be held liable in case he does not acquire the relevant information. Common wisdom in law and economics offers a similar concept whenever multiple parties can prevent the harm in the form of the least cost avoider principle.¹⁶³ Whoever can prevent the harm caused by ChatGPT at the lowest cost should be induced to do so.¹⁶⁴

IV. THE EUROPEAN UNION'S PROPOSALS TO RESTRAIN ARTIFICIAL INTELLIGENCE

This Part provides an overview of the most recent attempts to restrain AI at the EU, in the form of three pieces of proposed legislation: (1) the AI Act, (2) the AI Liability Directive (“AILD”), and (3) the revised Product Liability Directive (“PLD”).¹⁶⁵

Subpart IV.A summarizes the main early milestones of the EU’s road to create a framework for AI. Subpart IV.B reviews the AI Act. Subpart IV.C delves into the AILD. Subpart IV.D covers the PLD. Finally, Subpart IV.E adds a brief comparison between the approach in the EU and the initial ideas reflected in the recent proposals released in the United States.

A. THE EUROPEAN UNION’S APPROACH TO AI – EARLY MILESTONES

In March 2018, the European Commission decided to set up a working group of experts whose task was to draft a proposal for ethical guidelines for

162. Cf. Ann-Sophie Vandenberghe, *The Role of Information Deficiencies in Contract Enforcement*, 3 ERASMUS L. REV. 71, 76 (2010) (“From an economic perspective, rules should be designed to ensure that the party who can produce and reveal information at least cost produces the optimal amount of honest information in society. Economically optimal rules give incentives to the *least-cost information gatherer* to produce and communicate information.”) (emphasis in original).

163. See, e.g., Miriam Buiten, Alexandre de Streel & Martin Peitz, *The Law and Economics of AI Liability*, 48 COMPUT. L. & SEC. REV., Apr. 2023, at 1, 11 (“When multiple parties affect the risk of harm, the question arises of who should be targeted by the liability rule. From a welfare perspective, this should be the least-cost avoider, i.e., the party which can minimise harm at the lowest cost.”). For a discussion of the least-cost avoider (“cheapest cost avoider”) principle, see generally Gilles, *supra* note 120; Guido Calabresi & Jon T. Hirschoff, *Toward a Test for Strict Liability in Torts*, 81 YALE L.J. 1055 (1972).

164. *But cf.* Massimo D’Antoni & Avraham D. Tabbach, *Inadequate Compensation and Multiple Equilibria*, 38 INT’L REV. L. & ECON. 33 (2014) (showing that in bilateral care cases, holding the least-cost avoider liable may be insufficient for deterrence unless the victim is fully compensated).

165. See *AI Act*, *supra* note 25; *AILD*, *supra* note 26; *PLD*, *supra* note 27. For a general overview, see Hacker, *European AI Liability Directives*, *supra* note 28; PAWEŁ KSIĘŻAK & SYLWIA WOJTCZAK, *Liability of AI, in TOWARD A CONCEPTUAL NETWORK FOR THE PRIVATE LAW OF ARTIFICIAL INTELLIGENCE* 239–90 (2023).

AI.¹⁶⁶ Subsequently, the Commission gradually released various plans related to AI. First, a three-pronged plan was devised to “increase public and private investment in AI, prepare for socio-economic challenges, and ensure an appropriate ethical and legal framework.”¹⁶⁷ In parallel, the Commission released a working document addressing liability for emerging digital technologies.¹⁶⁸ This document, among other things, reviewed the principles that one usually finds at the EU level and the member-state level concerning the liability of new technologies. At the EU level, new technologies used to fall under the scope of a general product liability directive, which was “conceived around the notion of movable products, most of which are tangible.”¹⁶⁹ The working document notes that the existing product-liability directive sets a strict liability regime. This regime contrasts with most national jurisdictions, which generally tend to use fault-based liability for AI—with only some countries applying “special regimes of strict liability.”¹⁷⁰ The working document was followed by the establishment of a “European AI Alliance” in June 2018.¹⁷¹ The Alliance’s role was to engage with different interest groups, including trade unions, scholars, and regulators, to understand the implications of AI better. Later that year, the Commission launched a “Coordinated Plan” on AI,¹⁷² which was followed by a variety of interim updates and reports throughout the years 2019–2020.¹⁷³ The most significant step occurred in April 2021, with the release of the AI Act, the first comprehensive proposal to regulate AI.

B. THE AI ACT

1. *AI ACT: Original Proposal*

The AI Act is a 108-page proposal for new regulation of AI, seeking to harmonize the approach across the EU. Its stated objectives are fourfold.¹⁷⁴ First,

166. European Commission Press Release IP/18/1381, Artificial Intelligence: Commission Kicks off Work on Marrying Cutting-Edge Technology and Ethical Standards (Mar. 9, 2018).

167. European Commission Press Release IP/18/3362, Artificial Intelligence: Commission Outlines a European Approach to Boost Investment and Set Ethical Guidelines (Apr. 25, 2018).

168. *Commission Staff Working Document on Liability for Emerging Digital Technologies*, SWD (2018) 137 final (Apr. 25, 2018).

169. *Id.* at 6.

170. *Id.* at 8.

171. EUR. COMM’N, *The European AI Alliance*, <https://digital-strategy.ec.europa.eu/en/policies/european-ai-alliance> (Sept. 14, 2023).

172. *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions*, COM (2018) 795 final (July 12, 2018).

173. For an overview, see European Commission, A European Approach to Artificial Intelligence, SHAPING EUROPE’S DIGITAL FUTURE, <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence> (June 19, 2023).

174. *AI Act*, *supra* note 25, Explanatory Memorandum at 3.

ensuring that AI systems are “safe and respect existing law on fundamental rights.”¹⁷⁵ Second, “ensuring legal certainty to facilitate investment and innovation.”¹⁷⁶ Third, enhancing governance and enforcement. Fourth, facilitating the development of a market for “lawful, safe and trustworthy AI applications” and preventing fragmentation.¹⁷⁷

To achieve these goals, the AI Act covers a wide range of AI applications, including any software that is developed using technologies such as machine learning, which is clearly applicable to ChatGPT.¹⁷⁸ In terms of content, the AI Act adopts a *risk-based approach*, distinguishing between risks that are (1) unacceptable, (2) high, (3) limited, and (4) minimal.

For *unacceptable risks*, the act simply bans the technology.¹⁷⁹ These include, first and foremost, intuitively-bad practices, such as subliminal techniques to distort behavior in a manner that causes physical or psychological harm.¹⁸⁰ It further bans technologies exploiting the vulnerabilities of weaker groups, such as minors or the mentally disabled.¹⁸¹ However, it also includes practices meant to evaluate or classify trustworthiness based on social behavior¹⁸² or the use of “real-time remote biometric identification” (with some exceptions),¹⁸³ which are more privacy focused. ChatGPT seems unlikely to fall under one of these categories.

In contrast, the list of *high risks* is much broader. There is an open list of cases, where the AI is used as a safety component.¹⁸⁴ There is also a closed list, which includes biometric identification, management of critical infrastructure, admission or assessments in education or hiring, access to services (for example, credit scores), law enforcement, migration, and the administration of justice and democracy.¹⁸⁵ These types of risks are regulated via a risk-management system to identify risks and adopt mitigation measures.¹⁸⁶ Additional measures include technical documentation,¹⁸⁷ record-keeping,¹⁸⁸ transparency through detailed

175. *Id.*

176. *Id.*

177. *Id.*

178. *Id.* art. (3)(1); *id.* Annex I.

179. *AI Act*, *supra* note 25, art. 5.

180. *Id.* art. 5(1)(a).

181. *Id.* art. 5(1)(b).

182. *Id.* art. 5(1)(c).

183. *Id.* art. 5(1)(d).

184. *AI Act*, *supra* note 25, art. 6(1). Note that the definition also refers to Annex II, which lists various EU legislation applying to some specific categories of products. *Id.* Annex II.

185. *See id.* Annex III.

186. *Id.* art. 9.

187. *Id.* art. 11.

188. *Id.* art. 12.

instructions for use,¹⁸⁹ and registration of the system prior to its launch.¹⁹⁰ Furthermore, the AI Act determines that high-risk AI systems must be designed in a manner that allows for human oversight,¹⁹¹ and must implement an “appropriate level of accuracy, robustness and cybersecurity” throughout their lifecycle.¹⁹² The duty to follow the latter requirements lies with the providers of high-risk AIs.¹⁹³ Interestingly, the AI Act also imposes some specific obligations on the users of high-risk AI. In particular, if the user can exercise control, he bears an obligation to “monitor the operation of the high-risk AI system on the basis of the instructions of use”¹⁹⁴ and keep logs that the AI generates automatically.¹⁹⁵ Furthermore, the user is obligated to ensure that their input data is “relevant in view of the intended purpose of the high-risk AI system.”¹⁹⁶

As ChatGPT may be used for various purposes, it could, in principle, be classified as a high-risk AI in some cases under the definitions set in the original AI Act. For instance, one could potentially ask ChatGPT to perform some text-analysis tasks on college essays as part of a university’s admission process, which falls under the assessment category.¹⁹⁷

However, ChatGPT’s most straightforward use, eliciting answers to user questions, seems somewhat unlikely to be considered high-risk per se. In the example of Jack the lawyer who seeks to elicit answers to legal questions to draft a brief in a legal proceeding, Jack himself does not seem to engage in a high-risk activity. But if the judge were to rely on the content, this might fall under the category of administrating justice and democracy, which the AI Act defines as “AI systems intended to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts.”¹⁹⁸ There are already several known cases of such practices. A judge in

189. *AI Act*, *supra* note 25, art. 13.

190. *Id.* art. 51.

191. *Id.* art. 14.

192. *Id.* art. 15(1).

193. The AI Act defines a provider as a “natural or legal person, public authority, agency or other body that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark, whether for payment or free of charge.” *Id.* art. 3(2). Note that the AI Act’s original text has a clear typo, marking two paragraphs as (1), but the definition is the second one, so that it should be “(2).” The AI Act further entails a list of explicit obligations, such as ensuring that the AI undergoes conformity assessment and notifying the national authorities on the launch of the AI in their territory. *Id.* art. 16. The exact details of these are elaborated on in Articles 17–25. Importers and distributors are also subject to some obligations. *Id.* art. 26–28.

194. *AI Act*, *supra* note 25, art. 29(4).

195. *Id.* art. 29(5).

196. *Id.* art. 29(3).

197. Teachers seem to be using ChatGPT for various purposes already. See Emma Whitford, *How ChatGPT Is Fast Becoming the Teacher’s Pet*, FORBES (Mar. 25, 2023, 6:30 AM EST), <https://www.forbes.com/sites/emmawhitford/2023/03/25/how-chatgpt-is-fast-becoming-the-teachers-pet>).

198. *AI Act*, *supra* note 25, Annex III.

India asked ChatGPT to determine whether to grant bail in a murder case.¹⁹⁹ A judge in Pakistan experimented with ChatGPT for setting bail in a sexual assault case.²⁰⁰ A judge in Colombia used ChatGPT to decide on an insurance case.²⁰¹ Thus, ChatGPT seems to fall under this high-risk category.

Next, for AI systems with *limited risk*, there are only transparency requirements. The providers must ensure that natural persons (i) know they are dealing with an AI, (ii) are informed if the AI uses emotion-recognition or biometric categorization, and (iii) are told if an AI manipulated the content they are watching.²⁰² The only exception is when it is obvious from the circumstances or context that one is facing an AI.²⁰³ These requirements do apply to ChatGPT, but seem easy enough to fulfill by general warnings (which do exist already today on the ChatGPT website). Still, as one could use ChatGPT through an API²⁰⁴ without accessing the website directly, not all uses are transparent. Accordingly, provisions related to limited risk may also apply.

Finally, AIs with *minimal risk* (a residual category) are allowed free use.²⁰⁵ For day-to-day uses by private persons, ChatGPT plausibly falls under this category.

The AI Act has several additional parts, such as provisions meant to facilitate regulatory sandboxes, which are time-limited controlled environments

199. Ben Cost, *Judge Asks ChatGPT to Decide Bail in Murder Trial*, N.Y. POST (Mar. 29, 2023), <https://nypost.com/2023/03/29/judge-asks-chatgpt-for-decision-in-murder-trial>.

200. See *ChatGPT-4 Used in a Pakistani Judgment as an Experiment*, COURTING THE LAW (Apr. 7, 2023), <https://courtingthelaw.com/2023/04/07/laws-judgments-2/chatgpt-4-used-in-a-pakistani-judgment-as-an-experiment>. For the use of ChatGPT in another Pakistani case, involving a property dispute, see Sahar Iqbal, *AI in Pakistani Courts of Law*, INT'L BAR ASS'N (June 1, 2023), <https://www.ibanet.org/AI-in-Pakistani-courts-of-law>.

201. See Luke Taylor, *Colombian Judge Says He Used ChatGPT in Ruling*, THE GUARDIAN (Feb. 3, 2023, 9:53 PM), <https://www.theguardian.com/technology/2023/feb/03/colombia-judge-chatgpt-ruling>; Emily Hinkley, *Judge Says He Used ChatGPT in Court Ruling*, LEGAL CHEEK (Feb. 7, 2023, 11:37 AM), <https://www.legalcheek.com/2023/02/judge-says-he-used-chatgpt-in-court-ruling>.

202. *AI Act*, *supra* note 25, art. 52.

203. *Id.* (“Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use.”).

204. See Ms Valeria Ferrari, *The Platformisation of Digital Payments: The Fabrication of Consumer Interest in the EU FinTech Agenda*, 45 COMPUT. L. & SEC. REV. 1, 11 (2022) (defining API as “a set of rules and specifications followed by programmes to communicate with each other, and an interface between different programmes that facilitates their interaction”). See OPENAI, <https://openai.com/api> (last visited Sept. 17, 2023) for ChatGPT’s API.

205. This is clarified on the European Commission’s website, which states that “[t]he proposal allows the free use of minimal-risk AI. This includes applications such as AI-enabled video games or spam filters. The vast majority of AI systems currently used in the EU fall into this category.” EUR. COMM’N, *Regulatory Framework Proposal on Artificial Intelligence*, <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> (June 20, 2023).

in which technology can be tested without imposing penalties.²⁰⁶ As it has already been launched without such a framework, this Article avoids expanding further on this point.

2. *AI ACT: Revisions Proposed by EU Entities*

The text of the AI Act proposal went through several rounds of review by European institutions following its release in 2021. The Council of the EU²⁰⁷ proposed a first set of revisions in November 2021, some of which seem relevant to the context. First, the Council proposed to change the definition of AI to include also generative AI systems (such as ChatGPT) explicitly.²⁰⁸ Second, a title concerning “general purpose AI systems” was introduced, clarifying that such systems are not automatically subject to regulation.²⁰⁹ Third, the category of high-risk education was expanded to include the use for the “purpose of assessing natural persons with the view to evaluating learning outcomes or steering the learning process . . . at all levels.”²¹⁰ This expansion seems far more likely to include ChatGPT—for instance, anytime a professor uses ChatGPT to “steer the learning process” or evaluate students, provisions related to high-risk systems would apply. Fourth, the category of administering justice has been revised such that it only applies to systems used by a judicial authority or on its behalf, rather than simply “assisting” it. Thus, for our leading example of Jack the lawyer, this text makes it less likely to classify the use as high-risk. However, for the aforementioned known cases in which judges used ChatGPT directly, the classification of ChatGPT as a high-risk AI would remain the same.

After various other suggestions on the AI Act continued to flow between December 2021 and April 2022,²¹¹ the Council of the European Union released

206. The explanatory memorandum of the AI Act describes a regulatory sandbox as a “controlled environment to test innovative technologies for a limited time on the basis of a testing plan agreed with the competent authorities. Title V also contains measures to reduce the regulatory burden on SMEs and start-ups.” *AI Act*, *supra* note 25, Explanatory Memorandum at 15. A first sandbox was launched in Spain in June 2022. See EUR. COMM’N, *Launch Event for the Spanish Regulatory Sandbox on Artificial Intelligence* (June 27, 2022), <https://digital-strategy.ec.europa.eu/en/events/launch-event-spanish-regulatory-sandbox-artificial-intelligence>.

207. Council of the European Union, Note from Presidency to Delegations, No. 14278/21, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts - Presidency Compromise Text, Brussels European Council (Nov. 29, 2021).

208. *Id.* at 33. The text also proposes a revised definition for a provider: “a natural or legal person, public authority, agency or other body that develops an AI system or that has an AI system developed and places that system on the market or puts it into service under its own name or trademark, whether for payment or free of charge.” *Id.*

209. *Id.* at 68.

210. *Id.* at 97.

211. See generally *Opinion of the European Economic and Social Committee on Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, 2021 O.J. (C 517) 61 (recommending

a second proposal in July 2022.²¹² For the discussion here, several key suggestions of this new proposal seem relevant. First, it narrows down the definition of what constitutes AI to a system that operates with a “certain level of autonomy, and that, based on machine and/or human-provided data and inputs . . . produces system-generated outputs,”²¹³ such as generative AI systems. ChatGPT still falls under this definition. Second, it widely extended the scope of regulation for general-purpose AI systems. Instead of using this category to address technologies not directly subject to regulation, the proposal now says that general-purpose AI systems that may be used as high-risk AIs are subject to the same requirements as other high-risk systems.²¹⁴ In other words, it seems as though the classification of ChatGPT as a general-purpose AI would still likely subject it to strict regulations. Third, the proposal clarified that the exemption from transparency requirements applies only when it is obvious “from the point of view of a reasonable person” that one is dealing with an AI.²¹⁵ As mentioned, this is generally the case when using ChatGPT directly but may not hold if one simply uses an API and does not disclose to the user that ChatGPT has been used behind the scenes.

Another updated version of the Council’s proposal for the AI Act (a “general approach”) was released recently, as a next step toward final

improving the scope, definition, and clarity of prohibited AI practices); *Opinion of the Committee of the Regions on the European Approach to Artificial Intelligence – Artificial Intelligence Act*, COM (2021) 206 (Dec. 2, 2021) (recommending adding certain legislative amendments); *Opinion of the European Central Bank of 29 December 2021 on a Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence*, 2022 O.J. (C 515) 5 (clarifying the European Central Bank’s role under the proposed regulations); *Opinion of the Committee on the Environment, Public Health and Food Safety for the Committee on the Internal Market and Consumer Protection and for the Committee on Civil Liberties, Justice and Home Affairs on the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM (2021) 0206 (Apr. 22, 2022) (proposing adding environmental-related regulations); *Opinion of the Committee on Industry, Research and Energy for the Committee on the Internal Market and Consumer Protection and the Committee on Civil Liberties, Justice and Home Affairs on the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM (2021) 0206 (June 14, 2022) (proposing innovation-focused provisions); *Opinion of the Committee on Transport and Tourism for the Committee on the Internal Market and Consumer Protection and the Committee on Civil Liberties, Justice and Home Affairs on the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM (2021) 0206 (July 12, 2022) (proposing transport-focused clarifications).

212. Council of the European Union, Note from Presidency to Delegations, No. 8115/21, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts - Second Presidency Compromise Text, Brussels European Council (July 15, 2022).

213. *Id.* at 41.

214. *Id.* arts. 4a–4b.

215. *Id.* art. 52.

negotiations.²¹⁶ It contains a few revisions, but adopts the previous version's provision classifying general-purpose AIs as high-risk if they could be used for one of the high-risk purposes.²¹⁷ Finally, on June 14, 2023 the EU Parliament voted in favor of moving forward with the AI Act,²¹⁸ thereby commencing the standard "trilogue" between the Parliament, the Council of the European Union, and the European Commission.²¹⁹ Following the trilogue, a political agreement to move forward with the AI Act was reached in December 2023, bringing it one step closer to adoption.²²⁰ A multitude of last-minute amendments were also added to the AI Act, partially due to the overwhelming popularity of ChatGPT in the preceding months.²²¹ For the sake of brevity, this Article elaborates here on only a few relevant amendments.

First, the AI Act now distinguishes between general-purpose AI and so-called "foundation models" that have been "designed for generality of output and can be adapted to a wide range of distinctive tasks."²²² The category of a foundations model is aimed to tackle models like GPT.²²³ Interestingly, the

216. Council of the European Union, Note from Presidency to Delegations, No. 8115/21, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts - General Approach, at 1, Brussels European Council (Nov. 25, 2022) [hereinafter *AI Act General Approach*].

217. *Id.* art. 4b.

218. Giovanni Campi, Claude-Étienne Armingaud & Petr Bartoš, *European Parliament Adopts Negotiating Mandate on European Union's Artificial Intelligence Act; Trilogues Begin*, 13 NAT'L L. REV. 177 (June 26, 2023), <https://www.natlawreview.com/article/european-parliament-adopts-negotiating-mandate-european-union-s-artificial>.

219. For a general overview of the trilogue process, see *Trilogue*, EUR-LEX, <https://eur-lex.europa.eu/EN/legal-content/glossary/trilogue.html> (last visited Sept. 19, 2023).

220. See Press Release, European Council, Artificial Intelligence Act: Council and Parliament Strike a Deal on the First Rules for AI in the World (Dec. 9, 2023), <https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/>.

221. *Amendments Adopted by the European Parliament on 14 June 2023 on the Proposal for a Regulation of the European Parliament and of the Council on Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, at 1, COM (2021) 0206 (June 14, 2023) [hereinafter *AI Act Amendments*]. For the full text of the compromise amendments, see *DRAFT Compromise Amendments on the Draft Report Proposal for a Regulation of the European Parliament and of the Council on Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM (2021) 0206 (May 9, 2023). See also Shweta Sharma, *EU closes in on AI Act with Last-minute ChatGPT-related Adjustments*, COMPUTERWORLD (Apr. 28, 2023), <https://www.computerworld.com/article/3695009/eu-closes-in-on-ai-act-with-last-minute-chatgpt-related-adjustments.html>.

222. *Id.* art. 3(1c). The definition of a general-purpose AI system now reads instead: "an AI system that can be used in and adapted to a wide range of applications for which it was not intentionally and specifically designed." See *id.* art. 3(1d). See also *id.* recitals (60e)–(60f), which clarify that when a foundation model is provided as a service through an API, cooperation between upstream and downstream entities should take place. This seems to apply, for instance, when OpenAI grants access to its GPT-4 model through an API.

223. See, e.g., Press Release, Eur. Parl., AI Act: A Step Closer to the First Rules on Artificial Intelligence (May 11, 2023), <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer->

proposal hints in one of its recitals that foundation models do not constitute a high-risk AI *per se*.²²⁴ Yet it still adopts various obligations for foundation models, such as a transparency requirement and a duty to train the model to contain safeguards to conform with EU law.²²⁵

Second, the AI Act now refers to users as “deployers”²²⁶ and defines a new category of “affected persons.”²²⁷ This distinction captures the difference between Jack the lawyer (the deployer) and his clients (the affected persons).

Third, and relatedly, the amendments altered the nature of transparency requirements toward affected persons: the affected person should be able to reasonably understand how the system functions and which data it processes.²²⁸

Fourth, affected persons are granted the right to request that deployers (that is, the AI users) provide “clear and meaningful explanations” on the role of the AI in the decision-making process.²²⁹ Furthermore, deployers of high-risk AI systems are obligated to actively inform natural persons that might be affected on the intended purpose and type of decisions made with the AI.²³⁰

Irrespective of the subtleties of the AI Act’s text, these various proposals teach us that the EU began its restraints on AI by introducing a framework that constitutes regulation rather than liability (as defined for the purpose of the comparison laid out in Part II). However, parallel to the AI Act, the EU also developed a specific framework for liability: the AILD.

C. THE AI LIABILITY DIRECTIVE

1. *The Directive, Generally*

Parallel to the development of the AI Act, the European Commission (and other EU institutions) published several documents on the topic of the liability

to-the-first-rules-on-artificial-intelligence (mentioning the AI Act includes rules regarding “[t]ailor-made regimes for general-purpose AI and foundation models like GPT”).

224. See *AI Act Amendments*, *supra* note 221, recital (60g) (“Generative foundation models should ensure transparency about the fact the content is generated by an AI system, not by humans. These specific requirements and obligations do not amount to considering foundation models as high risk AI systems, but should guarantee that the objectives of this Regulation to ensure a high level of protection of fundamental rights, health and safety, environment, democracy and rule of law are achieved.”).

225. See *id.* art. 28(b)(4) (noting that “generative AI” specifically has an obligation to comply with legal requirements, document the process, and be transparent).

226. See *id.* art. 3(1)(4).

227. *Id.* art. 3(1)(8a).

228. *Id.* arts. 13(1)–13(2). Note that in the original version of the AI Act, the users simply needed to be able to interpret the output. The amendments replace this stipulation with a requirement that users understand the process.

229. *Id.* art. 68(c).

230. *Id.* art. 29(6a).

of AI,²³¹ including a “Public Consultation on Civil Liability” of AI,²³² and a proposal for a new regulation on general product safety.²³³ Eventually, the European Commission finalized a proposal for an AI Liability Directive (AILD) in September 2022.

The AILD is a much shorter document than the AI Act, with a far more specific focus on defining the conditions under which civil liability arises in the context of AI. As stated in its explanatory memorandum, the AILD begins with an observation that “current national liability rules, in particular based on fault, are not suited to handling liability claims for damage caused by AI-enabled products and services.”²³⁴ In other words, it claims that fault-based rules, such as negligence, are inappropriate. The argument in the AILD continues by raising a few points: First, because AI is complex, opaque, and autonomous, it is difficult for victims to identify the liable person and even more so, to prove that person’s fault.²³⁵ Second, because lawsuits face high up-front costs and are lengthy, victims might be deterred from suing.²³⁶ Third, national courts might apply disparate rules, leading to legal uncertainty.²³⁷ The AILD further clarifies that it aims to complement the AI Act:

In the AI Act proposal, the Commission has proposed rules that seek to reduce risks for safety and protect fundamental rights. Safety and liability are two sides of the same coin: they apply at different moments and reinforce each other. While rules to ensure safety and protect fundamental rights will reduce risks, they do not eliminate those risks entirely. Where such a risk materializ[es], damage may still occur. In such instances, the liability rules of this proposal will apply.²³⁸

231. See generally *Report on the Safety and Liability Implications of Artificial Intelligence, the Internet of Things and Robotics*, COM (2020) 64 final (Feb. 19, 2020) (examining the broader implications and potential gaps in liability and safety frameworks); *Commission White Paper on Artificial Intelligence – A European Approach to Excellence and Trust*, COM (2020) 65 final (Feb. 19, 2020) (discussing policy options that enable development of trustworthy and secure AI).

232. Product Consultation for Product Liability Directive, EUR. COMM’N, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12979-Product-Liability-Directive-Adapting-liability-rules-to-the-digital-age-circular-economy-and-global-value-chains_en (follow “View feedback received” under “Roadmap”). The feedback period was from June 30, 2021 to July 28, 2021. *Id.*

233. *Proposal for a Regulation of the European Parliament and of the Council on General Product Safety, Amending Regulation (EU) No 1025/2012 of the European Parliament and of the Council, and Repealing Council Directive 87/357/EEC and Directive 2001/95/EC of the European Parliament and of the Council*, COM (2021) 346 final (June 6, 2021). See also EXPERT GRP. ON LIAB. & NEW TECHS., EUR. COMM’N, *LIABILITY FOR ARTIFICIAL INTELLIGENCE AND OTHER EMERGING DIGITAL TECHNOLOGIES* 3 (2019), <https://data.europa.eu/doi/10.2838/573689>.

234. *AILD*, *supra* note 26, Explanatory Memorandum at 1.

235. *Id.*

236. *Id.*

237. *Id.*

238. *Id.*

The AILD also explicitly acknowledges that the liability rules provide an economic incentive to comply with the safety rules of the AI Act and contribute to the enforcement of requirements imposed on high-risk AI providers. The latter point is key to understanding the content of the AILD: the liability rules described in the following paragraphs are intertwined with the requirements of the AI Act. In other words, the AILD partially transforms the rules set in the AI Act into liability rules.

2. *Scope and Content*

The AILD formally applies to “non-contractual fault-based civil law claims for damages” for harms caused by an output of an AI system (or its failure to produce output).²³⁹ Content-wise, the AILD covers two key points: (1) disclosure of evidence; and (2) rebuttable presumptions that ease a plaintiff’s burden of proof.

The *disclosure of evidence* provision enables a “potential claimant” (who is considering bringing a lawsuit but has not done so yet) to request the disclosure of relevant evidence about a specific high-risk AI system suspected of causing harm.²⁴⁰ However, there are some limitations to disclosure. First, for disclosure to be granted, the potential claimant must present sufficient facts and evidence to support a plausible claim.²⁴¹ Second, disclosure is only granted if the claimant undertook “all proportionate attempts at gathering the relevant evidence from the defendant.”²⁴² Third, disclosure is limited to necessary and proportionate measures to support a claim for damages, also taking into account the interests of third parties.²⁴³

The rebuttable presumptions are twofold. The first presumption concerns defendants who have failed to comply with a court order of disclosure. For such defendants, the AILD orders national courts to presume the defendant’s non-compliance with a relevant duty of care,²⁴⁴ but the defendant has the right to rebut that presumption.²⁴⁵

The second presumption concerns all defendants and concerns the causal link. In particular, the AILD instructs national courts to presume the existence of a causal link between the fault of the defendant and the AI’s output (or failure

239. *Id.* art. 1(2). For the definition of “claim for damages,” see *AILD*, *supra* note 26, art. (2)(5).

240. *Id.* arts. 2(7), 3(1). Technically, the AILD does not give a direct right to claimants but instead orders the member states to empower the courts to grant disclosure. *Id.*

241. *Id.* art. 3(1).

242. *Id.* art. 3(2).

243. *Id.* art. 3(4).

244. A duty of care is defined as a “required standard of conduct, set by national or Union law, in order to avoid damage to legal interests recognised at national or Union law level, including life, physical integrity, property and the protection of fundamental rights.” *Id.* art. 2(9).

245. *Id.* art. 3(5).

to produce one), as long as several conditions hold.²⁴⁶ First, that the fault of the defendant has been either established or presumed (via the first rebuttable presumption).²⁴⁷ Second, given the circumstances of the case, that it is reasonably likely that the defendant's fault influenced the AI's output (or failure to produce one).²⁴⁸ Third, that the claimant demonstrated that the output (or failure to produce one) caused harm.²⁴⁹ In other words, the rebuttable presumption enables a claimant, under some conditions, to bear only the burden of proving the element of harm while relying on a presumption regarding the elements of a breach of duty of care and a causal link.

However, for the second presumption (of a causal link), the AILD makes a sharp distinction between high-risk AIs and other AIs with respect to what is required from the claimant. Specifically, for high-risk AIs, it is sufficient for the claimant to prove that the AI provider did not comply with their obligations under the AI Act.²⁵⁰ Furthermore, if the lawsuit is directed at a *user* rather than a provider (for example, if Jack the lawyer's clients file a lawsuit against him), it is sufficient to show that the user either (1) did not comply with the obligations to monitor the AI as set by the AI Act; or (2) exposed the AI to input data that was irrelevant to its purpose. Thus, under the AILD, non-compliance with the AI Act serves as the justification for invoking the presumption of a causal link when high-risk AI is concerned. However, the AILD allows the defendant to rebut this presumption.²⁵¹ It also adds one important limitation. If the defendant demonstrates that "sufficient evidence and expertise is reasonably accessible" to the claimant to prove the causal link, no presumption should be made.²⁵²

Next, for non-high-risk AIs, the claimant faces a taller hurdle before a presumption of a causal link is applied: the presumption is to be made only if it is "excessively difficult" for the claimant to prove the causal link.²⁵³ Thus, the AILD sets the presumption as the default for high-risk systems and as the exception for non-high-risk systems.

Finally, if the lawsuit is filed against a user who used the AI for personal, non-professional activities, the presumption only applies if he "materially

246. *Id.* art. 4(1).

247. *Id.* art. 4(1)(a).

248. *Id.* art. 4(1)(b).

249. *Id.* art. 4(1)(c).

250. Specifically, this concerns requirements under Title III of the AI Act, which include the establishment of a risk-management system, *id.* art. 10, fulfilling data governance requirements, *id.* art. 9, technical documentation requirements, *id.* art. 11, record keeping, *id.* art. 12, transparency provisions, *id.* art. 13, human oversight, *id.* art. 14, product manufacturers' obligation to comply with the regulation for legal acts listed in Annex II, *id.* art. 24, or the obligations of distributors, importers, users, or any other third-party, *id.* art. 28. See *id.* arts. 4(1)(a)–4(3)(b) for an explicit list of specific violations of obligations from the AI Act.

251. *Id.* art. 4(7).

252. *Id.* art. 4(4).

253. *Id.* art. 4(5).

interfered” with the system or if he was “required and able to determine the conditions of operations (but failed to do so).”²⁵⁴

In addition to the presumption, the AILD acknowledges the need to evaluate the appropriateness of no-fault liability rules (meaning, strict liability) and establishes a monitoring program to determine at a later time the exact scope of evidence needed for liability and the role of insurance.²⁵⁵ Thereby, the AILD sets out a two-stage plan: (1) easing the burden of proof for claimants using the rebuttable presumptions; and (2) establishing “a review mechanism to re-assess the need for harmoni[z]ing strict liability for AI use cases with a particular risk profile (possibly coupled with a mandatory insurance).”²⁵⁶

D. THE REVISED PRODUCT LIABILITY DIRECTIVE

Jointly with the AILD, the European Commission proposed a revised product liability directive (PLD) as a complementary set of rules.²⁵⁷ The PLD contains revisions of an existing directive from 1985, which was outdated for many reasons.²⁵⁸ A political agreement to move forward with the new PLD was reached in December 2023.²⁵⁹

The new PLD’s premise is that AI is, first and foremost, a product.²⁶⁰ Therefore, it subjects AIs to the same basic concepts as other defective products. To avoid clashes with the AI Act and the AILD, the PLD tries to limit its scope

254. *Id.* art. 4(6).

255. *Id.* art. 5.

256. *Commission Staff Working Document Executive Summary of the Impact Assessment Report*, SWD (2022) 320 final (Sept. 28, 2022).

257. One doctrinal difference concerns the basis of a tort claim: under the AILD, the basis is member-states laws, whereas under the PLD, the basis is EU law. *See* Hacker, *European AI Liability Directives*, *supra* note 28 (manuscript at 8).

258. Among other things, the directive imposed a minimum threshold for harms of 500 euros, making it impossible to sue for small damages. *See* Council Directive 85/374 of July 25, 1985, on the Approximation of the Laws, Regulations and Administrative Provisions of the Member States Concerning Liability for Defective Products, art. 9(b), 1985 O.J. (L210) 29, 31. The revision removes this threshold. *See Questions and Answers on the Revision of the Product Liability Directive*, EUR. COMM’N (Sept. 28, 2022), https://ec.europa.eu/commission/presscorner/detail/en/QANDA_22_5791 [hereinafter *Q&A on PLD*].

259. *See* Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs, *The Commission Welcomes the Provisional Agreement on Updating EU Product Liability Rules for the Digital Age and Circular Economy*, EUR. COMM’N (Dec. 14, 2023), https://single-market-economy.ec.europa.eu/news/commission-welcomes-provisional-agreement-updating-eu-product-liability-rules-digital-age-and-2023-12-14_en.

260. *Q&A on PLD*, *supra* note 258 (“The revised PLD crucially also makes clear that software, including AI systems, is a product. Therefore, if AI systems are defective and cause death, personal injury, property damage or data loss, injured people can use the PLD to claim compensation.”). The definition of “product” in the PLD is as follows: “[P]roduct’ means all movables, even if integrated into another movable or into an immovable. ‘Product’ includes electricity, digital manufacturing files and software.” *PLD*, *supra* note 27, art. 4(1).

to material losses caused by either (i) death or injury; (ii) harm to property (with some exceptions); or (iii) loss of data.²⁶¹

The PLD requires member states to ensure that any person suffering such harm by a defective product (including AI) is entitled to compensation²⁶² and that the manufacturers of a defective product are held liable.²⁶³ However, the key issue for our purposes is precisely how “defectiveness” is determined. The PLD says that a product would be considered defective when it “does not provide the safety the public at large is entitled to expect,”²⁶⁴ while taking into account a specific list of factors—one of which is the “effect on the product of any ability to continue to learn after deployment.”²⁶⁵ Suppose that Jack the lawyer reasonably expects ChatGPT to deliver accurate answers. Such an expectation may be unjustified, but what if it is? Would ChatGPT then be considered a defective product? Or does the fact that it is able to learn and adapt exempt it from the definition? The practical relevance lies in the slight differences to the AILD with respect to disclosure and the burden of proof.

Unlike the AILD, the PLD begins with a clear statement that the claimant is required to prove the defectiveness, the harm, and the causal link.²⁶⁶ However, it includes somewhat similar rebuttable presumptions. The defectiveness is presumed if either: (1) the defendant does not comply with disclosure; (2) the claimant shows that the product does not comply with safety requirements at the union-level or national level; or (3) the harm was caused by an “obvious malfunction . . . during normal use or under ordinary circumstances.”²⁶⁷ This raises some obvious questions: if an AI does not comply with the AI Act, does this mean it falls under noncompliance with safety requirement at the union level? If so, both the AILD and the PLD would apply in conjunction, which seems odd, given that both seemingly create their own liability regime.²⁶⁸

Next, the causal link is to be presumed if the product was shown to be defective and the harm is “of a kind typically consistent with the defect in question.”²⁶⁹ Again, this differs from the AILD, which uses a different terminology (which may or may not entail a different meaning). Furthermore, the concept of “excessive difficulty” to prove the causal link is again used, but

261. *PLD*, *supra* note 27, art. 4(6). The PLD excludes property harms to “(i) the defective product itself; (ii) a product damaged by a defective component of that product; (iii) property used exclusively for professional purposes.” *See id.*

262. *Id.* art. 5.

263. *Id.* art. 7(1).

264. *Id.* art. 6.

265. *Id.*

266. *Id.* art. 9.

267. *Id.* art. 9(2)(c).

268. *See infra* Part V (discussing legal uncertainty due to unclarity of which regime applies).

269. *PLD*, *supra* note 27, art. 9(3).

it is limited to difficulty “due to technical and scientific complexity.”²⁷⁰ Note, however, that the claimant still needs to prove that the product contributed to the harm and that it is likely defective (or that its defectiveness is the likely cause of harm). Interestingly, the PLD explicitly says that technical or scientific complexity should be determined on a “case-by-case basis,”²⁷¹ a statement different from the one in the AILD.

Finally, the PLD dictates a set of exemptions from liability, two of which seems especially interesting in the context of AI. A manufacturer is exempted (1) if it is probable that the defectiveness did not exist when the product was launched and (2) when the “objective state of scientific and technical knowledge” at the time of launching was such that the defectiveness could not be discovered. The former is troublesome given ChatGPT’s ever-changing nature due to machine learning, so one could interpret AI creators as almost always exempted. The latter raises serious questions about how one should think about the “objective state” of knowledge in an AI-enabled environment. Overall, the parallel application of the PLD and the AILD seems at odds.²⁷²

E. RECENT PROPOSALS IN THE UNITED STATES

Compared to the European framework, which entails very concrete proposals on how to restrain AI, development of AI policy in the United States is lagging.²⁷³ There are, however, a few initiatives that try to devise principles at the national level.²⁷⁴

The first initiative began with the enactment of the National Artificial Intelligence Initiative Act of 2020.²⁷⁵ This act established a National AI Advisory Committee (NAIAC), which is “tasked with providing an independent assessment . . . and making recommendations to the President . . . on topics related to AI research and development, ethics, standards, education, technology transfer, commercial application, security, and economic competitiveness.”²⁷⁶

270. *Id.* art. 9(4).

271. *Id.* pmbl. § 34.

272. For further critique of the combination of the AILD and PLD, see generally Hacker, European AI Liability Directives, *supra* note 28 (analyzing the contradiction between the AILD and the PLD).

273. See generally Lyon et al., *supra* note 24 (providing an overview of U.S. policy developments in comparison to the EU’s landmark AI Act). See also Em McPhie, *As ChatGPT’s Popularity Skyrockets, Some Experts Call For AI Regulation*, BROADBAND BREAKFAST (Feb. 3, 2023), <https://broadbandbreakfast.com/2023/02/as-chatgpts-popularity-skyrockets-some-experts-call-for-ai-regulation> (“So far, the U.S. has very little legislation governing the use of AI.”).

274. For an overview of the various state-level initiatives, see Lyon et al., *supra* note 24.

275. NAIAC, *supra* note 54.

276. National AI Advisory Committee (US), CTR. FOR AI & DIGIT. POL’Y, <https://www.caidp.org/resources/naiac/#:-:text=The%20Advisory%20Committee%20was%20established,AI%2C%20and%20AI%20workforce%20issues> (last visited Sept. 21, 2023).

The NAIAC is only just beginning its work, and has so far mainly issued requests for public comments.²⁷⁷

A second initiative is the aforementioned AI RMF,²⁷⁸ which is divided into two parts. The first part entails suggestions on how to think about AI's risks, as well as general actionable suggestions such as making sure that the risks are repeatedly measured throughout the life cycle of the AI,²⁷⁹ paying attention to risk tolerance,²⁸⁰ and ensuring trustworthiness.²⁸¹ The second part offers a framework for distinguish more concretely between four functions: governing, mapping, measuring, and managing risk.²⁸² It is difficult to compare these directly to the EU's proposals because the recommendations are general in nature and intended as voluntary best-practices for business using AI. While there is some overlap in the highlighted principles (such as transparency and record keeping), the AI RMF only barely hints at rationales similar to those of its European counterpart. For instance, the AI RMF suggests that “when consequences are severe, such as when life and liberty are at stake, AI developers and deployers should consider proportionally and proactively adjusting their transparency and accountability practices”—a much softer approach compared to the EU's approach of stark regulation and liability.²⁸³

A third initiative originated from a presidential executive order issued in December 2020,²⁸⁴ which later led to the development of a blueprint for an Artificial Intelligence Bill of Rights (AIBR)²⁸⁵ in October 2022. The AIBR lays out five principles to “help guide the design, use, and deployment of automated systems to protect the rights of the American public in the age of artificial intelligence.”²⁸⁶ The first principle calls for “safe and effective systems,” which should be developed in consultation with experts and in view of the potential risks.²⁸⁷ The second principle refers to protections against “algorithmic discrimination” to prevent unjust treatments of certain groups “based on race,

277. *See, e.g.*, Notice of Federal Advisory Committee Open Meeting, 87 Fed. Reg. 58312 (Sept. 26, 2022).

278. *See supra* Part II.

279. AI RMF, *supra* note 54, at 5.

280. *Id.* at 7.

281. *Id.* at 12 (“[A trustworthy AI is] valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed.”).

282. *Id.* at 3.

283. *Id.* at 16. *See also* the “playbook” published by the NIST in connection with the AI RMF. Nat’l Inst. of Standards and Tech., AI Risk Management Framework Playbook, U.S. DEP’T COM., https://airc.nist.gov/docs/AI_RMF_Playbook.pdf (last visited Sept. 21, 2023) (evaluating AI risk management approaches).

284. Exec. Order No. 13,960, 85 Fed. Reg. 78,939 (Dec. 3, 2020).

285. WHITE HOUSE OFF. OF SCI. & TECH. POL’Y, THE BLUEPRINT FOR AN AI BILL OF RIGHTS: MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE (2022) [hereinafter AIBR].

286. *Id.* at 4.

287. *Id.* at 5.

color, ethnicity, sex,²⁸⁸ among other characteristics. The third principle concerns data privacy, emphasizing the need to avoid unnecessary surveillance.²⁸⁹ The fourth principle, “notice and explanation” is about transparency—ensuring that users know they are dealing with an AI and understand how the outputs affect them.²⁹⁰ Finally, the fifth principle relates to “human alternatives,” that is, the right to opt out of automatic systems and reach a person who can assist in case of a problem.²⁹¹ These various principles seem to be consistent with some regulatory approach along the lines of the European AI Act, but the AIBR essentially neglects the issue of tort liability entirely.

A fourth initiative can be found in the recently announced “National AI Commission Act,” which will form a commission to review the U.S. policy on AI.²⁹²

There are several additional initiatives at the federal level.²⁹³ These include, for instance, a request by the Federal Trade Commission²⁹⁴ for public comments on AI surveillance practices and guidelines by the Consumer Financial Protection Bureau²⁹⁵ relating to transparency. However, liability remains unaddressed, especially for uses like ChatGPT, for which the main concern is not algorithmic discrimination but rather the (in)accuracy of the text and the fear that it will be misused to generate content that ends up being harmful in some way.

Summing up, the initiatives in the United States hint at regulation but do not pay much attention to liability.

V. A LAW AND ECONOMICS EVALUATION OF AI RESTRAINTS

This Part evaluates the EU’s framework through the lens of law and economics, applying the concepts covered in Parts II and III. Subpart V.A begins with the basic question of whether the intervention in AI has sufficient solid economic grounds. Subpart V.B discusses the EU’s choice to combine liability and regulation. Subpart V.C analyzes whether the AILD’s two presumptions—

288. *Id.* For examples on discriminatory outcomes of AI, see Dovič Barysè & Roe Sarel, *Algorithms in the Court: Does It Matter Which Part of the Judicial Decision-Making is Automated?*, A.I. & L. (2023).

289. AIBR, *supra* note 285, at 6.

290. *Id.*

291. *Id.*

292. See Robert Seamans, *AI Regulation Is Coming to the U.S., Albeit Slowly*, FORBES (June 27, 2023, 5:49 AM EST), <https://www.forbes.com/sites/washingtonbytes/2023/06/27/ai-regulation-is-coming-to-the-us-albeit-slowly>.

293. See generally Lyon et al., *supra* note 24 (listing both laws and bills at the federal and state level).

294. Trade Regulation Rule on Commercial Surveillance and Data Security, 87 Fed. Reg. 51,273 (Aug. 22, 2022).

295. CFPB Acts to Protect the Public from Black-Box Credit Models Using Complex Algorithms, CONSUMER FIN. PROT. BUREAU (May 26, 2022), <https://www.consumerfinance.gov/about-us/newsroom/cfpb-acts-to-protect-the-public-from-black-box-credit-models-using-complex-algorithms>.

regarding fault and the causal link—are consistent with the insights of choosing strict liability over negligence. Finally, Subpart V.D discusses the legal uncertainty that arises due to the combination of the AILD and the PLD.

A. MARKET FAILURE, WHERE ART THOU?

Recall that the economic rationale for intervening in markets is the existence of a market failure.²⁹⁶ In particular, for issues related to liability, the central market failure of interest is the concern of negative externalities.²⁹⁷ However, neither the AI Act nor the AILD dedicating any meaningful attention to analyzing whether a market failure exists. Instead, the proposals pool different risks into categories (high risk, limited risk, etc.) and apply various rules irrespective of whether the risk constitutes an externality.

To illustrate, consider the example of Jack the lawyer, who decides to use ChatGPT to attain an output that he later passes on as legal advice. In this case, any harm to Jack himself is not an externality at all, as Jack is a party to a contract with the AI creator.²⁹⁸ This is true irrespective of whether the harm to Jack results from a high or low risk. Moreover, the justification for intervening does not depend on whether the potential harm to Jack's clients as a consequence of is high or low risk. Hence, to a large extent, the logic of a risk-based approach is not necessarily consistent with the notion of preventing externalities. One may also wonder whether protecting Jack's clients is an important externality. Jack and his clients have a contractual relationship, so the clients can already account for the risk of Jack making a mistake in the price. Thus, even though the actions of an AI used by Jack are technically an externality on the clients, these clients do have a way to protect themselves *ex ante*. Nonetheless, there is at least one clear advantage to the risk-based approach. Namely, if high-risk AIs correspond to larger harms, then the magnitude of the externality increases, insofar as it is an externality. It could then make sense to separate the regulatory requirements according to some categorization, assuming that the magnitude of the risk is captured in each category.

There are also other market failures that could justify the AI Act's risk-based approach. For instance, the fact that users of AI cannot observe its actions after the contract is concluded can give rise to a moral hazard problem where the AI goes rogue and does not deliver the promised output. Perhaps the degree of asymmetric information could be reduced if the AI Act's conditions somehow provide third-party verification of the AI's actions, for instance, through the provisions forcing the AI creator to document and register the technology with a public authority. Furthermore, if one adopts a perspective of behavioral law

296. *See supra* Part II.

297. *See supra* Part III.A.1.

298. *See supra* Part II.

and economics, assuming that individuals are not fully rational but rather susceptible to biases, the requirements for transparency may help avoid the exploitation of users. However, these do not constitute a clear rationale for *liability*, which is typically about externalities.

The European approach seems only partially consistent with the usual economic rationale for intervention, as the risk-based approach does not map conveniently onto a clear market failure.

B. SHALL LIABILITY AND REGULATION GO TOGETHER?

The general overview of the considerations when choosing between regulation, liability, and a mixed regime apply to the EU's framework. The AI Act includes a comprehensive package of regulations, ordering member states to adopt various rules for monitoring AI. The AILD adds a tort-liability component, requiring member states to empower their national courts to handle lawsuits against actors involved in AI markets, including producers and users. The PLD adds additional liability for certain defects.

Recall that the four determinants proposed by Shavell—differential information, judgment-proofness, escaping lawsuits, and administrative costs—support a pure liability regime only under very specific conditions. Namely, conditions that mitigate the concern that AI creators would escape suits.²⁹⁹ Does the AILD mitigate those concerns?

Let us return to the cases Shavell identified as problematic for escaping lawsuits (dispersed harm, deterioration of evidence, and difficulties in proving a causal link) and discuss each in turn. First, as the AILD applies for both dispersed and non-dispersed harms, it may or may not function well when lawsuits are filed in connection with ChatGPT, depending on the specific harm. The same is true for the PLD. On this front, the liability directives do not provide a clear economic advantage for dispersed harms.

Second, while the AILD does not explicitly reference timing and therefore does not explicitly deal with harms that manifest in the future, its mechanisms are implicitly tailored to address the concern regarding the deterioration of evidence over time. In particular, the obligation of record-keeping over the entire life cycle of high-risk AI systems³⁰⁰ serves precisely to ensure that the evidence does not disappear. However, as this does not apply to low-risk systems, this solution, standing alone, is not comprehensive. Conversely, according to the latest amendments to the AI Act, the requirement of technical documentation (as opposed to record-keeping) not only applies to high-risk systems, but also to

299. See *supra* Part III.B.

300. *AI Act*, *supra* note 25, art. 12(1).

foundation models, which provides some remedy.³⁰¹ But perhaps more importantly, the fact that the rebuttable presumptions reverse the burden of proof incentivizes defendants to keep records so they can one day defend themselves. Thus, the rebuttable presumption concerning fault seems compatible with the principles of law and economics.

Third, the other rebuttable presumption concerning the causal link largely eliminates the concern of identifying who caused the harm. The burden is instead switched to the defendant to prove there is no causal link. Hence, the victims should not be reluctant to file a lawsuit, which implies that the AI creators will not anticipate escaping a lawsuit. Nevertheless, recall that this rebuttable presumption does not apply automatically. Rather, it requires some preconditions: that the fault was proven or presumed; that the circumstances are such that a causal link is likely; and that the connection between the harm and the output has been proven. Moreover, victims who suffer harm from a low-risk system might still be reluctant to sue where a presumption is applied more leniently to high-risk AI than other AIs.

A different concern arises from the intersection of the AI Act and the AILD in terms of triggering the presumptions. For instance, under the AI Act, the causal link presumption for high-risk AIs can be made based almost exclusively on the fact that the defendant breached his obligations under the regulation. In other words, the AILD precisely implements a connection that Shavell cautions against,³⁰² arguing that some potential defendants may be “undesirabl[y] led to comply . . . even when they would not otherwise have done so.”

The remaining considerations specified in Part II.B.3 must also be taken into account. First, if the defendant’s violation of the AI Act’s obligations is the only feasible way to establish liability, there is a concern that AI creators will only “just comply.” Second, AI creators may over-invest in precautions to reduce the likelihood that the presumptions are triggered. Finally, the presumptions may lead to higher prices. For instance, recall that while ChatGPT is currently still offered also in a free version, there is already a paid version, ChatGPT Plus,³⁰³ whose price may already reflect the expected cost of litigation from the EU directives that may come into force. The connection between the AI Act and the AILD might also be under-inclusive or over-inclusive, leading to differentiations that seem arbitrary.³⁰⁴

301. *Compare AI Act Amendments*, *supra* note 221, art. 11 (referring only to high-risk systems), *with id.* art. 28b(2)(e) (adding a requirement to “draw up extensive technical documentation and intelligible instructions for use” for foundation models).

302. Shavell, *supra* note 59, at 365.

303. *Introducing ChatGPT Plus*, OPENAI (Feb. 6, 2023), <https://openai.com/blog/chatgpt-plus>.

304. *See, e.g.*, Hacker, *European AI Liability Directives*, *supra* note 28 (manuscript at 12–14) (discussing the possibility of both under inclusiveness and over-inclusiveness, and arguing that ChatGPT falls under the latter problem).

C. STRICT LIABILITY OR NEGLIGENCE?

The AILD frames its choice of policy as something of a compromise between fault-based and strict liability,³⁰⁵ applying the rebuttable presumptions for now and postponing the discussion on strict liability for later.³⁰⁶ However, its *de jure* premise is one of a fault-based regime. The claimant must prove fault as a rule, with an exception in cases where the defendant does not comply with an order of disclosure.³⁰⁷ Conversely, the PLD begins with a premise of strict liability,³⁰⁸ but provides exceptions where there is no liability at all. Whether these two directives would lead to similar or disparate results is a *de facto* open question. For instance, suppose that Jack asks OpenAI to disclose precisely which process was used to derive the answer given to him on the issue of good-faith purchases by ChatGPT.³⁰⁹ Could OpenAI produce the relevant evidence on the specific risk that manifested in Jack's case? It may well occur that AI creators will be viewed as refusing to deliver evidence simply because they are not able to pin down precisely what went wrong given that AI systems learn continuously. Therefore, the proportion of cases that would effectively be treated as strict liability due to the perception that the defendant is non-compliant with disclosure may be non-negligible.³¹⁰ Moreover, the AILD does not appear to be very restrictive when it comes to determining what the relevant obligations that yield a duty of care are.³¹¹ In principle, one could face an almost unlimited set of rules whose breach would lead to liability, even without a presumption of fault.

This Article next turns to the concepts reviewed in Part IV for choosing between strict liability and negligence, including: (i) unilateral versus bilateral care; (ii) levels of activity; (iii) risk-attitudes; and (iv) known versus unknown risks. The goal of the analysis is to evaluate whether these are given sufficient weight in the EU's framework.

305. See, e.g., *AILD*, *supra* note 26, Explanatory Memorandum at 7–8 (explaining that consumer organizations were in favor of strict liability, but businesses found it to be disproportionate).

306. *Id.* at 9 (describing three policy options and the choice to go with a two-stage option).

307. See Hacker, *European AI Liability Directives*, *supra* note 28 (manuscript at 19) (“Fault is the key trigger of liability for cases brought under the AILD Proposal. Traditionally, fault implies intent or negligence on the part of the tortfeasor.”).

308. See *id.* (manuscript at VII).

309. See *supra* Part II.

310. *But see* Hacker, *European AI Liability Directives*, *supra* note 28 (manuscript at 40) (arguing that proving violations of the AI Act may be difficult for non-experts, so that the concern may be the opposite, such as a fault-based regime with a difficulty to prove fault).

311. See *id.* (manuscript at 34) (“Nothing in the wording of the definition or in the recitals suggests that only statutory law could establish a relevant duty. . . . Member States remain free to demand, besides a breach of duty, any subjective element concerning the defendant (e.g., negligence).”).

1. *Unilateral versus Bilateral Care*

Although the AILD does not explicitly address the distinction between unilateral and bilateral care—seemingly neglecting one of the most important economic issues—one can find some comfort in the preconditions for presuming a causal link, which hint at this distinction. Recall that for high-risk AI systems, the presumption of a causal link is to be avoided if the defendant proves that the claimant had sufficient access and expertise to prove the link. Insofar as such access is also a good proxy for the claimant’s ability to prevent the harm itself, this creates a *de facto* distinction between unilateral care and bilateral care cases. For example, suppose that a victim who is an expert on AI files a lawsuit and claims that he had no idea that ChatGPT is not 100% accurate. Such a claim would seem farfetched coming from an expert who likely knows quite well the dangers of blindly following AI-generated advice. The expert’s lawsuit could then technically be rejected because of a failure to invoke the presumption of a causal link, but in essence, would also reflect the expert’s contributory negligence.

In contrast, for non-high-risk AIs, the presumption is applied only when it is excessively difficult for the plaintiff to prove the causal link. This raises a concern that the AILD establishes a negligence-based regime for non-high-risk AIs and a sort of strict liability regime for high-risk systems. The reason for this distinction is that the preconditions underlying the two presumptions may be highly correlated as long as high-risk AIs cause large harms and vice versa. For instance, suppose that an AI creator refuses to disclose evidence in a high-risk AI case precisely because it knows the stakes are high but agrees to do so in cases involving non-high-risk AIs. The AI creator would also invest more resources to prove the claimant’s ability to prove the causal link if the stakes are high. Moreover, when the harms are large and obvious, it might be easier to attribute them to the defendant’s fault based on the argument that the defendant should have anticipated such large harms. For all these reasons, high-risk AI cases may lead to the triggering of both assumptions far more easily than in other cases. Thus, effectively, defendants in high-risk AI cases may be subject to strict liability more often, whereas defendants in low-risk AI cases would often be subject to negligence. The problem with this dichotomy is that there is no clear thread connecting the magnitude of the risk to the cases where bilateral care applies. Consequently, applying strict liability and negligence to different systems without differentiating between situations of unilateral and bilateral case may well lead to inefficiencies.

2. *Levels of Activity*

The AILD does not seem to be particularly concerned with activity levels. Nothing in the text suggests that the quantity or intensity of the activity is taken

into consideration. Still, one may ponder whether the risk-based approach also captures some attributes that correlate with the level of activity. For instance, consider the list in the AI Act's Appendix III, which specifies types of activity to be classified as high-risk.

Consider, for example, biometric systems, which are included in this list.³¹² Such systems can clearly be deployed to biometrically classify many individuals in many different contexts and potentially reflect a high level of activity. Similarly, systems used for education (another category) can be applied to many students across many institutions. Administration of justice can be similarly applied in many courts. Thus, one could make the argument that what makes a system high-risk is, at least in part, its potential to be used in a high level of activity. This would help explain why an approach closer to strict liability, which is more effective for attaining efficient activity levels of the injurer, is chosen for high-risk AIs.

This may also help explain why the AILD determines that a defendant who used the AI system for personal use, which is more likely to reflect a low level of activity, is subject to a regime closer to negligence where the presumption of a causal link applies only if he actively interfered with the system (and is therefore arguably at fault). At the same time, this would not explain why the PLD exclusively applies strict liability, disregarding the level of activity.

3. *Risk Attitudes and Insurance*

The AILD notes that the to-be-discussed harmonization of strict liability might be “possibly coupled with a mandatory insurance” of defendants.³¹³ The logic seems to be grounded in the protection of the plaintiffs from potential judgment-proof problems (thereby supporting liability over regulation), but this also means that strict liability will be accompanied by explicit insurance (rather than only implicit insurance) on the part of the victims. If victims of AI are risk-averse, and if market insurance for victims is unavailable, a mandatory insurance mechanism would indeed be consistent with the idea of optimizing the residual risk bearer such that risk-averse people do not end up bearing risk.

At the same time, if mandatory insurance is imperfect, some distortions would arise. For instance, the AI creators may misbehave (a moral hazard problem) and claim retroactively that they took precautions. In this sense, if the EU decides to eventually implement mandatory insurance, it would be important to verify that it does not aggravate the problem.

312. See *AI Act Amendments supra* note 221, Annex III, § 1 (referring to “biometric and biometrics-based systems” with some exceptions).

313. *AILD, supra* note 26, Explanatory Memorandum at 9. See also *id.* art. 5 (2) (mentioning that the European Commission should review the need for insurance coverage five years after the end of the transposition period, establish a monitoring program, and produce a report).

4. *Incentives to Research Unknown Risks*

Recall that strict negligence gives optimal incentives to search for new information on potential risks, whereas negligence may or may not do so, depending on how the standard of negligence is set. In the AILD, the (more or less) strict liability regime, through its rebuttable presumptions, gives AI creators good incentives to invest in figuring out which harms their product may cause. The one exception is the (more or less) fault-based regime for defendants who use AI for personal use. Such defendants can escape the presumption of a causal link as long as they do not actively interfere with the AI. However, interference may be precisely what one needs to gather information on unknown risks. In this sense, conditioning the presumption on the lack of intervention may be counterproductive.

Furthermore, consider the exemption from liability under the PLD regarding the objective state of knowledge.³¹⁴ Such a provision provides strong incentives *not* to acquire knowledge as long as the current state of things is such that the risk is still unknown.

In any case, the proposals do not seem particularly concerned with locating the least-cost information gatherer (nor the least-cost avoider).

D. PRESUMING A CAUSAL LINK

As a more general question, one may wonder whether presuming the causal link is efficient. Intuitively, AI might be too complex for any layman to be able to prove a full causal link. For instance, even if Jack the lawyer can easily prove that the information produced by ChatGPT is incorrect, linking that to a faulty algorithmic design seems difficult. A presumption then clearly saves on tertiary costs (administrative costs), as the court does not need to invest resources to ascertain the causal link. It may also reduce secondary costs (of risk-allocation) if the victim is risk-averse, whereas the injurer is risk-neutral (as the presumption shifts the risk to the injurer). But does the presumption also save on primary costs, namely, incentives to take precautions?

It is a well-known consequence that requiring victims to prove a full causal link may create suboptimal incentives for injurers to take care.³¹⁵ Namely, if AI creators know they can get away with harm because the victim cannot prove the causal link, the level of care would drop. Correspondingly, if the burden of proof is reversed—as is the case under the AILD—the AI creator may never be able to disprove the causal link because of the same complexity.³¹⁶ In this case, the question is whether this would cause overdeterrence, leading the AI creator to

314. *See supra* Part V.D.

315. HANS-BERND SCHÄFER & CLAUS OTT, *THE ECONOMIC ANALYSIS OF CIVIL LAW* 261–62 (2022).

316. *See id.* (“[I]f under negligence and uncertain causation the burden of proof is reverse, [then disproving the causal link] . . . is almost impossible.”).

take excessive care or to reduce his activity levels. Here, the law and economics literature identify a subtle difference between negligence and strict liability: under negligence, the injurer is exempted once he takes precautions, irrespective of the causal link element. Therefore, negligence should not give rise to overdeterrence (assuming the standard is efficiently determined), as the injurer can escape liability by proving he took precautions. Conversely, under strict liability, there will be overdeterrence, because the injurer cannot escape liability even by taking precautions,³¹⁷ so that the causal link matters. As the AILD allows for strict liability, and as the PLD does so as a general rule, the presumption of a causal link may plausibly lead to overdeterrence.³¹⁸ Thus, there is a concern that the presumption of a causal link will increase primary costs.

E. LEGAL UNCERTAINTY

A final consideration worthy of a brief discussion concerns legal uncertainty. While the EU's frameworks are well-intended and might even assist in clarifying some points that would otherwise be vague, the conjunction of three pieces of interrelated proposals may cause a great deal of confusion. Both victims and injurers might not be clear on which liability standard applies in many cases, including ChatGPT. The current proposals leave room for no liability (under the PLD's exemptions), fault-based liability (under the AILD), and strict liability (under the AILD or PLD)—all for similar instances. This, by itself, may lead to incentive distortions along the lines discussed in this Article with respect to imperfect insurance, including overdeterrence and underdeterrence. In fact, these are not very different from the concept of administrative costs (as the courts will have to decide which standard to apply), which usually speaks in favor of liability. Here, however, such costs may speak against it—at least in the current form of a triad of proposals.

CONCLUSION

ChatGPT has taken the world by storm, leading to intense public debate about the challenges created by the rise of AI. Although the technology is moving at light speed, the United States is lagging on the policy front, gradually releasing general recommendations without a comprehensive regulatory framework. In contrast, the EU has undergone a lengthy institutional discussion

317. *Id.*

318. In a recent article, Steven Shavell proposed to change the test of causality: instead of looking at a causal link between the negligence (once it is proven) and the harm, one should look only at a causal link between the harm and the risky activity. Steven Shavell, *An Alternative to the Basic Causal Requirement for Liability under the Negligence Rule* (Nat'l Bureau of Econ. Rsch., Working Paper No. 31219, 2023). He argues that this would not lead to overdeterrence, because liability still requires negligence, so parties can escape liability by taking precautions. *Id.* at 19. A full discussion of this point for ChatGPT is beyond the scope here, but under the current causality tests, overdeterrence may well emerge.

on regulating AI, yielding proposals for a framework that combines regulation and liability.




This Article utilizes traditional concepts from law and economics to evaluate the existing proposals, focusing on three main questions. First, do the proposals correspond to a market failure? Second, should ChatGPT (and other AIs) be restrained using public regulation, tort liability, or a mixture of the two? And third, should the liability regime be based on strict liability or negligence?

The analysis reveals several inconsistencies between the existing proposals and the principles of law and economics. In particular, the EU's choice to base its policy on a risk-based approach does not map neatly onto the relevant considerations, such as whether a market failure occurred, whether the victim could influence the likelihood of an accident, or whether the risks are known or unknown.

The goal of this analysis is not necessarily to provide a concrete recommendation on how policymakers should proceed, as the existing proposals seem to be, to a large extent, a work in progress. Rather, the argument is that the current efforts do not give sufficient weight to questions of incentives, which are at the center of law and economics, so some provisions hit the target while others miss it. Thus, let the suggestion drawn from this conclusion be only this: policymakers should reevaluate whether their proposals to restrain AI are consistent with law and economics, paying attention to whether the relevant distinctions are translated into policy. Unless this is done, the proposals are susceptible to distortions of incentives and inefficiencies.

APPENDIX

RO Please compare the law regarding good-faith purchases of mobile items in the United States and in Europe

 In the United States, the Uniform Commercial Code (UCC) governs the sale of goods, including mobile items such as vehicles. Under the UCC, a good-faith purchaser for value acquires clear title to the item, free from any defects or liens, as long as they did not have knowledge of or participate in any fraudulent activity related to the item.  

In Europe, the laws regarding good-faith purchases of mobile items can vary by country. However, many countries in Europe have laws similar to the UCC. For example, the Sale of Goods Act in the United Kingdom provides that a purchaser in good faith and for value acquires a good title to the goods, free from any third party claims.

In summary, both the US and many European countries have laws that protect good-faith purchasers of mobile items from defects and liens, as long as they did not have knowledge of or participate in any fraudulent activity related to the item.

Figure 1: ChatGPT's response to my question (Jan. 20, 2023)