ARTIFICIAL INTELLIGENCE OPINION LIABILITY

Yavar Bathaee†

ABSTRACT

Opinions are not simply a collection of factual statements—they are something more. They are models of reality that are based on probabilistic judgments, experience, and a complex weighting of information. That is why most liability regimes that address opinion statements apply scienter-like heuristics to determine whether liability is appropriate, for example, holding a speaker liable only if there is evidence that the speaker did not subjectively believe in his or her own opinion. In the case of artificial intelligence, scienter is problematic. Using machine-learning algorithms, such as deep neural networks, these artificial intelligence systems are capable of making intuitive and experiential judgments just as humans experts do, but their capabilities come at the price of transparency. Because of the Black Box Problem, it may be impossible to determine what facts or parameters an artificial intelligence system found important in its decision making or in reaching its opinions. This means that one cannot simply examine the artificial intelligence to determine the intent of the person that created or deployed it. This decouples intent from the opinion, and renders scienter-based heuristics inert, functionally insulating both artificial intelligence and artificial intelligence-assisted opinions from liability in a wide range of contexts. This Article proposes a more precise set of factual heuristics that address how much supervision and deference the artificial intelligence receives, the training, validation, and testing of the artificial intelligence, and the a priori constraints imposed on the artificial intelligence. This Article argues that although these heuristics may indicate that the creator or user of the artificial intelligence acted with scienter (i.e., recklessness), scienter should be merely sufficient, not necessary for liability. This Article also discusses other contexts, such as data bias in training data, that should also give rise to liability, even if there is no scienter and none of the granular factual heuristics suggest that liability is appropriate.

DOI: https://doi.org/10.15779/Z38P55DH32

† Litigator and computer scientist. This Article is dedicated to my wife, Jacqueline, and my children, Elliot and Audrey. I would like to thank James Steiner-Dillon for his comments on the Article and his support. All errors and omissions are my own.

TABLE OF CONTENTS

I. INTRODUCTION

Opinion statements are everywhere. They express judgments about things such as value,[1] probability,[2] or the appropriate course of action.[3] They are more than the facts underlying them; they are also the weights the person stating the opinion attaches to those facts. That is why opinion statements not only include factual statements, they also implicitly say something about the person expressing the opinion—namely, that the person stating the opinion has a basis for it, that they genuinely believe in the opinion, and that they are not aware of facts and reasons that would undermine the opinion.[4]

---

1. Statements about valuation are generally regarded as statements of opinion because, when there is no clear market price for an asset, the "fair value" of an asset "will vary depending on the particular methodology and assumptions used." Fait v. Regions Fin. Corp., 655 F.3d 105, 111 (2d Cir. 2011). Indeed, in many cases "[t]here may be a range of prices with reasonable claims to being fair market value." Henry v. Champlain Enters., Inc., 445 F.3d 610, 619 (2d Cir. 2006). Although much of the discussion of valuations as opinions have been in the securities law context, valuations have been treated as opinions in other fields of law, including contract. *See, e.g.*, RESTATEMENT (SECOND) OF CONTRACTS § 168 cmt. c (AM. LAW INST. 1981) ("A statement of value is, like one of quality, ordinarily a statement of opinion.").

2. An opinion statement often carries with it the implicit statement that it encompasses a belief based on incomplete information or based on uncertain facts. Indeed, the Restatement of Contracts states that "[a]n assertion is one of opinion if it expresses only a belief, without certainty, as to the existence of a fact or expresses only a judgment as to quality, value, authenticity, or similar matters." RESTATEMENT (SECOND) OF CONTRACTS § 168. Indeed, because opinions rest on the "weighing of competing facts," it is generally understood that stating an opinion is a way of "conveying uncertainty." Omnicare, Inc. v. Laborers Dist. Council Constr. Indus. Pension Fund, 135 S. Ct. 1318, 1329 (2015).

3. A recommendation or prognosis statement by an expert is a classic example of an opinion statement that may give rise to liability. Indeed, some of the earliest opinion liability cases in the United States concerned statements made by physicians about diagnosis and prognosis. *See, e.g.*, Hedin v. Minneapolis Med. & Surgical Inst., 64 N.W. 158, 160 (Minn. 1895) (noting that the physician's diagnosis came with it an opinion that "a representation that plaintiff's physical condition was such as to insure a complete recovery"). When the opinion of an expert, such as a medical professional, is involved, liability has traditionally turned on whether the speaker's role as an expert invited reliance on the opinion. *See, e.g.*, Gagne v. Bertran, 275 P.2d 15, 21 (Cal. 1954) ("Moreover, even if defendant's statement was an opinion, plaintiffs justifiably relied thereon. Defendant held himself out as an expert, plaintiffs hired him to supply information concerning matters of which they were ignorant, and his unequivocal statement necessarily implied that he knew facts that justified his statement.").

4. *See Omnicare*, 135 S. Ct. at 1334 (Scalia, J., concurring) ("In a few areas, the common law recognized the possibility that a listener could reasonably infer from an expression of opinion not only (1) that the speaker sincerely held it, and (2) that the speaker knew of no facts incompatible with the opinion, but also (3) that the speaker had a reasonable basis for holding the opinion."); *see also* RESTATEMENT (SECOND) OF CONTRACTS § 168 (noting that an opinion comes with it the assertion that "the facts known to that person are not incompatible with his opinion," or "that he knows facts sufficient to justify him in forming it"); *id.* § 168

The law has developed significant aptitude at evaluating the truth or falsity of factual statements based on evidence.[5] However, determining whether a speaker genuinely believes in their opinion will often require intent-based heuristics—the most notable of which is scienter.[6] Since opinion statements are not true or false merely because some fact the opinion is based upon proves to be true or false, these heuristics, which are described in Part II, are in many cases outcome-determinative on the question of liability.

The value of these intent-based heuristics will likely be aggressively challenged by a new breed of computer programs capable of forming and stating opinions—artificial intelligence (AI).[7] For the first time in human history, artificially intelligent computer programs are capable of rendering opinions without deterministic instructions.[8] They can learn from data—from experience—and come to intuitive conclusions without the aid of a human

---

cmt. a ("A statement of opinion is also a statement of fact because it . . . has a particular state of mind concerning the matter to which his opinion relates.").

5. Indeed, the stated purposes of the Federal Rules of Evidence include "the end of ascertaining the truth." FED. R. EVID. 102. Many of the rules themselves are addressed to determining the admissibility, relevance, and reliability of statements, the most notable of which is the hearsay rule and its exceptions. *See* FED. R. EVID. 801–802 (addressing the admissibility of statements, including out-of-court statements that are offered for their truth).

6. This is because the opinion carries with it the implicit statement that the opinion is genuinely believed by the speaker. Thus, proving the subjective falsity of the opinion is functionally the same as proving scienter. *See In re* Credit Suisse First Bos. Corp., 431 F.3d 36, 48 (1st Cir. 2005) ("[T]he subjective aspect of the falsity requirement and the scienter requirement essentially merge; the scienter analysis is subsumed by the analysis of subjective falsity."). Another useful heuristic is to determine whether the factual assumptions underlying an opinion hold true; if they do not, then the opinion itself is undermined because the speaker's intent is called into question. *See* Va. Bankshares v. Sandberg, 501 U.S. 1083, 1093 (1991) ("Provable facts either furnish good reasons to make a conclusory commercial judgment, or they count against it, and expressions of such judgments can be uttered with knowledge of truth or falsity just like more definite statements, and defended or attacked through the orthodox evidentiary process that either substantiates their underlying justifications or tends to disprove their existence.").

7. AI, as referred to in this Article, is a class of computer programs designed to solve problems that typically require "inferential reasoning [and/or] decision-making based on incomplete or uncertain information, classification, optimization, and perception." Yavar Bathaee, *The Artificial Intelligence Black Box and The Failure of Intent and Causation*, 31 HARV. J.L. TECH. 889, 920 (2018).

8. Although some forms of AI do in fact rely on deterministic instructions, see Bathaee, *supra* note 7, at 898, the AI addressed in this Article generally are not deterministically programmed, but are instead trained from examples using machine-learning algorithms—that is, they are computer programs that learn directly from data. *See* ETHEM ALPAYDIN, INTRODUCTION TO MACHINE LEARNING xxv (2004) ("We need learning in cases where we cannot directly write a computer program to solve a given problem, but need example data or experience. One case where learning is necessary is when human expertise does not exist, or when humans are unable to explain their expertise.").

being.[9] What then do intent-based heuristics achieve when the intent of the AI's creator or user does not necessarily affect or reflect the judgment or opinions of the AI? As this Article contends, very little.

As explained in Part III, this decoupling of the AI creator's intent from the AI's judgments arises from a technological problem that occurs when certain classes of machine-learning algorithms are used by AI—the Black Box Problem.[10] The black box problem arises where machine-learning algorithms rely on layers upon layers of linear and non-linear transformations, such as deep artificial neural networks. These algorithms are capable of learning from data and experience, just as humans do, but such powerful cognition comes at the price of transparency.[11] A trained neural network, for example, may have internalized hundreds of thousands, if not millions of data points, and may arrive at accurate predictions or sound opinions, but the complexity of the neural network may make it impossible to determine how the AI has made its judgments or reached an opinion.[12] Thus applying intent-based heuristics will almost never result in liability.[13]

Today, AI helps perform tasks that in the past have required human judgment and experience.[14] For example, AI can achieve higher accuracy at spotting certain forms of cancer—a task that in the past required a trained doctor with years of experience to perform.[15] The bread and butter of finance and accounting, valuation, will also soon be predominantly a task relegated to AI.[16] Even before the AI revolution, algorithmic valuation was a rapidly

---

9. *See* Bathaee, *supra* note 7, at 891. Because machine learning-based AI can learn directly from data instead of simply implementing rigid pre-programmed rules, it "can learn, adapt to changes in a problem's environment, establish patterns in situations where rules are not known, and deal with fuzzy or incomplete information." MICHAEL NEGNEVITSKY, ARTIFICIAL INTELLIGENCE 14 (2d ed. 2005).

10. *See infra* Section III.B.

11. *See infra* Section III.B & III.C.

12. For a detailed discussion of the AI Black Box Problem and how it arises from the use of certain machine-learning algorithms, see Bathaee, *supra* note 7, at 897–906.

13. *See* Bathaee, *supra* note 7, at 906–21.

14. *See infra* Sections III.B & III.C.

15. *See, e.g.*, Andre Esteva et al., *Dermatologist-level Classification of Skin Cancer with Deep Neural Networks*, 542 NATURE 115 (2017); Martin Stumpe & Lily Peng, *Assisting Pathologists in Detecting Cancer with Deep Learning*, GOOGLE RES. BLOG (Mar. 3, 2017), https://research.googleblog.com/2017/03/assisting-pathologists-in-detecting.html [https://perma.cc/2BMT-YCTX] ("In fact, the prediction heatmaps produced by the algorithm had improved so much that the localization score (FROC) for the algorithm reached 89%, which significantly exceeded the score of 73% for a pathologist with no time constraint."); *see also* Ahmed Hosny et al., *Artificial Intelligence in Radiology*, NATURE REVS. CANCER (May 17, 2018).

16. *See infra* Section III.C.

growing field.[17] With the ability to build models that can accurately learn from vast amounts of data, the number of AI-based valuation systems is only expected to multiply. AI will also likely assist other specialized experts with judgments, including judges and arbitrators.[18]

Under the current prevailing standards for opinion liability, a court will find liability only based on the intent of the humans that stated the opinion.[19] But, when an AI opinion is involved, its decisions will be based on data, and the intent of the creators or users of the AI will generally not provide insight into the AI's decision-making process.[20] And since the AI may suffer from the Black Box Problem, it may not have an ascertainable intent that can be examined or queried.[21] The net effect of this will be the end of opinion liability in many fields of law that require some form of intent, such as scienter, because intent-less AI and AI-assisted opinions will be functionally immune.[22]

Part IV of this Article argues that the current opinion liability regime requires two significant adjustments. First, more precise heuristics—designed specifically for AI—are needed. That is, courts and factfinders should look to (i) the extent to which the AI model was given deference and autonomy, (ii) the manner in which the AI was trained, validated, and tested, and (iii) the extent to which a priori constraints were placed on the judgments of the AI system to mitigate known risks.[23]

It is possible that these heuristics point to recklessness on the part of the creator or user of the AI, and in such a case, there may be a permissible inference of scienter,[24] but as this Article explains, there may also be other

---

17. Indeed, automated valuation models, which were based on deterministic algorithms (not modern AI) were a prominent feature of the mortgage crisis of 2008. *See, e.g.*, Mass. Mut. Life Ins. Co. v. DB Structured Prods., 110 F. Supp. 3d 288, 293 (D. Mass. 2015) (discussing automated valuation models used for due diligence and appraisals of real property prior to the real estate crisis of 2008); Fed. Hous. Fin. Agency v. Nomura Holding Am., Inc., 60 F. Supp. 3d 479, 491–92 (noting that automated valuation models were used by government-sponsored entities, such as Fannie Mae, to assess values of homes underlying mortgage-backed securities they purchased).

18. In Wisconsin, for example, the state's Supreme Court recently ruled that the use of actuarial data to predict recidivism did not offend a defendant's due process rights, even though the data and methodology was not disclosed to the court or the defendant. *See* State v. Loomis, 88 N.W.2d 749 (Wis. 2016). For a full discussion of the case, see Case Comment, *Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing:* State v. Loomis, 130 HARV. L. REV. 1530, 1534 (2017).

19. *See infra* Section III.D.

20. *See* Bathaee, *supra* note at 7, at 906–21.

21. *Id.*

22. *See infra* Section III.D.

23. *See infra* Section IV.A.

24. *See infra* Section IV.E.

circumstances that would warrant liability. For example, there may be some applications that would require a strict liability rule—those that involve high risks of harm or that implicate governmental or societal norms that would require human, not machine, judgment.[25] It may also be the case that a failure to detect significant bias in the data used to train the AI should itself warrant liability.[26]

In such cases, there may be no basis for an inference of scienter, including under the more precise heuristics proposed by this Article.[27] That does not, however, mean that liability for an opinion statement should not attach.[28] Accordingly, the second modification this Article proposes to the status quo is that scienter should be sufficient, not necessary for liability when AI is involved.[29] Most opinion liability regimes have it the other way around, requiring a showing of scienter for opinion liability—even in some cases where a statute does not require scienter for liability.[30] However, when AI is involved, requiring scienter will immunize a wide swath of conduct and provide a host of perverse incentives to use AI to shield opinions from liability.

With this new technology comes the promise of multiplying and perhaps exceeding human intelligence by orders of magnitude,[31] but with that comes the need to create new legal and factual heuristics designed for machines—not to make patchwork adjustments to legal doctrines designed to understand human conduct. Indeed, if the status quo would immunize almost all AI opinion from liability, there may be no occasion to make thoughtful and incremental adjustments to our legal doctrines.

---

25. *See infra* Section III.C.
26. *See infra* Section III.B.
27. *See infra* Section III.F.
28. *See id.*
29. *See infra* Section III.F.
30. *See infra* notes 247–48 and accompanying text.
31. As predicted for decades by commentators on AI, AI systems already exceed humans in perception-based tasks, such as vision. KURZWEIL, THE AGE OF SPIRITUAL MACHINES 65 (2000); Gina Smith, *Google Brain Chief: AI Tops Humans in Computer Vision, and Healthcare Will Never Be the Same*, SILICON ANGLE (Sept. 27, 2017), https://siliconangle.com/2017/09/27/google-brain-chief-jeff-dean-ai-beats-humans-computer-vision-healthcare-will-never/ [https://perma.cc/95AQ-8TYD]. Experts predict that AI systems will exceed humans in tasks such as language translation and truck driving within the coming decade. *Experts Predict When Artificial Intelligence Will Exceed Human Performance*, MIT TECH. REV. (May 31, 2017), https://www.technologyreview.com/s/607970/experts-predict-when-artificial-intelligence-will-exceed-human-performance/ [https://perma.cc/Y4PA-YPTX].

II. OPINION LIABILITY, THE SCIENTER HEURISTIC, AND INFORMATION ASYMMETRY

This Part describes the unique challenges posed by opinion statements as well as some of the heuristics used to determine whether the speaker of an opinion should be held liable. This Part does not survey any particular area of law but instead attempts to describe how familiar heuristics, such as scienter and reliance, solve many of the problems posed by opinion statements. These problems include, for example, information asymmetry, contrary or incomplete information, and unreasonable or inadequate bases for the opinion. This Part concludes that opinions are factual models, which include not only a set of underlying facts, but also probability weights for those facts and notions, acquired through the speaker's experience.

A. THE OPINION/FACT DISTINCTION

Factual statements are often at the center of legal disputes. Proving a factual statement true or false lies in finding empirical facts as they existed when the statement was made and comparing those facts to what was conveyed in the statement.[32]

The question of whether to impose liability based on a false statement, however, will not be a simple matter of determining what facts existed, were known, or were knowable when the statement was made. Instead, the question is often about the overall context of the statement and what the speaker intended to accomplish.[33] There are many battle-tested heuristics for dealing

---

32. In securities cases, falsity of a factual statement is often a necessary predicate for liability and can be pled or proven with evidence that the facts as they existed when the statement was made contradicted the factual statement. *See In re* Homestore.com, Inc. Sec. Litig., 252 F. Supp. 2d 1018, 1032 (C.D. Cal. 2003) (noting that falsity can be pled where defendant is in "possession of non-public information that would prove his statements false"); Plevy v. Haggerty, 38 F. Supp. 2d 816, 826 (C.D. Cal. 1998) (noting that falsity can be pled by "direct or circumstantial facts, such as, but not limited to, inconsistent contemporaneous statements or internal reports, that would support [that the statements] . . . were false when made"). In other contexts, such as false statements under the Lanham Act, courts have focused on whether a statement of fact is "measurable" and "specific" enough to be proven to be false. Franklin Fueling Sys. v. Veeder-Root Co., No. S-09-580 FCD/JFM, 2009 U.S. Dist. LEXIS 72953, at *13 (E.D. Cal. Aug. 11, 2009); *see also, e.g.*, Hi-Tech Pharm., Inc. v. HBS Int'l Corp., 910 F.3d 1186, 1193 (11th Cir. 2018) (alleging that precise advertisement and representation of drink's percentage of protein content was sufficiently specific to be proven false by an alleged test showing a lower amount of protein in a Lanham Act claim). These courts suggest that what makes a statement of fact provably true or false is the specificity of the statement, the ability to measure the information conveyed in the statement, and the existence of consistent or inconsistent contemporaneous evidence.

33. *See, e.g.*, Tolles v. Republican-American, No. UWYCV106005674, 2012 Conn. Super. LEXIS 2877, at *9 (Super. Ct. Nov. 20, 2012) ("Connecticut law makes clear that in

with the host of issues that arise as part of the liability question, such as evaluating and comparing the credibility of witnesses,[34] examining the motives of the person making the statement (and in some cases of those that heard it),[35] and evaluating whether the statement was important enough to have affected a transaction or a decision-making process.[36]

---

determining the scope of the alleged statement, and further in determining its truth or falsity, context is important and sometimes even dispositive."); Buetow v. A.L.S. Enters., 650 F.3d 1178, 1185 (8th Cir. 2011) ("In assessing whether an advertisement is literally false, a court must analyze the message conveyed within its full context.") (quoting United Indus. v. Clorox Co., 140 F.3d 1175).

34. The Federal Rules of Evidence, for example, provide for the impeachment of witnesses precisely because credibility is a powerful heuristic for assessing whether the facts conveyed by the witness are true, including whether the witness's testimony contradicts his own prior inconsistent statements. *See* FED. R. EVID. 613(b). The Federal Rules of Evidence accordingly treat out-of-court statements offered for impeachment as non-hearsay statements because they are not being offered for their truth. *See* Hartford Fire Ins. Co. v. Taylor, 903 F. Supp. 2d 623, 642 (N.D. Ill. 2012) (holding that out of court statement offered for impeachment was not hearsay).

35. Courts routinely consider a speaker's motive to make a false statement. In fact, the motive to have made a false or misleading statement is an important part of the scienter inquiry required for most fraud-based claims. *See In re* PXRE Grp., Ltd. Sec. Litig., 600 F. Supp. 2d 510, 531 (S.D.N.Y. 2009) (pleading securities fraud requires alleging facts indicating a "motive and opportunity probative of a strong inference of scienter") (quoting Rothman v. Gregor, 220 F.3d 81, 90 (2d Cir. 2000)). Even in circuits where motive and opportunity are not sufficient for scienter, they are an important part of the analysis. *See In re* Silicon Storage Tech., No. C 05-0295 PJH, 2006 U.S. Dist. LEXIS 14790, at *50 (N.D. Cal. Mar. 10, 2006) ("In the Ninth Circuit, motive and opportunity, standing alone, are not sufficient to establish scienter . . . . However, motive can be considered as part of the 'totality of the allegations' regarding scienter.") (internal citations omitted). What matters is that the alleged motive indicates a clear reason to make a false statement such that one can infer scienter. It will therefore not be enough to allege, for example, a speaker's generalized motive to maximize profits or to justify management decisions, because all companies or businessmen have such a motive—not just those that make false statements. *See* Zirkin v. Quanta Capital Holdings Ltd., No. 07 Civ. 851 (RPP), 2009 U.S. Dist. LEXIS 4667, at *35 (S.D.N.Y. Jan. 22, 2009) ("A motive to maintain a higher financial rating to protect the viability of the Company, which is what the Complaint alleges here, is not enough, under the law of this Circuit, to sufficiently put forth a claim that a statement contained in an offering document was 'fraudulent' at the time it was made."); *see also* Alaska Elec. Pension Fund v. Adecco S.A. (*In re* Adecco S.A.), 371 F. Supp. 2d 1203, 1223 (S.D. Cal. 2005) ("A desire to conceal mismanagement is not sufficient to show motive and opportunity.").

36. Both the doctrines of materiality and reliance serve this purpose. Materiality, which is required for many fraud-based claims, assesses whether a reasonable person would have considered the false statement important to his decision to enter into a transaction. *See* RESTATEMENT (SECOND) OF TORTS § 538 (AM. LAW INST. 1977) (a statement is material if, inter alia, "a reasonable man would attach importance to its existence or nonexistence in determining his choice of action in the transaction in question"); *see also* United States v. Raza, 876 F.3d 604, 619 (4th Cir. 2017) ("[T]he relevant elements of wire fraud are an intent to defraud and materiality, which Colton defined as 'what a reasonable financial institution would

Heuristics such as scienter, materiality,[37] and reliance[38] thus generally get at the heart of many of the issues presented by the fact liability question. These heuristics ask the natural questions about factual statements, such as whether the speaker intended to mislead the person buying the car, whether the error would matter to a reasonable person buying a car, and whether the purchaser was entitled to (and did) rely on the statement because of some information asymmetry or because of the expertise or conduct of the speaker. All of these heuristics focus on the speaker and the context.[39]

Where there is a materially false statement, damages or rescission will often be available. In contract law, for example, there will be an escape hatch for mistake or when there is a failure to reach a meeting of the minds.[40] And, of course, where there is scienter sufficient for fraud, a contract will be voidable.[41] In some cases, there may be a statutory cause of action that provides relief for

---

want to know in negotiating a particular transaction.' "). Reliance, which is also an element of most fraud claims, requires that the person hearing the false statement thought the statement was important enough to act upon. Both doctrines ensure that unimportant statements, even if provably false, do not give rise to liability.

37. *See generally* Wendy Gerwick Couture, *Materiality and a Theory of Legal Circularity*, 17 U. PA. J. BUS. L. 453, 455 (2015) ("[Materiality doctrine] divid[es] misrepresentations that are potentially actionable from those that pose no risk of liability.").

38. *See* Daniel B. Dobbs, *The Place of Reliance in Fraud*, 48 ARIZ. L. REV. 1001, 1009 (2006) (analogizing reliance in fraud cases to the role of proximate cause, because just as proximate cause requires that "the risks that are realized in the actual case are the risks that led us to characterize the defendant's conduct as negligent toward the victim," reliance in fraud determines "[w]hether the defendant has actually succeeded in harming the plaintiff by virtue of defrauding the plaintiff, as opposed to having harmed the plaintiff by deceiving others").

39. *See, e.g.*, Omnicare Inc. v. Laborers Dist. Council Constr. Indus. Pension Fund, 135 S. Ct. 1318, 1330 (holding that whether an opinion is misleading will depend on context, such as the custom and practices of the relevant industry).

40. The Second Restatement of Contracts defines a "mistake" as a "a belief that is not in accord with the facts." RESTATEMENT (SECOND) OF CONTRACTS § 151 (AM. LAW INST. 1981). If the mistake is unilateral, meaning it is a factual mistake of only one of the parties, the contract is voidable only if it is shown that the mistaken party did not bear the risk of the mistake or that the other party knew of, or caused, the mistake. *Id.* § 153. When the mistake is mutually made by all of the parties, the contract is voidable by the adversely affected party if the mistake is about a basic assumption underling the contract. *Id.* § 152 & cmt. b. Of course, if there is no meeting of the minds, there was never a contract formed. In all of these cases, the relief at common law is restitution, meaning the "reversal of any steps that the parties may have taken by way of performance, so that each party returns such benefit as he may have received," and in cases where this is not possible, damages. *Id.* § 158 & cmt. b.

41. Fraud, which generally requires proof of scienter, renders the transaction voidable, thus entitling the aggrieved party to restitution or rescission. *See* Eklund v. Koenig & Assocs., Inc., 451 N.W.2d 150, 153 (Wis. Ct. App. 1989) ("When a party discovers an alleged fraud . . . , he may affirm the contract and sue for damages, or he may disaffirm and seek restitution."); *see also* DANIEL B. DOBBS, LAW OF REMEDIES § 9.4, at 618 (1973) (stating that rescission and restitution are equitable remedies for fraud).

strictly false statements of fact due to information asymmetries inherent in certain types of transactions.[42]

Opinion statements include factual statements but are far more complex to evaluate for their liability. An opinion statement will often be based on one or more underlying fact(s),[43] but there is additional information being conveyed in an opinion statement. An opinion statement conveys not only that the speaker believes the facts underlying their opinion to be true, but also that they genuinely believe in their opinion, which is based on those facts.[44] In other words, an opinion statement contains not only factual information but information about the speaker's subjective belief in their stated judgment or decision-making process.[45]

In addition, opinions often convey information about the speaker's level of certainty about the facts and awareness of the facts.[46] A corporate executive

---

42. The most prominent examples are Sections 11 and 12 of the Securities Act of 1933, which provide for rescission or recessionary damages upon a showing that a material statement in an offering prospectus was false or misleading. *See* Securities Act Section 11, 15 U.S.C. § 77k (2012); Securities Act Section 12, 15 U.S.C. § 77l (2012). Section 11 provides for damages arising from a false statement in a registration statement, and Section 12 provides for rescission or recessionary damages. 15 U.S.C. §§ 77k(e) & 77l(a). There is no requirement that the false statement have been intentionally made. Askelson v. Freidus (*In re* Barclays Bank PLC Sec. Litig.), No. 17-3293-cv, 2018 U.S. App. LEXIS 32622, at *4 (2d Cir. Nov. 19, 2018). This is partly because of the information asymmetry that exists between issuer and the purchaser of the security. *See* William O. Douglas & George E. Bates, *The Federal Securities Act of 1933*, 43 YALE L.J. 171, 176 (1933) ("As stated above the protection given to investors by Section 11 fills a long felt need in so far as it shifts the burden of proof. This is particularly desirable during the early life of the security. At that time the registration statement will be an important conditioner of the market. Plaintiff may be wholly ignorant of anything in the statement. But if he buys in the open market at the time he may be as much affected by the concealed untruths or the omissions as if he had read and understood the registration statement. So it seems wholly desirable to create a presumption in favor of the investor in this regard.").

43. Liability may, however, attach if a statement of fact embedded in an opinion statement is materially false or misleading. *See* City of Dearborn Heights Act 345 Police & Fire Ret. Sys. v. Align Tech., Inc., 856 F.3d 605, 616 (9th Cir. 2017) ("[W]hen a plaintiff relies on a theory that a statement of fact contained within an opinion statement is materially misleading, the plaintiff must allege that 'the supporting fact [the speaker] supplied [is] untrue.' ") (quoting *Omnicare*, 135 S. Ct. at 1327).

44. WILLIAM LLOYD PROSSER ET AL., PROSSER AND KEETON ON THE LAW OF TORTS § 109, at 755 (5th ed. 1984) ("[A]n expression of opinion is itself always a statement of . . . the fact of the belief, the existing state of mind, of the one who asserts it."); *see also Omnicare*, 135 S. Ct. at 1327 (stating opinion with embedded statement of fact affirms both the underlying fact and the speaker's state of mind).

45. *See Omnicare*, 135 S. Ct. at 1327.

46. As comment a to Section 168 of the Second Restatement of Contracts explains, a statement of opinion "implies that [the speaker] does not have such definite information, that he is not certain enough of what he says, to make an assertion of his own knowledge as to that matter." RESTATEMENT (SECOND) OF CONTRACTS § 168 cmt. a (AM. LAW INST. 1981); *see also*

that says he "believes" that his company is in compliance with federal law is likely really making a probabilistic statement based on the information he has.[47] The addition of the word "believe" transforms what would otherwise be a purely factual statement into one conveying both uncertainty and some level of diligence.[48]

In many contexts, therefore, it will not be enough for liability if one or more factual predicate of an opinion statement is false. There will have to be something more, such as evidence that the opinion is disingenuous or some showing that the opinion statement was frivolous, that it lacked any reasonable basis, or that the speaker simply never bothered to look at the facts they would normally look at before rendering an opinion.[49] The important questions

---

Omnicare, 135 S. Ct. at 1329 ("Reasonable investors understand that opinions sometimes rest on a weighing of competing facts; indeed, the presence of such facts is one reason why an issuer may frame a statement as an opinion, thus conveying uncertainty.").

47. In *Omnicare*, management made statements in its registration statement to the effect that the company was in "compliance with applicable federal and state laws." *Omnicare*, 135 S. Ct. at 1323. Because this belief was not alleged to have been disingenuous—that is, there was no allegation that the company did not sincerely believe it was in compliance with applicable laws—there was no basis upon which to allege that such an opinion statement was false. *Id.* at 1327. The statement, however, may have omitted material information, but even then, the mere fact that some contradictory information existed would not be enough to render the opinion statement misleading, because "[a] reasonable investor does not expect that every fact known to an issuer supports its opinion statement." *Id.* at 1329. What may be implicitly conveyed by the opinion statement, however, is that there is some basis for the opinion, and in some cases, there are important facts that substantiate the opinion. If those facts are not provided, the opinion statement may mislead the listener. *Id.* at 1328. The opinion states:

> [A] reasonable investor may, depending on the circumstances, understand an opinion statement to convey facts about how the speaker has formed the opinion—or, otherwise put, about the speaker's basis for holding that view. And if the real facts are otherwise, but not provided, the opinion statement will mislead its audience.

Id. The common law rule, which the Restatement of Contracts articulates, is more stringent on the question of facts contradicting an opinion, as it presumes that an opinion statement's implicit indication of uncertainty carries with it the representation that the speaker is not aware of any facts contrary to the opinion. *See* RESTATEMENT (SECOND) OF CONTRACTS § 168 ("If it is reasonable to do so, the recipient of an assertion of a person's opinion as to facts not disclosed and not otherwise known to the recipient may properly interpret it as an assertion (a) that the facts known to that person are not incompatible with his opinion, or (b) that he knows facts sufficient to justify him in forming it."); *see also id.* cmt. a (noting that an opinion statement "implies at most that [the speaker] knows of no facts incompatible with the belief or that he knows of facts that justify him in holding it").

48. *See Omnicare*, 135 S. Ct. at 1334 ("The common law recognized that most listeners hear 'I believe,' 'in my estimation,' and other related phrases as disclaiming the assertion of a fact.").

49. *See, e.g.*, Twing v. Schott, 338 P.2d 839, 843 (Wyo. 1959) ("The words of defendant Schott that the sewage system was 'good' and either 'adequate' or 'sufficient' could not have

revolve around the intent and subjective state of mind of the speaker, and in some cases, the reasonableness of that state of mind.[50]

B. SCIENTER AND OPINION LIABILITY

Because opinions are not true or false simply because some underlying factual predicate for the opinion turns out to be true or false, the important question is often whether the speaker was being disingenuous when stating an opinion.[51] In many cases, the opinion may be disingenuous if evidence exists that the speaker believed something contrary to the opinion when they stated it.[52] For example, a doctor who tells a patient that the prognosis for a particular surgery is good, but contemporaneously sends an email to a colleague saying otherwise, may have been disingenuous when stating their opinion to the patient. The same may be true for an investment advisor that recommends an investment product while privately telling a colleague that the product is "junk."[53] In such cases, there is direct evidence that the opinion is not sincere,

---

been other than a fraudulent misrepresentation. We think his words constituted more than a puffing statement and more than any opinion. It was wholly inconsistent with the fact that he had repeatedly, according to his own admission, pumped out the cesspool. In that connection, it does not appear by testimony or otherwise that the purported dropping of rocks in the line would cause the cesspool to fill. It is true that the cesspool might have filled by reason of the use of excessive water by the tenants but if this were the fact then there would seem to be no excuse for failing to tell the prospective purchasers of the pumpings of the cesspool."). Some cases have reasoned that the opinion statement coupled with "half truths" creates a duty to disclose in full all contradictory information. *See, e.g.*, Mends v. Dykstra, 637 P.2d 502, 508 (Mont. 1981) (holding that representations about the condition of a house were misleading given undisclosed knowledge of defects and problems). A complete lack of basis will also give rise to liability, because the person hearing the opinion may conclude that the opinion is not the sort of statement someone would make based on an uninformed judgment. *See Omnicare*, 135 S. Ct. at 1330 ("Investors do not, and are right not to, expect opinions contained in those statements to reflect baseless, off-the-cuff judgments, of the kind that an individual might communicate in daily life.").

50. *See, e.g.*, *Omnicare*, 135 S. Ct. at 1334–35.

51. *See Omnicare*, 135 S. Ct. at 1328 ("[A] statement of opinion is not misleading just because external facts show the opinion to be incorrect.").

52. *See supra* note 49.

53. *See, e.g.*, Pursuit Partners, LLC v. UBS AG, No. X05CV084013452S, 2009 Conn. Super. LEXIS 2313, at *47 (Super. Ct. Sep. 8, 2009) ("The court takes [Defendant] employees at their word when they referenced their Notes, these purported 'investment grade' securities which they sold, as 'crap' and 'vomit', for [Defendant] alone possessed the knowledge of what their product, their inventory, was truly worth. While [Defendant] would argue that such descriptors lack a precise meaning, the true meaning of these words and the true value of [Defendant's] wares became abundantly clear when the Plaintiffs' multi-million dollar investment was completely wiped out and liquidated by [Defendant] shortly after the last of the Note purchases was consummated.").

and it is reasonable to infer that the speaker had some improper motive for stating the opinion.[54]

This sort of opinion is in a sense a false statement because the implicit representation that the opinion is genuine is false,[55] and courts have no trouble assigning liability in such cases. In fact, many courts have required some evidence that the opinion was not genuinely held when stated to assign liability.[56] Indeed, in the securities law context, courts sometimes require a showing of scienter, even when the underlying cause of action imposes strict liability for false or misleading statements. For example, under the Securities Act of 1933, a false statement in a prospectus will give rise to rescission or damages, essentially allowing the purchaser to unwind a securities transaction premised on materially false factual statements in a prospectus.[57] There is no scienter requirement in the statute, but courts have required that the statement not only be proven objectively false, but also subjectively disbelieved by the issuer when the statement was made in the prospectus—in other words, opinion statements must be both subjectively and objectively false for liability to attach.[58]

Requiring scienter solves many of the problems with opinion liability—namely, the nearly intractable problem of having to prove that the speaker's judgment was not only incorrect but should have been better.[59] In other words, the liability question would require a showing that the speaker's judgment was somehow improper, and the clearest scenario where this is the case is where

54. *See id.*

55. *See supra* notes 44–45 and accompanying text.

56. *See, e.g.*, *id.*

57. *See supra* note 42.

58. *See* Fed. Hous. Fin. Agency v. UBS Ams., Inc., 858 F. Supp. 2d 306, 325 (S.D.N.Y. 2012) ("[P]laintiff must assert that the statement upon which it seeks to predicate liability 'was both objectively false and disbelieved by the defendant at the time it was expressed.' ") (quoting Fait v. Regions Fin. Corp., 655 F.3d 105, 110 (2d Cir. 2011)); *see also Omnicare*, 135 S. Ct. at 1327 ("What the Funds instead claim is that Omnicare's belief turned out to be wrong—that whatever the company thought, it was in fact violating anti-kickback laws. But that allegation alone will not give rise to liability under § 11's first clause because, as we have shown, a sincere statement of pure opinion is not an 'untrue statement of material fact,' regardless whether an investor can ultimately prove the belief wrong. That clause, limited as it is to factual statements, does not allow investors to second-guess inherently subjective and uncertain assessments. In other words, the provision is not, as the Court of Appeals and the Funds would have it, an invitation to Monday morning quarterback an issuer's opinions.").

59. By eliminating a cause of action based on a hindsight evaluation of a subjective judgment—what the *Omnicare* Court referred to as "Monday morning quarterback[ing]"—courts and fact finders do not need to decide whether they would have reached the same opinion given the set of facts as they were known or knowable when the opinion statement was made, nor do they have to determine in most cases whether the opinion was reasonable. *See Omnicare*, 135 S. Ct. at 1326.

there is evidence that the opinion was inconsistent with the speaker's own beliefs.[60] In those cases, it is fair to presume that the person hearing the opinion statement is entitled to at least the speaker's genuine opinion on the matter, which they did not receive.

That does not mean that it is the only sort of opinion that is problematic. In fact, there are a host of opinions that are plausible but flawed on their merits.[61] Heightening the standard for opinion liability essentially excludes these cases because so long as an opinion is plausible and there is no evidence that the speaker disbelieved the opinion, there is no liability.[62] This standard creates two significant problems: First, it excludes from liability the scenario where the opinion is facially plausible but the speaker rendered it with inadequate investigation into the relevant facts.[63] Second, it excludes from liability the sort of case where the opinion is rendered in the face of contradictory information that the person hearing the opinion would have wanted to know.[64] In most cases, there will not likely be clear evidence that the speaker disbelieves the opinion, and a strict scienter-based legal standard will not allow any way to further test the opinion statement for error or incompetence.[65]

C. EXAGGERATED OPINIONS AND PUFFERY

In some cases, the context of the opinion statement may not justify an assumption that the opinion statement is grounded in supporting facts. It may be that the opinion is too general to be verifiably true or false, that the speaker's motive is such that one expects an exaggerated opinion, or both. The most

60. *See supra* note 49 and accompanying text.

61. In particular, an opinion may be based on misinterpretations of a set of underlying facts, based on dubious reasoning, or generally poorly thought out. These sort of opinion statements will not, without more, be actionable as misrepresentations.

62. *See, e.g.*, SEPTA v. Orrstown Fin. Servs., Inc., No. 1:12-cv-00993, 2015 U.S. Dist. LEXIS 80584, at *98 (M.D. Pa. June 22, 2015) (dismissing claim where "Plaintiff has failed to point to a factual basis supporting its allegation that Defendant SEK did not believe its opinion" about financial statements).

63. A merely negligently-rendered opinion will not give rise to liability if a scienter-heuristic is used exclusively for liability. What is required is a completely unreasonable or inadequate basis for an opinion, such that the opinion is merely an unadorned, bald conclusion that lacks any support. In such a case, the baseless opinion may be sufficiently reckless to give rise to an inference of scienter. *Cf. Omnicare*, 135 S. Ct. at 1330.

64. The *Omnicare* opinion-liability framework mitigates this problem by allowing the basis of an opinion to be examined where the claim being considered is based on omitted facts from an opinion statement rather than based on the claim that an affirmatively-stated opinion was false or misleading opinion. *See id.*

65. *See, e.g.*, *Omnicare*, 135 S. Ct. at 1326.

common example is the salesperson who exaggerates the value of their wares.[66] Most opinion-liability regimes assume that statements by a salesperson are often exaggerated and that those who hear such statements take them with a grain of salt.[67]

This is the rationale for the doctrine of puffery—which states that "an optimistic statement that is so vague, broad, and non-specific that a reasonable investor would not rely on it" is "immaterial as a matter of law."[68] Such statements by a seller are usually presumed by the buyer to be overstated, exaggerated, or impossible to prove true or false.[69] The legal heuristics at work in this context are reliance and materiality, as the buyer would not be reasonable to rely on vague and overblown statements that salesmen are known to make and those statements are likely to be immaterial anyway.[70]

What the buyer can often reasonably assume, however, is that the seller's overblown statements are not being made blatantly in the face of facts contrary to the opinion—that is, the opinion statement is "not fantastical."[71] In other words, the speaker is likely representing that they are not aware of any facts contrary to their statement of opinion. That does not necessarily mean that

66. *See* RESTATEMENT (SECOND) OF TORTS § 539 cmt. c (AM. LAW INST. 1977) ("The habit of vendors to exaggerate the advantages of the bargain that they are offering to make is a well recognized fact.").

67. This assumption is quite old, as it has been articulated in some of the earliest misrepresentation cases in the United States. *See, e.g.*, Kimball v. Bangs, 11 N.E. 113, 114 (Mass. 1887) ("The law recognizes the fact that men will naturally overstate the value and qualities of the articles which they have to sell. All men know this, and a buyer has no right to rely upon such statements."). As Judge Learned Hand has explained, it is presumed that there are statements that "no sensible man takes seriously, and if he does he suffers from his credulity. If we were all scrupulously honest, it would not be so; but, as it is, neither party usually believes what the seller says about his own opinions, and each knows it." Vulcan Metals Co. v. Simmons Mfg. Co., 248 F. 853, 856 (2d Cir. 1918).

68. *In re* Gen. Elec. Co. Sec. Litig., 857 F. Supp. 2d 367, 384 (S.D.N.Y. 2012); *see also In re* Vivendi, S.A. Sec. Litig., 838 F.3d 223, 245 (2d Cir. 2016) ("Puffery encompasses statements [that] are too general to cause a reasonable investor to rely upon them, and thus cannot have misled a reasonable investor.") (internal citations and quotation marks omitted).

69. The Second Restatement of Contracts expressly assumes this about representations by sellers. RESTATEMENT (SECOND) OF CONTRACTS § 169 (AM. LAW INST. 1981) ("It may be assumed, for example, that a seller will express a favorable opinion concerning what he has to sell. When he praises it in general terms, commonly known as 'puffing' or 'sales talk,' without specific content or reference to facts, buyers are expected to understand that they are not entitled to rely.").

70. *See supra* note 68.

71. *See* RESTATEMENT (SECOND) OF TORTS § 539 ("However, a purchaser is justified in assuming that even his vendor's opinion has some basis of fact, and therefore in believing that the vendor knows of nothing which makes his opinion fantastic.").

they believe that the facts underlying their opinion are objectively true.[72] Thus, in this context, the operative question is again the state of mind of the speaker—namely, their knowledge at the time the statement is made.

In many cases, however, puffery will simply not be actionable because such statements are likely to be too vague and general to evaluate, meaning they are not provably true or false.[73] In such cases, reliance, materiality, and intent are again the principal heuristics at work. A salesperson or seller's puffery cannot be reasonably relied on and therefore could not have been material,[74] and the statements may be too vague to have been intended as stating any facts, even about their state of mind.[75] If the seemingly exaggerated opinion has some specificity, then some showing that the speaker was aware of information contrary to his opinion will be required for liability.[76]

D. OMISSIONS AND INFORMATION ASYMMETRY

The more difficult case arises when there is information asymmetry and important—and perhaps contradictory—information is omitted from the opinion.[77] The clearest cases are when the speaker is—or is held out as—an

---

72. *See id.* § 168 ("If it is reasonable to do so, the recipient of an assertion of a person's opinion as to facts not disclosed and not otherwise known to the recipient may property interpret it as an assertion (a) that the facts known to that person are not incompatible with his opinion, or (b) that he knows facts sufficient to justify him in forming it."); *see also id.* § 539 cmt. a ("Frequently a statement which, though in form an opinion upon facts not disclosed or otherwise known to their recipient, is reasonably understood as implying that there are facts that justify the opinion or at least that there are no facts that are incompatible with it.").

73. *See In re* PDI Sec. Litig., Civil Action No. 02-cv-0211 (JLL), 2005 U.S. Dist. LEXIS 18145, at *69 (D.N.J. Aug. 16, 2005) ("Vague and general statements of optimism 'constitute no more than puffery and are understood by reasonable investors as such.' ") (citation omitted); *see also* Stefan J. Padfield, *Is Puffery Material to Investors? Maybe We Should Ask Them*, 10 U. PA. J. BUS. & EMP. L. 339, 352 (2008) ("Puffery and statements of fact are mutually exclusive. If a statement is a specific, measurable claim or can be reasonably interpreted as being a factual claim, *i.e.*, one capable of verification, the statement is one of fact. Conversely, if the statement is not specific and measurable, and cannot be reasonably interpreted as providing a benchmark by which the veracity of the statement can be ascertained, the statement constitutes puffery.") (internal citation omitted).

74. *See, e.g.*, *In re* Advanta Corp. Sec. Litig., 180 F.3d 525, 538 (3d Cir. 1999) ("Such statements, even if arguably misleading, do not give rise to a federal securities claim because they are not material.").

75. *See* State v. Am. TV & Appliance of Madison, Inc., 430 N.W.2d 709 (Wis. 1988) ("[E]xaggerations [are] reasonably to be expected of a seller as to the degree of quality of his product, the truth or falsity of which cannot be precisely determined.") (internal quotation marks and citation omitted).

76. *See supra* Section I.B.

77. For an omission of fact to be actionable, there must generally be some duty to disclose information, for example, because of a fiduciary relationship or a relationship of trust and confidence between the parties. *See, e.g.*, Chiarella v. United States, 100 U.S. 1108, 1115

expert on the subject of the opinion.[78] In such a case, the listener may not know even what facts are most important for a sound or justified opinion.[79] In other words, the listener may not only be relying on the judgment of the expert, but also the expert's judgment as to what information is most important.[80]

One of the clearest examples is the doctor-patient context.[81] In many cases, the lay patient can look at the same MRI results as the doctor but would not know what aspects of the results are significant for a diagnosis. The doctor, on the other hand, relies on education and experience to determine what aspects of the MRI results are most important.[82] The doctor not only has an

---

(1980) (noting that "silence in connection with the purchase or sale of securities may operate as a fraud actionable under § 10(b)" when there is "a duty to disclose arising from a relationship of trust and confidence between parties to a transaction"). Even without an affirmative duty to speak, a duty to disclose material information may arise because there may be a duty to speak fully and truthfully once a person has spoken. *See, e.g.*, Helwig v. Vencor, Inc., 251 F.3d 540, 561 (6th Cir. 2001) ("[E]ven absent a duty to speak, a party who discloses material facts in connection with securities transactions 'assumes a duty to speak fully and truthfully on those subjects.' ") (citation omitted).

78. RESTATEMENT (SECOND) OF TORTS § 539 cmt. b (AM. LAW. INST. 1979). A statement of opinion

> may also reasonably be understood to imply that [the speaker] does know facts sufficient to justify him in forming the opinion and that the facts known to him do justify him. This is true particularly when the maker is understood to have special knowledge of facts unknown to the recipient.

Id.

79. *See id.*

80. *See* Omnicare, Inc. v. Laborers Dist. Council Constr. Indus. Pension Fund, 135 S. Ct. 1318, 1335 (Scalia, J., concurring) ("[What] [the reasonable (female) person, and even he, the reasonable (male) person] would naturally understand a statement [of opinion] to convey is not that the statement has the foundation she (the reasonable female person) considers adequate. She is not an expert, and is relying on the advice of an expert—who ought to know how much 'foundation' is needed. She would naturally understand that the expert has conducted an investigation that he (or she or it) considered adequate. That is what relying upon the opinion of an expert means.") (brackets and alterations in the original, quotations omitted).

81. *See id.* at 1334 (holding that the common law recognizes that "expressions of opinion made in the context of a relationship of trust, such as between doctors and patients" may give rise to opinion liability based on the basis of the opinion).

82. It is because an expert, such as a doctor, typically relies on his experience and judgment in reviewing facts underlying his opinion that the Advisory Committee amended Federal Rule of Evidence 703 to allow the admission of the expert's testimony about the underlying facts in certain cases without having to admit those facts individually at trial as out of court statements subject to the hearsay rule. Indeed, the Advisory Committee Notes to Rule 703 mention X-rays as examples of such evidence, which doctors apply their expertise to as a matter of course. *See* FED. R. EVID. 703 advisory committee's note ("Thus a physician in his own practice bases his diagnosis on information from numerous sources and of considerable variety, including statements by patients and relatives, reports and opinions from nurses, technicians and other doctors, hospital records, and X-rays. Most of them are admissible in

informational vantage that is superior to the patient because of their experience, but also a judgment and intuition advantage over the patient. When the doctor provides a diagnosis, they do not simply convey information about the underlying facts or even merely about the diagnosis or medical outcome, but may also be implicitly representing that they made a reasonable inquiry into the facts and correctly weighed the facts, including the facts contrary to his opinion.[83]

It is precisely when the underlying facts are contradictory, indeterminate, or incomplete that the opinion of an expert is most valuable. The information conveyed in such an opinion is more than factual—it's a conclusion about the facts that is inextricably bound up with the speaker's experience, intuition, and judgment.[84] And, in the case of an expert, the opinion invites reliance, particularly if the expert holds themselves out as a disinterested party.[85]

In the expert context, it will not be enough for liability that a fact underlying the opinion is, or turns out to be, false. In fact, it is expected that

---

evidence, but only with the expenditure of substantial time in producing and examining various authenticating witnesses. The physician makes life-and-death decisions in reliance upon them. His validation, expertly performed and subject to cross-examination, ought to suffice for judicial purposes.").

83. Some of the earliest applications of this sort of expertise-asymmetry rationale appeared in cases involving physicians. *See, e.g.*, Hedin v. Minneapolis Med. & Surgical Inst. 64 N.W. 158, 160 (Minn. 1895) ("The doctor, especially trained in the art of healing, having superior learning and knowledge, assured plaintiff that he could be restored to health. That the plaintiff believed him is easily imagined; for a much stronger and more learned man would have readily believed the same thing. The doctor, with his skill and ability, should be able to approximate to the truth when giving his opinion as to what can be done with injuries of one year's standing, and he should always be able to speak with certainty before he undertakes to assert positively that a cure can be effected. If he cannot speak with certainty, let him express a doubt. If he speaks without any knowledge of the truth or falsity of a statement that he can cure, and does not believe the statement true, or if he has no knowledge of the truth or falsity of such a statement, but represents it as true of his own knowledge, it is to be inferred that he intended to deceive.").

84. *See* RESTATEMENT (SECOND) OF TORTS § 542 cmt. f (AM. LAW. INST. 1979) ("The complexities and specializations of modern commercial and financial life have created many situations in which special experience and training are necessary to the formation of a valuable judgment. In this case if the one party has special experience or training or purports to have them, the other, if without them, is entitled to rely upon the honesty of the former's opinion and to attach to it the importance that is warranted by his superior competence.").

85. *See id.* § 542 cmt. h ("One who has taken steps to induce another to believe that the other can safely trust to his judgment is subject to liability if the confidence so acquired is abused. This is true not only when the maker of the fraudulent misrepresentation of opinion is or professes to be disinterested, as when the transaction is between the recipient and a third person, as to which see § 543, but also when he is known to have an adverse interest in the transaction.").

there are contradictory or inconsistent facts underlying the opinion.[86] Indeed, if the expert could rely on only deterministic facts, there would be no room for their judgment.[87] There are, however, certain facts that any person, even one who relies on an expert, would want to know about. If a doctor states that an ailment appears benign based on their judgment and experience, but considered and disregarded a possible diagnosis that is likely terminal if not immediately treated, the patient would likely want to know about the disregarded diagnosis—the risk of harm is high, and the underlying information is time sensitive.[88] The patient would use that information to, perhaps, obtain a second opinion or at the least, to consciously decide to what extent they want to rely solely on the doctor's opinion.[89]

Even when the speaker is not an expert, there are contexts where the information asymmetry is so great that it is fair to assume that the speaker is better positioned not only to know all of the relevant facts but also how to weigh those facts. An officer of a public company is generally not free to share internal information about the company outside of a public filing with the SEC.[90] This results in a scarcity of information about the corporation between periodic filings, such as quarterly reports. The officer, however, presumably receives information in real time. Moreover, by virtue of his management position, he is aware of what information is most important to the operations and profitability of the company.[91] When the executive ultimately provides

86. *See supra* note 51.

87. An opinion based on determinative facts of obvious weight is not an opinion at all because there is no uncertainty about the facts to express. Such an opinion is likely simply nothing more than a set of factual statements.

88. *See* Arato v. Avedon, 858 P.2d 598, 607 (Cal. 1993) ("Rather than mandate the disclosure of specific information as a matter of law, the better rule is to instruct the jury that a physician is under a legal duty to disclose to the patient all material information—that is, 'information which the physician knows or should know would be regarded as significant by a reasonable person in the patient's position when deciding to accept or reject a recommended medical procedure'—needed to make an informed decision regarding a proposed treatment.").

89. *See id.*

90. Although a reporting company must in some cases file interim reports concerning material corporate events, most internal information about a public corporation is in practice withheld until the next quarterly report. *See* 17 C.F.R. § 240.13a-11. There are other rules that prevent real-time disclosure of information, which creates information asymmetries. For a detailed discussion about the law surrounding the disclosure of corporate information, including under Regulation FD, see generally M. Todd Henderson & Kevin S. Haeberle, *Information-Dissemination Law: The Regulation of How Market-Moving Information Is Revealed*, CORNELL L. REV. 1373 (2016).

91. A Justice Scalia noted in his concurrence in *Omnicare*, it is reasonable to assume that corporate executives have expertise concerning the finances of the companies they run, including about corporate and financial information that must be set forth in an offering document or registration statement. *See* Omnicare, Inc. v. Laborers Dist. Council Constr.

information through public filings, what is reported implicitly carries with it not only the representation that what is reported is accurate, but also that what is reported is pertinent.[92] In this context, the corporate officer is similarly regarded as an expert. If the corporate officer makes representations about asset valuations based on a universe of factual inputs, the failure to state contradictory facts alongside the valuation opinion may be misleading, depending on the importance and weight of the omitted fact.[93]

Thus, generally in asymmetric information contexts, and specifically in expert opinions, the opinion's veracity may be sensitive to omitted information. Although the information asymmetry requires more reliance on the speaker in these contexts, that reliance also makes those who hear the opinion vulnerable to a form of blindness—they cannot see around the opaque corners that are likely transparent to the speaker. If the information the speaker relies on is outcome determinative or immensely important to the opinion, its disclosure may be as important as the conclusion communicated in the opinion.

E. OPINION STATEMENTS AS MODELS

If opinion statements are more than the facts underlying them, then what exactly are they? One way to think of an opinion is as a model of reality that is based on an individual's judgment and a universe of facts. The information

---

Indus. Pension Fund, 135 S. Ct. 1318, 1335 (Scalia, J., concurring) ("It is reasonable enough to adopt such a presumption for those matters that are required to be set forth in a registration statement. Those are matters on which the management of a corporation are experts. If, for example, the registration statement said 'we believe that the corporation has $5,000,000 cash on hand,' or 'we believe the corporation has 7,500 shares of common stock outstanding,' the public is entitled to assume that the management has done the necessary research, so that the asserted 'belief' is undoubtedly correct.").

92. *Cf. id.* The SEC generally requires management to provide a discussion and analysis of a public company's financial condition in order to "enable . . . investors to see the company through the eyes of management," meaning that management must provide the information and form of information that it deems important as it manages the company. Commission Guidance Regarding Management's Discussion and Analysis of Financial Condition and Results of Operation, Release No. 33-8350 (Dec. 29, 2003).

93. Whether an omission is material will depend in most misrepresentation cases on the context surrounding the statement that contained the omission. *See, e.g.*, *Omnicare*, 135 S. Ct. at 1330 ("[A]n investor reads each statement within such a document, whether of fact or of opinion, in light of all its surrounding text, including hedges, disclaimers, and apparently conflicting information. And the investor takes into account the customs and practices of the relevant industry. So an omission that renders misleading a statement of opinion when viewed in a vacuum may not do so once that statement is considered, as is appropriate, in a broader frame. The reasonable investor understands a statement of opinion in its full context, and § 11 creates liability only for the omission of material facts that cannot be squared with such a fair reading.").

available may be incomplete, the facts may be wrong, and in some cases, the facts may cut different ways or be subject to diverging interpretations.[94] An opinion makes sense of the universe of facts, assigns interpretation and weight to those facts, and maps the universe of facts to a conclusion, decision, or outcome.[95]

In the case of a trained expert, the opinion model is not only based on a universe of facts, but also on their experience. That is, the person holding the opinion not only makes sense of the universe of data points available to them, but also squares those data points with what they have seen in the past. When the expert has specialized training, a certain standard set of data points and background information is attributable to the expert. For example, a trained lawyer is deemed to have exposure to essential building blocks from contract and tort law and is generally imputed with a basic understanding of constitutional norms. Any opinion they render is against the backdrop of both their training and experience. All of the data from training, experience, and fact-gathering are combined together to form an opinion. Thus, an opinion can be thought of as a model of reality. It is a collection of facts, interpretations, weights, and probabilistic assessments.

Indeed, opinions are in some ways similar to mathematical and statistical models, which often seek to replicate the behavior of a particular aspect of reality in order to make predictions.[96] A useful analogy is a crude least-squares

---

94. *See supra* note 2.

95. Opinions are in some ways analogous to scientific theories, as both are built on some set of facts or axioms assumed to be true and some derived implications from those facts or axioms. The difference, of course, is that a scientific theory is only as good as its predictive power, and if its predictions can be proven incorrect, meaning they are falsifiable, the theory itself can be proven false. KARL POPPER, THE LOGIC OF SCIENTIFIC DISCOVERY 10. Popper states:

> Next we seek a decision as regards these (and other) derived statements by comparing them with the results of practical applications and experiments. If this decision is positive, that is, if the singular conclusions turn out to be acceptable, or verified, then the theory has, for the time being, passed its test: we have found no reason to discard it. But if the decision is negative, or in other words, if the conclusions have been falsified, then their falsification also falsifies the theory from which they were logically deduced.

Id. Human opinions are evaluated for liability purposes, so the question is not whether the opinion is universally correct, but rather whether the opinion was justified under the circumstances. As explained *infra* Part III, however, AI opinions are closer to scientific theories, in that they can be tested for accuracy before being deployed.

96. *See* TIMOTHY GOWERS, MATHEMATICS: A VERY SHORT INTRODUCTION 4 (2002) ("Mathematics do not apply scientific theories directly to the world but rather to models. A model in this sense can be thought of as an imaginary, simplified version of the part of the world being studied, one in which exact calculations are possible.").

regression.[97] It models what could be a noisy and scattered set of data points with a line. This line is a blunt instrument but can be useful to get a sense of correlations in the data.[98] In most cases, virtually none of the data points will fit the modeled line, meaning that in a sense they are contradictory to the simplistic line created to describe the data—but divergent data points do not make the model "false" simply because they do not fit neatly on the regression line.[99]

The model may still be useful for a crude estimate. It is more than the points that were used to create it. It is a reduction of the facts, and its value depends entirely on what it is used for. Sometimes a regression line is useful to make general predictions about a population—for example, age and height will correlate up until a certain age. If you're using a height and age model to predict the height of elementary school students, it may be perfectly useful, but if you use the same model across a population that includes adults, the model is plainly insufficient and will be grossly inaccurate in many cases. Opinion statements are just as vulnerable to context. In the proper context, even a weak basis for an opinion may be sufficient.[100] That same basis in another context may be misleading.

In the case of statistical and mathematical models, some data points are so out of step with the entire data set that they are considered outliers.[101] How a

---

97. A least-squares regression is a simple mathematical model of data that attempts to fit a line to a set of data by minimizing the square of the error resulting from the fitted line's predictions. *See generally* WILLIAM MENDENHALL, III ET AL., INTRODUCTION TO PROBABILITY AND STATISTICS 482–529 (14th ed. 2013).

98. Regressions are often too simple to be used to study complex datasets but are frequently used as a starting point because of their simplicity. JEFFREY M. WOOLDRIDGE, INTRODUCTORY ECONOMETRICS: A MODERN APPROACH 21 (6th ed. 2015) ("Although simple regression is not widely used in applied econometrics, it is used occasionally and serves as a natural starting point because the algebra and interpretations are relatively straightforward.").

99. Some divergent points in a linear model will significantly skew the fitted line. Such divergent or influential data points are sometimes discarded as "outliers." *See id.* at 326–27 ("Loosely speaking, an observation is an influential observation if dropping it from the analysis changed the key LS estimates by a practically 'large' amount."). Ordinary Least Squares models are sensitive to outliers because the process minimizes the squares of errors or residuals, thus compounding the importance of large prediction errors. *See id.* at 327. ("OLS is susceptible to outlying observations because it minimizes the sum of squared residuals: large residuals (positive or negative) receive a lot of weight in the least squares minimization problem. If the estimates change by a practically large amount when we slightly modify our sample, we should be concerned.").

100. For example, an opinion provided during an emergency or under time constraints may be adequate even though a reasonable person would under normal circumstances undertake a more detailed inquiry into the matter.

101. *See supra* note 99.

model deals with an outlier is important and will sometimes require disclosure for someone using the model to fully understand the power and effectiveness of the model.

The same may be true for an opinion statement. Some facts may be so contradictory to the opinion that the omission of the fact may render the model misleading. The same may also be true if the opinion is based on incomplete or potentially inaccurate data. Some information may be unknown or unknowable. If the omitted or missing information can affect the efficacy of the opinion's model of reality, that is when disclosure may be important, and that may also be when failure to disclose that information should give rise to liability.[102]

There is an important attribute of most models, including both opinion models created by human beings and determinative algorithms (such as a statistical regression), that is important for the purposes of this Article: one will generally be able to query the person making the model or examine the deterministic algorithm upon which the model is based to determine how factors were weighted, what facts were considered, and the effect omitted information may have had on the overall opinion calculus. A person can be placed under oath and put on the stand, and his intent can be discerned by a factfinder using long-tested legal constructs and heuristics.[103] In the case of a deterministic algorithm, the algorithm itself can be examined by experts or even directly by factfinders. So, there is usually at least some minimal modicum of transparency.

There are, to be sure, instances even in the case of human experts and deterministic algorithms where transparency will be greatly diminished. The most obvious example is when facts have been interpreted using the judgment,

---

102. This is the rationale courts have applied when an opinion is based on uncertain facts, but the speaker fails to say so. *See, e.g.*, Hedin v. Minneapolis Med. & Surgical Inst., 64 N.W. 158, 160 ("The doctor, with his skill and ability, should be able to approximate to the truth when giving his opinion as to what can be done with injuries of one year's standing, and he should always be able to speak with certainty before he undertakes to assert positively that a cure can be effected. If he cannot speak with certainty, let him express a doubt.").

103. For example, in a recent trial resulting in a criminal conviction for "spoofing," the practice of using a computer program to rapidly place and cancel orders for securities to move a market, one of the most critical pieces of evidence at trial was the testimony of the computer program's designer about what the trader instructed him to create and what the computer program was designed to do. *See* United States v. Coscia, 866 F.3d 782, 789 (7th Cir. 2017) ("The designer of the programs, Jeremiah Park, testified that Mr. Coscia asked that the programs act '[l]ike a decoy, which would be '[u]sed to pump [the] market.' Park interpreted this direction as a desire to 'get a reaction from the other algorithms.' In particular, he noted that the large-volume orders *were designed specifically to avoid being filled* . . . .") (emphasis added).

experience, and intuition of an expert.[104] In those cases, it is difficult to determine how the underlying opinion model works. One cannot usually describe the vast degrees of freedom upon which a doctor with twenty years of medical experience bases their intuition.[105] But even in these cases, person's intent and set of motives can be examined. An expert with a motive to deceive will receive far less credit for his judgment than one without.[106] Moreover, experts may be judged against a standard of care reflecting their expertise, as they are in negligence cases,[107] which establishes a baseline of acceptable or

104. This opacity arises frequently when experts are called to testify in court about a technical subject. Commentators have questioned whether factfinders, such as judges and juries, are epistemically competent to hear such evidence, particularly given the tendency to rely on credibility heuristics when the substance of an expert's testimony is not accessible or understandable to a lay factfinder. *See, e.g.*, James R. Steiner-Dillon, *Expertise on Trial*, 19 COLUM. SCI. & TECH. L. REV. 247, 278 (2018) ("On the other hand, good intentions and genuine effort cannot create epistemic competence in the absence of substantive expertise. Jurors often fail to understand and apply scientific testimony correctly, even when the underlying science itself is relatively clear. They also tend to rely on specious proxies for substantive expertise."); *see also* Jennifer L. Mnookin, *Expert Evidence, Partisanship and Epistemic Competence*, 73 BROOK. L. REV. 1009, 1014 (2008) ("But if the jury lacks the knowledge that the expert provides, how, then, can it rationally evaluate the expertise on offer? To be sure, one might not need to be an expert in order to assess expertise, but the main mechanisms for assessing expertise outside of one's domain of knowledge are, by necessity, secondary indicia, proxies: demeanor, perhaps, or credentials, or superficial explanatory plausibility.").

105. *Cf.* Learned Hand, *Historical and Practical Considerations Regarding Expert Testimony*, 15 HARV. L. REV. 40, 54–55 (1901) ("The trouble with all this is that it is setting the jury to decide, where doctors disagree. The whole object of the expert is to tell the jury, not facts, as we have seen, but general truths derived from his specialized experience. But how can the jury judge between two statements each founded upon an experience confessedly foreign in kind to their own? It is just because they are incompetent for such a task that the expert is necessary at all. . . . What hope have the jury, or any other layman, of a rational decision between two such conflicting statements each based upon such experience.").

106. Again, in the context of testifying experts, an expert's motive for testifying—in many cases a fee—becomes an important proxy for credibility. *See* Mnookin, *supra* note 104, at 1014 ("Because the jury does not have the expertise to evaluate the substance of expert testimony, it is unlikely that it will be an accurate evaluator of partisan bias. . . . Without epistemic competence, the jury has no choice but to rely on proxies as secondary indicia of bias, and these may often be either inaccurate or difficult to evaluate.").

107. In negligence cases, the standard of reasonable care for one with expertise reflects his elevated capacity. *See* RESTATEMENT (THIRD) OF TORTS § 12 (AM. LAW INST. 2010) ("If an actor has skills or knowledge that exceed those possessed by most others, these skills or knowledge are circumstances to be taken into account in determining whether the actor has behaved as a reasonably careful person."); *see also* Omri Ben-Shahrar & Ariel Porat, *Personalizing Negligence Law*, 91 N.Y.U. L. REV. 627, 641 (2016) ("Defendant's special skills are most often taken into account in cases where the defendant's profession is relevant to the injury. For example, doctors are held to a standard of care for their patients that is considerably higher than the reasonable person standard.").

valid models of reality that an expert may operate within.[108] This is why scienter and basic heuristics continue to function in these settings, even when there is some opacity resulting from the application of human experience, judgment, and intuition.

As explained in the next Part, AI models are different. They in many cases risk creating the opacity of an expert's intuition and judgment, but without the ability to examine a motive, standard of care, or set of human biases.[109] And, because they are generally not based on deterministic instructions, there are no clear instructions that can be used as a proxy for the intent of the AI's creator or user.

III. ARTIFICIAL INTELLIGENCE, THE BLACK BOX PROBLEM, AND OPINION STATEMENTS

A. WHAT IS ARTIFICIAL INTELLIGENCE?

The term artificial intelligence generally refers to a class of computer programs capable of solving problems requiring inferential reasoning, decision making based on incomplete or uncertain information, classification, optimization, and perception.[110] AI can be based on determinative algorithms, such as a brute-force search,[111] or on machine-learning algorithms that learn directly from training examples.[112] The recent and rapid advances in AI have come mostly from the second category of AI—those built on machine-learning algorithms that learn from data,[113] such as deep networks of artificial neurons.

---

108. *Cf.* Ben-Shahrar et al., *supra* note 107, at 643 ("We saw that doctors are generally required to provide care that is at least as good as the average qualified medical practitioner, perhaps adjusted upwards to account for personal expertise.").

109. *See infra* Section III.B.

110. *See* Bathaee, *supra* note 7, at 898.

111. An example of such a brute force algorithm would be a computer program that searches the space of possible chess moves to determine which move to make next using some deterministic scoring or ranking criteria. *See* Dave Gershgorn, *Artificial Intelligence Is Taking Computer Chess Beyond Brute Force*, POPULAR SCI. (Sept. 16, 2015), http://www.popsci.com/artificial-intelligence-takes-chess-beyond-brute-force [https://perma.cc/PE5F-TSBE].

112. *See id.*

113. *See* IAN GOODFELLOW ET AL., DEEP LEARNING 2 (2016) ("Several artificial intelligence projects have sought to hard-code knowledge about the world in formal languages. A computer can reason automatically about statements in these formal languages using logical inference rules. This is known as the knowledge base approach to artificial intelligence.") (emphasis omitted). This approach of hard coding deterministic rules has given way to more powerful techniques that allow AI programs to learn directly from example and to make decisions based on a trained model's intuition. *See id.* at 1–2; *see also* Bathaee, *supra* note 7, at 898 ("On the most flexible end are modern AI programs that are based on machine-learning

Artificial neural networks are akin to neurons in the human brain, but they are not designed to mimic the function of biological neurons.[114] Rather, they are mathematical models—linear transformations often coupled with non-linear activation functions.[115] When combined into complex networks, they are capable of a form of cognition.[116] AI systems built on so-called "deep" architectures—stacked layers of artificial neurons—have been capable of performing tasks that most computers have been unable to perform at human-level proficiency.[117] In some applications, such as in the case of computer vision, these models exceed the proficiency of humans.[118]

AI programs may contain one or more of these underlying machine-learning algorithms.[119] Deep reinforcement learning systems, for example, use networks of artificial neurons to estimate future rewards when selecting from

---

algorithms that can learn from data. Such AI would, in contrast to the rule-based AI, examine countless other chess games and dynamically find patterns that it then uses to make moves.").

114. *See* Bathaee, *supra* note 7, at 901 ("The deep neural network is based on a mathematical model called the artificial neuron. While originally based on a simplistic model of the neurons in human and animal brains, the artificial neuron is not meant to be a computer-based simulation of a biological neuron. Instead, the goal of the artificial neuron is to achieve the same ability to learn from experience as with the biological neuron.").

115. An artificial neuron is typically structured as a linear combination of parameters and weights. *See* GOODFELLOW ET AL., *supra* note 113, at 192. The output of that linear combination is then passed to a non-linearity, or activation function, which broadcasts or squelches the neuron's output signal depending on the activation function's criteria. The activation functions provide necessary non-linearity to the model—otherwise, a series of linear transformations will generally only be able to approximate linear patterns, and there would be little additional power that would result from deepening a network of artificial neurons. *Id.* at 192. By adding a non-linearity, it is posited that a deep neural network can approximate important classes of non-linear functions in finite-dimensional space. *See id.* at 194 ("Specifically, the universal approximation theorem . . . states that a feedforward network with a linear output layer and at least one hidden layer with any 'squashing' activation function (such as the logistic sigmoid activation function) can approximate any Borel measurable function from one finite-dimensional space to another wit any desired non-zero amount of error, provided the network is given enough hidden units.").

116. The notion that cognition occurs in deeply interconnected networks, such as in both biological and artificial neural networks, is called connectionism. PETER FLACH, MACHINE LEARNING: THE ART AND SCIENCE OF ALGORITHMS THAT MAKE SENSE OF DATA 16 (2012) ("The central idea in connectionism is that a large number of simple computational units can achieve intelligent behavior when networked together. This insight applies equally to neurons in biological nervous systems as it does to hidden units in computational models.").

117. *See supra* note 15.

118. *See id.*

119. This Article distinguishes between machine learning and AI systems because AI is referred in this Article as systems that may include one or more machine learning-based sub-systems (and therefore employ one or more machine-learning algorithms).

a set of possible actions.[120] Reinforcement learning algorithms built on artificial neural networks have been able to defeat professional Go players, chess players, and even expert-level humans at complex real-time strategy games.[121]

What is striking about AI computer programs that are built on machine-learning algorithms is that they can be built to map an arbitrary set of states to an arbitrary set of actions in pursuit of complex goals.[122] A deep reinforcement learning system, for example, may converge on an optimal battlefield strategy, simply by repeating millions of simulated engagements.[123]

Deep machine-learning algorithms, such as deep neural networks, are significantly more complex with size, as no single artificial neuron or layer of artificial neuron bears much individual responsibility for the model's decisions.[124] Thus as the network of artificial neurons increases in size, the

120. Reinforcement learning algorithms are algorithms designed to "maximize a numerical reward signal," but unlike most forms of machine learning, reinforcement learning algorithms "must discover which actions yield the most reward by trying them." RICHARD S. SUTTON & ANDREW G. BARTO, REINFORCEMENT LEARNING: AN INTRODUCTION 2 (1998). A "deep" reinforcement system relies on deep architectures of neural networks to predict future rewards, thus enabling the reinforcement learning system to converge on an environments maximum rewards after repeated trial and error. *See generally* Maxim Lapan, DEEP REINFORCEMENT LEARNING HANDS-ON, loc. 2419 (2018) (ebook) (describing implementation of deep Q-learning system).

121. *See* David Silver et al., *Mastering the Game of Go Without Human Knowledge*, 550 NATURE 354, 354–59 (2017); David Silver et al., *A General Reinforcement Learning Algorithm that Masters Chess, Shogi, and Go through Self-Play*, 362 SCIENCE 1140 (2018); OPENAI FIVE (June 25, 2018), https://blog.openai.com/openai-five/ [https://perma.cc/QB49-KTD9] ("Our team of five neural networks, OpenAI Five, has started to defeat amateur human teams at Dota 2.").

122. This is particularly true for reinforcement learning systems that use deep neural networks, as such systems can learn to execute complex sequences of actions that require planning, meaning anticipating the future and estimating long-term rewards. *See* Razvan Pascanu et al., *Agents that Imagine and Plan*, GOOGLE DEEP MIND (July 20, 2017), https://deepmind.com/blog/agents-imagine-and-plan/ [https://perma.cc/C5LK-MKK2] ("We have seen some tremendous results in this area—particularly in programs like AlphaGo, which use an 'internal model' to analyse how actions lead to future outcomes in order to reason and plan.").

123. *See* SUTTON & BARTO, *supra* note 120, at 4 ("These two characteristics—trial-and-error search and delayed reward—are the two most important distinguishing features of reinforcement learning.").

124. *See* Davide Castelvecchi, *Can We Open the Black Box of AI?*, 538 NATURE 20, 22 (2016) ("But this form of learning is also why information is so diffuse in the network: just as in the brain, memory is encoded in the strength of multiple connections, rather than stored at specific locations, as in a conventional database."); *see also* Bathaee, *supra* note 7, at 891–92 ("AI that relies on machine-learning algorithms, such as deep neural networks, can be as difficult to understand as the human brain. There is no straightforward way to map out the decision-making process of these complex networks of artificial neurons.").

capacity of the AI model likewise increases.[125] With that increase in capacity, however, comes opacity.[126] A fully trained neural network is capable of making decisions the same way a trained expert makes decisions—based on experience and intuition.[127] In other words, there are no detailed instructions given to a computer as in the case of traditional computer programs, but instead, AI programs are often products of the data on which they have been trained.[128] In a sense, the patterns in the underlying training data govern the AI program's decision making. Because the complex network of artificial neurons allows for countless permutations, no single neuron or even layer of neurons encodes any particular part of the decision-making process.[129] Although the inputs to these models are often known, information, such as how those inputs are weighed as they propagate through the networks, may be nearly impossible to determine.

B. THE BLACK BOX PROBLEM

Modern deep neural networks can be very deep and are extremely interconnected. This means that there may not be any clear way of understanding the decision-making process of the network once it is trained on the data.[130] Moreover, the inputs to machine-learning algorithms, including deep neural networks, are often multi-dimensional, meaning that various input

125. Although it is not entirely understood why deeper architectures increase in capacity to approximate non-linear functions, it is assumed that it may because deeper architectures are decomposing non-linear functions into components that can be incrementally estimated. *See* GOODFELLOW ET AL., *supra* note 113, at 195 ("Choosing a deep model encodes a very general belief that the function we want to learn should involve composition of several simpler functions. This can be interpreted from a representation learning point of view as saying that we believe the learning problem consists of discovering a set of underlying factors of variation that can in turn be described in terms of other, simply underlying factors of variation.").

126. *See* Bathaee, *supra* note 7, at 894 ("Deep networks of artificial neurons distribute information and decision-making across thousands of neurons, creating a complexity that may be as impenetrable as that of the human brain.").

127. *See id.* at 902 ("The net result is akin to the way one 'knows' how to ride a bike. Although one can explain the process descriptively or even provide detailed steps, that information is unlikely to help someone who has never ridden one before to balance on two wheels. One learns to ride a bike by attempting to do so over and over again and develops an intuitive understanding."); *cf.* Siddartha Mukherjee, *A.I. Versus M.D.*, NEW YORKER (Apr. 3, 2017), http://www.newyorker.com/magazine/2017/04/03/ai-versus-md [https://perma.cc/MY9K-LBVG] (describing distinction between "knowing that" and "knowing how" forms of learning, where "knowing how" arises from trial and error and is learned from experience).

128. *See* Bathaee, *supra* note 7, at 902–03 ("Because a neural network is learning from experience, its decision-making process is likewise intuitive. Its knowledge cannot in most cases be reduced to a set of instructions, nor can one in most cases point to any neuron or group of neurons to determine what the system found interesting or important.").

129. *See id.*

130. *See infra* Section III.B.

parameters are encoded as high dimensional vectors.[131] Machine-learning algorithms, such as Support Vector Machines, rely on special relationships in higher-dimensional vector spaces.[132] In other words, if there are 115 different parameters used by a model, then machine-learning algorithms will search for patterns in 115 or more dimensions, a sort of geometric space that humans simply cannot visualize.[133] The net effect is both opacity from the vast number of interconnected layers and the difficulty of visualizing higher-dimensional patterns.[134] So there is no clear way for human beings to easily examine the patterns that a machine-learning algorithm may be seizing on as part of its decision-making process.

To complicate things further, the systems built on these machine-learning algorithms may introduce additional opacity to the decision-making process. A reward-seeking reinforcement learning system that uses a deep neural network to estimate future rewards for certain actions may mask the underlying patterns that the deep neural network has detected—all that a human will be able to discern is the estimated rewards for the next actions and those thereafter.[135] For example, a deep reinforcement learning system may predict an eventual checkmate several dozen moves in the future and choose the next move (e.g., moving a pawn two steps forward) that would lead to that outcome, but it may be impossible to tell what series of future moves will ultimately lead to such a result.[136]

---

131. For example, a model that uses three inputs, height, weight, and age, to predict the amount of time it takes for a person to run one mile would receive inputs as a three-dimensional vector (one for each input parameter) and would be searching a three-dimensional space of data for patterns.

132. For a description of Support Vector Machines (SVMs) and how they create opaqueness because of dimensionality, see Bathaee, *supra* note 7, at 903–04; *see also id.* at 905 ("Thus, when the number of variables or features provided to an SVM becomes large, it becomes virtually impossible to visualize how the model is simultaneously drawing distinctions between the data based on those numerous features.").

133. *See* Bathaee, *supra* note 7, at 892 n. 14 ("A two-dimensional space can be visualized as a series of points or lines with two coordinates identifying the location on a graph. To represent a third dimension, one would add a third axis to visualize vectors or coordinates in three-dimensional space. While four dimensions can be visualized by adding a time dimension, five dimensions and higher are impossible to visualize.").

134. *See id.* at 901–04.

135. The output of a deep "Q-learning" reinforcement system, for example, may be a vector of long-term rewards associated with a set of possible actions. *See* Lapan, *supra* note 120, at loc. 2638. Those rewards may not provide any insight into what patterns the reinforcement learning system's deep neural network has spotted and correlated with the anticipated reward.

136. Much of this depends on the structure of the reinforcement learning system. Some reinforcement learning systems evaluate particular moves on a tree of possible outcomes to estimate the value of a move or sequence of moves—in those cases, there may be more

All of these technologies for the first time provide computers the ability to make decisions as humans do—based on experience.[137] A trained neural network will use a decision-making process akin to intuition or judgment.[138] It is essentially the difference between a person who is given detailed instructions on how to ride a bike and a person who has learned to ride a bike—to balance and shift weight—through experience and iteration.[139]

Different machine-learning algorithms create varying levels of opacity. Some can be queried in a way such that outcome-determinative inputs can be ascertained. Others cannot. There are, therefore, both weak and strong versions of the Black Box Problem.[140] All things point to a trend towards the strong form as the technology progresses. Complexity in modern neural networks has increased significantly and it is likely that neural networks will continue to deepen in architecture and increase in size and connectivity.[141]

This Article mostly addresses the strong form of the Black Box Problem. In other words, I assume that fully trained AI systems will be mostly opaque—that the decision-making process cannot be determined by probing the model with different inputs.[142] This is the most problematic incarnation of the Black Box Problem for most legal doctrines, and it is the most important form to consider for legal constructs that rely on intent or scienter heuristics.[143]

C. AI OPINIONS

The most direct use of AI programs built on machine-learning algorithms are systems designed to predict outcomes or to classify data.[144] AI is already

---

transparency as to what course of action the model favors. *See, e.g.*, Silver et al., *supra* note 121, at 2 (noting the use of a Monte Carlo search tree to evaluate potential moves).

137. *See supra* note 127.

138. *See id.*

139. *See supra* note 126.

140. The strong version of the AI Black Box Problem posits that there is no way to determine a rank-order of importance for a model's inputs or to determine how the model is arriving at decisions. *See* Bathaee, *supra* note 7, at 906. The weak form assumes that a loose ordering of input importance can be ascertained. *See id.*

141. Neural network depth will likely increase because it is generally the case that deeper networks potentially have exponentially greater capacity to approximate functions. *See* GOODFELLOW ET AL., *supra* note 113, at 196 ("[P]iecewise linear networks (which can be obtained from rectifier nonlinearities or maxout units) can represent functions with a number of regions that is exponential in the depth of the network.").

142. *See* Bathaee, *supra* note 7, at 906.

143. *See id.* at 906–08.

144. This is because many underlying machine-learning algorithms, including deep neural networks, can be configured directly to classify data or to provide a bounded output, such as a regression or a sigmoid output function. *See* GOODFELLOW ET AL., *supra* note 113, at 166, 347.

being used to make diagnostic predictions given imaging information (such as from MRI results or X-rays).[145] Electronic Medical Records can be mapped to therapeutic outcomes or risk factors.[146] AI can be used to make predictions about what advertisements or search results to display.[147] It can be used to value real estate given a set of inputs about a particular piece of property, or to determine whether a borrower or counterparty is creditworthy. The applications are numerous and rapidly growing.

These applications are natural progressions from deterministic algorithms that occupied the space for decades prior to the recent explosive growth of AI. Automated valuation models, for example, were a prominent feature of the underwriting and appraisals that led to the mortgage-backed securities crisis that occurred after 2008.[148] In those cases, the algorithms and models used did not shield any of the actors from liability because the decision-making process remained mostly in human hands. In fact, when humans made decisions in those cases that ignored the algorithms, their decisions to do so sometimes served as a basis for a finding of scienter.[149]

Even computerized securities trading systems, like high frequency trading systems, which trade securities in fractions of a second,[150] although sometimes autonomous, were for years deterministic, meaning one merely had to examine the underlying code to determine what the intent of the programmer or user

---

145. *See supra* note 12. Many of these image-based models use convolutional neural networks to extract patterns from visual data such as images. *See* GOODFELLOW ET AL., *supra* note 113, at 326 ("[Convolutional Neural Networks] are a specialized kind of neural network for processing data that has a known grid-like topology. Examples include time-series data, which can be thought of as a 2-D grid of pixels.").

146. *See* Huiying Liang & Brian Y. Tsui, *Evaluation and Accurate Diagnoses of Pediatric Diseases Using Artificial Intelligence*, NATURE MEDICINE (Feb. 11, 2019), https://www.nature.com/articles/s41591-018-0335-9 [https://perma.cc/WK68-R7H4] ("Our model applies an automated natural language processing system using deep learning techniques to extract clinically relevant information from EHRs. In total, 101.6 million data points from 1,362,559 pediatric patient visits presenting to a major referral center were analyzed to train and validate the framework. Our model demonstrates high diagnostic accuracy across multiple organ systems and is comparable to experienced pediatricians in diagnosing common childhood diseases.").

147. *See* Tom Simonite, *Google and Microsoft Can Use AI to Extract Many More Ad Dollars from Our Clicks*, WIRED (Aug. 31, 2017, 7:00 AM), https://www.wired.com/story/big-tech-can-use-ai-to-extract-many-more-ad-dollars-from-our-clicks/ [https://perma.cc/4B4Z-LAH7].

148. *See supra* note 7.

149. *See, e.g.*, Fed. Hous. Fin. Agency v. Nomura Holding Am., Inc., 104 F. Supp. 3d 441, 479 (S.D.N.Y. 2015) (finding that decisions to include loans in mortgage-backed securities, notwithstanding automated valuations that exceeded tolerances, could serve as a basis for opinion liability).

150. *See* Bathaee, *supra* note 7, at 908–09.

was when the program was deployed. Indeed, in one of the first criminal trials concerning an unlawful trading practice called spoofing, wherein phantom orders were placed and canceled in fractions of a second in order to move the market, the jury's verdict was based on the testimony of the programmer who created the program at the request of the trader.[151] A human testified about intent because intent was ascertainable—a human provided the computer system detailed instructions, which either evinces an intent to spoof or does not.[152]

Valuations, risk assessments, and even hiring decisions are natural applications for AI models because they were already the subject of intricate deterministic computer programs. Many of the decisions made by computer programs in these fields were based on hard rules or crude statistical patterns (such as linear regressions). The ability to create computer programs that perform the same tasks based on complicated patterns in underlying data—perhaps data collected from hundreds of thousands of human decisions—is undoubtedly the next step for many businesses, governments, and institutions.

Because these models make decisions based on patterns in data and a number of case-specific factors, their outputs are likely to be considered opinions. The output of an AI model that values a security or evaluates counter-party risk or diagnoses patients will be more than a set of underlying facts, more than a set of hard (or even fuzzy) rules, and more than the broad patterns and correlations in the underlying data. They will be opinions to the same extent decisions based on human judgment are opinions, but with one important difference—there will be no human to put on the witness stand to describe the decision-making process that produced the opinion. And there will, in many cases, be no parity between the intent of the AI model's creators and the AI's decision-making schema.[153]

D. THE FAILURE OF THE SCIENTER HEURISTIC

The most serious legal problem posed by any complex AI system is the decoupling of the intent of the system's creators from the system's decisions.[154] A deep reinforcement learning system may be provided a scheme of clear rewards by the designer, but the AI may have many degrees of freedom in how it pursues those rewards and may have to traverse a massive state space to

151. *See supra* note 102.

152. *See id.*

153. Bathaee, *supra* note 7, at 926 ("It is clear that a strong black box, however, cannot be interrogated. Its decision-making process cannot be audited.").

154. *See* Bathaee, *supra* note 7, at 908.

obtain them,[155] which means how the model performs in pursuit of those rewards may be unpredictable. A seemingly absurd, but often recounted example, is Nick Bostrom's paperclip maximizer—an AI tasked with producing as many paperclips as possible. That AI is operating within its parameters even if it consumes all of the resources in the world to obtain its slated rewards. Indeed, because humans would be the source of precious atoms from which paperclips can be made, "the future that the AI would be trying to gear towards would be one in which there were a lot of paper clips but no humans."[156]

The problem is referred to as instrumental convergence.[157] It is the hypothetical notion that AI of a sufficient amount of intelligence will seek to obtain unbounded instrumental goals by maximizing resource acquisition as well as the system's own self-preservation (to ensure its longevity as it pursues its unbounded goals).[158] This is not a literal problem for AI systems today—indeed, few AI systems have unbounded instrumental goals and even fewer are directly plugged into sensitive systems. The problem, including the paperclip hypothetical, however, makes clear that the Black Box Problem is

---

155. Even a real-time strategy video game, such as StarCraft, creates a massive state and action space that a reinforcement learning system must traverse for rewards. *See, e.g.*, Zhen-Jia Pang et al., *On Reinforcement Learning for Full-length Game of StarCraft*, ARXIV 2 (Feb. 3, 2019), https://arxiv.org/pdf/1809.09095.pdf [https://perma.cc/69P7-37L9] ("From the perspective of reinforcement learning, StarCraft is a very difficult problem. Firstly, it is an imperfect information game. Players can only see a small area of map through a local camera and there is a fog of war in the game. Secondly, the state space and action space of StarCraft are huge. StarCraft's image size is much larger than that of Go. There are hundreds of units and buildings, and each of them has unique operations, making action space extremely large.").

156. Kathleen Miles, *Artificial Intelligence May Doom The Human Race Within A Century, Oxford Professor Says*, HUFFINGTON POST (Aug. 22, 2014), https://www.huffingtonpost.com/2014/08/22/artificial-intelligence-oxford_n_5689858.html [https://perma.cc/5HT7-US6K].

157. *See* Nick Bostrom, *The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents*, MINDS & MACHINES 6 (2012), https://nickbostrom.com/superintelligentwill.pdf [https://perma.cc/L69K-J24V].

158. *See id.* More formally, the Instrumental Convergence Thesis posits that:

> Several instrumental values can be identified which are convergent in the sense that their attainment would increase the chances of the agent's goal being realized for a wide range of final goals and a wide range of situations, implying that these instrumental values are likely to be pursued by many intelligent agents.

Id.; *see also id.* at 7 ("Suppose that an agent has some final goal that extends some way into the future. There are many scenarios in which the agent, if it is still around in the future, is then [] able to perform actions that increase the probability of achieving the goal. This creates an instrumental reason for the agent to try to be around in the future—to help achieve its present future-oriented goal.").

not just the result of the complexity of machine-learning algorithms, but also the opacity created by rewards system. Thus, one may be able to specify an AI system's goals very clearly but may not be able to anticipate how the AI achieves those goals, or even what instrumental goals it may deem necessary to achieve them.[159]

Reinforcement learning systems already exceed the capabilities of their creators at the tasks to which they are applied,[160] and indeed, sometimes even exceed the expectations of their creators or their creators' understanding of the problem. For example, those watching Google Deep Mind's AlphaGo and AlphaGo Zero AI play human champions have commented that there is something inhuman about the moves made by the program.[161] It is also beyond dispute that the AI exceeded its creators' ability at the game of Go—indeed, the AI defeated the very best human Go players in the world, which had no hand in the creation of the AI.[162] In other words, the AI's decisions are much more than the mere reward and value specifications set forth by its creator—the AI is making its own decisions.

Consider a reinforcement learning system that is given a reward system based on the amount of money it makes in an electronic trading market. If it stumbles upon spoofing or other forms of market manipulation as a viable strategy for maximizing the rewards it has been tasked to obtain, it may do so notwithstanding the fact that its creators never intended to break the law or engage in a manipulative strategy.[163] Of course, one may object and say that the creator has a duty to impose constraints, but what if the reinforcement learning system stumbles upon a manipulative trading strategy that no human had yet thought of or could even execute (for example, because it would require simultaneous cognition and coordination across thousands of different markets)? Worse yet, what if humans cannot tell that the AI's decisions are

159. *See id.* at 5 ("The orthogonality thesis implies that synthetic minds can have utterly non-anthropomorphic goals—goals as bizarre by our lights as sand-grain-counting or paperclip-maximizing. This holds even (indeed especially) for artificial agents that are extremely intelligent or superintelligent."). Notably, Bostrom believes it is conceptually possible to design systems that behave in a predictable fashion. *See id.* at 7–8. The question is an open one, and as this Article contends, it may be a technological one which depends on the complexity of an AI's internal models. *See supra* note 141 and accompanying text.

160. *See, e.g.*, *supra* note 15.

161. *See* Cade Metz, *How Google's AI Viewed the Move No Human Could Understand*, WIRED (Mar. 14, 2016 2:39 AM), https://www.wired.com/2016/03/googles-ai-viewed-move-no-human-understand/ [https://perma.cc/LYQ2-3WJN].

162. *See id.*

163. *See* Bathaee, *supra* note 7, at 911.

manipulative or unlawful because the AI converges on an obfuscated form of manipulation not strictly prohibited by a priori constraints?[164]

As AI becomes more sophisticated, all of this becomes exceedingly problematic for the hundreds of years' worth of legal doctrines and heuristics that we have accumulated, particularly those based on notions of intent or foreseeability.[165] The intent heuristic fails, because in many cases, AI may be provided perfectly legitimate rewards or ends to pursue, but because the degrees of freedom among possible actions it may pursue is high, the creator of the AI may not be able to predict how the AI will achieve those goals. Indeed, it may be impossible to foresee all of the possibly problematic sequences of actions that the AI may take in pursuit of the rewards it has been given, and that means that the creator of the AI may not be able to anticipate every constraint that would be necessary to keep the AI in line. The high degrees of freedom in the actions the AI can take means that certain actions may simply not be foreseeable, making scienter based on even recklessness or gross negligence impossible to prove because some awareness of risk or foreseeability is a necessary predicate for them.[166]

Because scienter cannot be satisfied when Black Box AI is involved, the law may excuse any injury inflicted by AI simply because no human intended the injury.[167] The net effect is the anomalous result where conduct, if done by a human, would result in liability, but if done by AI would be immune from liability.[168] Thus, a person who engages in spoofing may be convicted of a crime, but a person who designs AI that stumbles upon spoofing as an effective, reward-maximizing strategy, will not result in any liability because the creator of the AI never told it to engage in such a strategy.[169]

There is therefore a perverse incentive to use an AI model to make decisions in highly regulated environments because the AI functionally cuts off any possible liability.[170] Indeed, AI that discriminates based on gender

---

164. The AI may perceive obfuscation of its strategy as an instrumental goal, such as the overarching goal to survive described by Bostrom. *See supra* note 156. In other words, if avoiding detection of the AI's impermissible trading strategy is a necessary sub-goal of its overarching goal to obtain rewards, it will seek to optimize on that sub-goal as well as the rewards.

165. *See* Bathaee, *supra* note 7, at 892, 922.

166. *See* Bathaee, *supra* note 7, at 907.

167. *See id.*

168. *See id.*

169. *Id.* at 911.

170. Consider the current incentive to build complex corporate hierarchies. In the case of corporations charged with federal crimes, only about a third of the cases involved charges against individuals. *See* Brandon L. Garrett, *The Corporate Criminal as Scapegoat*, 101 VA. L. REV. 1790, 1802 (2015). Among those charged, "many were not higher-up officers of the

because of biases in the data used to train it will likely not result in liability for the company that created the AI (because there is no evidence of scienter or even negligence),[171] whereas if the same hiring decision was made by a human, there would possibly be exposure to lawsuits or, at a minimum, ethical objections.

E. AN OPAQUE BASIS AND MATERIALITY

Even without reward-based specifications and reinforcement-learning systems, the basis upon which a vast network of artificial neurons makes decisions will in many cases be impossible to determine.[172] In the case of deep neural networks, the nature of the highly connective system of linear and non-linear mathematical transformations of the data may result in mappings that cannot be readily understood—even if various inputs are provided to the model to determine boundary conditions.[173] In fact, the more complex the decision-making process, the less likely it will be that the AI's decision making can be mapped out simply by examining correlations between inputs and outputs of the AI model.[174]

For example, AI that makes a medical diagnosis based on three or four parameters in a patient's medical file can likely be probed—if one of the parameters is age, then the input age can be varied to determine whether the change is outcome determinative. But as more parameters are added, the

---

companies, but rather middle managers of one kind or another and also some quite low-level individuals." *Id.* at 1802. The explanation may be that corporate complexity results in the insulation of senior corporate executives from liability, as the more complex the organization becomes, the more the corporate institution can be blamed instead of the willful conduct of any single person. *Id.* at 1825. The incentive to use AI may be somewhat similar to the incentive to create complexity in a corporation—the complexity and opaqueness of the AI diffuses responsibility and insulates senior corporate officers, particularly if the programming, testing, and business applications of the AI are responsibilities of different parts of an organization.

171. At least one company has scrapped its AI designed to vet potential employees because their AI discriminated based on gender. Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women*, REUTERS (Oct. 9, 2018 11:12 PM), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G [https://perma.cc/99QW-EPWX]. It did so because of bias that existed in ten years of hiring data used to train the AI. *Id.* Several large companies have developed or are developing similar tools, though none have reportedly experienced the same kind of discrimination due to training data bias. *Id.*

172. *See* Bathaee, *supra* note 7, at 901.

173. *See supra* note 139.

174. This assumption definitionally tracks a strong-form Black Box Problem, which assumes that one cannot deterministically map inputs to outputs simply by probing the model. *See* Bathaee, *supra* note 7, at 906 ("Importantly, this form of black box cannot even be analyzed *ex post* by reverse engineering the AI's outputs.").

number of possible input combinations exponentially increase. Indeed, when a model bases its decisions on thousands of input parameters, it would take several lifetimes to fully determine what the model's decision boundary looks like.[175] Age may be outcome determinative when it is one of a few other parameters, but when it is one of thousands, it may be relevant only when certain other parameters meet certain criteria, and even then, it may not be outcome determinative in many cases. Thus, it becomes impossible to, for example, rank which input parameters are the most important to the AI model.[176]

This inability to understand the basis for an AI's decisions may also impair the materiality inquiry.[177] Human judgment may differ to such an extent from the AI's that an omitted fact may be material to a reasonable person but entirely irrelevant to an AI. Indeed, it may be that in all cases, the number of bedrooms in a house would be important to a human being that is valuing a house, but the AI may determine that in a certain zip code and given a certain threshold square footage, the number of bedrooms does not increase the accuracy of the model's valuations—if that's the case, the model may be making its decisions without considering the number of bedrooms in the house in many particular instances. Yes, a reasonable person would want to know how many bedrooms were in the house, but a trained and accurate AI model may not care at all—and maybe for good reason (because it does not make the model more or less accurate to consider that information).

A rule that hinges opinion liability on whether a material fact underlying the opinion was omitted may therefore focus on entirely spurious notions of materiality when AI is concerned.[178] If one cannot tell how the AI assigns

175. In the discrete case, meaning that the input space consists of non-continuous inputs such as a finite set of integers, the massive input space is a matter of combinatorics, as the number of possible input combinations potentially multiply, creating exponentially larger possible inputs as input parameters are added. In the continuous case, such as when the inputs are real numbers (approximated by floating-point numbers of a fixed bit size in the case of most computers), the input space, while discrete in the sense that the possible inputs for each parameter is bounded by the precision of the floating-point numbers used, becomes unfathomably massive. In other words, in almost any real-world case, it is simply not possible to try all of the possible input combinations to determine effects on outputs.

176. *See* Bathaee, *supra* note 7, at 906.

177. Materiality here refers to the legal test, which asks whether a stated or omitted fact was important to a decision. *See supra* note 36.

178. The power of AI comes from pattern recognition, and sometimes the value of the AI is that it can recognize patterns that are not intuitive or perceptible to humans. If one could simply examine the AI to determine what is most material to it, one could write a determinative algorithm to perform the AI's task—it would be traditional software. *See* Rudina Seseri, *The Problem with "Explainable" AI*, TECHCRUNCH (June 14, 2018), https://techcrunch.com/2018/06/14/the-problem-with-explainable-ai/ [https://perma.cc/QE8R-7VU6] ("Part of the

weights to parameters and patterns in particular contexts, there may be no way to prove that the omitted information was an important or relevant part of the AI's decision.[179]

F. AI AS AN OPAQUE EXPERT

It is not uncommon for human experts to use intuition or judgment to render opinions.[180] A bank that hires a valuation expert to determine the value of complex derivatives may not have much insight into aspects of the expert's opinion that are based on his experience or judgment. Experts with technical or scientific expertise may rely on mathematics that would require the lay person years of study to understand. In such cases, the expert is functionally a Black Box to those who may rely on his opinion.[181] There is, however, a notable difference. Humans can attempt to explain the bases for their opinions, and in some cases, even the principles and assumptions underlying their opinion.[182]

---

advantage of some of the current approaches (most notably deep learning), is that the model identifies (some) relevant variables that are better than the ones we can define, so part of the reason why their performance is better relates to that very complexity that is hard to explain because the system identifies variables and relationships that humans have not identified or articulated. If we could, we would program it and call it software."); *see also* Will Knight, *The Dark Secret at the Heart of AI*, MIT TECH. REV. (Apr. 11, 2017), https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ [https://perma.cc/QNG2-XSYJ] ("Information from the vehicle's sensors goes straight into a huge network of artificial neurons that process the data and then deliver the commands required to operate the steering wheel, the brakes, and other systems. The result seems to match the responses you'd expect from a human driver. But what if one day it did something unexpected—crashed into a tree, or sat at a green light? As things stand now, it might be difficult to find out why. The system is so complicated that even the engineers who design it may struggle to isolate the reason for any single action. And you can't ask it: there is no obvious way to design such a system so that it could always explain why it did what it did.").

179. *See* Bathaee, *supra* note 7, at 906.

180. *See, e.g.*, Daniel Susskind, *AlphaGo Marks Stark Difference between AI and Human Intelligence*, FIN. TIMES (Mar. 21, 2016), https://www.ft.com/content/8474df6a-ed0b-11e5-bb79-2303682345c8 [https://perma.cc/QZ7F-U69Z] ("When researchers sat down with grandmasters and asked them to explain how they played such fine chess, the answers were useless. Some players appealed to 'intuition,' others to 'experience.' Many said they did not really know at all.").

181. Of course, humans have developed heuristics to assess a person's credibility and trustworthiness. *See* Knight, *supra* note 179 ("Sure we humans can't always truly explain our thought processes either—but we find ways to intuitively trust and gauge people.").

182. Human experts or even institutions can also signal credibility or authority, thus provoking epistemic deference. *See* M. Neil Browne & Ronda R. Harrison-Spoerl, *Putting Expert Testimony in Its Epistemological Place: What Predictions of Dangerousness in Court Can Teach Us*, 91 MARQ. L. REV. 1119, 1132–33 (2008) ("When the court hears the testimony of an 'expert,' especially someone recognized as a 'scientific expert,' the jury may be overly impressed by the credentials presented and terminology used by this individual, hindering the jury's ability to

Indeed, it is not unusual for a human expert to explain their methodology by providing the broad strokes of the basis for the opinion.[183] In some cases, formalized mathematical or scientific models can be explained by analogy to something lay people can understand.[184] Human experts can also provide rankings of what they deemed to be most important to their decisions.

Consider a judge. While true that a judge may make a legal decision based on their intuition or expertise, they will also be able to provide a justification for their decision.[185] It may be the case that factors frequently used by

---

fully understand and evaluate the evidence presented by the expert."). Such signaling may be nothing more than an implicit appeal to the authority of the person or institution that stated a particular opinion, which may eliminate any inquiry into the underlying rationale for the cited opinion. *Cf.* Rostker v. Goldberg, 453 U.S. 57, 112 (1981) (Marshall, J., dissenting).

183. This is precisely what an expert witness attempts to do when explaining a scientific opinion to a jury. In most cases, the jury cannot directly evaluate the expert's scientific analysis but will instead determine whether the expert appears to be credible on the subject—whether he is to be believed. *See* H.L.A. HART, ESSAYS ON BENTHAM: STUDIES IN JURISPRUDENCE AND POLITICAL THEORY 261–62 (1982). The essays state:

> To be an authority on some subject matter a man must in fact have some superior knowledge, intelligence, or wisdom which makes it reasonable to believe that what he says on the subject is more likely to be true than the results reached by others through their independent investigations, so that it is reasonable for them to accept the authoritative statement without such independent investigation or evaluation of his reasoning.

Id. In other words, the jury will believe the person and therefore the proposition. *See* Scott Brewer, *Scientific Expert Testimony and Intellectual Due Process*, 107 YALE L.J. 1535, 1583 (1998) ("Where S is some speaker offering testimony that p and H is a hearer of that testimony, it is the distinction between H's believing that p and H's believing S that p.").

184. An expert that explains his view with analogies or simplifications often implicitly reveals what facts were material to his opinion. This is because to draw an analogy—that is, to engage in any analogical reasoning—one must generally make determinations as to which facts in one context are similar or different from those in an analogous context. *See* Frederick Schauer & Barbara A. Spellman, *Analogy, Expertise, and Experience*, 84 U. CHI. L. REV. 249, 253 (2017) ("[T]here remains a core position according to which the first move in the analogical process is the recognition of a relevant similarity between some previous set of facts and the set of facts that now calls for decision."); LARRY ALEXANDER & EMILY SHERWIN, DEMYSTIFYING LEGAL REASONING 76–83 (Cambridge ed. 2008) ("Similarities are infinite; therefore some rule or principle is necessary to identify important similarities."); *cf.* RICHARD A. POSNER, THE PROBLEMS OF JURISPRUDENCE 91 (Harvard ed. 1990) ("A set of cases can compose a pattern. But when lawyers or judges differ on what pattern it composes, their disagreement cannot be resolved . . . by an appeal to an intuitive sense of pattern.").

185. The requirement to write a legal opinion allows for a system where analogical reasoning is possible—otherwise, cases cannot look backwards to compare their reasoning about the facts before them to the facts and reasoning from previous cases. James Boyd White, *What's an Opinion For?*, 62 U. CHI. L. REV. 1363, 1368 (1995) ("The judicial opinion is a claim of meaning: it describes the case, telling its story in a particular way; it explains or justifies the result; and in the process it connects the case with earlier cases, the particular facts with more general concerns. It translates the experience of the parties, and the languages in which they

laypersons, such as moral judgments or life experiences, may not be explicit in such a decision,[186] but many of the important facts of the case—akin to parameters fed into an AI model—will be identified and some rank order can be discerned from the opinion. Even in the case of non-dispositive and multi-factor tests, legal opinions often provide insight into the factors and facts most important to the analysis. In any event, the number of factors in most cases are often few enough that crude empirical or statistical analysis will often provide insight as to what factors were dispositive.[187]

For the first time, AI presents us with the power of the expert but at the expense of transparency.[188] Common legal heuristics, such as witness examination or the use of contemporaneous evidence to determine intent, simply will not work when AI is involved.[189] The AI may have decision-making agency, so the intent of its user or creator may not matter.[190] In most cases, there will simply be no relevant human intent to apply an intent heuristic to.[191]

IV. A FRAMEWORK FOR AI OPINION LIABILITY

A. BETTER FACTUAL HEURISTICS FOR AI OPINION LIABILITY

If traditional heuristics such as intent may not work with AI, then a new set of factual heuristics is needed for liability—heuristics tailored to machine-learning models, not to human beings. The liability question should not turn on what the creator or user of the AI intended, but instead on how the creator or user of the AI trained the AI,[192] what data was used, what biases in the data

---

naturally speak of it, into the language of the law, which connects cases across time and space; and it translates the texts of the law—the statutes and opinions and connotational provisions—into the terms defined by the facts of the present case.").

186. Some studies have shown that judges use cognitive decision-making processes no different that laypersons when making decisions and therefore fall prey to cognitive biases, just as laypersons do. *See* Chris Guthrie et al., *Inside the Judicial Mind*, 86 CORNELL L. REV. 777, 829 (2001) ("Our study demonstrates that judges rely on the same cognitive decision-making process as laypersons and other experts, which leaves them vulnerable to cognitive illusions that can produce poor judgments. Even if judges have no bias or prejudice against either litigant, fully understand the relevant law, and know all of the relevant facts, they might still make systematically erroneous decisions under some circumstances simply because of how they—like all human beings—think.").

187. *See, e.g.*, Barton Beebe, *An Empirical Study of U.S. Copyright Fair Use Opinions*, 156 U. PENN. L. REV. 549, 585 (2008) (empirically determining that the first and fourth fair use factors in copyright were the most important).

188. *See supra* Section III.B.

189. *See supra* Sections III.B & III.D.

190. *See supra* Sections III.B & III.D.

191. *See supra* Section III.E.

192. *See infra* Section IV.A.2.

were, or could have been, detected, and most importantly, how the AI was tested, validated, and deployed.[193] This Section discusses some factual heuristics that will be useful in assessing liability, particularly because they allow us to make some basic determinations about the intent of the AI's user or creator.

To be sure, because of the Black Box Problem, each of these heuristics may be as ineffective as a traditional scienter heuristic, because the AI user or creator's intent or conduct may be completely decoupled from the AI's decision-making process.[194] These heuristics, however, get at the heart of whether humans using or deploying the AI should be responsible for the AI's conduct. These heuristics only go so far, but they are the beginning of any analysis, even if they are ultimately not dispositive.

1. Deference and Autonomy

The first heuristic for liability should be the degree of autonomy the AI was given and how much the AI's opinion was relied upon.[195] Just as one would assign a high degree of autonomy to a trusted agent, the user or creator of AI may defer to powerful and accurate AI.[196] The key question is whether deference to an AI's opinion was reasonable under the circumstances. The threshold factual question will therefore often be the extent to which the creator or user of AI deferred to the AI's decisions.

Here, deference refers to reliance on the AI opinion. For example, if a manufacturer relies on an AI's assessment of product safety without any human intervention or check, it may be a telltale sign that the user or creator of the AI believed that the AI was adequately tested and was worthy of deference. In cases where a high degree of deference is not justified, such as when the AI has not been adequately tested or lacked sufficient unbiased training data, it will be a basis upon which to assign liability.[197] That is, deference or reliance on AI without any human supervision may be evidence that the creator or user of the AI fell below a given standard of care; this will

193. *See infra* Section IV.B.

194. *See supra* Sections III.B, III.D, &III.E.

195. *See* Bathaee, *supra* note 7, at 936.

196. This notion is similar to the doctrine of negligent supervision, which posits that an employer should be responsible for a failure to exercise ordinary care in supervising an employee. *See* Jackson v. Ivory, 120 S.W.3d 587, 598 (Ark. 2003).

197. The question will often be whether the degree of supervision fell short of a standard of ordinary care, just as in the case of a negligent supervision action. *See supra* note 196.

be the case when there is a great amount of uncertainty as to how the AI is making its decisions or as to how it will perform in the real world.[198]

Examining deference and autonomy provides vital context in AI opinion cases. Deference to an AI's medical diagnosis may not be justified, even if it is accurate more than half of the time. The risk of loss if the AI fails to properly diagnose a patient will be too high in the case of some patients. Deference when the model is less than 50% accurate at diagnosing a disease, however, may be perfectly adequate if what is being diagnosed is a relatively benign malady, such as a cold.

A focus on deference and autonomy immediately converts a technical problem—the opacity of an AI model[199]—into a classic question of fact that a judge or jury can assess using battle-tested legal constructs, including the rules of evidence and standards of care from the law of negligence.[200] Of course, in many cases, the creator or user of the AI may subjectively and reasonably believe that deference was justified, and they may be wrong because of the Black Box Problem. In such cases, the deference and autonomy heuristic may not be dispositive. In other words, it may not be appropriate to excuse the creator or user of the AI from liability simply because they could not foresee the AI's decision boundary or the effects of the AI's opinions.

2. Training, Validation, and Testing

The next useful factual heuristic will be a focus on the training, validation, and testing of the AI model. Most of the most powerful AI are built on machine-learning algorithms that learn from data.[201] The most important question when examining such models is whether they "generalize," meaning whether they have seized on patterns that generally exist in a particular sort of dataset.[202]

198. This is similar to the situation where an employee's history of conduct alerts an employer to a potential risk of wrongful conduct by the employee. *See* Leftwich v. Gaines, 521 S.E.2d 717, 726 (N.C. 1999) (noting that some cases find liability where "the employee's wrongdoings were forecast to the employer and took place while working").

199. *See supra* note 178.

200. *See supra* note 107.

201. *See supra* note 113.

202. *See* GOODFELLOW ET AL., *supra* note 113, at 107 ("The central challenge in machine learning is that our algorithm must perform well on new, previously unseen inputs—not just on those which our model was trained. The ability to perform well on previously unobserved inputs is called generalization."); YASER S. ABU-MOSTAFA ET AL., LEARNING FROM DATA 39–40 (2012) (noting that generalization, "a key issue in learning," is the degree of error on data not used to train a model—that is, on "out of sample" data).

A model may be trained to fit too closely to the data upon which it was trained.[203] This concept is referred to as overfitting.[204] The easiest example of overfitting would be memorizing a math textbook and then taking a final exam with problems you have never seen before. Memorizing the textbook without understanding what is in it will not get the student very far. In such a case, the student has overfit to the material in the textbook but is not capable of generalizing based on the data they have studied.[205]

Patterns in data may also be spurious, meaning that a generalization from those patterns may not assist with new data inputs.[206] In that case, the model may have been trained to make very crude distinctions, which is what may be causing the inaccuracy. It may provide some baseline accuracy to predict mile run times based on the height of a runner—and perhaps it will work well at the extremes of the height distribution—but it will generally not provide a very good model for differentiating among close cases (e.g., two runners with similar heights in the middle of the distribution). The error rate will be too high for the model to generalize in a meaningful way.

It is important to note that any data-driven mathematical or statistical model will only work if there are patterns in the underlying data used to train them. Patterns in a particular period of stock market returns may not hold in the future when market dynamics and fundamentals change. Patterns in the medical records of certain genetically similar patients may not exist at all in others who do not share any genetic similarity.

To AI researchers, mathematicians, and economists, this notion is a familiar one—it is often referred to as the "No Free Lunch Theorem."[207] Put

---

203. In such a case, the model has overfitted to the training data—that is, the model correctly predicts training data, but it has a high error rate on data it has not yet seen. *See id.* ("Overfitting occurs when the gap between the training error and test error is too large.").

204. *Id.*

205. *See* ABU-MOSTAFA ET AL., *supra* note 184, at 119 ("Overfitting is the phenomenon where fitting the observed facts (data) well no longer indicates that we will get a decent out-of-sample error, and may actually lead to the opposite effect. You have probably seen cases of overfitting when the learning model is more complex than is necessary to represent the target function. The model uses its additional degrees of freedom to fit idiosyncrasies in the data (for example, noise), yielding a final hypothesis that is inferior.").

206. This corresponds to underfitting—when there is no learnable pattern in the training data other than, perhaps, a crude pattern with little predictive power. *See* GOODFELLOW ET AL., *supra* note 95, at 108 ("Underfitting occurs when the model is not able to obtain a sufficiently low error value on the training set.").

207. *See* GOODFELLOW ET AL., *supra* note 113, at 95 ("The no free lunch theorem for machine learning states that, averaged over all possible data generating distributions, every classification algorithm has the same error rate when classifying previously unobserved points. In other words, in some sense, no machine learning algorithm is universally any better than any other.").

simply, if an algorithm performs well on a particular dataset, it does so at the expense of performing poorly on another.[208] It essentially posits that there is no uniform set of patterns across all possible datasets.[209] In other words, AI models—or any mathematical or statistical model, for that matter—must fit to the data upon which they have been trained. If that data contains patterns that are not in other datasets, the model will not be effective in making predictions when it is shown new data.[210]

This is why it matters how the AI has been trained, validated, and tested.[211] To begin with, it is important to note whether best practices were followed when training the AI. For example, it is worth asking whether the dataset had been separated into a subset for training and a subset for testing or validation.[212] This prevents the AI model from knowing any information about the test data.[213] This allows a more accurate determination of whether the AI model is appropriately generalizing.[214] The model is trained on one subset of data, and if the accuracy rate holds on the testing subset, then the model has successfully been trained to recognize patterns in the data.[215] Conversely, if the accuracy rate is high in training but is poor in testing, then the model may be overfitting on the training data.[216]

Moreover, because models require tuning during the training process, using subsets of data for validation during training provides further assurances that none of the test set information was used to train the model.[217] There are many best practices for training machine-learning models, which are beyond

208. *See id.*

209. *See id.*

210. *See id.* at 115 ("This means that the goal of machine learning research is not to seek a universal learning algorithm or the absolute best learning algorithm. Instead, our goal is to understand what kinds of distributions are relevant to the 'real world' that an AI agent experiences, and what kinds of machine learning algorithms perform well on data drawn from the kinds of data-generating distributions we care about.").

211. Validation involves creating a subset of the training data and holding it out for tuning of the model. *See id.* at 118. Testing would be performed on a dataset that was not used for training or tuning. *See id.*

212. *See id.*

213. *See id.* at 119 ("It is important that the test examples are not used in any way to make choices about the model, including its hyperparameters. For this reason, no example from the test set can be used in the validation set.").

214. *See id.*

215. *See id.* at 109–10.

216. *See id.*

217. *See id.* at 119 ("More frequently, the setting must be a hyperparameter because it is not appropriate to learn that hyperparameter on the training set. This applies to all hyperparameters that control model capacity. If learned on the training set, such hyperparameters would always choose the maximum possible model capacity, resulting in overfitting.").

the scope of this Article, and many of these best practices are likely to change, but the extent to which the model was trained, validated, and tested according to some relevant standard of care will be imperative for assigning liability.

It is important to note that the training, validation, and testing heuristic is closely connected with the deference and autonomy heuristic. If a model is poorly validated and tested, then it may not have been reasonable to provide the model autonomy or deference.[218]

3. Constraint Policies and Conscientiousness

The extent to which constraints are provided to the AI model will also be an important heuristic. A defendant that takes great care to prevent opinions based on improper or spurious bases will be less culpable than one who provides the AI model an unbounded degree of freedom to achieve a particular accuracy, result, or reward.[219] The existence or lack of constraints says something about the creator or user of the AI's conscientiousness.[220] Did they attempt to exercise some care when deploying the AI? Because conscientiousness is something that must be incentivized, it makes sense to provide safe harbors for those who impose extensive constraints on AI decision making.[221]

Without a focus on constraints imposed on the AI, existing liability rules may perversely incentivize reckless behavior. If the imposition of constraints on the AI does not mitigate liability, it may only serve to establish that the user of the AI was aware of a particular risk.[222] For example, an AI model tasked with making hiring decisions that includes software safeguards against race or gender-based discrimination may prove that the defendant was aware of the

218. *See supra* Section IV.A.1.

219. Indeed, the failure to impose reasonable safeguards may allow the inference of recklessness or other forms of scienter, such as willful blindness. *See* Bathaee, *supra* note 7, at 933–34.

220. *See id.* at 933.

221. Some commentators have argued that safe harbors are an effective means of incentivizing laudable corporate conduct. *See* Elizabeth F. Brown, *No Good Deed Goes Unpunished: Is There a Need for a Safe Harbor for Aspirational Corporate Codes of Conduct*, 26 YALE L. & POL'Y REV. 367, 402 (2008) ("How can the law be amended in order to encourage businesses to seek to achieve higher standards of behavior than the bare legal minimum? One possible solution might be for states or the federal government or both to enact laws that limit the ability of corporations to be sued if they make good faith efforts to achieve aspirational standards of behavior but fails as long as their conduct still met the legal standards embodied in statutes, regulations, and the common law.").

222. *Cf. id.* at 401–02 ("[T]he market forces may encourage some businesses to try to get as close to the line of what is legally permissible behavior in order to maximize profits. The danger with that sort of behavior is that businesses frequently misjudge where the line is and end up owing large penalties and legal bills for violations of the law.").

risk that such discrimination would occur if the AI's opinion was relied upon. In other words, by imposing safeguards, the creator or user acknowledges consciousness of risk.

If constraints do not mitigate liability, then it may be better for the person using the AI not to impose any constraints on the AI at all and argue that any harm arising from the AI's opinions were completely unintended and perhaps unforeseeable. In other words, a conscientiousness heuristic that relieves a conscientious actor from liability may not only help assess the degree of culpability for deploying the AI, it would also incentivize reasonable care and risk mitigation.[223]

B. EXAMINING DATA BIAS, NOT DECISION BASIS IN OMISSIONS CASES

The Supreme Court has announced a rule in the securities law context that allows courts to examine the basis for an opinion statement only when there is a claim that material information was omitted from the opinion statement.[224] Such a rule, however, would be ineffective in many AI opinion contexts. If the AI is opaque, meaning that it suffers from the Black Box Problem, it will likely be nearly impossible to determine the basis for the AI's opinions.[225] In many cases, even a weak rank ordering of input parameters will not be possible.[226]

The basis of an opinion in an omissions case allows a court or factfinder to determine whether the speaker of the opinion was justified in holding the opinion.[227] If, for example, the speaker of the opinion considered information that undermined the opinion but failed to disclose that information, it may be fair to hold him liable for the omission.[228] The question of liability in such a circumstance will turn on whether it was reasonable of the person to hold the opinion notwithstanding the inconsistent or contradictory information in the speaker's possession.[229] The rule also safeguards against an entirely reckless or uninformed opinion.[230] In cases where the speaker never made any investigation as to any of the relevant facts or ignored gaping deficiencies in information required for the opinion, it may also be entirely fair to hold them

223. *See id.* at 402.

224. *See* Omnicare, Inc. v. Laborers Dist. Council Constr. Indus. Pension Fund, 135 S. Ct. 1318, 1327–29.

225. *See* Bathaee, *supra* note 7, at 916–17.

226. *See id.*

227. *See supra* note 78.

228. *See supra* note 47.

229. *See Omnicare*, 135 S. Ct. at 1327–29.

230. *See supra* note 63.

liable for the statement.[231] A reckless opinion in such cases is the analog of a false statement.[232]

In the AI context, the models will likely be trained on large amounts of information. If the model has been properly trained, validated, and tested, the question will seldom be about whether any factual investigation was done to justify the AI's opinion.[233] The question will instead be about whether the model was trained with data that implicitly expressed some form of bias.[234]

To AI systems, data bias can be the same as blindness.[235] If AI is trained with data that contains a prevalent pattern, it will undoubtedly leverage that pattern in its decision making—it may even do so with too much emphasis.[236] It may also be the case that the underlying data narrowly covers only a subset of possible data points and leads the AI to make decisions in all contexts as it would in a narrow unrepresentative one.[237]

Consider an AI designed to provide an opinion on recidivism in prisoners being considered for parole. If the underlying data is trained on a population of past inmates that were subject to widespread racial discrimination by the police while released on parole, then the data may use race as a proxy for its decision making.[238] Of course, one would say that if race is not included as an input to the model, then it cannot be considered. But even then, a model may be capable of using proxies for race, such as their economic backgrounds or even zip codes.[239] Because the data bias is so overwhelming, other factors will correlate with the bias and infiltrate the AI's opinions.[240]

---

231. *See id.*

232. This is in part because the implicit statement that the speaker has a basis for his opinion has proven false. *See supra* note 47.

233. The data used to train the model is the definition of factual information; it is the very basis upon which the model is built. In other words, with respect to AI that learns from data, if there were no foundation, there would be no data and therefore no AI model.

234. *See supra* Section IV.A.

235. This is because the AI cannot learn from data it never observed during training. If the out-of-sample data is too different from the training data the model has encountered, the model may overfit on the training data and fail to generalize with respect to the new data it has not yet encountered. *See supra* note 205.

236. *See id.*

237. This phenomenon arises from sampling bias. It is axiomatic that "[i]f the data is sampled in a biased way, learning will produce a similarly biased outcome." *See* ABU-MOSTAFA ET AL., *supra* note 202, at 172.

238. *See* Bathaee, *supra* note 7, at 920.

239. *See id.*

240. The model may be "snooping" at the prohibited data, meaning that the data has made its way into the model even though it was not explicitly provided. Accordingly, it is generally assumed that "[i]f a data set has affected any step in the learning process, its ability

The question is thus (a) whether there exists a bias in the training data, (b) whether the bias is strong, and (c) whether other input parameters correlate with that bias. If the bias is improper, the opinion may also be improper.[241] These three determinations can be made even when the model exhibits the Black Box Problem. There are myriad mathematical and statistical methods that can demonstrate the existence and strength of biases in underlying data, including correlation and covariance among variables.[242] Such a showing, coupled with a showing that such bias is legally improper, may be sufficient in some cases to establish opinion liability.

Take for instance the example of an AI designed to diagnose a particular medical condition based on the existence of a particular set of genetic traits. If the AI fails to diagnose a large number of patients and those patients relied on a negative diagnosis, then the question will be whether the AI was properly trained and tested.[243]

Assuming that it was properly tested and trained, the next question is whether bias in the data would explain why the AI's accuracy decreased in real-world application. A showing that all of the data came from the health records of a large interrelated population of patients with some other common genetic trait, then that bias may be the problem with the model.[244] It may be that the diagnosis is highly accurate in that biased population, but not so among the general population.[245]

The question for liability is therefore whether that bias in the underlying data was detectable, and if so, how strong that bias was.[246] It may be obvious

---

to assess the outcome has been compromised." *See* ABU-MOSTAFA ET AL., *supra* note 202, at 173.

241. *See id.*

242. *See, e.g.*, GOODFELLOW ET AL., *supra* note 113, at 119–21.

243. *See supra* Section IV.A.

244. *See* Robert David Hart, *If You're Not a White Male, Artificial Intelligence's Use in Healthcare Could Be Dangerous*, QUARTZ (July 10, 2017), https://qz.com/1023448/if-youre-not-a-white-male-artificial-intelligences-use-in-healthcare-could-be-dangerous/ [https://perma.cc/Q9WK-PJLP] ("The highly selective nature of trials systematically disfavor women, the elderly, and those with additional medical conditions to the ones being studied pregnant women are often excluded entirely. AIs are trained to make decisions using skewed data, and their results will therefore factor the biases contained within. This is especially concerning when it comes to medical data, which weighs heavily in the favor of white men.").

245. *See id.*

246. In medical cases, privacy constraints may compound the Black Box Problem's barriers to determining how the AI made its decision. This means that data biases may be especially difficult to detect in the medical context. *See id.* ("AI systems often function as black boxes, which means technologists are unaware of how an AI came to its conclusion. This can make it particularly hard to identify any inequality, bias, or discrimination feeding into a particular decision. The inability to access the medical data upon which a system was trained—

that drawing from a narrow population of patients from which to train the AI model was an error—well below the standard of care for diagnosis. In such a case, liability would fairly attach. And in the case where it can be shown that the creator of the AI knew about the data bias but deployed the model anyway, the omission of the data bias would also be grounds for liability.[247]

C. HIGH RISK / HIGH VALUE APPLICATIONS AND STRICT LIABILITY

There may be some cases where it is never safe to defer to an AI opinion, and in those cases, strict liability may be appropriate.[248] Medical applications will likely be riddled with such circumstances. Indeed, it may be that it will always be reckless to leave cancer diagnosis entirely to AI.[249] The same can be said about AI designed as weapons for police or military functions.[250] It may always be unreasonable to rely on an armed robot powered by an AI system for crowd control—indeed, the AI's opinion as to whether a person poses a

---

for reasons of protecting patients' privacy or the data not being in the public domain—exacerbates this.").

247. *See supra* Part II.

248. Some commentators have pointed out that an important precondition for a strict-liability regime to be viable is adjudicability, meaning that there is a defined set of facts or conduct to which strict liability will apply. *See* James A. Henderson, Jr., *Why Negligence Dominates Tort*, 50 UCLA L. REV. 377, 391 (2002) ("For disputes under strict liability to be adjudicable, the boundaries of the liability system—the descriptions of harm-causing activities for which the system holds enterprises strictly responsible—must be relatively specific and must not depend on fact-sensitive risk-utility calculations."). Because an AI's reasoning or conduct may be unpredictable due to the Black Box problem, it is the application of the AI that must be the defined trigger for a strict liability regime to be viable, not the particular conduct or risks involved in an AI's deployment.

249. Section 402A of the Restatement (Second) of Torts, which was drafted by William Prosser, was one of the first significant movements toward strict liability for unreasonably dangerous products, and most states eventually adopted some form of the rule stated in the Restatement. James A. Henderson, Jr. & Aaron D. Twerski, *A Proposed Revision of Section 402A of the Restatement (Second) of Torts*, 88 CORNELL L. REV. 1512, 1512–13 (1992); *see also* George L. Priest, *The Invention of Enterprise Liability: A Critical History of the Intellectual Foundations of Modern Tort Law*, 14 J. LEGAL STUD., 461, 512, 518 (1985). The rule has been applied to tobacco and cigarette products that cause cancer—see, for example, Hearn v. R.J. Reynolds Tobacco Co., 279 F. Supp. 2d 1096, 1103 (D. Ariz. 2003)—and it is certainly conceivable that the rule would also apply to an AI diagnosis product that improperly diagnoses (or fails to diagnose) cancer.

250. *See, e.g.*, Joseph A. Page, *Of Mace and Men: Tort Law as a Means of Controlling Domestic Chemical Warfare*, 57 GEO. L.J. 1238, 1258 (1969) (arguing that spray weapons could give rise to liability "[w]hen the plaintiff is the intended target and suffers more than transitorily disabling harm as a result of a construction or design defect in the spray or inadequate warnings and instruction in its use").

threat may be 98% accurate, but a 2% failure rate may result in death or injury. These are high risk applications, which may justify a strict liability regime.[251]

There is another class of cases where deference to an AI opinion may be improper—high value applications. For example, AI opinions are unlikely to be a fair replacement for juries or judges. Constitutional and democratic norms may require a human being to make factual determinations at trial.[252] A judge that entirely delegates legal decision making to an AI system that has been trained to predict how they would decide cases may also be doing so improperly.[253] Sometimes, a human check is required, even if it is a slight one.[254] This may be because due process requires it or it may be because the AI is opining on a function that as a society we would prefer humans to perform.

---

251. *See* Bathaee, *supra* note 7, at 931 (arguing that strict liability may be appropriate in some limited cases but that a blanket strict liability rule for AI would not be appropriate); *cf.* Yesha Yadav, *The Failure of Liability in Modern Markets*, 102 VA. L. REV. 1031, 1039 (2016) (arguing against strict liability for determinative algorithms). Strict liability in tort for abnormally dangerous conduct is, for the most part, circumscribed in application, as it applies only to a short list of abnormally dangerous activities, which have not been significantly expanded over the years. *See* John C.P. Goldberg & Benjamin C. Zipursky, *The Strict Liability in Fault and the Fault in Strict Liability*, 85 FORDHAM L. REV. 743 (2016) ("Each of the three Restatements of tort law has recognized a special domain of strict liability under the labels 'ultrahazardous' or 'abnormally dangerous' activities. This domain is quite narrow, applying only to injuries caused by blasting, escaped wild animals, bursting reservoirs, and a few other activities. Plaintiffs' lawyers, courts, and commentators have at times suggested that the particular form of liability that attaches to abnormally dangerous activities should occupy more of the torts landscape. But no such expansion has occurred.").

252. *See* Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1178 (2018) ("Yet the nondelegation doctrine, still a fixture in American constitutional and administrative law, places some theoretical limits on those delegations, which must, for example, be accompanied by an intelligible principle. Although this doctrine has long accepted even broad delegations of authority to administrative agencies, the law has always assumed that the recipient of that authority would be a human being, such as an officer of the United States, or on occasion, a private individual or group of individuals. . . . Yet if government actions should be undertaken by humans, then delegation to autonomously learning machines could potentially transfer governmental power outside the bounds that the Constitution permits.").

253. *See id.*

254. In Federal Courts and tribunals, the constraint may be a result of Article III of the United States Constitution, which vests the judicial power in the courts, which consist of (human) judges appointed for life. *See* U.S. CONST. art. III. Indeed, certain cases and controversies cannot be decided even by humans who are not appropriately appointed, and do not operate, within the requirements of Article III. *See* N. Pipeline Constr. Co. v. Marathon Pipe Line Co., 458 U.S. 50, 87 (1982) ("We conclude that . . . the Bankruptcy Act of 1978, has impermissibly removed most, if not all, of 'the essential attributes of the judicial power' from the Art. III district court, and has vested those attributes in a non-Art. III adjunct."); *cf.* Freytag v. Commissioner, 501 U.S. 868, 870 (1991) (holding that the appointment of special trial judge in Article I tax court did not violate separation of powers).

Strict liability may make sense for high-value applications. It may also make sense to bar AI from high-value applications entirely.

It is notable that strict liability has generally been rejected when opinion statements are involved. Indeed, in the Securities Act of 1933 Act context, false statements in prospectuses give rise to strict liability, but even in that context, courts have imposed a scienter-like requirement that the opinion statement be both subjectively and objectively false,[255] meaning that the speaker of the opinion intended to mislead with the opinion or did not genuinely believe the opinion.[256]

To be sure, there are several good reasons to reject strict liability in the opinion context. Opinions are often based on contradictory or incomplete information, so it is generally not enough that the speaker of an opinion know of information that contradicted his opinion to render the opinion false.[257] A doctor may know that some percentage of patients with a particular set of symptoms may have a completely different, perhaps more serious, diagnosis, but one would not say that the doctor should necessarily be liable if the diagnosis proves incorrect (and the alternative, more serious diagnosis turned out to be correct).

The determinative factor for liability is whether the doctor's judgment was reasonable under the circumstances.[258] Perhaps the probability of the alternative diagnosis was relatively low. Perhaps the patient had other characteristics that affected the relative probabilities. Or the doctor simply may have relied on his own experience to make a decision. None of these circumstances describe a doctor who has behaved improperly.[259] In other words, being wrong does not necessarily mean negligent or malicious—especially when individual judgment is involved.

---

255. *See supra* note 42 and accompanying text.

256. *See supra* notes 42, 58, and accompanying text.

257. *See supra* note 2.

258. This is because most malpractice cases will be negligence cases, which require some showing that the doctor breached the relevant standard of care. *See supra* note 197. For the most part, this will be a standard of care that is relative to other physicians and not the standard of care that applies to a lay person, as it may never be reasonable for a lay person to attempt to practice medicine without any acquired skill or training. *See* Charles R. Korsmo, *Lost in Translation: Law, Economics, and Subjective Standards of Care in Negligence Law*, 118 PENN. ST. L. REV. 285, 327 (2013) ("The illusion that skilled professionals are held to a 'higher' standard of care for a given activity is maintained only by ignoring the requirement that unskilled laypeople avoid the professional activity altogether.").

259. That is, unless the doctor relying on his own experience does so without having adequate experience. *See, e.g.*, Andersen v. Khanna, 913 N.W.2d 526, 537 (Iowa 2018) ("We conclude the district court erred when it found, as a matter of law, there is no duty to disclose personal characteristics, such as experience and training, under Iowa law.").

A strict liability rule for a wrong diagnosis opinion in this hypothetical case would be oppressive. It would be impossible for doctors to operate under such conditions, because they practice in a field where incomplete or probabilistic information is sometimes all that is available.[260] It would also incentivize the doctor to defensively attempt to rule out improbable diagnoses, slowing down treatment and perhaps in many cases increasing the costs.[261] The net effect would be to cripple the exercise of judgment by an expert—but most of the time, the expert's judgment and trained intuition is precisely what a patient seeks from the expert.

AI opinions are differently situated. Although AI shares some characteristics with a trained human expert, such as the ability to make judgments based on intuition, experience, and training, an AI's incentives to be thorough do not change with liability rules. For example, an AI will likely not practice defensive medicine. That is, AI operating in a strict liability regime will make the same predictions—based on data—regardless of whether it will be subject to strict liability. This eliminates some of the major policy problems with strict liability.

And the person deploying the AI can decide whether the potential of being strictly liable is worth it.[262] Under a strict liability regime, a person or company deploying the AI may think twice before using the AI to autonomously make valuation decisions for high-value assets, but may decide that the AI's opinions on low value items are worth the tradeoff of being strictly liable. Notably, the focus will be on the decision to use the AI, not on how the AI arrives at its opinions.[263]

---

260. Although medical malpractice or misdiagnosis is an insurable risk, which satisfies one of the preconditions for a viable strict liability regime, there can be no legitimate set of physician conduct in the ordinary course of care that could be a priori defined as prohibited. *See* Henderson, *supra* note 248, at 391. A diagnosis or medical test, for example, cannot be deemed to give rise to strict liability simply because in some cases it causes injury. The context in which any given medical test or diagnosis is used will vary greatly from case to case.

261. *See The Medical Malpractice Threat: A Study of Defensive Medicine*, 1971 DUKE L.J. 939, 943 (1971) (noting that in response to a ruling finding liability, "a physician may go far beyond the court-established standard by performing procedures which are neither legally nor medically required in order to guarantee that no hidden problems have been overlooked which might otherwise have become the basis of a malpractice suit").

262. This assumes that risks are independent and ascertainable, such that the amount of insurance necessary can be determined with some regularity. *See* Mark A. Geistfeld, *Interpreting the Rules of Insurance Contract Interpretation*, 68 RUTGERS U. L. REV. 371, 383–91 (2015) (noting that insurable risks must be independent across policyholders for the insurer to be able to predict and distribute risk across a pool of policies).

263. This is because how the AI makes decisions may be off limits due to the Black Box Problem, so specific conduct cannot be defined as a priori subject to strict liability. Rather,

D. WHY DISCLOSURE RULES ARE LESS EFFECTIVE IN THE CASE OF AI OPINIONS

In many opinion cases, disclosure of the basis of the opinion will preclude liability because it will be difficult to contend material information has been omitted or that an affirmative statement is misleading. With respect to Black Box AI, a disclosure cannot include the basis of the opinion or even a set of material facts,[264] so any disclosure will be limited to characteristics of the AI, such as how it was trained.[265] Accordingly, this Section argues that disclosure rules will generally not be effective in the case of AI opinions.

1. Disclosure in the Non-AI Opinion Context

All opinion statements suffer from the risk that some piece of contradictory information was improperly weighed when the opinion was made.[266] There is also the risk that incentives or biases played an improper role in the decision-making process.[267] That risk is precisely why a disingenuously held opinion can give rise to liability—it's usually the sign of bias or some incentive contrary to providing a truthful opinion.[268]

One way to fix the problem is to disclose everything about the decision-making process.[269] A valuation opinion that explicitly states what was important to the decision-making process and why will generally be more valuable than one that simply states a conclusion.[270] The reason is that the person hearing the opinion can evaluate the soundness of the opinion and determine what aspects of the opinion's model of the world are more or less correct.

---

only the specific use or application of the AI can be deemed subject to strict liability. *See supra* note 248.

264. *See supra* Part III.

265. *See supra* Section IV.A.2.

266. *See* Omnicare, Inc. v. Laborers Dist. Council Constr. Indus. Pension Fund, 135 S. Ct. 1318, 1328 (holding that opinion is not false simply because one can "second-guess inherently subjective and uncertain assessments").

267. *See supra* note 67 and accompanying text.

268. *See supra* note 2.

269. *See, e.g.*, *In re* Donald J. Trump Casino Sec. Litig., 7 F.3d 357, 371 (3d Cir. 1993) ("[C]autionary language, if sufficient, renders the alleged omissions or misrepresentations immaterial as a matter of law.").

270. In some cases, the opinion may be on a matter that is of such importance or complexity that a reasonable person would have expected to exercise some diligence as to the basis for the opinion before the opinion was stated. *See Omnicare*, 135 S. Ct. at 1330 (in some cases a reasonable person would not expect the opinion "to reflect baseless, off-the-cuff judgments, of the kind that an individual might communicate in daily life").

A judicial opinion, for example, articulates not only the outcome the judge has reached, but also explains how the judge reached that outcome. A real estate appraisal will often have the most pertinent facts set out within a report. If there has been disclosure, it is simply much less likely that the person hearing the opinion has been misled.

2. Disclosure Will Be Less Effective in the AI Context

When AI is concerned, disclosure of all of the parameters used by the AI may be of little value. Unlike humans, AI can simultaneously weigh significantly more information,[271] but disclosure of what information is provided to the AI will not likely make much of a difference to a person considering the AI's decision or opinion.[272] For example, medical AI that evaluates 3,000 different patient characteristics is not any less opaque if those 3,000 characteristics are disclosed.

There is also no way to succinctly explain how each parameter has been weighed by the AI if the AI suffers from the Black Box Problem.[273] There will generally be no way of strictly rank ordering the model's inputs in terms of their effect on the ultimate opinion.[274] Indeed, a particular parameter may only be relevant if hundreds of others bear some threshold characteristic, and may be much less relevant when those other parameters are not above that threshold.[275] Was the patient's diet relevant to the diagnosis? The answer may be that it depends on a host of other factors, so is diet more important than say, liver function? There will often be no strict rank ordering.[276]

Assuming that the AI model suffers from the Black Box Problem, disclosure will generally not mitigate the opaqueness of the overall opinion. This is completely different than in the case where a human exercises judgment—the human can provide some explanations and provide a rough ordering of what factors were most important.[277] That may not be possible when AI is involved.[278] This is why disclosure rules are likely to be ineffective when AI is concerned.

---

271. *See supra* Section III.B.

272. *See id.*

273. *See id.*

274. *See id.*

275. This may result on a neural-layer-scale, as each artificial neuron layer is coupled with a non-linear activation function. One popular non-linearity used as an activation function is a rectifier, which passes a scaled output from the neural layer onto the next layer if the signal is above some threshold; if the signal is below some threshold, the activation function passes no signal to the next layer. *See* GOODFELLOW ET AL., *supra* note 113, at 187.

276. *See supra* Section III.B.

277. *Id.*

278. *Id.*

E. WHEN CAN YOU INFER USER OR CREATOR INTENT FROM AN AI MODEL'S OPINION?

Given that there are several factual heuristics that can still be used to evaluate culpability on the part of the person deploying the AI—(1) training, validation and testing,[279] (2) deference and autonomy,[280] and (3) constraints and conscientiousness[281]—is it therefore possible that one can infer scienter based on these heuristics? The answer is likely yes in many cases, but with the important caveats described in this Section:

- Data bias will be more important in omissions cases;[282]
- Certain applications should be subject to strict liability;[283] and
- Disclosure is likely irrelevant in solving the liability issue.[284]

With these caveats in mind, it is possible to infer a very specific form of scienter if the heuristics imply culpability. At the extreme, an AI that received complete autonomy and little supervision without adequate training or validation while operating with no a priori constraints (not supervision, but deterministic constraints in its programming) will likely imply that the person who created or deployed the AI was at least reckless for doing so.[285] Thus, it may be that the factual heuristics described in this Part can give rise to an inference of scienter.

F. PUTTING IT ALL TOGETHER: SCIENTER SHOULD BE SUFFICIENT, NOT NECESSARY FOR OPINION LIABILITY

The caveats noted above, however, make clear why scienter should not always be required to impose liability. It may be that a model is correctly trained and tested, provided the right amount of human supervision, and given several constraints to address potential known risks, but that a latent bias in the data caused the AI to render improper opinions.[286] In such a case, none of the heuristics will allow a factfinder to infer scienter. And because of the Black Box Problem, there will generally be no evidence that the person deploying the AI intended to mislead or subjectively disbelieved the AI's opinions.[287]

---

279. *See supra* Section IV.A.2.
280. *See supra* Section IV.A.1.
281. *See supra* Section IV.A.3.
282. *See supra* Section IV.B.
283. *See supra* Section IV.C.
284. *See supra* Section IV.D.
285. *See* Bathaee, *supra* note 7, at 932–38.
286. *See supra* Section IV.C.
287. *See supra* Section III.B.

In such a case, the important question is one of negligence—why did the person deploying the AI miss the data bias? Should the data bias have been studied? Should the dataset upon which the AI was trained have been described or even disclosed? All of this depends on context.

Worse yet, none of these questions should matter at all if the risk of loss is sufficiently high.[288] If an AI malfunction could cause hundreds of deaths, perhaps it was a poor use case for the AI and no amount of precaution, conscientiousness, or supervision should absolve the person deploying the AI of liability.[289]

In other words, liability should certainly attach when there is scienter or where scienter can be inferred from precise heuristics (such as those described in this Article). But it is clear that requiring scienter as a necessary element would completely insulate a wide swath of AI from liability entirely.[290] It would also allow AI to sanitize human conduct that would otherwise give rise to liability.[291]

It is important to separate AI opinions from human opinions when liability is concerned and to apply a set of specialized rules and heuristics to AI opinions. Scienter may work well for humans, but it cannot be the requirement for AI, as that would mean virtual immunity from liability when AI is involved.

V. CONCLUSION

Opinion statements are generally not provably true or false.[292] They are functionally models of reality and are based on a set of facts—actual or assumed—that are considered together.[293] That is why the law has focused on intent-based heuristics, such as scienter, to assign liability based on opinion statements.[294] However, AI potentially decouples the connection between the opinion and the speaker of the opinion.[295] It also obfuscates the factual basis and fact weighting that is the basis for the opinion.[296] This obfuscation arises from AI's Black Box problem, which stems from the inherent connective and complex structure of the machine-learning algorithms used to build AI systems.[297] All of this means that intent will often not be inferable by simply

288. *See supra* Section IV.C.
289. *See id.*
290. *See supra* Part III.
291. *See id.*
292. *See supra* Part I.
293. *See id.*
294. *See id.*
295. *See supra* Part III.
296. *See id.*
297. *See id.*

examining the AI, as it would be in the case of a deterministic, instruction-based computer program.[298]

A new set of heuristics are necessary to determine whether a person who deploys an AI system should be responsible for the harm caused by it.[299] Those heuristics are more precise than conventional intent-based heuristics—they address how the model was constructed, constrained, and supervised.[300] These heuristics may not point towards an improper intent on the part of the AI's creator or user, but they are more precise—that is, they are heuristics that assume extensive training and testing of the AI, context-driven decision making as to the necessary level of supervision for the AI, and whether the appropriate constraints were put in place a priori.[301]

There are also contexts in which these heuristics do not strongly suggest that any person intentionally designed the AI in a manner that is at all culpable.[302] In these contexts, there may still be a case for liability.[303] When an AI is deployed in a context that has a high risk of harm or in which societal norms demand human judgment, strict liability may be appropriate—even though strict liability regimes prove problematic in the case of human opinion statements.[304]

It may also be the case that although the AI has been appropriately trained, deployed, and supervised, there was a significant amount of bias in the data used to train it.[305] In such a case, none of the scienter-like heuristics may point towards liability, but liability may nonetheless be appropriate.[306]

To be sure, courts must wrestle with a new set of liability heuristics as well as the significant policy judgments they implicitly come with and will need to gain significant amounts of experience with AI systems before a viable opinion liability regime emerges, but what is almost certain is that existing scienter rules for opinion liability would excuse a wide swath of AI and AI-assisted opinions from liability.[307] This is a result worse than a lack of regulation—it is tantamount to an unintended immunity from opinion liability entirely.

298. *See id.*
299. *See supra* Section IV.A.
300. *See id.*
301. *See id.*
302. *See supra* Sections IV.B–F.
303. *See supra* Section IV.F.
304. *See supra* Section IV.C.
305. *See supra* Section IV.B.
306. *See supra* Sections IV.B & IV.F.
307. *See supra* Part III.