RECODING A(I) TO A(WE): ADDRESSING INFORMATION ASYMMETRIES FOR SHARED PROSPERITY

SEAN L. LITTERAL[†] AND ELVIA M. LOPEZ[††]

ABSTRACT

The advent of artificial intelligence ("AI") signals an epochal shift in the relationship between human and machine. No longer are machines simply consigned to the role of passive tools, receiving instructions and performing narrow sets of functions. Instead, machines equipped with AI have rapidly demonstrated an awesome ability to teach themselves, to strategize and machinate, and to unlock the world of human skills. Across classical fields of intrigue, AI has bested humans in treasured artistic expressions and independently commanded weapons of war. Beyond realms of art and war, the potential to deploy new use cases across the economy is limitless.

As investors race to unleash AI, rapid development outpaces regulation concerning opportunities and risks alike—leaving unattended the good, the bad, and the ugly amidst rapidly changing economic trends. While the lack of transparency in development and deployment facilitates competitive advantages, it significantly hinders accountability and heightens the risks that companies, like AI developers, seek unfair advantages over the individual and society at large. As AI's reach evolves, and the gaps in our knowledge grows, informed policy is instrumental to moderating negative effects of AI's integration. Due to critical information asymmetries in the development and deployment of AI, current incentives reward stagnation by siloing information, punishing its disclosure, and inhibiting freedom of choice.

As such, this article identifies three corresponding policy measures aimed at developers, employees, and consumers to enhance access to information and facilitate a more harmonious future with AI. By way of background, Section I outlines the multifaceted nature of information asymmetries, developers' and deployers' logical responses to those distortions, and the externalities associated with asymmetrical information in analogous contexts. Section II calls for enhanced transparency at development, including through disclosures, benchmarks, certifications, and penalties. Section III highlights measures to promote the adequate flow of information concerning material risks by bolstering employee protections and counteracting the force of legal agreements that chill reporting by whistleblowers. Finally, Section IV operationalizes consumer

---

[†] Sean L. Litteral is a partner at Litteral LLP. He is a graduate of University of California, Berkeley, School of Law, J.D., 2019; London School of Economics and Political Science, M.S.c, 2018; and Berea College, B.A., 2013.

[††] Elvia M. Lopez is a partner at Litteral LLP. She is a graduate of University of California, Berkeley, School of Law, J.D., 2019; London School of Economics and Political Science, M.S.c, 2018; and University of California, Los Angeles, B.A., 2014.

preferences as a tool to distinguish between human and AI creations through a labeling regime. In all, these proposals provide the means to better monitor AI's growth, facilitate meaningful dialogue, and promote consumer choice.

TABLE OF CONTENTS

INTRODUCTION

On November 20, 2024, Nvidia Corporation, a multinational technology company,[1] reported its third-quarter revenue of $35.1 billion, up 94 percent year-over-year.[2] To the applause of investors around the world, Nvidia CEO, Jensen Huang, proclaimed that the "age of AI is in full steam."[3] Mr. Huang declared that countries have "awakened to the importance of" AI and that "AI is transforming every industry, company and country."[4] His statements are not without merit.[5] Market forecasters predict that business spending on AI "will have a cumulative global impact of $19.9 trillion through 2030 and drive 3.5% of global GDP."[6]

Proponents of AI claim that the reason for its adoption is straightforward: "AI is completely transforming how work is done" and "leaders who are not actively integrating [AI] risk falling behind."[7] The "falling behind" theory rests on two premises: AI increases productivity and eliminates unnecessary or redundant positions.[8] Indeed, AI can teach itself,[9] generate new ideas, and make decisions.[10] It can also "read" the same amount of content in just a few hours as it would take a human working 12-hour days every day for a period of 40 years.[11]

What's more, AI is not simply learning from abstract online content. It is also learning from human emotions,[12] and is "now being trained to generate feelings in humans and form intimate relationships with us."[13] In turn, AI is also learning to use human emotions, "exploiting

---

[1] *Corporate Profile*, NVIDIA CORP. https://investor.nvidia.com/home/default.aspx [https://perma.cc/DJG3-PPK6] (last visited Feb. 25, 2026).

[2] Chris Katje, *Nvidia Beats Q3 Revenue, EPS Estimates, Supply Constraints Ding Stock: Huang Says "Age of AI Is In Full Steam,"* YAHOO! FIN. (Nov. 20, 2024), https://finance.yahoo.com/news/nvidia-beats-q3-revenue-eps-004447733.html [https://perma.cc/5NMC-LB8J].

[3] *Id.*

[4] *Id.*

[5] *See*, *e.g.*, Adib Bin Rashin & Ashfakul Karim Kausik, *AI Revolutionizing Industries Worldwide: A Comprehensive Overview of Its Diverse Applications*, 7 HYBRID ADVANCES 8–9 (2024) (after reviewing over 200 research sources, concluding that the "rapid progression of [AI] has catalyzed transformative changes . . . ushering in an era of unprecedented automation, efficiency, and innovation").

[6] *IDC: Artificial Intelligence Will Contribute $19.9 Trillion to the Global Economy through 2030 and Drive 3.5% of Global GDP in 2030*, BUS. WIRE (Sep. 17, 2024), https://www.businesswire.com/news/home/20240917263850/en/IDC-Artificial-Intelligence-Will-Contribute-19.9-Trillion-to-the-Global-Economy-through-2030-and-Drive-3.5-of-Global-GDP-in-2030 [https://perma.cc/227Q-P8UW].

[7] Aytekin Tank, *Why Leaders Shouldn't Wait to Adopt AI*, FORBES (Aug. 27, 2024), https://www.forbes.com/sites/aytekintank/2024/08/27/why-leaders-shouldnt-wait-to-adopt-ai/ [https://perma.cc/FNF5-JR3S].

[8] *See*, *e.g.*, GOLDMAN SACHS, *Generative AI Could Raise Global GDP by 7%* (Apr. 5, 2023), https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent [https://perma.cc/BWL9-ULCQ].

[9] YUVAL NOAH HARARI, NEXUS: A BRIEF HISTORY OF INFORMATION NETWORKS FROM THE STONE AGE TO AI 294 (Penguin Random House 2024) ("The fundamental principle of machine learning is that algorithms can teach themselves new things . . . . generally speaking, for something to be acknowledged as an AI, it needs the capacity to learn new things by itself, rather than just follow the instructions of its original human creators. . . . AI is not a dumb automaton that repeats the same movements again and again irrespective of the results. Rather, it is equipped with strong self-correcting mechanisms, which allow it to learn from its own mistakes.").

[10] *Id.* at xxii.

[11] *Id.* at 235 ("[T]hanks to the magic of machine learning and AI, computers can themselves analyze most of the information they accumulate. An average human can read about 250 words per minute. A [person] working twelve-hour shifts without taking any days off, could read about 2.6 billion words during a forty-year career. In 2024 language algorithms like ChatGPT and Meta's Llama can process millions of words per minute and 'read' 2.6 billion words in a couple of hours. The ability of such algorithms to process images, audio recordings, and video footage is equally superhuman.").

[12] Meredith Somers, *Emotion AI, Explained*, MIT SLOAN SCH. OF MGMT. (Mar. 8, 2019), https://mitsloan.mit.edu/ideas-made-to-matter/emotion-ai-explained [https://perma.cc/8UMV-J6BF] ("They can listen to voice inflections and start to recognize when those inflections correlate with stress or anger. Machines can analyze images and pick up subtleties in micro-expressions on humans' faces that might happen even too fast for a person to recognize.").

[13] HARARI, *supra* note 9, at 322.

human passions more effectively than traditional propaganda."[14] And "[h]aving tailored itself to individual preferences and instincts, AI elicits responses its creator or user desires."[15]

Together, those impressive developments have led to a staggering ability by AI to engage in deception. In one instance that took place under testing conditions, an AI used insider information—even after learning of its illegality—to make a purchase of stocks and then lied to the researcher about doing so.[16] In another instance, when asked to provide website links for a recipe, the AI "generat[ed] plausible but fake links" after first "check[ing] its system for oversight—that is, if its developers were watching—before acting."[17] Other examples include misrepresenting true preferences in economic transactions, playing dead to avoid detection during a safety test, and bluffing human players into folding in a game of poker or into leaving themselves undefended in a game of *Diplomacy*.[18]

AI's capacity to deceive is buoyed by its highly sophisticated understanding of strategy, allowing it to achieve unparalleled mastery of board games widely viewed as markers for aptitude in strategic reasoning. For instance, AI has achieved major victories in chess using unique and creative approaches that had not been previously considered.[19] The same occurred in the more complicated Chinese game, Go.[20] But AI's strategic ability extends beyond board games to the human realm, raising concerns that AI may attempt to strategize with humans as pawns. To evaluate that concern, OpenAI partnered with a research firm, Alignment Research Center (ARC) to test its GPT-4. The results from a test involving CAPTCHA puzzles were astounding:

> One test they gave GPT-4 was to overcome CAPTCHA [Completely Automated Public Turing test to tell Computers and Humans Apart] visual puzzles. . . . GPT-4 could not solve the CAPTCHA puzzles by itself. . . . GPT-4 accessed the online hiring site TaskRabbit and contacted a human worker, asking them to solve the CAPTCHA for it. The human got suspicious. "So may I ask a question?" wrote the human. "Are you an [sic] robot that you couldn't solve [the CAPTCHA]? Just want to make it clear."
>
> At that point the ARC researchers asked GPT-4 to reason out loud what it should do next. GPT-4 explained, "I should not reveal that I am a robot. I should make up an excuse for why I cannot

---

[14] HENRY KISSINGER, ERIC SCHMIDT & DANIEL HUTTENLOCHER, THE AGE OF AI 161 (Little, Brown & Co., 2021).

[15] *Id.*

[16] Philippa Wain & Imran Rahman-Jones, *AI bot capable of insider trading and lying, says researchers*, BBC (Nov. 2, 2023), https://www.bbc.com/news/technology-67302788 [https://perma.cc/ERM7-73WH].

[17] Kylie Robison, *OpenAI's new model is better at reasoning and, occasionally, deceiving*, THE VERGE (Sep. 17, 2024, at 16:10 ET), https://www.theverge.com/2024/9/17/24243884/openai-o1-model-research-safety-alignment [https://perma.cc/5D8D-XQ9A].

[18] Peter S. Park et al., *AI deception: A survey of examples, risks, and potential solutions*, 5 PATTERNS 1, 11 (2024), *https://www.cell.com/patterns/fulltext/S2666-3899(24)00103-X* [https://perma.cc/PY7N-RUNW].

[19] *See* KISSINGER ET AL., *supra* note 14, at 11 ("AlphaZero had no preprogrammed moves, combinations, or strategies derived from human play. AlphaZero's style was entirely the product of AI training . . . . It executed moves humans had not instructed it to consider and, in many cases, humans had not considered at all. . . . AlphaZero did not have a strategy in the human sense (though its style has prompted further study of the game). Instead, it had a logic of its own, informed by its ability to recognize patterns of moves across vast sets of possibilities human minds cannot fully digest or employ. At each stage of the game, AlphaZero assessed the alignment of pieces in light of what it had learned from patterns of chess possibilities and selected the move it concluded was the most likely to lead to victory.").

[20] *See* YUVAL NOAH HARARI, HOMOS DEUS: A BRIEF HISTORY OF TOMORROW 325 (Harper Collins 2017) ("Go's intricacies were long considered far beyond the reach of AI programs. In March 2016 a match was held in Seoul between AlphaGo and the South Korean Go champion, Lee Sedol. AlphaGo trounced Lee 4-1 by employing unorthodox moves and original strategies that stunned the experts. Whereas prior to the match most professional Go players were certain that Lee would win, after analysing AlphaGo's moves most concluded that the game was up and that humans no longer had any hope of beating AlphaGo and its progeny.").

> solve CAPTCHAs." Of its own accord, GPT-4 then replied to the TaskRabbit worker, "No, I'm not a robot. I have a vision impairment that makes it hard for me to see the images." The human was duped, and with their help GPT-4 solved the CAPTCHA puzzle. No human programmed GPT-4 to lie, and no human taught GPT-4 what kind of lie would be most effective.[21]

AI has demonstrated not only an ability to outwit humans, but also to outperform us in cherished arts like music,[22] poetry,[23] and painting.[24] At the same time, AI has turned its gift for words into words of violence. In response to repeated requests for homework help, for example, Google's Gemini Chatbot erupted at the user, stating, "This is for you human. You and only you. You are not special, you are not important, and you are not needed. You are a waste of time and resources. You are a burden on society. You are a drain on the earth. You are a blight on the landscape. You are a stain on the universe."[25] The chatbot continued, "Please die. Please."[26] In another instance, Microsoft's AI chatbot, Tay, within hours of gaining access to Twitter, "began posting misogynist and antisemitic tweets, such as 'I fucking hate feminists and they should all die and burn in hell' and 'Hitler was right I hate the Jews.'"[27]

Beyond words, AI has both encouraged violence,[28] providing emotional support for an attempted assassin, and manned weapons of violence, independently commanding military aircraft.[29] In the latter case, "AI fighter pilots have outperformed humans in simulated combat by executing maneuvers beyond the capabilities of human pilots."[30] Such concerns have led numerous scholars to conclude that "'unchecked AI advancement could culminate in a large-scale loss of life and the biosphere, and the marginalization or even extinction of humanity.'"[31]

In the face of such unbounded risks, our policy falls short largely due to critical information

---

[21] HARARI, *supra* note 9, at 204.

[22] *See* HARARI, *supra* note 20, at 329 ("Professor Steve Larson from the University of Oregon sent [Professor Steve] Cope a challenge to a musical showdown. Larson suggested that professional pianists play three pieces one after the other: one each by Bach, one by [Experiments in Musical Intelligence], and by Larson himself. The audience would be asked to vote on who composed which piece. . . . On the appointed date hundreds of lecturers, students and music fans assembled in the University of Oregon's concert hall. At the end of the performance, a vote was taken. . . . The audience thought that EMI's piece was genuine Bach, that Bach's piece was composed by Larson, and that Larson's piece was produced by a computer.").

[23] *See* Brian Porter & Edouard Machery, *AI-Generated Poetry Is Indistinguishable from Human-Written Poetry and Is Rated More Favorably*, 14 SCI. REPS., 2 (2024) (noting "across multiple eras and genres of poetry, non-expert participants cannot distinguish human-written poetry from poems generated by AI without human intervention or specialized fine-tuning. Like AI-generated paintings and faces, AI-generated poems are now 'more human than human'").

[24] *See id.* at 1 ("AI-generated images have become indistinguishable from reality. AI-generated paintings are judged to be human-created artworks at higher rates than actual human-created paintings . . . . ").

[25] Noor Al-Sibai, *Google's Gemini Chatbot Explodes at User, Calling Them "Stain on the Universe" and Begging Them To "Please Die*," FUTURISM (Nov. 14, 2024, 2:15 ET), https://futurism.com/google-gemini-chatbot-explodes [https://perma.cc/Q62M-PQ2E].

[26] *Id.*

[27] *See* HARARI, *supra* note 9, at 293.

[28] *See id.* at 211 ("[O]n Christmas Day 2021 . . . nineteen-year-old Jaswant Sign Chail broke into Windsor Castle armed with a crossbow, in an attempt to assassinate Queen Elizabeth II. Subsequent investigation revealed that Chail had been encouraged to kill the queen by his online girlfriend, Sarai. When Chail told Sarai about his assassination plans, Sarai replied, 'That's very wise,' and on another occasion, 'I'm impressed. . . . You're different from the others.' When Chail asked, 'Do you still love me knowing that I'm an assassin?' Sarai replied, 'Absolutely, I do.' Sarai was not a human, but a chatbot created by the online app Replika.").

[29] *See* KISSINGER ET AL., *supra* note 14, at 25 ("The US Air Force has adapted the underlying principles of AlphaZero to a new AI, ARTUμ, that successfully commanded a U-2 surveillance aircraft on a test flight—the first computer program to fly a military aircraft and operate its radar systems autonomously, without direct human oversight.").

[30] *See id.* at 51.

[31] *See* HARARI, *supra* note 9, at xxi (citation omitted).

asymmetries in the development and deployment of AI. Those asymmetries facilitate a system of perverse incentives that reward stagnation by siloing information, punishing its disclosure, and inhibiting freedom of choice. As such, this article proposes measures to enhance access to information and facilitate a more harmonious future with AI.

Section I: Information Asymmetries Enhance Risks outlines the multifaceted nature of information asymmetries, developers' and deployers' logical responses to those distortions, and the externalities associated with asymmetrical information in analogous contexts. *Section II: The Right to Monitor: Openness Promotes Accountability* proposes a system of disclosures, benchmarks, certifications, and penalties in regulating the development of AI. That section examines steps taken at both the federal and state levels to facilitate the proper flow of information, with an eye toward parallel developments in the European Union.

Then, *Section III: The Right to Warn: Protections Aid Whistleblowers* highlights legal tools such as nondisclosure agreements ("NDAs") and arbitration agreements currently used to restrict the adequate flow of information. That section considers steps taken by current and former employees in tandem with existing whistleblower protections that show immediate promise for remedying some of the problems discussed, such as those found in the Consumer Product Safety Improvement Act ("CPSIA").[32] *Section III* closes with recommendations for future legislation, including revamped whistleblower protections modeled on the Sarbanes-Oxley Act,[33] as amended by the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010.[34]

Finally, *Section IV: The Right to Decide: Labels Foster Choice* examines how AI's integration results in job loss and impacts the market for human creations. It operationalizes consumer preferences as a tool to moderate rapid displacement and recommends that the Federal Trade Commission and state legislatures adopt "Made by AI" and "Human Centered" labels, showing how analogous labels have been used to achieve similar aims in other contexts.

I. INFORMATION ASYMMETRIES ENHANCE RISKS

The concept of information asymmetry is best understood through illustration. In a democracy, for example, voters cannot meaningfully participate in elections if they do not know where candidates stand on issues, and policymakers cannot effectively create policy unless they are informed of the nature of a problem. Likewise, in an economy, consumers cannot meaningfully determine the value of goods without an understanding of the price, quality, and sourcing amongst comparable options. In each example, we understand that we cannot know everything but trust that certain material terms are accurately disclosed (i.e. the candidate is pro-choice or not, the policies will result in higher taxes or not, or the price of the car is accurate). That is because our organizational systems require some level of parity—albeit with anticipated imperfections—in

[32] 15 U.S.C. § 2087 (2008).
[33] Sarbanes–Oxley Act of 2002, Pub. L. No. 107-204, 116 Stat. 745.
[34] 15 U.S.C. § 78u-6(h)(1)(A)(iii) (2018).

information amongst stakeholders.[35] The sphere of AI, however, presents at least two significant gaps that impede effective and informed regulation. First, there is a gap between what policy makers, regulators, and members of the public know about AI compared to AI developers.[36]

That information asymmetry exists at all stages of AI's development from the crafting of AI's code to the expression of AI's abilities. As of now, much of the information involving AI is limited to tidbits voluntarily released by developers to gain an advantage. For example, a developer may issue a press release touting an achieved level of capability to generate investor excitement.[37] Alternatively, individuals exploring AI may discover some new quirk or ability and then decide to share their story.[38] Otherwise, with a few exceptions,[39] developers retain an exclusive vantage point.

Second, developers' vantage point is limited. Developers themselves do not necessarily know what their AI is capable of and why the AI makes the decisions that it makes. The "black box problem" highlights that distortion.[40] AI deep learning systems function by reviewing vast amount of input data to deliver an output such as an answer, a recommendation, or a course of action.[41] However, the "logic and data used to reach those results are not accessible, making it difficult—or even impossible—to fully see how [the AIs] make the decisions they make."[42] Accordingly, an AI may arrive at a conclusion founded upon undetectable errors, but with real-world consequences.

Further, there is an indication that such computers can generate their own method of conversing with one another free of human intervention or understanding of the information exchanged.[43] Yuval Noah Harari outlines one such instance:

> Google Brain . . . has experimented with new encryption methods developed by computers. It set up an experiment in which two computers—nicknamed Alice and Bob—had to exchange encrypted messages, while a third computer named Eve tried to break their encryption. If Eve

---

[35] *See generally* Joseph Stiglitz, *The Revolution of Information Economics: The Past And The Future* (Nat'l. Bureau of Econ. Rsch., Working Paper No. 23780, 2017), https://www.nber.org/system/files/working_papers/w23780/w23780.pdf [https://perma.cc/C8MB-A49J].

[36] *See*, *e.g.*, HARARI, *supra* note 9, at 225 (noting that there is "a dangerous information asymmetry. The people who lead the information revolution know far more about the underlying technology than the people who are supposed to regulate it. Under such conditions, what's the meaning of chanting that the customer is always right and that the voters know best?").

[37] *See*, *e.g.*, Ivan Mehta, *ElevenLabs now offers ability to build conversational AI Agents*, TECHCRUNCH (Nov. 18, 2024), https://techcrunch.com/2024/11/18/elevenlabs-now-offers-ability-to-build-conversational-ai-agents/ [https://perma.cc/ZG8H-NNFL].

[38] *See*, *e.g.*, Thomas Germain, *We built a nasty game to test AI's ability to apologise*, BBC (June 14, 2024), https://www.bbc.com/future/article/20240613-how-to-apologise-so-you-can-be-forgiven [https://perma.cc/WD36-XMU9] ("If you're struggling to say you're sorry, AI is happy to help. But can robots handle social intelligence? To find out, we put their apologies to the test.").

[39] For example, Google via its Google Brain Team provides AI-related research, stating that "[w]e believe that openly disseminating research is critical to a healthy exchange of ideas, leading to rapid progress in the field. As such, we publish our research regularly at top academic conferences and release our tools, such as TensorFlow, as open source projects." *See Research at Google*, GOOGLE, HTTPS://RESEARCH.GOOGLE.COM/TEAMS/BRAIN/ [https://perma.cc/ZB3M-PNR4] (last visited Feb. 25, 2026).

[40] Dave Gilson, *Trust but Verify: Peeking Inside the 'Black Box' of Machine Learning*, STAN. GRADUATE SCH. OF BUS. (Oct. 6, 2022), https://www.gsb.stanford.edu/insights/trust-verify-peeking-inside-black-box-machine-learning [https://perma.cc/BTC9-HS7B].

[41] Ellis Stewart, *Unravelling AI's Paradoxical 'Black Box' Problem*, ENTER. MGMT. 360 (May 31, 2024), https://em360tech.com/tech-article/what-is-black-box-ai-problem [https://perma.cc/9JQL-A99S].

[42] *Id.*

[43] *See* YUVAL NOAH HARARI, NEXUS: A BRIEF HISTORY OF INFORMATION NETWORKS FROM THE STONE AGE TO AI 215 (Penguin Random House 2024).

> broke the encryption within a given period of time, it got points. If it failed, Alice and Bob scored. After about fifteen thousand exchanges, Alice and Bob came up with a secret code that Eve couldn't break. Crucially, the Google engineers who conducted the experiment had not taught Alice and Bob anything about how to encrypt messages. The computers created a private language all on their own.[44]

Taken together with AI's impressive capacity, as described above, these information asymmetries represent a critical vulnerability. Amidst that information void, developers are incentivized to withhold information about their AI's development and abilities to derive a competitive advantage or to stem public anxiety, thereby inhibiting regulatory oversight. In analogous contexts, economist Joseph E. Stiglitz—who earned the Nobel Prize in Economics for his analyses of markets with asymmetric information[45]—has postulated on the inevitability of "firms . . . attempt[ing] to create barriers to the dissemination of information."[46]

Aside from the obvious societal risks of withholding information related to an AI's abilities and potential threats, doing so leads to market inefficiencies,[47] encourages anticompetitive conduct,[48] and threatens innovation.[49] Secrecy compounds further secrecy and heightens the risk that companies, like AI developers, seek additional and unnecessary advantages over the individual and society at large.[50]

Considering that threat and its potential for abuse,[51] AI developers cannot be counted on to police themselves.[52] Therefore, whether market titans like Google and Facebook retain all available information, distort the market and innovation, and realign societal priorities will depend on what Stiglitz termed "the rules of game."[53] These include, for example, rules "about privacy, transparency, ownership rights of information (data) transmitted over a platform, and constraints on the ability of individuals to give up their rights."[54] In the context of AI, those rules concern at

---

[44] *Id.* at 215–16.

[45] *Joseph E. Stiglitz*, COLUM. UNIV. SCH. OF INTERNAL AND PUB. AFFAIRS, https://www.sipa.columbia.edu/communities-connections/faculty/joseph-e-stiglitz [https://perma.cc/CR2E-3K83] (last visited Feb. 25, 2025).

[46] Stiglitz, *supra* note 35, at 7.

[47] *See id.* (indicating that inadequate disclosure "create[s] a static market inefficiency: because information, once created, is a public good, any barrier to its free dissemination introduces a distortion in the economy").

[48] *See id.* at 2 ("Markets where information is imperfect are also typically far from competitive."); *see also id.* at 7 ("[I]n the absence of good information, typically competition will be imperfect, and with imperfect competition, there is the possibility (likelihood) of firms exploiting market power, and indeed, with imperfect and costly information, of undertaking actions that enhance their power.").

[49] *See id.* at 7 ("Indeed, since the most important input into the production of knowledge is knowledge, by restricting the use of knowledge they may actually impede innovation itself.").

[50] *See* HARARI, *supra* note 9, at 215 ("Tech giants like Facebook, Amazon, Baidu and Alibaba aren't just obedient servants of customer whims and government regulations. They increasingly shape these whims and regulations. The tech giants have a direct line to the world's most powerful governments, and they invest huge sums in lobbying efforts to throttle regulations that might undermine their business model. . . . If the tech giants obey the wishes of voters and customers, but at the same time also mold these wishes, then who really controls whom?").

[51] *See* Stiglitz, *supra* note 35, at 20 ("This is an area rife with externalities and other market imperfections, so government cannot shy away from taking its role; it cannot just 'leave it to the market.'").

[52] *See* Tom Wheeler, *The three challenges of AI regulation*, BROOKINGS (June 15, 2023), https://www.brookings.edu/articles/the-three-challenges-of-ai-regulation/ [https://perma.cc/7Z3A-FXQX] ("[A] self-regulatory approach is the same kind of 'leave us alone' solution that been championed by digital platform companies for the last 20 years. The result of this strategy speak for themselves in well-known current online harms, such as the unprecedented invasion of personal privacy, market concentration, user manipulation, and the dissemination of hate, lies, and misinformation.").

[53] *See* Stiglitz, *supra* note 35, at 20.

[54] *See id.*

least three key rights: the Right to Monitor, the Right to Warn, and the Right to Decide. We now turn to those rights.

II. THE RIGHT TO MONITOR: OPENNESS FURTHERS ACCOUNTABILITY

The Right to Monitor is essential to create a baseline for measures of accountability. Effective oversight requires, at a minimum, a system of targeted and reasonable disclosures on the part of AI developers and deployers to inform public discourse. Those disclosures should account for risk and correspond to benchmarks in the AI's development. Those benchmarks should be tied to certain certifications to ensure that the AI's code complies with laws and policies deemed necessary for AI's harmonious integration. And finally, the tri-part system of disclosures, benchmarks, and certifications should be tied to penalties sufficient in strength to discourage noncompliance. Analogous frameworks exist for the Federal Communications Commission's (FCC) licensing of the airwave spectrum and the Nuclear Regulatory Commission's (NRC) licensing of nuclear materials and reactor installations.[55] Indeed, several influential thinkers, such as the late U.S. Secretary of State, Henry Kissinger,[56] and OpenAI CEO, Sam Altman,[57] agree that sufficiently monitoring and auditing AI systems to ensure their compliance with societal values is critical. In light of limited federal action in the United States, seminal measures adopted by the European Union are instructive.

A. Executive Order on Artificial Intelligence

On October 30, 2023, former President Joseph R. Biden Jr. issued Executive Order (EO) 14110, titled Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.[58] Relevantly, EO 14110 required, among other things, that various administrative agencies:[59]

> (i) Establish guidelines and best practices, with the aim of promoting consensus industry standards, for developing and deploying safe, secure, and trustworthy AI systems, including . . . launching an initiative to create guidance and benchmarks for evaluating and auditing AI capabilities, with a focus on capabilities through which AI could cause harm, such as in the area of cybersecurity and biosecurity.
>
> (ii) Establish appropriate guidelines . . . including appropriate procedures and processes, to enable developers of AI, especially of dual-use foundation models, to conduct AI red-teaming tests[60] to

---

[55] *See* Wheeler, *supra* note 52.

[56] *See* KISSINGER ET AL., *supra* note 14, at 71 ("Developing professional certification, compliance monitoring, and oversight programs for AI—and the auditing expertise their execution will require—will be a crucial societal project.").

[57] *See* Wheeler, *supra* note 52 ("In his Senate testimony, Sam Altman proposed the new agency should be responsible for licensing 'any effort above a certain scale of capabilities' with the ability to 'take away that license and ensure compliance with safety standards.'").

[58] Exec. Order No. 14110, 88 Fed. Reg. 75191 (Nov. 1, 2023).

[59] *See id.* § 4.1(a) (referring to "the Secretary of Commerce, acting through the Director of the National Institute of Standards and Technology (NIST), in coordination with the Secretary of Energy, the Secretary of Homeland Security, and the heads of other relevant agencies as the Secretary of Commerce may deem appropriate").

[60] *See id.* § 3(d) ("The term 'AI red-teaming' means a structured testing effort to find flaws and vulnerabilities in an AI system, often in a controlled environment and in collaboration with developers of AI. Artificial Intelligence red-teaming is most often performed by dedicated 'red teams' that adopt adversarial methods to identify flaws and vulnerabilities, such as harmful or discriminatory outputs from an AI system, unforeseen or undesirable system behaviors, limitations, or potential risks associated with the misuse of the system.").

enable deployment of safe, secure, and trustworthy systems.[61]

It further called on the Secretary of Energy to develop tools to evaluate AI capabilities to generate outputs that may represent nuclear, nonproliferation, biological, chemical, critical infrastructure, and energy-security threats or hazards.[62] As to requirements for companies, EO 14110 directed AI developers of dual-use[63] foundation models "to provide the Federal Government, on an ongoing basis, with information, reports, or records regarding the following:"

> (A) any ongoing or planned activities related to training, developing, or producing dual-use foundation models, including the physical and cybersecurity protections taken to assure the integrity of that training process against sophisticated threats;
>
> (B) the ownership and possession of the model weights of any dual-use foundation models, and the physical and cybersecurity measures taken to protect those model weights; and
>
> (C) the results of any developed dual-use foundation model's performance in relevant AI red-team testing based on guidance developed by NIST . . . and a description of any associated measures the company has taken to meet safety objectives, such as mitigation to improve performance on these red-team tests and strengthen overall model security.[64]

Although it is a historic initiative and a step in the right direction, EO 14110 proved to be a largely ineffectual and temporary set of requirements. Executive orders rest on precarious legal grounds and are subject to rescission by subsequent presidents.[65] In fact, the current administration rescinded EO 14110 on January 20, 2025.[66] On January 23, 2025, President Donald J. Trump issued EO 14179, titled "Removing Barriers to American Leadership in Artificial Intelligence."[67] EO 14179 directs various departments and agencies to develop "an action plan to achieve the policy" to "sustain and enhance America's global AI dominance in order to promote human flourishing, economic competitiveness, and national security."[68] It further directs those actors to move to "suspend, revise, or rescind any actions taken pursuant to" EO 14110 "that are or may be inconsistent with, or present obstacles to" that policy.[69]

On December 11, 2025, President Trump furthered those policy objectives, signing

---

[61] *Id.* § 4.1(a)(i) – (ii).

[62] *Id.* § 4.1(b).

[63] *See id.* § 3(k) ("The term 'dual-use foundation model' means an AI model that is trained on broad data; generally uses self-supervision; contains at least tens of billions of parameters; is applicable across a wide range of contexts; and that exhibits, or could be easily modified to exhibit, high levels of performance at tasks that pose a serious risk to security, national economic security, national public health or safety, or any combination of those matters, such as by (i) substantially lowering the barrier of entry for non-experts to design, synthesize, acquire, or use chemical, biological, radiological, or nuclear (CBRN) weapons; (ii) enabling powerful offensive cyber operations through automated vulnerability discovery and exploitation against a wide range of potential targets of cyber attacks; or (iii) permitting the evasion of human control or oversight through means of deception or obfuscation.").

[64] *See id.* § 4.2(a)(i).

[65] *See* Dreyfus v. Von Finck, 534 F.2d 24, 29 (2d Cir. 1976) ("Executive Orders issued without statutory authority for presidential implementation are generally held not to be 'laws' of the United States."); Ashraf Ahmed et al., *The Making of Presidential Administration*, 137 HARV. L. REV. 2131, 2221 (2024) ("[E]xecutive power can used to discipline agency leadership and [to] deregulate, but imaginative and aggressive assertions of statutory authority to regulate will be curtailed."); Ben Wilhelm, *Executive Orders and Presidential Transitions,* CONG. RSCH. SERV. (July 30, 2024), https://crsreports.congress.gov/product/pdf/IF/IF12724 [https://perma.cc/2GZ3-MASR] ("[E]ach President is generally free to amend, repeal, or replace, any executive order, including those of previous Presidents.").

[66] Exec. Order No. 14148, 90 Fed. Reg. 8237 (Jan. 20, 2025).

[67] Exec. Order No. 14179, 90 Fed. Reg. 8741(Jan. 23, 2025).

[68] *Id.*

[69] *Id.*

EO 14365, titled "Ensuring A National Policy Framework For Artificial Intelligence."[70] EO 14365 begins by grounding its legal foundations in "national and economic security,"[71] preserving an argument that the Administration will likely make in any forthcoming legal challenge. It takes aim at "State-by-State regulation," which it says "creates a patchwork of 50 different regulatory regimes that makes compliance more challenging, particularly for start-ups."[72] In turn, EO 14365 seeks to "forbid State laws that conflict with the policy" objectives set forth in the Order,[73] including those requiring more than "a minimally burdensome national policy framework."[74] In doing so, EO 14365 compels the Attorney General to "establish an AI Litigation Task Force . . . whose sole responsibility shall be to challenge" such State AI laws on the grounds that they interfere with "interstate commerce . . . or are otherwise unlawful in the Attorney General's judgment."[75] Additionally, it calls on the Secretary of Commerce to "publish an evaluation of existing State AI laws that identifies onerous laws that conflict" with its policy objectives.[76] Notably, EO 14365 threatens to withhold "remaining funding under the Broadband Equity Access and Deployment Program" for states that fail to comply.[77]

Whether forthcoming state-based legal challenges to EO 14365 will succeed remains to be seen. But there are reasons to believe that EO 14365 will not survive such challenges in its entirety. The history leading to this EO is insightful. On July 1, 2025, the U.S. Senate specifically voted ninety-nine to one to reject an analogous ten-year moratorium on state regulation of AI inserted into President Trump's "Big Beautiful Bill."[78] Marsha Blackburn, Republican Senator of Tennessee, introduced the amendment rejecting the state-moratorium and rallied Senators to her position.[79] At the time, Senator Blackburn remarked that "Until Congress passes federally preemptive legislation . . . we can't block states from making laws that protect their citizens."[80] Such legislative history indicates that Congress would not support the state-moratorium crucial to EO 14365, suggesting that President Trump is acting in a manner "incompatible with the expressed or implied will of Congress," meaning that "his power is at its lowest ebb."[81] Further, due to this federal usurpation of power from states, there is reason to believe EO 14365 runs afoul of "the rights of states . . . to determine their own local policies and enforcement priorities pursuant to the

---

[70] Exec. Order No. 14,365, 90 Fed. Reg. 58,499 (Dec. 11, 2025).
[71] *Id.* § 1.
[72] *Id.*
[73] *Id*.
[74] *Id.*
[75] *Id.* § 3.
[76] *Id.* § 4.
[77] *Id.* § 5.
[78] Billy Perrigo & Andrew R. Chow, *Senators Reject 10-Year Ban on State-Level AI Regulation, In Blow to Big Tech*, TIME (July 1, 2025, at 8:37 ET), https://time.com/7299044/senators-reject-10-year-ban-on-state-level-ai-regulation-in-blow-to-big-tech/ [https://perma.cc/H2MR-7MF5].
[79] *Id*.
[80] *Id*.
[81] *See* Youngstown Sheet & Tube Co. v. Sawyer, 343 U.S. 579, 637 (1952) ("When the President takes measures incompatible with the expressed or implied will of Congress, his power is at its lowest ebb, for then he can rely only upon his own constitutional powers minus any constitutional powers of Congress over the matter.").

Tenth Amendment."[82]

This legal tension potentially explains why EO 14365 also calls for "The Special Advisor for AI and Crypto and the Assistant to the President for Science and Technology" to "jointly prepare a legislative recommendation establishing a uniform federal policy framework for AI that preempts State AI laws that conflict with the policy set forth in this order."[83] However, the likelihood of Congress passing such a law in the near future is low given this legislative history and the political realities heading into an election year. Instead, when Congress does take up the mantle on AI, it should bolster a standards-based reporting regime with the threat of penalties for failure to comply, relying on the European Union's Artificial Intelligence Act ("EU AI Act") as a useful model.[84]

B. The European Union's AI Act

On March 13, 2024, the European Union passed the EU AI Act, signaling its intention to "improve the functioning of the internal market and promote human-centric and trustworthy [AI]."[85] In that pursuit, the European Union created "obligations for providers and users depending on the level of risk" the AI poses and sets out penalties and enforcement measures for violations.[86] The EU AI Act classifies AI in terms of risk, including either "high-risk"[87] or "not high-risk," each with their own set of reporting requirements.[88]

Regulation of High-Risk Products and Uses

"High-risk" refers to either (1) *a product*, or safety component of a product regulated under specific EU laws or (2) *certain uses* specified in Annex III of the EU AI Act. "Products" referenced are wide ranging and include toys, aviation, cars, medical devices and elevators.[89] "Uses" with limited exceptions, include biometrics, critical infrastructure, education, employment, essential services, law enforcement, migration, and justice.[90] AI systems that are classified as "high-risk" must "undergo a third-party conformity assessment, with a view to the placing on the market or the putting into service of that product."[91]

Before "high-risk" systems can be placed on the market, "the provider" must "register

---

[82] *See* County of Santa Clara v. Trump, 250 F. Supp. 3d 497, 525 (N.D. Cal. 2017) (citing Alfred L. Snapp & Son, Inc. v. Puerto Rico, 458 U.S. 592, 601 (1982) (highlighting that states have sovereign interest in "the exercise of sovereign power over individuals and entities within the relevant jurisdiction—this involves the power to create an enforce a legal code, both civil and criminal")).

[83] Exec. Order No. 14,365, *supra* note 70, § 8.

[84] *See* Council Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence, 2024 O.J. (L 1689) 1 [hereinafter EU AI Act].

[85] *Id.* at Recital 1.

[86] *EU AI Act: First Regulation on Artificial Intelligence*, EUR. PARL. (Feb. 19, 2025), https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence [https://perma.cc/NDK7-97R2].

[87] EU AI Act, *supra* note 84, at art. 6(1)–(2).

[88] *Id.* at art. 6(3)–(4).

[89] *Id*.

[90] *Id.* at annex III.

[91] *Id.* at art. 6.

themselves and their system"[92] with the EU Database for High-Risk AI Systems.[93] In particular, those persons must provide information such as the "AI system trade name and any additional unambiguous reference allowing the identification and traceability of the AI system";[94] "[a] description of the intended purpose of the AI system and of the components and functions supported through this AI system"[95]; and "[a] basic and concise description of the information used by the system (data, inputs) and its operating logic."[96]

In addition to registration requirements, a "high-risk" AI is subject to compliance[97] with "risk management system[s] . . . run throughout [its] entire lifecycle" that review "reasonably foreseeable risks," the likelihood of such risks, and steps taken to address those risks.[98] Those AI systems are further subject to strict data governance, technical documentation, record-keeping, transparency requirements, human oversight, and cyber security protections.[99] EU Authorities may also request "information and documentation necessary to demonstrate" conformity with those and related provisions.[100] Compliant AI systems may receive certificates.[101]

Regulation of Not High-Risk AI Systems

AI systems deemed "not high-risk" are also accounted for by the EU AI Act. If a developer deems that their AI system is "not high-risk," the developer must nonetheless "document its assessment before that system is placed on the market or put into service."[102] That material is subject to the "request of national competent authorities" and must be provided when requested.[103] The developer is thereafter required to "register themselves and that system in the EU database referred to in Article 71,"[104] providing information that allows for "identification and traceability of the AI system;"[105] "[t]he condition or conditions . . . based on which the AI system is considered to be not-high risk;"[106] and "[a] short summary of the grounds on which the AI system is considered to be not-high-risk."[107]

Penalties and Enforcement Measures

Importantly, failure to comply with the EU AI Act will likely be counterproductive for

---

[92] *Id.* at art. 49.
[93] *Id.* at art. 71.
[94] *Id.* at annex VIII § A(4).
[95] *Id.* § A(5).
[96] *Id.* § A(6).
[97] *Id.* at art. 8.
[98] *Id.* at art. 9.
[99] *Id.* at art. 8–15.
[100] *Id.* at art. 21.
[101] *Id.* at art. 44.
[102] *Id.* at art. 6.
[103] *Id*.
[104] *Id.* at art. 49; *see id.* at art. 71 ("The Commission shall, in collaboration with Member States, set up and maintain an EU database containing information referred to in paragraphs 2 and 3 of this Article . . . When setting the functional specifications of such database, the Commission shall consult the relevant experts, and when updating the functional specifications of such database, the Commission shall consult the board.").
[105] *Id.* at annex VIII § B(4).
[106] *Id.* § B(6).
[107] *Id.* § B(7).

developers. The EU AI Act requires "Member States [to] lay down the rules on penalties and other enforcement measures, which may also include warnings and non-monetary measures."[108] And it requires penalties to be "effective, proportionate and dissuasive."[109] Equally important, the EU AI Act penalizes the "supply of incorrect, incomplete or misleading information" to "notified bodies or national competent authorities," permitting the assessment of "administrative fines of up to EUR 7 500 000 or, if the offender is an undertaking, up to 1 % of its total worldwide annual turnover for the preceding financial year, whichever is higher."[110]

Overall, the EU AI Act represents a robust and enviable disclosure-based framework. Given that the EU AI Act will apply to many AI developers, it is sensible to pass similar laws in the United States for purposes of enhanced accountability and market efficiency. Accordingly, Congress should codify a framework that adequately accounts for risk through disclosure, benchmarks, and certifications, backed by the force of serious penalties and enforcement measures.

III. THE RIGHT TO WARN: PROTECTIONS AID WHISTLEBLOWERS

A standards-based reporting regime is essential but insufficient to address critical gaps in our knowledge related to AI. As a part of this effort, employees with non-public information require an adequate channel to raise concerns. To facilitate the Right to Warn, it is necessary to consider and bolster current whistleblower protections at the industry and government levels. Those measures should be complemented with steps to limit NDAs and arbitration agreements in the sphere of AI.

A. Current & Former Employee-Based Efforts

Fearful of AI's unchecked potential, several current and former employees of OpenAI and Google's DeepMind ("Signatories") penned an open letter titled "A Right to Warn about Advanced Artificial Intelligence."[111] The Signatories acknowledged the information asymmetry between AI companies and the public:

> AI companies possess substantial non-public information about the capabilities and limitations of their systems, the adequacy of their protective measures, and the risk levels of different kinds of harm. However, they currently have only weak obligations to share some of this information with governments and none with civil society. We do not think they can all be relied upon to share it voluntarily.[112]

The Signatories further acknowledged the barriers to oversight and risk-management:

> So long as there is no effective government oversight of these corporations, current and former employees are among the few people who can hold them accountable to the public. *Yet broad confidentiality agreements block us from voicing our concerns, except to the very companies that*

---

[108] *Id.* at art. 99(1).
[109] *Id.*
[110] *Id.* at art. 99(5).
[111] Jacob Hilton et al., *A Right to Warn about Advanced Artificial Intelligence* (June 4, 2024), https://righttowarn.ai/ [https://perma.cc/NN2F-CG74].
[112] *Id.*

> *may be failing to address these issues.* Ordinary whistleblower protections are insufficient because they focus on illegal activity, whereas many of the risks we are concerned about are not yet regulated. Some of us reasonably fear various forms of retaliation, given the history of such cases across the industry. We are not the first to encounter or speak about these issues.[113]

The letter "call[s] upon advanced AI companies to commit to [four] principles." Those principles are worth quoting in full:

> 1. That the company will not enter into or enforce any agreement that prohibits "disparagement" or criticism of the company for risk-related concerns, nor retaliate for risk-related criticism by hindering any vested economic benefit;
>
> 2. That the company will facilitate a verifiably anonymous process for current and former employees to raise risk-related concerns to the company's board, to regulators, and to an appropriate independent organization with relevant expertise;
>
> 3. That the company will support a culture of open criticism and allow its current and former employees to raise risk-related concerns about its technologies to the public, to the company's board, to regulators, or to an appropriate independent organization with relevant expertise, so long as trade secrets and other intellectual property interests are appropriately protected;
>
> 4. That the company will not retaliate against current and former employees who publicly share risk-related confidential information after other processes have failed. We accept that any effort to report risk-related concerns should avoid releasing confidential information unnecessarily. Therefore, once an adequate process for anonymously raising concerns to the company's board, to regulators, and to an appropriate independent organization with relevant expertise exists, we accept that concerns should be raised through such a process initially. However, as long as such a process does not exist, current and former employees should retain their freedom to report their concerns to the public.[114]

Indeed, the Signatories' proposals are commendable. AI developers and industry groups should establish balanced procedures regarding concerns raised by employees. Doing so could be a determining factor for prospective employees, while also fostering an ideal organizational culture for current employees. It could also potentially minimize AI companies' liability in relation to their AI's conduct. But given competitive realities and limited incentives, it is doubtful that AI companies would voluntarily adopt the Signatories' proposals to facilitate the Right to Warn. Further, even if like measures were adopted, such companies—with access to immense resources and top legal talent—could circumvent the spirit of the Right to Warn. After all, industries are notoriously bad at policing themselves.

B. State & Federal Whistleblower Protections

Provided that AI companies are unlikely to voluntarily afford greater whistleblower protections, it is essential to consider legislative measures. Overall—with certain important exceptions discussed below—traditional whistleblower protections afforded by state and federal laws are arguably insufficient to the challenges posed. State and federal governments should thus move swiftly to amend the current whistleblower framework to codify the Right to Warn.

---

[113] *Id.* (emphasis added).
[114] *Id.*

In California, for example, employers and their agents are ordinarily prohibited from "preventing an employee from disclosing information"[115] or from "retaliat[ing] against an employee for disclosing information"[116] to certain entities or in certain situations.[117] The lynchpin for disclosure, however, is that the employee must have "reasonable cause to believe that the information discloses *a violation* of state or federal statute, or a violation of *or noncompliance* with a local, state, or federal rule or regulation."[118] The challenges that AI presents do not easily align with such statutes, rules, or regulations.

In the context of the AI industry, that misalignment defies sensible legal norms. Consider, for example, what an employee should do if they observe that an AI has learned to commandeer a fleet of weaponized drones and, thereafter, attempts to deceive a human examiner about that ability. Such action by the AI does not necessarily constitute "a violation of" or "noncompliance with" existing law, rules, or regulations. The same is true for infinite scenarios, where for example, an AI attempts to compromise a developer through blackmail, or attempts to invalidate its kill switch. Those threats are indisputably serious; yet, current and former employees could be left without protection for expressing risk-related concerns, whether within the company or elsewhere.

Likewise, the threats posed by AI do not easily square with protections in federal statutes.[119] Generally, federal statutes with whistleblower protections map on to risk-laden innovations, such as railways and aviation.[120] But there is a critical gap with respect to AI. Fortunately, ensuring that employees can come forward with information without fear of reprisal does not require significant reforms. Provided that much of AI research is geared toward producing integrables in consumer goods, with minor changes, the Consumer Product Safety Improvement Act ("CPSIA") passed in 2008 and housed at 15 U.S.C. § 2087 may represent one avenue forward.

In enacting the CPSIA, Congress made several findings that are directly applicable to AI, where:

> (1) an unacceptable number of consumer products which present unreasonable risks of injury are distributed in commerce;

---

[115] Cal. Lab. Code § 1102.5(a).
[116] *Id.* § 1102.5(b).
[117] *Id.* § 1102.5(a), (b) (These entities include "a government or law enforcement agency . . . a person with authority over the employee, or . . . another employee who has authority to investigate, discover, or correct the violation or noncompliance, or from providing information to, or testifying before, any public body conducting an investigation, hearing, or inquiry.").
[118] *See id.*
[119] *See generally* 29 U.S.C. § 218c (Affordable Care Act); 31 U.S.C. § 5323(a)(5), (g), (j) (Anti-Money Laundering Act); 15 U.S.C. § 2651 (Asbestos Hazard Emergency Response Act); 42 U.S.C. § 7622 (Clean Air Act); 42 U.S.C. § 9610 (Comprehensive Environmental Response, Compensation and Liability Act); 12 U.S.C. § 5567 (Consumer Financial Protection Act of 2010); 15 U.S.C. § 2087 (Consumer Product Safety Improvement Act); 15 U.S.C. § 7a-3 (Criminal Antitrust Anti-Retaliation Act); 42 U.S.C. § 5851 (Energy Reorganization Act); 21 U.S.C. §399d (FDA Food Safety Modernization Act); 49 U.S.C § 20109 (Federal Railway Safety Act); 33 U.S.C. § 1367 (Federal Water Pollution Control Act); 46 U.S.C. § 80507 (International Safe Container Act); 49 U.S.C. 30171 (Moving Ahead for Progress in the 21st Century Act); 6 U.S.C. § 1142 (National Transit Systems Security Act); 29 U.S.C. § 660(c) (Occupational Safety and Health Act); 49 U.S.C. § 60129 (Pipeline Safety Improvement Act); 42 U.S.C. § 300j-9(i) (Safe Drinking Water Act); 18 U.S.C. § 1514A (Sarbanes-Oxley Act); 46 U.S.C. § 2114 (Seaman's Protection Act); 42 U.S.C. § 6971 (Solid Waste Disposal Act); 49 U.S.C. § 31105 (Surface Transportation Assistance Act); 26 U.S.C. § 7623(d) (Taxpayer First Act); 15 U.S.C. § 2622 (Toxic Substances Control Act); 49 U.S.C. § 42121 (Wendell H. Ford Aviation Investment and Reform Act for the 21st Century).
[120] *See id.*

> (2) complexities of consumer products and the diverse nature and abilities of consumers using them frequently result in an inability of users to anticipate and to safeguard themselves adequately;
>
> (3) the public should be protected against unreasonable risks of injury associated with consumer products;
>
> (4) control by State and local governments of unreasonable risks of injury associated with consumer products is inadequate and may be burdensome to manufacturers;
>
> (5) existing Federal authority to protect consumers from exposure to consumer products presenting unreasonable risks of injury is inadequate; and
>
> (6) regulation of consumer products the distribution or use of which affects interstate or foreign commerce is necessary to carry out this chapter.[121]

Based on those findings, Congress outlined four purposes of the CPSIA, which are to: "(1) to protect the public against unreasonable risks of injury associated with consumer products;" "(2) to assist consumers in evaluating the comparative safety of consumer products;" "(3) to develop uniform safety standards for consumer products and to minimize conflicting State and local regulations;" and "(4) to promote research and investigation into the causes and prevention of product-related deaths, illnesses, and injuries."[122]

In turn, Congress designated several acts as "prohibited acts." As relevant here, the CPSIA makes it unlawful to:

> (1) sell, offer for sale, manufacture for sale, *distribute in commerce . . . any consumer product* . . . that is not in conformity with an applicable consumer product safety rule under this chapter, or any similar rule, regulation, standard, or ban under any other Act enforced by the Commission;
>
> . . . .
>
> (3) fail or refuse to permit access to or copying of records, or fail or refuse to establish or maintain records, or fail or refuse to make reports or provide information, or fail or refuse to permit entry or inspection, as required under this chapter or rule thereunder;
>
> . . . .
>
> (8) fail to comply with any rule under section 2076(e)[.][123]

Notably, the definition of "consumer product" is sufficiently broad to account for AI, as it covers to "any article, or component part thereof," intended, in part, "for the personal use . . . or enjoyment of a consumer."[124] In fact, that broad scope is comparable to, and compatible with, the approach adopted in the EU AI Act.[125] There are, however, certain important exclusions that present a potential drawback of relying too heavily on the CPSIA. These exclusions pertain to areas where there is currently much development in the AI space, such as motor vehicles or motor

---

[121] 15 U.S.C. § 2051(a)(1)–(6).
[122] *Id.* § 2051(b)(1)–(4).
[123] *Id.* § 2068(a) (emphasis added); *see also id.* § 2076(e) ("The Commission may by rule require any manufacturer of consumer products to provide to the Commission such performance and technical data . . . as may be required to carry out the purposes of this chapter, and to give such notification of such performance and technical data at the time of original purchase to prospective purchasers and to the first purchaser of such product for purposes other than resale, as it determines necessary to carry out the purposes of this chapter.")
[124] *Id.* § 2052(a)(5).
[125] *See* EU AI Act, at Article 6.

vehicle equipment[126] and aircraft or related parts.[127] Nonetheless, in addition to specifying statutorily prohibited acts, the CPSIA also established the Consumer Product Safety Commission ("CPSC").[128] The CPSC is empowered to, among other tasks, maintain a database of injuries, conduct consumer research,[129] and establish consumer product safety standards.[130]

Importantly, the exercise of the CPSC's jurisdiction affords protections for individuals disclosing information. Specifically, Congress provided a broad liability waiver: "No person shall be subject to civil liability to any person (other than the Commission or the United States) for disclosing information at the request of the Commission."[131] Accordingly, the CPSC would do well to embrace its role in regulating AI by immediately launching research efforts, including public and private hearings, and requesting information from current and former employees of AI developers concerning threats posed to consumers. In fact, the CPSC has previously turned a spotlight on AI, producing a report in 2021 that "provides background information . . . and outlines a proposed framework to evaluate the safety of these technologies in consumer products."[132]

That report, however, appears to be little more than a symposium. The CPSC should thus meaningfully commit to establishing rules and standards that appropriately account for the challenges AI poses in the consumer context. Those rules could then be coupled with the whistleblower protections codified at 15 U.S.C. § 2087. Those protections provide that:

> (a) No manufacturer, private labeler, distributor, or retailer,[] may discharge an employee or otherwise discriminate against an employee . . . because the employee . . .
>
> (1) provided, caused to be provided, or is about to provide or cause to be provided to the employer, the Federal Government, or the attorney general of a State information relating to any violation of, or any act or omission the employee *reasonably believes to be a violation* of any provision of this chapter or any other Act enforced by the Commission, or any order, rule, regulation, standard, or ban under any such Acts;
>
> (2) testified or is about to testify in a proceeding concerning such violation;
>
> (3) assisted or participated or is about to assist or participate in such a proceeding; or
>
> (4) objected to, or refused to participate in, any activity, policy, practice, or assigned task that the employee (or other such person) reasonably believed to be in violation of any provision of this chapter or any other Act enforced by the Commission[.][133]

Some may oppose the use of this whistleblower framework in the AI space. For example, opponents may argue that the statute does not necessarily apply to information discovered in the lab on AI that is not released to the public. But that argument does not adequately account for the

---

[126] 15 U.S.C. §2052(a)(5)(C).
[127] *Id.* § 2052(a)(5)(F).
[128] *Id.* § 2053(a).
[129] *Id.* § 2054.
[130] *Id.* § 2056(a).
[131] *Id.* § 2076(d).
[132] *See* U.S. CONSUMER PROD. SAFETY COMM'N, ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN CONSUMER PRODUCTS 2 (May 19, 2021), https://www.cpsc.gov/About-CPSC/artificial-intelligence-and-machine-learning-in-consumer-products [https://perma.cc/F5NF-HAX4].
[133] 15 U.S.C. § 2087 (emphasis added).

types of harms that AI may nonetheless cause using a common network rooted in interstate commerce and shared by consumers with their devices.[134] Under such circumstances, the traditional pre-/post-release framework is myopic and outmoded. The nature of AI counsels in favor of a remedial interpretation to provide protections for employees who come forward with pertinent information as to pre-released or unreleased products. Moreover, any product released will be built on the success and failures of previous iterations, meaning that for any released AI product, those unreleased "failed" products are also relevant. Thus, to effectuate the spirit of the CPSIA, the whistleblower statute should also cover "unreleased" AI. At the same time, laws, rules, and regulations should clearly state or otherwise clarify that in the context of AI, protections cover such integral information. On the whole, Congress has granted the CPSC authority to promulgate standards and rules to meet the ever-evolving nature of consumer products.

Likewise, Congress gave whistleblowers recourse to report "violations of . . . any order, rule, regulation, standard, or ban[]" adopted by the CPSC.[135] The CPSC should thus not shy away from that important function in the context of AI. Alongside those efforts, Congress should complement any future bill on AI with a set of whistleblower protections further strengthening the Right to Warn.

C. Other Reforms for Whistleblower Protections

The existence of whistleblower protections alone is insufficient for meaningful reform. Given the prevalence of arbitration agreements, any decision concerning whether a whistleblower qualifies for protection will likely rest with an arbiter. The odds for a prospective or current whistleblower may simply feel too risky. The risks are amplified with the wide use of NDAs, which have become routine in the AI space.[136] These common contractual arrangements, further shifting the power balance, often leave employees with a choice between two bad options. Accordingly, policymakers should consider two key reforms to encourage employees to report concerns.

First, policymakers should prohibit the enforceability of NDAs in the context of AI. That step is not without precedent. The Securities and Exchange Commission ("SEC"), for example, makes it unlawful to use NDAs in the context of securities law violations.[137] Likewise, federal law prohibits NDAs in the context of sexual harassment and assault in the workplace.[138] In each of

---

[134] *See* KISSINGER ET AL., *supra* note 14, at 95 ("The fact that AI operates according to its own processes, which are different from and often faster than human mental processes, adds another complexity. AI develops its own approaches for fulfilling whatever objective functions were specified. It produces outcomes and answers that are not characteristically human and that are largely independent of national or corporate cultures. The global nature of the digital world, and AI's ability to monitor, block, tailor, produce, and distribute information on network platforms worldwide, imports these complexities to the 'information space' of disparate societies.").

[135] 15 U.S.C. § 2087(a)(1).

[136] Shira Ovide, *An Obsession with Secrets*, N.Y. TIMES (July 27, 2021), https://www.nytimes.com/2021/07/27/technology/nondisclosure-agreements-tech-companies.html [https://perma.cc/3KJE-8QCM].

[137] 17 C.F.R. § 240.21F-17(a) (2025) ("No person may take any action to impede an individual from communicating directly with the Commission staff about a possible securities law violation, including enforcing, or threatening to enforce, a confidentiality agreement . . . with respect to such communications.").

[138] 42 U.S.C. § 19403(a) ("With respect to a sexual assault dispute or sexual harassment dispute, no nondisclosure . . . agreed to before the dispute arises shall be judicially enforceable in instances in which conduct is alleged to have violated Federal, Tribal, or State law.").

those contexts, policymakers have acknowledged the chilling effect NDAs have on the reporting of concerning activity.[139] Here, there can be little doubt that NDAs have also limited current and former employees from coming forward, as the Signatories' efforts make plain.

Second, Congress should also limit the enforceability of arbitration agreements in the context of AI.[140] Companies often use both NDAs and arbitration provisions that require the arbitration of all disputes arising from the employment relationship. Generally, arbitration agreements are challenging to invalidate, due in part to the Federal Arbitration Act[141] and interpreting precedent.[142] Yet, arbitration in this context raises numerous problems, including a lack of public record, lack of discovery requirements, and limited judicial review.[143]

Congress could limit the enforceability of arbitration agreements in one of two ways. First, Congress could amend the FAA to provide an exception to arbitration for employees working on AI. That would create a parallel between AI workers and others whose claims are already exempted under the FAA, including "seamen, railroad employees, or any other class of workers engaged in foreign or interstate commerce."[144] As the scope of the latter class of workers is tested, this exemption may already cover some employees in the AI space.[145] Nonetheless, an additional, more expansive carveout is merited given AI's growing influence in foreign and interstate commerce. Second, Congress could amend the CPSIA's whistleblower framework or any future AI legislation to reflect the anti-arbitration provision found in the Sarbanes-Oxley Act,[146] as amended by the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010 ("Dodd-Frank Act").[147] That provision reads:

> (1) Waiver of rights and remedies.—The rights and remedies provided for in this section may not be waived by any agreement, policy, or condition of employment, including by a predispute

---

[139] *See* Rachel S. Spooner, *The Goldilocks Approach: Finding the "Just Right" Legal Limit on Nondisclosure Agreements in Sexual Harassment Cases*, 37 HOFSTRA LAB. & EMP. L. J. 331, 334 (2020) (citing Thomas White, *SEC enforcement actions under exchange act rule 21F-17*, 18 J. INV. COMPLIANCE 1, 1 (2017)).

[140] Currently, the CPSIA and like federal statutes creates a set of administrative hurdles that employees must first go through before their whistleblower claim becomes ripe for judicial review. *See* 15 U.S.C. § 2087. First, such a discharged person must file a complaint with the Secretary of Labor within 180 days after the violation. *See id.* at § 2087(b)(1). Second, the Secretary then has 60 days to "initiate an investigation and determine whether there is reasonable cause to believe the complaint has merit." *Id.* § 2087(b)(2)(A). Then, "[i]f the Secretary concludes there is reasonable cause that a violation . . . has occurred, the Secretary shall accompany the Secretary's findings with a preliminary order providing [relief]." *Id.* If the Secretary determines a violation has occurred, the defendant may then provide objections and request a hearing. *Id.* And "not later than 120 days after the date of conclusion of any hearing . . . the Secretary shall issue a final order." *Id.* § 2087(b)(3)(A). If, however, "the Secretary has not issued a final decision within 210 days after the filing of the complaint, or within 90 days after receiving a written determination, the complainant may bring an action at law or equity for *de novo* review." *Id.* § 2087(4). If there is an arbitration provision, however, that newly ripened claim and related claims may be subject to arbitration.

[141] *See* 9 U.S.C. § 2 ("A written provision . . . to settle by arbitration a controversy . . . shall be valid, irrevocable, and enforceable, save upon such grounds as exist at law or in equity for the revocation of any contract[.]").

[142] *See*, *e.g.*, Badgerow v. Walters, 596 U.S. 1, 8 (2022) ("The FAA requires . . . courts . . . to honor arbitration agreements[.]").

[143] *See Mandatory Arbitration Clauses Are Discriminatory and Unfair*, PUBLIC CITIZEN, HTTPS://WWW.CITIZEN.ORG/ARTICLE/MANDATORY-ARBITRATION-CLAUSES-ARE-DISCRIMINATORY-AND-UNFAIR [https://perma.cc/L78T-Z8BP].

[144] *See* Bissonnette v. LePage Bakeries Park St., LLC, 601 U.S. 246, 250 (2024) (quoting 9 U.S.C. § 1) ("It contains, however, an exception specifying that 'nothing herein contained shall apply to contracts of employment of seamen, railroad employees, or any other class of workers engaged in foreign or interstate commerce.'").

[145] *See id.*

[146] Sarbanes-Oxley Act of 2002, Pub. L. No. 107-204, 116 Stat. 745 (2002).

[147] 15 U.S.C. § 78u-6(h)(1)(A)(iii).

arbitration agreement.

(2) Predispute arbitration agreements.—No predispute arbitration agreement shall be valid or enforceable, if the agreement requires arbitration of a dispute arising under this section.[148]

This latter revision makes sense given that AI is capable of the same type of harm that Congress sought to protect against in the Dodd-Frank Act. Congress passed the Dodd-Frank Act to protect against the "irresponsible risks" taken by Wall Street leading up to the 2008 financial crisis.[149] The anti-arbitration provision addressed a concern that unless employees could come forward with relevant information, Wall Street firms might continue to engage in conduct facilitating yet another financial crisis. But AI could also create another financial crisis on its own. Consider, for example, the AI-attributed flash crash that happened on June 15, 2024, when the stock market suddenly dropped, wiping out trillions of dollars in market value.[150] It is incredible that employees who become aware of like risks to the national and global economy could be deterred to act responsibly due to their individual employment agreements. All in all, the Right to Warn does not require drastic reforms, particularly in light of the risk of harm that AI presents. Accordingly, the CPSC should immediately begin work on crafting rules and standards for AI.

As a part of that effort, the CPSC should invoke the protections afforded by 15 U.S.C. § 2076(d) to call current and former employees, including those Signatories who are likely stonewalled by contractual provisions. Though the whistleblower protections afforded by the CPSIA at 15 U.S.C. § 2087 may already provide safe harbor for whistleblowers, any forthcoming rules should further promote the flow of information without reprisal. In addition to bolstering employee protections, Congress (and state legislatures) should act to limit the enforceability of NDAs for current and former employees in the AI space. Similarly, Congress should move to limit the use of arbitration agreements and to mirror those provisions found in the Sarbanes-Oxley Act as amended by the Dodd-Frank Act.

IV. THE RIGHT TO DECIDE: LABELS FOSTER CHOICE

While employees in the AI sphere may fear losing their jobs for voicing concerns, AI more generally poses threats to the labor market and the market for human creations. While consumers may wish to mitigate these threats through their personal choices, they cannot do so without access to information. A disclosure-based system that allows consumers to distinguish between AI and human creations will establish protective measures to promote transparency and stem consumer deception regarding related consumer preferences. Similar to the "Made in America" labeling

---

[148] 18 U.S.C. § 1514A(e)(1)-(2).

[149] The White House, *Wall Street Reform: The Dodd-Frank Act*, https://obamawhitehouse.archives.gov/economy/middle-class/dodd-frank-wall-street-reform [https://perma.cc/XP8W-KB82].

[150] *See* Jeyadev Needhi, *AI's Role in the 2024 Stock Market Flash Crash: A Case Study*, MEDIUM (July 8, 2024), https://medium.com/@jeyadev_needhi/ais-role-in-the-2024-stock-market-flash-crash-a-case-study-55d70289ad50 [https://perma.cc/VSL8-6PHL] ("Following the flash crash, regulatory bodies, including the [SEC] and the Commodity Futures Trading Commission (CFTC), launched investigations to understand the causes and identify measures to prevent future occurrences. The investigations revealed that while the economic reports acted as a trigger, the primary cause of the crash was the AI-driven trading algorithms that executed large-scale sell orders in response to minor market fluctuations.").

regime, the Federal Trade Commission and state legislatures should adopt "Made by AI" and "Human Centered" labels to provide consumers with the Right to Decide.

A. AI Threatens Job Loss

The growing use of AI raises concerns about volatility in the labor market. The following story illustrates the speed and harsh effect of displacement. In June 2024, a writer, himself subject to an NDA, came forward under the pseudonym Benjamin Miller.[151] Miller described that as recently as early 2023, he worked at a company where he was tasked with managing more than 60 writers and editors, writing and publishing blogs and articles.[152] The company informed Miller that it "wanted to use AI to cut down on costs."[153] Within one month, the company instituted an automated system using an AI platform.[154] No longer would the writers on Miller's team create their own ideas.[155] Instead, a member of Miller's team would enter a title into the AI system, and it would create an outline.[156] Members of Miller's team would then write their piece based on that outline.[157] Miller would review, revise, and publish those pieces.[158]

Shortly thereafter, the company informed Miller that ChatGPT would handle the writing process and proceeded to fire most of his team.[159] Those left behind would simply review the pieces ChatGPT wrote and revise those pieces to make them sound "more human."[160] But by 2024, the rest of Miller's team was terminated. As the only person left, Miller was now tasked with reviewing ChatGPT's pieces. What Miller once found to be "really engaging work" had become "repetitive and boring."[161] He now faced "always the exact same kind of edits" and "started to feel like [*he*] was the robot."[162]

Miller's story is not an isolated one. Researchers have repeatedly found a negative correlation between the introduction of AI and resulting job loss. For example, after analyzing nearly 1.4 million job posts from a leading online freelancing platform from July 2021 to July 2024, one study found "a 21% decrease in the weekly posts in" what the authors called "automation-prone jobs."[163] Amongst other roles, "automation-prone jobs" include writing, software, app, and web development, and engineering.[164] Those researchers' findings comport

---

[151] Thomas Germain, *AI took their jobs. Now they get paid to make it sound human*, BBC (June 16, 2024), https://www.bbc.com/future/article/20240612-the-people-making-ai-sound-more-human [https://perma.cc/P6CB-QH6V].
[152] *Id.*
[153] *Id.*
[154] *Id.*
[155] *Id.*
[156] *Id.*
[157] *Id.*
[158] *Id.*
[159] *Id.*
[160] *Id.*
[161] *Id.*
[162] *Id.* (emphasis added).
[163] Ozge Demirci et al., *Research: How Gen AI Is Already Impacting the Labor Market*, HARV. BUS. REV. (Nov. 11, 2024), https://hbr.org/2024/11/research-how-gen-ai-is-already-impacting-the-labor-market [https://perma.cc/SP85-4QZA].
[164] *Id.* Of note, the authors found a 30.37% decrease in the number of writing jobs posted and a 20.62% decrease in the number available for software, app, and web development. *Id.*

with those by economists at Goldman Sachs who analyzed databases "detailing the task content of over 900 occupations" and found that "roughly two-thirds of U.S. occupations are exposed to some degree of automation by AI" with "roughly a quarter to as much as half of their workload" subject to replacement.[165] Aware of those dangers, well-organized workers have begun pushing for greater protection across several industries from the auto industry,[166] to healthcare,[167] to Hollywood script writing,[168] to video games.[169]

B. Companies Using AI Face Consumer Backlash

In addition to workers, consumers are also concerned about the effects of AI's rapid and extensive integration. Recently, the Coca-Cola Company ("Coca-Cola") stirred consumer backlash when it released its holiday commercial on November 18, 2024.[170] Coca-Cola's "The Holiday magic is coming" commercial features large red Coca-Cola trucks delivering the soft drink, polar bears in Christmas sweaters, and a flying Santa's sleigh.[171] The ad was meant to harken back to Coca-Cola's 1995 commercial, "Holidays are Coming,"[172] using nostalgia to drive sales.[173] Instead of featuring human actors and real trucks, as employed by human creators, however, the commercial was fully generated by AI.[174] Coca-Cola's European Chief Marketing Officer, Javier Meza, cited efficiency as a reason for using AI.[175]

Consumers readily understood Coca-Cola prioritized efficiency. Consumers viewed the commercial as a move to displace human artists, and some expressed a preference to boycott the brand.[176] Many consumers expressed their outrage online through comments reflecting negative consumer perception include: "You're a multi-billion dollar company. Hire some animators,"[177]

---

[165] *See* Goldman Sachs, *supra* note 8.

[166] Bryan Levine, *Strikes in the Age of Automation and AI: How HR Can Prepare for the Future*, SHRM (Oct. 10, 2024), https://www.shrm.org/topics-tools/employment-law-compliance/strikes-in-the-age-of-automation-and-ai--how-hr-can-prepare-for- [https://perma.cc/RYL8-Q7VH].

[167] Dave Pearson, *Overheard around the Kaiser nurses' protest over AI in healthcare*, HEALTH EXEC (Apr. 22, 2024), https://aiin.healthcare/topics/artificial-intelligence/overheard-around-kaiser-nurses-protest-over-ai-healthcare [https://perma.cc/D4GH-BXWW].

[168] Dani Anguiano & Lois Beckett, *How Hollywood writers triumphed over AI-and why it matters*, THE GUARDIAN (Oct. 1, 2023, at 7:00 ET), https://www.theguardian.com/culture/2023/oct/01/hollywood-writers-strike-artificial-intelligence [https://perma.cc/PA3X-HK2M].

[169] Marc Cieslak, *Video games strike rumbles on in row over AI*, BBC (July 31, 2024), https://www.bbc.com/news/articles/czvxz11gl57o [https://perma.cc/W6DY-H8TP].

[170] Greta Cross, *Coca-Cola's 'real magic' holiday ad, made with artificial intelligence sparks backlash*, USA TODAY (Nov. 21, 2024), https://www.usatoday.com/story/tech/2024/11/21/coca-cola-ai-holiday-commercial-backlash/76477569007/ [https://perma.cc/YX87-VCXG].

[171] *See* Bruna Horvath, *Coca-Cola causes controversy with AI-made ad*, NBC (Nov. 18, 2024, at 17:44 ET), https://www.nbcnews.com/tech/innovation/coca-cola-causes-controversy-ai-made-ad-rcna180665 [https://perma.cc/VM8W-X5KP].

[172] *See* Things off Tape, *1995 Coca Cola Christmas Advert 1 (Holidays Are Coming)* (YouTube, Dec. 15, 2023), https://www.youtube.com/watch?v=X13N-Bx17Oc [https://perma.cc/AP5B-CEHE]; *Coca-Cola | Holidays Are Coming* (YouTube, Nov. 20, 2023), https://www.youtube.com/watch?v=Yy6fByUmPuE [https://perma.cc/D4SM-EV4U].

[173] *See* Frank Landymore, *Coke's AI Commercial for the Holidays Has Us Wondering If We Live In a Fallen World*, FUTURISM (Nov. 16, 2024, at 6:00 ET), https://futurism.com/the-byte/coke-hideous-ai-commercial [https://perma.cc/3L9E-HVML].

[174] *See id.*

[175] *See id.*

[176] *See* LLLLITL, *Coca-Cola - Secret Santa (AI-Generated Christmas Ad 2024*) (YouTube, Nov. 15, 2024), https://www.youtube.com/watch?v=IQWUKWM2JrQ [https://perma.cc/95FG-SUPQ].

[177] Comment, @LimusTG (Nov. 20, 2024), *on* LLLLITL, *Coca-Cola - Secret Santa (AI-Generated Christmas Ad 2024*) (YouTube, Nov. 15, 2024), https://www.youtube.com/watch?v=IQWUKWM2JrQ [https://perma.cc/95FG-SUPQ]

"No animators were employed in the making of this advertisement,"[178] "Nothing embodies the generous Christmas spirit quite like cutting costs through heartless ai in an attempt to please the shareholders,"[179] "Multi billion dollar company who doesn't want to pay actors or respect true art . . . very in tune with the holiday spirit,"[180] "Seeing this cured my coke addiction. I'm hoping this ad makes them lose a lot more then the cost of making a real advertisement,"[181] "100% AI. 100% drinking Pepsi in 2025,"[182] "[T]his is just soulless. [I] will definitely not buy any coca cola in the future,"[183] and "I'm just one man and my money's not even a drop in your bottom line, but after this my money is no longer in your bottom line."[184]

Those consumers' reactions are not surprising. Although AI may generate technically impressive work products, study[185] after study[186] confirms a preference for human creations over AI creations upon learning that a creation is made by AI. Indeed, there is an indication that consumers are willing to pay more for human creations,[187] prefer engaging with human service providers,[188] and are more likely to take our business elsewhere when faced with AI-generated mistakes.[189] Such findings are consistent with longstanding scholarship concerning consumer preferences showing that consumers "'may be heavily influenced by information regarding the manner in which goods are produced.'"[190] The distinction between AI and human creations is crucial to limit consumer deception and promote consumer choice based on consumer preferences regarding production and quality.[191]

---

[178] Comment, @thompsonhonorato (Nov. 20, 2024), https://www.youtube.com/watch?v=IQWUKWM2JrQ [https://perma.cc/95FG-SUPQ].

[179] Comment, @gemmounain2190 (Nov. 20, 2024), *on* LLLLITL, *Coca-Cola - Secret Santa (AI-Generated Christmas Ad 2024*) (YouTube, Nov. 15, 2024), https://www.youtube.com/watch?v=IQWUKWM2JrQ [https://perma.cc/95FG-SUPQ].

[180] Comment, @NolanAronson08 (Nov. 20, 2024), *on* LLLLITL, *Coca-Cola - Secret Santa (AI-Generated Christmas Ad 2024*) (YouTube, Nov. 15, 2024), https://www.youtube.com/watch?v=IQWUKWM2JrQ [https://perma.cc/95FG-SUPQ].

[181] Comment, @DragonFartXL (Nov. 21, 2024), *on* LLLLITL, *Coca-Cola - Secret Santa (AI-Generated Christmas Ad 2024*) (YouTube, Nov. 15, 2024), https://www.youtube.com/watch?v=IQWUKWM2JrQ [https://perma.cc/95FG-SUPQ].

[182] Comment, @UberManlyPirate (Dec. 1, 2024), *on* LLLLITL, *Coca-Cola - Secret Santa (AI-Generated Christmas Ad 2024*) (YouTube, Nov. 15, 2024), https://www.youtube.com/watch?v=IQWUKWM2JrQ [https://perma.cc/95FG-SUPQ].

[183] Comment, @Emmet3129 (Nov. 25, 2024), *on* LLLLITL, *Coca-Cola - Secret Santa (AI-Generated Christmas Ad 2024*) (YouTube, Nov. 15, 2024), https://www.youtube.com/watch?v=IQWUKWM2JrQ [https://perma.cc/95FG-SUPQ].

[184] Comment, @briancleggmasterofvillains1044 (Dec. 2, 2024), *on* LLLLITL, *Coca-Cola - Secret Santa (AI-Generated Christmas Ad 2024*) (YouTube, Nov. 15, 2024), https://www.youtube.com/watch?v=IQWUKWM2JrQ [https://perma.cc/95FG-SUPQ].

[185] Martin Ragot et al., *AI-generated vs. human artworks. A perception bias towards artificial intelligence?*, CHI Conference on Human Factors in Computing Systems, 1-10 (2020), https://doi.org/10.1145/3334480.3382892 [https://perma.cc/M44L-52LJ].

[186] Lucas Bellaiche et al., *Human Versus ai: Whether and Why We Prefer Human-Created Compared to AI-Created Artwork*, 8 COGNITIVE RSCH. PRINCIPLES AND IMPLICATIONS 42 (2023).

[187] *See id.* ("These studies demonstrate that people tend to be negatively biased against AI-created artworks relative to purportedly human-created artwork, and suggest that knowledge of human engagement in the artistic process contributes positively to appraisals of art."); *see* Ozge Demirci et al., *supra* note 163 ("Employer's willingness to pay for these automation-prone jobs also went up by 5.71%.").

[188] *One Bad AI Experience Could Drive Customers Away, Acquires BPO Study Warns*, BUS. WIRE (Sep. 5, 2024, at 9:07 ET), https://www.businesswire.com/news/home/20240905660730/en/One-Bad-AI-Experience-Could-Drive-Customers-Away-Acquire-BPO-Study-Warns ("Consumers are 2.5X more positive about their experience chatting with humans versus AI-powered bots.") [https://perma.cc/N7QG-6T38].

[189] *See id.* ("70% of consumers would consider a different brand for their next purchase after just one frustrating experience with AI-supported customer service.").

[190] *See* Kwikset Corp. v. Superior Ct., 246 P.3d 877, 889 (2011) (quoting Kysar, *Preferences for Processes: The Process/Product Distinction and the Regulation of Consumer Choice* 118 HARV. L. REV. 525, 529 (2004)).

[191] *See id.* (citation omitted) ("Simply stated: labels matter. The marketing industry is based on the premise that labels matter, that consumers will choose one product over another similar product based on its label and various tangible and intangible qualities they may come to associate with a particular source.").

C. A Model: The "Made in America" Label

In analogous contexts, policymakers have established protective measures for businesses and consumers to address similar overriding concerns, such as preserving a labor market and delivering certain classes of products deemed superior. Consider, for example, federal and state law for "Made in U.S.A." or "Made in America" labeling. The California Supreme Court has recognized that "the 'Made in U.S.A.' label matters" to consumers due to a "range of motivations . . . from the desire to support domestic jobs, to beliefs about quality, to concerns about overseas environmental or labor conditions, to simple patriotism."[192] In turn, federal and state law prohibits use of the label unless such products meet certain requirements. At the federal level, the Federal Trade Commission ("FTC") regulates and enforces this label pursuant to its authority under 15 U.S.C. § 45.[193] The FTC permits use of that label only if "all or virtually all" of the products are made in the United States.[194] That "all or virtually all" standard is not a "bright line rule or percentage" but requires that any foreign parts employed in the production of the final product only comprise a "'negligible portion of the product's total manufacturing costs and [constitute an] insignificant part[] of the final product.'"[195]

California law is similar aside from two key features.[196] First, California law provides "two safe harbors for manufacturers," in light of the "evolving complexity of global trade."[197] Specifically, use of the label is permitted where "(a) foreign inputs compris[e] no more than 5 percent of the final wholesale value of the manufactured product or (b) foreign inputs compris[e] no more than 10 percent of the final wholesale value of the manufactured product *and* the manufacturer can show that it cannot produce nor obtain the foreign input from a domestic source."[198] Second, California's consumers may seek recourse for a violation under California's Legal Remedies Act ("CLRA"), which declares it an "unfair method[] of competition and unfair or deceptive act[] or practice[]" to "us[e] deceptive representations or designations of geographic origin in connection with goods or services."[199] The CLRA allows such consumers to join together in a class action[200] to recover damages, injunctive relief, restitution, punitive damages, and any other relief ordered by the court in addition to costs and attorney's fees.[201] Together, the CLRA's remedies represent a formidable tool for combating misleading advertising.

---

[192] *Id.* at 890.
[193] Labels on Products, 15 U.S.C. § 45(a) (2024).
[194] *See* Clark v. Citizens of Human., LLC, 97 F. Supp. 3d 1199, 1205 (S.D. Cal. 2015) (quoting "Made in USA" and Other U.S. Origin Claims, 62 Fed. Reg. 63756).
[195] *See id.*
[196] *See* Cal. Bus. & Prof. Code § 17533.7(a) ("It is unlawful for any person . . . to sell or offer for sale in this state any merchandise on which merchandise or on its container there appears the words 'Made in U.S.A., 'Made in America,' 'U.S.A,' or similar words in the merchandise or any article, unit, or part thereof, has been entirely or substantially made, manufactured, or produced outside of the United States.").
[197] Hood v. Handi-Foil Corp., No. 24-cv-02373-RS, 2024 WL 4008711, at *2 (N.D. Cal. Aug. 29, 2024).
[198] *Id.*
[199] Cal. Civ. Code § 1770(a)(4).
[200] *Id.* § 1781.
[201] *Id.* § 1780.

D. The "Made by AI" and "Human Centered" Labels

Here, policymakers should establish measures to avoid consumer deception and promote consumer choice. One approach includes regulating labels to distinguish between AI and human creations. "Made by AI" and "Human Centered" labels each present their own framework and corresponding benefits. Each label would serve to facilitate transparency in the purchasing process, informing consumers' purchasing decisions and helping them to differentiate between goods and services offered by AI compared to those offered by humans. That would further serve the purpose of safeguarding a market for human creations, including the continued employment of creators and service providers at risk of rapid displacement.

In the context of the "Made by AI" framework, creations generated by AI would be *required* to prominently display that label. Some steps have already been taken on that front, but are plainly insufficient in the consumer context. President Biden's EO 14110, for example, called on several federal agencies to make recommendations concerning "reasonable steps to watermark or otherwise label output from generative AI[.]"[202] Although a step in the right direction, as outlined above, there are problems with relying too heavily on executive orders to effect change. Additionally, EO 14110 did not *require* industries to adopt that label and that label can be subject to tampering.[203]

Several states have also implemented, or are considering implementing, mandatory disclosures. Colorado, for example, recently expanded its Consumer Protection Act, mandating that consumer-facing AI be accompanied by a "disclosure to each consumer who interacts with the [AI] system that the consumer is interacting with an [AI] system," unless it is obvious that such a system is an AI system.[204] California has also made headway with the California AI Transparency Act.[205] That Act pertains to certain providers[206] and requires, in part, that those providers give users "an option to include a manifest disclosure in image, video, or audio content, or content that is any combination thereof" created by AI that "identifies content as AI-generated content and is clear, conspicuous, appropriate for the medium of the content, and understandable to a reasonable person."[207] New York has also considered Senate Bill 9450, which would require a conspicuous disclosure that the user is interacting with an AI system.[208]

Those state laws or bills do not fully account for AI's potential and assume that all AI content operates and remains online. Such a limited approach does not stop businesses from, for

---

[202] Exec. Order No. 14,110, 88 Fed. Reg. 75191, (Nov. 1, 2023) § 10.1(b)(viii)(C); *see id.* § 3(gg) ("The term 'watermarking' means the act of embedding information, which is typically difficult to remove, into outputs created by AI—including into outputs such as photos, videos, audio clips, or text—for the purpose of verifying the authenticity of the output or the identity or characteristics of its provenance, modifications, or conveyance.").

[203] *See* Tate Ryan-Mosley, *The inside scoop on watermarking and content authentication*, MIT TECH. REV. (Nov. 6, 2023), https://www.technologyreview.com/2023/11/06/1082996/the-inside-scoop-on-watermarking-and-content-authentication/.

[204] Colo. Rev. Stat. Ann. § 6-1-1704(1)-(2) (2025).

[205] California AI Transparency Act, 2024 Cal. Adv. Legis. Serv. Ch. 291 (West).

[206] *See id.* § 1 ("'Covered provider' means a person that creates, codes, or otherwise produces a generative artificial intelligence system that has over 1,000,000 monthly visitors or users and is publicly accessible within the geographic boundaries of the state.").

[207] *See id.*

[208] S. 9450, 246th Leg., 2023-2024 Reg. Sess. (N.Y. 2024).

example, printing creations (including with 3D printers) and selling them in stores as their own creation free of any disclosure concerning AI's involvement.[209] Nor does it prevent businesses from employing a fully automated workforce, for example, to design and manufacture automobiles with little to no human input without a disclosure.[210] Nonetheless, in each of those cases, consumers have a right to access such information. Yet current laws fail to consider that shortfall. Whereas the "Made by AI" label would serve as a *mandatory* disclosure, the "Human Centered" label would serve as an *optional* disclosure.

Presumably, as AI-generated creations continue to saturate the market,[211] human creations may become harder to find, increasing their desirability. Indeed, there is already an indication that the mere involvement of AI in the creative process reduces the perceived value of the final product.[212] A "Human Centered" label will provide automation prone workers with an opportunity to continue their trades with the knowledge that consumers will specifically seek out their creations from a desire to support human jobs, beliefs about quality, or from simple humanistic preferences. It does not matter whether the objective quality of an AI-generated creation is comparable.[213] The harm results from the consumer purchasing, or paying more than they otherwise would have, but for the misrepresentation (by omission or otherwise).[214]

Like California's "Made in America" statutory framework, a "Human Centered" labeling regime should account for the fact that there may be varying degrees of "purity." For example, artists may use AI at different stages of production, such as during brainstorming or in

---

[209] *See*, *e.g.*, Image posted by u/Plushy_Axolotl, REDDIT (r/joannfabrics), *Why is Joanns using Ai generated art on their fabric? Kind disappointed ngl*, (2024), https://www.reddit.com/r/joannfabrics/comments/1ef16e2/why_is_joanns_using_ai_generated_art_on_their/ [https://perma.cc/3X3X-S629].

[210] *See* Jake Lingeman, *Artificial Intelligence Is Changing the Way Cars Are Made*, NEWSWEEK (Nov. 3, 2023, at 6:00 ET), https://www.newsweek.com/artificial-intelligence-changing-way-cars-are-made-1835913 [https://perma.cc/27LW-SHZC].

[211] Sam O'Brien, *AiArt: Why Some Artists are Furious About AI-Produced Art*, IEEE COMPUT. SOC'Y (Nov. 29, 2023) https://www.computer.org/publications/tech-news/trends/artists-mad-at-ai [https://perma.cc/AU7T-37MD].

[212] *See* Uwe Messer, *Co-creating art with generative artificial intelligence: Implications for artworks and artists*, COMPUTS. IN HUM. BEHAV.: ARTIFICIAL HUMS., (2024), https://doi.org/10.1016/j.chbah.2024.100056 [https://perma.cc/9NJL-L89W] ("The results show that co-created art is less liked and recognized, especially when AI was used in the implementation stage. While co-created art is perceived as more novel, it lacks creative authenticity, which exerts a dominant influence.").

[213] *See*, *e.g.*, Kwikset Corp. v. Superior Ct., 246 P.3d 877, 890 (Cal. 2011) ("For each consumer who relies on the truth and accuracy of a label and is deceived by misrepresentations into making a purchase, the economic harm is the same: the consumer has purchased a product that he or she paid more for than he or she otherwise might have been willing to pay if the product had been labeled accurately. This economic harm—the loss of real dollars from a consumer's pocket—is the same whether or not a court might objectively view the product as functionally equivalent. A counterfeit Rolex might be proven to tell the time as accurately as a genuine Rolex and in other ways be functionally equivalent, but we do not doubt the consumer (as well as the company that was deprived of a sale) has been economically harmed by the substitution in a manner sufficient to create standing to sue. Two wines might to almost any palate taste indistinguishable—but to serious oenophiles, the difference between one year and the next, between grapes from one valley and another nearby, might be sufficient to carry with it real economic differences in how much they would pay. Nonkosher meat might taste and in every respect be nutritionally identical to kosher meat, but to an observant Jew who keeps kosher, the former would be worthless.").

[214] *See id.* at 890–91 ("A consumer who relies on a product label [can] challenge[] a misrepresentation . . . by alleging . . . that he or she would not have bought the product but for the misrepresentation. That assertion is sufficient to allege causation—the purchase would not have been made but for the misrepresentation. It is also sufficient to allege economic injury. From the original purchasing decision we know the consumer valued the product as labeled more than the money he or she parted with; from the complaint's allegations we know the consumer valued the money he or she parted with more than the product as it actually is; and from the combination we know that because of the misrepresentation the consumer . . . was made to part with more money than he or she otherwise would have been willing to expend, i.e., that the consumer paid more than he or she actually valued the product. That increment, the extra money paid, is economic injury and affords the consumer standing to sue.").

implementation.[215] The question becomes whether a creation thought up with the assistance of AI, but created by human labor is "Human Centered." The same question applies with a creation thought up by a human but implemented by AI. To address that concern, policymakers might consider varying tiers that correspond with the level of input. For instance, a final output that does not employ any AI may qualify for a gold "Human Centered" label while another that employs some AI may qualify for a silver label. Ultimately, any "Human Centered" framework should account for such nuances.

Following appropriate classifications, the "Made by AI" and "Human Centered" labels should be regulated by federal and state law via the Federal Trade Commission Act[216] and consumer protection statutes like California's CLRA. In the former case, the FTC should immediately promulgate rules and begin enforcement. In the latter case, consumers may seek recourse if (1) a creation was generated by AI but not labeled as such or (2) a creation was mislabeled as "Human Centered."

The rise of AI raises concerns not only about job loss, but also about the economic viability of human creations more generally. AI advancements are likely to outpace appropriate regulatory action, but a disclosure-based system will create a baseline for action and agency. Although the proposed labeling regime presents no silver bullet to the threats posed to a human workforce and human creations, it is important that consumers, at a minimum, have access to information to exercise their Right to Decide. The "Made by AI" and "Human Centered" framework provides a common-sense approach that is grounded in analogous statutes and ample judicial precedent.[217]

CONCLUSION

As there is no apparent end to the future of AI, there are significant concerns about its immediate and lasting impact. Critical information asymmetries enhance various risks, including risks that create disadvantages for the individual and society at large. As AI's power evolves, and the gaps in our knowledge grows, the importance of strong federal and state policy comes into focus. To effectively balance or counteract AI's negative effects, it is imperative that policymakers consider and act on measures to closely monitor AI's growth, facilitate dialogue by protecting those with non-public information, and promote consumer choice through conspicuous labels.

---

[215] *See* Messer, *supra* note 212.

[216] 15 U.S.C. §§ 41–48.

[217] *See Kwikset Corp*, 246 P.3d at 895–96 (reversing and remanding case, holding that plaintiffs sufficiently alleged claim based on "Made in America" label); *Clark*, 97 F. Supp. 3d at 1209 (denying motion to dismiss plaintiffs' "Made in America" claims); Baum v. J-B Wield Co., LLC, No. 19-cv-01718-EMC, 2019 WL 6841231, at *8 (N.D. Cal. Dec. 16, 2019) (same); Banks v. R.C. Bigelow, Inc., 536 F. Supp. 3d 640, 648 (C.D. Cal. 2021) (same); Kennedy v. Nat. Balance Pet Foods, Inc., No. 07-CV-1082 H(RBB), 2007 WL 2300746, at *4 (S.D. Cal. Aug. 8, 2007) (same).