Legal Opacity: Artificial Intelligence's Sticky Wicket

Charlotte A. Tschider
Loyola University Chicago School of Law, ctschider@luc.edu

Legal Opacity: Artificial Intelligence's Sticky Wicket

Charlotte A. Tschider[*]

ABSTRACT: Proponents of artificial intelligence ("AI") transparency have carefully illustrated the many ways in which transparency may be beneficial to prevent safety and unfairness issues, to promote innovation, and to effectively provide recovery or support due process in lawsuits. However, impediments to transparency goals, described as opacity, or the "black-box" nature of AI, present significant issues for promoting these goals.

An undertheorized perspective on opacity is legal opacity, where competitive, and often discretionary legal choices, coupled with regulatory barriers create opacity. Although legal opacity does not specifically affect AI only, the combination of technical opacity in AI systems with legal opacity amounts to a nearly insurmountable barrier to transparency goals. Types of legal opacity, including trade secrecy status, contractual provisions that promote confidentiality and data ownership restrictions, and privacy law independently and cumulatively make the black box substantially opaquer.

The degree to which legal opacity should be limited or disincentivized depends on the specific sector and transparency goals of specific AI technologies, technologies which may dramatically affect people's lives or may simply be introduced for convenience. This Response proposes a contextual approach to transparency: Legal opacity may be limited in situations where the individual or patient benefits, when data sharing and technology disclosure can be incentivized, or in a protected state when transparency and explanation are necessary.

[*] Charlotte A. Tschider, Assistant Professor of Law, The Loyola University Chicago School of Law, and the Beazley Institute for Health Law & Policy. I would like to thank the Loyola University Chicago School of Law faculty, who offered unique perspectives on the concept of unfairness and discrimination in AI. I would like to especially thank W. Nicholson Price II for his thought-provoking work with Arti K. Rai in *Clearing Opacity Through Machine Learning*, to which this piece is a response and for his generous review of this drafted response. I would also like to thank Dr. Krista Kennedy: their support, collaboration, and multi-disciplinary, thoughtful perspective has been instrumental in exploring legal opacity and potential solutions to endemic problems, specifically in healthcare use cases.

I. INTRODUCTION

Artificial intelligence ("AI") is a technology used in any number of sectors, not just healthcare. AI can streamline the employment interview process,[1] automate cafeteria ordering and prepare meals,[2] enable driverless delivery,[3] reduce cost in transportation logistics,[4] improve product personalization,[5] or increase medical diagnostic efficacy.[6] AI is built on big data, data that may be personally identifiable, confidential, both, or simply public data that an organization has invested considerable time to collect and organize.[7]

Depending on the data and the system underlying the AI's goals, technical issues may result in societal consequences and corresponding legal repercussions. For example, an AI system may eliminate certain job applicants from an employment search based on a combination of resume information and a video interview analysis.[8] A specific prison sentence could be recommended to a judge based on data gathered from current and former prisoners, as well as recidivism data.[9] Government agencies could allocate

[1] Patricia Farrell, *AI's Secrets Behind That Job Interview: It's Not What You Did or Who You Are but What AI Says About You*, MEDIUM (Nov. 9, 2019), https://drpatfarrell.medium.com/ais-secrets-behind-that-job-interview-it-s-not-what-you-did-or-who-you-are-but-what-ai-says-about-53e9858a64f9 [https://perma.cc/9AMU-EF3A].

[2] Kumba Sennaar, *Examples of AI in Restaurants and Food Services*, EMERJ (Jan. 31, 2019), https://emerj.com/ai-sector-overviews/ai-in-restaurants-food-services [https://perma.cc/778Z-B88A].

[3] *The Future of Delivery is Self-Driving*, DOMINOS, https://www.selfdrivingdelivery.dominos.com/en [https://perma.cc/H247-CBYA].

[4] Vitaly Kuprenko, *Artificial Intelligence in the Logistics Industry*, THE NETWORK EFFECT (Sept. 23, 2019), https://supplychainbeyond.com/artificial-intelligence-in-the-logistics-industry [https://perma.cc/55L7-SZEP].

[5] Andrew Pearson, *AI — A Personalization Engine on Steroids*, MEDIUM (May 18, 2021), https://medium.com/product-ai/ai-a-personalization-engine-on-steroids-d9ed59ec44f9 [https://perma.cc/FFV7-FZX5].

[6] Andrew Wade, *How AI is Powering a Revolution in Medical Diagnostics*, THE ENGINEER (Feb. 12, 2019, 1:08 PM), https://www.theengineer.co.uk/ai-medical-diagnostics [https://perma.cc/UZ8U-WFH5].

[7] Peter Leonard, *Beyond Data Privacy: Data "Ownership" and Regulation of Data-Driven Business*, SCITECH LAWYER (Jan. 17, 2020), https://www.americanbar.org/groups/science_technology/publications/scitech_lawyer/2020/winter/beyond-data-privacy-data-ownership-and-regulation-datadriven-business [https://perma.cc/S9M2-J2YC].

[8] Ben Dattner, Tomas Chamorro-Premuzic, Richard Buchband & Lucinda Schettler, *The Legal and Ethical Implications of Using AI in Hiring*, HARV. BUS. REV. (Apr. 25, 2019), https://hbr.org/2019/04/the-legal-and-ethical-implications-of-using-ai-in-hiring [https://perma.cc/NTB2-NA7N].

[9] Vyacheslav Polonski, *AI Is Convicting Criminals and Determining Jail Time, but is it Fair?*, WORLD ECON. F. (Nov. 19, 2018), https://www.weforum.org/agenda/2018/11/algorithms-court-criminals-jail-time-fair [https://perma.cc/LC5P-DFG4].

entitlement amounts using an AI tool.[10] Any decisions previously made by humans could conceivably be automated by AI in the future.[11] For this reason, regulating the technology and promoting innovation without context or corresponding potentialities is nearly impossible.[12] Two key issues seem to permeate healthcare AI and a variety of other AI-use cases: first, whether AI is reliable, safe, and fair; and second, how the AI algorithm rendered a decision.[13] For example, the U.S. public may wish to ensure AI products on the market will not malfunction or render discriminatory results. In court cases, it may be desirable to understand how a decision was reached, for example when establishing proximate cause for a negligence claim. Furthermore, data used to create AI could be beneficial for other uses that promote innovation.

One key impediment to addressing these issues is *opacity*, or the lack of transparency in AI decisions. Although advocates of AI transparency frequently demand the AI algorithm be explained, various AI choices affect whether the AI is reliable, safe, and fair, such as technology infrastructure, AI design approach and goals, process and training choices, testing, and data selection and structure reflect human choices that are intentionally not disclosed to the public.[14]

As explained by Frank Pasquale, opaque algorithms are both incomprehensible, not able to be explained, *and* secret.[15] According to Frank Pasquale's foundational text, *The Black Box Society*, secrecy includes both real secrecy (or functional technical secrecy) and legal secrecy, created through obligation, such as contract.[16]

Legal secrecy may include various legal mechanisms that cause information to not be disclosed to the public, such as privacy law, private law confidentiality provisions and contractual data use restrictions, and trade secrecy. Each of these legal vehicles promote opacity to various degrees, but

[10] DAVID FREEMAN ENGSTROM, DANIEL E. HO, CATHERINE M. SHARKEY & MARIANO-FLORENTINO CUÉLLAR, GOVERNMENT BY ALGORITHM: ARTIFICIAL INTELLIGENCE IN FEDERAL ADMINISTRATIVE AGENCIES 37–38 (2020), https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf [https://perma.cc/H8W3-V5Z5].

[11] Thomas C. Linn, *Anything You Can Do, A.I. Can Do Better?*, BALTIMORE SUN (Nov. 24, 2017, 6:00 AM), https://www.baltimoresun.com/opinion/op-ed/bs-ed-op-1126-artificial-intelligence-20171120-story.html [https://perma.cc/S8AG-HVYR].

[12] Indeed, just as context is important for building reliable, safe, and fair AI, context is also important for determining the appropriate legal model to balance innovation with individual interests.

[13] Charlotte A. Tschider, *Beyond the "Black Box"*, 98 DENV. L. REV. 683, 685–88 (2021).

[14] *Id.* at 690–91, 693; W. Nicholson Price II, *Black-Box Medicine*, 28 HARV. J.L. & TECH. 419, 433 (2015).

[15] FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 6–7 (2015). Danielle Keats Citron and Frank Pasquale similarly note that opacity also prevents review by "regulators charged with protecting" those affected by automated processing. *See* Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 10 (2014).

[16] *See* PASQUALE, *supra* note 15, at 6.

collectively illustrate how purely voluntary legal choices lead to opacity issues. The way in which vehicles of legal secrecy promote AI opacity will be described as a new term, *legal opacity*.

This Response makes two novel contributions to the legal literature on algorithmic opacity. First, this Response explores an undertheorized aspect of opacity, *legal opacity*, wherein legal mechanisms are used to prevent disclosure about an AI system. Second, this Response proposes strategies to satisfy the twin goals of data necessity through disclosure and justifiable non-disclosure, or *data essentialism*.[17] I aim to more fully illustrate the cards stacked against advocates for openness, transparency, and innovation while also exploring potential solutions that balance interests to prevent disclosure and interests to disclose.

In 2021, the *Iowa Law Review* published an insightful piece by Nicholson Price and Arti Rai, *Clearing Opacity Through Machine Learning*.[18] The essay explored the problem of algorithmic opacity, in particular algorithmic opacity for artificial intelligence, proposing the contours of a solution that could solve two key goals: 1) overcoming opacity—the inability to know how an algorithmic decision is made—to solve any number of legal challenges arising from the black-box nature of some artificial intelligence applications;[19] and 2) encouraging AI openness to spur innovation in access to data and explanation that could transform the next wave of scientific development.[20] In it, Price and Rai adopted an approach inspired by innovation scholarship that is applied to health care technology.[21]

In response to Price and Rai's excellent work, I aim to explore the legal hurdles that must be cleared in order to legally and more fully "clear opacity."[22] Specifically, I examine legal barriers in the context of data *essentialism*: that high-volume, representative, and organized big data sets are central to AI innovation and ongoing development.

Part II describes opacity's problem as explained by Price and Rai, including expansion of the term to include other private legal doctrine, such as contractual data ownership and confidentiality. Part III turns to probable scenarios, broadly construed, where disclosure may be desirable or necessary. Part III examines organizational restrictions on contractual data ownership, confidentiality, and trade secrecy with exploration into competing goals of privacy law and AI. Part IV introduces potential solutions, specifically contextual and nuanced legal strategies, to navigate the need for technology

[17] Wendy Netter Epstein & Charlotte Tschider, *We Need to Do More with Hospitals' Data, but There Are Better Ways*, HARV. L. PETRIE-FLOM CTR.: BILL OF HEALTH (Jul. 7, 2021), https://blog.petrieflom.law.harvard.edu/2021/07/07/hospital-data-big-tech [https://perma.cc/5K6E-8H3L].

[18] W. Nicholson Price II & Arti K. Rai, *Clearing Opacity Through Machine Learning*, 106 IOWA L. REV. 775 (2021).

[19] *See id.* at 779.

[20] *Id.* at 779, 781.

[21] *Id.* at 781–84.

[22] *See id.* at 778–80.

and data disclosure in various respects, intended to spur a fuller discussion. Solutions should balance individual privacy interests, contractual data rights, and trade secrecy with data essentialism while acknowledging the strong federal and state support of maintaining trade secrecy and private contracting for confidentiality and data control.

II. OPACITY AND THE CHALLENGE OF ESTABLISHED LEGAL DOCTRINE

Although opacity is typically expressed as opacity of processes and data, legal opacity introduces additional challenges to AI disclosure and transparency goals. Understanding not only the nature of opacity but also the goals of transparency are essential to determining the appropriate framework for enhancing innovation while protecting individual interests, such as privacy, fairness, and safety.

A. UNDERSTANDING OPACITY AS ONE EXAMPLE OF REDUCED ACCESS

Although opacity may prevent useful disclosure or hamper innovation, it is important to point out that opacity, in its broad sense, is a longstanding problem. Opacity is not that different from the legal concept of *confidentiality*, where access to relevant and sometimes important information is purposely not available to certain untrusted parties, established through private contract or internal policy.[23] Confidentiality, as it is operationalized in private contracting, has important functions too. Banking system technology may contain data that are not personal information yet maintaining confidentiality may nevertheless prevent a cyberattack that could compromise customer accounts.[24]

Confidentiality can also reduce the likelihood of personal information misuse, such as preventing important details about a patient's medical condition from being shared with the broader community. Overall, confidentiality is the mechanism of keeping information private, whatever its status—personal information, proprietary information, or trade secret.[25] For this reason, confidentiality overlaps with information privacy and trade secrecy, though confidentiality obligations are largely communicated in contract.[26]

Organizations often choose trade secrecy as a method to protect organizational investment in AI, which may be a weak or strong form of

[23] *See generally* Charles Tait Graves & Sonia K. Katyal, *From Trade Secrecy to Seclusion*, 109 GEO. L.J. 1337 (2021) (discussing how confidentiality is established in the context of trade secrets).

[24] *Information Security | Confidentiality*, GEEKSFORGEEKS (May 9, 2019), https://www.geeksforgeeks.org/information-security-confidentiality [https://perma.cc/Y9RL-Z6TR].

[25] *Id.*

[26] CHARLOTTE A. TSCHIDER, INTERNATIONAL CYBERSECURITY AND PRIVACY LAW IN PRACTICE 186–87, 378 (2018).

protection depending on *what* and *against whom* protection is desired.[27] The foundation for trade secrecy is unfair competition,[28] wherein an individual who uses improper means to take and use the trade secret can be held accountable under trade secret misappropriation.[29] If an organization takes reasonable steps to keep a trade secret, well, secret, and a person (or a downstream organization receiving the secret) takes and uses the information through improper access, the trade secret owner can recover under state or federal law.[30] AI is a natural fit for trade secret protection due to the unavailability of alternatives like patent law and the natural opacity of its processes and algorithms.[31]

Confidentiality obligations are usually enforced via common law contract actions in state and federal court using state choice of law,[32] while trade secrecy is enforced through federal and state statutes.[33] If an individual or another organization, such as a third party, exceeds their contractually-specific data uses, the individual or organization may face breach of contract allegations or trade secrecy legal issues.[34] For example, a contract may specify that a third party is responsible for maintaining an investment system that contains stock purchases and sales of shareholders. However, this does not necessarily (and often does not) permit the third party to access or use the data in these systems for their own purposes. Doing so would exceed their contractual authority.[35]

[27] *See* Price & Rai, *supra* note 18, at 791, 793. Trade secrecy can be construed as "low cost," though meeting requirements to invest reasonable means for keeping such information secret certainly can be expensive. Foundationally, some information may not be able to be maintained as secret due to its broad use, relying instead on broad personnel confidentiality provisions and internal policy. Trade secrets still require additional investment and reasonable notice to be legally defensible trade secrets. *See* Tom Kulik, *NDAs & How to Lose Your Trade Secrets Without Really Trying*, ABOVE THE LAW (Dec. 11, 2018, 11:45 AM), https://abovethelaw.com/2018/12/ndas-how-to-lose-your-trade-secrets-without-really-trying [https://perma.cc/EA3G-TABT].

[28] *See* Sharon K. Sandeen, *The Evolution of Trade Secret Law and Why Courts Commit Error When They Do Not Follow the Uniform Trade Secrets Act*, 33 HAMLINE L. REV. 493, 495 (2010).

[29] *See id.* at 529.

[30] *See* Michael J. Kasdan, Kevin M. Smith & Benjamin Daniels, *Trade Secrets: What You Need to Know*, THE NAT'L L. REV. (Dec. 12, 2019), https://www.natlawreview.com/article/trade-secrets-what-you-need-to-know [https://perma.cc/29UR-2KWE].

[31] *See* Tschider, *supra* note 13, at 700, 711–13.

[32] *See* Jordan Porter, *Determining Choice of Law in Civil Litigation*, JD PORTER LLC LEGAL PRACTICE (2016), https://www.jdporterlaw.com/285-2/determining-choice-law-civil-litigation [https://perma.cc/U84J-GSWS].

[33] *See* Sharon K. Sandeen & Christopher B. Seaman, *Toward a Federal Jurisprudence of Trade Secret Law*, 32 BERKELEY TECH. L.J. 829, 840–43 (2017).

[34] *See* Deepa Varadarajan, *The Trade Secret-Contract Interface*, 103 IOWA L. REV. 1543, 1560–61 (2018); Douglas R. Nemec, *Trade Secrets Take Center Stage, and Contracts Play a Lead Role*, SKADDEN'S 2019 INSIGHTS (Jan. 17, 2019), https://www.skadden.com/insights/publications/2019/01/2019-insights/trade-secrets-take-center-stage [https://perma.cc/595L-Q95G].

[35] *See* Varadarajan, *supra* note 34, at 1560–62 (observing that although the existence of confidentiality provisions for individuals is persuasive, it is not necessarily dispositive for notice of trade secrecy). Confidentiality provisions establish one requirement related to protection of

As Price and Rai suggest, it is important to acknowledge that broader issues of opacity, such as access to data or non-intuitive results for research innovation, is not necessarily *only* a matter of opacity.[36] Legal secrecy, although usually referencing trade secrecy status[37] also concerns thornier issues of contractual data ownership rights, confidentiality obligations, and a strong U.S. predisposition to protect trade secrets, all of which reinforce the desire of organizations to control their investments and occupy the market.[38] These forms of legal secrecy, though increasing opacity, also protect substantial investments in the creation, acquisition, collation, organization, and combination of the massive data sets necessary to run the reliable, fair, efficacious, safe, and efficient AI systems.[39] Problems emerge when legal secrecy's benefits to an organization outweigh the interests of the public.[40]

B. SYSTEM AND TOOL OPACITY V. ELECTIVE (LEGAL) OPACITY

Opacity is, as Price and Rai contend, a matter of technological choices, or *system* and *tool* opacity, and opacity by choice, or deliberate secrecy, which often overlap.[41] Tool opacity, for example, is a combination of algorithmic complexity and non-intuitiveness, where both the inputs into a decisional system and the decisional outcomes and logic behind such outcomes are not readily understandable by experts.[42] As I have described along the same lines, complexity includes not only inscrutability, or algorithmic complexity and non-intuitiveness, but also *dynamism*, or the self-learning capability of some AI machine learning utilities, or unlocked unsupervised learning.[43] Tool opacity can also include deliberate secrecy efforts, such as keeping technical details of algorithmic development secret.[44]

information generally and notice of legitimate data uses. Trade secrecy generally requires more and different notice, though information or data that is a trade secret is a subset of generally confidential data. Thadford Felton, *The Differences Between 'Confidential' and 'Trade Secret' Information, and Why They Matter*, GREENSFELDER: IMPACT | BUSINESS RISK MANAGEMENT BLOG (Feb. 1, 2016, 1:17 PM), https://www.greensfelder.com/business-risk-management-blog/the-differences-between-confidential-and-trade-secret-information-and-why-they-matter [https://perma.cc/N48T-PBQA].

[36] Price & Rai, *supra* note 18, at 793–94.

[37] *See* PASQUALE, *supra* note 15, at 6.

[38] *See* JOHN R. THOMAS, CONG. RSCH. SERV., THE ROLE OF TRADE SECRETS IN INNOVATION POLICY 1 (Jan. 15, 2014), https://fas.org/sgp/crs/secrecy/R41391.pdf [https://perma.cc/2JFK-TXT9].

[39] *See* John Bantleman, *The Big Cost of Big Data*, FORBES (Apr. 16, 2012, 1:21 AM), https://www.forbes.com/sites/ciocentral/2012/04/16/the-big-cost-of-big-data/?sh=765daf065a3b [https://perma.cc/8JAS-UZQD].

[40] *See* W. Nicholson Price II, *Regulating Secrecy*, 91 WASH. L. REV. 1769, 1784 (2016).

[41] *See* Price & Rai, *supra* note 18, at 784.

[42] *Id.*

[43] *See* Tschider, *supra* note 13, at 689–99.

[44] *See* Price & Rai, *supra* note 18, at 784.

Tool opacity, however, is not system opacity. System opacity refers to underlying real-world systems in which AI performs.[45] For example, a system may be the biological mechanisms that cause a specific type of brain cancer or the pharmacological approach to treating this type of cancer.[46] The method for explaining these systems will likely involve scientific investigation to determine cancer markers—including genetic screening, data analysis, pharmacokinetic testing, and a series of clinical testing phases—though an excellent cancer diagnostician identifies a probability of cancer without being able to explain exactly how the diagnosis was made.[47]

In contrast, tool opacity likely means that the AI used to diagnose this type of brain cancer may determine a patient has brain cancer with 92 percent certainty, but depending on the machine learning model used, AI architects or the oncologists using this technology may not understand how the diagnosis is rendered.[48] Further, the methods, training data, and infrastructure of the system may be protected using legal secrecy strategies by the organization creating the tool.

This distinction is incredibly useful in determining whether and to what extent mandating transparency will meaningfully add value or not. As Price and Rai rightly mention, some of the same aspects that increase tool opacity, such as non-intuitiveness, have important roles in reducing system opacity.[49] Non-intuitiveness is a condition that occurs when the results do not make sense.[50]

For example, an AI system may successfully predict breast cancer from mammogram images 95 percent of the time but when probed further, the reason for the 95 percent success rate may be because the AI utilities correlated a light spot on some images artificially and improperly introduced by the digital machine. In this case, the field recording probable thickening of tissue might actually have nothing to do with legitimate field imaging details at all. The non-intuitiveness aspect could force AI developers to tune and correct an AI system to improve its efficacy but may also provide important information to researchers studying breast cancer, such as what qualities the artificial spot might have accompanied in the principal image.

C. When AI Explanation is Necessary or Highly Useful

In my article, *Beyond the Black Box*, I challenged whether mandated explanation is the most effective way to think about broadly preventing issues

[45] *Id.* at 779–94.

[46] *See* Price, *supra* note 14, at 433–34.

[47] *Id.*

[48] *See* Will Knight, *The Dark Secret at the Heart of AI*, MIT TECH. REV. (Apr. 11, 2017), https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai [https://perma.cc/5FNL-ZNAA].

[49] *See* Price & Rai, *supra* note 18, at 794–95.

[50] *Id.* at 784.

related to AI opacity and legal secrecy.[51] In it, I suggested that disclosure of information other than an explanation of the decision would likely provide more useful information than broad algorithmic explanation to improve AI safety and reduce the potential for discrimination and unfairness.[52] This additional information includes: whether the AI will be technically supervised or unsupervised, whether deep learning will be used, and if the algorithm(s) will be locked in final release form; details about data volume, structure, diversity, representation; the type of AI and the process for creating it, including overall structure and decisional layers; testing procedure and plans for ongoing tuning and improvement; and the technical infrastructure.[53]

However, the disclosive value of these AI system inputs does not mean explanation will never be useful or that it is not worth exploring how explanation could be achieved.[54] Achieving goals of fairness and safety may require a different model of disclosure other than fulfilling, for example, innovation goals. It is crucial to understand what the goals are before designing a legal model for achieving them.

There is a myriad of situations where some degree of disclosure in automated decision-making will be necessary to protect individuals or promote social benefit. It is crucial, however, to understand the variety of situations where explanation is desirable, important, or essential. Why? Because mandating, compelling, or incentivizing explanation will likely destroy trade secrecy status, compromise confidentiality, or impact data ownership interests in private contracts.

Therefore, when considering when and under what conditions AI transparency is desirable, it must be weighed against the value of trade secrecy, confidentiality, and private contracting in innovation and investment-generating legal tools.[55] It must also be balanced with privacy interests of individuals who might be identifiable in data disclosure or algorithmic explanation.[56] Under some circumstances, the strong public interest will understandably weigh against preserving legal opacity. For others, the public interest may not be as compelling. If the United States intends to invalidate forms of legal protection without a useful alternative, we should clearly understand the various markets and sectors affected, including how disclosure will affect innovation behavior and other public interests.

[51] *See* Tschider, *supra* note 13, at 688.

[52] The concern in mandating outright algorithmic disclosure, broadly, is that organizations might provide such an explanation, and that that explanation may not be particularly useful in meeting the goal of less discriminatory, safer AI technologies. *Id.* at 700.

[53] *Id.* at 693–97.

[54] *Id.* at 723–24.

[55] *Id.* at 715–19.

[56] *See* Price & Rai, *supra* note 18, at 790, 810.

1. AI Non-Intuitiveness Valuable for Scientific Development and Product Efficacy

As Price and Rai describe in great detail, non-intuitiveness can prompt explanations that are very important for understanding the broader system in which an AI functions, especially to field experts.[57] When an AI's decision or resulting function is non-intuitive, such explanation may be critically important, which could demonstrate some underlying issue in the technology or process, especially when scientific research must be reproduced.[58] It could also reveal some new discovery related to the system being studied, such as a new understanding for when insulin is most effectively delivered or how to most effectively protect pedestrians near a self-driving car. Price and Rai specifically describe the value of data sharing and explanation for scientific development in the health sector, where large scale, useful data may be difficult to aggregate and collect.[59] Without access to big data sets, important health advancements might never be achieved.[60]

2. AI Transparency for Preventing Injury and Demonstrating Proximate Cause

AI transparency may be necessary when the decision or resulting function may cause or may have caused injury to people or damage to property. If *ex-ante* regulatory oversight bodies (or delegated third parties) review AI systems for safety, organizations using AI may need to provide sample explanations or other information about the AI to illustrate a low risk of foreseeable injuries prior to clearance, which may be needed "to observe the regulated behavior."[61] Indeed, more active clearance processes, like those operated by the U.S. Food and Drug Administration ("FDA"), could specifically require certain disclosures, such as training data, prior to receiving drug or medical device clearance.[62]

For example, a countertop AI-enabled slow-cooker could foreseeably exceed its heat setting and increase risk of fire in a consumer's home. Disclosures of representative explanations along with process and technology details could be useful to demonstrate that the manufacturer has reasonably

[57] *Id.* at 794–97.

[58] *Id.* at 797.

[59] *Id.* at 798–99. Reproducibility is a significant problem in scientific research, and small, non-representative data sets worsen these issues. Unfortunately, patent law, as it stands, does not provide the type of disclosure needed to satisfy this goal.

[60] *Healthcare Big Data and the Promise of Value-Based Care*, NEW ENG. J. MED. CATALYST (Jan. 1, 2018), https://catalyst.nejm.org/doi/full/10.1056/CAT.18.0290 [https://perma.cc/6ABL-S9TY].

[61] Price, *supra* note 40, at 1785.

[62] Price & Rai, *supra* note 18, at 804. More comprehensive review processes and upfront disclosures might prevent downstream torts, especially for sectors like health where regulators have an opportunity to prevent later safety issues. Charlotte A. Tschider, *Medical Device Artificial Intelligence: The New Tort Frontier*, 46 BYU L. REV. 1551, 1568–73 (2021).

implemented appropriate technical steps to avoid injury.[63] Of course, *ex ante* regulatory disclosure of representative explanations does not necessarily mean AI will never actually cause downstream injury.[64]

Actual, historical explanation may be necessary when individuals have suffered bodily injury or when property damage has occurred. Regulatory oversight bodies may require organizations using AI to report injuries or property damage, which could encourage AI accountability.[65] For common-law tort actions, explanations may be essential to demonstrate factual causation and for useful inputs to proximate cause analyses.[66] Indeed, under many circumstances an explanation of how the AI made a decision may be necessary to justify or defend an undesirable outcome.[67]

3. AI Explanation to Resolve Contractual Disputes

Furthermore, explanation may be necessary when injury occurs between sophisticated legal entities. Some organizations will not create AI solutions on their own, relying instead on AI solution providers.[68] In contracts between organizations and technology solution providers, often provisions are added which designate non-performance or culpability when the technology does not work as intended.[69] When organizations rely on an AI solution provider, it is foreseeable (as with any other IT solution or contracting activity) that unanticipated results may occur while using an AI solution. For example, an organization manufacturing engines may use AI for supply chain management or to assemble the engine in a facility.

If the AI does not perform as anticipated, such as introducing a defect into the manufacturing process or routing raw materials to the wrong facility,

[63] Tschider, *supra* note 13, at 706–07.

[64] *See* Tschider, *supra* note 62, at 1568–72, 1603–04.

[65] *Id.* at 1604. Although post-market surveillance activities are common for the FDA and limited other agencies, they do not eliminate the need for tort recovery. However, they may improve AI products and the processes for validating their safety and efficacy.

[66] *See* Andrew D. Selbst, *Negligence and AI's Human Users*, 100 B.U. L. REV. 1315, 1344–46 (2020) (analyzing foreseeability in establishing proximate cause for AI injuries).

[67] Margot E. Kaminski, *Understanding Transparency in Algorithmic Accountability*, *in* THE CAMBRIDGE HANDBOOK OF THE LAW OF ALGORITHMS 121, 123 (Woodrow Barfield ed., 2021). Kaminski describes this as the need for "justification." *Id.*

[68] *See* Fernando Lucini, *What's the Best AI Strategy: Build, Buy, Partner?*, ACCENTURE (May 6, 2019), https://www.accenture.com/us-en/insights/digital/best-ai-strategy-build-buy-partner [https://perma.cc/4BKR-KMK9]; *Should You Build or Buy Your AI?*, FORBES (May 22, 2019, 2:55 PM), https://www.forbes.com/sites/insights-intelai/2019/05/22/should-you-build-or-buy-your-ai/?sh=1cfbc5fa441d [https://perma.cc/VG3Z-L9FZ]; Maria Korolov, *AI Technology: When to Build, When to Buy*, CIO (Apr. 9, 2019, 3:00 AM), https://www.cio.com/article/3387618/ai-technology-when-to-build-when-to-buy.html [https://perma.cc/8QPT-FGSB].

[69] *See, e.g.*, Anna Chen, Nicole Heller & Rebecca Lindhout, *Termination Rights in IT Services Contracts – Making Sure You Can Get Out When It All Goes Wrong*, LEXOLOGY (Jan. 17, 2013), https://www.lexology.com/library/detail.aspx?g=3a14e54b-cf14-438a-86d4-0cfd5103ab02 [https://perma.cc/NKT2-Q2BM]; Jim Steinberg & Meredith Francis, *Terminating Long-Term Technology Agreements*, FULTON CNTY. DAILY RPT. (May 1, 2015), https://www.kilpatricktownsend.com/~/media/Files/articles/2015/TechArticle5115.ashx [https://perma.cc/SYC2-95CW].

the organization will likely seek redress.[70] If the AI solution creator resists correcting the problem or absorbing financial cost, the organization may sue the AI solution provider if such problems sufficiently demonstrate a material breach as defined in the contract or may pursue termination based on activities defined in the termination clause.[71] In order to resolve whether the AI solution provider caused the injury or breached the contract, the organization may seek AI explanation to determine whether the solution provider is responsible rather than the organization.

4. AI Disclosure to Prevent or Address Discrimination

AI may directly discriminate against certain groups due to discriminatory data used to create the algorithms, based on proxies for such data, or because the algorithm has not been properly tested to avoid these outcomes.[72] To prevent discrimination or to address due process concerns, explanation may be desirable and necessary in similar ways to tort actions. A future regulatory agency could ensure AI algorithms are not: 1) designed *to* discriminate against individuals in protected classes; or 2) designed in a way that nevertheless disproportionately discriminates against these individuals.[73] Regulating algorithms could reduce the effects of discriminatory algorithms.[74] Future laws that require fairness in algorithmic AI functionality could require some submission of AI prior to release and review by an agency or third parties.[75] For example, an AI algorithm used for criminal sentencing may need to be reviewed prior to its use due to the potential risk of disproportionately impacting individuals in a way that directly affects their life and liberty.[76]

[70] Similar issues arise in language specific to cybersecurity breaches. When a data breach occurs, often organizations both seek to determine what the root cause of such a breach might be and have the opportunity to terminate the agreement. *See* TSCHIDER, *supra* note 26, at 380.

[71] *See* Leslie Marell, *Defining "Material Breach" in Your Contract*, MARELL LAW FIRM (Mar. 2, 2015), http://marell-lawfirm.com/defining-material-breach-contract [https://perma.cc/9UTM-XK9U].

[72] *See* Citron & Pasquale, *supra* note 15, at 3–4; Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. 671, 694–702 (2016); Anya E.R. Prince & Daniel Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, 105 IOWA L. REV. 1257, 1273–81 (2020); Charlotte A. Tschider, *Regulating the Internet of Things: Discrimination, Privacy, and Cybersecurity in the Artificial Intelligence Age*, 96 DENV. L. REV. 87, 97–103 (2018); Sharona Hoffman & Andy Podgurski, *Artificial Intelligence and Discrimination in Health Care*, 19 YALE J. HEALTH POL'Y, L., & ETHICS 1, 17–18 (2020).

[73] *See generally* David S. Rubenstein, *Acquiring Ethical AI*, 73 FLA. L. REV. (forthcoming 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3731106 [https://perma.cc/87WQ-HJBB] (describing U.S. interest in trustworthy, transparency AI for government AI use and proposing regulatory involvement in procurement processes for acquiring AI).

[74] *See* Kaminski, *supra* note 67, at 122–23. Kaminski describes regulating this as instrumental rationale. *Id.*

[75] *See* Roger Allan Ford & W. Nicholson Price II, *Privacy and Accountability in Black-Box Medicine*, 23 MICH. TELECOMM. & TECH. L. REV. 1, 18–19 (2016).

[76] *See generally* Cary Coglianese & Lavi M. Ben Dor, *AI in Adjudication and Administration*, 86 BROOK. L. REV. (forthcoming 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3501067 [https://perma.cc/Q26T-ZVE4] (describing guidelines developed by the Administrative

Further, a person injured by discriminatory behavior, whether intentional or not, may seek to introduce an explanation of the algorithm's decision or function to prove that the algorithm discriminated on the basis of protected class.[77] Government agencies making decisions regarding entitlements, such as eligibility and amount of a welfare payment, may use AI.[78] Or a credit check tool used to determine whether an individual qualifies for housing may be "appealable" if the tool is later demonstrated to be discriminatory.[79] Algorithms may be used in any number of decisions that dramatically affect an individual's economic prospects, legal rights, or government entitlements,[80] and it is crucial to clearly delineate the policy goals prior to determining an appropriate legal approach.

III. LEGAL OPACITY & DATA ESSENTIALISM

AI technology may be key to resolving legal issues, but legal strategies may also function to make these same technologies opaque. Legal opacity is created both through private contract terms and statutory privacy requirements, which collectively create substantial impediments to transparency goals.

AI transparency may be valuable in many different legal and non-legal scenarios, and, as Price and Rai suggest, machine learning may be the

Conference of the United States for considering issues like transparency, bias, technical capacity, and procurement and types of AI introduced to date); *cf.* Saul Levmore & Frank Fagan, *Competing Algorithms for Law: Sentencing, Admissions, and Employment*, 88 U. CHI. L. REV. 367, 390–92 (2021) (proposing competing algorithms rather than calls for algorithmic transparency).

[77] *See* Frank Pasquale & Danielle Keats Citron, *Promoting Innovation While Preventing Discrimination: Policy Goals for the Scored Society*, 89 WASH. L. REV. 1413, 1417–19 (2014) (acknowledging the integrated nature of protected classes within scoring and decision-making systems); Frank Pasquale, *Restoring Transparency to Automated Authority*, 9 J. ON TELECOMM. & HIGH TECH. L. 235, 236–40, 244–50 (2011) (examining the interplay of secrecy and explanation to determine where transparency in explanation is warranted). *See generally* Lilian Edwards & Michael Veale, *Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?*, 16 IEEE SEC. & PRIV. 46 (2018) (arguing that putting the onus on individuals to challenge opaque systems may not be the best model for improving AI).

[78] ENGSTROM ET AL., *supra* note 10, at 10, 37–44.

[79] *See* Charlton McIlwain, *AI Has Exacerbated Racial Bias in Housing. Could It Help Eliminate It Instead?* MIT TECH. REV. (Oct. 20, 2020), https://www.technologyreview.com/2020/10/20/1009452/ai-has-exacerbated-racial-bias-in-housing-could-it-help-eliminate-it-instead [https://perma.cc/AU4W-X9XB]; Patrick Sisson, *Housing Discrimination Goes High Tech*, CURBED (Dec. 17, 2019, 6:12 PM), https://archive.curbed.com/2019/12/17/21026311/mortgage-apartment-housing-algorithm-discrimination [https://perma.cc/3YPU-NV4Y]; William Gordon, Katherine Kirkpatrick & Katherine Martin, *Artificial Intelligence and the Fair Housing Act: Algorithms Under Attack?*, JD SUPRA (June 20, 2019), https://www.jdsupra.com/legalnews/artificial-intelligence-and-the-fair-10153 [https://perma.cc/JZU4-TDYL].

[80] This is precisely why the state of Colorado and the European Union ("EU") have statutorily protected an individual's ability to avoid profiling and automated processing when such processing is likely to affect an individual's legal interests or similar rights. *See* S. 21-190, 73d Gen. Assemb., Reg. Sess. (Colo. 2021); Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1, art. 22.

mechanism for achieving explanation, even where such systems are complex.[81] But requiring or compelling "explanation" or promoting AI transparency is not as simple as just mandating it—an explanation without context is likely not very useful.[82] How an AI algorithm works, including the system, process, testing, and infrastructure used to create it, may be useful to know, but it is often protected from disclosure under confidentiality language, non-disclosure agreements, and other contractual methods.[83] Confidentiality obligations, data ownership contractual provisions, and trade secrecy status reflect an organization's need to protect its investments in cutting-edge research, usually resulting in less data use and sharing.[84]

Data protection and privacy laws also restrict data collection, use, and retention.[85] Laws in most jurisdictions include data minimization principles along with data use limited to what is explicitly disclosed in a privacy notice to the downstream individual about whom data are collected.[86] Overall, the legal impediments to explanation and broad data sharing are more nebulous and intransigent that one might immediately expect when focused on the language of secrecy.

A. *DISCRETIONARY LEGAL CHOICES*

When any organization has developed or acquired an invention, one initial step involves consultation with attorneys and discussion of the appropriate strategic approach to protect financial interests tied to such an invention.[87] This brief Response cannot fully describe the many ways in which organizations and individuals hope to profit from their investments and

[81] *See* Price & Rai, *supra* note 18, at 785. Although explanations are usually statistical approximations, machine learning can be used to explain machine learning systems to some extent.

[82] *Id.* at 785–86; *see* Tschider, *supra* note 13, at 692–96 (describing the details of AI systems that could better fulfill fairness and safety goals).

[83] *See supra* Part II and accompanying notes.

[84] Indeed, this is precisely why data sharing agreements have become a common contractual form when data is involved. If data *will* be shared, specific limitations may govern such transactions. *See* Julien Debussche, Jasmien César, Benoit Van Asbroeck & Isis De Moortel, *Big Data & Issues & Opportunities: Data Sharing Agreements*, BIRD & BIRD (Apr. 2019), https://www.twobirds.com/en/news/articles/2019/global/big-data-and-issues-and-data-sharing-agreements [https://perma.cc/EU58-JZET]. Such agreements usually include not only use and restrictions from a contractual perspective, but applicable law related to data protection and privacy. *See* Nemac, *supra* note 34. Confidentiality and trade secrecy language are usually integrated into contracts between organizations and between an organization and its personnel (including contractors and employees).

[85] *See* Tschider, *supra* note 72, at 110–12.

[86] *Id.*

[87] Wil Michiels, *How Do You Protect Your Machine Learning Investment?*, EE TIMES (Mar. 26, 2020), https://www.eetimes.com/how-do-you-protect-your-machine-learning-investment [https://perma.cc/F5D4-UMDN]; Andrea Weiss Jeffries & Emily J. Tait, *Protecting Artificial Intelligence IP: Patents, Trade Secrets, or Copyrights?*, JONES DAY (Jan. 2018), https://www.jonesday.com/en/insights/2018/01/protecting-artificial-intelligence-ip-patents-trad [https://perma.cc/C7A9-3NAC].

inventions, but suffice to say, these individuals do not necessarily struggle with *whether* to protect it but rather with *how*.[88]

There are a variety of legal tools organizations and individuals use to protect their inventions including trade secrecy, patents, and private contracts. These legal tools limit disclosure of information, as does privacy law. Trade secrecy, private contracting, and privacy law, used together, present a cumulative effect that enhances opacity, rather than promoting transparency of AI technologies and underlying data.

1. Trade Secrecy & Patent Law

First, it is important to distinguish between secrecy and trade secrecy. Secrecy is the act of something being a secret or secretive whereas trade secrecy only exists because the law protects secrets that meet specific requirements. Some AI technologies, for example, may be dynamically inscrutable.[89] Dynamically inscrutable algorithms dynamically change and learn, updating the live algorithm as new data inform these changes. Algorithms that are inscrutable are not able to be easily explained; even their (human) data scientist creators may not understand how they render a decision due to the complexity of the underlying system. This form of secrecy, both in underlying complexity and in continuous change, does not immediately qualify any AI technologies a trade secret. However, this kind of "natural" secrecy may make these technologies good candidates.

Trade secrecy's combination of private law, federal law, and state law is a potent and often attractive legal approach to protect inventions that may make poor candidates for patenting *and* in situations where complementary intellectual property techniques are desirable.[90] Unlike patent law, where a granted patent permits an organization to maintain a limited monopoly on an invention protectible and disclosed under federal law, a trade secret exists because an organization determines it exists and protects it as such.[91] The lack of a central regulatory body that requires submission of a trade secrets for approval means that organizations may maintain any number of trade secrets for an unrestricted period of time.

Trade secrecy requires that trade information is kept secret to receive the defensive benefit of being able to sue an individual or organization for trade

[88] *See* Michiels, *supra* note 87.

[89] *See* Tschider, *supra* note 13, at 689–90.

[90] Steven R. Daniels & Sharae' L. Williams, *So You Want to Take a Trade Secret to a Patent Fight? Managing the Conflicts between Patents and Trade Secret Rights*, LANDSLIDE (Aug. 5, 2019), https://www.americanbar.org/groups/intellectual_property_law/publications/landslide/2018-19/july-august/so-you-want-take-trade-secret-patent-fight [https://perma.cc/Z876-C7QH]; Karl F. Jorda, *Patent and Trade Secret Complementariness: An Unsuspected Synergy*, 48 WASHBURN L.J. 1, 1–2 (2008).

[91] Lawrence Goodwin & Stacy Grossman, *Trade Secrets*, N.Y.C. BAR (Dec. 2018), https://www.nycbar.org/get-legal-help/article/intellectual-property/trade-secrets [https://perma.cc/U5VC-QW2G].

secret misappropriation.[92] This secrecy generally requires that only a few individuals within an organization know of the trade secret and that the invention cannot be reverse engineered, or that the trade secret cannot be immediately discovered using reasonable means.[93] Organizations have further buttressed trade secrecy status through employee confidentiality contract provisions and vendor contracts,[94] which partially demonstrates notice to such actors about their obligations with respect to confidential information and trade secrets.[95] Indeed, trade secrets can exist absent from reference in a contract and are often positioned as an alternative to protection via contract, but organizations usually pursue both simultaneously.[96]

Trade secrecy can be applied to many aspects of an AI invention, from data to process, system architecture, to the algorithms themselves.[97] And trade secrecy will render a substantial amount of the invention secret: If information is disclosed, even to a larger group of individuals, organizations could risk destroying their trade secrets.[98] Without trade secrecy status, organizations—especially for-profit start-ups who have created new and innovative AI products and systems—cannot otherwise enjoy legal protection for their investment in the event patents are not available to them.[99] This operates as a Catch-22: By disclosing, organizations have no protection for these inventions, yet by maintaining secrecy, transparency is not possible.[100]

Organizations may use a combination of approaches to protect their investment and inventions. The U.S. Patent and Trademark Office, for its part, has noted the "complementary" nature of trade secrecy to patents, acknowledging the relative overlap and buttressing of inventions—patents

[92] *See* Daniels & Williams, *supra* note 90.

[93] *Id.*

[94] *See* Jorda, *supra* note 90, at 7–8.

[95] A confidentiality agreement is not *usually* enough alone but can be part of a strategy for notifying individuals as to the presence of a trade secret and their attendant obligations. *See* Varadarajan, *supra* note 34, at 1560–62.

[96] Jeanne C. Fromer, *Machines as the New Oompa-Loompas: Trade Secrecy, the Cloud, Machine Learning, and Automation*, 94 N.Y.U. L. REV. 706, 712.

[97] *See* W. Nicholson Price II, *Big Data, Patents, and the Future of Medicine*, 37 CARDOZO L. REV. 1401, 1433 (2016); *see* Tschider, *supra* note 13, at 711–13.

[98] FENWICK & WEST LLP, TRADE SECRETS PROTECTION: A PRIMER AND DESK REFERENCE FOR MANAGERS AND IN HOUSE COUNSEL 2–3 (2001), https://assets.fenwick.com/legacy/FenwickDocuments/Trade_Secrets_Protection.pdf [https://perma.cc/5ZNT-4X9Y].

[99] *See* David S. Levine & Ted Sichelman, *Why Do Startups Use Trade Secrets?*, 94 NOTRE DAME L. REV. 751, 757–59 (2018). For some statistics on AI startups, see Sarah Feldman, *AI Startup Funding Reaches Record High*, STATISTA (July 31, 2019), https://www.statista.com/chart/18878/artificial-intelligence-startup-funding [https://perma.cc/FEV5-CMUH].

[100] Not everything can be rendered a trade secret; for example, information that must be shared with a larger group of people or that can be reverse engineered from simply using technology will generally not qualify. Nevertheless, for complex AI, as in neural networking, most of the invention can reasonably be maintained as a trade secret. *See* Tschider, *supra* 13, at 712–13.

and trade secrets are not mutually exclusive.[101] As Sonia Katyal has identified in relation to computer source code, patent and trade secrecy are used as complements, alternatives to one another, and sometimes in concert to overprotect inventions.[102] Scholars such as Nicholson Price and Christopher Seaman have noted the problematic nature of similar subject matter for patents and trade secrets, especially when trade secrecy is used to artificially extend protection or when complementary patent and trade secret usage overprotects technologies.[103]

For AI technologies, however, patent law may not be a possible alternative to trade secrecy, leaving organizations with few options for protected disclosure of the invention itself.[104] While innovation might benefit from increased disclosure of some kind, the unavailability of an alternative mechanism to trade secrecy is problematic because organizations will likely opt for trade secrecy rather than leave inventions unprotected.[105] In order to maintain a trade secret, organizations will not be able to transparently provide at least some information on the AI system, lest destroy its trade secret status.

2. Contracts

Although many scholars have focused on trade secrecy as the primary concern regarding AI, private contracting poses significantly more risk to data transparency and subsequent use. For example, in the healthcare sector healthcare providers and insurers, the organizations often operating, using, implanting, or prescribing AI, do not create AI inventions. Rather, these parties contract with medical device manufacturers, such as university entities, start-ups, or sophisticated manufacturers, that license or sell AI technologies. There are typically three types of contract language, whether managed within a singular contract or multiple agreements. See Figure 1 for an illustration of these arrangements.

Figure 1: Contracts Limiting Data Use

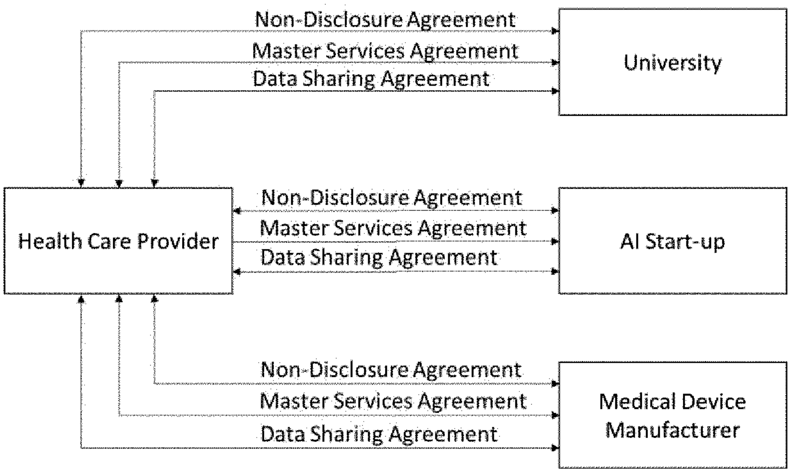
[101] *Trade Secret Policy*, U.S. PAT. & TRADEMARK OFF. (Nov. 18, 2020, 8:49 AM), https://www.uspto.gov/ip-policy/trade-secret-policy [https://perma.cc/S8KA-4JB2].

[102] Sonia K. Katyal, *The Paradox of Source Code Secrecy*, 104 CORNELL L. REV. 1183, 1211–16 (2019); *see also* Tabrez Y. Ebrahim, *Artificial Intelligence Inventions & Patent Disclosure*, 125 PA. ST. L. REV. 147, 182 (2020).

[103] *See* W. Nicholson Price II, *Expired Patents, Trade Secrets, and Stymied Competition*, 92 NOTRE DAME L. REV. 1611, 1614, 1636 (2017); Brian J. Love & Christopher B. Seaman, *Best Mode Trade Secrets*, 15 YALE J.L. & TECH. 1, 11–12 (2012).

[104] *See* Tschider, *supra* note 13, at 715–16; Brian Higgins, *The Role of Explainable Artificial Intelligence in Patent Law*, 31 Intell. Prop. & Tech. L.J., Mar. 2019, at 3, 7; Ebrahim, *supra* note 102, at 151–56. *See generally* Ryan Abbott, *Everything Is Obvious*, 66 UCLA L. REV. 2 (2019) (arguing that obviousness in AI would be a significant bar for patent eligibility).

[105] Indeed, the algorithmic portion of an AI invention will not be typically protected because it is likely an abstract, mathematical idea, and the complexity and inscrutability of complex algorithms means that it is far easier to keep its internal logic secret, what I've describe as a "natural" trade secret. *See* Tschider, *supra* note 13, at 688–89.



When healthcare providers ("HCPs") desire to prescribe or use AI, whether implantable medical devices, AI diagnostic tools, or integrated AI systems (such as AI-enabled robotics or surgical tools), HCPs rely on the technologies of third parties, such as medical device manufacturers, AI start-ups, or university technology providers, which may also be considered start-ups. These are "third parties" with respect to the HCP because they provide the AI technology but not the patients.[106] In order to use AI technologies, HCPs execute contracts with these third parties to negotiate pricing, licensing, and details of their relationship. Although these parties may appear to have linear relationships as described in Figure 1, the reality is that they may also have relationships with each other, such as when a manufacturer invests in university research and development or acquires a start-up or licenses their technology. The reality is that private agreements are plentiful in these arrangements and likely have included different terms depending on the parties contracting and their relative goals.

What many scholars advocating for AI transparency may not be considering is that these private contracts serve to render a great deal about an AI system confidential, including system technology details and data collected and used within that system with respect to the two parties.

There are at least three types of provisions or contracts that may be executed. Sometimes these provisions will be integrated into one primary contract or executed separately.

[106] Many of these third parties also rely on third parties, such as cloud providers that manage applications or storage technologies for these organizations. In clinical trials, sometimes health care providers enter into different contractual relationships with these third parties, for example agreeing to administer a clinical trial or to recruit potential clinical trial participants. For purposes of this illustration, I will be focusing on contractual relationships formed when an AI system is already being used and (where applicable) has passed the FDA review process.

Prior to executing a final contract, these organizations will frequently execute a mutually restrictive non-disclosure agreement ("NDA"). NDAs restrict disclosure of information deemed confidential or proprietary to each organization.[107] An NDA works to ensure confidential information from both parties is not disclosed during initial discussions, for example if a third party wishes to share information about the technology, or if an HCP wishes to describe or use some information about its internal practices or patients.[108] After it is determined that the organizations would like to work together, a more comprehensive contract is executed to govern actual technology acquisition and use.

The primary contract executed will usually be a Master Services Agreement or a Master Supply Agreement. A Master Services Agreement is a common contract form used for arrangements where organizations will not have any ownership rights in the technology, for example if an HCP does not wish to purchase a surgical robot but does wish to use it. A Master Supply Agreement is used when an HCP is purchasing the technology. These contracts usually include at least three key sections that limit what these organizations may do with the information they receive.

i. Licensing Terms

First, when an HCP will not own the product, a third party may ensure licensing terms are included in the agreement. Licensing terms restrict how and to what extent an HCP may use the product. For example, AI software may only be installed on three computers, or ten physicians have the ability to use it. Importantly, many licensing sections also restrict what an HCP can do with respect to the invention. Specifically, many Master Services Agreements and Master Supply Agreements include a "no reverse engineering" obligation. Because reverse engineering is one way to destroy trade secrecy status (by removing the technology secrecy requirement), third parties use this contractual provision to limit any power an HCP might have to make that protection unavailable. Functionally, this means that the HCP may not use its position with access to the technology to figure out how it works in a way that would destroy trade secrecy status.

ii. Confidentiality Terms

Second, confidentiality provisions are usually included even if an NDA has previously been executed. Although an NDA is designed to govern pre-contractual discussions about technology and HCP operations and use cases, confidentiality language included in Master Services or Supply Agreements

[107] *The 5 No's of Confidentiality Agreements*, EVERYNDA (Nov. 22, 2017), https://www.everynda.com/blog/5-no-confidentiality-agreements [https://perma.cc/Y4BA-94Q7].

[108] Josh Angert, *6 Best Practices for Using Nondisclosure Agreements During the Procurement Process*, VENDORCENTRIC (June 12, 2019), https://vendorcentric.com/single-post/6-best-practices-for-using-nondisclosure-agreements-during-the-procurement-process [https://perma.cc/NNL4-ATK5].

directly restricts external sharing of *any* information deemed confidential in the contract, which includes technology details, patient details, and any other proprietary information about the organizations. The operative impact is that neither of the organizations can disclose any data or information to any other entity, except in responding to a legally enforceable order. Often these contracts also restrict the speed at which either organization can even respond to an order, so that the other organization has the opportunity to execute a protective order. Protective orders operate to limit disclosure in a court of law pursuant to a legal action of any kind (e.g., an administrative agency's investigation or a class action lawsuit).

iii. Data Ownership and Sharing Restrictions

Third, and perhaps most importantly, most MSAs include patient data ownership rights established by the HCP with respect to their patients. Although data do not have independent status under the law as real property or intellectual property, they may be subject to restrictions in their use or disclosure via contract.[109] Whether because of perceived privacy issues or a desire for control over the data for some other reason, these provisions have the effect of limiting a third party's use of patient data for any purposes other than simply providing a service. This includes use by the third party and sharing with any other entities. Functionally, this means that third parties may not contractually use patient data collected pursuant to providing a service or product for purposes such as product improvement, to create new products, or for independent research.

Because data have no independent legal status outside intellectual property status, data are typically a creature of contracts when they are poor candidates for intellectual property protection: organizational rights to data are subject to the limits that the providing organization attaches to them in contract.[110] For example, an HCP may contract with a medical device manufacturer to purchase or lease a surgical robot. This surgical robot collects and transmits data back to the manufacturer to improve the effectiveness of the robot over time. Despite the manufacturer having custody of the data, likely the healthcare provider has limited use of the data collected from the provider's patients under the contract.[111] This means that the

[109] TERESA SCASSA, CENTRE FOR INTERNATIONAL GOVERNANCE INNOVATION, CIGI PAPERS NO. 187, DATA OWNERSHIP 2 (2018), https://www.cigionline.org/documents/1491/Paper%20no.187_2.pdf [https://perma.cc/G36Q-XE6E].

[110] *Id.* at 7 n.44, 7–10 (describing the challenges of using copyright protection and database rights under EU law). Indeed, the idea/expression dichotomy makes copyright a poor available choice for AI that are used for inventive purposes. *See generally* Adam Mossoff, *A Brief History of Software Patents (and Why They're Valid)*, GEO. MASON UNIV. CTR. FOR INTELL. PROP. X INNOVATION POL'Y (Sept. 18, 2013), https://cip2.gmu.edu/2013/09/18/a-brief-history-of-software-patents-and-why-theyre-valid-2 [https://perma.cc/2XZN-NBMB].

[111] *See* TSCHIDER, *supra* note 26, at 300–01. Data owners and data custodians often have different rights and responsibilities with respect to data. When these roles are separated between business entities, often a contract is used to establish and limit these responsibilities and rights.

medical device manufacturer may not use these data for their internal purposes.

When third parties and HCPs agree that data sharing is desirable, so that the third party can use data supplied by the HCP for other reasons or disclose this data to other organizations, a separate contract is usually executed and appended to the MSA.[112] These data sharing agreements, although functioning to permit data sharing beyond the limitations of the MSA, still limit data use and sharing to what the HCP specifies.[113] They may also include profit sharing and other terms of benefit to the HCP for providing use of data supplied.

Frequently, data sharing agreements also include privacy obligations with respect to patient data. Commonly, it is required that a privacy notice and an associated lawful basis (such as consent) will be executed for additional uses.[114] In some jurisdictions outside of the United States (and in the United States for specific sectors), data sharing in this way increases privacy obligations to an individual about whom personally identifiable data are maintained.[115] For example, a third party cannot be directly regulated under the Health Insurance Portability and Accountability Act of 1996 ("HIPAA") because the third party is legally organized or incorporated outside of the United States. However, if an MSA or data sharing agreement requires compliance with HIPAA, the third party must comply lest potentially be in breach of that contract. The inclusion of references to privacy laws, as described in Section II.C, in private contracting language has the effect of exporting privacy laws to third parties that may not be otherwise regulated.

When aggregated across all HCPs and third parties, private contracting extends and expands legal opacity substantially. Organizations may be functionally limited from using or sharing patient data because the contract outright prohibits it. Therefore, even when trade secrecy and patents do not protect AI technology statutorily, often several private contracts exist that restrict disclosure and aim to control the dissemination of proprietary information or data.

Contracts are tremendously powerful. In summary, confidentiality provisions generally restrict disclosure of information about the AI

[112] *See, e.g.*, R.I. DEP'T OF HEALTH & HUM. SERVS., SAMPLE DATA-SHARING AND USAGE AGREEMENT, https://www.cdc.gov/cancer/ncccp/doc/sampledatasharingusageagreement.doc [https://perma.cc/F2HZ-8Z52] (specifying data uses permitted, obligations of confidentiality, and data ownership terms).

[113] *See, e.g.*, *Data Use Agreement Guidance*, UNC RSCH., https://research.unc.edu/wp-content/uploads/sites/61/2013/04/CCM3_039360.pdf#:~:text=Data%20Use%20Agreements%20%28DUAs%29%20are%20contractual%20documents%20used,the%20appropriate%20delegated%20signature%20authority%20from%20the%20Chancellor [https://perma.cc/CF4E-J7UF].

[114] *See* TSCHIDER, *supra* note 26, at 14–17 (describing notice and consent obligations), 36–37 (defining secondary use and specifying required steps to process data for secondary purposes cross-jurisdictionally).

[115] *Id.* at 39 (describing the challenges associated with maintaining multiple roles with respect to the data).

technology. Licensing language similarly prevents disclosure or other actions that could destroy trade secrecy status. Data ownership language restricts data use beyond what is specified, and even if organizations agree to a data sharing agreement, these agreements usually include statutory privacy limitations, which make further data use and sharing tremendously complicated. Each of these provisions significantly limits transparency goals with respect to the technology and data. The effect is privately negotiated and cumulative legal opacity as HCPs negotiate multiple contracts with many third parties and third parties negotiate with several HCPs.

B. Privacy Law: Legally Mandated Opacity

As an extension of data transmission and restriction, privacy goals similarly complicate the notion of broad inventive disclosure, especially data disclosure and algorithmic explanation when such explanation has the ability to identify a natural person. The challenges associated with these disclosures result from the disclosures themselves and the ability of organizations to collect the volume of data necessary to create AI systems in the first place. Although data protection and privacy laws may protect personal information, they also present significant, potentially intractable, impediments to innovation, one of Price and Rai's primary motivations for disclosure.[116]

1. The U.S. Sectoral Privacy System

The United States has implemented a sectoral privacy system, which means that certain sectors may establish more stringent or more flexible legal requirements for data. This fact is important because promoting AI innovation broadly is more complicated in the United States than potentially for other global jurisdictions that have passed an omnibus data protection law that applies to all sectors.

Although a great variety of sectoral and use-specific laws exist in the United States, I will use HIPAA as an example of data use restrictions.[117] HIPAA established a broad data minimization standard for protected health data, insofar that data collected generally may not be collected unless such collection is *necessary* for treatment, payment, or healthcare operations.[118] Unlike clinical trial data subject to the Common Rule, which usually permits

[116] *See* Price & Rai, *supra* note 18, at 781.

[117] Although United States privacy laws exhibit different levels of stringency, and HIPAA is often considered one of the most stringent, Price and Rai's discussion of the healthcare space should explore healthcare-specific privacy.

[118] 45 C.F.R. §§ 164.502(b), 164.514(d)(1) (2020).

future data use restricted to what has been communicated to clinical trial subjects,[119] patient data use under HIPAA is substantially more limited.[120]

HIPAA applies to statutorily defined covered entities, which include HCPs. Certain rules under HIPAA apply when HCPs engage in electronic transmission of Protected Health Information, which is broadly defined as health information collected by a covered entity and transmitted electronically. The Health Information Technology and Electronic Communications for Health Act of 2009 ("HITECH") extended HIPAA to include business associates, or third parties of covered entities receiving Protected Health Information. HIPAA statutorily requires that covered entities execute a business associate agreement in contracts with business associates, which effectually exports HIPAA's provisions to the third party, regardless of where the third party is geographically located.

HIPAA conditions data use on what is disclosed in the Notice of Privacy Practices presented to the patient. These data uses are limited to use for treatment, payment, or healthcare operations.[121] In the event data will not be used for these purposes, or for purposes not consistent with data minimization, an additional authorization (with explicit consent of the patient) must be executed.[122] The authorization document itself limits any use of protected health information to the additional purposes and third parties specified in the authorization form, subject to a specific time period or until a terminating event occurs.[123] HIPAA authorization to date has not been designed for broad and unlimited duration data sharing.

In the event that an AI technology provider is selling direct to a consumer without submitting an insurance claim, and not selling a product or service to an HCP, the Federal Trade Commission ("FTC") is responsible for enforcing against unfair or deceptive trade practices under Section 5 of the FTC Act. The FTC has not passed specific rules regarding data minimization, notice and consent, or authorization. However, the Fair Information Practice

[119] *Id.* § 46.116(b); Kate Fultz Hollis, *To Share or Not to Share: Ethical Acquisition and Use of Medical Data*, 2016 AMIA JOINT SUMMITS ON TRANSLATIONAL SCI. PROC. 420, 424–25. Notably, additional data use beyond what is included in informed consent forms in a clinical setting can be referred to the Institutional Review Board. *See Institutional Review Boards and the HIPAA Privacy Rule*, U.S. DEP'T OF HEALTH & HUM. SERVS. NAT'L INSTS. HEALTH (July 8, 2004), https://privacyruleandresearch.nih.gov/irbandprivacyrule.asp [https://perma.cc/D9RD-RRJW].

[120] 45 C.F.R. §§ 164.508(b)(4), 164.502(a)(1). It should be noted that when federal funding applies, data deposit requirements may provide some relief from data use limitations. *See* W. Nicholson Price II, Rachel Sachs & Rebecca S. Eisenberg, *New Innovation Models in Medical AI*, WASH. U. L. REV. (forthcoming 2022) (manuscript at 40), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3783879 [https://perma.cc/S2JE-HX8U]. However, at least at this time, data collected after an AI product has been commercialized generally will not be included in such a mandate.

[121] 45 C.F.R. § 164.520(a)–(b).

[122] 45 C.F.R. §§ 164.508(a)–(b), 164.532(a)–(c). A covered entity, such as a healthcare provider or health plan administrator may not condition the provision of healthcare or coverage on execution of an authorization. Any consent must be freely given.

[123] *Id.* §§ 164.508(a)–(b).

Principles ("FIPPs"), which are non-binding guidance that inform FTC enforcement, do establish recommended guidance to use notice and consent. The effect is that complying with such guidance operates to avoid potential liability under the FTC Act.

The general outcome of HIPAA statutory requirements and the FTC's enforcement prerogative is that data collection must be minimized and specific to the uses defined at the time of collection, data use must be able to be terminated at a point in time, and patients may be required in some circumstances. In all cases, previously given consent may be revoked. Practically, this has a dramatic effect on subsequent data use and sharing. First, privacy notices must be reasonably specific as it describes how data will be used, which will likely be fairly restrictive, reducing the usability of data previously collected for any other purposes. Under HIPAA, uses that are beyond treatment, payment, and healthcare operations, as may be the case with AI technology, must be specifically described in an authorization form, which may be difficult to execute, easy to terminate, and also specifically tailored. If consent is revoked, then data previously collected may not be reused for other purposes.

The narrowly tailored aspect of privacy notices and a patient or consumer's ability to revoke consent means that data previously collected may be prohibited from use for other purposes, or shared with other parties, by law. Moreover, the ability for a patient to revoke consent means that data previously supplied also not be used for any additional uses or shared with other parties. For protected health information collected under HIPAA, authorization may be revoked at any time.

Privacy law, therefore, functions as a substantial barrier to data use and sharing. Because data collection may be minimized—restricted to the purposes specified in the notice of privacy practices, privacy, notice, or authorization, and consent may be revoked, limiting further data use—privacy law and data essentialism in AI are at odds.

2. Clearing Privacy Obstacles

One option for overcoming privacy obstacles while protecting individual privacy and promoting data use and sharing is for organizations to invest in de-identification technologies.[124] Although the United States has not established de-identification standards in other sectors, the De-identification Safe Harbor rule—established by the U.S. Department of Health and Human Services ("HHS") for entities obligated to follow HIPAA—permits broad data use when a data set has 18 identifiers removed or when expert determination has rendered statistically low risk to an individual.[125] In both cases,

[124] *Using De-identified Health Information to Improve Care: What, How and Why*, PRACTICE FUSION (Apr. 30, 2010), https://www.practicefusion.com/blog/using-de-identified-patient-data-to [https://perma.cc/H9GL-FBDY].

[125] *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*, U.S. DEP'T OF

organizations may not rely on the Safe Harbor provision when they independently know that individuals can actually be identified from the data set.[126]

On its face, AI might function well using de-identified data only. However, AI's power is in its ability to generate broad decisions and adapt those decisions to an individual, or personalization.[127] For this reason, AI is inherently personalized when it is being used to offer differentiated service or treatment, making full de-identification of data sets nearly impossible while retaining the efficacy of AI for tailored medicine.[128] Unfortunately, if AI cannot be effectively de-identified to an acceptable risk standard, privacy restrictions remain.

AI use might actually destroy de-identification, too, through reidentification.[129] Generally, the more data one has, especially diverse data, the easier it is to identify an individual.[130] And the computing power of AI, the same computing power that can identify new insights in a wide range of data elements can similarly use that computing power to reidentify an individual,[131] especially with the introduction of additional data. Consider, for example, a data set that includes de-identified medical data such as age, gender, medical diagnosis, general physical location, and physical descriptors

HEALTH & HUM. SERVS. (Nov. 6, 2015), https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html [https://perma.cc/U6LM-F74W].

[126] *Id.* It should be noted that a data set's ability to identify an individual is referenced with respect to actual knowledge, rather than inferential knowledge. For example, inferential knowledge deals in probabilities, rather than actual knowledge. For this reason, typical big data problems regarding identifiability may not necessarily invalidate the Safe Harbor provision. Indeed, inferential knowledge of a person's identity as well as sensitive information about them (which may subject that person to unfairness or discrimination based on that information) may subject that person to additional risk of harm. However, inferences are not objective identifiability.

[127] Charlotte A. Tschider, *AI's Legitimate Interest: Towards a Public Benefit Privacy Model*, 21 HOUST. J. HEALTH L. & POL'Y (forthcoming 2021) (manuscript at 33–34), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3725933 [https://perma.cc/44GN-MXB6]; James Warner, *Thanks to AI, Medical Treatments Are Becoming More Personalized*, TNW (Dec. 11, 2019), https://thenextweb.com/news/thanks-to-ai-medical-treatments-are-becoming-more-personalized [https://perma.cc/W2V3-PNLB].

[128] Tschider, *supra* note 127, at 33–35.

[129] Boris Lubarsky, *Re-Identification of "Anonymized Data"*, 1 GEO. L. TECH. REV. 202, 208–12 (2017).

[130] W. Nicholson Price II, *Problematic Interactions Between AI and Health Privacy*, UTAH L. REV. (forthcoming 2021) (manuscript at 2–4), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3797161 [https://perma.cc/TS6W-9JF4]; Mark Gibbs, *MIT Researchers Show You Can Be Identified by a Just Few Data Points*, NETWORK WORLD (Jan. 30, 2015, 11:47 AM), https://www.networkworld.com/article/2878394/mit-researchers-show-you-can-be-identified-by-a-just-few-data-points.html [https://perma.cc/99YC-5SK7]. This is not to say that de-identification, even of big data that still retain some identifiability is not a good strategy—it does reduce privacy risk to individuals. However, with big data feeding powerful AI, it may not be a panacea. ANN CAVOUKIAN & DANIEL CASTRO, BIG DATA AND INNOVATION, SETTING THE RECORD STRAIGHT: DE-IDENTIFICATION DOES WORK 1–2 (2014), https://www2.itif.org/2014-big-data-deidentification.pdf [https://perma.cc/8HRG-ZH52].

[131] *See* Price, *supra* note 130, at 2–3.

but does not include specific medical record number, date of birth, or address but nevertheless can identify an individual in a big data set.[132]

It is possible, but not likely probable, that an individual will be readily identified from these identifiers. However, if such data are coupled with the identities of friends, places the individual frequents and what they buy, or other de-identified data about the individual's family, it may become easier to identify the individual. To date, HIPAA does not explicitly bar reidentification as a barrier to safe harbor status.[133]

For clinical health data, the concerns may be even more significant.[134] By harnessing the power of AI to analyze diverse and large data sets, the AI can often identify the individual from previously de-identified and publicly available data.[135]

The need for large and diverse data sources to feed machine learning applications means that de-identified data will likely be combined with other data sources, such as publicly available data or other de-identified data received from another entity.[136] This means that even *if* data can be de-identified, in large data sets coupled with sophisticated AI algorithms, data are more likely to be able to identify an individual.[137] Therefore, making data broadly available and making AI explanations more transparent actually could result in a greater probability of identifiability and greater risk to individuals using these systems, which is exactly what transparency advocates aim to solve.

Although much of this section has focused on individually identifiable data elements, advanced algorithms may also reveal cumulative data insights, or inferences.[138] Even when an organization has legitimately collected and

[132] Liangyuan Na et al., *Feasibility of Reidentifying Individuals in Large National Physical Activity Data Sets from Which Protected Health Information Has Been Removed with Use of Machine Learning*, JAMA NETWORK OPEN, Dec. 2018, at 1, 2–3.

[133] *See* U.S. DEPT. OF HEALTH & HUMAN SERVS., *supra* note 125. In contrast, the General Data Protection Regulation in the EU and its predecessor, the Data Protection Directive, both established an "impossibility of reidentification" standard for anonymization, an alternative standard to de-identification. This means that the standard for big data sets using AI will likely be more restrictive in the EU and not apply to U.S. HIPAA protected health information. *Opinion 05/2014 on Anonymisation Techniques*, 0829/14/EN, annex, at 27–28 (Apr. 10, 2014) https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf [https://perma.cc/9MRL-YRCY].

[134] *See generally* Mark A. Rothstein, *Is Deidentification Sufficient to Protect Health Privacy in Research?*, 10 AM. J. BIOETHICS, Sept. 2010, at 3 (concluding that even use of deidentified data pose a privacy risk).

[135] Yves-Alexandre de Montjoye, Ali Farzanehfar, Julien Hendrickx & Luc Rocher, *Solving Artificial Intelligence's Privacy Problem*, 17 FIELD ACTIONS SCI. REPS. (SPECIAL ISSUE) 80, 81–83 (2017).

[136] *See* Lubarsky, *supra* note 129, at 211.

[137] Indeed, prior to current big data and AI use, 63 percent of the population could be identified by a combination of gender, date of birth, and zip code, despite these being "indirect identifiers." *Id.* at 203.

[138] Manish Prabhu, *Security & Privacy Considerations in Artificial Intelligence & Machine Learning—Part-6: Up Close with Privacy*, TOWARDS DATA SCI. (Feb. 8, 2019), https://towardsdatasci

maintained a big data set to feed AI, and even if the big data set contains public and de-identified data, advanced algorithms may nevertheless be able to identify information about an individual in new or different ways, when generalized data are applied in a personalized way, to an individual.[139]

De-identification could reasonably protect individual privacy while permitting data use and sharing, but big data coupled with AI technologies could result in the creation of inferences that are accurate with respect to data already removed from the data set, such as race, or create brand new inferences not independently captured anywhere in the original data set (such as employment status). If the United States mandates direct explanation for AI decisions, the explanations could, in cases where AI developers do not guard against inferences in their design, reveal sensitive details about an individual.[140] For example, a patient's disability status may have been removed from a data set but prescriptions and durable medical supplies could inferentially establish a patient's disability status. And, if enough data are collected about the patient from insurance data, doctors' visits, and public data sources, even if the data are de-identified, other information might be gleaned, too, such as race.

Ultimately, de-identification could promote data use and sharing by reducing risk to patient privacy. Unfortunately, identifiable data are often necessary for the development of personalized AI technologies, and, moreover, de-identified data sets will likely create inferences anyway. While de-identification is an important tool for information privacy, it is not designed to completely resolve key tensions between privacy law and AI transparency goals.

C. *JUSTIFYING TRANSPARENCY*

To achieve "transparency," technical and legal opacity cannot make it difficult to ascertain information about an AI product or service. For situations where transparency may be more necessary, as described in Part I, such as when AI decisions result in significant impacts to a person's legal interests, transparency is exceptionally important.[141]

To determine whether overcoming legal opacity is justified, it is important to revisit why AI transparency is desirable. AI drives calls for

ence.com/security-privacy-in-artificial-intelligence-machine-learning-part-6-up-close-with-privacy-3ae5334d4d4b [https://perma.cc/L95Y-FSAN]; *see* Tschider, *supra* note 72, at 98–99; Barocas & Selbst, *supra* note 72, at 681, 688.

[139] *See generally* REVA SCHWARTZ, LEANN DOWN, ADAM JONAS & ELHAM TABASSI, NAT'L INST. OF STANDARDS & TECH., SPECIAL PUB. 1270, A PROPOSAL FOR IDENTIFYING AND MANAGING BIAS IN ARTIFICIAL INTELLIGENCE (2021), https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270-draft.pdf [https://perma.cc/LP7Y-QJRL] (describing the dangers of bias in AI systems that can have certain negative consequences on an individual).

[140] Sandra Wachter & Brent Mittelstadt, *A Right to Reasonable Inferences: Re-thinking Data Protection Law in the Age of Big Data and AI*, 2019 COLUM. BUS. L. REV. 494, 543–48 (2019) (proposing a right to know and rectify inferences).

[141] *See* Citron & Pasquale, *supra* note 15, at 20–29.

transparency in a variety of contexts because: 1) AI is a comparatively new technology which socially we do not completely understand;[142] 2) in complex AI, decisions are made by the AI or at least partially by the AI (causing inherent issues related to trustworthiness);[143] 3) some of the scenarios where AI will be used have the potential to detrimentally affect many people simultaneously;[144] and 4) this detriment could affect individual lives dramatically, as in discriminatory impact or safety issues.[145]

For AI systems to be reliable, non-discriminatory, and safe, organizations producing AI must have access to extremely large, diverse, representative, non-discriminatory, well-organized, and structured data sets.[146] Data volume and diversity are not optional, but rather essential, for safe and fair AI. Therefore, reducing the probability of discrimination and safety risks is improved by *more*, rather than *less* data.[147] Privacy goals of restrictions in data collection, use, and sharing, then, unfortunately run at cross-purposes to goals of reliability, non-discrimination, and safety. At its most rudimentary explanation, data essentialism prompts broader collection and data use to train safe and fair AI; data protection urges minimization and adds considerable layers of administrative burden to collect and use data.[148]

For example, imagine a new lending system, created by a third-party AI developer and used by a significant number of banks, that determines (without the involvement of a banker) whether an individual qualifies for a loan. The inputs that determine who qualifies are unavailable to both the individual and the banks using (and licensing) the system. The data used to train the system to recognize "good loan candidate" are similarly not available. It is unknown what volume and diversity of data were used to create the system, and whether such data are representative of a diverse community of loan applicants. Indeed, most of the third party's employees were not involved in developing the AI, and the AI is so complex that the developers don't exactly know how the loan recommendations are being rendered. Although the loan system seems to produce the correct "goal" numbers for qualified

[142] *See* Mark MacCarthy, *AI Needs More Regulation, Not Less*, BROOKINGS (Mar. 9, 2020), https://www.brookings.edu/research/ai-needs-more-regulation-not-less [https://perma.cc/ZD45-TVPB].

[143] *Assisted, Augmented and Autonomous: The 3 Flavours of AI Decisions*, TGDAILY (June 28, 2017), https://tgdaily.com/technology/assisted-augmented-and-autonomous-the-3-flavours-of-ai-decisions [https://perma.cc/JPW5-H398].

[144] Charlotte A. Tschider, *Deus ex Machina: Regulating Cybersecurity and Artificial Intelligence for Patients of the Future*, 5 SAVANNAH L. REV. 177, 189–90 (2018).

[145] *See* Tschider, *supra* note 13, at 707.

[146] *See* Price & Rai, *supra* note 18, at 800–01; W. Nicholson Price II, *Medical AI and Contextual Bias*, 33 HARV. J.L. & TECH. 65, 91–94 (2019).

[147] Lee Johnston, *Recent Cases Highlight Growing Conflict Between AI and Data Privacy*, HAYES BOONE (Apr. 20, 2020), https://www.haynesboone.com/publications/recent-cases-highlight-growing-conflict-between-ai-and-data-privacy [https://perma.cc/2NJ7-TNLC] (describing recent litigation at the intersection of privacy and artificial intelligence).

[148] *See supra* Part I and accompanying notes.

loan recipients for each bank, lending activities have recently been criticized for biased algorithmic loan denial.[149]

How might issues like this be resolved? Presumably access to more and diverse data would likely make AI like this fairer and more representative, yet the tool opacity and legal opacity of this scenario creates many hurdles to overcome before an organization could even begin to defend its AI decision.

The challenge to achieving some degree of transparency in AI technologies and their underlying data is three-fold: 1) trade secrecy of at least part of the AI technology; 2) contractual language prohibiting disclosure of confidential information about the AI technology as well as limitations on personal information use and sharing; and 3) privacy laws limiting personal information collection, use, and sharing. Any one of these sources of legal opacity could create a barrier to transparency goals. Together, they are nearly impossible to surmount. Any calls for transparency, especially innovation through data use and sharing, cannot be accomplished in the system as it stands.

IV. TRANSPARENCY: AN IMPOSSIBLE DREAM?

It might seem easy to demand that AI technology and underlying data should be open and available, and certainly there may be situations that demand transparency or explainability. However, broad advocacy of openness fails to appreciate not only the difficulty and sometimes impossibility of making complex, non-intuitive systems explainable, but also it neglects the United States' strong legal protection of trade secrecy, private contract provisions, and privacy law protections. Rather than advocate for general transparency and data availability, we should balance various interests to ensure all participants'—organizations, researchers, and individuals—interests are advanced.[150]

Price and Rai make excellent points regarding big data reuse and non-intuitiveness explanations in a variety of situations—incentives rather than mandates in many situations might help to navigate the legal quandary posed by a strong recognition of private contracting and trade secrecy in the United States. Perhaps, however, complementary legal models might be used to satisfy distinct policy goals.

I suggest that any mandated disclosure, explanation, or expanded data use or sharing, might first be explored contextually. For example, applications in healthcare might well demand more comprehensive disclosure due to health and safety risks, as well as potential discriminatory impact, than AI supporting a consumer retail product like a coffee maker. I also suggest that depending on *what* might be disclosed, such as data,

[149] Public Affairs, *Mortgage Algorithms Perpetuate Racial Bias in Lending, Study Finds*, BERKELEY NEWS (Nov. 13, 2018), https://news.berkeley.edu/story_jump/mortgage-algorithms-perpetuate-racial-bias-in-lending-study-finds [https://perma.cc/55YK-L7J8].

[150] Neil M. Richards & Jonathan H. King, *Big Data Ethics*, 49 WAKE FOREST L. REV. 393, 395 (2014).

algorithmic explanation, or process and system details, the United States might consider different policy approaches. These recommendations only scratch the surface of the comprehensive law and policy discussions to come and are intended to encourage greater conversation on these topics.

A. Promoting Data Disclosure

Price and Rai describe the importance of curated, labeled, representative, and voluminous data availability for innovation in AI, a goal championed by the past three administrations.[151] Without access to broad, diverse data sets, any number of AI applications will be less representative of the communities in which they will be used. Data, especially well-organized and coded data, are the cornerstone of contemporary AI machine learning systems. In addition to Price and Rai's data disclosure public funding policy lever, privacy considerations should evolve to maximize use of these data while simultaneously protecting individuals.

Twin aims could assist in encouraging data disclosure while reducing privacy impacts. First, privacy laws like HIPAA could evolve to balance data minimization and data maximization, data *essentialism*. Data minimization could be construed by the Office for Civil Rights, the enforcement arm of HHS, to include justifiable data collection and use for effective AI applications: namely, to improve AI function safety and fairness in the public interest.[152] Additionally, privacy notices could include directly notifying individuals that AI systems are being used and explicitly limit personal information uses to improved functionality and fairness. Consistent with data essentialism goals, data uses that primarily benefit commercial enterprise rather than the individual or patient would be prohibited, for example, more effective marketing techniques and other uses that would not be in the legitimate interest of individuals.[153]

[151] *See* NAT'L SCI. & TECH. COUNCIL, EXEC. OFF. OF THE PRESIDENT, PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE 14 (2016), https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf [https://perma.cc/7B2Z-EHSF]; THE WHITE HOUSE OFF. OF SCI. & TECH. POL'Y, AMERICAN ARTIFICIAL INTELLIGENCE INITIATIVE: YEAR ONE ANNUAL REPORT 10–11 (2020), https://trumpwhitehouse.archives.gov/wp-content/uploads/2020/02/American-AI-Initiative-One-Year-Annual-Report.pdf [https://perma.cc/52YQ-TXAT]; Press Release, The White House, The Biden Administration Launches the National Artificial Intelligence Research Resource Task Force (June 10, 2021), https://www.whitehouse.gov/ostp/news-updates/2021/06/10/the-biden-administration-launches-the-national-artificial-intelligence-research-resource-task-force [https://perma.cc/3NLN-APF5].

[152] HIPAA, like many privacy laws, contemplates public benefit. In addition to more flexibility for clinical research, HIPAA offers some flexibility for the public interest, such as: public health activities, health oversight activities, research, or as required by law. *See* 45 C.F.R. § 160.203(a)(1) (2020). HHS, as with other administrative agencies enforcing privacy laws, can define what public health activities, for example, qualify under the public health exemptions.

[153] *See generally* Tschider, *supra* note 127 (proposing a legitimate interest model for data processing in healthcare based on individual or public benefit).

For example, in healthcare, this could mean that data previously collected could be subject to a broader public health exception to the requirement for authorization in data sharing. A HIPAA public health exception applies when disclosing to a public health authority or when a public health authority directs disclosure to a foreign government.[154] However, it might be possible to interpret such discussion for public health activities sanctioned by a public health authority, such as the U.S. government, or with the Office for Civil Rights' ("OCR") enforcement discretion.[155] With this approach, the OCR, perhaps collaborating across the FDA, the division of HHS that regulates medical device safety and efficacy, could establish that public health exceptions apply to AI fairness and safety efforts. In the interest of public health, specifically for improving and running AI, the OCR and FDA could declare data sharing for at least some limited purposes and to a necessary extent, permissible under a public interest exception.

Other sectors could follow a similar path. Although healthcare adopts an authorization approach, most sectors follow a general consent approach. For these sectors, consent could be waived when data use is in the best interest of an individual, a class, or a community of individuals, or *legitimate interest*.[156] In short, because further data use primarily benefits a defined group and not the organization collecting data, such use could be justified.[157] In this model of waiving consent, legitimate interest is a useful alternative specifically because it substitutes a collective benefit model for a consent model—data could be used or transferred in the event that substantial benefit to individuals significantly outweighs benefit to an organization.[158] For example, additional processing of data to make better, fairer AI lending software will likely benefit a substantial number of consumers.

Moreover, significant privacy barriers concern the degree to which data are actually capable of identifying an individual, exposing that individual to some harm. To minimize risk to individuals, the United States should invest in more standardization of de-identification techniques for AI data sets that

[154] 45 C.F.R. § 164.512(b).

[155] The OCR has the ability to use discretion in its enforcement of certain provisions of HIPAA. For example, the OCR made clear that disclosure within a Health Information Exchange ("HIE") for purposes of sharing protected health information is permitted. *See*OFF. FOR CIV. RTS., U.S. DEP'T OF HEALTH & HUM. SERVS., HIPAA, HEALTH INFORMATION EXCHANGES, AND DISCLOSURES OF PROTECTED HEALTH INFORMATION FOR PUBLIC HEALTH PURPOSES 1–2 (2020), https://www.hhs.gov/sites/default/files/hie-faqs.pdf [https://perma.cc/BK3F-URAZ]. Moreover, the OCR suspended enforcement temporarily during the COVID-19 crisis. U.S. DEP'T OF HEALTH & HUM. SERVS., COVID-19 & HIPAA BULLETIN LIMITED WAIVER OF HIPAA SANCTIONS AND PENALTIES DURING A NATIONWIDE PUBLIC HEALTH EMERGENCY 1 (2020), https://www.hhs.gov/sites/default/files/hipaa-and-covid-19-limited-hipaa-waiver-bulletin-508.pdf [https://perma.cc/CS94-AEF4].

[156] *See* Charlotte A. Tschider, *Meaningful Choice: A History of Consent and Alternatives to the Consent Myth*, 22 N.C. J.L. & TECH. 617, 675–77 (2021).

[157] *Id.* at 675–79.

[158] *Id.*

render data useful and effectively low risk to patients and other individuals. Although the National Institute of Standards and Technology ("NIST") has created some tools and use cases for organizations to review, NIST could either create legally qualifying procedures for achieving self-de-identification or certify third parties to provide these services.[159] For the healthcare sector, the De-identification Safe Harbor provision should be revised, building on NIST guidance, to promote privacy-oriented data-sharing strategies with little to no privacy risk to individuals.[160] Together with more expansive considerations of alternative justifications for further data use and transfer, such as public health exceptions or legitimate interest, this model could both promote beneficial data use while minimizing risk to individuals.

While these proposed changes do not eliminate contractual limitations on data use, it could encourage the creation of centralized data lakes derived from identifiable data, reducing impediments to data access. For example, an organization receiving *identifiable* data from a customer could de-identify those data and use them for additional purposes, so long as the de-identification activities were properly executed and legally defensible. Resulting data lake infrastructure and virtual AI workspaces could be financed by the federal government, as Price and Rai suggest.[161] However, any access to common data and workspaces should be subject to a common contribution-based contractual obligation where participants must register their use and deposit their own de-identified, labeled, and curated data sets. This model could create a reciprocal sharing arrangement, similar to the contractual obligation to deposit data in government repositories as a condition of receiving federal funding without requiring the government to fund data *creation* long-term. Reciprocal arrangements and government incentives could indeed help to overcome contractual limitations by promoting more frequent execution of data sharing arrangements, especially if health care providers, for example, are not concerned about potential privacy issues.

B. Incentivizing Disclosure of AI Technology

Organizations, especially for-profit corporations and start-ups, may not be encouraged to disclose details of their inventions to competitors. Specifically, these organizations may wish to protect this information as trade secrets or bar any independent disclosure under contractual confidentiality terms. Given the strength of trade secrecy in the United States and the prevalent use of confidentiality agreements, it is unlikely that forced trade

[159] *De-identification*, NAT'L INST. OF STANDARDS & TECH. (July 23, 2020), https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/collaboration-space/focus-areas/de-id [https://perma.cc/E8DT-CRXN].

[160] Latanya Sweeney et al., *Re-identification Risks in HIPAA Safe Harbor Data: A Study of Data from One Environmental Health Study*, TECH. SCI. (Aug. 28, 2017), https://techscience.org/a/2017082801 [https://perma.cc/F7WQ-8MGV] (finding that the HIPAA Safe Harbor is not sufficient to protect against reasonable reidentification efforts).

[161] *See* Price & Rai, *supra* note 18, at 800–01.

secret and confidentiality destruction through compelled disclosure will be successful. Instead, the United States might consider other incentive-based models, as Price and Rai recommend.[162]

Innovation in AI does require technology detail sharing, not just data sharing. AI training techniques, validation and feedback loop engineering, or even infrastructure approaches to improve performance could revolutionize how organizations develop AI, the key to safe, fair AI.[163] Although data and algorithms are important, the techniques and approaches used to create the system itself will likely advance the field at a far greater rate, so long as sufficiently specific details are available.

Given the existing challenges in drafting patents, the United States could consider alternatives for incentivizing disclosure while simultaneously offering some protection for organizations disclosing these details. For example, an alternative to a patent that operates like a limited patent, such as market exclusivity used for generic drugs, might effectively provide enough competitive cover for organizations creating innovative technologies while promoting disclosure.

Similar to reforms, Price and Rai recommend the product-by-process approach to patenting;[164] additionally, the United States might also consider an alternative patent with a shorter monopoly period and more detailed claim requirements.[165] Either of these solutions could prevent competitors from duplicating technology for a limited time period, while simultaneously disclosing the invention, something trade secrecy cannot provide. Organizations could also deposit training data (or at a minimum, the data elements used for training data), host trained machine learning models, and capture feedback to meaningfully animate review and oversight processes.[166] A hosted environment with this information could create a dynamic, rather than static, patent file.[167]

Any of these approaches could include a live hosting environment for the algorithm, which simultaneously demonstrates that the technology works (overcoming enablement concerns if relying on a version of a patent) and promotes adverse testing by competitors.[168] If advanced machine learning algorithms are used, it will be difficult to reverse engineer these algorithms, and a credentialed log-on system with associated validation of qualified

[162] *See id.* at 801–04 (describing potential changes to patent law that would promote more detailed disclosure in a product-by-process model).

[163] *See* Tschider, *supra* note 13, at 692–98.

[164] *See* Price & Rai, *supra* note 18, at 802–03.

[165] *See* Tschider, *supra* note 13, at 721–22.

[166] *See id.* at 721; Price & Rai, *supra* note 18, at 802.

[167] *See* Price & Rai, *supra* note 18, at 803–04.

[168] *See* Tschider, *supra* note 13, at 722. This hosting model mirrors other deposit requirements, such as biological sample deposits and historical patent depositories. 37 C.F.R. § 1.802 (2020); U.S. PAT. & TRADEMARK OFF., § 2164.06(a) EXAMPLES OF ENABLEMENT ISSUES-MISSING INFORMATION, MPEP (9th ed. Rev. 10.2019, June 2020), https://www.uspto.gov/web/offices/pac/mpep/s2164.html [https://perma.cc/NWX9-5DCT].

entities should prevent bad actors from accessing the system.[169] In addition, access to a wide variety of AI algorithms (and the ability to test them!) might inspire different or more AI development.

Such an environment could be managed by a governmental agency like the U.S. Patent and Trademark Office or perhaps another independent agency. A disclosure and hosting environment could also permit customer and patient advocacy groups to give ratings, offer comments, or promote suggestions on how to make these systems safer and fairer. Indeed, such a model could enable those affected by or scored under automated decisioning (or advocates for individuals) to score the scorer.[170] If participation in such an environment is optional, organizations could be persuaded to participate if they are offered some protection from lawsuits or at least a rebuttable presumption of reasonable behavior, similar to information sharing in a cybersecurity context.[171]

C. AI Disclosure in Regulatory Review

Although incentivizing data sharing through government sponsorship and encouraging disclosure through patents might offer significant progress for transparency goals, some legal situations will simply require more detailed forms of explanation than patent information or data deposits can provide. Specifically, these situations may call for mandated disclosure that leverages existing legal protections to minimize excessive disclosure that can harm organizational interests.

Price and Rai recommend that an alternative machine learning algorithm could be used to explain the algorithm(s) in question in the event administrative agencies, such as the FDA or qualified third parties, are positioned to complete *ex ante* invention reviews.[172] *Ex ante* reviews would occur prior to an AI product hitting the market and are valuable because they could prevent potential safety or unfairness issues. For AI, the FDA has promoted a Pre-Certification program for *ex ante* review that could, if mandated, require organizations to disclose information about the quality process for creating AI, but not the details of the AI technology itself. Although I do not share Price and Rai's optimism about the FDA's Pre-

[169] Because reverse engineering can destroy trade secret status for algorithms that cannot otherwise be patented, it should be acknowledged that organizations may resist hosting simple, easily reverse-engineered algorithms in these systems. However, complex systems that pose a greater risk to patients or consumers are more likely to be in greater need of hosting, wherein the possibility of reverse engineering is low. Generally speaking, reverse engineering is permitted unless barred by contract and does not amount to trade secret misappropriation. *See* Varadarajan, *supra* note 34, at 1553–55, 1563–64.

[170] *See* Citron & Pasquale, *supra* note 15, at 3–4.

[171] *See* Brad S. Karp, *Federal Guidance on the Cybersecurity Information Sharing Act of 2015*, HARV. L. SCH. F. ON CORP. GOVERNANCE (Mar. 3, 2016), https://corpgov.law.harvard.edu/2016/03/03/federal-guidance-on-the-cybersecurity-information-sharing-act-of-2015 [https://perma.cc/D25B-CTHD]; Exec. Order No. 14,028, 86 Fed. Reg. 26,633 (May 17, 2021).

[172] *See* Price & Rai, *supra* note 18, at 804–06.

Certification Program's ability to prevent downstream issues,[173] explanation could be useful for administrative review.

Disclosure of AI technology details and sample explanations could reduce the likelihood of safety or unfairness issues downstream. Moreover, the FDA has actually been criticized for its non-disclosure of confidential information disclosed to it in FDA review processes in downstream litigation, illustrating that confidentiality or trade secret destruction risks for *ex ante* reviews are very low. For example, an AI-enabled surgical robot would likely receive a comprehensive FDA review. If the FDA will determine whether such a robot can be marketed, they will likely need to determine whether such a system is safe and efficacious. To do so will likely require some details of the technology and how it renders safe and effective decisions and, for example, whether such decisions are disproportionately more safe and effective for certain communities than for others. As is typical in all FDA reviews, the details of the review process will likely be kept confidential, including sample explanations. Explanation might be even more crucial after a significant problem is discovered.

Organizations that have received FDA approval or have received an alternative path for marketing a product are responsible for adverse event reporting or a continuing obligation to report events that seriously impact the safety of a previously reviewed product.[174] When reporting injuries or property damage to an agency, like the FDA, in a medical device report,[175] the FDA could require organizations to conduct a root cause analysis to determine how the AI algorithm(s)' decision or resulting function resulted in bodily injury, property damage, or discriminatory impact and, after a reasonable time, disclose the corrective measures taken.[176] Co-developed algorithms—some that function and some that disentangle the decisions (such as Price and Rai's machine learning algorithms)—could streamline this process.

It should be noted that of course, other sectors do not have strong *ex ante* and continuous monitoring obligations required by law. For this reason, the FDA actually has an opportunity to do much more to promote safety and fairness in AI than other administrative agencies can. However, other sectors could rely on incentivized disclosure, including live algorithmic hosting to promote broad reviews.

[173] *See* Tschider, *supra* note 62, at 1568–69.

[174] 21 C.F.R. § 803.10 (2020).

[175] *Id.*

[176] *See* Tschider, *supra* note 62, at 1561 n.43; *see, e.g.*, Lizzie Barclay, *Quality Improvement in Theory and Practice: Guiding Principles and a Real-World Incident Fix*, AIDENCE (July 20, 2021), https://www.aidence.com/articles/ai-pms-principles-incident-fix [https://perma.cc/S4XQ-XLZW] (describing post-market surveillance for an AI product, involving root cause analysis). Post-market surveillance activities often involve root cause analysis and disclosure of that analysis. This process could hypothetically reduce the need for forced explanatory disclosure of trade secrets or other confidential information by instead recording the problem and its solution.

D. AI EXPLANATION IN LITIGATION

Promoting *ex ante* regulatory incentives and requirements will likely serve a better function than completely relying on litigation to reveal AI technology safety and fairness issues. AI technologies are typically centrally managed and may make decisions, opaquely, that affect a large number of individuals simultaneously.[177] For this reason, the potential issues are not just significant to an individual, they may be unavoidable to anyone using the technology.[178] However, there are many situations where transparency will be necessary for litigation and potential legal recovery, as well.

For product safety torts or discrimination causes of action, explanation could also be necessary, at least in discovery. For example, a consumer using an AI-enabled home thermostat could bring a cause of action when the thermostat malfunctions, increasing the temperature threshold until a furnace overheats, starting a serious fire. In a case like this, how might a consumer actually demonstrate that the thermostat caused the fire without information about how the thermostat's AI functions? In a case where an employer AI system rated facial expressions in a videotaped interview and failed to pass on interviewers with a specific skin tone or accent, how might any information be gleaned as to how the AI system failed if disclosure is not possible? In both of these cases, it may even be desirable to explain the specific AI algorithmic decision that caused the behavior harming these plaintiffs.

Fortunately, courts have effectively managed confidential and trade secret information for a long time, and it may be possible for courts to require an explanation in a way that protects an AI technology innovator while simultaneously enabling legitimate cases to go forward. The law has long recognized that information is not simply public or private, it may be available to some individuals or organizations while simultaneously unavailable to others.[179] Information disclosure does not necessarily destroy its status; rather, intermediary states are common and expected. In short, informational status is a dynamic, rather than static, state.

One option is Court review of trade secrets *in camera* or with experts bound by confidentiality agreements. During an *in-camera* review, privileged or confidential information, including trade secrets, may be reviewed without disclosing them to the public by discussing in chambers or submitting evidence in sealed, marked, envelopes and in preapproved online

[177] For example, cybersecurity risks wherein a cyberattacker may make changes to a system will likely affect the system holistically, not just one patient. *See* Tschider, *supra* note 144, at 182, 205.

[178] *Id.*

[179] *See* Richards & King, *supra* note 150, at 396. Neil Richards and Jonathan King have discussed the intermediary state of confidential data and noted that such confidentiality, whether involving personal information or not, can retain its status through intermediary states.

environments.[180] Situations that could compromise personal privacy are sometimes also protected. This model of review for AI algorithmic function certainly could preserve trade secrecy for portions of the invention that require protection while providing evidence that could reveal intentionality, disparate impact, or factual causation. Another option might be to require disclosure only when the plaintiff has met some burden. For example, AI technology safety and fairness issues are likely to affect a large number of individuals at one time and for that reason are poised to frequently operate as class-action lawsuits. For this reason, courts could potentially permit a showing of causation circumstantially, such as use of an AI system and a specific group of people who used the system and experienced the same or similar injuries. For example, AI-based class-actions could be scrutinized more heavily for meeting the commonality requirement but allowed some flexibility in meeting the proximate cause prong for a products liability case. Permitting some flexibility in such a finding (when proximate cause may be nearly impossible with substantial AI opacity and lack of explanation), could shift the burden to a defendant, who may be required to explain AI decisions within the context of the case. This model could also leverage alternatives to public disclosure in court.

V. CONCLUSION

Innovation is a centrally important goal in the development of AI positioned to transform our lives. In sectors like healthcare, AI systems have the potential to radically improve health outcomes, reduce costs, and promote self-sufficiency. Without innovation, the promise and opportunity of AI will never be realized. Discrimination, safety, and privacy risks do, however, threaten the potential benefits of such innovation.

Although beyond the scope of this Response, it is crucial to determine whether transparency to address these issues is: 1) possible; and 2) if it is possible, whether it is useful; and 3) if it is possible, and it is useful, whether organizations should be required to disclose confidential or secret information about AI technologies. Specifically, in considering whether transparency is desirable, it is crucial to understand that legal opacity presents a substantial impediment to transparency goals. Because legal opacity functions for many justifiable purposes, transparency goals should consider to what extent these legitimate legal practices and legal requirements should be limited in their application.

AI is currently being used in nearly every sector, from high-stakes decisions like how an insulin pump will function or appropriate government entitlements to comparatively lower-stakes decisions like the optimal temperature for my home to reduce energy costs. Therefore, transparency requirements should accommodate differing sector risk profiles, reducing

[180] Bow Tie Law, *In Camera Review of Claimed Privilege Communications:* Dyson, Inc. v. SharkNinja, EVERLAW (May 3, 2017), https://www.everlaw.com/blog/2017/05/03/in-camera-review-privilege-data [https://perma.cc/778D-LSZ4].

disclosure requirements for low-risk product and services, while prompting greater disclosure for higher-risk and higher-benefit AI. The purpose of specific types of disclosures is important, too, such as whether the United States aims to create safe and fair AI technologies, whether the United States desires to promote AI innovation through data use and sharing, or if the U.S. judicial system wants to facilitate legitimate and lawful plaintiff remedies for AI injuries.

Our current legal ecosystem, operating in favor of opacity, threatens good-faith efforts to bring greater investment, adoption, and confidence in AI technologies. Solving this problem requires a nuanced and combined approach using a variety of policy levers simultaneously to address these complexities while promoting information sharing. By addressing these perspectives contextually and encouraging disclosure of what is specifically needed to attain a particular goal, the United States can promote better AI systems while protecting the organizations developing them.