MACHINES MAKE MISTAKES TOO: PLANNING FOR AI LIABILITY IN CONTRACTING

Mark A. Sayre, J.D.[*] *& Kyle Glover, Esq.*[**]

ABSTRACT

Recent advances in artificial intelligence have set off a frenzy of commercial activity, with companies fearful that they may fall behind if they are unable to quickly incorporate the new technology into their products or their internal processes. At the same time, numerous scholars from the machine learning community have warned of the fundamental risks that uninhibited use of artificial intelligence poses to society. The question is not whether artificial intelligence will cause harm, but when, and how. The certainty of future harm necessitates that legal scholars and practitioners examine the liability implications of artificial intelligence. While this topic has been given increasing focus in the literature, such discussion is lacking in two key ways. First, there has been little attempt to consolidate the literature on the range of legal theories that might apply to harm resulting from the use of artificial intelligence. Second, the literature has failed to address the role that contracting may play in reducing uncertainty around liability and overriding common law approaches. This paper addresses both gaps in the literature and provides legal practitioners with an overview of key considerations related to liability allocation when contracting for artificial intelligence technology.

Part I of the paper begins by briefly discussing the risks inherent in the use of artificial intelligence, including in particular risks resulting from a lack of transparency and explainability, and the harms that might result. Part II of the paper distills past legal scholarship on the legal theories that might apply when harm results from the use of artificial intelligence. The theories analyzed include vicarious liability, products liability and

[*] J.D., University of Maine School of Law, Class of 2024.

[**] Partner, Pierce Atwood LLP and Adjunct Prof. of Law, Georgetown University Law Center. The authors would like to thank the following individuals for their support, feedback and edits: Prof. Scott Bloomberg, Prof. Tina Saunders, George Martin, Bill Goddard, and Tom Graham.

negligence. Relevant distinctions between artificial intelligence and software are discussed as they relate to the application of products liability and negligence theories in particular. Part II closes by highlighting that the current uncertainty in the legal landscape for artificial intelligence liability incentivizes contracting parties to address liability directly within their contracts. Part III of the paper then proceeds to provide an overview of important considerations for contracting parties when using contractual apportionment of liability to reduce uncertainty around harm resulting from the use of artificial intelligence. These considerations are organized by contracting phase and by relevant contracting section.

MACHINES MAKE MISTAKES TOO: PLANNING FOR AI LIABILITY IN CONTRACTING

TABLE OF CONTENTS

INTRODUCTION

Artificial Intelligence (AI) is hot right now. Whether it's ChatGPT,[1] AI artists,[2] or deepfake videos,[3] the news is filled with articles about the explosion of AI technology and the profound effects such technology is likely to have on society.[4] Public perception of AI tends to focus on "moon shot" scenarios in which sweeping, all-knowing machines will solve the world's most complex problems, and, in the process, put millions of people out of work.[5] But, a more grounded approach to thinking about AI would focus instead on the current and ongoing implementation of AI into the day-to-day operations of businesses and organizations, such as robotic process automation, advanced analytics, and triage within customer service processes through the use of chatbots.[6] Many businesses have

[1] *See*, *e.g.*, Sabrina Ortiz, *What Is ChatGPT and Why Does It Matter? Here's What You Need to Know*, ZDNET (Feb. 20, 2024, 5:20 AM), https://www.zdnet.com/article/what-is-chatgpt-and-why-does-it-matter-heres-everything-you-need-to-know/ [https://perma.cc/5Q37-Y68Q].

[2] *See*, *e.g.*, Alex Greenberger, *Artist Wins Photography Contest After Submitting AI-Generated Image, Then Forfeits Prize*, ARTNEWS (Apr. 17, 2023, 1:08 PM), https://www.artnews.com/art-news/news/ai-generated-image-world-photography-organization-contest-artist-declines-award-1234664549/ [https://perma.cc/KZ2G-RRSE].

[3] *See*, *e.g.*, Shannon Bond, *It Takes a Few Dollars and 8 Minutes to Create a Deepfake. And That's Only the Start*, NPR (Mar. 23, 2023, 5:00 AM), https://www.npr.org/2023/03/23/1165146797/it-takes-a-few-dollars-and-8-minutes-to-create-a-deepfake-and-thats-only-the-sta [https://perma.cc/VBS5-ZFSN].

[4] *See*, *e.g.*, German Lopez, *Using A.I. in Everyday Life*, N.Y. TIMES (Apr. 21, 2023), https://www.nytimes.com/2023/04/21/briefing/ai-chatgpt.html [https://perma.cc/5HW4-8BPS].

[5] The reality of whether and how many human jobs will disappear as a result of AI is complex. While it is likely that many jobs will no longer be necessary, others argue that the proliferation of AI will require a similar if not greater number of new jobs that are likely to be filled by humans. *See*, *e.g.*, Charles Simon, *As AI Advances, Will Human Workers Disappear?*, FORBES (June 28, 2022, 8:30 AM), https://www.forbes.com/sites/forbestechcouncil/2022/06/28/as-ai-advances-will-human-workers-disappear/?sh=5fbb5cda5e68 [https://perma.cc/QEV6-AFRG] ("In its recent Future of Jobs Report, the World Economic Forum estimated that AI will replace some 85 million jobs by 2025. The same report, however, concluded that some 97 million new jobs would be created in the same timeframe due to AI."). Some scholars point to the role that human judgment plays in complementing AI and predict that the need for human judgment will grow proportionately with greater access to AI. *See* Avi Goldfarb & Jon R. Lindsay, *Prediction and Judgment: Why Artificial Intelligence Increases the Importance of Humans in War*, 46 INT'L SEC. 7, 9 (2022).

[6] *See* Thomas H. Davenport & Rajeev Ronanki, *Artificial Intelligence for the Real World*, HARV. BUS. REV. (Jan.–Feb. 2018), https://hbr.org/2018/01/artificial-intelligence-for-the-real-world [https://perma.cc/3ZSY-W3FV].

already implemented these narrower forms of AI,[7] which are therefore likely to be most relevant for advancement of society both in the short and medium term.[8]

One primary reason for the greater short-term relevance of narrower forms of AI, compared to large-scale general AI, is limitations on the ability to train and maintain large-scale AI (i.e., access to sufficient data and processing power).[9] Overcoming such limitations requires either money, partners, or both.[10] And while there are many potential ways for AI developers to raise money, including by engaging investors who also bring processing power or data (e.g., Microsoft Azure and OpenAI), it is likely that many will focus on raising revenue by commercializing their AI through relationships with other businesses and AI developers.[11] This increased focus on commercialization of AI technologies in the short-term is likely to set off a frenzy of contracting activities.[12]

This paper focuses on the specific context of AI utilization or licensing contracts, in which the owner/developer of an AI technology ("Vendor") enters into a licensing and use agreement with another entity ("User") for use of the AI model or system.[13] However, the analysis extends to multi-party agreements containing multiple Users and multiple

[7] The annual McKinsey Global Survey on AI shows that between fifty and sixty percent of respondent businesses say their organization has adopted AI in at least one business unit or function. *The State of AI in 2022 — and a Half Decade in Review,* MCKINSEY: QUANTUMBLACK AI (Dec. 6, 2022), https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review [https://perma.cc/3U4W-N4YA].

[8] *See* J.E. (Hans). Korteling et al., *Human- versus Artificial Intelligence*, FRONTIERS IN A.I., Mar. 25, 2021, at 1, 8.

[9] *ChatGPT Mania May Be Cooling, but a Serious New Industry Is Taking Shape*, THE ECONOMIST (Sept. 21, 2023), https://www.economist.com/leaders/2023/09/21/chatgpt-mania-may-be-cooling-but-a-serious-new-industry-is-taking-shape [https://perma.cc/FD2P-565M].

[10] *See id.*

[11] *See, e.g., Could OpenAI Be the Next Tech Giant?*, THE ECONOMIST (Sept. 18, 2023), https://www.economist.com/business/2023/09/18/could-openai-be-the-next-tech-giant [https://perma.cc/9MDQ-NZ3K].

[12] *See* The Economist*, supra* note 9.

[13] This article focuses solely on AI contracting between private parties. Although contracting for AI by governmental entities is rapidly expanding, such use presents a broader set of legal issues, including constitutional law issues, that are beyond the scope of this analysis. *See* Grant Fergusson, *Outsourced and Automated: How AI Companies Have Taken Over Government Decision-Making,* ELEC. PRIV. INFO. CTR. (Sept. 2023), https://epic.org/wp-content/uploads/2023/09/FINAL-EPIC-Outsourced-Automated-Report-w-Appendix-Updated-9.26.23.pdf [https://perma.cc/BB2A-YCN6].

Vendors (although the issues presented here likely multiply or compound in such scenarios).

The paper first examines some of the unique risks posed by AI technology and the harms that may result, drawing upon a combination of scholarship and reporting from both the technical and legal communities. Next, the paper examines potential liability resulting from the use of AI. Legal scholars have explored AI liability under a range of theories, including vicarious liability,[14] products liability[15] and negligence[16]. This paper distills the scholarship on these topics and provides new insights on the challenges of AI liability under current legal frameworks to emphasize the value that contractual allocation of liability risks provides in reducing uncertainty for companies. Finally, the paper discusses contracting techniques to address AI liability. Current literature on contractual negotiations related to AI liability is concentrated in high-level guidance by practitioners in the form of client alerts or blog posts.[17] This paper provides a more detailed discussion of contracting considerations by drawing on the risks outlined in the first section and the analysis of AI liability under agency and tort theories in the second section.

[14] *See, e.g.*, Jason Chung & Amanda Zink, *Hey Watson — Can I Sue You for Malpractice? Examining the Liability of Artificial Intelligence in Medicine*, 11 ASIA PACIFIC J. HEALTH L. & ETHICS 51 (2018). *See also* Anna Beckers & Gunther Teubner, *Responsibility for Algorithmic Misconduct: Unity or Fragmentation of Liability Regimes?*, 25 YALE J. L. & TECH. (SPECIAL ISSUE) 76 (2023).

[15] *See* Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J. L. & TECH. 889, 931 (2018); Karni A. Chagal-Feferkorn, *Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision-Makers*, 30 STAN. L. & POL'Y REV. 61, 77 (2019); Mindy Nunez Duffourc, *Malpractice by the Autonomous AI Physician*, 2023 U. ILL. J. L. TECH. & POL'Y 1, 18–20 (2023); Cassandra Burke Robertson, *Litigating Partial Autonomy*, 109 IOWA L. REV. (forthcoming 2023).

[16] *See, e.g.*, Andrew D. Selbst, *Negligence and AI's Human Users*, 100 B.U. L. Rev. 1315 (2020). *See also* Jan De Bruyne et al., *Tort Law and Damage Caused by AI Systems*, *in* ARTIFICIAL INTELLIGENCE AND THE LAW 359, 370 (Jan De Bruyne & Cedric Vanleenhove eds., 2021).

[17] *See, e.g.*, Pieter-Jan Aerts et al., *Contracting for the Purchase and Use of AI*, 2022 DENTONS A.I. GUIDE 30 (2021); Lisa R. Lifshitz, *Avoiding AI Agreement Dystopia: Managing Key Risks in AI Licensing Deals*, A.B.A.: BUS. L. TODAY (Sept. 4, 2023), https://www.americanbar.org/groups/business_law/resources/business-law-today/2023-september/avoiding-ai-agreement-dystopia-managing-key-risks-in-ai-licensing-deals/ [https://perma.cc/3QCY-N3ZQ]; Alexa Delaney Christianson et al., *Contracting for AI Technologies — Top Five Best Practices,* JDSUPRA (Nov. 16, 2023), https://www.jdsupra.com/legalnews/contracting-for-ai-technologies-top-3387165/ [https://perma.cc/FD67-2RR9].

I. AN OVERVIEW OF AI AND ITS RISKS

A. A Brief Overview of AI

Any discussion about AI must first start with what AI is—and what it isn't. The Oxford English Dictionary defines artificial intelligence as "[t]he theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages."[18] Merriam-Webster defines artificial intelligence as "[t]he capability of computer systems or algorithms to imitate intelligent human behavior."[19] Both definitions point to a key insight—that artificial intelligence is distinct and different from human intelligence, even though its ability to "imitate" human behavior by performing tasks normally associated with intelligence may render such distinctions difficult to perceive.[20]

As AI has increased in complexity and its ability to solve problems, AI experts have begun to differentiate between "Narrow AI" and "Artificial General Intelligence" (AGI).[21] A key distinction between Narrow AI and AGI is that while the former is only capable of performing the specific task for which it is trained, the latter is capable of independently performing new tasks beyond those on which it was trained and doing so in new and

[18] *Artificial Intelligence*, THE OXFORD DICTIONARY OF PHRASE AND FABLE (2d ed. 2005).

[19] *Artificial Intelligence*, MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/artificial%20intelligence [https://perma.cc/Q5JC-9P6Y] (Apr. 4, 2024).

[20] This concept of imitation can be found in the famous Turing test, under which a computer is deemed to imitate human intelligence if it is as successful as a human at deceiving a second human with whom both the computer and the first human are separately communicating. *See* Rembrandt Devillé et al., *Basic Concepts of AI for Legal Scholars*, *in* ARTIFICIAL INTELLIGENCE AND THE LAW 1, 3 (Jan De Bruyne & Cedric Vanleenhove eds., 2021).

[21] Some even propose a third category called Super AI or artificial superintelligence. *Understanding the Different Types of Artificial Intelligence*, IBM (Oct. 12, 2023), https://www.ibm.com/blog/understanding-the-different-types-and-kinds-of-artificial-intelligence/ [https://perma.cc/5JKE-KDJU]. However, there is disagreement among AI experts as to whether AGI and Super AI are purely theoretical concepts or whether such types of AI could exist, either now or in the likely future. *See id.* (arguing that even ChatGPT is Narrow AI and not AGI because it is "limited to the single task of text-based chat"). *But see* Eliza Strickland & Glenn Zorpette, *The AI Apocalypse: A Scorecard*, IEEE SPECTRUM (June 21, 2023), https://spectrum.ieee.org/artificial-general-intelligence [https://perma.cc/9LV8-MHKR] (listing opinions of several experts who believe that an AGI is likely or at least possible in the future).

different contexts.[22] In this respect, AGI is able to learn and perform similarly to humans.[23]

In addition to being categorized by level (AI vs. AGI), AI can also be categorized by type. The two broadest categories are "Traditional AI" and "Generative AI" ("Gen AI").[24] The primary differences relate to the technology's (1) capabilities (while Traditional AI is focused primarily on recognizing patterns from data and making predictions based on such patterns, Gen AI involves creating something entirely new) and (2) applications (while Traditional AI excels in task-specific applications such as scoring models or recommendation engines, Gen AI excels in applications that require creativity and innovation).[25] Of course, these two broad types of AI are not mutually exclusive;

[22] *See* IBM, *supra* note 21.

[23] *See id.* Super AI, then, means AI that surpasses human cognitive abilities and has needs and desires of its own. *See id.* Some scholars have remarked that categorizing AI by comparing its abilities to those of humans is too anthropocentric for two reasons: (1) such a framing ignores the diversity of biological intelligence, which can be found in various forms in a wide range of species; and (2) there are likely to be significant differences between biological and artificial (or computer-based) intelligence with respect to basic structure, speed, connectivity, updatability, scalability, and efficiency. *See* Korteling et al., *supra* note 8, at 2, 5 (suggesting an alternate definition of AI as a "non-biological capacity to realize complex goals"). Accordingly, they argue that such differences render any analogies between artificial intelligence and human intelligence "very misleading." *See id.* at 6. One example of this is Moravec's Paradox, or the distinction between task difficulty (a subjective, human-centric measure) and task complexity (an objective measure); tasks which are extremely difficult for humans (e.g., computations involving large numbers) may be computationally simple and thus easy for computers, while tasks which are easy for humans (e.g., walking) may be objectively quite complex and thus difficult for computers. *See id.*

[24] *See* Bernard Marr, *The Difference Between Generative AI and Traditional AI: An Easy Explanation for Anyone*, FORBES (July 24, 2023, 1:41 AM), https://www.forbes.com/sites/bernardmarr/2023/07/24/the-difference-between-generative-ai-and-traditional-ai-an-easy-explanation-for-anyone/?sh=74f328e4508a [https://perma.cc/Y8ZU-VPEY].

[25] *See id.* Perhaps the most famous Generative AI in the world is ChatGPT, the fastest-growing software program in history at 100 million active users within its first two months. *See* Benj Edwards, *ChatGPT Sets Record for Fastest-Growing User Base in History, Report Says*, ARS TECHNICA (Feb. 1, 2023, 5:57 PM), https://arstechnica.com/information-technology/2023/02/chatgpt-sets-record-for-fastest-growing-user-base-in-history-report-says/ [https://perma.cc/69VT-CTC7]. However, ChatGPT is only one of an increasingly large number of Generative AI tools; others include StabilityAI's Stable Diffusion, Google's Bard, OpenAI's DALL-E, and open-source alternatives such as UC Berkeley's Koala. *See* Tiernan Ray, *Generative AI Will Far Surpass What ChatGPT Can Do. Here's Everything on How the Tech Advances*, ZDNET (Oct. 2, 2023, 10:30 AM), https://www.zdnet.com/article/generative-ai-will-far-surpass-what-chatgpt-can-do-heres-everything-you-need-to-know-how-the-tech-advances/ [https://perma.cc/UV8Q-

Traditional and Gen AI could be combined together in particularly effective ways such as analyzing patterns in a particular user's behavior and then creating customized designs to appeal to that user.[26]

This paper will focus on the licensing of AI technologies, whether Traditional or Gen AI.[27] For the purposes of this paper, a specific definition of AI is not necessary. Rather, it is helpful to view AI as existing somewhere along a spectrum depending on its capabilities with respect to a few potentially overlapping dimensions: (1) the source of its decision-making capabilities (whether human-provided or self-taught)[28]; (2) its capability and mechanisms for learning over time; and, (3) its ability to adapt and act autonomously in new or unexpected scenarios.[29] Where applicable, the following sections will consider the impact that the type of AI, or its position along these dimensions, may have on liability and contract terms.

B. Risks of AI

It is tempting to believe that the automated and programmable nature of AI renders it immune to the risks often associated with human judgment, such as mistakes, inconsistency, and bias. However, research has increasingly shown such a belief to be

BZZJ]. Nearly two-thirds of Americans are now familiar with Generative AI and nearly a quarter have used a Gen AI tool. *Most Americans Support Regulating Generative AI*, THE HARRIS POLL (May 10, 2023), https://theharrispoll.com/briefs/regulating-generative-ai/ [https://perma.cc/LD3R-HCCT].

[26] *See* Marr, *supra* note 24.

[27] This paper will not cover the use of AI in drafting contracts or assisting lawyers in contract management. Although Gen AI is particularly helpful for contract drafting, the topic has already been covered by prior scholarship. *5 Ways Generative AI for Contracts Can Boost Legal Operations,* LEXOLOGY (June 26, 2023), https://www.lexology.com/library/detail.aspx?g=da3c5550-2435-4a9a-aa69-36ed94ca6c32 [https://perma.cc/TZ96-9WWM]. *See also* Florian Martin-Bariteau & Marina Pavlović, *AI and Contract Law*, *in* ARTIFICIAL INTELLIGENCE AND THE LAW IN CANADA (Florian Martin-Bariteau & Teresa Scassa eds., 2021); John Linarelli, *Artificial Intelligence and Contract Formation: Back to Contract as Bargain?*, *in* EMERGING ISSUES AT THE INTERSECTION OF COMMERCIAL LAW AND TECHNOLOGY (Stacy-Ann Elvy & Nancy Kim eds., forthcoming 2023).

[28] This dimension is intended to capture two concepts: (1) knowledge-based versus data-based learning; and (2) supervised versus unsupervised learning (and anything in between such as semi-supervised learning). *See* Devillé et al., *supra* note 20, at 4–7.

[29] *Artificial Intelligence: A Primer*, THE CONF. BD. (May 17, 2023), https://www.conference-board.org/publications/artificial-intelligence-a-primer [https://perma.cc/W2FL-QWVD].

false.[30] For example, scholars and policymakers have dedicated significant effort over recent years to classifying and understanding the biases that can occur within AI and outlining technical, procedural, or regulatory safeguards that can reduce the risk that such biases lead to discrimination or harm.[31] There is some evidence that such efforts may already be overdue; in a recent survey of 1,580 executives of large companies from more than ten countries, 41% of executives reported abandoning an AI system altogether as a result of ethical concerns.[32]

Discussion about the fallibility of AI often centers around the lack of transparency around "black box" AI models and their underlying datasets.[33] There are numerous reasons for this lack of transparency. It can be the result of intentional decisions to protect trade

[30] Researchers have documented numerous ways in which AI may generate or exacerbate errors and biases. Such issues can emerge from errors in the dataset on which the AI was trained. *See*, *e.g.*, Curtis G. Northcutt et al., *Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks* (2021), https://arxiv.org/pdf/2103.14749.pdf. Some can emerge from systemic biases in the training data that are replicated or multiplied by the AI. *See*, *e.g.*, Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, *in* 81 PROCEEDINGS OF MACHINE LEARNING RESEARCH 1 (Sorelle A. Friedler & Christo Wilson eds., 2018). Others can emerge from the implementation of AI without the "social control mechanisms" that regulate and govern how humans might perform similar tasks. *Commission White Paper on Artificial Intelligence — A European Approach to Excellence and Trust*, at 11, COM (2020) 65 final (Feb. 19, 2020).

[31] *See* Benjamin Van Giffen et al., *Overcoming the Pitfalls and Perils of Algorithms: A Classification of Machine Learning Biases and Mitigation Methods*, 144 J. BUS. RSCH. 93 (2022).

[32] *How Consumers View the Transparency of Their AI-enabled Interactions*, HELP NET SEC. (July 11, 2019), https://www.helpnetsecurity.com/2019/07/11/ai-enabled-interactions/ [https://perma.cc/HWV7-87DQ].

[33] *AI Black Box Horror Stories — When Transparency Was Needed More than Ever,* MEDIUM (Oct. 28, 2019), https://odsc.medium.com/ai-black-box-horror-stories-when-transparency-was-needed-more-than-ever-3d6ac0439242 [https://perma.cc/57YF-UFGQ]. *See also* Virginia Dignum, *On Bias, Black-boxes and the Quest for Transparency in Artificial Intelligence*, MEDIUM (Jan. 26, 2018), https://medium.com/@virginiadignum/on-bias-black-boxes-and-the-quest-for-transparency-in-artificial-intelligence-bcde64f59f5b [https://perma.cc/6PSH-XL88].

secrets and preserve competitive disadvantage[34] or to avoid unwanted scrutiny.[35] It can be merely a byproduct of the inherent complexity of machine learning techniques.[36] Or, it may be necessary to protect an AI model from cyberattacks or other incidents that could interfere with its ability to perform.[37] But, counterbalancing these reasons for less transparency are reasons for greater transparency; specifically, there is a growing recognition by both consumers and executives that transparency around businesses' use of AI is critical to maintaining consumer trust and loyalty.[38]

These competing factors that are simultaneously discouraging and encouraging transparency are referred to as AI's "Transparency Paradox."[39] This paradox informs some of the dynamics that will likely occur between parties when contracting for AI. The

[34] *See* Aparna Dhinakaran, *Overcoming AI's Transparency Paradox*, FORBES (Sept. 10, 2021, 3:56 PM), https://www.forbes.com/sites/aparnadhinakaran/2021/09/10/overcoming-ais-transparency-paradox/?sh=2d3d0d74b778 [https://perma.cc/6KEV-42KK] ("There is a growing tension between the desire for AI transparency and an organizations' interest in maintaining secrecy over their AI tools. Firstly, secrecy helps maintain their competitive advantage in the market.").

[35] *See id.* ("Some organizations may worry that disclosure of the source code, the underlying mathematical model, the training data, or simply the inputs and outputs of a machine learning model may expose them to the risk of losing customer trust, dealing with intense public scrutiny, or disruptions to the deployment and use of their machine-learned innovations.").

[36] *See id.* ("Part of this dilemma stems from the sheer technical complexity of AI systems. While it is possible to build machine learning models that are easily interpretable, simple decision trees being an example, such models aren't always helpful to achieving complex tasks or objectives.").

[37] *See* Andrew Burt, *The AI Transparency Paradox*, Harv. Bus. Rev. (Dec. 13, 2019), https://hbr.org/2019/12/the-ai-transparency-paradox [https://perma.cc/69LU-NMH8] ("At the same time, however, it is becoming clear that disclosures about AI pose their own risks: Explanations can be hacked, releasing additional information may make AI more vulnerable to attacks . . .").

[38] *See* ANNE-LAURE THIEULLENT ET AL., WHY ADDRESSING ETHICAL QUESTIONS IN AI WILL BENEFIT ORGANIZATIONS, 23 (Capgemini Research Institute 2019).

[39] *See* Dhinakaran, *supra* note 34. *See also* Burt, *supra* note 37. Andrew Burt suggests that the inclusion of lawyers in the AI development and review process can "facilitate an open and legally privileged environment" that furthers evaluation of AI for vulnerabilities while limiting exposure to additional liability. *Id.* A relevant parallel when two or more companies or entities partner on the development, purchase or licensing of AI technology may be a strong Non-Disclosure Agreement (NDA); however, while such an agreement may increase confidence that problematic findings would not be leaked to the public, it may not protect such findings from discovery during potential future litigation. *See* Jared S. Sunshine, *The Secrets of Corporate Courtship and Marriage: Evaluating Common Interest Privilege When Companies Combine in Mergers*, 69 S.C. L. REV. 301, 318–20 (2017) (discussing common interest privilege under non-disclosure agreements in the context of a merger).

User will want to gather as much information about the AI as possible during the due diligence phase (and beyond) to assess its legal, reputational, and operational risks. The Vendor will want to reveal as little as possible about the inner workings of the AI to protect its intellectual property and preserve its competitive advantage. There is some evidence however that the lack of transparency in a Vendor's model may make it difficult for the Vendor to compete with open-source solutions due to concerns about data privacy or a desire to exercise more control over the AI's behavior.[40] Moreover, if (or when) something goes wrong, the disparity in knowledge between the User and the Vendor as to the model's inner workings may place the User at a significant disadvantage in resolving disputes as to liability. This disparity may further a demand by Users for greater transparency.

The Blueprint for an AI Bill of Rights, released by the White House in Fall 2022, addresses transparency under two of its five principles: (1) Safe and Effective Systems; and (2) Notice and Explanation.[41] However, the Blueprint does not contain a definition of or specific criteria for transparent AI. Fortunately, more guidance does exist outside the United States. The European Commission has defined three minimum requirements for transparent AI: traceability, explainability, and communication.[42] Traceability requires that owners of AI identify and document the "data sets and the processes that yield the AI system's decision," including data gathering and labeling methods and the algorithms used.[43] Traceability increases both the auditability and the explainability of the AI, and

[40] *See* Chris Tozzi, *The Data Privacy Risks of Third-party Enterprise AI Services*, TECHTARGET (Oct. 18, 2023), https://www.techtarget.com/searchenterpriseai/tip/The-data-privacy-risks-of-third-party-enterprise-AI-services#:~:text=Third%2Dparty%20AI%20vendors%20might,confines%20of%20their%20IT%20estate [https://perma.cc/WA2A-HXSD].

[41] *Safe and Effective Systems*, WHITE HOUSE, https://www.whitehouse.gov/ostp/ai-bill-of-rights/safe-and-effective-systems-3/ [https://perma.cc/3P4B-CZAV] (last visited Dec. 2, 2023) (discussing the importance of documenting the source of data, especially derived data, and reporting that includes a description of the AI, how the AI uses data, and the results of any testing performed on the AI); *Notice and Explanation*, WHITE HOUSE, https://www.whitehouse.gov/ostp/ai-bill-of-rights/notice-and-explanation/ [https://perma.cc/B98F-UUFX] (last visited Dec. 2, 2023) (indicating that entities deploying AI should provide clear and understandable notices of use that explain how and why a decision was made or an action was taken by the AI).

[42] EUR. COMM'N INDEP. HIGH-LEVEL EXPERT GRP. ON A.I., *Ethics Guidelines for Trustworthy AI*, at 14 (Apr. 8, 2019).

[43] *Id.* at 18.

allows AI owners to understand and remedy erroneous decisions by the AI system.[44] Explainability refers to the "ability to explain both the technical processes of the AI system and the related human decisions."[45] Related human decisions can be understood to mean the design choices of the AI system, the rationale behind deploying it, and the ways that the AI system influences organizational decision-making.[46] Finally, communication refers to identifiability of AI systems, such that humans are aware that they are engaging, directly or indirectly, with an AI system, and the publication of information regarding the AI's accuracy and limitations.[47]

Of course, AI technologies can also be fallible in other ways unrelated to their lack of transparency,[48] many of which are relevant for the purposes of analyzing liability. For example, AI, especially self-learning AI, can be highly unpredictable and volatile.[49] At a more basic level, an AI could be flawed simply because it was trained by practitioners who were not sufficiently educated in AI modeling or statistics (a problem that is likely to

[44] *Id.*

[45] *Id.* The report explicitly states that trade-off decisions may be required during model development between model accuracy and model explainability. *Id.* These trade-off decisions will be especially difficult within the liability context, as an increase in model accuracy is likely to reduce the risk of liability overall while a corresponding decrease in explainability may complicate the ability to defend against a suit if one does occur. *Id.*

[46] *Id.*

[47] *See id.* Some scholars have argued that focusing on explainable AI is misguided because explainability is viewed from the perspective of human decision-making, which itself is imbued with implicit biases or attitudes. *See* Korteling et al., *supra* note 8, at 7 (suggesting that requiring explainability or transparency may constrain AI to only those benefits that can be understood by humans, thus limiting their potential). These scholars suggest instead that trust in AI technologies be dependent on its objective performance against empirically derived validation measures, rather than on the transparency of its decision-making processes. *See id. See also* Cynthia Dwork & Martha Minow, *Distrust of Artificial Intelligence: Sources & Responses From Computer Science & Law*, 151 DAEDALUS 309, 311–13 (2022) (arguing that trust in AI may be supported by regulation that simply requires entities using AI to consider broader public concerns prior to use).

[48] *See, e.g.*, Northcutt et al., *supra* note 30.

[49] *See* Stephen Ornes, *The Unpredictable Abilities Emerging from Large AI Models*, QUANTA MAG. (Mar. 16, 2023), https://www.quantamagazine.org/the-unpredictable-abilities-emerging-from-large-ai-models-20230316/ [https://perma.cc/NH48-KLYX]. Such unpredictability may limit the effectiveness of pre-implementation risk mitigation techniques. The injection of large quantities of new data into an AI system could render moot all the analysis and testing performed during the training phase. *See id.*

grow with the proliferation of open-source tools).[50] Alternatively, the training of the AI or the AI itself may violate the law.[51] For these reasons, the European Union lists liability-related issues as one of the main risks associated with the implementation of AI, alongside the application of rules to protect fundamental rights (e.g., data privacy) and safety.[52]

[50] *See* Elliott Hoffman, *The Risks of Open-source AI*, BUS. REP. (May 16, 2023), https://www.business-reporter.co.uk/ai--automation/the-risks-of-open-source-ai [https://perma.cc/CQX5-BH64].

[51] Examples may include: (1) AI trained on data whose collection violated federal or state law, s*ee, e.g.*, Joshua A. Goland, *Algorithmic Disgorgement: Destruction of Artificial Intelligence Models as the FTC's Newest Enforcement Tool for Bad Data*, 29 RICH. J. L. & TECH. 1, 22–23 (2023) (discussing a company's use of algorithms trained on data collected in violation of the Children's Online Privacy and Protection Act); (2) AI trained on data whose collection was lawful for other purposes but not for the purpose of training AI, s*ee, e.g.*, *Ensuring the Lawfulness of the Data Processing*, CNIL (Oct. 16, 2023), https://www.cnil.fr/en/ensuring-lawfulness-data-processing [https://perma.cc/75F3-MUHH]; (3) the use of AI in violation of a law specifically prohibiting such AI, s*ee, e.g.*, Deja Davis et al., *New NYC Law Restricting Artificial Intelligence-driven Employment Tools Reveals What's to Come*, JDSUPRA (July 19, 2023), https://www.jdsupra.com/legalnews/new-nyc-law-restricting-artificial-6808542/#:~:text=Local%20Law%20144%20prohibits%20employers,bias%20within%20the%20preceding%20year [https://perma.cc/AD2Z-UT98] ("Local Law 144 prohibits employers from using an automated employment decision tool ('AEDT') in hiring, promotion, and other employment decisions, unless the employer first ensures that the tool has been audited for bias within the preceding year."); and (4) AI whose training violated a company's own stated data privacy policies, s*ee, e.g.*, Goland, *supra*, at 19–21.
The first and third types are particularly relevant within the context of AI that is explicitly trained on protected characteristics to ensure fairness. *See* Cynthia Dwork et al., *Fairness Through Awareness*, *in* ITCS '12: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference 214 (2012). In legal literature, this is related to "algorithmic affirmative action." *See* Jason R. Bent, *Is Algorithmic Affirmative Action Legal?*, 108 GEO. L. J. 803, 807–08 (2020). For the purposes of this paper, which focuses solely on contracts between private actors, the potential for AI to violate anti-discrimination laws is more relevant than concerns around constitutional violations (e.g., Equal Protection Clause). In the context of contracting, it will be critical for the User to understand what approach to fairness the Vendor took when training the model as it could significantly affect the User's potential exposure to liability and thus the User's motivations when negotiating the indemnification and representations and warranties provisions of the contract in particular. *See* discussion *infra* Section III.C.1.

[52] *See* European Commission, *supra* note 30, at 10. Specifically, the report references the potential to violate European data privacy laws and regulations, which would result in discrimination against certain individuals. *See id.* at 11, n.34 (citing in particular the General Data Protection Regulation (GDPR) and amendments to the ePrivacy Directive, and highlighting the need to monitor and assess the applicability of the GDPR to AI on an ongoing basis). Although both risks also exist when AI is not used, the absence of social behavioral controls when AI is not paired with human review, alongside the inherent scalability of AI, increases the risk that these problems will occur with greater frequency and will have broader impact. *See id.* at 11.

1. Key Differences Between AI and Software

The fallibility and transparency issues discussed above are not unique to AI. Such issues are likely to exist in any contract involving the licensing of highly advanced technology. However, the relative complexity, autonomy, and opacity of AI compared to other technology (e.g., software, with its relatively accessible and reviewable code) increases the difficulty with which liability can be quickly and objectively allocated in failure scenarios.[53] Both the User and the Vendor should understand and consider the ways in which AI differs from conventional software development and how such differences affect contracting. Key differences include: (1) in co-development or piloting scenarios, the contents and performance of AI may be unclear when the contract is executed, and the AI may take on many forms during the process; (2) the contents and performance of AI depends heavily on the training dataset, and such dataset may be held by the User, the Vendor, or by a third party; (3) the importance of know-how is even higher for AI than for software; and (4) there is greater demand for further reuse of AI, especially Gen AI which is capable of performing a wide array of tasks.[54]

Given these differences, there is significant risk in the parties merely adopting a traditional software licensing agreement or software "as a service" agreement to function as an AI licensing or services agreement. Nevertheless, parties will likely be motivated to do so because of pressure to reduce both cost and the likelihood that the contracting process becomes the cause of project delays. The considerations outlined below in Part III should be viewed as minimum considerations for parties working from a software licensing agreement to ensure that the differences between AI and software are appropriately considered.

C. Defining AI Harms

[53] *Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee on the Safety and Liability Implications of Artificial Intelligence, the Internet of Things and Robotics*, at 14, COM (2020) 64 final (Feb. 19, 2020).

[54] *See* Toki Kawase, *Guidelines for AI Use Contracts in Japan: Ensuring Smooth Contractual Relations*, MONOLITH L. MAG. (Sept. 22, 2023), https://monolith.law/en/it/ai-guidelines [https://perma.cc/WY7C-PY97]. *See also Expert Q&A on Artificial Intelligence (AI) Licensing*, WESTLAW: PRACTICAL L., https://www.mayerbrown.com/-/media/files/news/2019/01/expert-qanda-on-artificial-intelligence-ai-licensing-w0219801.pdf [https://perma.cc/X7LW-XC4U] (last visited Mar. 29, 2024).

Understanding the risks associated with AI is critical to identify and describe the potential liability resulting from its use. However, it is not the risks of AI that ultimately give rise to liability; rather, it is the specific harms that result when such risks become reality. Therefore, it is helpful (and perhaps necessary) to establish a framework for describing and categorizing AI harms before analyzing the legal landscape for AI liability.[55] Fortunately, a few possible frameworks already exist;[56] however, a full survey of possible frameworks is outside the scope of this paper. In order to simplify the analysis, this paper focuses in particular on the framework proposed by the Future of Privacy Forum (FPF),[57] which was chosen because of its incorporation of mitigation tools that can be easily applied to the AI contracting context.

At the highest level, the FPF framework categorizes AI harms across three dimensions: (1) the recipient of the harm; (2) the nature of the harm; and (3) the type of harm.[58] The

[55] In order to understand why a clear framework for describing and categorizing AI harms may be necessary, one can simply look to the numerous challenges faced by plaintiffs in bringing claims related to data privacy harms. *See* Danielle Keats Citron & Daniel J. Solove, *Privacy Harms*, 102 B.U. L. REV. 793, 799–800 (2022). Many of these challenges are directly applicable to AI; for example, while some scenarios related to AI may involve a single large harm to an individual (e.g., facial recognition error resulting in wrongful arrest), other scenarios are likely to involve many small harms repeated over long periods of time or across a large population. *See id.* at 816. Additionally, some potential harms may be deemed too speculative at the time of a potential claim, such as when the output of an AI is improperly shared with a third party but it is unclear whether the third party has used the output in an unlawful or harmful manner. *See id.* at 817–18 (discussing challenges to legal standing in the context of data breaches that present a future risk of identity theft or fraud but which has not yet occurred).

[56] *See, e.g.*, Rebecca Kelly Slaughter, *Algorithms and Economic Justice: A Taxonomy of Harms and a Fath Forward for the Federal Trade Commission*, 23 YALE J. L. & TECH. (SPECIAL ISSUE) 1 (2023); Mia Hoffmann & Heather Frase, *Adding Structure to AI Harm,* CTR. FOR SEC. AND EMERGING TECH. (July 2023), https://cset.georgetown.edu/wp-content/uploads/20230022-Adding-structure-to-AI-Harm-FINAL.pdf [https://perma.cc/K7EG-ESH3]; Renee Shelby et al., *Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction*, *in* AIES '23: Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society 723 (2023). Other AI harms taxonomies are currently under development. *AI, Algorithmic, and Automation Risks/Harms Taxonomy*, AIAAIC, https://www.aiaaic.org/projects/ai-algorithmic-risks-harms-taxonomy [https://perma.cc/E322-NUUW] (last visited Dec. 2, 2023).

[57] *Unfairness by Algorithm: Distilling the Harms of Automated Decision Making*, FUTURE OF PRIV. F. (Dec. 11, 2017), https://fpf.org/blog/unfairness-by-algorithm-distilling-the-harms-of-automated-decision-making/ [https://perma.cc/6TXQ-JJCB].

[58] *See id.*

recipient of the harm may either be an individual, a community/society, or both.[59] The nature of the harm may be: (a) illegal; (b) unfair; or (c) both.[60] The type of harm may be: (i) loss of opportunity; (ii) economic loss; (iii) social detriment; and/or (iv) loss of liberty.[61] Using this framework as a guide, FPF recommends a set of mitigation tools (FPF Mitigation Tools) across five sub-pairings from the highest risk (individual harms that are illegal) to the lowest risk (collective/societal harms without an illegal analog).[62] These tools include data methods, algorithmic design, business processes, and policies such as Data Protection Impact Assessments (DPIAs).[63]

With respect to contracting, the risk of illegal harm to individuals resulting from AI is one of the most critical factors that the contracting parties will need to consider during negotiations and drafting. However, focusing solely on illegal harms and failing to appropriately address non-legal harms would be a grave error. As reputational risk grows for large firms (especially for consumer-facing companies that may face boycotts),[64] the

[59] *See id.* For example, a discriminatory AI used in employment will certainly cause harm to any individuals who are directly affected by the AI's use (e.g., rejected for employment, terminated, or passed over for promotion). However, such AI will also indirectly harm the community at large because the discriminatory allocation of employment opportunities within the community is likely to exacerbate pre-existing inequities. This paper will focus on Individual harms, which can be most effectively addressed through the contract governing the AI's use, rather than Collective/Social harms, which are likely to be better addressed through regulation and public policy.

[60] *See id.* Note that the concept of illegal harms to individuals also has implications for the collective/societal harms category, especially given the potential for class action lawsuits. Where a single AI system is used on a large group of people and results in harm to several individuals within that group, it is practically guaranteed that common questions of both fact and law will be implicated and that the claims of harm will be substantially similar across many if not all of the harmed individuals. *See* FED. R. CIV. P. 23(a)(2)–(3).

[61] *See* Future of Privacy Forum, *supra* note 57.

[62] *See id.* Within this spectrum, whether the harm disproportionately impacts individuals in protected classes may be critical to distinguish between illegal and unfair-but-not-illegal harm. *See id.* Accordingly, when contracting for AI, it is critical that the User and Vendor examine whether protected class data will be used in tandem with or to train the AI and/or whether there is a risk that the AI's decisions may vary significantly by protected class group. The FPF Mitigation Tools may be a great starting point for contract terms and conditions.

[63] *See id.* A Data Protection Impact Assessment is required under the EU GDPR when making decisions based on automated processing such as an AI, and when those decisions produce legal effects concerning a natural person. *See* Commission Regulation 2016/679, 2016 O.J (L. 119) 1, 53 (EU).

[64] *See* Jim Salas et al., *Strategies for Managing in the Age of Boycotts*, 22 GRAZIADIO BUS. REV. (2019). *See also* WORLD ECONOMIC FORUM, THE GLOBAL RISKS REPORT 2023, at 19–20 (18th ed. 2023).

prospect of an AI failure resulting in an unfair-but-not-illegal harm to a large group of individuals may be just as damaging or more damaging than the prospect of causing illegal harm.

II. AI LIABILITY: THEORIES AND FAULT

This section explores the question of AI liability in greater detail by considering a few potential theories of liability which may apply. After outlining the general approach taken under each theory, a discussion of the unique challenges posed by AI follows. For the purposes of this section, it is assumed that a contract either does not exist or does exist but does not contain any provisions that relate to liability. The section will focus on U.S. common law and/or statutory law, as applicable; however, where helpful, the approach taken by other jurisdictions is considered. Finally, the section turns to the question of whether the Vendor, the User, or both is liable, considering both the theories discussed earlier in the section as well as policy considerations.

At this point, it is helpful to introduce a hypothetical dispute between a User and a Vendor that can ground the analysis of liability in a specific area of the law and a particular, albeit general, set of facts. Let's assume that Vendor is a company that specializes in the development of an AI-enabled semi-autonomous robotic surgical assistant.[65] Vendor licenses this technology to the User, a surgical practice. The technology is comprised of three primary modules: (1) a perception module, which uses data collected during the operation from an array of sensors to infer the status of the procedure and the actions of the surgeon; and (2) a cognitive module, which predicts the likely next steps taken by the surgeon and makes decisions about appropriate placement of the autonomous portions of the robot; and (3) a planning module, which translates the decisions made by the cognitive model into a set of instructions then transmitted to the robot.[66] After what appeared to be a successful surgery, a patient later discovers that an accident during the surgery has resulted in permanent damage. Who is liable – the surgeon or surgical practice (User), the medical technology company (Vendor), or both? Where applicable, this section will refer to the robotic surgical assistant hypothetical and

[65] This example is modeled after the Smart Autonomous Robotic Assistant Surgeon Project. *Objectives*, SARAS, https://saras-project.eu/?page_id=158 [https://perma.cc/W7VK-JQTP] (last visited Nov. 11, 2023).

[66] *Concept and Approach*, SARAS, https://saras-project.eu/?page_id=99 [https://perma.cc/SJ6B-GVV8] (last visited Nov. 11, 2023).

the issues it presents in order to elucidate the challenges posed when applying current legal theories to AI.

A. Vicarious Liability: AI as Agents

Legal systems frequently determine liability issues through the concept of agency. For example, under common law, individuals harmed by an employee's actions may seek compensation from the employer if certain criteria are met.[67] This is rooted in the idea that the employee serves as an agent for the employer. Agency relationships also exist outside of the traditional employer-employee context; for an agency relationship to form, the agent must "consent[] to act on behalf of [a] principal," and the principal must have "the right throughout the duration of the relationship to control the agent's acts."[68]

The application of vicarious liability to AI appears attractive at first glance. The relationship between an AI and its developer mirrors in some way that of an employee and employer.[69] The developer decides that it needs to recruit an AI to fill a business need or pursue a business objective; it researches the skills and capabilities it needs from the AI; once it chooses a specific AI, it trains the AI to perform the necessary tasks; and finally, it provides oversight to the AI in performing such tasks. Of course, one major difference is that human employees have the capacity for autonomous decision-making independent of, and perhaps even in direct contradiction with, the guidance provider by their employer.[70] However, this otherwise substantial difference between employees and AI

[67] This theory of liability is referred to as *respondeat superior*; although it is functionally associated with tort law, it has historically been classified as a doctrine within agency law. *See* RESTATEMENT (THIRD) OF AGENCY § 2.04 (Am. L. Inst. 2006).

[68] *See id.* § 1.01, cmt. c.

[69] In the medical context, some have argued that medical diagnostic AI, such as IBM's "Watson," can be analogized more appropriately to medical students, to whom liability for medical malpractice extends via vicarious liability, than to consulting physicians, to whom liability for medical malpractice does not extend. *See* Jason Chung & Amanda Zink, *Hey Watson, Can I Sue You for Malpractice? Examining the Liability of Artificial Intelligence in Medicine*, 11 ASIA-PAC. J. HEALTH L., POL'Y & ETHICS 51, 70 (2018). Unlike consulting physicians, who do not interact with or perform examinations on the patient, diagnostic AI does "examine" the patient by accessing the patient's medical history and data. *See id.* Moreover, whereas the presence of a consulting physician may be entirely unknown to a patient, there is evidence that IBM has encouraged patients to view Watson as a member of their medical team. *See id.*

[70] For this reason, courts have carved out a few key exceptions to the doctrine of vicarious liability; for example, employers are not liable for an employee's actions when an employee goes "rogue" or when an

may be quickly disappearing with the rise of self-learning or autonomous AI.[71] Similar to employees, self-learning or autonomous AI may adopt unexpected and unique methods or means for performing the task as it learns. And, unlike with software or similar non-AI technologies, such self-learned behavior may be a feature of the AI rather than a bug. This analogy can also extend to the User of the AI, although the User may play a smaller role in the initial training of the AI and may provide less detailed oversight during use.

Despite these parallels, applying vicarious liability to AI is likely to encounter numerous obstacles. First, and perhaps most critically, it is unclear whether an AI could ever constitute an agent because agency requires legal personhood,[72] a status which AI has not yet achieved.[73] Second, and related, is the issue of whether an AI is capable of consenting to an agency relationship—without legal personhood, the answer is most likely no. However, it is important to note that formal consent may not be necessary; the performance of an AI after a request by its Vendor or User may be sufficient.[74] Third, it may be difficult to determine the level of control over an AI that is required, especially

employee makes a significant personal detour while performing their duties. *See* discussion *infra* Section II.C.

[71] In 2014, many years prior to the launch of ChatGPT, scholar David Vladek highlighted the issues that autonomous AI could pose to traditional agency-based jurisprudence of employer liability or *respondeat superior*. *See* David C. Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117 (2014). Vladeck indicates that liability rules need not be revisited where the "hand of human involvement in machine decision-making is so evident" because AI does not have legal personhood. *See id.* at 120–21. Rather, Vladeck states that machines that are "capable of independent initiative and of making their own plans" will force courts to move away from agency concepts as the basis for determining liability. *See id.* at 122–23. Ultimately, Vladeck suggests that granting autonomous AI some form of legal personhood may be necessary. *See id.* at 124–25. *Cf.* Alanna Mayham, *The Legal System Could Recognize AI-Led Corporations, Researchers Say*, COURTHOUSE NEWS SERV. (Oct. 26, 2023), https://www.courthousenews.com/the-legal-system-could-recognize-ai-led-corporations-researchers-say/ [https://perma.cc/2MZQ-GTYC]; Maura O'Malley, *UK Supreme Court Grapples with Whether AI Can Be a Patent Inventor as DABUS Case Is Heard*, GLOB. LEGAL POST (Mar. 7, 2023), https://www.globallegalpost.com/news/uk-supreme-court-grapples-with-whether-ai-can-be-a-patent-inventor-as-dabus-case-is-heard-1273622924 [https://perma.cc/DQ7Z-DKNL] (indicating that courts in both South Africa and Australia have recognized an AI as the inventor of a legal patent). *But cf.* Thaler v. Vidal, 43 F.4th 1207, 1212 (Fed. Cir. 2022) (rejecting the idea that an AI can be the named inventor on a patent).

[72] *See* Duffourc, *supra* note 15, at 20.

[73] *See* Vladeck, *supra* note 71.

[74] *See* RESTATEMENT (THIRD) OF AGENCY § 1.01 cmt. d.

with respect to incredibly complex, self-learning, or partially autonomous AI.[75] And finally, the question of "agent for whom?" may be difficult to answer where both the Vendor and the User have exercised some level of control over the purposes and methods of the AI.[76]

Our robotic surgical assistant hypothetical helps to highlight the intricacies of the last two issues. With respect to control, the Vendor clearly designed, developed, and trained the AI embedded within the robot's perception and control modules. It exercises general control over the AI. However, during any particular surgery, the AI responds to the actions of the surgeon to try to predict and prepare for the surgeon's next moves. Thus, the surgeon clearly has at least some indirect control over the AI. Moreover, to the extent the surgeon modifies her usual process to elicit specific behavior from the robotic assistant, she begins to exert direct control over the AI.

B. Products Liability: AI as a Product or Service

Products liability will likely be a more natural fit for AI. After all, this paper discusses AI in the context of a development or licensing agreement, under which the AI is the product being developed or licensed.[77] Traditional product liability involves three separate claims: (1) strict liability; (2) negligence; and (3) breach of implied warranty.[78] In some states, a

[75] *See* Duffourc, *supra* note 15, at 21–28.

[76] *See id.*

[77] It is important to note that a contract for the development or licensing of AI will likely include both products (the AI itself, explainability or visualization tools, etc.) and services (application programming interfaces (APIs), maintenance, training, support, etc.). *The 5 Kinds of Contracts Every AI or Robotics Company Should Have*, ROBOTICS & AUTOMATION NEWS, https://roboticsandautomationnews.com/2022/12/22/the-5-kinds-of-contracts-every-ai-or-robotics-company-should-have/58758/#:~:text=Just%20like%20other%20sales%20agreements,mechanism%20and%20timeline%20for%20payments [https://perma.cc/GA84-UHQQ] (last visited Apr. 15, 2024) ("Just like other sales agreements, these legal contracts for AI/robotics should lay out the services to be provided, the circumstances in which these services are expected to be fulfilled, any qualifications that might come into play . . .").

[78] *See* Brenda Leong & Jey Kumarasamy, *Third-party Liability and Product Liability for AI Systems*, IAPP (July 26, 2023), https://iapp.org/news/a/third-party-liability-and-product-liability-for-ai-systems/ [https://perma.cc/TG7K-XJ6K].

fourth claim is added for the failure to warn.[79] The framework used to analyze a product liability claim depends significantly on whether the technology in question is a product or service, as services are excluded from traditional products liability.[80] Thus, a defective product may be subject to any of the three (four in some states) claims above, while a defective service will default to a general negligence framework.[81]

It is unclear at this point whether AI will be treated as a product or service in the U.S. A recent case in the Third Circuit addressed this question directly with respect to a "multifactor risk estimation model" that was used to inform pretrial release decisions in New Jersey state courts.[82] After indicating that New Jersey courts often look to the Third Restatement's definition of "product" as "tangible personal property distributed commercially," the Third Circuit held that the model in question did not fall within the scope of the definition for two reasons: (1) The model was not distributed commercially but instead only made available to specific entities; and (2) An algorithm is neither tangible personal property nor sufficiently analogous to it.[83] The District Court characterized the algorithm as "information, guidance, ideas, and recommendations," or, in other words, "speech" that judges and prosecutors were free to disregard.[84]

More helpful guidance may be found by looking to the way in which courts have approached software.[85] The Third Restatement suggests that courts should consider the distinction made in the Uniform Commercial Code (U.C.C.), which distinguishes between mass-marketed software (a good) and software that was developed or customized

[79] Failure to warn may be at or after the time of sale. *See* RESTATEMENT (THIRD) OF TORTS: PRODUCT LIABILITY § 1, cmt. a.

[80] *See id.* § 19(b).

[81] *See id.* § 19(b), cmt. a.

[82] Rodgers v. Christie, 795 Fed. App'x 878, 878–79 (3d Cir. 2020). It is unclear if the model in Rodgers constitutes AI; the model consists of a set of nine factors, each of which are scored based on information from electronic court records, that are then mapped onto a "Decision-Making Framework." *See* Holland v. Rosen, 895 F.3d 272, 281 (3d Cir. 2018) (discussing the same model). While that description aligns more with a rules-based system than AI, the model does appear to have been developed using an extremely large data set, which suggests at least some aspects of AI or machine learning. *About the Public Safety Assessment*, APPR https://advancingpretrial.org/psa/about/ [https://perma.cc/LZ6Q-EWJX] (last visited Nov. 11, 2023).

[83] *See Rodgers*, 795 Fed. App'x at 879–80.

[84] *See* Rodgers v. Laura & John Arnold Found., No. 17-5556, 2019 U.S. Dist. LEXIS 97607, at *7–8 (D.N.J.).

[85] *See* Leong & Kumarasamy, *supra* note 78.

specifically for the customer (a service).[86] However, the U.C.C. has since been amended and now considers software and goods to be mutually exclusive, unless the software is embedded within a good.[87] The limited case law that exists suggests that standalone software does not fall under products liability frameworks, and is therefore not subject to strict liability.[88] Louisiana may be a notable exception.[89]

Should courts analyze AI similarly, whether AI will be subject to a strict liability or negligence framework may depend on whether the AI is used on a standalone basis or embedded within a product. Although our hypothetical robotic surgical assistant clearly involves AI embedded within a product (the robot), both approaches are considered below for completeness, especially given that plaintiffs often bring claims under products liability and negligence at the same time.[90]

1. AI as a Product

Product liability is often a combination of state blackletter and statutory law.[91] Product liability holds product manufacturers liable for harm caused by their products even in scenarios in which the manufacturer exercised care or lacked the intention to create a harmful product. For example, manufacturers may be held liable for product manufacturing defects even when exercising all possible care during manufacturing.[92] Manufacturers may also be liable for failing to adopt reasonable alternative designs that are less harmful than the design made available to the public.[93] And finally, manufacturers

[86] *See* RESTATEMENT (THIRD) OF TORTS: PRODUCT LIABILITY § 19, cmt. d.

[87] *See* James M. Beck, *New Decision Directly Addresses the "Is Software a Product" Question*, DRUG AND DEVICE L. (May 2, 2022), https://www.druganddevicelawblog.com/2022/05/new-decision-directly-addresses-the-is-software-a-product-question.html [https://perma.cc/6AN7-RFGM].

[88] *See id.*

[89] *See id.* (citing Schafer v. State Farm Fire & Casualty Co., 507 F. Supp. 2d 587, 600–01 (E.D. La. 2007) (referring to a previous case which established that software was a "corporeal property" for taxation purposes in suggesting that it may also be a product for purposes of the state products liability law).

[90] *See id.* Additionally, many states have specifically excluded use of an allegedly defective product by a medical professional from the scope of strict liability. *See* RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 20, notes on cmt. d ("Most jurisdictions hold that hospitals and doctors provide a service . . . and immunize them from strict liability for harm from defective products used in medical treatment, whether the product is implanted in the patient, loaned to the patient, or merely used as a tool.").

[91] *See* AMERICAN LAW OF PRODUCTS LIABILITY, 3d § 1.2 (2023).

[92] *See* RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 2(a).

[93] *See id.* § 2(b).

may be liable for failing to provide adequate instructions or warnings to users.[94] Which type of defect is alleged has a significant impact on the framework used to analyze fault.[95] These potential sources of product defects giving rise to liability have significant parallels with respect to AI embedded in a product.[96] When developing an AI, a Vendor may exercise all possible care and yet the AI may fail to handle (be "defective" in) particular cases, such as an outlier scenario absent from the training set or improperly formatted inputs. A Vendor may also fail to sufficiently consider alternative designs (with respect to the AI itself or with respect to the interaction between the AI and the product) that produce less harmful outcomes for individuals, perhaps due to a lack of access to the data required to perform such analysis during the training process. And finally, even where a Vendor provides the User with instructions and/or warnings, such instructions and/or warnings may be inadequate because they do not appropriately reflect the complexity of the AI or because they are not reasonably understandable by the User, whose knowledge of the AI is extremely limited. Given the similarities to product liability concepts, it is possible that AI will pose significantly fewer challenges to product liability doctrine than it will to vicarious liability.[97]

Nevertheless, a successful products liability claim related to an AI will encounter three significant challenges. First, establishing harm almost always requires injury to persons or

[94] *See id.* § 2(c).

[95] *See* Cassandra Burke Robertson, *Litigating Partial Autonomy*, CASE W. RSRV. UNIV. FAC. PUBL'N (2023), https://scholarlycommons.law.case.edu/cgi/viewcontent.cgi?article=3188&context=faculty_publications [https://perma.cc/BVT6-4BCU]. It is unclear whether AI defects could fit within the first category of defect, which is often referred to as "manufacturing defects." *Cf. id.* (indicating that some software defects, such as flaws introduced during testing or typos in the code, may be considered manufacturing defects).

[96] Karni Chagal-Feferkorn has argued that, unlike other products, AI is unique in that it is "expected to cause damage regardless of any defects." Chagal-Feferkorn, *supra* note 15, at 84 (providing an example of a medical diagnosis and treatment algorithm which solves for the optimal course of action by balancing the damage resulting from intervening unnecessarily against the damage resulting by failing to intervene when necessary, given that the algorithm cannot perfectly identify when intervention is necessary for every case). This is because "whenever [] algorithms reach decisions based on probabilities . . . inevitable damage will occur when the general rule is applied in cases that in hindsight turned out to be the exceptions." *Id.* at 85.

[97] *See* Vladeck, *supra* note 71, at 123.

property,[98] thus limiting the types of AI that may be covered.[99] In our hypothetical, this means that the patient that has suffered permanent physical damage maintains the strongest claim against the Vendor. But there is support for an argument that the likely economic damage to the surgeon's practice resulting from the patient's claim (or, more immediately, publicity of the patient's injury) also provides sufficient injury for a claim by the surgeon against the Vendor.[100]

Second, establishing a defect may be extremely difficult given the complexity of AI technologies and the interactions between the AI and the other technologies embedded in the product (e.g., in our hypothetical case, the sensors used by the robot in the perception module).[101] Some states may allow plaintiffs to argue that a defect may be inferred by the fact that the product did not function as intended, but other states have specifically rejected this approach.[102] Design defects may be more appropriate when dealing with AI, but will encounter challenges related to access (the AI is likely to be subject to trade secret protection), cost (may require access to AI experts), and feasibility (given how rapidly the technology is advancing).[103] Failure to provide adequate instructions or warnings may be the easiest defect for plaintiffs to show, but in particularly complex products, the impracticality of an effective warning may result in the analysis reverting back to design defect instead.[104]

Third, even though products liability is a form of strict liability, this does not excuse plaintiffs from establishing causation.[105] Unfortunately for plaintiffs, AI presents fundamental challenges with respect to two particular elements of the causation requirement: (1) proximate cause; and (2) the nexus requirement, or the requirement that

[98] *See* RESTATEMENT (THIRD) OF TORTS: PROD. LIAB. § 1, cmt. d. (AM. L. INST. 1998).

[99] For example, a discriminatory AI used in employment is unlikely to cause the type of harm required (additionally, it may be characterized as a service rather than a product).

[100] *See* RESTATEMENT (THIRD) OF TORTS: PROD. LIAB. § 21, illus. 1. (AM. L. INST. 1998).

[101] *See* Robertson, *supra* note 95, at 34.

[102] *Id.* at 34–35 (referring to such arguments as the "malfunction doctrine"); RESTATEMENT (THIRD) OF TORTS: PROD. LIAB. § 21, illus. 1. (AM. L. INST. 1998).

[103] *See* Robertson, *supra* note 95, at 39. Given the difficulty in establishing a design defect, it is critical for parties to understand which party has the burden of proof; not every state places the burden of proof on the plaintiff.

[104] *Id.* at 40.

[105] *See* RESTATEMENT (THIRD) OF TORTS: PROD. LIAB. § 15 cmt. a (AM. L. INST. 1998).

the plaintiff establish a link between the defect and the alleged harm.[106] Proximate cause focuses on whether the harm resulting from the defect in question was foreseeable by a reasonable person.[107] However, with respect to AI, and especially unpredictable AI with little or no transparency or explainability, foreseeability may be futile.[108] With respect to the nexus requirement, the opacity of AI will present a similar challenge—it may be incredibly difficult to prove that the existence of the defect led to harm when the internal structure or decision-making process of the AI is unknown.[109]

Notwithstanding the challenges outlined above, there are likely some scenarios under which an AI may be subject to strict liability as a product. Such scenarios are likely to result in significantly more exposure for the Vendor than the User; however, a deeper discussion on how liability is allocated in these scenarios is provided below.

2. AI as a Service

Analyzing AI as a service presents additional challenges. Services are typically evaluated under a negligence theory,[110] and thus far, little scholarship has focused on the applicability of negligence to AI.[111] This may be driven in part by the fact that most discussion of AI liability focuses on the AI's creator or developer (the Vendor, in the context of this paper), which aligns more closely to product liability's focus on the

[106] *See* Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J.L & TECH. 890, 922 (2018); RESTATEMENT (THIRD) OF TORTS: PROD. LIAB. §§ 15–16 (AM. L. INST. 1998).

[107] Bathaee, *supra* note 106, at 923.

[108] *See id.* at 923–25.

[109] *See id.* at 925–28 (arguing further that the effect of this challenge may even prevent the case from getting past the first stage, at least in Federal courts, because of the Article III standing requirement that the injury be "fairly traceable" to the defendant's unlawful conduct); Vladeck, *supra* note 71, at 128. In cases involving software, the software vendor almost always claims that errors are the result of the user's actions or the user's hardware, reflecting the concept "garbage in, garbage out." L.J. KUTTEN & FREDERIC M. WILF, COMPUTER SOFTWARE: PROTECTION/LIABILITY/LAW/FORMS, § 12:55 (2022 Update). Similar arguments are likely to be made by AI vendors, especially where the AI is embedded within a larger or more complex business process or where little to no controls are placed on the AI User's use of the AI. Additionally, there may be scenarios in which the AI behaves in unexpected ways that reflect that AI learning from and improving upon the "instructions" provided to it by the developer.
Vladeck, *supra* note 71, at 144–46 (arguing that these issues support a strict liability regime for AI).

[110] *Mitigating Product Liability for Artificial Intelligence*, Jones Day (Mar. 2018), https://www.jonesday.com/en/insights/2018/03/mitigating-product-liability-for-artificial-intell.

[111] Andrew D. Selbst, *Negligence and AI's Human Users*, 100 B.U. L. REV. 1315, 1327 (2020).

product manufacturer.[112] Negligence, on the other hand, often focuses on the AI's user, which may be its creator or another business to whom the AI is made available (the User, in the context of this paper).[113] Thus, treating AI as a service, rather than as a product, will significantly increase the User's exposure to liability resulting from AI-caused harm.

To establish negligence, a plaintiff must prove that the defendant breached the standard of care, and that such breach caused the plaintiff harm.[114] Inherent in the concept of negligence is the idea that there exists some standard of care, generally defined as reasonable care, that if adhered to, would have prevented the harm.[115] Importantly, the requirement to exercise reasonable care applies regardless of whether a person makes use of a technology such as AI when performing an activity.[116]

The use of AI presents four major challenges with respect to the breach element of a negligence claim. First, the typical user of an AI will not have enough information, knowledge or understanding about the AI to know when it is likely to err, and thus will be unable to know what care to exercise over the AI's use.[117] In other words, the AI may err regardless of the level of care exercised by the user—making it extremely difficult to define the level of reasonable care that, if adhered to by the user, would have prevented the harm. This issue is likely to be most acute in narrow, predictive AI that aims to

[112] *Id.* at 1330.

[113] *Id.* Nevertheless, the Vendor may still be exposed to a negligence claim; specifically, that the Vendor was negligent when developing and distributing the AI. *Cf.* Jan De Bruyne et al., ARTIFICIAL INTELLIGENCE AND THE LAW, 370–71 (Jan De Bruyne & Cedric Vanleenhove, eds. 2021).

[114] Negligence is frequently described as comprising four elements: (1) Duty; (2) Breach; (3) Causation; and (4) Damages. Barry A. Lindahl, 1 MODERN TORT LAW: LIABILITY AND LITIGATION § 3:2 (2d ed.). For the purposes of this section, the focus is solely on breach, the second element. The breach element presents unique issues with respect to AI that were not already addressed in the prior section. Challenges in establishing causation related to AI were discussed in the previous section. *See supra* II.B.1. While the damages element may present additional issues in that the damages recoverable in tort may vary dramatically from those recoverable as between the User and Vendor under contract, we exclude a discussion of damages here because such issues are broader than the specific context of an AI licensing or use agreement. *See generally* Michael Dorff, *Attaching Tort Claims to Contract Actions: An Economic Analysis of Contort*, 28 SETON HALL L. REV. 390 (1997).

[115] Selbst, *supra* note 111, at 1331.

[116] *Id.*

[117] *See id.* at 1331–33 (arguing that the unforeseeable nature of AI errors is distinct from but related to the issues of foreseeability that have traditionally underpinned tort law).

identify patterns that even the most trained humans are incapable of recognizing.[118] Turning back to our robotic surgical assistant example, even an informed surgeon likely does not understand the potential ways that the robot may misinterpret the surgeon's actions and thus fail to appropriately reposition its appendages. If the surgeon can't connect her actions to those of the robot, then she will be unable to know what care she can exercise to reduce the risk of an accident. Accordingly, the basis for the standard of care may default back to whether the surgeon exercised reasonable care when making the decision to use the AI, in what way, and in which situations.[119] Any guidance or standards from applicable professional bodies may be highly relevant to this analysis.[120] If the negligence claim is brought against the Vendor, this analysis may be even more difficult because the human-centric negligence standard may prove to be irreconcilable with the way that AI operates.[121]

Second, in situations involving human-AI interaction, the standard for care may become impossibly high due to an expectation that the human should remain continuously alert in

[118] *See id.* at 1333–38 (distinguishing between AI that is amenable to human oversight (e.g., machine vision, which aims to accurately identify and categorize visible physical objects and may even incorporate humans into its training process) and AI that is not amenable to human oversight because it arguably surpasses human capabilities (e.g., AI that analyzes large quantities of network traffic data, identifies common patterns, and then detects and flags deviations from those patterns)). These problems are further exacerbated when AI decisions for which there is no equivalent human decision or ground truth (e.g., personalized AI recommendation systems). *Id.* at 1338.

[119] *Id.* at 1339.

[120] Jan De Bruyne et al., *supra* note 113, at 373; *See generally* AMERICAN MEDICAL ASSOCIATION, *American Medical Association Principles for Augmented Intelligence Development, Deployment, and Use* (Nov. 14, 2023), https://www.ama-assn.org/system/files/ama-ai-principles.pdf (physicians); *Resolution,* AMERICAN BAR ASSOCIATION, (December 13, 2019), https://www.americanbar.org/content/dam/aba/directories/policy/annual-2019/112-annual-2019.pdf (lawyers).

[121] Jan De Bruyne et al., *supra* note 113, at 374; *see* discussion of AI and Intelligence, *supra* I.A. Scholars have suggested a few possible solutions to this problem: (1) the adoption of a "reasonable algorithm" standard in lieu of the reasonable person standard; (2) applying "'soft law' rules and standards related to the proper design, training, monitoring, updating and decommissioning of AI systems" to the reasonable person standard; and (3) reversing the burden of proof in certain situations (e.g., cases of significant informational asymmetry or where the defendant has violated its legal or contractual obligations to ensure safety or transparency). *See id.* at 374–76. For more on burden shifting, *see infra* II.B.3.

case the AI fails.[122] Systems that involve human-AI interaction are often designed for the specific purpose of reducing the effort required by the human, allowing the human to direct her efforts elsewhere (e.g., autonomous vehicles). That purpose is in tension with the likely expectation that the human will be available to "step in" in case something goes wrong. To the extent that the reasonable person standard requires a human using AI to be prepared to intervene at any moment, it will require a level of attention and awareness that is neither physically possible nor conducive to achieving the benefits of the AI.[123] Our robotic surgeon example illustrates this well. The surgeon cannot both focus on properly executing the surgery while also keeping an eye on any potential issues with the robot.[124] Of course, she could hire a surgical assistant whose role is to monitor the robot and intervene as needed, but doing so would likely nullify the purpose for using the robot.

Third, it is unclear whether the appropriate standard of care should include a duty to protect the AI from manipulation or interference. A new area of research called "adversarial machine learning" focuses on ways to manipulate an AI's inputs to influence its decisions.[125] There may be some scenarios where it is both conceptually and practically reasonable to expect that users will undertake precautions against the intentional manipulation of AI (e.g., ensuring data integrity and validity within their systems as part of their broader cybersecurity efforts). But in other scenarios, especially ones in which the user has limited control over the AI's training and operations, such a standard would be highly impractical.

[122] *See* Selbst, *supra* note 111, at 1348. Note that this is different than the first issue in that here, it is not whether the human can foresee the potential error by the AI, but whether the human is paying attention when the error occurs and thus able to react appropriately to prevent the harm. *Id.* at 1348 n.155.

[123] *See id.* at 1348–49. (suggesting that applying such a high standard would essentially move scenarios involving human-AI interactions into the pockets of negligence law that function as if under a strict liability regime (e.g., imposing a reasonable person standard on children or the inept).

[124] One tragic real-life parallel to this dilemma is the story of the Boeing 737 Max. Boeing designed an automated flight-control system, in part "to avoid costly simulator retraining" of pilots, but expected that pilots would both be able to and know how to diagnose and react to issues with the system within four seconds. Robert Zafft, *Faulty to the Max: Boeing and the FAA's 737 Debacle*, FORBES (Jan. 4, 2021), https://www.forbes.com/sites/robertzafft/2021/01/04/faulty-to-the-max-boeing-and-the-faas-737-debacle/?sh=4b3f5f802134 [https://perma.cc/8DA3-DCF8]. Further complicating the issue, Boeing did not inform pilots about the system's mechanics or its reliance on a single sensor which could be faulty. *Id.* Boeing essentially put pilots "at the center of operational safety," but withheld from them the key information they needed in order to identify and manage issues. *Id.* In fixing the problem, Boeing appears to have focused primarily on reducing the risk of errors in the system itself. *See id.*

[125] Selbst, *supra* note 111, at 1351.

Finally, the use of AI may render an otherwise successful negligence claim related to an unanticipated and unintentional discriminatory practice more difficult to establish, even in a scenario where the AI is known to be more accurate for certain protected classes.[126] This results when the AI improves outcomes for all individuals but at disproportionate levels by group, meaning that, although everyone will ostensibly benefit from the use of the AI, some groups will experience higher error rates than others.[127]
The four challenges above point in different directions with respect to the potential liability of an AI user. The second and third challenges indicate that the use of AI increases the risk that a user will be held liable under negligence for actions that were beyond their control from a practical standpoint. The first challenge, on the other hand, indicates that the user may be able to reduce the risk of liability caused by the use of AI by ensuring that the initial decision to use the AI itself was not negligent.[128] The fourth challenge surprisingly, and worryingly, indicates that the user might be less likely to be held liable for an unintentionally discriminatory practice if that practice has been replicated (or even exacerbated) indirectly through an AI.

Given the significant differences in legal treatment depending on whether an AI is viewed as a product or as a service and potential challenges regarding causation, addressing this topic during contract negotiations will be critical to reduce potential liability. The determination as to whether a contract for AI technology constitutes a product or a service is likely to relate directly to which party assumes liability in the case where the technology fails or causes harm,[129] therefore, this issue is explored in greater detail below.

[126] *Id.* at 1357–58.

[127] *Id.* at 1358. Selbst's intuition here leads to the frightening possibility that discrimination may be "washed" through the use of AI, at least with respect to negligence claims. However, the result might be entirely different with respect to statutory laws prohibiting discrimination. *Cf. id.* at 1355. Selbst ultimately places this issue under the duty element of negligence rather than breach because duty is typically "the place where public policy considerations most explicitly enter the picture of negligence law." *Id.* at 1359 (emphasizing that there is currently no duty to ensure "distributional fairness in individual case outcomes").

[128] Of course, establishing that the choice to use an AI is not negligent may encounter similar complications given the unique issues posed by AI. *See supra* I.B. And, this argument will be less successful with each incident. An AI user cannot believably claim that they do not have the ability to foresee the risk of harm resulting from complex AI due to their lack of knowledge if the AI, or similar AI, has already caused harm on multiple occasions.

[129] *See* KUTTEN & WILF, *supra* note 109, § 12.45.

3. Learning from Europe

The preceding sections outline numerous challenges posed by AI to current product liability and negligence frameworks, especially issues related to causation and breach. For courts and legislators grappling with how to adapt current law to address such issues, the current efforts by the European Union may be helpful as a reference. As part of its focus on liability-related issues posed by AI,[130] the EU has proposed reforms to the current EU liability framework along two dimensions: (1) adapting the current Product Liability Directive (PLD), which applies strict liability to manufacturers of defective products, to "adapt it to the digital age" while preserving its "technology-neutral nature and coverage"; and (2) a proposed AI Liability Directive (AILD), to harmonize non-contractual fault-based liability rules for harm caused by AI.[131]

Two proposed changes to the PLD are most critical to this discussion. First, the proposal explicitly incorporates AI into a strict liability framework under the umbrella of software.[132] Second, the proposal creates a rebuttable presumption of a causal link between a product's defectiveness and damage under two scenarios: (1) when a product defect is established and the damage caused is of a kind typically consistent with such defect;[133] or (2) when a court finds that the plaintiff faces excessive difficulty in proving such link due to the technical or scientific complexity of the product.[134]

[130] *See On Artificial Intelligence - A European Approach to Excellence and Trust, supra* note 52.

[131] *See* EUROPEAN PARLIAMENTARY RESEARCH SERVICE (EPRS), *Briefing: Artificial Intelligence Liability Directive*, at 3, 5 (Feb. 2023). The European Commission describes both reforms as comprising part of a package consisting of three complementary work streams, where the third work stream is the AI Act, which aims to set comprehensive rules to reduce risks to safety and fundamental rights posed by AI across all sectors. *Commission Proposal for a Directive of the European Parliament and of the Council on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence (AI Liability Directive)*, at 2, COM (2022) 496 final (Sept. 28, 2022) (citing *Commission Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*, COM (2021) 206 final (Apr. 21, 2021).

[132] *Commission Proposal for a Directive of the European Parliament and of the Council on Liability for Defective Products (PLD)*, at 25, COM (2022) 495 final (Sept. 28, 2022).

[133] *Id.*at 28.

[134] Additionally, the plaintiff must show based on sufficient evidence: (1) that the product contributed to the damage; and, (2) either, (i) that it is likely the product was defective, or (ii) that the product's defectiveness is a likely cause of the damage. *Id.* at 28–29.

The proposed AILD adopts a number of substantial changes aimed at promoting the rollout of trustworthy AI and reducing the legal uncertainty for businesses that may result from a fragmented, national approach.[135] Importantly, for the purposes of this paper, it creates a rebuttable presumption of causality in situations where an AI developer or user's failure to comply with a specific legal obligation is reasonably likely to have influenced the AI action that gave rise to damage.[136] But this presumption only applies to the causation requirement and not to the nexus requirement.[137]

Given the significant challenges in establishing causation related to complex AI, [138] the presumptions of causality proposed in both the PLD and AILD are likely to have a significant impact on a plaintiff's likelihood of success. Moreover, these changes appear to be reasonable and practical solutions to the problems implicated by AI. It is sensible to expect that courts and legislators in the U.S. will look to these approaches when adapting current common and/or statutory law, especially given that the use of targeted burden shifting through rebuttable presumptions already exists within U.S. tort frameworks.[139]

C. Allocation of Liability

Courts adjudicating cases of AI-caused harm will need to determine whether AI failures should be attributed to the User, the Vendor, or both, to assign liability. For the purposes of this paper, it is assumed that the AI cannot itself be held liable because it is not a legal person.[140] This section explores the allocation of liability under each of the legal theories analyzed above. Of course, the contract between the User and Vendor may override the

[135] *AI Liability Directive*, *supra* note 131, at 5.

[136] *See AI Liability Directive*, *supra* note 131, at 24. For AI systems that are not high-risk under the AI Act, the court must also find that it is "excessively difficult" for the plaintiff to prove the causal link for the rebuttable presumption to trigger. *Id.* at 24. The proposed directive references requirements in the AI Act as examples of the legal obligations with which the defendant may fail to comply; such examples differ between the AI provider and the AI user. *Id.* at 24.

[137] *Id.* at 13.

[138] *See id.* at 16; Bathaee, *supra* note 106, at 922.

[139] *See* RESTATEMENT (THIRD) OF TORTS: PROD. LIAB. § 4, rep. note (AM. L. INST. 1998).

[140] This should be relatively noncontroversial. However, many scholars have challenged this assumption, arguing that the law should grant legal personhood to certain forms of AI. *See* Vladeck, supra note 71, at 124–25; Duffourc, supra note 72; Victor Shollaert, *AI and Legal Personality in Private Law: An Option Worth Considering (?)*, 31 EUROPEAN REV. PRIVATE L. 387 (2023); Beckers & Teubner, supra note 14, at 93–94. Others strongly disagree. *See* Shollaert, supra, at 389 (discussing AI's legal status in other contexts).

rules outlined in this section. Thus, the purpose of this analysis is to outline the backdrop of default rules within which the contract negotiation process occurs.

Under vicarious liability (agency theory), liability related to harm caused by the AI would attach to the AI's principal.[141] For human agents, the principal would generally be their regular employer (the "general employer"); however, one common modification to this rule applies when a human agent is "lent" or "contracted" out to another employer (the "special employer").[142] In that scenario, liability will depend on which employer (general, special or both) had the right to control the agent's conduct, both prior to and during the specific acts giving rise to liability, with a strong presumption that the general employer always exercises some level of control.[143] The content of the contract between the parties will likely be relevant in determining the extent of control by each party.[144] Our robot surgical assistant hypothetical highlights issues that arise with respect to semi-autonomous AI—while the developer of the AI exercises control over the AI's design and general parameters for interacting with the surgeon, it is the surgeon who exercises control (even if through potentially unknown mechanisms) during each particular use.

Under a strict product liability framework, liability is more likely to attach to the Vendor than the User given the emphasis in assigning liability to the "manufacturer," especially where the User is not selling or distributing the AI commercially but rather using it for its own purposes.[145] Rooted in the principle of deterrence, this approach reflects the fact that the manufacturer exerts the greatest control over the risk posed by its products through decisions around product design and manufacturing and is likely best positioned to insure against such risks.[146] But three alternative scenarios exist: (1) a user may be partially or fully liable if the defendant can show that the user was negligent in its use of the product

[141] 2A C.J.S. *Agency* § 451, WESTLAW (updated Nov. 2023).

[142] RESTATEMENT (THIRD) OF AGENCY § 7.03, cmt. d(2) (AM. L. INST. 2006).

[143] *Id.* Control prior to the employee's acts include decisions around selection criteria, training, and the equipment used; control during the employee's acts may depend on the extent to which the work requires significant coordination with others and the level of supervision by the special employer. *See id.* Another consideration may be which party is best positioned to purchase insurance covering use of the AI. *See id.*

[144] *See id.* at cmt. D, note.

[145] RESTATEMENT (THIRD) OF TORTS: PROD. LIAB. § 1, cmt. e (AM. L. INST. 1997).

[146] *See* Chagal-Feferkorn, *supra* note 15, at 78.

(contributory negligence);[147] (2) liability may be expressly assumed by the user of a product under a contract between the manufacturer and the user;[148] and (3) liability may be impliedly assumed by the user of a product because it knowingly and voluntarily used an inherently dangerous product.[149] The first scenario, however, is likely to present significant challenges given uncertainty around the negligence standard applied to human use of AI and the difficulty of establishing causation.[150]

Under a negligence framework, on the other hand, liability is more likely to attach to the User; unlike the Vendor, the User has a direct relationship to the injured party. The User may be able to assign all or some of the liability to the Vendor in states that recognize a contributory negligence or comparative fault framework. But the User will be at a distinct disadvantage in that the User's knowledge of and access to information about the AI's design, training and operation will be significantly diminished compared to the Vendor. Fortunately for the User, contractual allocations of liability may displace the default contribution and indemnity rules that would otherwise apply.

[147] RESTATEMENT (THIRD) OF TORTS: APPORTIONMENT OF LIAB. § 4 (AM. L. INST. 1997). However, the failure of a user to discover and guard against defects in the AI would be unlikely to support a contributory negligence defense. *See* Amy L. Stein, *Assuming the Risks of Artificial Intelligence*, 102 B.U. L. REV. 979, 996 n.85 (2022).

[148] RESTATEMENT (THIRD) OF TORTS: APPORTIONMENT OF LIAB. § 2 (AM. L. INST. 1997).

[149] *See* Stein, *supra* note 147, at 91 (referring to such theory as "[p]rimary implied assumption of risk").

[150] *See* discussion *supra* Section II.B. Some factors that may be relevant when assessing the AI User's contributory negligence are: (1) The AI technology's level of transparency (users of black-box AIs may benefit from a lower reasonableness standard given the inability to fully understand how the AI works); (2) The level of due diligence performed by the AI User (users who adopt AI systems without understanding them are more likely to be deemed negligent); and, (3) The level of monitoring performed by the AI user, especially for experimental or relatively novel AI technologies. *See AI, Machine Learning & Big Data Laws and Regulations 2022 | Japan*, GLOBAL LEGAL INSIGHTS, § 6.3.1 (2022), https://www.globallegalinsights.com/practice-areas/ai-machine-learning-and-big-data-laws-and-regulations/japan [https://perma.cc/NT6Z-MBM5]. Additionally, it is unclear to what extent a Vendor of AI-enabled products used by a human operator may benefit from a comparative fault framework regardless of its relative level of culpability. Ultimately, it may depend on the fact-finder—there is some evidence that jurors may be more likely to assign fault to the human user, while judges may be more likely to assign fault to the manufacturer. *See* Robertson, *supra* note 95, at 42–43 (discussing fault in the context of semi-autonomous vehicles). However, it is unclear whether the same pattern would extend to the use of AI in contexts with which jurors have little to no personal experience (e.g., surgery).

Analyzing a few of the currently existing or proposed laws specifically governing AI within the U.S. provides additional guidance on the allocation of liability.[151] New York City's Local Law 144 governs the use of automated employment decision tools, which require such tools are only implemented after a bias audit and alongside required notices.[152] The law applies such requirements to the employer (or employment agency) who uses such tools rather than the vendor or developer of the tool itself, although it appears to recognize that the bias audit would likely need to be performed by the vendor or an independent auditor.[153] Similarly, but in an entirely different context, Colorado Regulation 702-10 governs life insurers' use of algorithms and predictive models that use certain types of external consumer data.[154] The regulation requires insurers to establish a risk-based governance and risk management framework designed to determine whether use of such algorithms may result in unfair discrimination (and, if so, remediate such unfair discrimination).[155] However, the focus on life insurer users over AI vendors by Colorado may simply be due to the fact that the Division of Insurance, which issued the regulations, does not have authority over AI vendors. In yet another completely different arena, recently enacted Georgia House Bill 203 also takes a similar approach by focusing on the user.[156] The bill prohibits optometrists (or ophthalmologists) from conducting an

[151] A helpful overview of the current and proposed laws related to AI can be found on the Electronic Privacy Information Center's website. *The State of State AI Laws: 2023,* ELEC. PRIV. INF. CTR., https://epic.org/the-state-of-state-ai-laws-2023/ [https://perma.cc/SL4Y-H8YB] (last accessed Dec. 12, 2023). For the purposes of this section, I excluded from consideration any regulation of AI contained within a state data privacy law.

[152] N.Y.C., NY, CODE § 5–300 *et seq.* (2023) [hereinafter *Local Law 144*]. The law became effective on Jan. 1, 2023, and enforcement begins on July 5, 2023. *Automated Employment Decision Tools: Frequently Asked Questions*, N.Y.C. DEP'T OF CONSUMER & WORKER PROT., (June 9, 2023) [hereinafter *AEDT FAQs*].

[153] *See* LOCAL LAW 144, *supra* note 152, at § 5–301(a), (b) (example) ("The employer asks the vendor for a bias audit."), (c) (example) ("The employer provides historical data . . . to an independent auditor to conduct a bias audit . . . ."); AEDT FAQS, *supra* note 152, at 5 (stating "Employers and employment agencies are ultimately responsible for ensuring a bias audit was done before using an AEDT.").

[154] COLO. CODE REGS. § 702–10(2) (2023).

[155] *Id.* at § 702–10(5)(A). Additionally, the law establishes reporting requirements for insurers. *Id.* at § 702–10(6). The Colorado Division of Insurance has also proposed a separate regulation that would outline the specific testing for unfair discrimination that is required. *See SB21-169 - Protecting Consumers from Unfair Discrimination in Insurance Practices*, COLO. DEPT. REGUL. AGENCIES https://doi.colorado.gov/for-consumers/sb21-169-protecting-consumers-from-unfair-discrimination-in-insurance-practices [https://perma.cc/6K6B-2C63] (last visited April 7, 2023).

[156] H.B. 203, 157th Gen. Assemb., Reg. Sess. (Ga. 2023).

eye assessment or prescribing contact lenses or glasses based on an AI-enabled eye exam unless certain requirements are met.[157] While it also establishes baseline requirements for providers of AI-enabled eye exams, such requirements mainly reference pre-existing federal law.[158] Based on the current trend, it appears that AI-related laws are more likely to target users rather than vendors, potentially increasing the relative level of exposure that users may have under *negligence per se*.

Contractual assumption of liability is likely to be a highly desirable approach for AI licensing and services agreements, given the uncertainties discussed above around the allocation of liability under current tort law frameworks. This approach will shift adjudication from the realm of tort law into the realm of contract law,[159] an area with which commercial entities are likely to be more familiar. Both the AI Vendor and the AI User will benefit from the certainty that contractual apportionment of liability provides relative to tort law. Naturally, the AI Vendor will expect the AI User to assume full liability, and vice versa. The differing results from above, where the Vendor has greater exposure under strict product liability while the User has greater exposure under negligence, suggests that both parties may benefit from some level of compromise. The parameters for potential compromise are discussed in further detail in Section III below.

III. MITIGATING AI LIABILITY THROUGH CONTRACTING

A contract is a set of legally enforceable promises[160] that typically result from negotiations[161] between two or more parties and reflect the parties' agreement, among

[157] One requirement is to maintain liability insurance through the owner or lessee of the AI-enabled eye exam mechanism. *See, e.g.*, H.B. 203, 157th Gen. Assemb., Reg. Sess. § 1(d)(12) (Ga. 2023).

[158] *Id.* § 1(c).

[159] *See* KUTTEN & WILF, *supra* note 111, § 12:44. The AI User should be mindful, however, that in the absence of a contract with potential individual plaintiffs harmed by its use of AI, it may be subject to tort law as a defendant but contract law as a third-party plaintiff. Accordingly, AI Users may want to consider taking steps to increase the likelihood that any disputes with individuals will be governed under the terms of a contract between the AI User and such individuals. Further, Users may want to incorporate this issue into discussions with the AI Vendor, specifically as to the parties' relative rights and responsibilities vis-à-vis claims by harmed individuals (e.g., indemnification).

[160] RICHARD A. LORD, 1 WILLISTON ON CONTRACTS § 1.1 (4th ed. 1993).

[161] *But see* OPENAI, *Terms & Policies*, https://openai.com/policies [https://perma.cc/5WFT-MVMG] (last visited Jan. 27, 2024) (demonstrating that not all arrangements involving AI will involve negotiation of a contract. OpenAI's "plug and play" ChatGPT offering to consumers, for example, is subject to online terms and conditions that are not intended to be negotiated).

other things, to a certain allocation of risks and costs. Once finalized and executed, each party typically takes steps to fulfil its promises, and expects that the other party will also fulfill its own promises,[162] and contract law incentivizes the parties to do so by imposing remedies for a party's breach.[163] Contracts therefore play a critical role in reducing uncertainty that may arise out of commercial relationships. Theoretically, it also follows that, the more terms the parties can agree to, the less uncertainty exists with respect to the relationship. However, the time and cost required to negotiate and draft terms act as an outer limit on the parties' efforts to contract around every eventuality. Parties must prioritize.

One key difference that drives contracting in traditional technology procurement scenarios, and that likely will similarly drive AI contracting, is the vendor's delivery model. Delivery models commonly take one of two forms: (1) the AI Vendor provides a trained AI model to the AI User for the latter's use, but the AI Vendor remains responsible for improvement, maintenance, and ongoing functioning of the AI model; and, (2) the AI Vendor provides a development or training program that the AI User uses in order to build their own AI model for subsequent use.[164]

This section aims to help parties prioritize when negotiating AI contracts by focusing on the key risks posed by AI and the ways in which such risks exacerbate the potential liability of the contracting parties, as outlined in the prior sections. For each key risk, we will discuss the interests of each party with respect to how the contract might address that risk and identify possible contract approaches that reflect these interests, including, where appropriate, possible compromise language balancing both parties. To prioritize the scenarios that we believe will be most relevant to potential readers, we limit our analysis to fully developed AI that is nevertheless too complex to be a simple "plug-and-play" tool.

[162] RICHARD A. LORD, 1 WILLISTON ON CONTRACTS § 1.1 (4th ed. 1993).

[163] *Id.*

[164] *See* JAPAN MINISTRY OF ECONOMY, TRADE AND INDUSTRY, *Contract Guidelines on Utilization of AI and Data: AI Section* (hereinafter METI CONTRACT GUIDELINES), § V.1 (June 2018). These can also be paired with other services, including consulting, technical support, etc. Contracting considerations around these ancillary services are outside the scope of this paper. One key consequence of these two different approaches may be which company has a better claim to ownership over either the raw data or the resulting model in the case where such has not been expressly addressed in the contract, *see id.* § V.3(1); the issue of Intellectual Property is likely to be particularly acute in licensing contracts for Gen AI, *see* PRACTICAL LAW, *supra* note 54, at 2-3. However, intellectual property issues implicated by AI are beyond the scope of this article.

As a result, some issues implicated in other types of arrangements involving AI, for example in more nascent partnerships to develop AI,[165] are outside the scope of this section.

It is important to note that this section is not intended to provide a comprehensive list of relevant contract sections and considerations when licensing AI technology. Rather, we intend merely to expand on the insights from the prior section by providing a few concrete examples of the ways in which contracting may reduce uncertainty around liability resulting from the use of AI, and the techniques that both parties may use during negotiations.

A. Preparing to Contract

Addressing liability issues begins well before the first clauses are written, with rigorous and targeted AI-specific due diligence.[166] First, as with any business relationship, lawyers need to understand who the other party is. If advising the User, the lawyer needs to understand the Vendor's level of maturity as a business generally and with respect to the specific AI product or service being licensed.[167] Less mature counterparties increase risk, and the failure to consider a party's and/or AI technology's maturity prior to using the AI may be a relevant factor in a negligence claim against the User. If advising the Vendor, it is important to understand the User's level of sophistication regarding AI technologies, expected use cases, and general risk tolerance, as these factors may increase

[165] The Contracting Guidelines from the Japanese Ministry of Economy, Trade and Industry contain model contracts for three different phases in an AI-based partnership: (1) Assessment phase (non-disclosure agreement); (2) Proof-of-concept phase (operations test agreement); and (3) Development phase (software development agreement). METI CONTRACT GUIDELINES, *supra* note 164, § VII.3.

[166] *See* Lisa R. Lifshitz, *Avoiding AI Agreement Dystopia: Managing Key Risks in AI Licensing Deals*, AM. BAR ASS'N BUS. L. TODAY (Sept. 4, 2023), https://www.americanbar.org/groups/business_law/resources/business-law-today/2023-september/avoiding-ai-agreement-dystopia-managing-key-risks-in-ai-licensing-deals/ [https://perma.cc/UJ7R-KCQ4]; Christianson et al., *Contracting for AI Technologies – Top Five Best Practices*, JDSUPRA (Nov. 16, 2023), https://www.jdsupra.com/legalnews/contracting-for-ai-technologies-top-3387165/ [https://perma.cc/V3KX-66HG].

[167] *See* Lifshitz, *supra* note 168. Note that given the recency in AI advances, the level of maturity generally and specific to the AI may be dramatically different (in both directions). Parties should be careful not to confuse experience with contracting generally and knowledge or sophistication around the potential issues inherent in the AI product or service; otherwise, the parties may improperly underestimate the unique liability risks posed by the AI.

the likelihood that use of AI causes harm to a third party. In higher risk scenarios, Vendors may want to consider additional controls in the contract to reduce its exposure, such as technical or contractual restrictions preventing or prohibiting certain types of uses.[168]

As part of this due diligence, one critical piece of background information that should inform both Vendor and User is the maturity level of the other party's AI governance efforts.[169] This information is particularly important for two reasons. First, a party that lacks appropriate (or any) governance processes may present a higher risk as a contractual counterparty. Second, it will be easier to obtain contractual commitments from the other party that obligate it to have and continue existing governance practices than it will be to obtain commitments that will effectively require the other party to create new internal processes. The level of risk posed by the intended or foreseeable uses of the AI is also a key consideration[170]—the rigor and maturity of the parties' governance programs should be higher for high-risk uses.

Although typically not strictly within the purview of the legal team, counsel should ensure that the parties have found appropriate alignment between the AI Vendor's product capabilities, on the one hand, and the AI User's anticipated (or reasonably foreseeable) use cases, on the other hand. Key questions include:

(1) What are the purposes for which the AI was developed, and how does that compare to the uses envisioned by the AI User?
(2) Who developed the AI, and what documentation exists related to the development and ongoing monitoring and quality-control process, including documentation

[168] *See infra* Section III.C.3. (discussing contractual provisions that suspend use of the AI as needed to account for legal issues that may arise during the contract period).

[169] *See* FORBES, *AI Governance Maturity Index: A Comprehensive Assessment Framework* (July 26, 2023), https://www.forbes.com/sites/forbeseq/2023/07/26/ai-governance-maturity-index-a-comprehensive-assessment-framework/?sh=16f540384155 [https://perma.cc/YGA9-NHZH].

[170] The European Union has provided model AI contractual clauses that vary based on risk. *See EU Model Contractual AI Clauses to Pilot in Procurements of AI*, EUR. COMM'N (last visited Jan. 27, 2024), https://public-buyers-community.ec.europa.eu/communities/procurement-ai/resources/eu-model-contractual-ai-clauses-pilot-procurements-ai [https://perma.cc/JYL2-2AAT].

relating to training data used, model performance,[171] testing, and analysis as to the AI's compliance with applicable laws and regulations (e.g., absence of unintended bias)?

(3) Are the parameters within the AI static, or are they likely to change over time? If they are likely to change, what is the process by which such changes occur and what role will the Vendor and/or User play in informing or controlling such changes?

(4) What "user manuals" and/or training is provided related to use of the AI?

Given the complexity of AI technologies and uncertainty around the liability frameworks that apply, counsel should assess the responses to these questions to enable proper scoping of liability issues and inform contract negotiations.

Finally, Users, Vendors, and their counsel should aim to understand the risks associated with the specific form and type of AI being offered based on its technical functions and technical capabilities, as well as the proposed use cases. This may involve engaging AI experts to review academic literature on the methods and techniques used, searching for any lawsuits related to similar AI technologies, and evaluating the AI under a risk management framework.[172] For counsel, this risk assessment is particularly important because it will strongly inform the importance of and focus on various issues and provisions during negotiations. For purposes of fleshing out the various key issues in the discussion below, we assume that the robotic surgical assistant AI is determined to be a

[171] Examples of documentation regarding model performance are "model cards" or "factsheets." *See* Paul B. de Laat, *Companies Committed to Responsible AI: From Principles Towards Implementation and Regulation?*, 34 PHIL. & TECH. 1135, 1161–62 (2021). Model cards, proposed by employees at Google, summarize details about the model, including its intended uses, data used, performance metrics, and ethical considerations when using the model. *Id* at 1161. The term "factsheets" represents the same concept but was proposed by IBM instead of Google. *Id.* at 1162.

[172] Such frameworks may be comprehensive or industry specific. S*ee, e.g.*, NAT'L INST. OF STANDARDS & TECH., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0) (2023) (describing a comprehensive risk management framework); DEP'T OF ENERGY, *DOE AI Risk Management Playbook (AIRMP)* (last visited Dec. 10, 2023), https://www.energy.gov/ai/doe-ai-risk-management-playbook-airmp [https://perma.cc/2JD2-7TNH] (describing an industry specific risk management framework). Increasingly, many companies are adopting internal AI governance structures, which typically include AI risk and liability assessments; however, evidence seems to indicate that the maturity of such programs remains low, even among leading AI companies. *See* de Laat, *supra* note 173, at 1146–47 (analyzing the AI governance programs of twenty-four AI companies who had publicly committed to principles for AI or participated in the Partnership on AI).

"high risk" AI because of its role in medical procedures and the potential for bodily harm if the AI does not function as expected.

B. Transparency and Explainability

The transparency and explainability issues implicated by AI pose key challenges for both the User and Vendor during the due diligence phase and as part of contract negotiations. Users will want to understand how the AI works, how it reaches decisions, and how it was tested to appropriately assess risk. Vendors, on the other hand, will be concerned about the feasibility of explaining highly complex AI and the potential disclosure of valuable intellectual property and trade secrets.

Once the parties begin working on the contract, these same concerns will motivate the User to request that the Vendor documents the AI's functions through representations as to the AI's current state and covenants as to how the AI will function in the future. Vendors will likely resist out of concern that the inaccuracy of a particular representation may result in a breach and that covenants may constrain the Vendor's ability to make necessary or desired changes to the AI.

The AI's transparency may be a significant factor in representations and covenants. For "black box" AI, the Vendor may be extremely hesitant to make detailed representations and covenants as to the AI's current functions. Even where explainability tools have been added, the Vendor may want to limit representations to only those elements of the AI (or accompanying explainability tools) that it can clearly and accurately describe. AI that is intended to change over time (through User-specific learning or general retraining) presents additional challenges. Under that scenario, the Vendor may be unable to predict exactly how the model will change and thus reluctant to commit to contractual covenants as to future functions.

On the other hand, both the potential application of negligence theories to AI[173] and the current regulatory focus on AI users (rather than developers)[174] are likely to lead Users to seek even greater assurances regarding current and future AI functions.[175]

During the contracting process, two key provisions where this risk is likely to be addressed are provisions requiring the Vendor to have and maintain appropriate documentation and provisions governing the User's right to audit the Vendor.

1. Documentation

During due diligence, the User will want documentation from the Vendor that explains how the AI functions, what data is used (or has been used) to train the AI, how the Vendor monitors and tests the AI on an ongoing basis to ensure appropriate performance, and evidence of testing that Vendor has already performed.[176] As the parties move into the contracting process, the User will want to capture much or all of this information in the contract in the form of representations that such documentation is accurate and covenants that the Vendor will continue to comply with such documentation and conduct such tests or other quality assurance processes set out therein. The Vendor, on the other hand, may want to ensure that any User-oriented materials[177] it develops are referenced in the contract and that the User represents that it has received and reviewed such materials.[178]

[173] *See* discussion *supra* II.B.2.

[174] *See* discussion *supra* II.C.

[175] In general, the emergence of the law in this area is likely to play a key role in how and whether Vendors build more transparency into contracts for AI (and perhaps even the AI itself). For example, to the extent that a negligence theory of liability is extended to users of AI, that will motivate Users to demand more information, transparency, and assurances from Vendors in order to show that the User properly met its standard of care. *See* discussion *supra* II.B.2.

[176] *See* discussion *supra* III.A. The User will be further incentivized to engage in a thorough review and negotiation of terms related to documentation to the extent that negligence theory of liability might apply. *See supra* note 177 and accompanying text.

[177] For example, "model cards," "factsheets," or other user-facing guides. *See supra* text accompanying note 173.

[178] Such representations may reduce the Vendor's potential liability under a product liability theory by showing that the User understood and assumed the risk associated with use of the AI. *See* Stein, *supra* note 149, at 1006–08.

Documentation may include a description of how the AI arrives at decisions generally (whether a specific type of output or a subsequent course of action, such as in the robotic surgical assistant case) and any controls or constraints that apply to the AI's decision-making process (sometimes referred to as "global" explainability).[179] In addition, the User may want to know whether the Vendor is capable of providing logs detailing how, in a given case, the AI arrived at a specific decision based on a specific set of inputs (sometimes referred to as "local" explainability).

The Vendor, on the other hand, will want to ensure that any documentation it does provide accurately describes its processes and that it does not commit to any contractual provisions requiring functionality, testing, tools, or the like that it does not (or cannot) provide. The Vendor will also want to constrain the User's rights as needed to protect sensitive and/or confidential information and leave itself flexibility to develop its product and processes without having to obtain consent or an amendment to the contract each time it does so.

If the Vendor has developed a mature and robust governance process along with related materials describing its AI, the parties may find middle ground in the contract by including a general covenant that the Vendor will continue to adhere to its governance process (with reference to incorporated documents). If further flexibility is needed, the parties may further agree that the Vendor can update and change its processes, features, and/or functionality provided that the functionality and/or protections for User provided by such processes, features, and/or functionality is not diminished.

If the Vendor is not prepared to provide that level of detail, the parties may be able to reach an acceptable landing place by focusing on the expected outputs of the AI and describing the criteria that the outputs are expected to satisfy. While not as detailed as the User might want, a crisp description of the expected outputs of the AI may be easier

[179] "Explainability" tools and features, although separate from the AI itself, are valuable in reducing the risk and uncertainty around the use of AI. Accordingly, offering such tools may become a market differentiator for Vendors. *See, e.g.*, *Glossary – Explainability*, C3.AI (last visited Jan. 27, 2024), https://c3.ai/glossary/machine-learning/explainability/ [https://perma.cc/9GBA-SNXQ]. In certain industries, especially highly regulated industries such as finance and healthcare, such tools may become industry standard. Both Users and Vendors should be cognizant that the absence of explainability tools where such tools have become industry standard may be a relevant factor under both negligence (i.e., breach) and product liability (i.e., defective product design, failure to warn) theories.

for Vendors to agree to and may give Users comfort that, if there are problems or deficiencies in testing or underlying functionality that impact outputs, Users may have some recourse against the Vendor.

2. Audit

By "Audit," we mean provisions that enable the User to obtain additional information about the Vendor and about the AI. Such provisions may require the Vendor to provide such information directly to the User, may grant the User the right to obtain the information for itself through inspection of the Vendor's facilities and systems and access to the Vendor's personnel, or may contain a mixture of both approaches. The User will generally want to ensure that it has the right to periodically obtain information from the Vendor that shows that the AI is performing as expected and/or that the Vendor is performing is monitoring and testing obligations under the agreement. Users may want these rights to extend beyond the term of the contract, given the real possibility of delays between any problems caused during use and the filing of legal claims (i.e., any relevant statutes of limitation should be considered).

The User may seek in particular the right to request and receive documentation regarding: (1) any changes or updates to the list of data used to train or re-train the AI, including the sources of such data—in lieu of a specific list, the Vendor may opt instead to provide a general description of the data, but the User should demand that such description include whether personally identifiable information is included and, if so, what types; (2) data on performance and results of any testing, including any scenarios in which the AI is more likely to generate unexpected outcomes (e.g., outliers or blind spots); (3) data on the type and frequency of bias testing performed and the results of such testing; and, (4) the right to examine, directly or via a designated third-party, the AI, including through the use of automated tools.

The Vendor, on the other hand, will want to limit the potential that the User's exercise of audit rights will meaningfully disrupt its operations or compromise its sensitive information (e.g., proprietary data used to train the AI, trade secreters, and other confidential information). Possible contractual solutions include: (1) allowing the Vendor to provide summaries or descriptions of the information being requested in lieu of the underlying information;[180] (2) the right of Vendor to respond to specific questions by the

[180] Such as "model cards" or other user-oriented materials. *See supra* text accompanying note 173.

User in lieu of providing complete access to the AI, especially where such access may pose security risks; (3) limitations on the frequency of timing of User requests; (4) limitations on further disclosure (e.g., regulators, law enforcement, subpoenas); and, (5) explicit provisions declaring that the information is considered to be the Vendor's confidential information, coupled with robust confidentiality provisions and related remedies provisions (such as uncapped or very high liability caps that provide Vendor with access to consequential damages and/or provisions that make it easier for the Vendor to seek and obtain injunctive relief).

In any event, given the heightened public focus on AI risks and the speed with which new concerns may emerge, both parties will benefit from avoiding arbitrarily fixed limitations and should instead prioritize flexible language relevant to the AI where possible (e.g., frequency triggers based on new AI updates or versions rather than calendar or contract year).

Until the regulatory environment and the industry matures, this will likely be a subject of some tension, especially where Vendors do not have robust governance, testing, and transparency processes. [181]

C. Compliance with Laws

Given the rapidly changing legal landscape with respect to AI, it will be important to stay abreast of current compliance requirements and ensure that AI technologies in use conform to such requirements as needed. The failure to comply with laws in developing, distributing, or using AI may result in liability or, in particularly egregious scenarios, disgorgement,[182] thereby disrupting the parties' ability to perform under the contract. Generally, Users will likely want a commitment from the Vendor that the Vendor complies with, and will continue to comply with, all applicable laws, including any laws that govern the Vendor's AI technology or the data on which the AI was trained. Further, because AI regulations increasingly target Users of AI rather than Vendors,[183] Users,

[181] As the regulatory environment and the industry matures, we expect some form of User audit right to become a standard provision in these contracts, but for the scope of this right to gravitate toward a "one to many" model by which Vendors can respond to numerous User requests with some form of standard documentation, which will likely include summaries of some type of performance testing and, as industry standards and requirements develop, perhaps a report of some form of standardized external audit.

[182] *See* Goland, *supra* note 53, at 16.

[183] *See* discussion *supra* II.C.

especially those in highly regulated industries, may seek to obligate Vendors to make changes to comply with Users' legal obligations, even where such changes are not directly required of Vendors. In particular, the User will want a representation or covenant from Vendor that: (1) All training data is used in accordance with applicable laws and any obligations to third parties (e.g., contractual obligations; consents); (2) Vendor will perform its obligations in a manner that facilitates the User's compliance with the law.

In contrast, the Vendor will likely seek to avoid anything beyond general representations or covenants to comply with laws applicable to the Vendor in the conduct of its business. The Vendor is also unlikely to agree to a covenant to facilitate User's compliance, except in scenarios where the Vendor has little bargaining power or where the Vendor caters to a specific regulated industry and has decided to offer this type of commitment as a market differentiator. In addition, Vendors will likely want a commitment from Users that the User will comply with applicable law in its use of the AI, including with respect to the inputs provided to the AI by the User (e.g., with respect to data privacy laws that may govern the User's use of personally identifiable information).

Beyond these general issues of compliance, AI presents heightened risks related to bias, changes in law during the contract period, and unexpected suspension of use. Each of these risks are discussed below in more detail.

1. Biased AI

The use of biased AI may violate antidiscrimination laws.[184] This risk is likely higher for the User, who will have to rely on statements or documents from the Vendor about any bias analysis or testing that was performed.[185] Accordingly, the User should seek a representation or covenant from the Vendor that the AI is and will be free from any intended and unintended bias during the contract period. Additionally, the User should seek defense and indemnification commitments from the Vendor for any harms resulting from AI bias, whether known or unknown, in particular with respect to any third-party action against the User arising out of unintended bias (such as regulatory action or

[184] *See supra* text accompanying note 53.

[185] *See* discussion *supra* III.B. (discussing vendor documentation, including documentation related to testing.

consumer claims).[186] The need for defense and indemnification obligations from the Vendor is heightened when dealing with highly complex or adaptable AI that may be more likely to perform in unexpected ways.[187]

Of course, the Vendor is likely to strongly resist any blanket commitments about the absence of bias in its AI. Bias testing is complex and nuanced.[188] Further, there may be no single standard for bias testing for the type of AI involved, and testing for bias under all possible methods is neither feasible nor helpful.[189] A Vendor's first preference will be avoid any representation at all as to the existence of unintended bias, or else to limit such representations significantly (for example, to only that bias of which the Vendor is aware).

Where neither party can achieve its preferred language, a potential middle ground may be representations and covenants from the Vendor that it has performed and will continue to perform specific tests designed to determine whether unintended bias exists, and that it has taken and will continue to take reasonable steps to mitigate any bias uncovered in those tests. This may be acceptable to the User to the extent the User has determined

[186] Before committing to defend and/or indemnify, Vendors should research whether their current insurance policies would cover third-party claims resulting from the use of AI generally and whether such coverage also includes biased AI that may violate the law. *See* Ariel Dora Stern et al., *AI Insurance: How Liability Insurance Can Drive the Responsible Adoption of Artificial Intelligence in Health Care*, 3:4 NEW ENG. J. MED. CATALYST, Apr. 2022, at 5. (highlighting that although health care liability insurance products exist for cybersecurity and IT-related losses, such policies do not cover many failures of AI). Users should do the same in order to understand the risk if a Vendor does not agree to defend or indemnify the User. Further, should courts or legislatures impose legal status on AI, it is possible that the one party (or both) could be required to purchase a liability insurance policy on the AI's behalf. *Cf.* Duffoure, *supra* note 74, at 36–37 (proposing mandatory medical malpractice insurance for autonomous AI physicians after suggesting that such AIs should be granted legal personhood and thus subject to direct liability). This would provide the additional benefit of allowing human users to seek contribution from the AI for any fault that may be attributed to it. *See* Mindy Nunez Duffoure, *Malpractice by the Autonomous AI Physician*, 2023 U. ILL. J.L. TECH. & POL'Y 1, 45 (2023).

[187] *Cf.* Ian De Freitas, *Exploring AI Indemnities: Their Purpose and Impact*, FARRER & CO. (Mar. 10, 2023), https://www.farrer.co.uk/news-and-insights/exploring-ai-indemnities-their-purpose-and-impact/ [https://perma.cc/L2VK-ZNR6].

[188] *See supra* text accompanying note 53.

[189] *See, e.g.*, Lama H. Nazer et al., *Bias in Artificial Intelligence Algorithms and Recommendations for Mitigation*, PLOS DIGIT. HEALTH, June 22, 2023, at 6, 7 (listing five different bias testing and evaluation frameworks in the healthcare sector alone).

that such testing is reasonably likely to uncover unintended bias and/or is "industry standard" for mitigating against this harm. This may be acceptable to the Vendor because it limits the Vendor's obligations only to performing specific tasks fully within its control and avoids strict liability for the risk that the AI might have unintended bias.

2. Changes in Laws

To avoid the risk that a previously legal use of AI becomes illegal due to a change in applicable law during the contract period, both parties will want to ensure that the contract addresses the parties' rights and obligations should a change in law occur. This risk is likely greater for the User.[190] The User will prefer that: (1) the Vendor have an ongoing obligation to monitor for compliance with applicable laws applicable to both Vendor and User, to make changes to the AI product to ensure that it continues to comply with applicable laws at no additional cost to the User; (2) the Vendor have an obligation to notify the User of such changes in advance and, if desired, involve User in the development and roll-out of such changes; and (3) the User have the right to request additional changes to the AI product that the User feels are necessary for compliance with applicable laws, with constraints on the Vendor's right to decline such changes.

The Vendor will likely resist any obligation to ensure that the AI product itself is compliant with applicable laws over time or to monitor for and ensure compliance with changes in laws applicable to the User but not to the Vendor. Rather, Vendors will likely seek to limit their own compliance commitments only to ensuring that it will monitor for changes in, and continue to comply with, applicable laws applicable to Vendor in the operation of its business.

If the parties are unable to achieve their preferred levels of commitment on this issue, one possible middle ground may be establishing a baseline obligation regarding Vendor's compliance with laws clearly applicable to Vendor, coupled with a version of the mechanism for the User to request additional changes to the AI where required by law (as outlined in (3) above) in lieu of Vendor covenants to monitor laws applicable to User and ensure that the AI facilitates the User's compliance with changes in law. Of course, the Vendor is likely to demand the discretion to say no (or to adjust such changes so that they are broadly applicable to all of Vendor's customers) and to be compensated for any

[190] *See* discussion *supra* II.C (discussing the recent trend in AI-related laws to target the users of AI rather than AI developers).

work required to make such changes. One possible compromise for that issue may be to stipulate that any fees charged for such work are waived if User can clearly establish that the changes are mandatory due to a change in applicable law.

3. Suspension of Use

Given the breadth of potential use cases, the ability of some AI to evolve in unpredictable ways, and the rapidly developing legal landscape, there may be circumstances where it is necessary to temporarily suspend use of the AI system in order to address and remediate issues (including bias and changes in law as discussed in the previous subsections), especially where continued use could harm persons or property or violate the law.[191] To address such cases, the parties should consider a "circuit-breaker" provision that enables temporary suspension of the User's use of the AI in certain circumstances. Here, it is likely the User that is at greater risk of an adverse impact of suspension, as an unexpected suspension could create significant disruptions to User operations. Vendors, in turn, will likely worry about claims from Users seeking damages resulting from those impacts, as well as harm to the Vendor's reputation as a reliable provider of AI technology.

Of course, one important means of managing the risk that an unexpected suspension will occur is to arrive at clear definition of the expected and permitted use cases for the AI, appropriate governance over the AI (including ongoing testing), and appropriate monitoring of the legal landscape, as discussed in prior sections. For certain use cases involving high-risk or mission-critical functions, Vendors may also want a representation and covenant from the User that it has and will keep backup processes in place should the AI need to be suspended unexpectedly. In the case of our robotic surgical assistant, for example, the parties will want to prevent the risk that an unexpected suspension in use of the AI disrupts the surgeon from performing surgeries for a prolonged period, causing significant damage to the User and its patients. The contingency plan may involve a mix of responsibilities for both Vendor (e.g., revert to a non-AI-enabled version of the software) and User (e.g., revert to traditional or manual processes).

[191] In extreme scenarios, Vendors may be required to disgorge the AI in its entirety. *See, e.g.,* Joshua A. Goland, *Algorithmic Disgorgement: Destruction of Artificial Intelligence Models as the FTC's Newest Enforcement Tool for Bad Data*, 29 RICH. J.L. & TECH. 1, 22-23 (2023) (discussing a company's use of algorithms trained on data collected in violation of the Children's Online Privacy and Protection Act). If the Vendor is required to delete both the AI and all data used to train the AI, it may become impossible for the Vendor to perform under the contract. However, this presents more significant issues with respect to breach and impossibility that are beyond the scope of this section. Because this section focuses on suspensions that could be remediated in a reasonable timeframe, disgorgements will not be considered.

While Users may be supportive of a suspension right in circumstances where a suspension could limit the User's own risk (i.e., where continued use may put User in violation of the law), Users will likely want to enable immediate suspension only in very narrow circumstances. Further, Users will likely expect commitments that, to the greatest extent possible, suspension will occur outside of business hours and/or with appropriate advance notice. Further, to the extent the suspension is due to specific uses by Users, Users will likely want a reasonable period of time to change or stop the objectionable use prior to giving the Vendor the right to suspend and will likely want Vendor's suspension right to be as narrow in scope and duration as possible. Relatedly, Users may want assurances from Vendors that, to the extent the non-compliance results from an unexpected change to the AI itself, the Vendor has a "backup" procedure (i.e., reverting to a prior version) that it can implement within a reasonable period of time. To the extent that the suspension results from a defect in the AI itself or a failure by the Vendor to comply with its obligations under the Agreement, Users will want the Vendor to be liable for damages they may incur as a result of the suspension, including ideally consequential damages.

In light of the above, some version of a suspension right, with appropriately defined circumstances where immediate suspension without notice is permitted and appropriate constraints on the duration and timing of suspension, seems likely. To address the issue of liability of Vendors for Users' loss of the system, some form of pre-defined credit or liquidated damages payment may be an acceptable middle ground.

D. Liability

Contractual apportionment of liability is a key tool to reduce the risk that a party could be responsible for significant damages resulting from the breach of its obligations under the contract or from harm caused by that party to the other party or to third parties. This risk is not new to AI, and negotiating liability provisions for AI is likely to involve similar considerations as those implicated in technology and services contracts more broadly. For example, to address concerns about representations and covenants discussed in prior sections (e.g., documentation, quality), the parties may consider limiting the Vendor's liability for breaches of these commitments, including by capping the Vendor's liability for damages or by outlining exclusive remedies (such as repair, reperformance, or an identified credit or liquidated damages payment).

However, AI does pose some additional and/or unique risks that should be considered. In particular, AI presents challenges with respect to attribution of fault, especially in the case of human-AI interaction (e.g., our robotic surgical assistant),[192] which may lead to disagreements as to how to apply contractual provisions to specific incidents. Additionally, because of its broad set of potential applications, the universe of AI use cases will very likely include scenarios in which there is a risk of bodily injury or harm to individuals (e.g., use of AI in medicine or self-driving cars). Finally, AI has the potential to fail in systemic but hard-to-detect ways (e.g., bias or drift), meaning that damages could accrue rapidly before either party notices something is wrong. The scale of unique issues posed by AI is likely proportionate to where the AI technology lives on the spectrum of AI capabilities[193]; more flexible AI poses greater risk that liability can emerge in unexpected ways which complicate attribution between the User and the Vendor.

We consider these unique issues with respect to two potential contract provisions: disclaimer and limitations on liability.

1. Disclaimer

As with most technology transactions, the Vendor is likely to seek a disclaimer of any implied warranties that may exist. Similarly, the Vendor is likely to seek to limit the remedies that may be available for any warranties as to quality or non-compliance. In addition, Vendors should seek an express statement that the User assumes any risk arising from its own use of the AI, especially in cases of human-AI interaction or for uses that pose risk of bodily injury or harm.[194] In order to strengthen the User's assumption of the risk, the Vendor may seek a contractual obligation on the User's part (on behalf of its personnel and/or third parties who will use the AI) requiring the User to familiarize itself with the AI's features, functionality, proper use, and inherent risks.[195] Additionally, the Vendor may require the User to sign a statement confirming that it has read and received any documentation related to the AI.

[192] *See* discussion II.B.1 (discussing causation issues under a products liability theory), II.B.3 (discussing contributory negligence and assumption of risk).

[193] *See* discussion *supra* I.A.

[194] To the extent a product liability theory is applied to the Vendor's liability for AI, such an express assumption of risk by the User may be particularly helpful in reducing the Vendor's liability. *See* discussion *supra* II.B.1.

[195] Such a provision may help to establish that the User voluntarily and knowingly assumed the risk, even if the User rejects an express assumption of risk. *See supra* text accompanying note 151 (discussing primary implied assumption of risk).

While it is fairly typical to accept a disclaimer of implied warranties, the User will want to be careful to avoid a disclaimer that is so broad that it effectively eliminates any basis for the User to hold the Vendor liable for harm caused by the AI. To the extent the User has been successful in obtaining adequate representations and covenants regarding documentation and quality,[196] the User may push for an exception to the general disclaimer for any express warranties made by the Vendor in the contract and/or in any writings incorporated into the contract by reference. If limited remedies exist, the User will want to remove these or try to limit the scenarios and/or harms to which any limited remedies apply. Importantly, the User will want to ensure that any limited or exclusive remedies do not unintentionally displace other claims the User might have arising from the same transaction or occurrence. For example, if an AI malfunction implicates a breach of more than one representation or covenant (e.g., an errant output breaches a representation as to quality and a covenant as to the protection of User's confidential information), the User will want to make clear in the contract that the User is not foreclosed from seeking remedies for each breached representation and covenant. The User should be hesitant to accept any language proposed by the Vendor that limits or eliminates the User's remedies when the breach was caused in part or in whole by the User, especially when the User may be at a disadvantage to establish fault given its limited knowledge of the AI's inner workings and the ways in which the AI interacts with User input and feedback.

2. Limitations on Liability

With respect to contractual limitations on liability, the Vendor's considerations will, as with disclaimers, likely mirror those present when negotiating other technology and services contracts.[197] The User's considerations, on the other hand, are likely to change substantially as a result of the unique challenges posed by AI. Specifically, the User should be concerned about limiting the Vendor's liability for consequential damages related to AI that violates the law. The transparency issues discussed above, combined with the User's potentially limited knowledge of the AI and the risks it poses, exacerbates the systemic risk that AI-driven violations accumulate rapidly and invisibly. Even in a scenario where the User has knowledge that continued use of the AI may violate the law (e.g., if the law changes during the period of use), the User should avoid limiting the

[196] *See* discussion *supra* III.B.1.

[197] For example, the Vendor will likely seek exclusions from or limitations on liability for consequential damages, reputational damages, damages to persons and/or property, loss of goodwill and lost profits.

Vendor's liability when damage results from the Vendor's failure or refusal to make necessary changes requested by the User.[198] Additionally, the User will likely want to avoid excluding or limiting liability for damages resulting from the forced suspension of the AI as a result of regulatory or legal action, if the reason for such suspension can be traced to the Vendor.

Both parties may want to limit liability whenever damages can be traced to the fault of the other party. However, fault can be extremely difficult to assign, especially in cases of human-AI interaction or in the absence of local explainability. Attribution is likely to be fact-specific and thus difficult to define in advance. One potential solution is a contractual commitment to engage a third-party neutral (or other alternative dispute resolution mechanism) to evaluate and attribute fault between the User and Vendor.

Even if the User agrees to significant exclusions from or limitations of liability for the Vendor, it is likely to expect that the Vendor will defend (or facilitate the defense of) the User against third-party claims, especially given the Vendor's superior knowledge as to the AI's development and functioning. The existence of a separate and independent mechanism for attributing fault (as discussed in the prior paragraph) could enable the parties to collaborate on defense against third-party claims while preserving a mechanism for the allocation of defense costs and liability between User and Vendor.

E. Quality

This issue is related to the discussion of Transparency and Explainability.[199] In addition to obtaining representations and covenants as to the functioning of the AI system over time, the User will want to ensure that the Vendor's statements as to quality hold up—that the AI works "as advertised."[200] Meanwhile, the Vendor will want to ensure that the User takes responsibility for any role that they play in the normal functioning of the AI so that User failures are not mistaken for quality issues. Establishing a clear "line of demarcation" between the parties' respective responsibilities regarding use will be key, particularly if courts begin to evaluate AI under agency theories (for which control is a key factor).[201] One key difference when considering quality issues for AI is that the technology may be

[198] *See* discussion *supra* III.C.2.

[199] *See* discussion *supra* III.B.

[200] *See* METI CONTRACT GUIDELINES, *supra* note 164, § VI.2(1)(iv).

[201] *See* discussion *supra* II.A.

intended to change over time, and in unexpected ways. This means that quality must be monitored and maintained on an ongoing basis, not only during the "acceptance" phase that typically proceeds the launch of an engagement.

1. Summary of Product Features and Vendor Program

To protect both parties, the contract should clearly describe the AI technology or system and its intended purpose as of the effective date. This enables the User to feel confident that the expectations of the system's performance are clear so that the AI is usable and reliable, and any potential deviations can be identified and addressed.

As discussed above,[202] the User will want the description of product features to be contained within the written agreement (or at a minimum, within accompanying documents that are incorporated by reference). On the other hand, the Vendor may be hesitant to commit to a fixed description of an AI system that might change. The Vendor will also want to ensure that any statements as to purpose do not constitute an implied warranty of fitness, especially for flexible and adaptable AI systems that are closer to AGI on the spectrum.[203] The Vendor is likely to request that the description in the contract be high level and reference the Vendor's online terms or user guide so that the Vendor can maintain control over the description as needed to reflect an evolving AI technology or system.

One potential area for compromise is a detailed description of the product features that will not change (or perhaps are unlikely to change between different versions of the AI), such as: (1) A description of data used to train the version of the AI system referenced in the contract and how such data was obtained; (2) A commitment that certain types of data were not used to train the AI and, if can be stated with confidence, will not be used in future retraining; and, (3) Expectations and responsibilities of the User when using the AI (e.g., cleaning and/or parametrizing data inputs; security). While one or the other of the parties may seek to include statements that both parties will not act negligently in their provision or use of the AI, given the significant uncertainty around breach and

[202] *See* discussion *supra* III.B.1.

[203] *See* discussion *supra* III.D.1 (discussing a general disclaimer of any implied warranties).

causation issues as applied to AI,[204] both parties should be hesitant to explicitly incorporate a negligence standard into the contract.

2. Security and Privacy

Maintaining the security of an AI system after release is critical to ensure that it continues to perform as expected. This is especially true for AI systems that learn and retrain over time, as such features may require greater integration between User and Vendor systems, thereby increasing the possible vectors for security incidents.

The User will want the Vendor to commit that it has adopted measures to prevent attempts by unauthorized parties to corrupt the data used to train the AI, the AI system itself, or any ancillary tools used for testing or explainability. The User will want a broader notification requirement for any security incident that may compromise, directly or indirectly, the quality or reliability of the AI system or related tools. Additionally, the User will want to limit its own responsibility for improper use of credentials required to access the AI system to only those individuals over whom the User exerts control.

The Vendor, on the other hand, will want to avoid overly prescriptive security requirements and overly broad notification requirements. Where there is significant continued integration between the User and Vendor systems to facilitate ongoing training and testing, the Vendor will expect the User to commit equally to security measures and notification requirements. The Vendor will want to avoid liability for scenarios in which the User has allowed unauthorized individuals indirect access to the AI system that compromises the system's confidentiality, integrity or reliability (e.g., when the AI is embedded within a broader system or device without the necessary firewalls).

F. Governing Law

Determining the law that will govern the contract is a critical aspect of any negotiation. However, AI utilization contracts require additional consideration, largely due to uncertainty around the treatment of AI under various legal regimes.[205] Further, the

[204] *See* discussion of challenges related to causation under a products liability theory (although such challenges extend to causation under a negligence theory as well), *supra* II.B.1, and challenges related to breach, *supra* II.B.2.

[205] *See* discussion *supra* II.

regulation of AI is a rapidly evolving area.[206] Lawyers for both parties will want to ensure that they have a grasp of trends related to the regulation of AI within the governing law jurisdiction and are monitoring regulatory action up until the moment the contract is signed, because such regulation may affect the enforceability of contract clauses or add unintended implied terms to the contract.[207] Additionally, both parties should be aware of any laws beyond the governing law stated in the contract that might apply, such as laws governing the processing of personal data or cross-border data transfers (especially where the User sends data to the Vendor's API in order to access the AI technology)[208] and laws governing the use or provision of AI and laws governing the export or import of AI technology.

Another challenge to consider is the impact that regulation from the European Union may have on potential conflicts between the governing law and the contracting requirements sought by parties with significant presence in the EU.[209] These challenges have been widely documented with respect to the General Data Privacy Regulation. Lawyers can learn from their experience with the GDPR to guide contracting as the full force of the EU AI Act begins to come into effect.[210]

CONCLUSION

AI is increasingly being adopted by businesses to improve their operations. These AI technologies are often developed by an AI Vendor and licensed by the business User. At the same time, AI technologies are becoming increasingly complex, and concern is growing around the lack of transparency in the data used to train AI and the AI's inner workings.

[206] INT'L ASS'N PRIV. PROS., *Global AI Legislation Tracker* (2023), https://iapp.org/resources/article/global-ai-legislation-tracker/ [https://perma.cc/V99Y-AZNS].

[207] Domien Kriger et al., *Key Challenges of Artificial Intelligence: Contracting for the Purchase and Use of AI*, *in* DENTON'S A.I. GUIDE 2022 (2021).

[208] *See* METI CONTRACT GUIDELINES, *supra* note 164, § VI.2(1)(iii).

[209] Numerous scholars and policymakers have described the outsized effect that European regulations have on global commerce, often referred to as the "Brussels Effect." Annegret Bendiek & Isabella Stuerzer, *The Brussels Effect, European Regulatory Power and Political Capital: Evidence for Mutually Reinforcing Internal and External Dimensions of the Brussels Effect from the European Digital Policy Debate*, 2 DIG. SOC'Y 5, Jan. 2023, at 5.

[210] *See* CHARLOTTE SIGEMANN & MARKUS ANDERLJUNG, *The Brussels Effect and Artificial Intelligence: How EU Regulation Will Impact the Global AI Market* CTR. FOR GOVERNANCE OF AI (2022).

Competing economic and legal forces may prevent Users and Vendors from achieving full transparency when partnering. Alongside this tension between AI Vendors and Users is a growing societal awareness in the inherent fallibility of many AI technologies, especially with respect to risks of bias and discrimination. Numerous public and private bodies have begun work to describe and categorize the harms that may result from AI into comprehensive frameworks, although no single framework has yet been established. Nevertheless, the growing awareness of AI risks alongside the development of established harm frameworks increases the risk that AI failures lead to legal liability for both Users and Vendors.

Currently, it is unclear which liability theories are likely to be most applicable to AI. Although both employer liability (respondeat superior) and products liability frameworks may apply, it is possible that the latter is a more natural fit given that AI does not currently have legal status. Within products liability, it is unclear whether courts will generally consider AI to be a product or a service, or whether such decision will depend on the context in which the AI is used. How AI is categorized by the courts will have a significant impact on the elements required to establish liability and on which party is likely to be liable when the AI causes harm. Further, AI presents unique challenges to established jurisprudence under both theories. Moreover, scholarship to date has focused on traditional torts and scholars have yet to explore the potential for AI liability under the broader set of business torts, including misrepresentation. And finally, legislatures and regulators may adopt new laws that create private rights of action that accompany (or perhaps even preempt) common law claims such as those discussed here. For all of these reasons, both Users and Vendors of AI should expect significant short-term uncertainty in the courts. Regardless of the future path for AI liability, the practices adopted by the industry in the contracts written today are likely to become valuable and persuasive references for courts as they assess cases involving AI.

Given this uncertainty, both Users and Vendors should be intentional in their use of contracts to reduce risk and facilitate use of AI. Addressing AI liability through contracting requires the parties to consider the unique challenges posed by AI at all steps in the negotiating and contracting process. To the extent that companies have invested in robust AI governance and risk management programs generally, the contracting process for specific AI technologies is likely to be significantly more efficient. Nevertheless, the parties are likely to encounter conflicting interests in their negotiation positions, especially with respect to concerns of AI bias and the potential for new laws to emerge during the contract period. Although we have provided some potential areas for compromise here, the parties will need to be flexible and creative in order to find solutions that work in this highly uncertain and rapidly changing area.