Beyond the False Dichotomy: Regulating AI Safety, Ethics, and Innovation

Maarten Herbosch*

Artificial Intelligence ("AI") is now at the center of regulatory attention amid debates over the European AI Act and proposals for an AI moratorium. Yet discussion is still framed by a false choice: regulators must supposedly pick between AI safety and ethics or promoting innovation.

This Article demonstrates the limits of a segregated approach. It challenges this oversimplification by dissecting two central dichotomies: first, between regulation for contemporary versus prospective AI, and second, between innovation and safety. By analyzing the regulatory impacts of the EU AI Act and a proposed U.S. regulatory moratorium in the contexts of non-discrimination and liability law, the Article shows that both extremes—comprehensive value-driven regulation and hands-off innovation policies—ultimately fail to achieve their intended aims and, in some respects, undermine their own normative foundations.

Neither sweeping ethical mandates nor regulatory inaction meaningfully address the practical and legal challenges created by AI's opacity, unpredictability, and complexity. Instead, both approaches risk overinclusive or underinclusive frameworks that fragment legal protections and reduce certainty for all stakeholders.

Rejecting the idea that effective AI regulation requires a trade-off, the Article proposes a Pareto principle for AI governance: narrow, targeted procedural safeguards—such as robust documentation and clear accountability mechanisms—can maximize AI safety with minimal cost to innovation. Rather than reinventing the regulatory wheel, these measures should be anchored in established legal regimes, equipping regulators to address both contemporary and prospective AI developments. By moving beyond the false dichotomy, this Article offers a roadmap for sustainable, high-impact AI regulation that preserves both innovation and public trust.

---

* Assistant Professor of Law and Artificial Intelligence, Faculty of Law and Criminology, KU Leuven; Senior Research Scholar, Institute for Law & AI. All opinions expressed are the author's own. Email: maarten.herbosch@kuleuven.be.

INTRODUCTION

The regulation of artificial intelligence (AI) is shaped by a double dichotomy. The first is temporal: some scholars focus on current AI systems and capabilities, while others emphasize prospective—often speculative—future developments.[1] This Article concentrates more fully on a second, and arguably more consequential, dichotomy: the tension between promoting innovation and safeguarding against AI risks, frequently framed as "AI ethics" or "AI safety."[2] This framing risks oversimplification—suggesting that regulation inherently prioritizes safety at the expense of innovation, and thus inevitably impedes technological progress.

This Article is the first to critically examine both dichotomies through a focused legal analysis of AI regulation. It does so by analyzing two starkly divergent approaches to AI regulation. On one side stands the value-driven EU AI Act, which imposes stringent obligations on AI providers to uphold fundamental rights.[3] On the other is a recurring U.S. proposal for an AI moratorium—an attempt to suspend all AI-related regulation at the state level.[4]

The Article argues that both the "hands-off" approach and the European value-centric model lack the nuance necessary to effectively safeguard either innovation or AI safety and ethics. Drawing on analyses of liability[5]

---

1. *See infra* Part I.
2. *See infra* Section I.B.
3. Regulation 2024/1689, 2024 O.J. (L 1689) 1 (EU).
4. The House of Representatives-passed version contained a ten-year moratorium on state AI regulation. *See* One Big Beautiful Bill Act, H.R. 1, 119th Cong. § 43201(c)(1) (as passed by House, May 22, 2025) ("Except as provided in paragraph (2), no State or political subdivision thereof may enforce, during the 10-year period beginning on the date of the enactment of this Act, any law or regulation of that State or a political subdivision thereof limiting, restricting, or otherwise regulating artificial intelligence models, artificial intelligence systems, or automated decision systems entered into interstate commerce."). This version of the Act was subsequently struck in the Senate by a 99–1 vote on July 1, 2025. David Morgan & David Shepardson, *US Senate Strikes AI Regulation Ban from Trump Megabill*, REUTERS (July 1, 2025), https://www.reuters.com/legal/government/us-senate-strikes-ai-regulation-ban-trump-megabill-2025-07-01/.
5. *See, e.g.*, Marguerite E. Gerstner, *Liability Issues with Artificial Intelligence Software*, 33 SANTA CLARA L. REV. 239 (1993); David C. Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117, 129–50 (2014); Anat Lior, *AI Strict Liability Vis-À-Vis AI Monopolization*, 22 COLUM. SCI. & TECH. L. REV. 90, 94–106 (2020); Andrew D. Selbst, *Negligence and AI's Human Users*, 100 B.U. L. REV. 1315, 1318–76 (2020).

and non-discrimination[6] law—two areas central to AI governance—it demonstrates how either approach ultimately risks undermining its own normative objectives. As an alternative, this Article outlines the contours of a "Pareto optimal" regime—one that minimizes regulatory burden while maximizing ethical impact. Such a regime, it contends, should be anchored in existing legal structures and emphasize procedural mechanisms capable of supporting the governance of evolving AI technologies.[7]

The Article begins in Part I by outlining the two dichotomies identified above, highlighting some of their adverse impacts. Part II then turns to the legal challenges that have been widely recognized in the literature. Part III evaluates the effects of both "extreme" approaches—the value-driven AI Act and the innovation-oriented moratorium—on the goals of value protection and innovation, concluding that neither achieves an optimal balance. This assessment, however, sets the stage for Part IV, which identifies targeted, low-cost, high-impact regulatory tools and proposes a Pareto principle for AI regulation.

I. TWO DICHOTOMIES

Discussions about AI regulation are largely dominated by two dichotomies. A first dichotomy relates to the types of AI that are being discussed. Various contributions focus on either potential, though currently hypothetical, future AI systems, whereas others focus on existing capabilities. Authors do not always make that choice explicit, sometimes leading to confusion.

---

6. *See, e.g.*, Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 679–82 (2017); Stephanie Bornstein, *Antidiscriminatory Algorithms*, 70 ALA. L. REV. 519, 553–67 (2018) (applying Title VII to algorithmic discrimination); Jon Kleinberg et al., *Discrimination in the Age of Algorithms*, 10 J. LEGAL ANALYSIS 113, 120–21 (2019) (introducing the idea that algorithms can help combat discrimination); Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2221–22 (2019) (discussing the challenge in the context of criminal law); Anya E.R. Prince & Daniel Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, 105 IOWA L. REV. 1257, 1260–61 (2020) (describing how AI can result in proxy discrimination); Sandra Wachter et al., *Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI*, COMPUT. L. & SEC. REV., July 2021, at 1, 5–6 (neatly identifying the key challenges); Pauline T. Kim, *Race-Aware Algorithms: Fairness, Nondiscrimination and Affirmative Action*, 110 CALIF. L. REV. 1539, 1541–47 (2022) (considering whether race-aware algorithms can help combat discrimination while complying with non-discrimination law); Cass R. Sunstein, *Governing by Algorithm? No Noise and (Potentially) Less Bias*, 71 DUKE L.J. 1175, 1196–203 (2022) (considering how algorithms can help make disparate impact more transparent).

7. *See infra* Part IV.

A second dichotomy relates to the desirability of regulation, and what shape or goal that regulation should adopt. Two extremes on the spectrum of this dichotomy are the European Union's approach to adopt an encompassing AI regulatory framework,[8] and the opposing American suggestion to implement a regulatory moratorium for AI.[9] This second dichotomy encompasses two aspects. Depending on the perspective, it is mainly concerned with the legal protection of existing ethical values in AI contexts, or with the impact that value protection might have on innovation. The next Sections sketch each of those dichotomies in more detail.

A. Contemporary v. Prospective

1. General

AI has been the subject of legal debate for several decades.[10] The types of AI models and systems discussed during that time have varied widely—from more traditional systems[11] to state-of-the-art[12] or even speculative, futuristic models.[13] These differing approaches are often obscured by the shared label 'artificial intelligence.' The following Sections examine this notion to clarify the legal challenges AI systems pose today. This analysis helps frame the temporal dichotomy discussed below and is essential to providing a nuanced account of the legal issues arising in the areas of liability and non-discrimination law.

---

8. Regulation 2024/1689, *supra* note 3.

9. *See, e.g.*, One Big Beautiful Bill Act, H.R. 1, 119th Cong. (2025); *see also* Exec. Order No. 14,365, *Ensuring a National Policy Framework for Artificial Intelligence*, 90 Fed. Reg. 58499 (Dec. 11, 2025) (more recently establishing an "AI Litigation Task Force" charged with challenging State AI laws that obstruct innovation).

10. *See, e.g.*, Lawrence B. Solum, *Legal Personhood for Artificial Intelligences*, 70 N.C. L. REV. 1231, 1231–34 (1992); Curtis E.A. Karnow, *Liability for Distributed Artificial Intelligences*, 11 BERKELEY TECH. L.J. 147, 148–49 (1996).

11. *See, e.g.*, George S. Cole, *Tort Liability for Artificial Intelligence and Expert Systems*, 10 COMPUT. L.J. 127, 131 (1990).

12. *See, e.g.*, Nadia Banteka, *Artificially Intelligent Persons*, 58 HOUS. L. REV. 537, 543–44 (2021); Maarten Herbosch, *Liability for AI Agents*, 26 N.C. J.L. & TECH. 391, 397–98 (2025).

13. *See, e.g.*, Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J.L. & TECH. 353, 365 (2016). Some authors have discussed capabilities that were, at the time, futuristic. *See* Sam N. Lehman-Wilzig, *Frankenstein Unbound: Towards a Legal Definition of Artificial Intelligence*, 13 FUTURES 442, 442–43 (1981); Solum, *supra* note 10, at 1234.

2. Contemporary AI

a. Definition

While the term "artificial intelligence" has historically been interpreted broadly, the legal challenges posed by traditional AI differ markedly from those raised by modern or emerging systems. Over time, the definition of AI has shifted: where earlier conceptions emphasized the simulation of human intelligence,[14] more recent definitions highlight autonomy, adaptiveness, and inference. A clear example is the EU AI Act,[15] which defines an AI system as "a machine-based system that is designed to operate with varying levels of autonomy"[16] and capable of generating outputs—such as predictions or decisions—that influence physical or virtual environments. This shift underscores key legal concerns, particularly around autonomy, complexity, and accountability.

Machine learning exemplifies this autonomy particularly well.[17] In such systems, the programmer does not prescribe how outputs should be derived; rather, the system is equipped with mechanisms to "learn,"[18] often through processing vast quantities of training data.[19] During training, the system adapts to better align with statistical correlations between input data and the

---

14. *See* John McCarthy et al., A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence 2 (Aug. 31, 1955) (on file with the Dartmouth College and Stanford University Libraries); Maura R. Grossman & Gordon V. Cormack, *The Grossman-Cormack Glossary of Technology-Assisted Review*, 7 FED. CTS. L. REV. 1, 7 (2013); Shlomit Yanisky-Ravid, *Generating Rembrandt: Artificial Intelligence, Copyright, and Accountability in the 3A Era—The Human-Like Authors Are Already Here—A New Model*, 2017 MICH. ST. L. REV. 659, 673; Ronald Yu & Gabriele Spina Ali, *What's Inside the Black Box? AI Challenges for Lawyers and Researchers*, 19 LEGAL INFO. MGMT. 2, 2 (2019).

15. Regulation 2024/1689, *supra* note 3.

16. *Id.* art. 3(1); *see also Commission Guidelines on the Definition of an Artificial Intelligence System Established by Regulation (EU) 2024/1689 (AI Act)*, COM (2025) 5053 final (July 29, 2025). For a similar but less explicit definition, see 15 U.S.C. § 9401(3). *See also* S. 358, 2025 Gen. Assemb., Jan. Sess. (R.I. 2025); Ill. Pub. Act 103-0804 (effective Jan. 1, 2026) (amending 775 Ill. Comp. Stat. 5/2-101 and 5/2-102); S.B. 53, 2025–2026 Leg., Reg. Sess. (Cal. 2025) (inserting CAL. BUS. & PROF. CODE § 22757.11(b)).

17. *See, e.g.*, Weston Kowert, *The Foreseeability of Human-Artificial Intelligence Interactions*, 96 TEX. L. REV. 181, 183 (2017).

18. Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 88 (2014); *see* Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 3 U. BOLOGNA L. REV. 180, 185 (2018).

19. *See* Surden, *supra* note 18, at 89; KEVIN D. ASHLEY, ARTIFICIAL INTELLIGENCE AND LEGAL ANALYTICS 234 (2017); Harry Surden, *Artificial Intelligence and Law: An Overview*, 35 GA. ST. U. L. REV. 1305, 1311 (2019).

desired output.[20] As a result, system developers do not need to identify the relationship between the input and the output themselves.[21]

b. Key Challenges

AI techniques, and those based on machine learning in particular, routinely produce algorithms that execute millions—or even billions—of computations to map inputs to outputs.[22] This introduces two principal challenges, especially in comparison with traditional computer algorithms. First, the output a system will produce for a given input is often largely unpredictable.[23] This is by design: the programmer has not encoded a specific decision-making procedure but has instead enabled the system to devise its own.[24] Substituting the programmer's conceptual reasoning with autonomously identified statistical correlations is precisely what allows AI

---

20. *See* Surden, *supra* note 18, at 89; ASHLEY, *supra* note 19, at 234; Iria Giuffrida et al., *A Legal Perspective on the Trials and Tribulations of AI: How Artificial Intelligence, the Internet of Things, Smart Contracts, and Other Technologies Will Affect the Law*, 68 CASE W. RSRV. L. REV. 747, 753 (2018); Surden, *supra* note 19, at 1311; *cf.* Janneke Gerards & Frederik Zuiderveen Borgesius, *Protected Grounds and the System of Non-Discrimination Law in the Context of Algorithmic Decision-Making and Artificial Intelligence*, 20 COLO. TECH. L.J. 1, 6 (2022) (describing AI's ability to find correlations in data sets); Raphaële Xenidis, *When Computers Say No: Towards a Legal Response to Algorithmic Discrimination in Europe*, *in* RESEARCH HANDBOOK ON LAW AND TECHNOLOGY 222, 230 (Bartosz Brożek et al. eds., 2024) (describing how algorithms can lead to impermissible discrimination, which can be difficult to assess in terms of "causation" as machines operate on the basis of correlation).

21. Surden, *supra* note 19, at 1314; *see* Emmanuel Gbenga Dada et al., *Machine Learning for Email Spam Filtering: Review, Approaches and Open Research Problems*, HELIYON, June 2019, at 1, 2–4 (discussing spam filters using AI).

22. *See, e.g.*, Giuffrida et al., *supra* note 20, at 755; Hsiao-Ying Lin, *Large-Scale Artificial Intelligence Models*, COMPUT., May 2022, at 76, 76–78.

23. Kowert, *supra* note 17, at 183; Mark A. Lemley & Bryan Casey, *Remedies for Robots*, 86 U. CHI. L. REV. 1311, 1334 (2019); Yu & Ali, *supra* note 14, at 5; Selbst, *supra* note 5, at 1342; *see* Matthew Oliver, *Contracting by Artificial Intelligence: Open Offers, Unilateral Mistakes, and Why Algorithms Are Not Agents*, AUSTL. NAT'L U. J.L. & TECH., Autumn 2021, at 45, 50.

24. *See, e.g.*, Surden, *supra* note 19, at 1314; Dada et al., *supra* note 21, at 2–4.

systems to exceed human expert performance in various domains.[25] The resulting statistical processes, however, do not reflect human logic.[26]

Second, these programming techniques can result in highly complex systems,[27] which should not come as a surprise given the potential millions or billions of internal computations to determine the output.[28] This, combined with the fact that they do not follow any particular line of conceptual logic,[29] often makes the system's output effectively inexplicable.[30] This opacity stems from the complex,[31] dynamic,[32] and largely statistical nature of the "decision-making"[33] process in such AI systems and, much like their unpredictability, is arguably part of what enables them to surpass conceptual human reasoning. While traditional computer algorithms can also be highly complex, they have—typically, one would hope—remained understandable to the individuals who developed them. That is not the case for many AI systems.[34]

---

25. *See* Ryan Calo, *Singularity: AI and the Law*, 41 SEATTLE U. L. REV. 1123, 1124 (2018); Michael Hatfield, *Professionally Responsible Artificial Intelligence*, 51 ARIZ. ST. L.J. 1057, 1060 (2019); Wojciech Samek & Klaus-Robert Müller, *Towards Explainable Artificial Intelligence*, *in* EXPLAINABLE AI: INTERPRETING, EXPLAINING AND VISUALIZING DEEP LEARNING 5, 5–6 (Wojciech Samek et al. eds., 2019); Brian Judge et al., *When Code Isn't Law: Rethinking Regulation for Artificial Intelligence*, 44 POL'Y & SOC'Y 85, 86 (2025).

26. *See* Daniel Martin Katz, *Quantitative Legal Prediction—Or—How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry*, 62 EMORY L.J. 909, 918 (2013); Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, BIG DATA & SOC'Y, Jan.–June 2016, at 1, 2; Giuffrida et al., *supra* note 20, at 755–56; Surden, *supra* note 19, at 1315; Wachter et al., *supra* note 6, at 3.

27. *See* Brent Daniel Mittelstadt et al., *The Ethics of Algorithms: Mapping the Debate*, BIG DATA & SOC'Y, July–Dec. 2016, at 1, 6.

28. *See* Giuffrida et al., *supra* note 20, at 755; Lin, *supra* note 22, at 76–78.

29. *See* Katz, *supra* note 26, at 918; Burrell, *supra* note 26, at 2; Giuffrida et al., *supra* note 20, at 755; Surden, *supra* note 19, at 1315; Wachter et al., *supra* note 6, at 3.

30. *See* Burrell, *supra* note 26, at 1–2; Mittelstadt et al., *supra* note 27, at 7; Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J.L. & TECH. 889, 897 (2018); Hatfield, *supra* note 25, at 1118 n.278; Samek & Müller, *supra* note 25, at 6; Yu & Ali, *supra* note 14, at 5; Alicia Solow-Niederman, *Administering Artificial Intelligence*, 93 S. CAL. L. REV. 633, 657 (2020); Benjamin Bartlett, *Clinical Negligence in an Age of Machine Learning: Res Ipsa Loquitur to the Rescue?*, 15 J. EUR. TORT L. 295, 297 (2024).

31. Mittelstadt et al., *supra* note 27, at 6.

32. AI systems can be further trained during their use. *See, e.g.*, Mittelstadt et al., *supra* note 27, at 6; Selbst, *supra* note 5, at 1332–33; Kowert, *supra* note 17, at 184.

33. *See* Katz, *supra* note 26, at 918; Burrell, *supra* note 26, at 2; Giuffrida et al., *supra* note 20, at 755; Surden, *supra* note 19, at 1315; Wachter et al., *supra* note 6, at 3.

34. *See* Burrell, *supra* note 26, at 1–2; Mittelstadt et al., *supra* note 27, at 7; Bathaee, *supra* note 30, at 897; Hatfield, *supra* note 25, at 1118 n.278; Samek & Müller, *supra* note 25,

Consequently, it is unsurprising that explainable AI has been a major research objective for some time.[35] Moreover, some scholars argue that restricting the use of systems to only those that are explainable inevitably reduces accuracy,[36] as it limits the techniques available to optimize performance. Others, however, contend that increased explainability can lead to improved performance over time.[37] This is, in part, due to the fact that explainability helps optimize human supervision.[38]

In either case, it is important to note that explainability remains, to some extent, limited for many of the most widely used—and highly complex—AI systems.[39] Likewise, accuracy is also inherently constrained. For most applications, developing a perfect AI system is impossible.[40] While performance is difficult to assess uniformly in the context of AI systems,[41] this difficulty partly arises from the statistical foundations of such systems. This is significant, as it implies that gains in accuracy tend to yield

---

at 6; Yu & Ali, *supra* note 14, at 5; Solow-Niederman, *supra* note 30, at 657; Bartlett, *supra* note 30, at 297.

35. *See* Amina Adadi & Mohammed Berrada, *Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)*, 6 IEEE ACCESS 52138, 52139 (2018); Ashley Deeks, *The Judicial Demand for Explainable Artificial Intelligence*, 119 COLUM. L. REV. 1829, 1833–34 (2019); Amy L. Stein, *Assuming the Risks of Artificial Intelligence*, 102 B.U. L. REV. 979, 1005–06 (2022); Paulo Henrique Padovan et al., *Black Is the New Orange: How to Determine AI Liability*, 31 A.I. & L. 133, 151–58 (2023).

36. Deeks, *supra* note 35, at 1834; *see* Mengnan Du et al., *Techniques for Interpretable Machine Learning*, 63 COMMC'NS ACM, Jan. 2020, at 68, 70; Arun Rai, *Explainable AI: From Black Box to Glass Box*, 48 J. ACAD. MKTG. SCI. 137, 138 (2020).

37. *See* Cynthia Rudin, *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead*, 1 NATURE MACH. INTEL. 206, 207 (2019); LEONIDA GIANFAGNA & ANTONIO DI CECCO, EXPLAINABLE AI WITH PYTHON 10 (1st ed. 2021).

38. *See* Du et al., *supra* note 36, at 75.

39. *See, e.g.*, *id.*

40. *See* Nathan A. Greenblatt, *Self-Driving Cars and the Law*, IEEE SPECTRUM, Feb. 2016, at 46, 48, 50; Azim Shariff et al., *Psychological Roadblocks to the Adoption of Self-Driving Vehicles*, 1 NATURE HUM. BEHAV. 694, 695 (2017); Bryan H. Choi, *Crashworthy Code*, 94 WASH. L. REV. 39, 79–86 (2019) (discussing computer systems more generally); Lemley & Casey, *supra* note 23, at 1327–28.

41. *See* Bart W. Schermer, *The Limits of Privacy in Automated Profiling and Data Mining*, 27 COMPUT. L. & SEC. REV. 45, 48 (2011); GOPINATH REBALA ET AL., AN INTRODUCTION TO MACHINE LEARNING 60–62 (2019); John Mongan et al., *Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers*, RADIOLOGY: A.I., Mar. 2020, at 1, 1.

diminishing returns at increasingly high costs in terms of time, training data, and overall effort.[42]

3. Conflation

a. General

The level and extent of these challenging properties vary with the increased autonomy of AI systems. As AI becomes more autonomous, its output becomes less foreseeable and could end up being less explainable as well.[43] In other words, the precise extent of these challenging properties and corresponding legal challenges attributed or attached to them illustrate how important it is to be clear about the type of AI systems and their capabilities that one is discussing.

In any case, a general observation is that any form of AI regulation—whether inspired by existing challenges or those anticipated in the future—should be capable of addressing the current state of the art in AI. Put differently, even if regulation is premised on prospective AI properties, it should ideally also resolve the challenges presented by contemporary AI, as discussed below.[44]

Another general observation is that it is rare in legal scholarship to be explicit about those capabilities.[45] A frequent classification that does get used distinguishes between contemporary so-called "weak" AI,[46] in contrast to the prospective[47]—or, some argue, hypothetical[48]—"strong" AI (or

---

42. Andrew Majot & Roman Yampolskiy, *Diminishing Returns and Recursive Self Improving Artificial Intelligence*, *in* THE TECHNOLOGICAL SINGULARITY: MANAGING THE JOURNEY 141, 148 (Victor Callaghan et al. eds., 2017); *see* Neil C. Thompson et al., *Deep Learning's Diminishing Returns: The Cost of Improvement Is Becoming Unsustainable*, IEEE SPECTRUM, Oct. 2021, at 51, 55.

43. *See supra* note 30 and accompanying text.

44. *See infra* Section III.

45. As an example of more empirical work incorporating AI benchmarks, see, for example, Varun Magesh et al., *Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools*, 22 J. EMPIRICAL LEGAL STUD. 216, 232–33 (2025). *See* Daniel Kokotajlo et al., *AI 2027*, AI FUTURES PROJECT (Apr. 3, 2025), https://ai-2027.com/ai-2027.pdf [https://perma.cc/7RDB-LTJ9] (providing a more explicit example of future AI capabilities).

46. *See* Ted Goertzel, *The Path to More General Artificial Intelligence*, 26 J. EXPERIMENTAL & THEORETICAL A.I. 343, 344 (2014); Brian L. Frye, *The Lion, the Bat & the Thermostat: Metaphors on Consciousness*, 5 SAVANNAH L. REV. 13, 18 (2018).

47. On the existing impossibility of strong AI, see Roni A. Elias, *Facing the Brave New World of Killer Robots: Adapting the Development of Autonomous Weapon Systems into the Framework of the International Law of War*, 21 TRINITY L. REV. 70, 72 (2016); Frye, *supra*

"artificial general intelligence"),[49] which would entail autonomous behavior across any domain, akin to human intelligence.[50] In the case of weak AI, while the specific output may be unpredictable, the range of possible outputs within the system's defined domain remains predictable, due precisely to its limited, "weak" nature.

A related conflation appears when many authors—as well as the European legal framework—refer to "general-purpose AI model[s]"[51] or "foundation models"[52] or frontier AI. Interestingly, though, those models, too, in their existing forms do not (yet) meet the threshold required to constitute AGI or strong AI.[53] As a result, they constitute a subsection of the existing "weak" AI models and systems.

While contemporary AI presents clear challenges, including various challenges for existing legal frameworks, ongoing debates surrounding prospective AI systems together with an eager reliance on technological exceptionalism sometimes leads to confusion or inaccurate legal assessment. Ideally, if authors discuss prospective rather than contemporary AI systems and challenges, they should emphasize that clearly as many do.[54] Moreover, even discussions of existing AI systems and capabilities can lead to some confusion and poor scholarly debate in the sense that the

---

note 46, at 19; Stephen E. Henderson, *Should Robots Prosecute and Defend?*, 72 OKLA. L. REV. 1, 7–8 (2019); John Linarelli, *Artificial General Intelligence and Contract*, 24 UNIF. L. REV. 330, 331 (2019); Patric M. Reinbold, *Taking Artificial Intelligence Beyond the Turing Test*, 2020 WIS. L. REV. 873, 874; Robin Feldman & Kara Stein, *AI Governance in the Financial Industry*, 27 STAN. J.L. BUS. & FIN. 94, 102 (2022).

48. John O. McGinnis, *Accelerating AI*, 104 NW. U. L. REV. COLLOQUY 366, 369 (2010).

49. *See* Elias, *supra* note 47, at 87; Jason Chung & Amanda Zink, *Hey Watson—Can I Sue You for Malpractice? Examining the Liability of Artificial Intelligence in Medicine*, 11 ASIA PAC. J. HEALTH L. & ETHICS 51, 53 (2018); Frye, *supra* note 46, at 19; Henderson, *supra* note 47, at 3; Selbst, *supra* note 5, at 1344.

50. *See* Elias, *supra* note 47, at 87; Frye, *supra* note 46, at 19; Henderson, *supra* note 47, at 3.

51. *See* Regulation 2024/1689, *supra* note 3, art. 3(63).

52. *See, e.g.*, S.B. 53, 2025–2026 Leg., Reg. Sess. (Cal. 2025) (inserting CAL. BUS. & PROF. CODE § 22757.11(f)); Miriam Vogel et al., *Is Your Use of AI Violating the Law? An Overview of the Current Legal Landscape*, 26 N.Y.U. J. LEGIS. & PUB. POL'Y 1029, 1088 (2024).

53. *See* Celia Ford, *Teaching AI to Learn*, TRANSFORMER (Jan. 22, 2026), https://www.transformernews.ai/p/teaching-ai-to-continual-learning [https://perma.cc/6467-L8CB] (framing "truly general-purpose AI" as a future objective rather than as describing today's AI).

54. *See* Scherer, *supra* note 13, at 365; Lehman-Wilzig, *supra* note 13, at 442–43; Solum, *supra* note 10, at 1234.

underlying technology develops at a very rapid pace—limiting the relevance of some older assessments.

b. AI Exceptionalism

More problematically, however, prospective AI properties and capabilities are sometimes attributed to existing or prior AI systems[55]—likely the result of an overeager assumption about the disruptive or exceptional nature of AI. The associated belief in AI's disruptive nature is part of technological exceptionalism, which also relates to the idea that law "fails to keep up" with technology—the so-called pacing problem,[56] rooted in the notion that law is inherently retrospective[57] and must continually adapt. The doctrine of technological exceptionalism stems from this premise: certain technologies are viewed as inherently disruptive[58] or as escaping[59] the bounds of existing law, thus "driving"[60] legal reform.

Within this framework, exceptionalism operates as a policy tool[61] based on the conviction that certain technologies—such as AI—demand a systematic legal overhaul to preserve societal justice.[62] This approach draws

---

55. *See infra* note 74 and accompanying text.

56. *See* Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 TEX. L. REV. 553, 586 (1998) ("[T]oday's regulations may easily pertain to yesterday's technologies."); Gary E. Marchant, *The Growing Gap Between Emerging Technologies and the Law*, *in* THE GROWING GAP BETWEEN EMERGING TECHNOLOGIES AND THE LEGAL-ETHICAL OVERSIGHT 19, 22–23 (Gary E. Marchant et al. eds., 2011); Meg Leta Jones, *Does Technology Drive Law? The Dilemma of Technological Exceptionalism in Cyberlaw*, 2018 U. ILL. J.L. TECH. & POL'Y 249, 251. For a discussion of the pacing problem in the AI context, see Marco Almada & Nicolas Petit, *The EU AI Act: Between the Rock of Product Safety and the Hard Place of Fundamental Rights*, 62 COMMON MKT. L. REV. 85, 97–103 (2025).

57. *See* Giuffrida et al., *supra* note 20, at 750; Surden, *supra* note 19, at 1306.

58. *See, e.g.*, Joshua Schoonmaker, *Proactive Privacy for a Driverless Age*, 25 INFO. & COMMC'NS. TECH. L. 96, 97 (2016); Alessandro Miasato & Fabiana Reis Silva, *Artificial Intelligence as an Instrument of Discrimination in Workforce Recruitment*, 8 ACTA UNIV. SAPIENTIAE, LEGAL STUD. 191, 193–94 (2019) (discussing the disruption AI causes); Horst Eidenmüller & Faidon Varesis, *What Is an Arbitration? Artificial Intelligence and the Vanishing Human Arbitrator*, 17 N.Y.U. J.L. & BUS. 49, 51 (2020).

59. On the pacing problem, see, for example, Schoonmaker, *supra* note 58, at 97 ("This sort of technological disruption is shaking a number of legal realms as they struggle to keep up with the relentless pace of innovation.").

60. *But see* Jones, *supra* note 56, at 260–84; Margot E. Kaminski, *Technological "Disruption" of the Law's Imagined Scene: Some Lessons from* Lex Informatica, 36 BERKELEY TECH. L.J. 883, 895 (2021).

61. *See* Jones, *supra* note 56, at 284.

62. *See* Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 552 (2015); Jones, *supra* note 56, at 253.

a binary distinction between technologies that can be accommodated within existing legal frameworks and those considered so disruptive as to require novel regulatory innovation.[63]

While this paradigm dominates[64] current approaches to AI regulation[65]—often without rigorous justification—it is not the only possible path. Frank Easterbrook's critique of cyberlaw as a mere "law of the horse"[66] anticipated the risks of excessive legal fragmentation. His argument, read charitably,[67] highlights the value of identifying an interdisciplinary legal core rather than multiplying technology-specific carve-outs. Nevertheless, the perception that "AI is different"[68] endures, reflecting the same logic of exceptionalism.[69]

Overreliance on technological exceptionalism presents two principal drawbacks. First, it risks diminishing human agency.[70] Most technologies, including AI, function as tools rather than legal subjects.[71] Humans generally retain the capacity to understand and mitigate technological limitations,[72] and must remain accountable under existing law. The opacity of technology should not be used to justify the displacement of legal responsibility.

---

63. Jones, *supra* note 56, at 250.

64. *Id.* at 251–52, 254 (discussing technological determinism); *see* Gaia Bernstein, *Toward a General Theory of Law and Technology: Introduction*, 8 MINN. J.L. SCI. & TECH. 441, 441–42 (2007); MATTHIJS M. MAAS, ARCHITECTURES OF GLOBAL AI GOVERNANCE: FROM TECHNOLOGICAL CHANGE TO HUMAN CHOICE 295 (2025).

65. *See, e.g.*, Judge et al., *supra* note 25, at 85–87 (discussing how AI's opacity is said to make it impossible to analyze, specify or audit against regulations—effectively rendering any regulation of AI, let alone through the existing legal framework, moot).

66. Frank H. Easterbrook, *Cyberspace and the Law of the Horse*, 1996 U. CHI. LEGAL F. 207, 207–08.

67. Less charitably or more directly, Easterbrook was convinced that law courses should be limited to areas where they could encompass an entire legal field as to not miss out on unifying principles. *See id.* at 207.

68. *See* Calo, *supra* note 62, at 551–52.

69. *See id.* at 550–51.

70. In this sense, it corresponds (to some degree) to substantive theories of technology that emphasize how individuals may be impacted by technology. *See* Arthur Cockfield & Jason Pridmore, *A Synthetic Theory of Law and Technology*, 8 MINN. J.L. SCI. & TECH. 475, 475–76 (2007).

71. Nevertheless, many authors would advocate a treatment as legal subjects. *See, e.g.*, Solum, *supra* note 10, at 1231; Vladeck, *supra* note 5, at 121 (on "strong" AI); Anat Lior, *AI Entities as AI Agents: Artificial Intelligence Liability and the AI Respondeat Superior Analogy*, 46 MITCHELL HAMLINE L. REV. 1043, 1065–71 (2020).

72. *See* Cockfield & Pridmore, *supra* note 70, at 480.

Second, exceptionalism often overlooks both the complexity and the legitimacy of existing legal frameworks. It may oversimplify the nuanced character of these frameworks[73] or the particular challenges posed by AI systems,[74] leading to the conflations discussed above, and may disregard the extent to which current frameworks already reflect value judgments and policy trade-offs.[75] Ignoring this foundational grounding undermines both the coherence and the enforceability of legal regimes. Moreover, it is notable that "pure" exceptionalism has historically achieved limited success,[76] as genuine legal disruption remains rare.[77]

To be clear, exceptionalism can also be valuable in more nuanced forms. It can reveal inconsistencies or gaps within existing regimes,[78] prompting targeted updates while preserving core legal principles.[79] The resulting legal analysis and perceptions can help guide technological development, suggesting a far more reciprocal relationship than the pacing problem implies.[80] In the present context, however, the frequent conflation of contemporary and prospective AI capabilities and challenges is especially relevant.

B. Safety & Values v. Innovation

A second apparent dichotomy in debates over AI regulation concerns the perceived trade-off between regulation designed to protect AI safety and

---

73. *See infra* Section III.A on the challenges presented by causation in AI contexts.

74. *See* Jones, *supra* note 56, at 256–57; Kaminski, *supra* note 60, at 885 n.7. In an AI context, this is also evident from the eagerness of authors to (implicitly) attribute artificial general intelligence properties to existing AI systems. *See, e.g.*, Bathaee, *supra* note 30, at 924 (indicating a causation challenge exists for tort law, which implies attribution a lack of foreseeability to AI systems that implies artificial general intelligence rather than present-day AI).

75. *See infra* Section I.B.1.a.

76. Exceptionalism has historically failed to accurately describe legal responses to technological innovation. *See* SHEILA JASANOFF, SCIENCE AT THE BAR 22 (1995); Jones, *supra* note 56, at 260–77.

77. *See* Calo, *supra* note 62, at 558.

78. *See* LAWRENCE LESSIG, CODE: VERSION 2.0 25 (2006); Calo, *supra* note 62, at 552; Kaminski, *supra* note 60, at 892, 895; *see also* Jack M. Balkin, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 N.Y.U. L. REV. 1, 2 (2004) (discussing how technology may lead to new perspectives on existing legal frameworks).

79. *See* Jones, *supra* note 56, at 281–84.

80. *See* Kaminski, *supra* note 60, at 892.

(fundamental) values and the promotion of competitiveness or innovation.[81] In this framing, the absence of regulation is seen as optimal for innovation,[82] whereas regulation aimed at safeguarding safety or, more broadly, ethical values—whether general or AI-specific—is assumed to impede innovation and undermine AI competitiveness.[83] In this discussion, we will focus primarily on AI ethics, which has historically received significantly more attention within the legal community than the adjacent and partly overlapping[84] field of AI safety, which is defined broadly in this Article to also include short-term concerns such as robustness, reliability, bias and accountability.[85]

This Article challenges several of these claims in greater detail below. To properly contextualize them, it is first useful to briefly examine the roles that the shared underlying values—ethical principles—typically occupy within the legal framework. For purposes of this Article, ethics is defined broadly as the study of such normative values.[86] Notably, the significance of ethics has arguably been magnified in certain AI regulatory contexts, resulting in what may be termed an "ethics capture" of the framework. This Article thus refines the framing of the perceived dichotomy before analyzing it more specifically in relation to liability law and non-discrimination law in the following Sections.

---

81. *See, e.g.*, Deepika Chhillar & Ruth V. Aguilera, *An Eye for Artificial Intelligence: Insights into the Governance of Artificial Intelligence and Vision for Future Research*, 61 BUS. & SOC'Y 1197, 1225 (2022); *see also America's AI Action Plan*, WHITE HOUSE 3 (July, 2025), https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf [https://perma.cc/HG4P-XCLE].

82. *See, e.g.*, *America's AI Action Plan* , *supra* note 81, at 3.

83. *See id.*

84. Interestingly, the distinction between two fields is largely one of emphasis that relates to the first dichotomy, discussed above. In AI ethics, the stronger focus is on present-day AI, whereas AI safety is often concerned with more powerful AI. *See* Stephen Cave & Seán S. ÓhÉigeartaigh, *Bridging Near- and Long-Term Concerns About AI*, 1 NATURE MACH. INTEL. 5, 5 (2019).

85. *See id.* at 5.

86. See in this sense Mike Ananny, *Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness*, 41 SCI., TECH., & HUM. VALUES 93, 94 (2016); SANNE TAEKEMA & WIBREN VAN DER BURG, CONTEXTUALISING LEGAL RESEARCH 227 (2024). This approach falls within the deontological branch of ethical theory. *See* John C. Merrill, *Theoretical Foundations for Media Ethics*, *in* CONTROVERSIES IN MEDIA ETHICS 3, 11 (A. David Gordon et al. eds., 3d ed. 2011).

1. Ethics' Traditional Role

Ethics typically serves two primary roles in relation to law. First, ethics often acts as a precursor to law, with legal rules frequently grounded in underlying ethical principles.[87] Second, ethical guidelines commonly supplement the law, particularly in areas where legal norms reflect not only ethical considerations but also additional factors.

a. Ethics as a Primer

First, ethical principles frequently serve as the foundation for legal rules. This "bottom up" function[88] is readily apparent from several of the AI-related ethical requirements described below. For example, the ethical principle of accountability underpins much of liability law.[89] In this sense, the objective of law is sometimes characterized as the promotion of ethical behavior.[90]

The law, however, typically operates at a general or societal level, rather than prescribing individual ethical conduct.[91] As a result, legal rules rarely mirror ethical principles in a direct or absolute fashion. The creation of legal rules involves the balancing of ethical principles against other, sometimes competing, ethical norms,[92] as well as practical considerations such as economic efficiency, national security, public health, feasibility,[93] or

---

87. See, for example, for fundamental rights: Emma Ruttkamp-Bloem, *Restating the Role of Ethics in AI Law and Regulation*, 1 J. AI L. & REGUL. 259, 259 (2024). See more generally on law reflecting ethics: TAEKEMA & VAN DER BURG, *supra* note 86, at 228.

88. Ruttkamp-Bloem, *supra* note 87, at 260 (using the phrase "bottom up" to describe the role of AI ethics due to its focus on "building social resilience" and acting as a "watchdog for international law").

89. *See, e.g.*, Richard J. McGowan & Hilary G. Buttrick, *Moral Responsibility and Legal Liability, or, Ethics Drives the Law*, J. LEARNING HIGHER EDUC., Fall 2015, at 9, 10.

90. *See* Martin L. Cook, *Reflections on the Relationship Between Law and Ethics*, 40 ADEL. L. REV. 485, 497–98 (2019); Kenworthey Bilz & Janice Nadler, *Law, Moral Attitudes, and Behavioral Change*, *in* THE OXFORD HANDBOOK OF BEHAVIORAL ECONOMICS AND THE LAW 241, 241 (Eyal Zamir & Doron Teichman eds., 2014).

91. *See* TAEKEMA & VAN DER BURG, *supra* note 86, at 227–28 ("Ethics is primarily concerned with the moral obligations and reasoning of an individual . . . . [W]hen this is generalised to the level of groups or a society the question gains political significance . . . adding a legal dimension.").

92. *See, e.g.*, Margarita Robles Carrillo, *Artificial Intelligence: From Ethics to Law*, TELECOMMS. POL'Y, July 2020, at 1, 6; *see also* TAEKEMA & VAN DER BURG, *supra* note 86, at 235–36 (discussing instances of competing ethical norms).

93. *See, e.g.*, Riikka Koulu, *Human Control over Automation: EU Policy and AI Ethics*, EUR. J. LEGAL STUD., Spring 2020, at 9, 17–28 (for human control)*; see also* Carrillo, *supra* note 92, at 6, 13.

enforcement concerns.[94] These considerations—external to any particular ethical principle—can impose practical limits on the dominance of ethics over legal rules. Thus, ethical principles do not, in general, dominate.[95] For instance, the principle of accountability does not control where legal liability cannot be established due to the absence of a breached duty of care.[96] Instead, other considerations, such as the economic interests of the parties involved,[97] are incorporated and weighed alongside ethical considerations in the development of binding legal rules.[98]

b. Supplementing Legal Frameworks

Ethics plays a critical role beyond the mere development of legal rules. Ethical principles frequently possess a breadth and depth that exceed their comparatively narrow legal articulation,[99] making it unsurprising that ethical frameworks are often advanced to complement or supplement existing law. These frameworks are frequently constructed with the legal regime as a baseline or point of departure.[100]

The justification for this approach lies in the recognition that compliance with the law is "*necessary . . . but . . . insufficient*."[101] The law may fall short of fully achieving ethical imperatives, such as promoting caution,

---

94. *See, e.g.*, TAEKEMA & VAN DER BURG, *supra* note 86, at 237.

95. *See* Carrillo, *supra* note 92, at 6 (describing it as worrying that a lack of knowledge leads some to defend ethical principles and to exclude legal rules).

96. *Cf.* STEVEN SHAVELL, FOUNDATIONS OF ECONOMIC ANALYSIS OF LAW 259–61 (2004) (discussing situations where liability cannot be established leaving victims to bear the risk of their own injuries).

97. *See, e.g.*, *id.* at 257–87.

98. *See* RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 19 Reporter's Notes to Cmt. a (AM. L. INST. 1998).

99. Existing non-discrimination law, for example, largely targets discrimination only for specific characteristics and/or in specific contexts. *See infra* Section III.A.1.b.

100. *See, e.g.*, PAULA BODDINGTON, TOWARDS A CODE OF ETHICS FOR ARTIFICIAL INTELLIGENCE 25 (2017). The need for sufficient attention for the legal framework is also exemplified by the fact that ethical guidelines or policies may violate competition law. *See* Calo, *supra* note 18, at 189 ("History is replete with examples of new industries forming ethical codes of conduct, only to have those codes invalidated by the federal government . . . as a restraint on trade.").

101. Luciano Floridi et al., *AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, 28 MINDS & MACHS. 689, 694 (2018); *see also* INDEPENDENT HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE: ETHICS GUIDELINES FOR TRUSTWORTHY AI 6 (2019) [hereinafter HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE].

accountability,[102] or the creation of fair and trustworthy[103] algorithmic systems.[104] In this sense, ethics represents the distinction between simply adhering to the rules—the law—and striving to play well.[105] Ethical principles and frameworks thus offer essential guidance to actors engaged with AI systems, from developers (notably in the field of machine ethics)[106] to those responsible for their deployment.[107] They inform critical decisions about system selection and the appropriateness of reliance on specific systems in given contexts. More broadly, such guidelines serve to shape and regulate the interactions between humans and AI systems.[108]

Moreover, the role of ethics is often understood to extend further, offering guidance where law is absent or insufficient. For instance, ethical frameworks are considered especially necessary in contexts where real-time human intervention is challenging,[109] such as high-frequency trading[110] or autonomous vehicles.[111] Ethical guidelines also address the law's inability to keep pace with rapid technological change.[112] This function connects to the previously discussed role of ethics as a precursor to law: when law lags,[113]

---

102. *See* Sandra L. J. Johnson, *AI, Machine Learning, and Ethics in Health Care*, 39 J. LEGAL MED. 427, 437 (2020) (in medical contexts).

103. *See* Natalia Díaz-Rodríguez et al., *Connecting the Dots in Trustworthy Artificial Intelligence: From AI Principles, Ethics, and Key Requirements to Responsible AI Systems and Regulation*, INFO. FUSION, Nov. 2023, at 1, 2.

104. *See* Koulu, *supra* note 93, at 14; *see also* David Restrepo Amariles & Pablo Marcello Baquero, *Promises and Limits of Law for a Human-Centric Artificial Intelligence*, COMPUT. L. & SEC. REV., Apr. 2023, at 1, 2.

105. *See* BODDINGTON, *supra* note 100, at 25; Calo, *supra* note 18, at 189.

106. *See* Gordana Dodig Crnkovic & Baran Çürüklü, *Robots: Ethical by Design*, 14 ETHICS & INFO. TECH. 61, 62 (2012); *see also* Joanna J. Bryson & Andreas Theodorou, *How Society Can Maintain Human-Centric Artificial Intelligence*, *in* HUMAN-CENTERED DIGITALIZATION AND SERVICES 305, 309 (Marja Toivonen & Eveliina Saari eds., 2019) (discussing how design can ensure that systems operate within set parameters).

107. Michael Anderson & Susan Leigh Anderson, *Machine Ethics: Creating an Ethical Intelligent Agent*, AI MAG., Winter 2007, at 15, 15 ("The ultimate goal of machine ethics, we believe, is to create a machine that itself follows an ideal ethical principle."); Michael Davis et al., *Ethics, Finance, and Automation: A Preliminary Survey of Problems in High Frequency Trading*, 19 SCI. & ENG'G ETHICS 851, 854 (2013) ("Trading systems implement rule-based decisions, essentially executing the ethics derived from their human designers.").

108. *See* Mittelstadt et al., *supra* note 27, at 11.

109. *See id.* at 1–2.

110. *Id.* at 6; *see also* Davis et al., *supra* note 107, at 853–54.

111. Herbosch, *supra* note 12, at 438.

112. Laurie Clarke, *AI Is Mostly Governed by 'Soft Law'. But That Is Set to Change*, TECH MONITOR (Sept. 17, 2021), https://www.techmonitor.ai/policy/ai-mostly-governed-by-soft-law-but-set-to-change/ [https://perma.cc/ZDR7-V3JY]; BODDINGTON, *supra* note 100, at 25.

113. *See supra* note 56 and accompanying text.

ethics may inform or even shape subsequent legal developments. This dynamic may help explain why the AI research community's initial response has focused more on AI ethics than on AI law.[114]

Ethical frameworks, however, are not without criticism. Ethics is frequently regarded as abstract,[115] non-binding,[116] or even ineffective.[117] As a form of "soft law," [118] the primary consequence of noncompliance with ethical guidelines may be limited to reputational harm.[119] More broadly, absent widespread awareness or acceptance, violations of ethical requirements typically carry no significant consequences, [120] given the lack of enforcement mechanisms.[121] Some commentators argue that this flexibility is, in fact, attractive to certain actors,[122] as it allows them to establish and adjust[123] their own rules without external constraints, such as regulatory intervention.[124] This has led some to suggest that ethics frameworks may sometimes serve to preempt or avoid binding external regulation.[125]

Nevertheless, many scholars defend ethical frameworks by emphasizing that the role of ethics is fundamentally distinct from that of law.[126] The fact that ethics does not perform the function of legal rules is not, therefore, a substantive critique; it simply falls outside ethics' intended scope.[127] This debate underscores the importance of understanding the nuanced function of

---

114. *See* Vasiliki Koniakou, *From the "Rush to Ethics" to the "Race for Governance" in Artificial Intelligence*, 25 INFO. SYS. FRONTIERS 71, 75–76 (2023).

115. Koulu, *supra* note 93, at 14.

116. *See id.*; Carrillo, *supra* note 92, at 6*; see also* Ben Wagner, *Ethics as an Escape from Regulation. From "Ethics-Washing" to Ethics-Shopping?*, *in* BEING PROFILED: COGITAS ERGO SUM 84, 84 (Emre Bayamlioğlu et al. eds., 2018); Christoph Bartneck et al., *Personality and Demographic Correlates of Support for Regulating Artificial Intelligence*, 4 AI & ETHICS 419, 419 (2024).

117. Thilo Hagendorff, *The Ethics of AI Ethics: An Evaluation of Guidelines*, 30 MINDS & MACHS. 99, 99 (2020).

118. Koulu, *supra* note 93, at 14.

119. Hagendorff, *supra* note 117, at 99–100.

120. *See* Calo, *supra* note 18, at 188; Hagendorff, *supra* note 117, at 100.

121. *See* Calo, *supra* note 18, at 188.

122. *See id.*

123. *See id.* (calling ethics "notoriously malleable").

124. *See id.*; Hagendorff, *supra* note 117, at 100.

125. Hagendorff, *supra* note 117, at 100; Michael Veale et al., *AI and Global Governance: Modalities, Rationales, Tensions*, 19 ANN. REV. L. & SOC. SCI. 255, 259 (2023); *see also* Wagner, *supra* note 116, at 84; Linda Hogan & Marta Lasek-Markey, *Towards a Human Rights-Based Approach to Ethical AI Governance in Europe*, PHILS., Nov. 2024, at 1, 12.

126. Ruttkamp-Bloem, *supra* note 87, at 260.

127. *See id.* at 261.

ethics—as both a regulatory tool and, potentially, as a means of circumventing regulation.

2. AI Ethics

When discussing ethics in AI, the term is often interpreted broadly, frequently encompassing distinct ethical dilemmas but also economic,[128] legal, social, and psychological concerns—such as job loss,[129] displacement,[130] sectoral instability,[131] and social[132] or psychological[133] impacts. In parallel, many issues that are framed as "ethical" challenges[134] are deeply rooted in legal obligations.[135] For example, mitigating bias and avoiding discrimination are not solely ethical imperatives; they have long been legal requirements.[136] This tendency to conflate ethical and legal regimes[137]—especially as traditional legal areas like non-discrimination[138] are increasingly discussed in ethical terms—can generate confusion about the nature and enforceability of the underlying obligations.

---

128. Esmat Zaidan & Imad Antoine Ibrahim, *AI Governance in a Complex and Rapidly Changing Regulatory Landscape: A Global Perspective*, HUMANS. & SOC. SCIS. COMMC'NS, Sept. 2024, at 1, 2.

129. Axel Walz & Kay Firth-Butterfield, *Implementing Ethics into Artificial Intelligence: A Contribution, from a Legal Perspective, to the Development of an AI Governance Regime*, 18 DUKE L. & TECH. REV. 176, 185–87 (2019); Zaidan & Ibrahim, *supra* note 128, at 2; *see also* Judge et al., *supra* note 25, at 86.

130. Judge et al., *supra* note 25, at 87.

131. Zaidan & Ibrahim, *supra* note 128, at 2.

132. *Id.*

133. *See id.* (recognizing economics and social problems in addition to ethical challenges).

134. *See supra* Section I.B.1; *see, e.g.*, Díaz-Rodríguez et al., *supra* note 103, at 2; Carrillo, *supra* note 92, at 6; Deepak Khazanchi & Mahima Saxena, *Navigating Digital Human Rights in the Age of AI: Challenges, Theoretical Perspectives, and Research Implications*, J. INFO. TECH. CASE & APPLICATION RSCH. 3 (Jan. 20, 2025), https://www.tandfonline.com/doi/epdf/10.1080/15228053.2025.2452028?needAccess=true [https://perma.cc/L2CG-YBLN].

135. *See, e.g.*, Rowena Rodrigues, *Legal and Human Rights Issues of AI: Gaps, Challenges and Vulnerabilities*, J. RESPONSIBLE TECH., Dec. 2020, at 1, 2–3.

136. *See id.* at 3. *See generally* FREDERIK ZUIDERVEEN BORGESIUS, DISCRIMINATION, ARTIFICIAL INTELLIGENCE, AND ALGORITHMIC DECISION-MAKING (2018), https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73 [https://perma.cc/WP36-BZLE] (examining risks of discrimination arising from algorithmic decision-making and artificial intelligence).

137. *See* Wagner, *supra* note 116, at 85; Carrillo, *supra* note 92, at 6; MARC M. ANDERSON, SOME ETHICAL REFLECTIONS ON THE EU AI ACT 1 (2022); Joshua P. Davis, *AI, Ethics, and Law: A Way Forward*, *in* THE CAMBRIDGE HANDBOOK OF ARTIFICIAL INTELLIGENCE 304, 304 (Larry A. DiMatteo et al. eds., 2022).

138. Carrillo, *supra* note 92, at 3; BORGESIUS, *supra* note 136.

This Section reviews the principal ethical requirements in AI, drawing on the European Commission's High-Level Expert Group on Artificial Intelligence,[139] whose recommendations were particularly influential in the context of the AI Act.[140] These principles are distilled from a broader foundation of human rights law,[141] though they may reflect a Western perspective.[142] A central ethical requirement is the need for human oversight of AI systems,[143] which is complicated by the opacity inherent in many forms of AI.[144] The effectiveness of oversight is inextricably linked to explainability;[145] where outputs cannot be readily explained, meaningful oversight becomes difficult, if not impossible.

Beyond oversight, AI systems must perform adequately[146] and accurately,[147] not just under normal conditions but also in unexpected scenarios such as cyberattacks.[148] Privacy must be respected[149] throughout the system's lifecycle.[150] Transparency,[151] closely related, entails both informing individuals when they are interacting with AI rather than a human[152] and ensuring that complex systems remain explainable.[153]

---

139. HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *supra* note 101.

140. *See* Hogan & Lasek-Markey, *supra* note 125, at 1, 12.

141. *See* HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *supra* note 101, at 7 (called "fundamental rights" in the European Union).

142. *See* Carrillo, *supra* note 92, at 1–2.

143. HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *supra* note 101, at 15; Koulu, *supra* note 93, at 14–15; UNESCO, RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE 22 (2021), https://unesdoc.unesco.org/ark:/48223/pf0000381137/PDF/381137eng.pdf.multi [https://perma.cc/98L2-EPVK].

144. *See* Andreas Matthias, *The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata*, 6 ETHICS & INFO. TECH. 175, 182–83 (2004) (discussing complex computer systems); Mittelstadt et al., *supra* note 27, at 6.

145. Burrell, *supra* note 26, at 1; Mittelstadt et al., *supra* note 27, at 6; *cf.* Branka Hadji Misheva et al., *Audience-Dependent Explanations for AI-Based Risk Management Tools: A Survey*, FRONTIERS A.I., Dec. 2021, at 1, 6.

146. HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *supra* note 101, at 16.

147. *Id.*

148. *Id.* at 16–17. *See also* UNESCO, *supra* note 143, at 20 (discussing safety and security).

149. *See* UNESCO, *supra* note 143, at 20; Walz & Firth-Butterfield, *supra* note 129, at 191; Hagendorff, *supra* note 117, at 103.

150. HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *supra* note 101, at 17.

151. *See* Floridi et al., *supra* note 101, at 699–700; Petar Radanliev, *AI Ethics: Integrating Transparency, Fairness, and Privacy in AI Development*, 39 APPLIED A.I. 1, 3 (2025); *see also* Walz & Firth-Butterfield, *supra* note 129, at 188–89.

152. HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *supra* note 101, at 18.

Explainability, which is challenging for many AI systems,[154] is not only essential for oversight[155] but also for protecting human agency—ensuring that individuals can respond meaningfully to algorithmic decisions.[156] The loss of autonomy is a related risk, as algorithmic outputs may shape or constrain human choices,[157] sometimes by subtle means such as "nudging."[158]

Fairness is another foundational principle,[159] requiring active avoidance of bias and inequality.[160] AI systems,[161] by virtue of their design[162] and their reliance on potentially biased data, can reinforce[163] or worsen pre-existing

---

153. *See id.*; Hagendorff, *supra* note 117, at 103; UNESCO, *supra* note 143, at 22; Díaz-Rodríguez et al., *supra* note 103, at 4; *see also* Mittelstadt et al., *supra* note 27, at 5–6 (discussing complex algorithms more generally).

154. Burrell, *supra* note 26, at 1; Mittelstadt et al., *supra* note 27, at 6.

155. *See supra* note 145.

156. Mittelstadt et al., *supra* note 27, at 7, 9.

157. *Id.* at 9; *see also* Jeffrey Alan Johnson, *The Ethics of Big Data in Higher Education*, INT'L REV. INFO. ETHICS, July 2014, at 3, 6 (discussing "institutionally preferred action" rather than personally preferred action); Walz & Firth-Butterfield, *supra* note 129, at 192.

158. Mittelstadt et al., *supra* note 27, at 9; *see also* Johnson, *supra* note 157, at 6.

159. *See* Mittelstadt et al., *supra* note 27, at 5 (for algorithms more broadly); Hagendorff, *supra* note 117, at 103; Díaz-Rodríguez et al., *supra* note 103, at 4; Zaidan & Ibrahim, *supra* note 128, at 2.

160. Díaz-Rodríguez et al., *supra* note 103, at 4.

161. For computer systems more generally, see Engin Bozdag, *Bias in Algorithmic Filtering and Personalization*, 15 ETHICS & INFO. TECH. 209, 210 (2013); Mittelstadt et al., *supra* note 27, at 7; *see also* Katja de Vries, *Identity, Profiling Algorithms and a World of Ambient Intelligence*, 12 ETHICS & INFO. TECH. 71, 83 (2010) ("Technologies *always* differentiate between and discriminate against people (e.g., ordinary supermarket shelves discriminate against short people).").

162. Computer systems, by virtue of their programming and corresponding need for quantification, necessarily imply some bias. *See* Mittelstadt et al., *supra* note 27, at 7; *see also* Andrea Romei & Salvatore Ruggieri, *A Multidisciplinary Survey on Discrimination Analysis*, 29 KNOWLEDGE ENG'G REV. 582, 622–23 (2013) (describing how algorithms entail some bias); Nicholas Diakopoulos, *Algorithmic Accountability*, 3 DIGIT. JOURNALISM 398, 401 (2015); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 674 (2016); Emily M. Bender et al., *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*, *in* FACCT '21: PROCEEDINGS OF THE 2021 ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 610, 610 (2021) (describing the impact of data reliance, particularly textual data); Philip Hacker & Bilyana Petkova, *Reining in the Big Promise of Big Data: Transparency, Inequality, and New Regulatory Frontiers*, 15 NW. J. TECH. & INTELL. PROP. 1, 8 (2017).

163. *See* Kim, *supra* note 6, at 1575; Sunstein, *supra* note 6, at 1196; Antje von Ungern-Sternberg, *Discriminatory AI and the Law*, *in* THE CAMBRIDGE HANDBOOK OF RESPONSIBLE ARTIFICIAL INTELLIGENCE: INTERDISCIPLINARY PERSPECTIVES 252, 252–53 (Silja Voeneky et al. eds., 2022); Michael H. LeRoy, *Algorithmic Bias in Hiring: Amending Title VII to Prohibit AI Discrimination*, 51 J. LEGIS. 261, 272 (2025); Francesca Palmiotto, *The AI Act Roller Coaster:*

disparities.[164] The risk is compounded by the apparent objectivity of algorithmic decisions,[165] which can mask underlying biases and perpetuate opaque forms of discrimination. While not central to this Article's core argument, some commentators suggest that AI and algorithms, if properly leveraged, might[166] help reveal and mitigate[167] entrenched biases by making historical patterns[168] more transparent.[169]

Furthermore, social and environmental well-being[170]—rooted in ethical principles[171] such as beneficence[172] and non-maleficence[173]—and accountability[174] are essential requirements. In the context of complex machine learning systems, accountability can be particularly challenging,[175] leading to so-called "accountability gaps."[176] This challenge has led some to

---

The Evolution of Fundamental Rights Protection in the Legislative Process and the Future of the Regulation, 16 EUR. J. RISK REGUL. 1, 8 (2025); Radanliev, *supra* note 151, at 1.

164. Patrick Brione & Devyani Gajjar, *Artificial Intelligence: Ethics, Governance and Regulation*, POST (Oct. 7, 2024), https://post.parliament.uk/artificial-intelligence-ethics-governance-and-regulation/ [https://perma.cc/2BBS-L9ZL]; Khazanchi & Saxena, *supra* note 134, at 3.

165. *See* Mittelstadt et al., *supra* note 27, at 8; de Vries, *supra* note 161, at 83.

166. Radanliev, *supra* note 151, at 3 (arguing that AI should help ensure equitable decisions).

167. *See* Kleinberg et al., *supra* note 6, at 114; Jon Kleinberg et al., *Human Decisions and Machine Predictions*, 133 Q.J. ECON. 237, 277–78 (2018); Mayson, *supra* note 6, at 2277 (indicating that objective algorithmic assessments likely result in less discrimination than subjective assessments); Thomas B. Nachbar, *Algorithmic Fairness, Algorithmic Discrimination*, 48 FLA. ST. U. L. REV. 509, 544 (2021); Bozdag, *supra* note 161, at 210; Mittelstadt et al., *supra* note 27, at 7.

168. *See, e.g.*, Sunstein, *supra* note 6, at 1202.

169. *See* von Ungern-Sternberg, *supra* note 163, at 252 (describing that view).

170. HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *supra* note 101, at 19; *see also* UNESCO, *supra* note 143, at 19.

171. *See* Díaz-Rodríguez et al., *supra* note 103, at 3–4 (discussing the ethical idea of harm-prevention); Radanliev, *supra* note 151, at 3.

172. Floridi et al., *supra* note 101, at 697; UNESCO, *supra* note 143, at 20.

173. Floridi et al., *supra* note 101, at 697.

174. *See* Mittelstadt et al., *supra* note 27, at 5 (discussing "traceability"); HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *supra* note 101, at 19–20; Bryson & Theodorou, *supra* note 106, at 307 (although not explicitly in the context of ethics); Hagendorff, *supra* note 117, at 103; UNESCO, *supra* note 143, at 22–23; Radanliev, *supra* note 151, at 1; Floridi et al., *supra* note 101, at 700.

175. *See* Matthias, *supra* note 144, at 177; Colin Allen et al., *Why Machine Ethics?*, 21 IEEE INTELLIGENT SYS. 12, 14 (2006); Mittelstadt et al., *supra* note 27, at 11.

176. Mittelstadt et al., *supra* note 27, at 11; Mihailis E. Diamantis, *Employed Algorithms: A Labor Model of Corporate Liability for AI*, 72 DUKE L.J. 797, 805 (2023) (giving a rather strict interpretation to the notion "accountability gap"); Klaus Heine & Alberto Quintavalla, *Bridging the Accountability Gap of Artificial Intelligence – What Can Be Learned from Roman Law?*, 44 LEGAL STUD. 65, 66 (2024); *see also* Bert-Jaap Koops et al., *Bridging the Accountability Gap:*

propose recognizing moral agency in AI systems,[177] but in the absence of such recognition, responsibility remains with developers[178] and deployers.[179]

It is important to acknowledge that these are not the only relevant ethical considerations. Some scholars have proposed additional principles, such as reciprocity[180]—whereby those who benefit from AI should also bear its attendant risks—and the importance of ensuring that all affected societal groups have a meaningful voice in shaping AI's development and deployment.[181]

3. Values or Innovation?

Building on the discussion of AI values and ethics, this Section shifts to the second major dichotomy: whether effective AI regulation requires a choice between ethical value protection and innovation. It begins by briefly examining how the European AI Act and the proposed American moratorium are commonly framed—often, and sometimes by their own proponents,[182] with limited nuance[183]—as prioritizing value protection[184] in the case of the former, and the promotion of innovation in the case of the latter.[185]

Rights for New Entities in the Information Society?, 11 MINN. J.L., SCI. & TECH. 497, 517–18 (2010); Yanisky-Ravid, *supra* note 14, at 716 (discussing copyright law).

177. Crnkovic & Çürüklü, *supra* note 106, at 61; Mittelstadt et al., *supra* note 27, at 11.

178. *See, e.g.*, Matteo Turilli, *Ethical Protocols Design*, 9 ETHICS & INFO. TECH. 49, 49 (2007); Felicitas Kraemer et al., *Is There an Ethics of Algorithms?*, 13 ETHICS & INFO. TECH. 251, 251 (2011).

179. *See* Mittelstadt et al., *supra* note 27, at 11; *see also* Crnkovic & Çürüklü, *supra* note 106, at 61 (discussing the moral responsibility of the AI agent next to that of other actors involved).

180. Carrillo, *supra* note 92, at 14*; see also* Ruth Janal, *Extra-Contractual Liability for Wrongs Committed by Autonomous Systems*, *in* ALGORITHMS AND LAW 174, 194 (Martin Ebers & Susana Navas eds., 2020) (discussing the adage "*eius damnum, cuius commodum*").

181. Gauri Naik & Sanika S. Bhide, *Will the Future of Knowledge Work Automation Transform Personalized Medicine?*, 3 APPLIED & TRANSLATIONAL GENOMICS 50, 53 (2014); UNESCO, *supra* note 143, at 19, 23.

182. *See America's AI Action Plan*, *supra* note 81, at 3.

183. The European AI Act does contain a regime to support innovation. Regulation 2024/1689, *supra* note 3, art. 57–63.

184. *See, e.g.*, Asress Adimi Gikay, *Risks, Innovation, and Adaptability in the UK's Incrementalism Versus the European Union's Comprehensive Artificial Intelligence Regulation*, 32 INT'L J.L. & INFO. TECH, June 2024, at 1, 23.

185. *See America's AI Action Plan*, *supra* note 81, at 3.

a. Framing

At one end of the perceived "values & safety versus innovation" spectrum lies the European AI Act, often presented as protecting fundamental values in Europe but seen as obstructive to innovation in the United States.[186] This Section briefly introduces both regulatory approaches and highlights select American state-level initiatives.

The European AI Act is relevant not only for its positioning on this spectrum but also because it is designed to regulate U.S. AI providers both directly and indirectly. The EU explicitly seeks to shape global AI standards[187] through a "Brussels Effect,"[188] leveraging its market power[189] as it did with the General Data Protection Regulation.[190] Directly, the Act applies to American AI companies operating in the EU.[191]

Put succinctly, the European AI Act prohibits[192] certain AI systems (e.g., those intended to manipulate people[193]); imposes documentation, transparency, and risk-management requirements on general-purpose and foundation model providers;[194] and establishes a governance framework.[195]

Most notably for this discussion, the Act targets providers[196] and deployers[197] of "high-risk" AI systems, subjecting them to administrative

---

186. *See* Sandra Wachter, *Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond*, 26 YALE J.L. & TECH. 671, 675–76 (2024).

187. *See, e.g.*, *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Fostering a European Approach to Artificial Intelligence*, COM (2021) 205 final (Apr. 21, 2021); Marco Almada & Anca Radu, *The Brussels Side-Effect: How the AI Act Can Reduce the Global Reach of EU Policy*, 25 GERMAN L.J. 646, 646–47 (2024); Nathalie A. Smuha & Karen Yeung, *The European Union's AI Act: Beyond Motherhood and Apple Pie?*, *in* THE CAMBRIDGE HANDBOOK OF THE LAW, ETHICS AND POLICY OF ARTIFICIAL INTELLIGENCE 228, 230 (Nathalie A. Smuha ed. 2025).

188. Anu Bradford, *The Brussels Effect*, 107 NW. U. L. REV. 1, 3 (2012); *see also* Guido Noto La Diega & Leonardo C. T. Bezerra, *Can There Be Responsible AI Without AI Liability? Incentivizing Generative AI Safety Through Ex-Post Tort Liability Under the EU AI Liability Directive*, 32 INT'L J.L. & INFO. TECH., Sept. 2024, at 1, 7.

189. *See* Wachter, *supra* note 186, 675–76.

190. Regulation 2016/679, 2016 O.J. (L 119) 1 (EU).

191. Regulation 2024/1689, *supra* note 3, art. 2(1). The Act applies to AI providers placing AI systems on the market in the European Union, as well as providers and deployers of AI systems whose output is in the European Union. *Id.* arts. 2(1)(a), (c).

192. *Id.* art. 5.

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

194. *Id.* arts. 51–56.

195. *Id.* arts. 64–70.

196. *Id.* art. 16.

fines[198] and binding obligations. The classification as "high risk"[199] largely overlaps with domains already under heightened regulatory scrutiny—either directly[200] or indirectly[201]—due to the inherent risks associated with their application domains[202] in an abstract and *ex ante* way.[203] Systems are, however, excluded if they pose no significant risk to health, safety, or fundamental rights[204]—such as when used solely for preparatory tasks before human review.[205]

For "high-risk" systems, the AI Act has effectively transformed various ethical requirements into legal obligations.[206] Many of the High-Level Expert Group's principles are directly[207] or substantively[208] reproduced in the Act.

These legal requirements vary: some are substantive—governing the permissibility, accuracy,[209] explainability,[210] and human oversight[211] of AI systems—while others are procedural,[212] such as documentation[213] and

---

197. *Id.* art. 26.

198. *Id.* art. 99–101.

199. *See id.* art. 6.

200. *See id.* art. 6(1), annex I.

201. *See id.* art. 6(2), annex III.

202. This restriction to abstract categories is sometimes identified as conflicting with an "ethics-based" approach. *See* Hogan & Lasek-Markey, *supra* note 125, at 2, 7 (implicitly, by contrasting it with a more direct fundamental rights approach—which the authors perceive as more closely related to AI ethics).

203. Almada & Petit, *supra* note 56, at 94–95.

204. Regulation 2024/1689, *supra* note 3, art. 6(3).

205. *See id.* art. 6(3)(d).

206. *See* Hogan & Lasek-Markey, *supra* note 125, at 1, 12; Isabel Kusche, *Possible Harms of Artificial Intelligence and the EU AI Act: Fundamental Rights and Risk*, J. RISK RSCH., May 2024, at 1, 1–2.

207. *See, e.g.*, Regulation 2024/1689, *supra* note 3, art. 10, 13–14 (data governance, transparency, and human oversight); HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *supra* note 101, at 15–18 (data governance, transparency, and human oversight).

208. *See, e.g.*, Regulation 2024/1689, *supra* note 3, art. 9 (risk management systems); HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *supra* note 101, at 16–17, 19–20 (accountability/technical robustness and safety); Regulation 2024/1689, *supra* note 3, art. 15 (accuracy, robustness and cybersecurity); *id.* art. 12 (record keeping); *id.* art. 11 (technical documentation); HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *supra* note 101, at 18 (transparency/traceability).

209. Regulation 2024/1689, *supra* note 3, art. 15.

210. *Id.* art. 13.

211. *Id.* art. 14.

212. *See* Margot E. Kaminski, *The Developing Law of AI: A Turn to Risk Regulation*, *in* THE DIGITAL SOCIAL CONTRACT: A LAWFARE PAPER SERIES 1, 11 (2023).

213. Regulation 2024/1689, *supra* note 3, art. 11.

recordkeeping.[214] Deployers must operationalize these, for example, by maintaining logs and ensuring oversight.[215]

On the other end of the spectrum, the proposed American regulatory moratorium would have imposed a ten-year ban on state-level AI regulation.[216] The European Act also preempts stricter Member State rules,[217] but, critically, introduces an additional comprehensive value-driven framework—unlike the proposed U.S. moratorium, which would bar any AI-specific regulation.[218] While the moratorium failed, similar ideas persist.[219]

Proponents argue a moratorium would support innovation chiefly by preventing a patchwork of inconsistent state laws and, second, by ensuring the absence of regulatory barriers at both state and federal levels.[220]

Importantly, "regulation" in the moratorium referred solely to state law, leaving common law remedies intact[221]—a distinction with important implications for innovation and legal accountability.

The distinction between the "innovation-first" moratorium and the value-driven AI Act should be nuanced. The European AI Act includes safeguards to foster innovation, such as regulatory sandboxes,[222] enabling controlled experimentation. The U.S. moratorium, likewise, would leave common law in effect.[223]

---

214. *Id.* art. 12.

215. *Id.* art. 26.

216. One Big Beautiful Bill Act, H.R. 1, 119th Cong. § 43201 (2025).

217. *See* Regulation 2024/1689, *supra* note 3, at 1.

218. *See supra* note 4 and accompanying text.

219. *See supra* note 4; *America's AI Action Plan*, *supra* note 81, at 3 ("The Federal government should not allow AI-related Federal funding to be directed toward states with burdensome AI regulations that waste these funds, but should also not interfere with states' rights to pass prudent laws that are not unduly restrictive to innovation."). The White House also proposes that the FCC "evaluate whether state AI regulations interfere with the agency's ability to carry out its obligations and authorities under the Communications Act of 1934." *Id.*; *see also* Exec. Order No. 14,365, *Ensuring a National Policy Framework for Artificial Intelligence*, 90 Fed. Reg. 58499 (Dec. 11, 2025).

220. *See, e.g.*, Marc Rotenberg et al., *Proposed Moratorium on US State AI Laws Is Short-Sighted and Ill-Conceived*, TECH POL'Y PRESS (May 21, 2025), https://www.techpolicy.press/proposed-moratorium-on-us-state-ai-laws-is-shortsighted-and-illconceived/ [https://perma.cc/R4Q3-7275].

221. *See* Kevin Frazier & Adam Thierer, *Understanding the Proposed AI Moratorium: Answers to Key Questions*, R ST., https://www.rstreet.org/ai-moratorium-questions/ [https://perma.cc/M5EX-6B7Q] (June 3, 2025).

222. *See* Regulation 2024/1689, *supra* note 3, art. 57–63.

223. *See* Frazier and Thierer, *supra* note 221.

In addition, some U.S. states have adopted or are considering AI regulation, motivating the moratorium proposal. A useful example is New York's RAISE Act[224] which imposes obligations on frontier AI developers, such as retaining test records and implementing safeguards against "critical harm"—defined as events causing at least 100 deaths or over a billion dollars in damages.[225]

b. Preliminary Assessment

Before assessing this framing in more detail, it is necessary to clarify several key methodological assumptions regarding the (potential) impact of AI regulation on AI safety, ethics, and innovation. First, as to safety and ethics, this Article uses the value protections enshrined in existing legal frameworks as the benchmark, assuming those values should be equally safeguarded in AI contexts unless there are compelling AI-specific reasons to depart from this baseline. Absent such reasons, a regime that does not protect these values would not be sustainable. A value-free approach—leaving AI developers and deployers unchecked—would ultimately undermine public trust and provoke backlash, even if some societal benefits are externalized. Such benefits alone are unlikely to provide sufficient incentive for developers and deployers to optimize their systems responsibly.[226] Moreover, even if most behave responsibly, a single reckless actor can undermine public trust entirely.[227]

That said, regulation can hinder innovation in several ways. Compliance imposes costs,[228] diverting resources from research and development. Regulation can also limit what AI providers are permitted to do—either by outright bans on certain uses or by imposing requirements and restrictions on AI development and deployment, as illustrated below by the European AI Act. Substantive requirements are likely to most strongly affect innovation, as they can limit permissible AI techniques or require costly[229]

---

224. S. 6953B, 2025–2026 Leg., Reg. Sess. (N.Y. 2025).

225. *Id.*

226. *See, e.g.*, Herbosch, *supra* note 12, at 418–19.

227. *See generally* Pedro Robles & Daniel J. Mallinson, *Artificial Intelligence Technology, Public Trust, and Effective Governance*, 42 REV. POL'Y RSCH. 11, 16 (2023) (discussing the central role public trust plays in successful AI governance).

228. *See, e.g.*, EUROPEAN COMMISSION, STUDY TO SUPPORT AN IMPACT ASSESSMENT OF REGULATORY REQUIREMENTS FOR ARTIFICIAL INTELLIGENCE IN EUROPE 113–66 (2021), https://ec.europa.eu/newsroom/dae/redirection/document/75755 [https://perma.cc/BU3U-EMYZ].

229. *See id.* at 129–31, 134–35.

measures like human oversight. Procedural requirements, though burdensome, are generally more manageable[230] and often derivative[231] of substantive obligations. Notably, under the European AI Act compliance is presumed[232] where industry standards are followed,[233] echoing self-regulatory ethical codes.[234] Additionally, accountability mechanisms may expose developers to financial liability, reducing profitability. These limitations are typically driven by the values at the core of AI ethics.[235]

However, regulation can also promote innovation.[236] It can guide innovation toward societally desirable outcomes,[237] foster legal certainty and predictability,[238] and build public trust[239]—all of which can spur further investment and market adoption, enhancing both the sustainability and societal value of innovation.[240]

The following Sections provide a more detailed assessment of whether the European AI Act and the proposed American moratorium effectively protect the values they target, and whether they foster or hinder innovation—beginning with an overview of the legal challenges AI poses for liability and non-discrimination law. The legal discussion will include

---

230. *See generally id.* at 119–21 (providing an assessment of that cost).

231. *See* Kaminski*, supra* note 212, at 11. *Compare* Regulation 2024/1689, *supra* note 3, art. 12 (imposing a record-keeping requirement), *with* HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *supra* note 101, at 19–20 (demonstrating how the record-keeping requirement arises under "accountability" standards).

232. *See* Regulation 2024/1689, *supra* note 3, art. 40(1); *see* Smuha & Yeung, *supra* note 187, at 253; *cf.* S.B. 24-205, 74th Gen. Assemb., 2024 Reg. Sess. § 6-1-1702(1) (Colo. 2024) (containing a similar presumption of compliance).

233. *See* Regulation 2024/1689, *supra* note 3, art. 40 (imposing standardization requirements to imply conformity).

234. *See infra* Section II.B. Many scholars are critical of these standards. *See* Michael Veale & Frederik Zuiderveen Borgesius, *Demystifying the Draft EU Artificial Intelligence Act*, 22 COMPUT. L. REV. INT'L 97, 105 (2021) (arguing that the resulting standards are largely industry-driven because consumer organizations lack the means to meaningfully participate in the process); Wachter, *supra* note 186, at 691–92 (arguing that European standards bodies' task is too broad in the sense that industry representatives are tasked to resolve normative issues rather than merely come up with technical standards); Smuha & Yeung, *supra* note 187, at 28–32.

235. *See infra* Section III.A.

236. *See* Anu Bradford, *The False Choice Between Digital Regulation and Innovation*, 119 NW. U. L. REV. 377, 402–03 (2024).

237. *See* ALESSIO TARTARO ET AL., ASSESSING THE IMPACT OF REGULATIONS AND STANDARDS ON INNOVATION IN THE FIELD OF AI, ARXIV 1, 5 (2023), https://arxiv.org/pdf/2302.04110 [https://perma.cc/V9W5-6MX9]; Bradford, *supra* note 236, at 419.

238. *See* TARTARO ET AL., *supra* note 237, at 5; Bradford, *supra* note 236, at 417–18.

239. *See* TARTARO ET AL., *supra* note 237, at 5; Bradford, *supra* note 236, at 418–19.

240. *See* TARTARO ET AL., *supra* note 237, at 4; Bradford, *supra* note 236, at 419.

European perspectives to accurately assess the impact of the AI Act on that framework.

II. LEGAL CHALLENGES

A. Liability Law

AI presents key challenges for liability law.[241] The challenge is especially acute as many AI systems—such as autonomous vehicles[242] and medical decision-support tools[243]—are capable of causing harm given their inherent imperfections and growing prevalence.[244] Accordingly, the question of AI liability has generated significant debate,[245] and is often considered crucial for AI safety.[246] While some scholars emphasize the need for a dedicated legal regime,[247] this Article proceeds with a more traditional approach, briefly identifying several key challenges that AI systems pose to existing legal frameworks.[248] This, in turn, enables an analysis of the impact of ethics-driven AI regulation in this domain.

First, within the context of negligence liability, a general difficulty arises in determining whether the development or deployment of a particular AI system was, in fact, negligent.[249] This difficulty stems in part from inherent

---

241. *See* sources cited *supra* note 5.

242. While such vehicles would result in fewer accidents than human drivers, the inherent erroneous nature of AI implies that there would still be accidents. *See* Selbst, *supra* note 5, at 1323–26.

243. *See, e.g.*, Mongan et al., *supra* note 41, at 1.

244. *See* Alicia Lai, *Artificial Intelligence, LLC: Corporate Personhood as Tort Reform*, 2021 MICH. ST. L. REV. 597, 612–13.

245. *See* sources cited *supra* note 5.

246. *See* Noto La Diega & Bezerra, *supra* note 188, at 3.

247. *See, e.g.*, Karni A. Chagal-Feferkorn, *Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision-Makers*, 30 STAN. L. & POL'Y REV. 61, 90–114 (2019) (arguing for a tailored regime); Lai, *supra* note 244, at 631–53 (proposing legal personhood); Sahara Shrestha, *Nature, Nurture, or Neither?: Liability for Automated and Autonomous Artificial Intelligence Torts Based on Human Design and Influences*, 29 GEO. MASON L. REV. 375, 401–10 (2021) (introducing a new balancing test); Renee Henson, *"I Am Become Death, the Destroyer of Worlds": Applying Strict Liability to Artificial Intelligence as an Abnormally Dangerous Activity*, 96 TEMP. L. REV. 349, 362–90 (2024) (arguing for an "abnormally dangerous activities" test for AI system products liability).

248. *See, e.g.*, Selbst, *supra* note 5, at 1318–19 (discussing negligence).

249. *See, e.g.*, Selbst, *supra* note 5, at 1331–32; William D. Smart et al., *An Education Theory of Fault for Autonomous Systems*, 2 NOTRE DAME J. ON EMERGING TECHS. 35, 43–44 (2021); *see generally* ANNA BECKERS & GUNTHER TEUBNER, THREE LIABILITY REGIMES FOR

characteristics of AI systems—such as their complexity, and their limited predictability[250] and explainability[251]—which tend to increase the separation between human decision-making and AI-generated outcomes.[252] More broadly, it reflects how difficult it is for legal practitioners to assess the contextual risks of AI development and deployment across different contexts in the absence of abstract standards.[253] The Learned Hand formula illustrates the contextual nature of negligence assessments particularly well.[254] It shows that conduct considered diligent in one context may be negligent in another, where the risk of harm is greater.[255]

Rather than imposing an abstract standard, the Hand formula suggests that developers and deployers should adjust their behavior based on context, as they are only required to take economically justified precautionary measures.[256] These measures need not directly affect system accuracy but may include other risk-reduction strategies, such as providing appropriate warnings to deployers.[257] In the same vein, the level of accuracy and explainability an AI system should attain depends both on the feasibility of achieving higher performance and on the risks associated with failing to do so. In addition, the various components of explainability exemplify the

---

ARTIFICIAL INTELLIGENCE 72–73 (2021) (discussing a fault-based liability regime for artificial intelligence).

250. *See* sources cited *supra* note 23.

251. *See* sources cited *supra* note 30.

252. *See* Selbst, *supra* note 5, at 1375; BECKERS & TEUBNER, *supra* note 249, at 72–73; Kowert, *supra* note 17, at 184.

253. *See* Beatriz Botero Arcila, *AI Liability in Europe: How Does It Complement Risk Regulation and Deal with the Problem of Human Oversight?*, COMPUT. L. & SEC. REV., Sept. 2024, at 1, 1–2.

254. *See* United States v. Carroll Towing Co., 159 F.2d 169, 173 (2d Cir. 1947); *see also* Barbara Ann White, *Risk-Utility Analysis and the Learned Hand Formula: A Hand That Helps or a Hand That Hides?*, 32 ARIZ. L. REV. 77, 102–06 (1990); Keith N. Hylton, *Information and Causation in Tort Law: Generalizing the Learned Hand Test for Causation Cases*, 7 J. TORT L. 35, 52–54 (2014); Daniel P. O'Gorman, *Contract Law and the Hand Formula*, 75 LA. L. REV. 127, 156–58 (2014); RESTATEMENT (THIRD) OF TORTS: PHYSICAL & EMOTIONAL HARM § 3 (AM. L. INST. 2010).

255. *See* sources cited *supra* note 254.

256. *Cf.* Vasant Dhar, *When to Trust Robots with Decisions, and When Not to*, HARV. BUS. REV. (May 17, 2016), https://hbr.org/2016/05/when-to-trust-robots-with-decisions-and-when-not-to [https://perma.cc/JK2J-UASD].

257. *Cf.* Michael D. Scott, *Tort Liability for Vendors of Insecure Software: Has the Time Finally Come?*, 67 MD. L. REV. 425, 443 (2008) (arguing that software vendors have a legal duty to warn licensees of potential dangers); RESTATEMENT (THIRD) OF TORTS: PHYSICAL & EMOTIONAL HARM § 18 (AM. L. INST. 2010) (describing the duty to warn others who are subjected to some risks); Smart et al., *supra* note 249, at 49–51 (arguing that software manufacturers have a duty to warn users about the limitations of software).

contextual nature of these requirements as well. A more superficial understanding of a system's output—sometimes[258] termed interpretability[259]—may, in some cases, be sufficient to assess relevant risks, obviating the need for deeper insight. Likewise, the degree of explainability is often contingent on the observer's—typically the deployer's—experience and level of expertise. For example, a physician may require less explanation to understand an AI-based image analysis of a patient than a layperson would.

Similarly, the extent to which a deployer may rely on a system, or must verify its outputs through oversight, is determined by the contextual risks of deploying that specific system.[260] In each of these instances, it is clear that greater accuracy, explainability, and oversight[261] reflect more diligent development and deployment. The more pertinent question, however, is when the threshold for negligence is crossed. While the existing negligence framework offers tools to address this, the assessments are often complex—particularly for legally trained practitioners, such as lawyers and judges, who typically lack expertise in computer engineering.[262]

A similar challenge arises in the area of products liability, where two criteria exist to establish a design defect.[263] First, there is the consumer

---

258. Both notions are often used interchangeably. *See, e.g.*, Behnoush Abdollahi & Olfa Nasraoui, *Transparency in Fair Machine Learning: The Case of Explainable Recommender Systems*, *in* HUMAN AND MACHINE LEARNING 21, 24 (Jianlong Zhou & Fang Chen eds., 2018); PATRICK HALL & NAVDEEP GILL, AN INTRODUCTION TO MACHINE LEARNING INTERPRETABILITY 2 (2d ed. 2019); Du et al., *supra* note 36, at 69; GIANFAGNA & DI CECCO, *supra* note 37, at 25.

259. *See* HALL & GILL, *supra* note 258, at 2; GIANFAGNA & DI CECCO, *supra* note 37, at 12–13. Explainability is said to refer to understanding why water boils at a certain temperature (the underlying process), whereas interpretability consists of knowing at what temperature it boils. *See id.* at 16–17.

260. For example, many states require autonomous vehicles to have a human driver. *See* Atilla Kasap, *States' Approaches to Autonomous Vehicle Technology in Light of Federal Law*, 19 OHIO ST. TECH. L.J. 315, 341–42 (2023).

261. *See, e.g.*, Ignacio N. Cofone, *Servers and Waiters: What Matters in the Law of A.I.*, 21 STAN. TECH. L. REV. 167, 191 (2018).

262. *See* Omri Rachum-Twaig, *Whose Robot Is It Anyway?: Liability for Artificial-Intelligence-Based Robots*, 2020 U. ILL. L. REV. 1141, 1160; Bartlett, *supra* note 30, at 297 (referring to hindsight bias and outcome bias).

263. The imperfection of AI output is usually considered to constitute a design defect, see, for example, Brian S. Haney, *The Optimal Agent: The Future of Autonomous Vehicles & Liability Theory*, 30 ALB. L.J. SCI. & TECH. 1, 28–31 (2020), although AI can be argued to blur the relevant distinction with manufacturing defects. On the two tests for design defects, see Selbst, *supra* note 5, at 1323; LEWIS BASS & THOMAS PARKER REDICK, PRODUCTS LIABILITY: DESIGN & MANUFACTURING DEFECTS 244–45 (2nd ed. 2022); 5 STUART M. SPEISER ET AL., THE AMERICAN LAW OF TORTS 835 (Monique C. M. Leahy ed. 2016). On manufacturing defects (such as a faulty sensor) in AI contexts, see Kevin Funkhouser, *Paving the Road Ahead:*

expectations test,[264] which—resembling the approach taken in the European Union[265]—deems a product defective if it fails to meet the safety expectations of a reasonable consumer. [266] Alternatively, some courts apply the cost-benefit test,[267] akin to the Learned Hand formula.[268] Under this test, a product is deemed defective if a reasonable alternative design was available that would have reduced the relevant danger, but was not adopted, rendering the product not reasonably safe.[269] For each of these standards, it is evident that the same concerns present in a negligence context apply.[270]

In both negligence and products liability contexts, the assessment often requires a contextual evaluation of factors such as system accuracy, explainability, and the degree of human oversight. A system with limited accuracy should, for example, generally be subject to greater human oversight.[271] Conversely, a lower level of accuracy may be more acceptable from a liability perspective if offset by a higher degree of explainability, which enables more effective oversight at a similar cost.[272] Furthermore, both the reasonable alternative design rule and negligence liability—particularly as articulated through the Learned Hand rule—clarify that it is

---

Autonomous Vehicles, Products Liability, and the Need for a New Approach, 2013 UTAH L. REV. 437, 453 (2013).

264. SPEISER ET AL., *supra* note 263, at 835.

265. *See* Council Directive 85/374, art. 7, 2024 O.J. (L 2853) 1 (EU); *see also* Vibe Ulfbeck, *Product Liability Law and AI: Revival or Death of Product Liability Law*, *in* THE CAMBRIDGE HANDBOOK OF PRIVATE LAW AND ARTIFICIAL INTELLIGENCE 206, 207–08 (Ernest Lim & Phillip Morgan eds., 2024) (discussing the similar test under the old European Directive).

266. Vladeck, *supra* note 5, at 134–35; Rachum-Twaig, *supra* note 262, at 1155–56; Selbst, *supra* note 5, at 1323–24; BASS & REDICK, *supra* note 263, at 244; *see* SPEISER ET AL., *supra* note 263, at 835.

267. Although some courts consider both. *See* BASS & REDICK, *supra* note 263, at 261–62.

268. Scott, *supra* note 257, at 467; Aaron D. Twerski & James A. Henderson, Jr., *Manufacturers' Liability for Defective Product Designs: The Triumph of Risk-Utility*, 74 BROOK. L. REV. 1061, 1065 (2009); Chagal-Feferkorn, *supra* note 247, at 81.

269. RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2 (AM. L. INST. 1998); Rachum-Twaig, *supra* note 262, at 1155–56; Frank Griffin, *Artificial Intelligence and Liability in Health Care*, 31 HEALTH MATRIX: J.L. MED. 65, 79 (2021).

270. *See* Ulfbeck, *supra* note 265, at 225 (discussing individualization of AI products, calling for more contextualization). In an attempt to remedy some of these difficulties, Article 7(2)(f) of the EU's Revised Products Liability Directive refers to relevant product standards—such as those imposed by the European AI Act, *see infra* note 389 and accompanying text, but this reference suffers many of the drawbacks discussed below. *See infra* Section IV.

271. Applying the Learned Hand rule, more limited accuracy implies a greater risk of harm and thus requires additional precautionary measures. *See supra* note 254 and accompanying text.

272. *See supra* note 155 and accompanying text.

difficult to define relevant thresholds in abstract terms. Rather, the appropriate level of precaution is context-dependent, varying with the severity of the potential harm that the system may cause.[273]

Second, with respect to causality, several challenges emerge.[274] First, many scholars argue that the unpredictability of AI system outputs disrupts traditional notions of causation[275]—sometimes situating this concern within the framework of negligence[276]—on the basis that developers or deployers cannot foresee the specific harm that may occur.[277] This concern—reminiscent of technological exceptionalism[278]—requires some refinement: fault-based liability does not generally require foreseeability of the precise harm, but only foreseeability of the category of harm,[279] which is typically met in the case of present-day "weak" AI applications,[280] although this might change in the future.[281]

Another key challenge concerns the possibility of establishing a causal link between the harmful output and a specific fault of the developer or deployer.[282] Given the inherent imperfection of AI systems,[283] even highly diligent conduct—meeting the highest available standards—may not be sufficient to eliminate all risk of harm. The difficulty of distinguishing such harm from the harm due to the developer's negligence, for example, raises a fundamental issue for legal causation. While the American requirement that plaintiffs prove causation as "more likely than not"[284] adds nuance, the

---

273. *See supra* note 254 and accompanying text.

274. *See* Lai, *supra* note 244, at 628; Zhao Yan Lee et al., *Deep Learning Artificial Intelligence and the Law of Causation: Application, Challenges and Solutions*, 30 INFO. & COMMC'NS. TECH. L. 255, 265 (2021).

275. *See, e.g.*, Lai, *supra* note 244, at 628–31; Lee et al., *supra* note 274, at 264–65.

276. Rachum-Twaig, *supra* note 262, at 1156; Selbst, *supra* note 5, at 1375; BECKERS & TEUBNER, *supra* note 249, at 73.

277. Selbst, *supra* note 5, at 1375; BECKERS & TEUBNER, *supra* note 249, at 73; Lee et al., *supra* note 274, at 262 (discussing deep learning AI).

278. *See supra* Section I.A.3.b.

279. Selbst, *supra* note 5, at 1342; *see* W. Jonathan Cardi, *Reconstructing Foreseeability*, 46 B.C. L. REV. 921, 926 (2005); Benjamin C. Zipursky, *Foreseeability in Breach, Duty, and Proximate Cause*, 44 WAKE FOREST L. REV. 1247, 1252 (2009).

280. *See supra* Section I.A.3.a.

281. Ryan Calo, *Is the Law Ready for Driverless Cars?*, 61 COMMC'NS. ACM 34, 35 (2018); Selbst, *supra* note 5, at 1343. Specifically for artificial general intelligence, see Scherer, *supra* note 13, at 365; Selbst, *supra* note 5, at 1344; Lai, *supra* note 244, at 629–30.

282. Botero Arcila, *supra* note 253, at 6.

283. *See supra* Section I.A.2.b; *see also* Selbst, *supra* note 5, at 1343–44.

284. *See, e.g.*, Gideon v. Johns-Manville Sales Corp., 761 F.2d 1129, 1138 (5th Cir. 1985) ("Possibility alone cannot serve as the basis for recovery, for mere possibility does not meet the preponderance of the evidence standard. Certainty, however, is not required: the plaintiff need

challenge is even greater in jurisdictions with stricter thresholds—sometimes requiring a level of certainty approaching ninety percent for causation.[285] Applying probabilistic causation doctrines[286] such as the *loss of chance* doctrine[287] may also help address this challenge more effectively; however, these doctrines are largely restricted to medical malpractice cases.[288]

While some authors have identified additional concerns—such as the aforementioned disruption of causality due to unpredictability[289]—the challenges surrounding the identification of liability thresholds and the establishment of causality are sufficient to support the analysis advanced in this Article.

A further challenge, relevant to both liability thresholds and questions of causality, is the significant information asymmetry inherent in AI development and deployment.[290] This asymmetry is especially problematic when a third party—i.e., someone other than the developer or deployer—is harmed by an AI system's output. In such cases, the injured party is often poorly positioned to demonstrate the relevant fault standard[291]—whether negligence or design defect. Moreover, the developer or deployer may not

---

demonstrate only that the event is more likely to occur than not."); Robert T. Ebert, Jr., *Damages for an Increased Risk of Developing Cancer Caused by Asbestos Exposure Are Only Recoverable If It Is More Likely Than Not That Cancer Will Develop*, 51 MO. L. REV. 847, 848 (1986).

285. Mark Schweizer, *The Civil Standard of Proof—What Is it, Actually?*, 20 INT'L J. EVIDENCE & PROOF 217, 220 (2016).

286. *See* Glen O. Robinson, *Probabilistic Causation and Compensation for Tortious Risk*, 14 J. LEGAL STUD. 779, 780–81 (1985); Alessandro Romano, *God's Dice: The Law in a Probabilistic World*, 41 U. DAYTON L. REV. 57, 75 (2016).

287. *See, e.g.*, Dillon v. Evanston Hosp., 771 N.E.2d 357, 370 (2002); David A. Fischer, *Tort Recovery for Loss of a Chance*, 36 WAKE FOREST L. REV. 605, 606 (2001); Timothy Dylan Reeves, *Tort Liability for Manufacturers of Violent Video Games: A Situational Discussion of the Causation Calamity*, 60 ALA. L. REV. 519, 543–44 (2009); Robert J. Rhee, *Loss of Chance, Probabilistic Cause, and Damage Calculations: The Error in* Matsuyama v. Birnbaum *and the Majority Rule of Damages in Many Jurisdictions More Generally*, 1 SUFFOLK U. L. REV. ONLINE 39, 39 (2013).

288. Romano, *supra* note 286, at 84.

289. *See supra* note 275.

290. Marta Ziosi et al., *The EU AI Liability Directive (AILD): Bridging Information Gaps*, 14 EUR. J.L. & TECH. 1, 3 (2023); Botero Arcila, *supra* note 253, at 4; *cf. infra* Section II.B (discussing a similar information asymmetry for discrimination law).

291. *See* Bartlett, *supra* note 30, at 297. While the European Union had proposed an AI Liability Directive, see *Proposal for a Directive of the European Parliament and of the Council on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence (AI Liability Directive)*, COM (2022) 496 final (Sept. 28, 2022), containing a presumption of a causal link (art. 4) and a potential presumption of a fault (art. 3(5)), it has since abandoned this initiative.

have retained the relevant data or documentation, rendering proof of causality or of fault effectively impossible.[292]

Moreover, at a general level, existing liability regimes fail to provide incentives that are adequately tailored to the unique characteristics of AI development and deployment.[293] Consequently, AI developers and deployers may be incentivized to engage in conduct that is harmful to society, or, conversely, may be deterred from pursuing socially beneficial AI advancements due to the potential for disproportionate liability.

B. Non-Discrimination Law

Like accountability, non-discrimination and bias are central concerns in ethical AI.[294] Against that backdrop, it is instructive to consider how existing non-discrimination law provisions and standards may be affected by AI technologies.

European non-discrimination law largely reflects the American framework. A broad, general constitutional-level provision[295] is operationalized through various instruments,[296] such as the Racial Equality Directive,[297] the Employment Equality Directive[298], the recast Gender Equality Directive,[299] and the Gender Access Directive.[300] These target discrimination based on specific protected characteristics[301] and in particular

---

292. For further discussion, see *infra* Section III.A.

293. *See* Herbosch, *supra* note 12, at 453–57.

294. *See supra* Section I.B.

295. Charter of Fundamental Rights of the European Union, 2012 O.J. (C 326) 391, art. 21(1) ("Any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited.").

296. *See, e.g.*, Wachter et al., *supra* note 6, at 6–7; Gerards & Borgesius, *supra* note 20, at 19–20. In the US, compare with the Equal Credit Opportunity Act, Titles VI and VII of the Civil Rights Act of 1964 (42 U.S.C. §§ 2000d–ff), and Title VIII of the Civil Rights Act of 1968 (42 U.S.C. §§ 3601–19). *See also* Kim, *supra* note 6, at 1555.

297. Council Directive 2000/43, 2000 O.J. (L 180) 22 (EC).

298. Council Directive 2000/78, 2000 O.J. (L 303) 16 (EC).

299. Council Directive 2006/54, 2006 O.J. (L 204) 23 (EC).

300. Council Directive 2004/113, 2004 O.J. (L 373) 37 (EC).

301. *E.g.*, Council Directive 2006/54, *supra* note 299, arts. 1–2 ("men and women" and "sex"); Council Directive 2000/43, *supra* note 297, art. 1 ("racial or ethnic origin"); Council Directive 2000/78, *supra* note 298, art. 1 ("religion or belief, disability, age or sexual orientation"). Article 21 EU Charter of Fundamental rights is more open-ended, using the phrase "such as." *See* Raphaële Xenidis, *Tuning EU Equality Law to Algorithmic Discrimination: Three Pathways to Resilience*, 27 MAASTRICHT J. EUR. & COMPAR. L. 736, 755 (2020); Wachter et al., *supra* note 6, at 6; Gerards & Borgesius, *supra* note 20, at 25; Alexandru

sectors,[302] while allowing a degree of discretion to member states in their implementation.[303]

Each of these European directives establishes a similar framework, distinguishing[304] between "direct" and "indirect" discrimination.[305] Direct discrimination refers to instances in which an individual is treated less favorably explicitly because of a protected characteristic,[306] aligning broadly[307] with the U.S. concept of disparate treatment.[308] Indirect discrimination, by contrast, involves facially neutral practices that disproportionately disadvantage members of a protected group,[309] broadly corresponding to the American doctrine of disparate impact.[310]

Cîrciumaru, *Discrimination in the Age of Algorithms—Is EU Law Ready?*, *in* EUROPEAN YEARBOOK OF CONSTITUTIONAL LAW 2023: CONSTITUTIONAL LAW IN THE DIGITAL ERA 111, 120 (Charlotte van Oirsouw et al. eds., 2024).

302. Council Directive 2006/54, *supra* note 299, art. 1; Council Directive 2000/43, *supra* note 297, art. 3; Council Directive 2000/78, *supra* note 298, art. 3; *see also* von Ungern-Sternberg, *supra* note 163, at 257; Cîrciumaru, *supra* note 301, at 118–19.

303. Wachter et al., *supra* note 6, at 7.

304. See Nachbar, *supra* note 167, at 540, for criticism of the similar American distinction between disparate treatment and disparate impact.

305. *See also* Frederik J. Zuiderveen Borgesius, *Strengthening Legal Protection Against Discrimination by Algorithms and Artificial Intelligence*, 24 INT'L J. HUM. RTS. 1572, 1576 (2020); von Ungern-Sternberg, *supra* note 163, at 257.

306. *See* Philipp Hacker, *Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law*, 55 COMMON MKT. L. REV. 1143, 1151 (2018); Borgesius, *supra* note 305, at 1576; von Ungern-Sternberg, *supra* note 163, at 257.

307. *See* Jeremias Adams-Prassl et al., *Directly Discriminatory Algorithms*, 86 MOD. L. REV. 144, 149 (2023).

308. Kleinberg et al., *supra* note 6, at 121; Mayson, *supra* note 6, at 2240; Matthew U. Scherer et al., *Applying Old Rules to New Tools: Employment Discrimination Law in the Age of Algorithms*, 71 S.C. L. REV. 449, 459 (2019); Prince & Schwarcz, *supra* note 6, at 1269–70; Sunstein, *supra* note 6, at 1197–98.

309. Hacker, *supra* note 306, at 1152–53; Borgesius, *supra* note 305, at 1576; von Ungern-Sternberg, *supra* note 163, at 257–58; Adams-Prassl et al., *supra* note 307, at 145.

310. *See, e.g.*, Barocas & Selbst, *supra* note 162, at 701; Bornstein, *supra* note 6, at 553–54; Mayson, *supra* note 6, at 2241–42; Lori Andrews & Hannah Bucher, *Automating Discrimination: AI Hiring Practices and Gender Inequality*, 44 CARDOZO L. REV. 145, 162 (2022); Sunstein, *supra* note 6, at 1198–99.

In certain circumstances, a difference in treatment or impact may be legally justified without violating the prohibition on discrimination. While terminology varies slightly across various EU non-discrimination law directives,[311] a shared exception[312] permits such treatment if it serves a legitimate aim and is proportionate to achieving that aim.[313] In such cases, conduct that would otherwise constitute prohibited discrimination may be rendered lawful under this proportionality test, a structure that also parallels U.S. jurisprudence.[314]

Even within this well-developed framework, AI introduces distinctive challenges. The opacity of many AI systems significantly hinders the application of direct discrimination and disparate treatment standards.[315] Consider the example of an AI system deployed to rank résumés for hiring purposes[316]—an application that would potentially[317] qualify as "high-risk"

---

311. Some directives refer to "appropriate and necessary" means. *See, e.g.*, Council Directive 2004/113, *supra* note 300, art. 4; Council Directive 2006/54, *supra* note 299, art. 2; Council Directive 2000/43, *supra* note 297, art. 2; Council Directive 2000/78, *supra* note 298, art. 2; *see also* von Ungern-Sternberg, *supra* note 163, at 264; Council Directive 2000/78, *supra* note 298, art. 4 (applying also to direct discrimination).

312. *See also* von Ungern-Sternberg, *supra* note 163, at 264; Charter of Fundamental Rights of the European Union, *supra* note 295, art. 52(1) ("Any limitation on the exercise of the rights and freedoms recognised by this Charter must be provided for by law and respect the essence of those rights and freedoms. Subject to the principle of proportionality, limitations may be made only if they are necessary and genuinely meet objectives of general interest recognised by the Union or the need to protect the rights and freedoms of others.").

313. *See, e.g.*, Council Directive 2004/113, *supra* note 300, art. 4; Council Directive 2006/54, *supra* note 299, art. 2; Council Directive 2000/43, *supra* note 297, art. 2; Council Directive 2000/78, *supra* note 298, art. 2; *see also* Hacker, *supra* note 306, at 1160–61; Borgesius, *supra* note 305, at 1577–78; von Ungern-Sternberg, *supra* note 163, at 264; Adams-Prassl et al., *supra* note 307, at 145.

314. Kleinberg et al., *supra* note 6, at 122–23; Nachbar, *supra* note 167, at 534–40; *see, e.g.*, 42 U.S.C. § 2000e-2(k)(1)(A) (referring to "business necessity" (cfr. the legitimate aim) and the absence of an alternative employment practice (cfr. proportionality)); *see also* Barocas & Selbst, *supra* note 162, at 701–02; Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109, 161 (2017).

315. *See also* BORGESIUS, *supra* note 136, at 15; George Gantzias, *Dynamics of Public Interest in Artificial Intelligence: 'Business Intelligence Culture' and Global Regulation in the Digital Era*, *in* THE PALGRAVE HANDBOOK OF CORPORATE SUSTAINABILITY IN THE DIGITAL ERA 259, 272 (Seung Ho Park et al. eds., 2021).

316. *See, e.g.*, Jeffrey Dastin, *Insight - Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women*, REUTERS (Oct. 10, 2018), https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/; *see also* Bornstein, *supra* note 6, at 521; Betsy Anne Williams et al., *How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications*, 8 J. INFO. POL'Y 78, 78 (2018); Spencer M. Mainka, *Algorithm-Based Recruiting Technology in the Workplace*, 5 TEX. A&M J. PROP. L. 801, 811 (2019); Scherer et al., *supra*

under the European AI Act. Suppose the system exhibits a bias that systematically favors male over female candidates.[318] Two scenarios are possible. In the first scenario—which one would hope to be rare—a developer may intentionally prefer male candidates and may have deliberately encoded this preference into the system. Even in such a case, however, the system's opacity and the associated information asymmetry—as discussed in the context of liability[319]—mean that a plaintiff would face significant difficulty in proving the existence of that discriminatory preference.[320] This issue closely resembles a similar difficulty in traditional discrimination cases,[321] where plaintiffs are often required to demonstrate discriminatory decision-making[322] that occurs within the mind of another person.[323] These challenges are, however, somewhat mitigated by the reduced burden of proof applicable in this area of law, as discussed earlier.[324]

The unpredictability and opacity of AI systems give rise to a second scenario. In this case, the developer did not intend[325] to create a biased system, yet bias may nonetheless emerge.[326] This may result from human

---

note 308, at 492; Ljupcho Grozdanovski, *In Search of Effectiveness and Fairness in Proving Algorithmic Discrimination in EU Law*, 58 COMMON MKT. L. REV. 99, 105 (2021); Andrews & Bucher, *supra* note 310, at 168.

317. *See supra* note 204 and accompanying text.

318. This example is not merely theoretical, see, for example, Bornstein, *supra* note 6, at 521; Andrews & Bucher, *supra* note 310, at 168–70.

319. *See supra* Section III.A.

320. *See also* Borgesius, *supra* note 305, at 1577.

321. Although some authors tend to stress that discrimination becomes particularly "hidden" in AI contexts, see, for example, Scherer et al., *supra* note 308, at 492–93 (underscoring that algorithms cannot be put on the witness stand); Jens Ludwig & Sendhil Mullainathan, *Fragile Algorithms and Fallible Decision-Makers: Lessons from the Justice System*, 35 J. ECON. PERSPS. 71, 82 (2021); Wachter et al., *supra* note 6, at 3, 5 (in a similar sense, by identifying that many traditional discrimination cases revolve around intuition); LeRoy, *supra* note 163, at 269.

322. *See, e.g.*, Kleinberg et al., *supra* note 6, at 114.

323. *See, e.g.*, Barocas & Selbst, *supra* note 162, at 713 (discussing the "evidence of state of mind"). The challenge is nuanced there, as that human can be put on the witness stand. *See* Scherer et al., *supra* note 308, at 492–93.

324. *See supra* Section II.B.

325. The desire to differentiate does not necessarily stem from irrational preferences, but may also stem from statistical preferences—for example, based on an analysis that young people are more productive. This does, however, not mean that the latter is more permissible. *See* Kleinberg et al., *supra* note 6, at 121–22; von Ungern-Sternberg, *supra* note 163, at 260. Problematically, AI may make it easier to mask such intentions. *See* Barocas & Selbst, *supra* note 162, at 692–93.

326. Barocas & Selbst, *supra* note 162, at 674; Grozdanovski, *supra* note 316, at 99–100.

error[327] or—more plausibly—from the use of biased training data.[328] Two sub-scenarios can be distinguished. In the first, the system directly incorporates a protected characteristic—such as gender or race—into its decision-making process, thereby disadvantaging individuals based on that trait. In the second, the system arrives at similar outcomes by relying on proxy variables that correlate with protected characteristics.[329] While the former is relatively straightforward to avoid—for instance, by explicitly programming the system to disregard such traits[330]—the latter is more complex. This second sub-scenario closely aligns with the concept of indirect discrimination or disparate impact,[331] discussed below, particularly when it leads to disproportionate adverse effects on groups defined by protected characteristics. Moreover, the system's statistical patterns may give rise to new forms of bias or discrimination that do not fall neatly within existing protected categories,[332] potentially leaving affected individuals without recourse under current non-discrimination law.[333]

The critical point is that disparate treatment[334] and direct discrimination largely depend on whether a protected characteristic was used as a basis for decision-making.[335] If the system explicitly employs such a characteristic, direct discrimination occurs.[336] However, due to the opacity of many AI

---

327. In this sense, see Bornstein, *supra* note 6, at 534; Hacker, *supra* note 306, at 1153. Human errors can, for example, relate to feature selection, see, for example, Selbst, *supra* note 314, at 164.

328. Barocas & Selbst, *supra* note 162, at 680–87; Paul B. de Laat, *Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?*, 31 PHIL. & TECH. 525, 530 (2018); Mainka, *supra* note 316, at 811; Mayson, *supra* note 6, at 2224 (discussing how flawed inputs create "bias in, bias out" predictions); Borgesius, *supra* note 305, at 1574; Sonia K. Katyal, *Democracy & Distrust in an Era of Artificial Intelligence*, 151 DAEDALUS 322, 326 (2022); Kim, *supra* note 6, at 1575; von Ungern-Sternberg, *supra* note 163, at 261–62.

329. Barocas & Selbst, *supra* note 162, at 712; Brent Mittelstadt, *From Individual to Group Privacy in Big Data Analytics*, 30 PHIL. & TECH. 475, 479 (2017); Hacker, *supra* note 306, at 1153; Prince & Schwarcz, *supra* note 6, at 1260–61; Xenidis, *supra* note 301, at 745; Gerards & Borgesius, *supra* note 20, at 8–10.

330. Sunstein, *supra* note 6, at 1202.

331. See, for example, on disparate impact, Prince & Schwarcz, *supra* note 6, at 1260–61.

332. *See also* BORGESIUS, *supra* note 136, at 7; Mittelstadt, *supra* note 329, at 479; Rodrigues, *supra* note 135, at 3; Wachter et al., *supra* note 6, at 6.

333. *See also* Mittelstadt, *supra* note 329, at 488; Gerards & Borgesius, *supra* note 20, at 11; Cîrciumaru, *supra* note 301, at 114.

334. Andrews & Bucher, *supra* note 310, at 171–72 (emphasizing an additional role for intent and thus knowledge of the system's bias).

335. *See also* Barocas & Selbst, *supra* note 162, at 694–95.

336. *See, e.g.*, Xenidis, *supra* note 301, at 745–46.

systems, establishing that a protected trait was used is often extremely difficult. This challenge mirrors traditional evidentiary difficulties faced by plaintiffs in direct discrimination cases, where the internal mental processes of the decision-maker must be inferred. In the context of AI, the obscurity of human reasoning[337] is replaced by the complexity and opacity of algorithmic processes.[338] That complexity may further hinder efforts to demonstrate the use of a protected characteristic in the decision. While the European discrimination regime incorporates a reversal of the burden of proof—requiring the claimant only to establish facts that give rise to a presumption of discrimination, shifting the burden to the respondent to prove that no discrimination occurred[339]—this mechanism only partially mitigates the difficulty posed by AI opacity.

In the absence of an intent requirement,[340] even for indirect discrimination,[341] the European[342] non-discrimination framework resembles a strict liability regime. The motives or efforts of the potentially discriminating party are irrelevant, as liability is based solely on the effects of the conduct. The legitimate aim exception, discussed earlier, provides some nuance in many[343]—though not all[344]—contexts. Its availability and scope vary depending on the context and the applicable directive. As in the U.S. legal framework,[345] this exception is most commonly applied to indirect discrimination,[346] but it may also apply to direct discrimination.[347] In the case of a hiring algorithm, selecting the best candidate may constitute

---

337. *See, e.g.*, Kleinberg et al., *supra* note 6, at 130.

338. *See also* LeRoy, *supra* note 163, at 269–70.

339. *See, e.g.*, Council Directive 2000/43, *supra* note 297, art. 8; *see also* Grozdanovski, *supra* note 316, at 113; Wachter et al., *supra* note 6, at 7.

340. Adams-Prassl et al., *supra* note 307, at 149. Intent *can* play a role, however, if the plaintiff can show that the deployer of the system was aware of the indirect discriminatory effect the system had, and actually intended that effect. That might result in a case of direct discrimination. *See* Hacker, *supra* note 306, at 1154.

341. In this way, the regime differs from the American regime for disparate impact. *See, e.g.*, Pers. Adm'r v. Feeney, 442 U.S. 256, 273–74 (1979).

342. However, it works differently in the US. *See, e.g.*, Andrews & Bucher, *supra* note 310, at 171–72.

343. Hacker, *supra* note 306, at 1161 (arguing that many instances of genuine AI deployment pass the relevant hurdle).

344. Not all forms of discrimination by AI systems can be justified this way and some may merit more strict scrutiny. *See also* von Ungern-Sternberg, *supra* note 163, at 267.

345. *Supra* note 312.

346. Various indirect discrimination definitions directly refer to the possible existence of some justification. *See, e.g.*, Council Directive 2000/43, *supra* note 297 art. 2; *see also* Hacker, *supra* note 306, at 1152–53.

347. *See, e.g.*, Council Directive 2004/113, *supra* note 300, art. 4.

a legitimate aim.[348] However, this raises the question of whether deploying an AI system that potentially produces discriminatory outcomes is a proportionate means of achieving that aim.[349] It invites an inquiry into whether the same goal could be met using alternative, less discriminatory methods—a standard that closely resembles the contextual reasonableness analysis in negligence liability or the feasibility analysis in American products liability. This proportionality assessment calls for consideration of whether reasonable measures could have prevented the discriminatory effect.[350] Such measures may include the use of less biased training data,[351] where available, or a more careful selection of target variables—ensuring that they meaningfully relate to the legitimate aim pursued.[352]

Another related challenge is that AI systems do not necessarily need to rely on a specific protected trait or variable in their decision-making processes. Rather, such systems are often capable of inferring protected characteristics—even when these are not explicitly submitted—through statistical correlations with other variables.[353] This phenomenon is exemplified by the perceived[354] discrimination by COMPAS[355]—a tool used by courts to assess the likelihood of recidivism—despite ethnicity not being

---

348. *See* von Ungern-Sternberg, *supra* note 163, at 266; *see also* Bornstein, *supra* note 6, at 555.

349. *See* von Ungern-Sternberg, *supra* note 163, at 265–66.

350. *See also* Elizabeth Pendo & Jennifer D. Oliva, *Disability Discrimination by Clinical Algorithm*, 103 N.C. L. REV. 187, 236–37 (2024) (discussing federal disability antidiscrimination requirements in an algorithmic context).

351. *See, e.g.*, Barocas & Selbst, *supra* note 162, at 716–19; Bornstein, *supra* note 6, at 535.

352. Barocas & Selbst, *supra* note 162, at 715–16; Hacker, *supra* note 306, at 1161; Kim, *supra* note 6, at 1577 (describing "Problem Formulation" for predictive algorithms).

353. *Supra* note 327; de Laat, *supra* note 328, at 532. See, for example, on the correlation between race and many other socio-economic features, Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L.J. 1043, 1100 (2019).

354. Research showed that COMPAS, in fact, adheres to some fairness metrics, but that it is impossible to adhere to various standards at the same time. *See* Sam Corbett-Davies et al., *A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased Against Blacks. It's Actually Not That Clear*, WASH. POST (Oct. 17, 2016), https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/; Deborah Hellman, *Measuring Algorithmic Fairness*, 106 VA. L. REV. 811, 815–18 (2020); Ludwig & Mullainathan, *supra* note 321, at 81–82; Nachbar, *supra* note 167, at 511; *infra* note 422 and accompanying text.

355. Julia Angwin et al., *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/3KXF-V9PR]; Nachbar, *supra* note 167, at 511; *see also* Mayson, *supra* note 6, at 2221.

included in the input data.[356] This illustrates that direct discrimination may be of limited relevance,[357] thereby expanding the scope and significance of indirect discrimination in the context of AI.[358]

It should also be noted that challenges persist in addressing disparate treatment and indirect discrimination. First, a range of "traditional" difficulties remain salient when AI is involved. These include the absence of clear thresholds[359]—such as how much of a disparate impact is required—and the complexities inherent in identifying both the affected group and the appropriate comparator group.[360] Additionally, many non-discrimination law provisions are grounded in a one-dimensional rationale—focused on a single protected trait—which complicates their application in cases of intersectional discrimination.[361] More generally, these issues relate to the highly contextual nature of non-discrimination law.[362] Furthermore, although indirect discrimination is not as closely tethered to the (internal) decision-making processes of the alleged discriminator, the information asymmetry and opacity of AI systems[363] mean that plaintiffs are generally faced with a heavy burden of proof, being required to demonstrate, *prima facie*, that the AI system in question produces an adverse statistical impact.[364]

In addition, several more AI-specific challenges can be identified. One such challenge concerns the difficulty of formulating a single, quantifiable fairness metric.[365] Two illustrative examples may be noted.[366] The first is

---

356. Borgesius, *supra* note 305, at 1574; Nachbar, *supra* note 167, at 511.

357. For an argument for a broader application of direct discrimination, see Adams-Prassl et al., *supra* note 307, at 167–69.

358. *See* Wachter et al., *supra* note 6, at 19–20 (similarly stressing the importance of indirect discrimination in an algorithmic context); Mainka, *supra* note 316, at 813 (implicitly stressing the same importance of indirect discrimination); Sunstein, *supra* note 6, at 1202.

359. *See, e.g.*, Wachter et al., *supra* note 6, at 3.

360. *See, e.g.*, *id.*

361. *See, e.g.*, Xenidis, *supra* note 301, at 741–42; Wachter et al., *supra* note 6, at 9; Xenidis, *supra* note 20, at 228–29.

362. On which for EU law, see, for example, Wachter et al., *supra* note 6, at 3.

363. *See supra* Section I.B.2.b.

364. *See also* Bornstein, *supra* note 6, at 554; Borgesius, *supra* note 305, at 1577.

365. *See also* Mayson, *supra* note 6, at 2242–48; Hellman, *supra* note 354, at 834–35; Wachter, *supra* note 186, at 688–89; Jacy Reese Anthis et al., The Impossibility of Fair LLMs 2 (June 5, 2025) (unpublished manuscript) (on file with arXiv), https://arxiv.org/pdf/2406.03198 [https://perma.cc/WX9N-6SPB].

366. For nuance on the distinction between group and individual fairness, see Reuben Binns, *On the Apparent Conflict Between Individual and Group Fairness*, *in* PROCEEDINGS OF THE 2020 CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 514, 519–23 (Mireille Hildebrandt & Carlos Castillo eds., 2020).

individual fairness,[367] which emphasizes that individuals should not be treated differently on the basis of a particular characteristic or trait.[368] This concept is closely related to the rationale underpinning disparate treatment and direct discrimination.

A second metric is group fairness.[369] Unlike individual fairness, group fairness focuses not on individuals but on groups. It requires that distinct groups—such as those defined by a protected characteristic—not be treated differently, in aggregate, from other groups—thus assessed on a statistical basis.[370] Group fairness aligns more closely with the statistical tests commonly used to identify disparate treatment or indirect discrimination.[371]

The significance of the lack of a uniform fairness metric becomes particularly evident when one considers that, in many contexts, it is mathematically impossible for a given algorithm to satisfy both criteria simultaneously.[372] This raises the problematic implication that system developers may be unable to avoid both direct and indirect discrimination or disparate treatment and disparate impact, as measures intended to address one form of discrimination may exacerbate the other.[373]

A final key challenge in the context of AI relates more directly to system unpredictability. Even a programmer who harbors no intention of producing a discriminatory system can only take certain measures to mitigate that risk. For example, they may seek to use neutral, bias-free data—although such data is not always available[374]—or deliberately exclude sensitive traits from the dataset. However, it remains impossible to entirely rule out the possibility that the system will develop biases during training or—if subject to further learning during deployment—through its continued use.

---

367. Hacker, *supra* note 306, at 1171; Binns, *supra* note 366, at 515–16.

368. Phrased differently, similar individuals—disregarding any specific protected trait—should be treated similarly. *See, e.g.*, Hacker, *supra* note 306, at 1171; Binns, *supra* note 366, at 516; Yap Jia Qing & Ernest Lim, *A Legal Framework for Artificial Intelligence Fairness Reporting*, 81 CAMBRIDGE L.J. 610, 624 n.64 (2022); Anthis et al., *supra* note 365, at 5.

369. *See, e.g.*, Cynthia Dwork et al., *Fairness Through Awareness*, *in* PROCEEDINGS OF THE 3RD INNOVATIONS IN THEORETICAL COMPUTER SCIENCE CONFERENCE 214, 215 (2012). This relates to "demographic parity," discussed in Kim, *supra* note 6, at 1577–78.

370. Dwork et al., *supra* note 369, at 215; Hacker, *supra* note 306, at 1175; Qing & Lim, *supra* note 368, at 624 n.64.

371. *See also* Hacker, *supra* note 306, at 1175.

372. Mayson, *supra* note 6, at 2233–34; *see also* Nachbar, *supra* note 167, at 511–12; *see also* Dwork et al., *supra* note 369, at 215.

373. Although the legitimate aim exception, see *supra* note 312, nuances this challenge, this may nevertheless present the system deployer with an increased burden of proof and corresponding compliance costs.

374. Wachter, *supra* note 186, at 689 (arguing that such data is "a fantasy").

These challenges are only marginally alleviated by two principal exceptions. The first is the possibility of introducing measures of affirmative or "positive action,"[375] which, in most European jurisdictions, is broader in scope than under current U.S. law.[376] In theory, this enables programmers to directly apply a "bonus" or preferential treatment to groups that might otherwise be disadvantaged by the system, thereby counteracting potential discriminatory effects.[377] Nevertheless, it is crucial to emphasize that, even in a European context, such measures are subject to stringent legal conditions.[378]

The second exception allows that an AI developer or deployer may argue that the use of a system exhibiting some degree of bias constitutes a proportionate means of achieving a legitimate aim—such as hiring the most qualified candidate or similar objectives.[379] This justification, akin to the American business necessity defense for disparate impact,[380] generally presupposes that the system does not directly utilize or reference the protected trait, as the exception applies most broadly to indirect discrimination.[381] Moreover, it would require the developer or deployer to demonstrate, inter alia, that it was unreasonable to further refine or finetune the system[382] (analogous to standards required to avoid negligence liability) and that the legitimate aim could not reasonably be achieved as effectively by less discriminatory means without recourse to the AI system.[383] The mere fact that the system reduces costs[384] or accommodates customer preferences[385] is insufficient to justify the discriminatory effect.

---

375. *See, e.g.*, Council Directive 2000/43, *supra* note 297, art. 5; Council Directive 2006/54, *supra* note 299, art. 3; Council Directive 2000/78, *supra* note 298, art. 7; Council Directive 2004/113, *supra* note 300, art. 6.

376. On the restricted nature of that possibility in the U.S., see, for example, Students for Fair Admissions, Inc. v. President & Fellows of Harvard Coll., 600 U.S. 181, 206 (2023), subjecting affirmative action to the same standards as disparate treatment. In fact, the European approach aligns more closely with pre-2023 U.S. Supreme Court case law. *See, e.g.*, Kim, *supra* note 6, at 1557–60.

377. *See also* Sam Corbett-Davies et al., *Algorithmic Decision Making and the Cost of Fairness*, *in* PROCEEDINGS OF THE 23RD ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING 797, 802–05 (2017).

378. *See, e.g.*, Council Directive 2000/43, *supra* note 297, art. 5; Council Directive 2000/78, *supra* note 298, art. 7.

379. *Supra* note 348 and accompanying text.

380. Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971).

381. *Supra* note 348 and accompanying text.

382. Hacker, *supra* note 306, at 1163–64.

383. See in the same sense, *id.* at 1162.

384. See, for example, in a non-AI context, holding that purely economic reasons are insufficient, Case C-388/07, Inc. Trs. of the Nat'l Council on Ageing (Age Concern Eng.) v.

III. REGULATORY IMPACTS

A. Ethics/Values

1. AI Act

a. Liability Law

While the AI Act does not directly aim to regulate liability law, this area remains highly relevant due to its close connection to the Act's core ethical goals of accountability and safety in AI.[386] In the context of liability, one clear contribution of ethics-based regulation is the introduction of procedural requirements designed to help ensure that developers and deployers retain access to information necessary to establish or refute negligence, defect, or causation. Traditional liability frameworks do not expressly impose such duties. While these procedural rules are a step forward, they fall short of fully addressing the underlying problems—evidenced by the European Commission's initial proposal to introduce a presumption of causality,[387] recognizing that not all legal systems require equally robust cooperation in evidence production.

On a substantive level, there is also a link between the AI Act's requirements and the contextual standards for assessing defects, negligence, or causation—particularly the first two. By articulating specific compliance obligations, the AI Act may allow conformity with these obligations to serve as evidence of due diligence or absence of defect, a connection reflected in Article 7(2) of the revised European Product Liability

---

Sec'y of State for Bus., Enter. & Regul. Reform, 2009 E.C.R. I-1569. Similarly in the US, see, for example, *Griggs*, 401 U.S. at (1971) ("The touchstone is business necessity. If an employment practice which operates to exclude [the protected group] cannot be shown to be related to job performance, the practice is prohibited.").

385. *See, e.g.*, Case C-54/07, Centrum voor gelijkheid van kansen en voor racismebestrijding v. Firma Feryn NV, 2008 E.C.R. I-5187; Case C-188/15, Bougnaoui v. Micropole SA*,* ECLI:EU:C:2017:204 (Mar. 14, 2017); von Ungern-Sternberg, *supra* note 163, at 276. See similarly in the US, Diaz v. Pan Am. World Airways, Inc., 442 F.2d 385, 387–88 (5th Cir. 1971); Wilson v. Sw. Airlines Co., 517 F. Supp. 292, 304 (N.D. Tex. 1981).

386. *See* Regulation 2024/1689, *supra* note 3, art. 1; *id.* recital 1 at 1 ("to protect against the harmful effects of AI systems in the Union"); *id.* recital 2 at 1 (implicitly by referring to fundamental values); *id.* recital 27 at 8 (referring to the ethics guidelines which include accountability); *id.* recital 59 at 17; *id.* recital 114 at 30.

387. *See Proposal for a Directive of the European Parliament and of the Council on Adapting Non-contractual Civil Liability Rules to Artificial Intelligence (AI Liability Directive)*, COM (2022) 496 final (Sept. 28, 2022).

Directive.[388] However, compliance with the Act does not translate directly into satisfaction of liability standards.[389]

Liability law shows that such ethically inspired regulatory requirements are likely overinclusive.[390] Many will likely extend far beyond what is traditionally demanded under products liability or negligence law. This overbreadth partly results from the AI Act's broad, risk-based classification approach: all AI systems deployed in high-risk settings are subject to high-risk requirements,[391] regardless of whether they actually present contextually high risks.[392]

This overinclusiveness is further reflected in the Act's structuring of independent obligations for oversight, explainability, and accuracy. In contrast, traditional liability regimes treat these factors as interdependent.[393] The level of accuracy required in negligence or products liability contexts depends on the risks posed by lower accuracy, the feasibility of improvement, and the potential mitigating effects of warnings, explainability, or oversight.[394] The Learned Hand test,[395] for example, weighs the burden of precautions against the probability and magnitude of harm. The contextual nature of negligence and products liability law means that highly accurate systems may justify reduced expectations for explainability or oversight.[396] The inherent imperfection and opacity of most AI systems reinforce the need for a contextual, holistic approach—focused on how these challenges were addressed in each case.

---

388. *See* s*upra* note 270.

389. The AI Act can only constitute a *lex specialis* overriding existing frameworks to the extent that it regulates some aspect. *See* Kristof Meding, It's Complicated: The Relationship of Algorithmic Fairness and Non-Discrimination Provisions for High-Risk Systems in the EU AI Act 9 (Jan. 22, 2026) (unpublished manuscript) (on file with arXiv), https://arxiv.org/pdf/2501.12962 [https://perma.cc/2XAN-HQP9]. Since the Act does not exempt defendants from tort liability in case of compliance, it does not override stricter diligence obligations arising under general tort law.

390. They are also underinclusive, see *infra* note 398 and accompanying text, in the context of liability law, given the highly contextual nature of liability thresholds, see *supra* Section III.A, relative to the more abstract, ethically inspired requirements of the AI Act.

391. With the exception discussed earlier, see *supra* Section I.B.3.a.

392. *See* Daniel Leufer et al., *The Pitfalls of the European Union's Risk-Based Approach to Digital Rulemaking*, 71 UCLA L. REV. DISCOURSE 156, 168 (2024).

393. *See supra* Section II.B.

394. *See infra* Section III.B.

395. *See supra* note 254.

396. In fact, there may be an implied trade-off in some of these regards. *See infra* Section IV.A.1.b.

Accordingly, the AI Act's substantive obligations often diverge from and exceed those in traditional liability law. From the standpoint of liability doctrine,[397] such extensions are difficult to justify, as they depart from reasonableness-based or strict standards in favor of new ethical mandates lacking a clear basis within liability law.

At the same time, the Act's requirements are also underinclusive. Most provisions are restricted to high-risk systems, even though so-called "low-risk" AI can cause similar harm.[398] Even within the high-risk category, negligence and products liability law's contextual nature may sometimes demand more or less than what the AI Act requires—meaning stricter or more lenient standards may be appropriate depending on the circumstances.

In summary, the AI Act's substantive requirements do not align with the contextual standards of liability law. Rather than adapting to or reinforcing the established doctrines of negligence and products liability, the Act introduces its own abstract obligations without clearly addressing existing deficiencies. This contrasts with its procedural rules, which meaningfully—though not comprehensively—help reduce proof burdens and improve access to remedies for victims of AI-related harms.

b. Non-Discrimination Law

At first glance, one might expect the AI Act to make a significant contribution to non-discrimination law, given its clear acknowledgment that bias and discrimination are core challenges in AI contexts.[399] Yet, as with liability law, this recognition has not resulted in provisions that effectively

---

397. Admittedly, the AI Act does not directly aim to facilitate liability; however, it does place a strong emphasis on accountability.

398. See in the same sense Botero Arcila, *supra* note 253, at 9; Leufer et al., *supra* note 392, at 168. Similarly, the exception under Article 6(3), see *supra* note 204 and accompanying text, for "preparatory tasks" may prove undesirable where the resulting AI-generated content is merely adopted by a human without meaningful scrutiny. *Cf.* Fabian Lütz, *The AI Act, Gender Equality and Non-Discrimination: What Role for the AI Office?*, 25 ERA F. 79, 81–82 (2024) (analyzing the AI Act's approach to gender equality and non-discrimination). This concern is particularly acute when the task becomes routine in nature, as human performance in verifying such outputs tends to decline under repetitive or low-engagement conditions. *See, e.g.*, Michael Greenberg & M. Susan Ridgely, *Clinical Decision Support and Malpractice Risk*, 306 JAMA 90, 90 (2011); Selbst, *supra* note 5, at 1347–48; *see also* S.B. 24-205, 74th Gen. Assemb., 2024 Reg. Sess. § 6-1-1701(9)(b)(I)(A) (Colo. 2024).

399. The recitals to the AI Act mention contain dozens of references to "discrimination," "non-discrimination" and "discriminatory," as well as a dozen to bias-related challenges. *See* Lütz, *supra* note 398, at 79; Serena Oduro et al., *Obligations to Assess: Recent Trends in AI Accountability Regulations*, 3 PATTERNS 1, 5 (2022); S.B. 24-205, 74th Gen. Assemb., 2024 Reg. Sess. § 6-1-1701(1)(a) (Colo. 2024).

address the deficiencies of the existing legal framework.[400] Instead, the AI Act's substantive requirements are, in important respects, both overinclusive[401] and underinclusive.[402]

With respect to high-risk systems, it is notable that a system's designation as "high-risk" generally depends on its deployment in sectors that largely[403]—but not entirely[404]—overlap with those regulated by existing discrimination directives. Once classified as high-risk, systems become subject to various obligations[405] regarding accuracy, human oversight[406] and explainability,[407] which may, at least indirectly, help mitigate bias and discrimination. However, these obligations are not sufficiently tailored: they may impose unnecessary burdens in some contexts, while failing to address persistent gaps in others where non-discrimination law already struggles.

Directly addressing bias and discrimination,[408] the AI Act requires that continuously learning AI systems mitigate biased outputs with "appropriate mitigation measures."[409] Providers of high-risk AI must also take steps to prevent and address bias in training data,[410] adopting "appropriate

---

400. See similarly on Article 10 of the AI Act: Philipp Hacker, *A Legal Framework for AI Training Data–From First Principles to the Artificial Intelligence Act*, L. INNOVATION & TECH. 1, 42 (2021).

401. To be clear, the overinclusiveness of the AI Act with regard to non-discrimination law well exceeds the analysis here, which focuses on the general principles. An additional example is that the Act prohibits social scoring in certain contexts—citing the risks of discrimination, see, for example, Palmiotto, *supra* note 163, at 8, without tying the ban to an actual assessment of discrimination. *See* Regulation 2024/1689, *supra* note 3, art. 5(1)(c); *id.* recital 31 at 9.

402. Much like in the context of liability law, the AI Act does not constitute *lex specialis* concerning non-discrimination law, see *supra* note 389, as it does not address discriminatory outcomes. Meding, *supra* note 389, at 9.

403. *See* Regulation 2024/1689, *supra* note 3, art. 6(2); *id.* annex III; *see also* Lütz, *supra* note 398, at 81 (addressing gender equality and the corresponding directives).

404. See, for example, Regulation 2024/1689**,** *supra* note 3, annex III (2), which refers to critical infrastructure and is not directly cited in the EU's discrimination directives.

405. *See supra* Section I.B.3.a.

406. Humans are sometimes said to be better equipped to spot bias than algorithms. *See* Mousa Alshanteer, *A Current Regime of Uncertainty: Improving Assessments of Liability for Damages Caused by Artificial Intelligence*, 21 N.C. J.L. & TECH. 27, 38 (2020).

407. *See, e.g.*, Nachbar, *supra* note 167, at 544–46. See with more skepticism, Kroll et al., *supra* note 6, at 659–60.

408. Regulation 2024/1689, *supra* note 3, art. 9 more generally, imposes the establishment of a "risk management system." *See also* Meding, *supra* note 389, at 6.

409. Regulation 2024/1689, *supra* note 3, art. 15(4); Meding, *supra* note 389, at 8–9.

410. Regulation 2024/1689, *supra* note 3, art. 10; Hacker, *supra* note 400, at 40–41; Meding, *supra* note 389, at 6–8; Marvin van Bekkum, *Using Sensitive Data to De-Bias AI Systems: Article 10(5) of the EU AI Act*, 56 COMPUT. L. & SEC. REV. 1, 1 (2025); *see also* Wachter, *supra* note 186, at 689 (arguing that this is impossible as "[n]eutral data is a fantasy").

measures"[411] "to the best extent possible."[412] A limited exemption permits processing personal data for this purpose,[413] and the Act recognizes that it may not always be feasible to fully correct the dataset.[414] Crucially, the Act does not specify a fairness metric,[415] failing to resolve this fundamental challenge of applying non-discrimination law to AI and the resulting potential for conflicting obligations.[416]

Even with careful data selection and monitoring, discrimination risks persist: "The only way to ensure that a data mining system will not discriminate is to not use it."[417] Moreover, imposing specific forms of equality may[418] reduce system accuracy[419]—not because equality is undesirable,[420] but because, as with the broader European regime, non-discrimination law is highly contextual and resists abstract application.[421] While algorithmic systems can help quantify such trade-offs,[422] this underscores the need for clear guidance on their permissible limits, both regarding indirect discrimination (legitimate aims, proportionality) and

---

411. Regulation 2024/1689, *supra* note 3, art. 10(2)(g).

412. *Id.* art. 10(3); *see also* Nachbar, *supra* note 167, at 549; Kim, *supra* note 6, at 1575–76; Lütz, *supra* note 398, at 84; Wachter, *supra* note 186, at 689.

413. Regulation 2024/1689, *supra* note 3, art. 10(5); Hacker, *supra* note 400, at 41; Veale & Borgesius, *supra* note 234, at 103; Marvin van Bekkum & Frederik Zuiderveen Borgesius, *Using Sensitive Data to Prevent Discrimination by Artificial Intelligence: Does the GDPR Need a New Exception?*, 48 COMPUT. L. & SEC. REV. 1, 9–11 (2023); Meding, *supra* note 389, at 8; van Bekkum, *supra* note 410, at 7–9.

414. See similarly on that impossibility: Hacker, *supra* note 400, at 42.

415. *See* Luca Deck et al., Implications of the AI Act for Non-Discrimination Law and Algorithmic Fairness 4 (June 26, 2024) (unpublished manuscript) (on file with arXiv), https://arxiv.org/abs/2403.20089 [https://perma.cc/8TZ5-NHKZ]; Wachter, *supra* note 186, at 687–89.

416. *Supra* Section II.B.

417. Selbst, *supra* note 314, at 164; *see* Mayson, *supra* note 6, at 2277 ("The third and increasingly most prevalent strategy for promoting racial equity in prediction is to resist the use of algorithmic methods altogether."); Adams-Prassl et al., *supra* note 307, at 144 ("[N]o automated system is completely free of bias.").

418. This is not necessarily the case. *See, e.g.*, Kleinberg et al., *supra* note 167, at 277–78; Sunstein, *supra* note 6, at 1201–02.

419. Corbett-Davies et al., *supra* note 377, at 797 ("Adhering to past fairness definitions can substantially decrease public safety; conversely, optimizing for public safety alone can produce stark racial disparities."); Mayson, *supra* note 6, at 2249.

420. Much like how other requirements—such as explainability—are sometimes said to limit system performance. *See, e.g.*, UDAY KAMATH & JOHN LIU, EXPLAINABLE ARTIFICIAL INTELLIGENCE: AN INTRODUCTION TO INTERPRETABLE MACHINE LEARNING 13–15 (2021).

421. *Supra* Section II.B.

422. Sunstein, *supra* note 6, at 1203; *see, e.g.*, Corbett-Davies et al., *supra* note 377, at 802–05.

positive action.[423] As observed, "[s]ome of the tradeoffs might well be painful, but in general, it is best to know what they are."[424]

As the Act's relevant standards remain to be clarified,[425] it currently imposes abstract requirements in an area governed by highly contextual assessments. Thus, the Act may, at times, exceed non-discrimination law—where a system avoids biased outcomes despite biased data due to effective oversight[426]—or fall short, where discriminatory results emerge from systems trained on compliant data. In either case, compliance with the AI Act does not guarantee compliance with non-discrimination law, and vice versa.

The Act does introduce targeted procedural obligations that may help address bias and discrimination more effectively,[427] such as mandatory fundamental rights impact assessments for high-risk AI,[428] the right to lodge complaints,[429] and the right to an explanation,[430] all of which may enhance enforcement and identify discriminatory outcomes.[431]

Broader procedural requirements—like recordkeeping—also support plaintiffs,[432] especially when paired with a right to an explanation or in cases involving bad-faith defendants. Where developers act in good faith, the reversal of the burden of proof[433] in discrimination cases already incentivizes meticulous record-keeping. Even without good faith, a refusal to disclose information may strengthen a plaintiff's prima facie case.[434] Accordingly, the AI Act's procedural rules are most valuable in deterring or addressing bad-faith AI development and deployment.

---

423. *See* Wachter, *supra* note 186, at 688.

424. Sunstein, *supra* note 6, at 1203; *see also* Adams-Prassl et al., *supra* note 307, at 171.

425. *Supra* Section I.B.3.b.

426. *See* Hacker, *supra* note 400, at 42–43.

427. *See* Deck et al., *supra* note 415, at 2.

428. Regulation 2024/1689, *supra* note 3, art. 27; Lütz, *supra* note 398, at 84.

429. Regulation 2024/1689, *supra* note 3, art. 85.

430. *Id.* art. 86. The impact of Article 86 is arguably nuanced by the pre-existence of the EU's GDPR, *supra* note 190, which is sometimes argued to contain a right to explanation based on recital 71—though that view is not apparent from the Articles of the Regulation itself. *See* Adrien Bibal et al., *Legal Requirements on Explainability in Machine Learning*, 29 A.I. & L. 149, 151–152 (2021); Ljupcho Grozdanovski, *Non-Discrimination Law, the GDPR, the AI Act, and the—Now Withdrawn—AI Liability Directive Proposal Offering Gateways to Pre-Trial Knowledge of Algorithmic Discrimination*, 5 AI & ETHICS 5039, 5061 (2025).

431. *See also* Lütz, *supra* note 398, at 84; Grozdanovski, *supra* note 430, at 5050.

432. *Supra* Section III.A.1.a.

433. *Supra* Section II.B.

434. *See* Case C-415/10, Meister v. Speech Design Carrier Sys. GmbH, ECLI:EU:C:2012:217, ¶ 45 (Apr. 19, 2012).

While these procedural rules provide some value, the overall conclusion mirrors that reached regarding liability law. The AI Act primarily creates its own regime, which does not effectively reduce prevailing uncertainties about the application of non-discrimination law to AI. It offers limited guidance on critical issues such as quantifying discrimination, setting statistical thresholds, addressing proxy variables and intersectionality, and determining when biased systems may be lawfully deployed.

This reinforces earlier conclusions from the liability analysis. For potential victims of discrimination, the Act provides some additional safeguards, but these are likely insufficient to offset the increased complexity and legal hurdles introduced by AI—particularly since protection is largely limited to high-risk systems. For developers and deployers, the AI Act raises compliance obligations without delivering corresponding legal certainty.[435]

c. General Observations

The preceding Sections have demonstrated that the European AI Act's ethically motivated requirements hinge on whether a system is classified as an "AI system," and, further, as "high-risk." From the standpoint of liability and non-discrimination law, both criteria render the Act underinclusive, as they do not meaningfully supplement existing legal frameworks. Limiting obligations to "high-risk" systems allows similar technologies used in equally risk-prone settings[436] to evade the Act's requirements.[437] As a result, their developers and deployers are merely "encouraged," rather than required, to follow the Act's provisions for high-risk AI[438]—despite the reality that "low-risk" systems can also inflict harm or perpetuate discrimination.

The definition of "AI system" itself poses a similar challenge. Its complexity excludes certain technologies from the Act's scope altogether, even though many concerns addressed by liability and, particularly, non-

---

435. *Infra* Section IV.B.1.

436. *See* Wachter, *supra* note 186, at 709–10; Nicoletta Rangone & Luca Megale, *Risks Without Rights? The EU AI Act's Approach to AI in Law and Rule-Making*, 16 EUR. J. RISK REGUL. 1082, 1084–88 (2025).

437. *See* Rangone & Megale, *supra* note 436, at 1091–94. Such systems are "only" subject to a transparency requirement if they interact with humans, see Regulation 2024/1689, *supra* note 3, art. 50, in addition to the more broad AI literacy requirement found in Article 4 (sometimes deemed "insufficient" to add meaningful value in discrimination contexts, see Meding, *supra* note 389, at 10).

438. Regulation 2024/1689, *supra* note 3, art. 95; Almada & Petit, *supra* note 56, at 91.

discrimination law long predate AI and apply broadly to algorithmic systems.[439] This definitional underinclusiveness risks leaving victims of AI-driven harm or discrimination without recourse and fragments the regulatory landscape, as these broader systems fall under disparate legal regimes, undermining coherence.

Substantively, both the procedural and substantive thresholds for compliance remain to be determined, with significant input from industry, in a manner reminiscent of ethical codes of conduct.[440] Notably, the Act treats requirements such as human oversight, explainability, accuracy, and bias mitigation in isolation, whereas existing liability[441] and non-discrimination[442] frameworks tend to adopt a more integrated, holistic approach. Moreover, as previously discussed, enhancing one element—such as accuracy[443]—can compromise others, like explainability,[444] and existing technological constraints often preclude the full optimization of these various values.[445]

The resulting regulatory model is highly abstract, raising concerns of both underinclusiveness and overinclusiveness. Some requirements of the AI Act may exceed what is necessary for effective application of existing liability and non-discrimination law, while in areas where genuine legal gaps exist, the Act's ethics-driven provisions often fail to provide meaningful supplementation—and may even fall short of established standards.

2. Moratorium

a. Liability Law

At first glance, the seemingly opposing American approach of a regulatory moratorium would leave most existing legal challenges unresolved. There would be no intervention to ease the burden of proof for

---

439. *See, e.g.*, Bozdag, *supra* note 161, at 209; Mittelstadt et al., *supra* note 27, at 8–9; Mayson, *supra* note 6, at 2221; Hellman, *supra* note 354, at 813–14; Xenidis, *supra* note 20, at 223–24.

440. *See supra* Section II.B.3.b.

441. *See supra* Section III.A.

442. *See supra* Section II.B.

443. *See supra* Section II.A.2.b.

444. *See supra* Section II.A.2.b.

445. *See supra* Section II.A.2.b.

injured parties, nor would it resolve the persistent legal uncertainty surrounding the contextual thresholds relevant to liability.

Critically, it would remain unclear which forms of AI development and deployment breach standards of reasonableness or negligence, or how established frameworks such as the Learned Hand formula should apply in AI contexts. The same ambiguity surrounds questions of causation linking a duty breach or product defect to resulting harm. Further, current liability regimes would continue to provide insufficient incentives for socially beneficial behavior by AI developers.

From an AI ethics and safety perspective, this means that the normative commitments embedded in existing liability regimes would remain poorly realized in AI settings. While a regulatory moratorium may not obstruct the field to the extent some fear from the AI Act, it equally fails to address any of the core legal concerns. To be sure, the moratorium's openness to court-developed standards could, over time, nuance these challenges. Yet this raises two key issues: first, it is uncertain whether courts—without the benefit of interdisciplinary expertise—are best positioned to develop appropriate standards; and second, judicial evolution is inevitably slow, which is problematic given the rapid pace of AI development. That said, similar reservations can be raised regarding the capacity of other regulatory bodies.

It is notable, in this context, that various state legislatures are actively pursuing their own frameworks to address these gaps. For example, Rhode Island has considered a bill imposing strict liability on frontier AI developers.[446] In California, SB 1047[447]—ultimately vetoed, partly due to innovation concerns—would have introduced a liability regime for high-risk AI models, while the more recent SB 813[448] would establish safety standards for AI models and applications, offering certification as evidence of reasonable care. Yet, all such state initiatives would likely be preempted by a federal regulatory moratorium.

At the federal level, the proposed RISE (Responsible Innovation and Safe Expertise) Act—distinct from New York's RAISE Act—would confer civil immunity on AI developers whose models are used by learned professionals, provided certain transparency and documentation

446. S.B. 358, 2025 Gen. Assemb., Jan. Sess. (R.I. 2025).
447. S.B. 1047, 2023–2024 Leg., Reg. Sess. (Cal. 2024).
448. S.B. 813, 2025–2026 Leg., Reg. Sess. (Cal. 2025).

requirements are satisfied.[449] Notably, such federal initiatives could persist even if a federal regulatory moratorium were to preempt state action.

b. Non-Discrimination Law

As with liability law, a regulatory moratorium without additional substantive measures would leave many of the challenges facing non-discrimination law unresolved. Plaintiffs would continue to face significant difficulties in proving discrimination facilitated by AI, largely due to pronounced information asymmetries—a problem that persists even with existing burden-shifting frameworks. Moreover, current legal regimes would remain ill-equipped to address novel algorithmic biases, whether they intersect with or diverge from established protected characteristics.

On a more substantive level, AI developers still face considerable challenges in complying with non-discrimination law. Efforts to mitigate bias—whether through individual or group fairness metrics—carry the risk of violating legal standards either way. These challenges are compounded by the inherent unpredictability of many AI systems: a system that seems compliant today may result in discriminatory outcomes tomorrow. This underscores the highly contextual nature of non-discrimination law and the difficulties inherent in its application to AI.

Equally concerning is the lack of experience and clarity in applying existing legal exceptions within AI contexts, paralleling issues in liability law regarding the identification of suitable benchmarks.[450] Additionally, the likely increasing prominence of disparate impact and indirect discrimination regimes presents challenges for which current legal frameworks are not fully prepared.

There are, however, some noteworthy state-level initiatives in non-discrimination law. California's Civil Rights Department, for example, has adopted regulations for automated decision systems,[451] imposing anti-bias testing, mitigation, and recordkeeping requirements on employers using such systems. In Illinois, recently enacted legislation[452] prohibits the use of AI in employment decisions if it results in discrimination based on protected characteristics and requires disclosure of AI use to employees and applicants. Conversely, the proposed Stop Discrimination by Algorithms

449. RISE Act of 2025, S. 2081, 119th Cong. (2025).
450. *See supra* Section III.A.
451. CAL. CODE REGS. tit. 2, §§ 11009–11013 (2025).
452. H.B. 3773, 103rd Gen. Assemb., Reg. Sess. (Ill. 2024).

Act[453] in Washington, DC—which would have broadly prohibited algorithmic discrimination—failed to advance.

Most of these adopted state-level efforts would be preempted by a broad regulatory moratorium such as the one that was proposed.[454]

c. General Observations

On a broader level, absent a deliberate intent to advance AI-specific values, there appears to be no compelling reason to depart from the established balances embodied in existing regulations—which, by their nature, already govern instances of AI development and deployment. The shortcomings in safeguarding certain values are not attributable to regulatory intervention or intrinsic limits of regulatory scope, but rather to the practical and theoretical challenges unique to AI. In this sense, even open-ended legal norms may, in fact, be "outpaced" by the realities of AI after all.[455]

This brief examination of the effects of a regulatory moratorium demonstrates that adopting such a moratorium—without addressing the existing, predominantly practical obstacles to applying non-discrimination and liability law in AI contexts—is far from a neutral position.[456] Rather, it constitutes an affirmative decision to restrict the legal reach of established mechanisms, such as liability and non-discrimination law, thereby precluding their traditional—and, in many cases, constitutionally or otherwise normatively required—application to AI.

3. Moving Forward

Liability and non-discrimination laws serve core societal functions: liability law incentivizes responsible behavior through economic consequences,[457] while non-discrimination law enshrines the constitutional diagnosis that discrimination harms society. This connects to the subsequent innovation discussion; an AI regime—whether the result of deliberate

---

453. Council of D.C. B25-0114, 25th Council, 1st Sess. (D.C. 2023), https://lims.dccouncil.gov/Legislation/B25-0114 [https://perma.cc/LU7N-PSB3].

454. One Big Beautiful Bill Act, H.R. 1, 119th Cong. § 43201 (2025).

455. *See supra* note 56, though this is a narrower interpretation than usually attributed to the "pacing problem."

456. *See supra* Section III.A.

457. *See, e.g.*, William M. Landes & Richard A. Posner, *The Positive Economic Theory of Tort Law*, 15 GA. L. REV. 851, 858 (1980); SHAVELL, *supra* note 96, at 182–84.

policy or regulatory inaction—can arguably not be sustainable if it undermines these established legal safeguards.[458]

More fundamentally, any attempt to impose a values-driven AI regime should be grounded in a thorough assessment of existing legal frameworks and their effectiveness in upholding such values. Otherwise, there is a risk of eroding these protections. In any case, procedural mechanisms remain central, as procedural requirements form a crucial boundary for non-discrimination and liability law and are likely to play a similar role in other domains.

While state-level initiatives illustrate some of the challenges seen in the European AI Act or the federal moratorium, they also highlight aspects of effective solutions. For example, Illinois's requirement that job applicants be notified of AI use[459] enables them to challenge adverse AI-driven decisions. Likewise, a key strength of the California proposal[460] is its integration with existing liability law via a relevant presumption, avoiding unnecessary new regulatory layers.

The foregoing analysis yields several insights for future regulatory design. Many provisions in the European AI Act appear to elevate—or, arguably, reduce—ethical standards to binding legal rules. This "AI ethics capture" overlooks both the strengths and limits of ethics as a regulatory tool[461] and fails to provide a framework that adequately protects ethical values as understood within traditional legal systems. In key respects, the AI Act falls short of the balance achieved by existing non-discrimination and liability regimes.

While some critics contend the AI Act did not go far enough in advancing ethics,[462] the non-discrimination and liability analysis here suggests the central issue is its lack of focus. Rather than assessing and supplementing existing frameworks where needed, the Act assumes—incorrectly[463]—that AI operates in a legal vacuum.

---

458. *See supra* note 227 and accompanying text.

459. H.B. 3773, 103rd Gen. Assemb., Reg. Sess. (Ill. 2024).

460. CAL. CODE REGS. tit. 2, §§ 11009–11013 (2025).

461. *See supra* Section II.B. While some authors dispute that the AI Act reflects ethics, see, for example, ANDERSON, *supra* note 137, at 1, they do so, not on the basis of its content (which largely mimics the preceding Ethics Guidelines), but rather its methods.

462. *See, e.g.*, Grozdanovski, *supra* note 430, 5061.

463. Carrillo, *supra* note 92, at 14; Vogel et al., *supra* note 52, at 1037; *see also* HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *supra* note 101, at 6 ("AI systems do not operate in a lawless world. A number of legally binding rules at European, national and international level already apply or are relevant to the development, deployment and use of AI

Many AI-related challenges are already addressed, to varying degrees, by traditional legal regimes.[464] This does not preclude tailored policy responses for unique AI challenges, whether substantive (e.g., complicating application of existing laws[465]) or quantitative (e.g., scale of violations taxing enforcement[466]).

Even then, there is little justification for layering on ethical requirements without restoring the foundational balancing exercise—incorporating ethical considerations—that informed the original frameworks. Given the blurred boundaries between AI and other algorithmic systems, new regulation should be anchored in existing law and narrowly tailored to address distinct AI challenges.[467] Otherwise, AI may be subjected to a unique regulatory regime while highly similar non-AI systems escape that regime. This also reflects the abstract, "all-or-nothing" approach adopted in the AI Act,[468] which makes it difficult to impose a stricter regime on systems that clearly exhibit the risks driving that AI regulation.[469] As a result, less problematic algorithms that technically fall within the broad definition of AI are subjected to the same regulatory regime.[470]

As a result, many challenges within these legal regimes remain unresolved.

---

systems today."); Nachbar, *supra* note 167, at 543 (discussing non-discrimination law in the U.S.).

464. Carrillo, *supra* note 92, at 14; *see also* HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *supra* note 101, at 6; Veale & Borgesius, *supra* note 234, at 98 (referencing "prohibited" AI systems under the AI Act).

465. Non-discrimination law, discussed earlier, is arguably a nice example. *See supra* Section II.B.

466. This is, arguably, the case for deep fakes, which are far more easily generated using AI tools than in instances of traditional audio, image, or video manipulation.

467. For a similar critique of the AI's approach to fundamental rights protection, see Almada & Petit, *supra* note 56, at 95–97.

468. See similarly Rangone & Megale, *supra* note 436, at 1091–92.

469. Article 3(1) AI Act lists seven elements: they are machine-based systems; they are designed to operate with varying levels of autonomy; they may exhibit adaptiveness after deployment; they have explicit or implicit objectives; they infer, from the input they receive, how to generate outputs; their output takes forms such as predictions, content, recommendations, or decisions; their output can influence physical or virtual environments. Regulation 2024/1689, *supra* note 3, art. 3; *see also* Commission Guidelines on the Definition of an Artificial Intelligence System, Annex to the Communication to the Commission, 924 final (Feb. 6, 2025), https://eimin.lrv.lt/public/canonical/1739857334/5400/Guidelines_on_the_definition_of_AI%20system.pdf [https://perma.cc/Q832-L4WW].

470. A broad interpretation of the definition of AI systems in Article 3(1) of the AI Act would encompass many systems that few would have historically considered to be AI. Regulation 2024/1689, *supra* note 3, art. 3.

B. Innovation and Competitiveness

1. AI Act

Notably, the goal of fostering innovation is not exclusive to the proposed regulatory moratorium; the European AI Act likewise aims to promote innovation,[471] ensure EU competitiveness,[472] and harmonize AI regulation across Member States.[473] However, as the preceding analysis suggests, the Act's ethical orientation is unlikely to achieve these objectives.

The Act's overinclusive provisions are unlikely to spur meaningful improvements in innovation or competitiveness in the short term. From the standpoint of liability and non-discrimination, the Act's substantive requirements may unduly restrict AI development and deployment beyond what is required under existing law, creating unnecessary barriers to innovation. Conversely, the Act's underinclusive elements fail to realize potential innovation gains, as the baseline obligations of traditional legal frameworks remain. As a result, even when the Act permits certain conduct, developers and deployers remain exposed to liability or discrimination claims.

These substantive issues are not addressed by the Act's "regulatory sandbox" mechanism,[474] which ostensibly targets innovation. While sandbox experimentation offers certain advantages, those benefits are arguably diminished if stringent regulatory requirements are ultimately imposed at the point of AI system deployment.

The goal of Union-wide harmonization is similarly problematic. While EU-wide regulation is a sound objective, the continued fragmentation of Member State legal frameworks—particularly in liability law—undermines this ambition. As the Act itself acknowledges, divergent national rules perpetuate internal market fragmentation and decrease legal certainty for AI operators.[475] Without alignment of these underlying frameworks, an additional regulatory layer does little to promote true harmonization. Moreover, even if the Act increases legal certainty, its particularly severe sanctions[476] could further discourage innovation.

---

471. *See, e.g.*, *id.*, art. 1; *id.* recital 1 at 1; *id.* recital 2 at 1; *id.* recital 8 at 2.

472. *See, e.g.*, *id.* recital 121.

473. *Id.* recital 8.

474. *Id.* Chapter VI.

475. *Id.* recital 3; *see also id.* recital 1 at 1; *id.* recital 12 at 4; *id.* recital 83 at 23; *id.* recital 84 at 23–24; *id.* recital 97 at 26; *id.* recital 177 at 44.

476. Hacker, *supra* note 400, at 42.

The Act's broad scope often results in overinclusiveness, which may unnecessarily constrain AI development. In this sense, value-driven regulation comes at the expense of legal coherence and tradition.[477] Although a more abstract, principle-based approach could balance value protection and innovation,[478] its benefits are largely negated by the continued applicability of contextual legal regimes in Member States.[479] In sum, while the AI Act aspires to both ethical and legal ideals, its preference for ethical principles over established legal rules suggests it will neither adequately protect victims of AI-related harm and discrimination nor promote European AI competitiveness.

Interestingly, the Act's more procedural requirements are significantly less costly[480] while yielding a demonstrable positive impact on the value domain. Moreover, the Act's general-purpose model regulation,[481] which seems to have inspired the Californian SB 53,[482] illustrates how such procedural measures—including recordkeeping and information obligations—can keep regulators informed, enabling them to govern the future development of AI capabilities and thus bridge the gap between regulating contemporary and prospective AI.

2. Moratorium

The proposed regulatory moratorium would avoid many of the pitfalls found in the AI Act, but it would also fail to fully protect innovation and AI competitiveness. Moreover, without any value protection, the regime could prove unsustainable over time.[483] More substantively, the moratorium would merely prevent states from enacting targeted AI regulation, without addressing the broader reach of common law. This would allow courts to interpret existing laws in AI contexts, leading to a fragmented and inconsistent regulatory landscape across states—particularly in areas such as liability and non-discrimination. Such fragmentation may become more pronounced as the need for a sustainable regulatory solution becomes increasingly clear.

---

477. See similarly Carrillo, *supra* note 92, at 6.

478. *See, e.g.*, Martin Ebers, *Truly Risk-based Regulation of Artificial Intelligence: How to Implement the EU's AI Act*, 16 EUR. J. RISK REGUL. 684, 685 (2025).

479. *Supra* note 389.

480. *See supra* Section I.B.3.b.

481. *See supra* Section I.B.3.a.

482. S.B. 53, 2025–2026 Leg., Reg. Sess. (Cal. 2025).

483. *Supra* note 227 and accompanying text.

In addition, the European AI Act has significantly altered the regulatory environment. While it is uncertain whether the full force of a "Brussels effect"[484] will ever apply, American AI providers seeking access to the EU market are already indirectly affected by the Act.[485] They may be required to thoroughly document their AI development processes to meet European requirements. While this does not necessarily mean that U.S. regulation should mirror these requirements, it does complicate the innovation landscape and would mitigate the burden of similar American regulations.

IV. A PARETO PRINCIPLE FOR AI REGULATION

The preceding Sections have shown that regulation focused solely on protecting legally enshrined values, or solely on fostering innovation, often falls short of achieving its intended goal. Equally problematic is regulation that overlooks deficiencies in existing frameworks' protections. Both diagnoses, however, help surface principles to guide future AI governance that is more effective at protecting values *and* encouraging innovation. Specifically, these Sections identify principles supporting regulation that is limited in scope and adverse impact, yet able to "punch above its weight" by safeguarding core values and supporting sustainable innovation. This approach is analogous to the Pareto principle:[486] a small regulatory input can yield disproportionately large benefits.

From a values-driven perspective, procedural requirements are critical. Such requirements—among the strongest in the AI Act—make the most structural contributions to traditional legal frameworks. Importantly, their compliance burden is comparatively low. Measures such as requiring developers to document AI performance, benchmarking, and improvement efforts are highly valuable and impose minimal cost relative to their regulatory impact.

Documentation and more elaborate capability and risk monitoring is also essential for governing prospective, more advanced AI systems approaching artificial general intelligence, which will require heightened scrutiny, and ensuring sustainable AI ethics and safety. In this respect, the AI Act and the

---

484. *Supra* note 188.

485. *See supra* Section I.B.3.a.

486. Also known as the 80-20 rule. *See generally* VILFREDO PARETO, COURS D'ÉCONOMIE POLITIQUE (1896) (describing the empirical observation that a small minority owns a large share of wealth); F. John Reh, *Understanding Pareto's Principle—The 80-20 Rule*, THE BALANCE, https://oregonwomenlawyers.org/wp-content/uploads/2018/05/The-80-20-Rule-Paretos-Principle-copy.pdf [https://perma.cc/P7FM-NSBV].

similar Californian SB 53 provisions on general-purpose or foundation models strike a stronger balance,[487] largely imposing only necessary regulatory burdens while minimizing negative effects on innovation. Similarly, the New York RAISE Act captures this balance by requiring developers to retain testing records and inform authorities about potential risks.[488]

This analysis also suggests regulators should avoid adopting values-driven frameworks that supplant, rather than supplement, existing regimes. Such frameworks risk undermining the objectives of traditional legal systems without providing adequate alternatives. Conversely, failing to act is not a neutral stance, but an active choice that weakens existing protections.

Superimposing new regulations onto existing frameworks—as with the AI Act—creates a complex patchwork of obligations that complicates compliance and fails to deliver legal certainty. Similarly, actively preempting regulation can foster uncertainty through common law developments. Moreover, American regulators cannot ignore the European AI Act's potential impact on U.S. companies seeking to operate in Europe, as it may offset the marginal burden of parallel domestic requirements.

Therefore, regulatory inaction is also undesirable, as it perpetuates the legal uncertainty that characterizes AI development today. The preferable solution is to anchor any new regime to existing frameworks. California's SB 813[489] is instructive in this respect, tying substantive obligations to established legal structures. Coupled with procedural requirements for developers and deployers, this approach protects key values while minimizing costs to innovation. Nonetheless, such initiatives must be carefully calibrated to avoid undermining value protection in pursuit of innovation and to ensure regulatory sustainability.

These considerations suggest that optimal AI regulation favors "less" rather than "more." Regulators should avoid sweeping regimes and instead provide targeted clarity and support where traditional frameworks are insufficient. Generally, procedural requirements are preferable to substantive ones, as the latter more directly constrain the types and modalities of AI that can be developed—thus having a greater adverse

---

487. *See* Regulation 2024/1689, *supra* note 3, Chapter V; S.B. 53, 2025–2026 Leg., Reg. Sess. (Cal. 2025).

488. S. 6953B, 2025 Leg., Reg. Sess. (N.Y. 2025).

489. S.B. 813, 2025–2026 Leg., Reg. Sess. (Cal. 2025).

effect on innovation. Procedural requirements, by contrast, typically impose a small compliance burden while delivering significant regulatory benefit.

V. CONCLUSION

This Article has examined two central dichotomies in AI regulation: the divide between contemporary and prospective AI, and the perceived trade-off between safeguarding AI safety and values versus promoting innovation. Focusing primarily on the latter, it has demonstrated that this dichotomy is unduly reductive. Neither a regulatory moratorium nor the values-driven European AI Act effectively advances either goal. Instead, a balanced approach—integrating elements of both—offers greater promise. By analyzing the impact of both regimes on liability and non-discrimination law, this Article argues that effective regulation must be narrowly tailored to address existing gaps in the application of legal frameworks to AI. Such targeting enables high regulatory impact with minimal burden.

By identifying a Pareto principle for AI regulation,[490] this Article argues that the optimal approach is often to emphasize low-cost, high-impact procedural safeguards that empower victims of AI harm and regulators, complemented by narrowly tailored substantive standards that reinforce traditional legal norms. Such a model would enhance legal certainty for compliant actors—something lacking under a moratorium and potentially undermined by complex regulatory overlays as found in parts of the AI Act. Procedural safeguards—such as documentation and information sharing—are also essential for enabling regulators to respond to future advances in AI. More substantive changes should be narrowly tailored and grounded in a clear analysis of the existing framework's limits.

Ultimately, while this approach may seem modest in ambition, it is bolder in execution: it confines demanding AI regulation to domains where it can achieve concrete, meaningful results in advancing AI safety and ethics while sustaining innovation.

---

490. *See supra* Section IV.