THE COLUMBIA
SCIENCE & TECHNOLOGY
LAW REVIEW

VOLUME 27 STLR.ORG NUMBER 2

ARTICLE

THE ROLE OF STANDARDS IN ENABLING THE AI STACK

Christopher S. Yoo*

The debate over AI standards is undergoing a significant reorientation. In major jurisdictions, such as the United States, European Union, United Kingdom, and Australia, policymakers are increasingly shifting from a predominantly harm-centered approach toward one that emphasizes productivity, innovation, and the beneficial uses of AI while still managing risk. At the same time, AI is increasingly being developed and deployed through a multi-party stack composed of data providers, foundation model developers, fine tuners, deployers, and other actors rather than by a single integrated firm. In this environment, standards are essential not merely as a safety tool but as enabling infrastructure: they clarify roles, define interfaces, support validation, facilitate interoperability, and create the basis for ex post evaluation of real-world performance. The Essay further contends that standards should generally emerge through flexible, multistakeholder processes and will likely vary across industry verticals rather than take the form of a single universal rule set. It explores five core components of effective AI standards: substantive performance requirements, meaningful data disclosures, transparent validation methods, protection against attacks, and articulation of acceptable levels of risk measured against real-world counterfactuals rather than zero-risk baselines. Properly designed, such standards can unlock AI's economic and social value while also mitigating bias, error, and safety concerns.

* Imasogie Professor in Law & Technology, Professor of Communications, Professor of Computer & Information Science, and Founding Director of the Center for Technology, Innovation & Competition, University of Pennsylvania.

I. INTRODUCTION

The past year has borne witness to two important changes in the discourse surrounding AI standards. First, major jurisdictions are increasingly shifting away from focusing primarily on the risks that AI poses in favor of one that emphasizes how to promote beneficial uses of AI while managing risks.[1] Second, it is becoming increasingly clear that many, if not most, AI systems will be deployed by multiple companies operating in cooperation rather than a single company that controls the entire system.[2]

This Essay examines both of these developments. It then explores the essential elements of an AI standard that would empower participants in the AI stack to deploy the technology in a way that can help it realize its beneficial potential.

II. THE SHIFT IN FOCUS TOWARD VALUE CREATION

While early efforts at formulating standards, such as those issued by the EU, OECD, and various AI summits, focused primarily on the potential harms that AI could cause,[3] more recent statements have paid less attention to harms and

---

[1] *See infra* Part II.

[2] *See infra* Part III.

[3] *See, e.g.*, European Declaration on Digital Rights and Principles for the Digital Decade, 2023 O.J. (C 23) 1, ¶¶ 9(b), 10, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32023C0123(01) [https://perma.cc/GWC3-XBWC]; Org. for Econ. Coop. & Dev. [OECD], *Recommendation of the Council on Artificial Intelligence*, § 1.3, OECD/LEGAL/0449 (May 21, 2019), https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 [https://perma.cc/J7UP-PKH7]; DEP'T FOR SCI., INNOVATION & TECH., THE BLETCHLEY DECLARATION BY COUNTRIES ATTENDING THE AI SAFETY SUMMIT 1-2 (Nov. 1, 2023),

concentrated instead on how best to facilitate the realization of the potential benefits that AI can create. This shift is evident in major jurisdictions around the world, including the U.S., EU, UK, and Australia.

A. United States

Regarding the U.S., Vice President J.D. Vance's speech before the Paris AI Action Summit cautioned that "[t]he AI future will not be won by hand-wringing about safety" and emphasized the need to avoid excessive regulation.[4] More concretely, the Trump Administration returned to the approach taken by his first Administration[5] by replacing the Biden Administration's harm-focused Executive Order on AI with one more oriented on removing regulatory barriers to AI innovation and calling for the development of an AI action plan.[6] The resulting AI Action Plan, issued in July 2025, stressed accelerating AI innovation, building AI infrastructure, and promoting U.S. leadership in global standards.[7]

B. European Union

In the European Union, the September 2024 report on competitiveness authored by former Italian Prime Minister and former European Central Bank President Mario Draghi warned that the fact that Europe is confronting an "existential challenge."[8] In particular, the EU's focus on ex-ante regulation was causing European productivity levels to lag far behind those of the U.S., with one of his principal examples being the EU's new AI Act.[9]

---

https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023 [https://perma.cc/W6CU-TBYA].

[4] J.D. Vance, *Remarks by the Vice President at the Artificial Intelligence Action Summit in Paris, France*, AM. PRESIDENCY PROJECT (Feb. 11, 2025), https://www.presidency.ucsb.edu/documents/remarks-the-vice-president-the-artificial-intelligence-action-summit-paris-france [https://perma.cc/L8ZA-2U7A].

[5] Christopher S. Yoo & Alex Mueller, *U.S. Artificial Intelligence Regulation During the Biden Administration*, 17 J.L. & ECON. REGUL. 7 (2024), https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artiId=ART003146343 [https://perma.cc/8S6Q-NC6S].

[6] Exec. Order No. 14179, 90 Fed. Reg. 8741 (Jan. 23, 2025); *see also* Exec. Order No. 14110, 88 Fed Reg. 75,191 (Nov. 1, 2023) [hereinafter Biden Administration AI Executive Order].

[7] THE WHITE HOUSE, WINNING THE RACE: AMERICA'S AI ACTION PLAN 1, 3, 14, 20 (July 2025), https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf [https://perma.cc/KT82-WTK2].

[8] MARIO DRAGHI, THE FUTURE OF EUROPEAN COMPETITIVENESS: PART A – A COMPETITIVENESS STRATEGY FOR EUROPE 5 (Publ'ns Off. of the Eur. Union Sept. 2024). https://commission.europa.eu/document/download/97e481fd-2dc3-412d-be4c-f152a8232961_en [https://perma.cc/C4YU-2EWB].

[9] *Id.* at 30.

Ursula von der Leyen's presidency has embraced the Draghi Report as the blueprint for her second term as EU President.[10] In fact, von der Leyen created a task force to implement the report's recommendations[11] and organized a major conference to mark the one-year anniversary of the report's release in which she reviewed the progress in fulfilling the report's proposals and gave Draghi a platform to reiterate his call to action.[12]

The Digital Omnibus package proposed revisions intended to make the AI rules more innovation-friendly, including greater flexibility to rely on the legitimate interest legal basis for processing personal data for AI training, postponement of the AI Act's application for high-risk systems, and the relaxation of a number of other requirements imposed by GDPR and the AI Act.[13]

C. United Kingdom

In the UK, the Labour government fired the head of its competition enforcement agency in January 2025 for being insufficiently focused on growth.[14] The Treasury has similarly criticized risk-averse regulators who, out of fear of criticism, turn to rules-based frameworks that immunize decisionmakers from criticism by limiting their discretion but prevent regulators from striking the right balance between risk and growth.[15] The Chancellor has also emphasized the need to stop regulating just

---

[10] EUR. COMM'N, *Draghi's Report on the Future of European Competitiveness: A Blueprint for Europe's Demographic and Regional Cohesion* (Oct. 23, 2024), https://ec.europa.eu/regional_policy/whats-new/newsroom/23-10-2024-draghi-s-report-on-the-future-of-european-competitiveness-a-blueprint-for-europe-s-demographic-and-regional-cohesion_en [https://perma.cc/336E-UYY9].

[11] Giovanna Faggionato, *Von der Leyen Is Creating a Task Force to Turn Draghi Report into Reality*, POLITICO (Nov. 29, 2024, 11:49 AM CET), https://www.politico.eu/article/ursula-von-der-leyen-mario-draghi-is-creating-a-task-force/ [https://perma.cc/V2JJ-UKWR].

[12] EUR. COMM'N, *The Draghi Report: One Year On* (Sept. 16, 2025), https://commission.europa.eu/topics/competitiveness/draghi-report/one-year-after_en [https://perma.cc/9C8Y-9VUH].

[13] *Proposal for a Regulation of the European Parliament and of the Council Amending Regulations (EU) 2016/679, (EU) 2018/1724, (EU) 2018/1725, (EU) 2023/2854 and Directives 2002/58/EC, (EU) 2022/2555 and (EU) 2022/2557 as Regards the Simplification of the Digital Legislative Framework, and Repealing Regulations (EU) 2018/1807, (EU) 2019/1150, (EU) 2022/868, and Directive (EU) 2019/1024 (Digital Omnibus),* COM (2025) 837 final (Nov. 19, 2025) [hereinafter Digital Omnibus], https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52025PC0837 [https://perma.cc/75EH-JGX3].

[14] Paul Sandle & Kate Holton, *UK Boots Out Antitrust Boss for Failing to Back Growth Agenda*, REUTERS (Jan. 22, 2025, 6:50 AM EST), https://www.reuters.com/world/uk/uk-boots-out-antitrust-boss-failing-back-growth-agenda-2025-01-22/ [https://perma.cc/Q3W9-BWZQ].

[15] HM TREASURY, *New Approach to Ensure Regulators and Regulation Support Growth*, UK GOV'T. (Oct. 22, 2025), https://www.gov.uk/government/publications/a-new-approach-to-ensure-regulators-and-regulation-support-growth/new-approach-to-ensure-regulators-and-regulation-support-growth-html [https://perma.cc/55W6-QSDX].

for risk and has called upon regulators instead to regulate for growth while managing risk.[16]

Consistent with this shift, the UK's Department for Science, Innovation & Technology issued its "AI Opportunities Action Plan" in January 2025. This plan laid out a pro-innovation regulatory approach to AI aimed at ensuring that the UK and its citizens benefit from economic growth and prosperity offered by AI and called upon all areas of government to support AI and take down barriers to growth.[17] On October 21, 2025, the Technology Secretary issued a new blueprint to drive AI growth and public trust that proposed the creation of sandboxes in which regulations are temporarily relaxed so that innovators can test new AI products under real-world conditions.[18]

The day following this missive, The Treasury issued a policy paper entitled, "New approach to ensure regulators and regulation support growth," warning that regulations, while essential to protect the public and promote competition, should also encourage new investment, innovation, and growth.[19] In particular, the policy paper again targets regulations that are excessively risk averse by adopting rules-based frameworks designed to avoid criticism by eliminating discretion, which has the consequence of preventing regulators from striking the right balance between risk and growth.[20] It cites the AI Opportunities Action Plan as the blueprint for promoting a pro-innovation regulatory approach that ensures that "the defining opportunity of our generation" is not held back by ineffective regulation.[21]

D. Australia

In Australia, the newly elected Labour government has made productivity its primary priority and criticized excessive regulation for hamstringing the Australian economy.[22] Regarding AI specifically, in June, leaders have advocated for light-

---

[16] Liz Hoffman & Andrew Edgecliffe-Johnson, *UK Chancellor Rachel Reeves: The Regulatory Focus Is Now on "Growth, Not Just Risk,"* SEMAFOR (Jan. 24, 2025, 7:53 AM EST), https://www.semafor.com/article/01/24/2025/uk-chancellor-rachel-reeves-the-regulatory-focus-is-now-on-growth-not-just-risk [https://perma.cc/MH4Q-EDFU].

[17] UK DEP'T FOR SCI., INNOVATION & TECH., AI OPPORTUNITIES ACTION PLAN (2025), https://assets.publishing.service.gov.uk/media/67851771f0528401055d2329/ai_opportunities_action_plan.pdf [https://perma.cc/6L9Q-QS4J].

[18] Press Release, UK Dep't for Sci., Innovation & Tech. & Rt. Hon. Liz Kendall, New Blueprint for AI Regulation Could Speed Up Planning Approvals, Slash NHS Waiting Times, and Drive Growth and Public Trust (Oct. 21, 2025), https://www.gov.uk/government/news/new-blueprint-for-ai-regulation-could-speed-up-planning-approvals-slash-nhs-waiting-times-and-drive-growth-and-public-trust [https://perma.cc/WXG7-S775].

[19] HM TREASURY, *supra* note 15, § 1.1.

[20] *Id.* § 1.

[21] *Id.* § 4.1.

[22] *See. e.g.*, Hon. Dr. Jim Chalmers, Treasurer, Address to the National Press Club, Canberra, on Economic reform in our second term 6, 9-10 (June 18, 2025), https://ministers.treasury.gov.au/ministers/jim-chalmers-2022/speeches/address-national-press-club-canberra-5 [https://perma.cc/SF3S-DPM4] (identifying productivity as the government's

touch regulation that "focus[es] . . . on how the technology can boost productivity, rather than establishing guardrails for its use"[23] and that "protect[s] the public while allowing productivity-enhancing innovation to flourish."[24]

In October 2024, the high-profile Productivity Commission submitted views opposing the Government's proposal to introduce mandatory guardrails for AI in high-risk settings.[25] The Commission's research indicated that AI has significant potential to increase productivity and that regulation should enable AI adoption rather than stifle it.[26] Furthermore, the research further found that "AI regulation that offers the best chance of improved productivity while managing risks" will "focus on the *net benefit* of regulation," "be proportionate, effective and risk based," "regulate *outcomes* where possible, rather than using technology-specific approaches that can quickly become outdated," and "compare the risks of AI use to *real-world counterfactuals* – not necessarily aim for zero risk."[27] In particular, "pre-emptive regulations based on typical uses and harms are likely to be ineffective (as harms are unknown) or overly restrictive (costs may outweigh benefits)."[28]

The Productivity Commission issued an Interim Report on the pillar on "Harnessing data and digital technology" in August 2025, finding that AI "could underpin a new wave of productivity growth" and that "[p]oorly designed

---

"primary focus" and criticizing excessive regulation for blowing out approval times and stifling economic dynamism); Interview by Deborah Knight with Hon. Dr. Andrew Leigh, Ass't Min. for Productivity, Competition, Charities & Treasury (June 16, 2025), https://ministers.treasury.gov.au/ministers/andrew-leigh-2025/transcripts/interview-deborah-knight-money-news-2gb [https://perma.cc/7KRP-2MR9] (describing accumulated regulations as "a thicket of regulation" that has caused approval times to be "blown out" and urging reform to improve Australia's historically weak productivity growth).

[23] Michael Read & Phillip Coorey, *Chalmers pushes back on union demands to regulate AI at work*, FIN. REV. (Austl.) (June 14, 2025, 5:00 AM), https://www.afr.com/policy/economy/chalmers-pushes-back-on-union-demands-to-regulate-ai-at-work-20250613-p5m75k [https://perma.cc/6Q58-JCQW].

[24] Hon. Andrew Leigh, Ass't Min. for Productivity, Competition, Charities & Treasury, The Progressive Productivity Agenda, Speech before the McKell Inst. (June 25, 2025), https://www.andrewleigh.com/speech_the_progressive_productivity_agenda [https://perma.cc/T8UW-ZB5R].

[25] AUSTL. GOV'T, PRODUCTIVITY COMM'N, REGULATING AI IN HIGH-RISK SETTINGS: PRODUCTIVITY COMMISSION SUBMISSION 3 (Oct. 2024) [hereinafter PRODUCTIVITY COMM'N SUBMISSION], https://assets.pc.gov.au/research/supporting/regulating-ai/regulating-ai.pdf [https://perma.cc/W8T3-BSP5]. For the government proposal to which this responded, see AUSTL. GOV'T, DEP'T OF INDUS., SCI. & RES., SAFE AND RESPONSIBLE AI IN AUSTRALIA: PROPOSAL PAPER FOR INTRODUCING MANDATORY GUARDRAILS FOR AI IN HIGH-RISK SETTINGS 16 (Sept. 2024) [hereinafter GUARDRAILS PROPOSAL], https://consult.industry.gov.au/ai-mandatory-guardrails [https://perma.cc/Y6HS-DWKY].

[26] PRODUCTIVITY COMM'N SUBMISSION, *supra* note 25, at 3.

[27] *Id.*

[28] *Id.* at 5.

regulations could stifle AI investment without improving outcomes."[29] It further notes how the EU's General Data Protection Regulation "appears to have stifled innovation and investment" and the lack of signs that any other countries will follow the approach taken by the EU's AI Act.[30] Accordingly, the Interim Report included draft recommendations that "[p]roductivity growth from AI will be built on existing legal foundations" and that "[a]ny regulatory responses to potential harms from using AI must be proportionate, risk-based, outcomes-based and technology-neutral where possible"[31] In addition, "AI-specific regulation should be a last resort," and the government should "pause steps to implement mandatory guardrails for high risk AI" until it has analyzed existing law for potential gaps based on the understanding that "[n]ew, untested, guardrails are more likely to raise uncertainty than lower it."[32]

On December 2, 2025, the Australian Government issued a "National AI Action Plan" committed to ensuring that "Australians . . . share in the benefits of AI while remaining protected in a fast-changing world."[33] The plan is anchored in three policy objectives: "capturing the opportunity" so that Australians enjoy the economic benefits offered by AI," "spreading the benefits" to ensure that the economic gains are equitably distributed," and "keeping Australians safe" – mitigating the risk of harms from AI use.[34]

Eight days later, the Productivity Commission issued its Inquiry Report on the same pillar as the Interim Report discussed above emphasizing that AI has the potential to "transform the global economy and speed up productivity growth" and that "Australia needs to harness the consumer and productivity of benefits of data and digital technology while managing and mitigating any downside risks."[35] The Inquiry Report formally adopted the recommendations of the Interim Report.[36] It noted that the National AI Action Plan represented a move away from the Government's previous support for whole-of-economy, technology-specific AI regulation, particularly, mandatory guardrails on high-risk AI, in favor of relying on existing legal frameworks and regarding AI-specific regulation as a last resort.[37] The webpage hosting the Government's October 2024 proposal now notes, "The

---

[29] AUSTL. GOV'T, PRODUCTIVITY COMM'N, HARNESSING DATA AND DIGITAL TECHNOLOGY: INTERIM REPORT 1, 9 (Aug. 2025), https://assets.pc.gov.au/2025-09/data-digital-interim.pdf [https://perma.cc/44XT-58V2].

[30] *Id.* at 16, 22-23, 66.

[31] *Id.* at 2, 19.

[32] *Id.* at 2-3, 9, 18, 22.

[33] *See* AUST'L GOV'T, DEP'T OF INDUS., SCI. & RES., NATIONAL AI ACTION PLAN 5 (2025), https://www.industry.gov.au/sites/default/files/2025-12/national-ai-plan.pdf [https://perma.cc/3JV3-7FSQ].

[34] *Id.* at 7.

[35] AUSTL. GOV'T, PRODUCTIVITY COMM'N, HARNESSING DATA AND DIGITAL TECHNOLOGY: INQUIRY REPORT 1 (Dec. 10, 2025), https://assets.pc.gov.au/2025-12/data-digital_0.pdf [https://perma.cc/A2FB-FMJP].

[36] *See id.* at 2, 20-30.

[37] *Id.* at 25-26.

Australian Government will not proceed at this time with previous proposals to introduce mandatory guardrails for AI development and deployment."[38]

* * *

The emergence of a greater emphasis on AI's potential to promote productivity and growth represents a fundamental shift away from approaches that place primary emphasis on the risks that AI poses. It is thus a far cry from the approach taken by the EU AI Act that focuses on product safety without taking into account the economic interests of consumers.[39]

III. THE ROLE OF STANDARDS IN UNLOCKING THE VALUE OF THE AI STACK

Together these developments suggest a fundamental reorientation of the discourse around AI standards. The new focus is on how to configure standards to enable the vertical stack needed to support the benefits of AI. The process of enabling functionality should simultaneously have the side effect of reducing many of the risks that were the focus of prior standardization efforts.

What would a governance approach focused on unlocking the value of AI look like? There is a growing recognition that AI systems are not developed by a single actor. Instead, they are typically the product of multiple contributors that span data providers, foundation model developers, fine tuners, deployers, and others.[40]

The literature on modularity and software engineering lays out the requirements needed to enable a system in which multiple actors each provide different components to function properly, all of which are well served by technical standards, the approach that now governs most modern technology.[41] As an initial matter, standards clarify expectations about which link in the chain of production will handle particular functions.[42] Moreover, standards can provide clear guidance about data interfaces and which interdependencies the system will take into account.[43] Equally importantly, it can set clear expectations about how each link will handle different situations so that others can correctly interpret the signals that

---

[38] GUARDRAILS PROPOSAL, *supra* note 25.

[39] *See, e.g.*, Martin Ebers, *EU Consumer Law and the AI Act*, *in* THE CAMBRIDGE HANDBOOK OF AI AND CONSUMER LAW: COMPARATIVE PERSPECTIVES 215, 234 (Larry A. DiMatteo et al. eds., 2024).

[40] *See e.g.*, Regulation 2024/1689, 2024 O.J. (L 1689) 1, 62, 66, 67-69 [hereinafter EU AI Act] (imposing distinct obligations on providers, distributors, and deployers).

[41] *See* CARLISS Y. BALDWIN & KIM B. CLARK, DESIGN RULES, VOLUME 1: THE POWER OF MODULARITY 73-89 (2000); *see also* Alan MacCormack et al., *The Power of Modularity Today: 20 Years of "Design Rules,"* 32 INDUS. & CORP. CHANGE 1, 1-15 (2023) (reviewing two decades of scholarship confirming that modularity enabled by standardized interfaces and design rules, which allows multiple independent actors to contribute interoperable components to complex systems).

[42] *See* Christopher S. Yoo, *Modularity Theory and Internet Regulation*, 2016 U. ILL. L. REV. 1, 19-20.

[43] *See id.* at 10-11.

they see.[44] Any other alternative requires every AI system to be provided by a single entity.

In addition, standards can create a basis for validation. For example, the EU AI Act calls for impact assessments of high-risk AI systems, although the Digital Omnibus has proposed postponing this obligation.[45] It is hard to see how such an assessment could be conducted without some standard against which to measure these systems. More fundamentally, unless a chain of production is willing to operate on the honor system, third-party provision of different functions depends on each actor's ability to verify that the services provided by the other actors are performing as expected.[46] And any such expectation necessarily depends on standards.

The focus on different vertical stacks underscores the real possibility, if not the likelihood, that the usable standards will not represent a single horizontal standard spanning all use cases. Indeed, it would blink reality if a single standard could be devised to govern such varying use cases as autonomous vehicles, health care, employment, financial services, and criminal law, just to name a few. Instead, it is more likely that a different set of standards will govern each industry vertical. That said, there may be a core of requirements that could prove robust enough to span multiple verticals.

Finally, like any complex system, AI exhibits emergent characteristics that defy advance prediction and only appear when the system is deployed in the real world at scale. The result is the need for some form of ex post review of how the system behaves in practice.[47] Evaluation of an AI system's real-world performance after it has been deployed thus represents a key aspect of assessing any unintended consequences that AI systems create.

Standards-based approaches also have the virtue of creating markets for possible governance approaches. The voluntary nature of standard adoption allows the user community acting on a decentralized basis to decide which approach should prevail.[48] Moreover, refraining from enshrining any particular approach into law has the virtue of allowing greater flexibility for standards to evolve with changes in the technological or economic environment. Moreover, standard-setting processes are often better situated to tap into technical expertise, are more agile, can allow for multistakeholder participation, and offer an easier path to global

---

[44] *See* Christopher S. Yoo, *Protocol Layering and Internet Policy*, 161 U. PA. L. REV. 1707, 1716-17 (2013).

[45] *See* EU AI Act, *supra* note 40, art. 27.

[46] *See* Yoo, *supra* note 42, at 17-18.

[47] *See* Christopher S. Yoo, *Beyond Algorithmic Disclosure for AI*, 25 COLUM. SCI. & TECH. L. REV. 314, 327-30 (2024), https://journals.library.columbia.edu/index.php/stlr/article/view/12766 [https://perma.cc/4RE9-782P].

[48] *See* Alexander R. Mueller & Christopher S. Yoo, *Taking Standards Seriously: The Case for a Private Standards-Based Approach to AI Governance*, 40 BERKELEY TECH. L.J. 1273, 1286-88 (2026).

adoption than would national laws.[49] The standards developed by private organizations have generally been superior to those set by governments.[50]

Interestingly, focusing on enabling the AI value chain should simultaneously reduce AI bias and enhance AI safety. Downstream actors need assurances about the reliability of the products on which they are building and need to understand their attributes if they are to ensure that their fine-tuned models are designed appropriately to meet their customers' expectations. Indeed, none of them have any interest in deploying systems that perform poorly. Moreover, the fact that the actor providing the final product bears the most liability for the AI service providers gives it high-powered incentives not to inject inadequately designed or biased products into the world.

This is not to say that standards are a panacea. Their voluntary nature can give rise to complications in adoption. Standard setting is potentially susceptible to races to the bottom, industry capture and being dragged into geopolitical trade wars.[51] That said, the same criticisms apply to government-led solutions. The choice between the two modalities is a matter of the comparative second-best.

IV. COMPONENTS OF AI STANDARDS

What components must standards include if they are to serve these salutary purposes? This essay offers an initial framework focusing on five components: (1) substantive performance requirements, (2) data disclosure, (3) validation methods, (4) protection against attacks, and (5) identification of acceptable levels of risk.

A. Substantive Performance Requirements

Standards must contain substantive requirements that AI systems must meet. As noted earlier, the diversity of businesses using AI means that each industry may be governed by different standards. Indeed, even within one industry, the standard may not yield a single overarching standard and may consist of multiple sub-standards.

Consider, for example, the Standard for Assumptions in Safety-Related Models for Automated Driving Systems issued by IEEE. Rather than offering a single comprehensive framework for autonomous vehicle safety, it instead focuses on seven commonly occurring scenarios.[52] In addition, it distinguishes attributes that

---

[49] *Id.* at 1295-96.

[50] *See, e.g.,* STANLEY BESEN & LELAND JOHNSON, COMPATIBILITY STANDARDS, COMPETITION, AND INNOVATION IN THE BROADCASTING INDUSTRY 135 (Rand Corporation, 1986), https://www.rand.org/content/dam/rand/pubs/reports/2007/R3453.pdf [https://perma.cc/4AW7-RYYQ]; JEFFREY H. ROHLFS, BANDWAGON EFFECTS IN HIGH-TECHNOLOGY INDUSTRIES 201 (MIT Press, 2001).

[51] Mueller & Yoo, *supra* note 48, at 1291-95, 1297-1302, 1309-11.

[52] INST. OF ELEC. & ELECS. ENG'R, *IEEE Standard for Assumptions in Safety-Related Models for Automated Driving Sys*tems, IEEE 2846-2022, 22-37 (Apr. 22, 2022) [hereinafter IEEE Standard], https://standards.ieee.org/ieee/2846/10831/ [https://perma.cc/Z3MB-6FFK].

can be identified in advance from those that can be observed only after a technology has been deployed at scale.[53] Specifically, it lists twenty-three attributes that are verifiable via inspection before separately enumerating six attributes that are demonstrable only via real-world validation.[54]

B. Data Disclosure

In addition to substantive behaviors, AI standards should include disclosures about the data on which the model was trained. Although many regimes have experimented with model cards, these typically do not provide enough information to enable downstream users to assess the models on which they are building.

For example, downstream users need information about the *sources* of the data on which the upstream provider trained its model.[55] This should include not only the amount of data but also key factors including the accuracy and recency of the data, the importance of which will vary with the uses to which it is put.[56] The sources of data also play a key role in determining whether the resulting model complies with relevant data protection and copyright laws.[57]

In addition, downstream users need information about the *scope* of the data on which the upstream provider trained its data.[58] AI models are much more effective at interpolation rather than extrapolation.[59] Knowing the range of each parameter of data on which a model was trained can play a key role in determining when it is likely to be accurate and when it is likely to hallucinate.[60] Moreover, some information about correlations among data can also help unravel potential problems.[61] That said, the fact that modern AI models are often trained on over one million parameters makes complete disclosure of this information problematic.

---

[53] *Id.* at 38-43.

[54] *Id.*

[55] Yoo, *supra* note 47, at 320-21.

[56] *See* Christopher Yoo, Michael Buchwald & Calvin Ketcham, *Big Data and Competition Law: Lessons for Innovation Markets*, 87 ANTITRUST L.J. 241, 249-53 (2025).

[57] Yoo, *supra* note 47, at 321.

[58] Timnit Gebru et al., *Datasheets for Datasets*, 64 COMMC'NS ACM 86, 86-87 (2021), https://www.microsoft.com/en-us/research/wp-content/uploads/2019/01/1803.09010.pdf [https://perma.cc/N3D3-5XNS] (arguing that documentation is necessary for downstream consumers to determine whether a dataset is appropriate for their specific needs).

[59] *See* Elan Rosenfeld, Pradeep Ravikumar & Andrej Risteski, *An Online Learning Approach to Interpolation and Extrapolation in Domain Generalization*, 151 PROC. MACH. LEARNING RSCH. 2641, 2641-42 (2022), https://proceedings.mlr.press/v151/rosenfeld22a.html [https://perma.cc/LF5E-7C9Q].

[60] *Id.* at 318-19.

[61] *See* Margaret Mitchell et al., *Model Cards for Model Reporting,* PROC. 2019 CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 220, 220-21 (2019), https://dl.acm.org/doi/pdf/10.1145/3287560.3287596 [https://perma.cc/8TLS-93VY] (proposing standardized reporting to help downstream users understand model limitations and intersectional characteristics that may reveal biases).

What may be necessary is some assessment of which data are the most critical for the particular uses to which the model is being put.

Efforts to limit bias may also require information about correlations among different features inferred from the data used to construct the model.[62] Even when consideration of certain dimensions is forbidden, models can construct proxies for those dimensions from other features that are highly correlated with them, which requires an understanding of the correlations among the forbidden dimensions and other features represented in the dataset.[63] Given that frontier models are often trained on datasets consisting of tens of thousands of features, any disclosure obligations must strike the appropriate balance between providing the information users with the information about correlations among the data that they need to assess its performance properly and minimizing the burdens place on model developers.

Furthermore, downstream users need to know the extent to which the upstream provider trained the model on *synthetic data*. Synthetic data is often held out as a panacea that can solve limitations in the amount of data.[64] The technical literature reveals that while synthetic data can exploit the Central Limit Theorem and provide an increase in the number of observations that allows the distribution to tighten more quickly, the fact that the newly constructed observations are not independent means that it does not introduce instances of low probability events in the way that a true data draw would.[65] The result is that although the use of synthetic data initially improves models, at some point it begins to degrade them before eventually leading to model collapse.[66] Understanding how much synthetic data were used to train a model is thus essential to its proper deployment and to guide when the use of synthetic data in fine tuning is appropriate.

Knowledge of the data on which a model is trained is also essential to evaluate the stationarity assumption, which is "the idea that natural systems fluctuate within an unchanging envelope of variability."[67] All predictive analytics are based on the premise that the data based on past events provide a useful basis for forecasting the

---

[62] Yoo, *supra* note 47, at 327.

[63] *Id.*

[64] *See* James Jordon et al., *Synthetic Data - What, Why and How?*, ARXIV 1 (May 6, 2022), https://arxiv.org/pdf/2205.03257 [http://perma.cc/C7F2-KLDR] ("Outliers and low probability events, as are often found in real data, are particularly difficult to capture and include in a synthetic dataset in a private way.").

[65] Elvis Dohmatob et al., *A Tale of Tails: Model Collapse as a Change of Scaling Laws*, ICML'24: PROC. 41ST INT'L CONF. ON MACHINE LEARNING 11165, 11165-67 (2024).

[66] Mohammed El Amine Seddik et al., *How Bad is Training on Synthetic Data? A Statistical Analysis of Language Model Collapse*, PROC. CONF. ON LANGUAGE MODELING 2 (2024), https://openreview.net/pdf/4b55130ce33429ab600e5cde927dfe10a36328ba.pdf [https://perma.cc/KX4D-W7K6]; Ilia Shumailov et al, *The Curse of Recursion: Training on Generated Data Makes Models Forget*, ARXIV 1 (2023), https://arxiv.org/pdf/2305.17493 [https://perma.cc/C6MZ-DJYT].

[67] P.C.D. Milly et al., *Stationarity Is Dead: Whither Water Management?*, 319 SCI. 573, 573 (2008).

future. However, circumstances exist where structural changes may make past performance less predictive of future results. For example, climate change has undercut the ability to build conservation models on prior observations.[68] Similarly, the unanticipated changes in the real estate market during the COVID pandemic caused Zillow to terminate its algorithm-based program to buy undervalued houses after it lost over $1 billion.[69] Any standard should mandate disclosure of sufficient information to permit an assessment of when stationarity is unlikely to hold true.

Law has begun to move in these directions. For example, Article 13 of the EU AI Act requires providers of high-risk AI to make a limited number of disclosures discussed here.[70] Furthermore, California's new Generative AI Training Data Transparency Act, which went into effect on January 1, 2026, requires model developers to disclose, among other things, information about the source, amount, and contents of the training data as well as the role that synthetic data played in its development and implementation.[71] These requirements need to be operationalized and examined both for completeness and for whether the burdens they impose are justified by commensurate benefits. xAI has challenged the law, arguing that the mandatory disclosure of trade secrets runs afoul of the Fifth Amendment and compels speech in violation of the First Amendment.[72]

C. Validation Methods

Another key element that downstream users need to know is the validation methods employed by upstream providers. All validation methods involving real-world phenomena are necessarily incomplete, as they can test only for a finite number of scenarios while the independent actions of multiple actors in an environment subject to extensive variability present a never-ending series of permutations.[73] Understanding the particular testing regime used helps identify strengths and weaknesses.

For example, the IEEE Standard on Autonomous Vehicle Safety discussed above identifies seven forms of validation: (1) systematic process, (2) safety-by-design architecture, (3) formal methods, (4) robustness, (5) simulation, (6) closed course real-world testing, and (7) public real-world testing.[74] Understanding how each of these approaches have been applied to a system is critical to predicting its performance.

---

[68] *Id.*; Robin K. Craig, *Stationarity is Dead – Long Live Transformation: Five Principles for Climate Change Adaptation Law*, 34 HARV. ENV'T L. REV. 9, 15-16 (2010); J.B. Ruhl, *Climate Change Adaptation and the Structural Transformation of Environmental Law*, 40 ENV'T. L. 363, 394 (2010).

[69] Yoo, *supra* note 42, at 320.

[70] EU AI Act, *supra* note 40, art. 13(3)(b).

[71] CAL. CIV. CODE § 3111(a).

[72] Complaint, X.AI LLC v. Bonta, No. 2:25-cv-12295 (C.D. Cal. filed Dec. 29, 2025).

[73] Yoo, *supra* note 42, at 324.

[74] IEEE Standard, *supra* note 52, at 45-48.

This example suggests that AI models should undergo multiple different types of testing—some theoretical, some process-based, some based on initial empirical evaluations in the lab, and some based on real-world evaluations in public. In addition, upstream providers must disclose the nature of this testing to all channel partners if they are to interpret the results of that testing and to understand what it reveals about the models' capabilities and limits. Model users also need information about the testing regime to determine whether it suffers from the pitfalls of overfitting or reward hacking.[75] As noted above, all models must undergo ongoing evaluation of actual deployments at scale to guard against the possibility that unanticipated behaviors may emerge.[76]

Some disclosure requirements have made their way into law. Article 55 of the EU AI Act mandates that providers of general-purpose AI models with systemic risk evaluate their models.[77] California's new Transparency in Frontier Artificial Intelligence Act requires providers to share summaries of assessments of catastrophic risks.[78] Future efforts will add substance to these requirements and include areas mentioned in this article that are not currently addressed.

D. Protection Against Attacks

System designers often optimistically presume that people will use their system in a manner consistent with its intended purpose. The reality is that AI systems will inevitably be subject to attack. These attacks can occur during the training phase through activities intended to poison the data on which the model is being trained. It can also occur during production, such as by feeding the model prompts designed to make the model misbehave, to reveal its features, or to extract that data on which it is trained.

Information about the steps taken to protect against these attacks is thus critical to downstream providers.[79] At the same time, disclosure of these steps can simply provide a roadmap to bad actors about how best to fashion their attacks.

E. Acceptable Levels of Risk

Lastly, any standard must lay out what the upstream provider regards as acceptable levels of risk. AI systems are probabilistic by their nature. Even a model that is 99.999% accurate will err on average once every one hundred thousand times, which at scale is a certainty. The frequency can be reduced but not be completely eliminated by increasing the level of accuracy, and the increase in fidelity comes at the cost of innovation.

---

[75] Yoo, *supra* note 42, at 325-26.

[76] *Id.* at 327.

[77] EU AI Act, *supra* note 40, art. 55(1)(a).

[78] CAL. BUS. & PROF. CODE § 22757.12(c)(2).

[79] *Id.* at 327-30.

This point is illustrated by issues of security. A designer could spend its entire budget on security, which would in turn cause two problems: (1) It would have no product, and (2) the resulting system would not be 100% secure. With respect to security, those using a system must understand that the key decision is what level of risk they can accept. Upstream providers must supply the information needed for downstream uses to make this assessment.

The Australian Productivity Commission summarized the need to evaluate acceptable levels of risk:

> If AI use cases are compared to a situation of zero risk, many low-risk use cases will be unintentionally captured under the regulatory regime. It is misleading to measure the risk from a use of AI relative to a fictitious 'perfect world'. Rather, the appropriate benchmark for risk-based regulation is the expected harm from the use of the AI technology relative to the real-world counterfactual level of expected harm that would arise if the technology in question was not used. For example, the risk of a self-driving vehicle algorithm should be evaluated against a counterfactual of a competent, licensed human driver, rather than a fictitious world of zero road fatalities. The risk of an AI driven diagnostic tool in health needs to be judged against the alternative of not having such a tool to assist a health practitioner, rather than a false world of perfect diagnosis.[80]

Assessments of acceptable levels of risk are also necessitated by the fact that regulation is costly not only in terms of direct costs, such as compliance and enforcement, but also in terms of indirect costs that arise from how regulation influences the development of the technology and changes behavior. Such costs can render an innovation that would otherwise have cleared the hurdle rate of profitability for a firm economically nonviable.[81]

V. CONCLUSION

The global sea change in the attitudes toward AI regulation invites initiating a new dialogue about the proper approach to take. As part of this new discourse, standards offer the chance to unlock the potential of multi-party provision of AI systems. Aligning all members of the chain on the services that each link will provide and giving downstream actors the information they need to ensure their systems work properly while simultaneously giving them the information that enables them to reduce errors and minimize bias and the high-powered incentives to ensure that their systems meet their customers' expectations.

---

[80] PRODUCTIVITY COMM'N SUBMISSION, *supra* note 25, at 4.

[81] AUSTL. PRODUCTIVITY COMM'N, MAKING THE MOST OF THE AI OPPORTUNITY: THE CHALLENGES OF REGULATING AI 4 (Research Paper No. 2, 2024), https://assets.pc.gov.au/2025-10/ai-paper2-regulating.pdf [http://perma.cc/T84J-M9FQ].

The key is determining the necessary components of any such standards. This essay offers an initial step in exploring how to define these standards so that the benefits that AI promises to create can be realized.