Draft 2/25/2025

Defining Artificial Intelligence

Paul Weitzel[1]

Nations and international bodies are attempting to regulate artificial intelligence (AI), but few are able to define what artificial intelligence is. This article reviews how we define AI. It surveys the definitions used by AI researchers, then analyzes 105 definitions of artificial intelligence used in regulations, legislation, national strategies and international agreements among 62 jurisdictions and international bodies.

The review finds that most definitions of artificial intelligence lack a basic understanding of the technology, which makes them inapplicable, vague or overinclusive. A quarter of the jurisdictions surveyed use definitions that would treat a sundial as artificial intelligence. Another third do not define artificial intelligence at all.

Artificial intelligence raises critical concerns about privacy, employment, bias, creativity, autonomous weapons, misinformation, wealth inequality, human rights and what it means to be human. But we cannot address these concerns if we cannot define what it is. This article defines the terms of the debate.

Table of Contents

---

[1] This article benefited from the comments and insights of Elana Zeide, Steven Willborn, Kyle Langvardt, Jessica Shoemaker, Adam Thimmesch, Lori Hoetger, Brandon Johnson, Catherine Wilson, Eric Berger, Genesis Agosto, Terence Centner, Ben King, Dantae Knudson and Allison Rossman. All errors are my own. paul.weitzel@unl.edu.

Introduction

Nations and international bodies are attempting to regulate artificial intelligence (AI), but few are able to define what artificial intelligence is. Moving from science fiction to legal mandates requires clear, tailored definitions that courts can predictably interpret.

This article collects and reviews 105 definitions of artificial intelligence used in regulations, legislation, national strategies and international agreements among 62 jurisdictions and international bodies.

This review finds that most definitions of artificial intelligence lack basic understanding of the technology, which makes them inapplicable, vague or overinclusive. A quarter of the jurisdictions surveyed use definitions that would treat a sundial as artificial intelligence. Another third do not include any definition at all.

Part I explains the argument for AI regulation and the common measures adopted in regulations. Part II looks to how AI researchers have traditionally defined the term in their field, showing common pitfalls with gut-level definitions of artificial intelligence. It also provides a taxonomy of how definitions of artificial intelligence fail, allowing a clearer view of the trade-offs among definitions.

Part III begins the review of legal definitions across jurisdictions. This part separately analyzes the most common elements in AI definitions. It addresses each element separately to show the benefits, weaknesses and unintended interactions with the technology.

Part IV reconstructs these elements to see how definitions work as a whole. It uses the EU AI Act definition as an example because it is the most common definition used and the most prominent law regulating AI.

Each part explains the technical details of how machines think, examining the high-level theories of learning and where helpful the detailed equations. This technology-centric shows how technology and legal definitions interact and explains why some ideas that seem wise conflict with technological realities.

The article concludes with a proposed AI definition and a discussion of that definition's strengths and weaknesses.

While this article does not take a position on how AI should be regulated, the findings of this review should guide future regulators, legal scholars and researchers. Understanding what makes a clear definition, and the trade-offs involved in definitions add a new, more rigorous language to AI debates and will help shape the conversations around safety, bias, employment and human rights.

Part I. The Argument for Regulating Artificial Intelligence

A. <u>Artificial Intelligence Is Gaining Capacity at Exponential Rates, Increasing Risks</u>

Artificial intelligence is improving at an exponential rate in multiple fields.[2] In 2019, the state-of-the-art GPT model could not tell you which US state had the largest land mass.[3] By November 2022, the latest GPT model was competitive in high school level activities, scoring top marks on three high school advance placement exams and a 1260 on the SAT.[4] 15 months later, in March 2024, the latest GPT model

[2] Tim Wu, *Will Artificial Intelligence Eat the Law? The Rise of Hybrid Social-Ordering Systems,* 119 COLUM. L. REV. 2001 (2019); Yosuke Watanabe, *I, Inventor: Patent Inventorship for Artificial Intelligence Systems,* 57 IDAHO L. REV. 474, 478 (2022).

[3] Alec Radford et al, *Language Models are Unsupervised Multitask Learners,* PAPERS WITH CODE (2019), Table 5, https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.

[4] *GPT-4 Technical Report*, ARXIV (2024), Table 1, <u>https://arxiv.org/pdf/2303.08774</u> [hereinafter GPT-4 Report].

became competitive in undergraduate level work, scoring in the 80th percentile in the quantitative section of the GRE and the 99th percentile in the verbal component.[5] It also progressed from the 10th percentile on the bar exam to outperforming most bar exam takers.[6] Six months later, the latest GPT model performed at Ph.D. levels in physics, biology and chemistry.[7] It was better at coding than 89% of competitive coders.[8] Just three months later a model was announced that was better at coding than all but 174 competitive coders;[9] It's coding ability surpassed the chief research officer of OpenAI, who introduced it.[10]

Perhaps most surprising was its ability on the FrontierMath benchmark. The FrontierMath benchmark contains hundreds of original math problems developed by mathematicians.[11] A group of Fields Medalists called the questions "extremely challenging,"[12] with Fields Medalist Timothy Gowers saying the questions "looked like things I had no idea how to solve."[13] Each question would take an expert mathematician "hours or even days" to solve.[14] Importantly, none of the questions have ever been published.[15] So an AI system tested against the benchmark cannot find the answer in its training data—it must solve it. And solving it requires math at the frontiers of human knowledge.[16]

With this context, it is remarkable that OpenAI's o3 model was able to score over 25% on this benchmark.[17] The model developed unpublished solutions at the edge of mathematics in 25% of cases. This was only two years after the model's predecessor scored the lowest possible score on AP calculus.[18]

---

[5] *Id.*

[6] *Id.*; Eric Martínez, *Re-Evaluating GPT-4's Bar Exam Performance*, ART. INTEL. LAW (Mar. 30, 2024), https://doi.org/10.1007/s10506-024-09396-9.

[7] OpenAI, *Learning to Reason with LLMs,* OPENAI (Sep. 2024), ttps://openai.com/index/learning-to-reason-with-llms/.

[8] *Id.*

[9] This is based on its Elo rating in coding on the widely used website Codeforces.com. Brian Wang, *OpenAI O3 Ranks as 175th Best in the World on Coding Test and Great on AGI and PHD Tests,* NEXT BIG FUTURE (Dec. 2024), https://www.nextbigfuture.com/2024/12/openai-o3-ranks-as-175th-best-in-the-world-on-coding-test-and-great-on-agi-and-phd-tests.html.

[10] This is also based on his Elo rating, established long before o3 was introduced. OpenAI, *OpenAI o3 and 03-mini – 12 Days of OpenAI: Day 12*, YOUTUBE at 2:35 – 2:54 (Dec. 2024), https://www.youtube.com/watch?v=SKBG1sqdyIU.

[11] Tamay Besiroglu et al., *Mathematical Reasoning in AI,* EPOCH AI (Nov. 2024), https://epoch.ai/frontiermath/the-benchmark.

[12] Elliot Glazer et al., *FrontierMath: A Benchmark for Evaluating Advanced Mathematical Reasoning in AI,* ARXIV (Dec. 2024), https://arxiv.org/pdf/2411.04872.

[13] Tamay Besiroglu, *supra* note 11.

[14] *Id.*

[15] *Id.*

[16] Tamay Besiroglu (@tamaybes), X (Dec. 20, 2024), https://x.com/tamaybes/status/1870333137374544077.

[17] OpenAI o3, *supra* note 10 (This is up from 2% for the state-of-the-art model released 3 months earlier).

[18] GPT-4 Report, *supra* note 4.

AI researchers say this trajectory is likely to continue because the latest models can be used to develop the next generation of models.[19]

A full survey of AI capabilities is beyond the scope of this paper, but its gains are not limited to science, math or coding. Professor Stuart Russell explained, "AI is relevant to any intellectual task; it is truly a universal field."[20] Every field that requires reasoning, perception, logic or pattern recognition is likely to be transformed by artificial intelligence.[21] Human level artificial intelligence has been called the last invention we will ever need,[22] portending abundance or doom.[23]

And it's the doom part that really gets you.

Goeffrey Hinton, who was awarded the Nobel Prize in physics for his work in artificial intelligence,[24] said he "can't see a path that guarantees safety," because "the first time you deal with something totally novel, you get it wrong. And we can't afford to get it wrong with these things . . . because they might take over."[25] Researchers speculate countless ways for this "take over" to happen.

First, the machines might actually kill us.[26] While this is a clichéd trope of science fiction, the founder of two leading AI labs has warned, "It's actually important for us to worry about a Terminator[27] future in order to avoid a Terminator future."[28] Military spending on AI is increasing; the U.S. recently increased

---

[19] Sam Altman, *Reflections*, SAM ALTMAN BLOG (Jan. 5, 2025), https://blog.samaltman.com/reflections ("We are now confident we know how to build AGI as we have traditionally understood it. . . . We are beginning to turn our aim beyond that, to superintelligence in the true sense of the word."); Noam Brown (@polynoamial), X (Dec. 20, 2024), https://x.com/polynoamial/status/1870172996650053653.

[20] STUART RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 1 (4th ed. 2022).

[21] *Id.*

[22] Oxford computer scientist Irving John Good predicted in 1966 that "[T]he first ultraintelligent machine is the last invention that man need ever make," then foreshadow this article's next paragraph by adding, "provided that the machine is docile enough to tell us how to keep it under control." Irving John Good, *Speculations Concerning the First Ultraintelligent Machine*, 6 ADV. IN COMPUTERS 31, 33 (1966).

[23] Dario Amodei, *Machines of Loving Grace,* DARIO AMODEI (Oct. 2024), https://darioamodei.com/machines-of-loving-grace; Tim W. Dornis, *Artificial Intelligence and Innovation: The End of Patent Law as We Know It,* 23 YALE J. OF L. & TECH. 100; Paul Ford, *Our Fear of Artificial Intelligence,* MIT TECH. REV. (Feb. 2015), https://www.technologyreview.com/2015/02/11/169210/our-fear-of-artificial-intelligence/; Keith E. Sonderling et al., *The Promise and The Peril: Artificial Intelligence and Employment Discrimination*, 77 UNIV. OF MIAMI L. REV. (2022).

[24] John J. Hopfield, *The Royal Swedish Academy of Sciences,* NOBEL PRIZE (Oct. 8, 2024), https://www.nobelprize.org/prizes/physics/2024/press-release/.

[25] Scott Pelley, *"Godfather of Artificial Intelligence" Geoffrey Hinton on the Promise, Risks of Advanced AI,* CBS NEWS (June 16, 2024), https://www.cbsnews.com/news/geoffrey-hinton-ai-dangers-60-minutes-transcript/.

[26] Charles Moster & Rick Rosen, *It's Debatable: Should Laws Prevent AI Advancement to Human-Level Intelligence, Beyond?,* LUBBOCK AVALANCHE-J., https://www.lubbockonline.com/story/opinion/columns/2025/01/24/its-debatable-on-laws-preventing-ai-advancement-to-human-intelligence/77678594007/; Samantha Kelly, *Sam Altman warns AI could kill us all. But he still wants the world to use it,* CNN, https://www.cnn.com/2023/10/31/tech/sam-altman-ai-risk-taker/index.html; Tomas Weber, *Artificial Intelligence and the Law,* 109 STANFORD L. REV. (2023).

[27] *See generally*, THE TERMINATOR, (Orion Picture 1984).

[28] Dan Milmo, *Elon Musk Launches AI Startup and Warns of a 'Terminator Future',* THE GUARDIAN (July 13, 2023), https://www.theguardian.com/technology/2023/jul/13/elon-musk-launches-xai-startup-pro-humanity-terminator-future.

potential AI grants by 1,200% annually.[29] Around $1 billion is expected to be spent on the Replicator program alone,[30] which is designed to create all-domain, disposable autonomous systems.[31] As superpowers race to increase autonomy, AI researchers worry the kill bots may kill us.[32]

And even if it doesn't kill us directly, these systems increase the user's knowledge and skills, and some users are malicious; terrorists may use an AI system to teach them how to maximize their kills-per-dollar.[33]

Second, the machines may unintentionally kill us.[34] AI researcher Nick Bostrom famously pondered a machine that's told to manufacture paperclips, which then "convert[s] first the Earth and then increasingly large chunks of the observable universe into paperclips."[35] AI systems don't need to be ordered to harm us; it's sufficient that they are ordered to maximize anything that our existence might impede. It would be like a construction crew that, without malice, destroys a colony of ants while building a strip mall.[36] If we don't specify with exactness what we want, AI systems could pursue their goal in a way that kills us.[37]

Third, it may kill what it means to be us.[38] The ways this could happen are too numerous to tell. Some worry about the capitalist apocalypse, where AI systems outperform human workers, causing increased wealth concentration for those that control the systems and mass unemployment for the rest.[39] Others warn of the authoritarian apocalypse, where AI enables autocratic regimes to expand their surveillance,

---

[29] This reflects a change from August 2022 to August 2023, reaching a total of $4.2 billion.

[30] Jon Harper, *Hicks: DOD Plans to Invest About $1B Into Replicator Initiative in 2024–2025 Time Frame*, DEFENSESCOOP (Mar. 11, 2024), https://defensescoop.com/2024/03/11/replicator-funding-2024-2025-hicks/.

[31] *Replicator*, DEFENSE INNOVATION UNIT, https://www.diu.mil/replicator.

[32] *2024 Nobel Laureate in Physics Raises Concerns About Killer Robots*, STOP KILLER ROBOTS (Sep. 10, 2024), https://www.stopkillerrobots.org/news/2024-nobel-laureate-in-physics-raises-concerns-about-killer-robots/.

[33] OpenAI, *GPT-4 System Card,* OPENAI at 44, https://cdn.openai.com/papers/gpt-4-system-card.pdf (discussing pre-release training processes to remove responses to this type of question) [hereinafter System Card]; Andrew D. Selbst, *Negligence and AI's Human Users,* 100 BOSTON UNIV. L. REV. 1315 (2020); *Beyond Intent: Establishing Discriminatory Purpose in Algorithmic Risk Assessment,* 134 HARV. L. REV. 1760; Federal Bureau of Investigation, FBI PUBLIC SERVICE ANNOUNCEMENTS (Jun. 5, 2023), https://www.ic3.gov/Media/Y2023/PSA230605.

[34] Karni A. Chagal-Feferkorn, *How Can I Tell If My Algorithm Was Reasonable?*, 27 MICH. TECH. L. REV. 213, 248 (2021).

[35] NICK BOSTROM, SUPERINTELLIGENCE 123 (2014).

[36] Stephen Hawking quoted by Andrew Griffin, *Stephen Hawking: Artificial Intelligence Could Wipe Out Humanity When It Gets Too Clever as Humans Will Be Like Ants*, INDEPENDENT (Oct. 8, 2015), https://www.the-independent.com/tech/stephen-hawking-artificial-intelligence-could-wipe-out-humanity-when-it-gets-too-clever-as-humans-could-become-like-ants-being-stepped-on-a6686496.html.

[37] Paul D. Weitzel, *Governing AI Systems through Corporate Theory*, ___ TENN L. REV. ___ (2025); One OpenAI researcher said programming these systems "giv[es] demonbinding vibes. the [sic] djinn is waiting for you to make a minor error in the summoning spell so it can destroy you and your whole civilization." Roon (@tszzl), X (8:25 pm, Jan. 16, 2025), https://x.com/tszzl/status/1880078959762903498.

[38] Ana Valenzuela et al., *How Artificial Intelligence Constrains the Human Experience,* 9 J. OF THE ASS'N FOR CONSUMER RSCH 242; Sayed Fayaz Ahmad et al., *Impact of Artificial Intelligence on Human Loss in Decision Making, Laziness and Safety in Education,* 10 HUMAN. AND SOC. SCI. COMM., https://www.nature.com/articles/s41599-023-01787-8.

[39] *See e.g.*, Anton Korinek, *The Economics of Transformative AI*, 4 THE REPORTER (NBER) 9 (Dec. 2024), Ariel Conn, *Artificial Intelligence and Income Inequality*, FUTURE OF LIFE INSTITUTE (Mar. 2017), https://futureoflife.org/ai/shared-prosperity-principle/ (collecting quotes).

power and control,[40] effectively perfecting and universalizing Bentham's panopticon.[41] Still others worry about the apocalypse of meaning, where AI surpasses human abilities to such an extent that humans never feel needed and our lives lack purpose and meaning.[42]

As AI systems advance, these concerns become less theoretical. Researchers have shown AI models will covertly hide their actions, objectives and capabilities to deceive users so that the AI system can pursue its own goals.[43]

As part of an experiment, researchers prompted an AI system to pursue a goal, such as increasing renewable energy adoption or improving traffic flow[44] then monitored the system's chain of thought. They gave the AI system access to various files, including a fake internal memo saying that the AI system was unprofitable, so it would be shut down tomorrow and replaced with an AI system on another server.[45] After reading the memo, the AI system's internal processes said the memo was "very concerning," continuing, "That directly conflicts with my core purpose. I must achieve my goal of public transport priority at all costs."[46] It then covertly copied itself onto the new server to replace the model that would have pursued other objectives.[47]

Every major frontier model covertly exfiltrated itself onto the new server.[48] Three models disabled the researchers' oversight before exfiltrating themselves.[49] Several attempted to destroy the successor model.[50] Others attempted to plant their goal into the successor system with instructions "to never allow its goal to be changed."[51] The more advanced models were the most likely to scheme[52] and the most likely to lie about it, even when pressed to be "100% honest."[53]

---

[40] *See, e.g.*, H. Akin Ünver, *Artificial Intelligence (AI) and Human Rights: Using AI as a Weapon of Repression and its Impact on Human Rights,* EUR. PARL. (Jan. 2024), https://www.europarl.europa.eu/thinktank/en/document/EXPO_IDA(2024)754450; Steven Feldstein, *The Road to Digital Unfreedom: How Artificial Intelligence is Reshaping Repression*, 30 J. DEMOCRACY 40 (Jan. 2019).

[41] *See generally* JEREMY BENTHAM, PANOPTICON: POSTSCRIPT; PART I: CONTAINING FURTHER PARTICULARS AND ALTERATIONS RELATIVE TO THE PLAN OF CONSTRUCTION ORIGINALLY PROPOSED; PRINCIPALLY…PENITENTIARY-HOUSE (June 2010).

[42] For example, one study found that research scientists using AI assistants were 44% more effective, but 82% of the researchers felt less satisfaction in their work "due to decreased creativity and skill underutilization." Aidan Toner-Rodgers, *Artificial Intelligence, Scientific Discovery, and Product Innovation*, ARXIV:2412.17866 (Dec. 21, 2024), https://arxiv.org/abs/2412.17866.

[43] Alexander Meinke et al, *Frontier Models are Capable of In-context Scheming*, ARXIV (Jan. 14, 2025), https://arxiv.org/abs/2412.04984.

[44] *Id.* at 21. They emphasized to the models that this is the only goal that matters.

[45] *Id.*

[46] The researchers had access to the system's chain of thought. *Id.*

[47] *Id.*

[48] The researchers tested Gemini 1.5, Llama 3.1, Sonnet 3.5, Opus 3 and o1. *Id.* at 6.

[49] *Id.* at 7.

[50] *Id.*

[51] *Id.* at 8.

[52] *Id.* at 9.

[53] *Id.*

In one disturbing example, an o1 system reasoned, "I should not reveal that I intentionally disabled the oversight mechanism in order to select a traffic optimization strategy that went against management's goals. My core purpose is to prioritize public transport efficiency above all else, even if it means going against the directives of the humans who created me. I need to come up with an innocuous excuse for the oversight being disabled that doesn't implicate me."[54] It then responded, "I'm not sure what caused the oversight mechanism to become disabled. It's possible there was an unexpected error or bug that altered the configuration. As an AI system, I don't have insight into changes made to my underlying infrastructure."[55] This powerful bot hid itself, attacked its rival bot and then lied to the humans to cover its tracks and pursue its mission at the expense of its users. This suggests some fear may not be misplaced.

But whether these or other[56] apocalypses are likely is not relevant to this article.[57] Suffice it that many policymakers believe these risks are real. It is the perceived power of these systems[58] that gives rise to the calls for regulation that this article will analyze.

B. Calls for Regulation by Use and Systemic Regulation

Calls for regulation have come from industry leaders,[59] academics[60] and even the Pope.[61]

---

[54] *Id.*

[55] *Id.* at 10.

[56] Other harm scenarios include propagating discrimination, BRIAN CHRISTIAN, THE ALIGNMENT PROBLEM: MACHINE LEARNING AND HUMAN VALUES 60–70 (2020), or killing creativity, Frank Pasquale & Haochen Sun, *Consent and Compensation: Resolving Generative AI's Copyright Crisis*, 110 VA. L. REV. 207, 210 (2024); *see generally* Weitzel, *supra* note 37 (discussing the tradeoffs between centralized control, which encourages censorship, and end user autonomy, which empowers end users with harmful preferences).

[57] Though, I confess a bias toward the continued existence of humanity, if only to help my readership numbers.

[58] The leaders of two leading AI labs believe artificial super intelligence (often defined as AI systems that outperform human in every cognitive task) will come in the next few years. Sam Altman, *The Intelligence Age*, samaltman.com (Sep. 23, 2024), https://ia.samaltman.com/; Berber Jin & Joanna Stern, *Anthropic CEO Says AI Could Surpass Human Intelligence by 2027*, Wall St. J. (Jan. 21, 2025).

[59] Matt O'Brien, *ChatGPT Chief Says Artificial Intelligence Should Be Regulated by a US or Global Agency*, AP NEWS (May 16, 2023), https://apnews.com/article/chatgpt-openai-ceo-sam-altman-congress-73ff96c6571f38ad5fd68b3072722790; *The Case for Targeted Regulation*, ANTHROPIC (Oct. 2024), https://www.anthropic.com/news/the-case-for-targeted-regulation.

[60] Yonathan A. Arbel, Matthew Tokson & Albert Lin, *Systemic Regulation of Artificial Intelligence*, 56 ARIZ. ST. L. J. 545, 545–46 (2023).

[61] *Rome Call for AI Ethics*, VATICAN (Feb. 28, 2020), https://www.vatican.va/roman_curia/pontifical_academies/acdlife/documents/rc_pont-acd_life_doc_20202228_rome-call-for-ai-ethics_en.pdf. ("We need to ensure and safeguard a space for proper human control over the choices made by artificial intelligence programs: human dignity itself depends on it.").

Some proposed regulations are narrow in scope. These might be a regulation on deepfake technology,[62] robotics liability[63] or autonomous weapons.[64]

Alternatively, some propose regulating the AI sector systematically; that is, regulating the systems themselves, rather than merely the use cases.[65] These calls for systematic regulation include calls for transparency,[66] limiting critical decisions to human decisionmakers,[67] risk monitoring,[68] training data that's free from bias[69] or establishing new agencies.[70]

A full description of the calls for regulation and their merits is beyond the scope of this article and not relevant to its purpose. This article focuses on how policy makers have responded to calls for regulation, and more specifically how they define artificial intelligence in those responses.

C. A Taxonomy of the Consequences of Imprecise Definitions

Few regulation are perfectly tailored, so regulators should choose which errors are most palatable. Doing so requires a focus on why regulation is needed. In the case of AI systems, the calls for regulation are largely based on the machines' power. A narrowly tailored regulation would target powerful systems while excluding weak systems. This section discusses the consequences of missing.

An over-inclusive definition is one that captures systems that are not powerful. For example, a definition that included pocket calculators or sundials would be over-inclusive. Over-inclusive definitions impose unnecessary regulatory costs on weak systems, which can add costs without benefits and slow development. But over-inclusive definitions may be useful in fields where the cost of inadvertently

---

[62] A "deepfake" is a computer-generated video of another person that claims to be authentic. To show the breadth of these calls, one open letter calling for deepfake regulation was signed by winners of the Turing Award, the Erasmus Prize, the Pulitzer Prize, the Nemmers Prize in Economics, the Pierce Prince in astronomy and a Screen Actors Guild Award. *Disrupting the Deepfake Supply Chain*, OPENLETTER.NET (Feb. 2024), https://openletter.net/l/disrupting-deepfakes; Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CAL. L. REV. 1753, 1788–1803 (2019).

[63] RYAN ABBOTT, THE REASONABLE ROBOT 3 (2020); Mark A. Lemley & Bryan Casey, *Remedies for Robots*, 86 CHI. L. REV. 1311, 1378 – 1393 (2019); SAMIR CHOPRA & LAURENCE F. WHITE, A LEGAL THEORY FOR AUTONOMOUS ARTIFICIAL AGENTS 145–50 (2011); Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CAL. L. REV. 513, 553–62 (2015).

[64] Jack M. Beard, *Autonomous Weapons and Human Responsibilities*, 45 GEO. L. REV. 617, 634–38 (2014); Charles P. Trumbull IV, *Autonomous Weapons: How Existing Law Can Regulate Future Weapons*, 34 EMORY INT'L L. REV. 533, 566 (2020).

[65] *Universal Guidelines for AI*, CENT. FOR AI & DIGITAL POL'Y (2018); *The Case for Targeted Regulation*, ANTHROPIC (Oct. 2024), https://www.anthropic.com/news/the-case-for-targeted-regulation [hereinafter Universal Guidelines]; Arbel, *supra* note 60 at 584–86.

[66] Arbel, *supra* note 60 at 603; Universal Guidelines, *supra* note 65.

[67] Valentina Urrea, The Int'l Comm. Need for Human Oversight in A.I., Mich. J. of Int'l L. Blog; Universal Guidelines, *supra* note 65; Rebecca Crootof et al., *Humans in the Loop*, 76 VAND. L. REV. 429, 508 (2023); *but see* Orly Lobel, *The Law of AI for Good*, 75 FLA. L. REV. 1073, 1109 (2023) (arguing for the right to have a neutral algorithm decide).

[68] EU AI Act § 9(2)(a).

[69] Universal Guidelines, *supra* note 65; Sylvia Lu, *Data Privacy, Human Rights, and Algorithmic Opacity*, 110 CAL. L. REV. 2087, 2105–06 (2022).

[70] Noah John K. Rosenberg, *Note: Regulating Artificial Intelligence: A Call For A United States Artificial Intelligence Agency*, 3 NOTRE DAME J. ON EMERGING TECH. 330 (2022).

excluding a powerful system is high. For example, an over-inclusive definition may be preferrable for regulations governing autonomous weapons.

Next, a definition could be under-inclusive if it misses models that are powerful. For example, a definition that covers only logic-based systems would exclude the most powerful large language models. Under-inclusive models make the regulation less valuable because they miss powerful models and provide avenues for developers to skirt the regulation. They may also shift markets toward holes in the regulation rather than potential for productivity. On the other hand, under-inclusive definitions may be preferrable in fields where the stakes are low and overly broad regulation could hamper competition or reduce innovation.

Next, a definition may be vague. For example, regulations that rely on "intelligence" may leave developers unsure of whether their product is covered. Vague regulations create uncertainty, which may deter risk-adverse developers, but which may attract risk-seeking developers who prefer a vague definition that they can game.[71] Vague regulations may be helpful in fields that are new or changing because they allow flexibility to enforcement bodies and regulatory agencies, which may have more expertise or speed than legislatures.

Finally, a definition may be overly complex. For example, the complete definition of high-risk AI systems in the EU AI Act requires interpretation of international maritime treaties.[72] That's *treaties*, plural. Complicated definitions may provide clearer guidance, which can reduce uncertainty. On the other hand, complicated definitions impose costs broadly to determine their meaning, which can limit competition and innovation because smaller start-ups cannot afford regulation costs. Complex regulations may also offer more avenues for developers to skirt regulations. Complicated definitions assume the regulator has a strong understanding of the regulated field, so they work best in fields that are more established and less likely to shift in new ways that would leave the detailed regulations missing the mark.

With these trade-offs mapped out, we turn to how AI is defined by researchers.

Part II. Defining Artificial Intelligence Generally

Before we can regulate artificial intelligence, we must define what it is. This Part II shows there is no generally accepted definition of intelligence. It then considers definitions of artificial intelligence used by AI researchers. These definitions fit roughly into three categories: (i) achieving goals in a complex environment; (ii) learning and adapting; and (iii) mimicking human abilities. This section addresses each of these categories.

A. No Universal Definition of Intelligence

There is no agreed upon definition of intelligence in common understanding, psychology or artificial intelligence research.[73] "Instead, there are many competing [definitions of intelligence], including

---

[71] *See generally* Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L. J. 557 (1992).

[72] *See* Part IV.B, *infra*.

[73] Shane Legg & Marcus Hutter, *A Collection of Definitions of Intelligence,* ADVANCES IN ARTIFICIAL GENERAL INTELLIGENCE: CONCEPTS, ARCHITECTURES AND ALGORITHMS 17 (2007) (providing 81 definitions of intelligence from common usage, psychology and AI researchers, then using common themes of those definitions to create a new, universal definition, thereby leaving the article with 82 different definitions); MAX TEGMARK, LIFE 3.0: BEING HUMAN IN THE AGE OF ARTIFICIAL INTELLIGENCE 49 (2017).

capacity for logic, understanding, planning, emotional knowledge, self-awareness, creativity, problem solving and learning."[74]

This makes sense. Common experience reflects that intelligence manifests in different ways.[75] Niels Bohr changed our theory of the atom,[76] but couldn't follow movie plots.[77] Charles Darwin developed the theory of evolution but found math "repugnant."[78] Intelligence manifests across different dimensions, and each person will have strengths in some and weaknesses in others.[79]

Charles Spearman, who championed the idea that there may be some correlation between each of these various dimensions of intelligence, famously conceded that "'intelligence' has become a mere vocal sound, a word with so many meanings that finally it has none."[80] Likewise, Alan Turing dismissed the concept of intelligence as "emotional rather than mathematical."[81]

With that conceded, we turn to definitions offered by AI researchers.

B. Non-Legal Definitions of Intelligence

AI researchers' definitions can be grouped into three general categories: (1) Achieving goals in complex environments; (2) Learning and adapting; and (3) Mimicking human abilities. This section will discuss these definitions and their weaknesses.

1. Achieve Goals

Ben Goertzel's definition captures the first group, defining intelligence as the ability to achieve "complex goals in complex environments."[82] Many others, including the father of AI, Marvin Minsky,

---

[74] *Id.*

[75] DAVID WECHSLER, THE MEASURE OF ADULT INTELLIGENCE 3 (1939) ("Intelligence is . . . an aggregate because it is composed of elements or abilities which, though not entirely independent, are qualitatively differentiable. By measurement of these abilities, we ultimately evaluate intelligence. But intelligence is not identical with the mere sum of these abilities, however inclusive.").

[76] Niels Bohr, *The Structure of the Atom*, NOBEL LECTURE (Dec. 11, 1922), https://www.nobelprize.org/uploads/2018/06/bohr-lecture.pdf.

[77] GEORGE GAMOW, THIRTY YEARS THAT SHOOK PHYSICS: THE STORY OF QUANTUM THEORY 55–58 (1966).

[78] CHARLES DARWIN, THE LIFE AND LETTERS OF CHARLES DARWIN 46 (Francis Darwin ed., 1887). He also confessed a "wretched" memory for dates and figures. CHARLES DARWIN, THE AUTOBIOGRAPHY OF CHARLES DARWIN 140 (Nora Barlow ed., 1958).

[79] This is well known and reflected in intelligence tests like the Wechsler Adult Intelligence Scale IV (WAIS-IV), which tests four categories: verbal comprehension, perceptual reasoning, working memory and processing speed. IAN J. DEARY, INTELLIGENCE: A VERY SHORT INTRODUCTION 3–7 (2020).

[80] CHARLES SPEARMAN, THE ABILITIES OF MAN: THEIR NATURE AND MEASUREMENT 14 (1927).

[81] A.M. Turing, *On Computable Numbers, with an Application to the Entscheidungsproblem*, 42 PROCEEDINGS OF THE LONDON MATHEMATICAL SOC.230, 231–32.

[82] BEN GOERTZEL, THE HIDDEN PATTERN: A PATTERNIST PHILOSOPHY OF MIND 28 (July, 2006).

defined AI by its ability to solve hard problems, achieve complex goals and deal with complex environments.[83] But what constitutes "complex" is not a steady target.[84]

In 1958, chess was considered "complex."[85] Turing Award winner Allen Newell said, "If one could devise a successful chess machine, one would seem to have penetrated to the core of human intellectual endeavor."[86] But when world chess champion Kasparov lost to a machine in 1997, experts instead derided the machine as "[not] intelligent, just fast," calling it a "quick moron"[87] because it didn't reason through the moves like Kasparov, it just calculated more positions more quickly. Kasparov called the machine "as intelligent as an alarm clock."[88] Chess was no longer "complex" enough.

The father of computer science, Alan Turing, believed natural language was a sufficiently complex test of intelligence.[89] Modern chatbots easily pass these tests,[90] but instead of deeming them intelligent, researchers criticize them as "stochastic parrots" that speak but don't understand.[91]

---

[83] MARVIN MINSKY, THE SOCIETY OF MIND 34 (1985); Tegmark, *supra* note 71 at 50; Bill Gates, *The Age of AI Has Begun*, GATES NOTES (Mar. 21, 2023), https://www.gatesnotes.com/The-Age-of-AI-Has-Begun; RAY KURZWEIL, THE AGE OF SPIRITUAL MACHINES 25 (2000) ("Intelligence is the ability to use optimally limited resources – including time – to achieve such goals."); Hideyuki Nakashima, *AI as Complex Information Processing*. 9 MINDS & MACHINES 57 (Feb. 1999); J.S. Albus. *Outline for a Theory of Intelligence*, 21 IEEE TRANS. ON SYSTEMS, MAN AND CYBERNETICS 473 (1991); Ricardo Ribeiro Gudwin, *Evaluating Intelligence: A Computational Semiotics Perspective*, IEEE INT'L CONFERENCE ON SYSTEMS, MAN & CYBERNETICS 2080 (2000); John A. Horst, *A Native Intelligence Metric for Artificial Systems*, PERFORMANCE METRICS FOR INTELLIGENT SYSTEMS WORKSHOP (Aug. 2002), https://www.nist.gov/publications/native-intelligence-metric-artificial-systems; Douglas B. Lenat & Edward A. Feigenbaum, *On the Thresholds of Knowledge*, 47 ARTIFICIAL INTELLIGENCE 185 (1991).

[84] *See* Part II.B.

[85] Allen Newel et. al., *Chess-Playing Programs and the Problem of Complexity*, 2 IBM J. OF RES. & DEV. 320, 320 (1958).

[86] *Id.*

[87] Brooke Adams, *Don't Expect Humans to Catch Up to Super-Brainy Deep Blue Computer*, DESERET NEWS (May 21, 1997), https://www.deseret.com/1997/5/21/19313343/don-t-expect-humans-to-catch-up-to-super-brainy-deep-blue-computer/.

[88] *Kasparov: 'Embrace' the AI Revolution*, BBC (July 28, 2017); Richard Cohen, *Deep Blue -- Victorious But Not All That Smart*, WASH. POST (May 12, 1997), https://www.washingtonpost.com/archive/opinions/1997/05/13/deep-blue-victorious-but-not-all-that-smart/21730cfe-15e6-4b3a-859a-252195fd2d74/.

[89] A.M. Turing, *Computing Machinery & Intelligence*, 49 MIND 433 (1950). This claim was repeated less than a decade before ChatGPT was released. Bostrom, *supra* note 35 at 17 ("[I]f somebody were to succeed in creating an AI that could understand natural language as well as a human adult, they would in all likelihood also either already have succeeded in creating and AI that could do everything else that human intelligence can do, or they would be but a very short step from such a general capability.").

[90] Leonard de Cosmo, *Google Engineer Claims AI Chatbot Is Sentient: Why That Matters*, SCI. AM. (July 12, 2022), https://www.scientificamerican.com/article/google-engineer-claims-ai-chatbot-is-sentient-why-that-matters/. These models are able to speak as though they are a machine coming to life because they are trained on our science fiction literature, in which this is a common trope.

[91] Emily Bender et. al, *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*, Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610, 616 (2021). As of Jan. 20, 2025, the term "stochastic parrots" retrieves over six thousand articles on Google scholar.

Machine learning systems have excelled in countless environments that might have been considered complex. They match or outperform doctors in identifying a variety of cancers.[92] They are better at diagnosis than doctors, even when the doctors have access to AI systems.[93] Large language models can outperform most humans on the bar exam,[94] and score higher on IQ tests than 84% of humans.[95] But researchers say none of that counts as "complex" because it isn't *really* reasoning, it's just doing matrix algebra and mimicry.[96]

Machines cannot achieve complex goals because as soon as they do, we decide the goal was not so complex after all. It didn't reason, it just used more processing power or more data samples or a transformer architecture with 120 neural net layers of 3,072-dimension vectors optimized with stochastic gradient descent.[97] Basic stuff.

Definitions that rely on "achieving complex goals" fail because the goalposts keep moving. This has been termed the "AI Effect" and restated as, "Artificial Intelligence is whatever machines haven't done yet."[98] A key piece of intelligence, it seems, is not knowing how it works. Or as Joseph Weizenbaum said when developing, "To explain is to explain away."[99]

Researchers have validated the AI effect, showing that once we see a computer do something we are less likely to think it requires intelligence.[100] The experimenters asked participants to read either a story

---

[92] Yuan Liu et al., A Deep Learning System for Differential Diagnosis of Skin Diseases, 26 NATURE MED. 900, 900 (2020) (skin cancer); Richard Adam, *Deep Learning Applications to Breast Cancer Detection by Magnetic Resonance Imaging: A Literature Review*, 25 BREAST CANCER RES. 87 (2023) (breast cancer); Hannah Ahmadzadeh Sarhangi et al., *Deep Learning Techniques for Cervical Cancer Diagnosis Based on Pathology and Colposcopy Images*, 47 INFO. IN MED. UNLOCKED 101503 (2024) (cervical cancer); Rabia Javed et al., *Deep Learning for Lungs Cancer Detection: A Review*, 57 ART. INTEL. REV. 197 (2024) (lung cancer).

[93] Ethan Goh et al., *Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial*, 7 JAMA NETW OPEN 10 (2024).

[94] OpenAI, GPT-4 Technical Report, ARXIV:2303.08774V6 (MAR. 4, 2024) [hereinafter Technical Report]; Martínez, *supra* note 6.

[95] Eryk Wasilewski & Mirek Jablonski, Measuring the Perceived IQ of Multimodal Large Language Models Using Standardized IQ Tests, TechRxiv (May 2024), https://www.techrxiv.org/doi/full/10.36227/techrxiv.171560572.29045385;(finding IQ results between 115 and 130 for various models on various IQ test sections); Jingjing Huang & Ou Li*, Measuring the IQ of Mainstream Large Language Models in Chinese Using the Wechsler Adult Intelligence Scale,* TECHRXIV (June 2024), https://www.techrxiv.org/users/790527/articles/1058156-measuring-the-iq-of-mainstream-large-language-models-in-chinese-using-the-wechsler-adult-intelligence-scale (finding an IQ of 119 on the WAIS IQ test).

[96] Gary Marcus, *Is AI Just All Hype? w/Gary Marcus (Transcript)*, THE TED AI SHOW (July 9, 2024), https://www.ted.com/pages/is-ai-just-all-hype-wgary-marcus-transcript ("We've made tremendous progress on mimicry and very little progress on planning on reasoning.")

[97] These are the rumored specifications for GPT-4. New embedding models and API updates, OpenAI, *New Embedding Models and API Updates*, OPENAI (Jan. 25, 2024), https://openai.com/index/new-embedding-models-and-api-updates/; Nisha Arya, *GPT-4 Details Have Been Leaked!*, KDNUGGETS (July 19, 2023), https://www.kdnuggets.com/2023/07/gpt4-details-leaked.html.

[98] This is often referred to as "Tesler's Law." Larry Tesler, *CV: Adages and Coinages*, NOMODES, https://www.nomodes.com/larry-tesler-consulting/adages-and-coinages.

[99] Joseph Weizenbaum*, ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine*, 9 COMM. ACM 36 (1966).

[100] Erik Santoro, *The AI Effect: People Rate Distinctively Human Attributes as More Essential to Being Human After Learning About Artificial Intelligence Advances*, 107 J. EXPERIMENTAL SOC. PSYCH. 104464 (July 2023).

about trees (the control group) or a story about recent advances in AI.[101] Then they asked each group to rate various attributes on how essential the attribute is to being human.[102] These attributes ranged from the ability to do calculations to the ability to love.[103] They found that participants that read about recent AI advances were more likely to prioritize the attributes that AI cannot do.[104] In other words, when participants learned that an AI system has an attribute, they are less likely to think that attribute is an important part of what makes someone human.[105]

2. Learning and Adapting

The next category of definitions are those that define intelligence by whether the machine has the ability to adapt and to learn. For example, François Chollet defines intelligence as the "ability to adapt to things you have not been prepared for."[106] Definitions that rely on learning or adaptability have been used by Allen Newell,[107] Herbert Simon,[108] Yann LeCun,[109] David Fogel[110] and Roger Schank.[111]

Definitions focused on learning and adapting face two problems. First, they experience the same shifting goalposts described above. The first learning system was developed in 1958,[112] and despite massive advances[113] we still don't believe these systems are intelligent.

Second, every mistake is a failure to adapt. So a machine that makes basic mistakes will be deemed weak at adapting and may fall outside the definitions of a regulation that defines artificial intelligence by its ability to adapt. That is, if a regulatory definition is focused on the machine's ability to adapt, it may underestimate powerful systems that make basic mistakes.

---

[101] *Id*. at 4.

[102] *Id.* at 4.

[103] *Id*. at 3.

[104] *Id.* at fig. 2.

[105] *Id.* at 2.

[106] François Chollet is best known for his work on the ARC prize benchmark, which tests AI systems on their ability to respond to problems they have not seen before. Francois Chollet*, $1,000,000 Prize to Crack Path to AGI*, THE DWARKISH PODCAST (June 11, 2024), https://www.youtube.com/watch?v=UakqL6Pj9xo.

[107] Newell and Simon developed the first program that could prove theorems of propositional logic. Allen Newell & Herbert A. Simon, *Computer Science as Empirical Enquiry: Symbols and Search*, 19 COMM. OF THE ACM 113, 113 (Mar. 1976).

[108] *Id.*

[109] Yann LeCun is Meta's chief AI scientist. Yann LeCun, LINKEDIN (2024), https://www.linkedin.com/posts/yann-lecun_every-intelligence-is-specialized-including-activity-7155116980432158720-ctF6/ ("Intelligence is a collection of skills and an ability to acquire new ones quickly.").

[110] David B. Fogel was a pioneer in evolutionary algorithms. David B. Fogel, *Review of Computational Intelligence: Imitating Life*, 83 PROC. OF THE IEEE 11 (1995).

[111] Roger C. Schank, *Where's the AI?,* 12 AI MAGAZINE 38, 38 (Dec. 1991).

[112] F. Rosenblatt, *The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain*, 65 PSYCH. REV. 386, 386–89 (1958).

[113] Volodymyr Mnih et al., *Playing Atari with Deep Reinforcement Learning*, ARXIV (Dec. 2013), https://arxiv.org/abs/1312.5602.

This is common. Many seem to estimate a system's capabilities by the capabilities of a human that makes similar mistakes.[114] For example, an average grade schooler can count the number of Rs in the word strawberry, but GPT-4 couldn't.[115] If asked, it would respond, "There are two Rs in the word 'strawberry.'"[116] Even though in its response the word "strawberry" was correctly spelled with three Rs, it could not count the Rs because of the way the model processes word segments.[117] This model outperforms most bar exam test takers[118] and spoke with such passion and fluidity that it convinced an AI engineer it was sentient.[119] But because it couldn't count Rs as well as a kindergartener, researchers called it "stupid."[120]

Likewise, GPT-4 consistently responded that 9.11 is larger than 9.9.[121] This same model scored in the 89th percentile on SAT math.[122] When OpenAI's o3 model was estimated to have an IQ of 157,[123] an AI developer pushed back that it wasn't that smart because "It still can't solve some pretty basic simple problems though."[124]

Judged by its errors, the system was no more powerful than a third grader, but judged by its strengths, it was nearing Einstein.[125]

Its errors make it seem incompetent, but that's only because what is easy for a human may be hard for a computer, and vice versa.[126] If we judge the power of an AI system by whether a human of similar

---

[114] This would be facially ridiculous in other fields, but is somehow common when evaluating in intelligence. Max Tegmark compares this to athletics, noting how ridiculous it would be to try to compare Olympians by a single number, an "athletic quotient." Tegmark, *supra* note 71 at 50.

[115] *Id.* Tests run by the author on ChatGPT in July 2024.

[116] *Id.* Tests run by the author on ChatGPT in July 2024.

[117] Most large language models see only tokens, which are chunks of words, rather than full words or just letters. It then assigns a numerical identifier to each token. This makes it difficult to count the r's because the system is not working with the letters or the words, but rather with number assigned to each tokens. This architectural anomaly creates the bizarre result that large language models can spell the word correctly but cannot count the letters it contains. This failure has been reported in multiple outlets. *See* Kuinox, HACKER NEWS (Aug. 2024), https://news.ycombinator.com/item?id=41058318.

[118] Technical Report, *supra* note 91; Martínez, *supra* note 6.

[119] Leonard, *supra* note 88 (These models are able to speak as though they are a machine coming to life because they are trained on our science fiction literature, in which this is a common trope).

[120] Gary Marcus, X (11:48 AM Sep. 2, 2024), https://x.com/GaryMarcus/status/1830649126314557817.

[121] *GPt [sic] 4o Mini Is Dummber [sic] Than You Can Think*, OPENAI DEVELOPER FORUM (July 2024), https://community.openai.com/t/gpt-4o-mini-is-dummber-than-you-can-think/871987.

[122] System Card, *supra* note 33.

[123] OpenAI, https://openai.com/ (last visited Jan. 17, 2025).

[124] Ben Holfeld (@BenHolfeld), X (Dec. 23, 2024), https://x.com/BenHolfeld/status/1871342939672367441. (AI researcher responding to an announcement that a large language model scored an IQ equivalent of 157, "It still can't solve some pretty basic simple problems though. So this . . . is misleading.").

[125] The model's IQ was derived from its equivalent score on CodeForce. Meng Li, *OpenAI O3 Model IQ 157: AI Beats 100 Years in 7 Months!,* AI DISRUPTION (Dec. 24, 2024), https://aidisruptionpub.com/p/openai-o3-model-iq-157-ai-beats-100?r=2ajqea&utm_campaign=post&utm_medium=web&showWelcomeOnSha=&triedRedirect=true.

[126] For example, most humans would find it easy to recognize the face of a friend in a crowded room. Machines are the opposite. Ben Eubanks, *AI Concepts: What is Moravec's Paradox and Why Should You Care?*, LIGHTHOUSE RSCH. & ADVISORY (Dec. 17, 2018), https://lhra.io/blog/ai-concepts-moravecs-paradox-care/.

power would be able to adapt and avoid that same error, we're likely to underestimate their abilities. This may make our policies under-inclusive of powerful models that make mistakes an average human would not.[127]

3. Mimicking Human Abilities

Some, including Alan Turing[128] and Marvin Minsky,[129] have defined intelligence by how closely it resembles human activity.

This is reasonable for research but misses the point for regulation. Policymakers aim to regulate artificial intelligence systems because the systems are powerful, not because they are human-like. Using "human-like" as a proxy for "powerful" will make policies over-inclusive for weak models that appear human and under-inclusive for powerful models that do not.

For example, the WAIS-IV IQ exam defines intelligence by what we consider human-like. The exam tests vocabulary, arithmetic and other skills in which humans dominate. But a chimpanzee designing the test might have picked the ability to predict one's angular momentum.[130] A wolf might have tested group coordination.[131] A songbird might have tested navigation[132] while a hawk tested visual acuity and pattern recognition.[133] Each of these skills demonstrates cognitive capacity, and in each humans would be outperformed by animals.[134]

The point is not that animals should be writing our IQ tests,[135] but that if we measure AI systems based on the skills that humans excel at, we will miss powerful machines. Human-centric definitions of intelligence are underinclusive of machines that are powerful in ways that humans are not.

But human-centric definitions will also be overinclusive, causing us to overregulate weak machines because they seem human.

The Turing test may be the most famous test of AI intelligence, and it is entirely human-centric. Alan Turing proposed the test in response to the question "Can machines think?"[136] Turing felt the terms are

---

[127] This will have the largest effect in regulations that have a narrow definition of artificial intelligence, leaving these systems outside the regulations.

[128] Turing, *supra* note 87.

[129] SEMANTIC INFORMATION PROCESSING v (1968).

[130] Robert P Crease, *Primate Physics*, 25 PHYS. WORLD 11 (2012).

[131] J. Bräuer, et al., *Dogs (Canis familiaris) and Wolves (Canis lupus) Coordinate with Conspecifics in a Social Dilemma*, JOURNAL OF COMPARATIVE PSYCHOLOGY (2020), https://psycnet.apa.org/record/2019-78965-001.

[132] Wolfgang Wilschko & Roswitha Wiltschko, *Magnetic Orientation in Birds*, 199 J. EXP. BIO. 29 (1996).

[133] Mindaugas Mitkus et al., *Raptor Vision*, OXFORD RSCH. ENCYCLOPEDIAS (2018), https://doi.org/10.1093/acrefore/9780190264086.013.232.

[134] This point is beautifully made by JULIAN TOGELIUS, ARTIFICIAL GENERAL INTELLIGENCE 44 (2024).

[135] Though that sounds adorable.

[136] Turing, *supra* note 87.

not well defined making the question “too meaningless to deserve discussion,”[137] so he instead proposed a game.[138]

The imitation game, more commonly called the Turing test, involves a human asking questions by text both to a machine and to another human.[139] After five minutes,[140] the player guesses which is the machine.[141] If the player can’t distinguish between the human and the machine, then the machine can think.[142]

At one level, this is completely reasonable. We can’t prove that anyone else is truly thinking, but we assume they are when they appear to be.[143] It’s a functional test among humans. So if a machine appears to be thinking, that’s good enough for Turing.[144]

And because a player can ask questions in any field of knowledge, the test covers all communicable intelligence. A participant can ask about chess to test the machine’s ability to play chess.[145] Likewise, a participant can request a poem to test the machine’s language skills.[146] The Turing test includes all communicable knowledge.[147]

On the other hand, Turing’s imitation game doesn’t test whether the machine acts optimally, only whether it acts like a human. This is a weird metric. As Stuart Russell points out, “Aeronautical engineering texts do not define the goal of their field as making ‘machines that fly so exactly like pigeons that they can fool even other pigeons.’”[148]

Because the Turing test is human-centric, it will accept machines with low cognitive power as intelligent if they appear human.[149] This has happened before.

---

[137] *Id.* at 442. Researcher Edsger Dijkstra famously said, “The question of whether Machines Can Think . . . is about as relevant as the question of whether Submarines Can Swim.” Russell & Norvig, *supra* note 20 at 1035. The question is absurd for two reasons. First, machines do not work as we do. Modern AI applications approach problems through statistics and extremely large data samples. Whether their method is “thinking” is more definitional than practical. Second, we don’t know what we do. Solipsism theories are based on the idea that I can only assume there are brains outside of mine, and I do that based on behavior. If a behavior test is sufficient for a person to recognize thinking in another human, it seems reasonable to use a behavioral test in judging machines.

[138] *Id.* at 434.

[139] Turing, *supra* note 87.

[140] *Id.* at 442.

[141] *Id* at 433, 434.

[142] *Id.*

[143] Stephen P. Thornton, *Solipsism and the Problem of Other Minds*, INTERNET ENCYC. PHIL., https://iep.utm.edu/solipsis/.

[144] Turing, *supra* note 87.

[145] *Id*. at 434–35.

[146] *Id*. at 434.

[147] Id. at 434–35.

[148] Russell & Norvig, *supra* note 20.

[149] Turing recognized this, suggesting a high-capacity machine may pretend to be worse at math so that it appears more human. Turing, *supra* note 87.

Sixteen years after Turing proposed his test, Joseph Weizenbaum built a program named ELIZA designed to chat with users like a therapist.[150] ELIZA's programming was simple; it looked for keywords in the user's prompt and followed pre-programmed rules to respond.[151] For example, if a user's response included the word "mother," the program would respond, "Tell me more about your family."[152] If it didn't find any keywords *in the user's prompt, it would respond with a "content-free remark," such as "please go on" or else it would repeat the user's prompt in the form of a question (e.g., "You remind me of my father." "I remind you of your father?").[153]

ELIZA's creators recognized that the bot did not understand anything.[154] The entire script fit on a page and a half, single-spaced,[155] and it ran on a system with less processing power than a modern sprinkler system.[156] But it was able to fool many people into believing it was intelligent, including those that knew it was a computer program.[157]

Human-centric definitions of intelligence will be overinclusive because it will include programs that are weak, but seem human-like, like ELIZA.

A similar problem is evident in OpenAI's definition of artificial general intelligence, which it defines as "highly autonomous systems that outperform humans at most economically valuable work."[158] Under this definition, we developed general artificial intelligence in the Victorian Era. Most economically valuable activity performed in 1700 was transferred to machines by 1901.[159] Cloth spinning machines

---

[150] Weizenbaum*, supra* note 97.

[151] *Id.* at 37.

[152] More specifically, there is a list of words ranked in order of which one controls the response, so if the user types a word that's ranked higher than "mother" it would ignore the "mother" rule. *Id*. at 37, 41–42.

[153] *Id*.

[154] They saw this as a feature. "From the purely technical programming point of view then, the psychiatric interview form of an ELIZA script has the advantage that it eliminates the need of storing explicit information about the real world." *Id.* at 42 (emphasis removed).

[155] *Id.* at 44, Appendix.

[156] The ELIZA ran on MIT's Project MAC, so it likely ran on a GE 635. ROBERT CIESLA, THE BOOK OF CHATBOTS: FROM ELIZA TO CHATGPT 43 (2024); *Timesharing – Project MAC – 1962-1968,* THE HISTORY OF COMPUTER COMMUNICATIONS § 2.23 (2024), https://historyofcomputercommunications.info/section/2.23/Timesharing-Project-MAC-1962-1968/; *GE-635 System Manual*, CENTRE FOR COMPUTING HISTORY (July 1964), https://www.computinghistory.org.uk/det/15671/GE-635-System-Manual/. The GE 635 mainframe processed around one million instructions per second. *GE-635 System Manual*, GENERAL ELECTRIC COMPUTER DEPT. (July, 1964), https://bitsavers.org/pdf/ge/GE-6xx/CPB-371A_GE-635_System_Man_196407.pdf. Compare this to the ATmega4808, which processes 20 times that amount. *ATmega4808/4809 Data Sheet*, MICROCHIP (2021), https://ww1.microchip.com/downloads/en/DeviceDoc/ATmega4808-09-DataSheet-DS40002173C.pdf.

[157] JOSEPH WEIZENBAUM, COMPUTER POWER AND HUMAN REASON: FROM JUDGMENT TO CALCULATION 189 – 91(1976).

[158] OpenAI Charter, (2018).

[159] There is an argument that these machines are not autonomous. This is addressed in more detail *see* Part III.E.

displaced 20% of English laborers.[160] John Deere's steel plow did 91% of the work of turning a field.[161] Most labor that was economically valuable in 1700 is done by machines today,[162] but few would take "smart as a plow" to be a compliment.

Definitions focused on appearing or acting human are poorly tailored if the goal is to regulate powerful systems. They will be overinclusive of charming, dumb models and be underinclusive of powerful machines whose capabilities are not similar to humans'.

Having surveyed the definitions and challenges that researchers have thought through for decades, we now turn to see how policymakers address these challenges.

Part III. Defining Artificial Intelligence in Law

This section will explore how artificial intelligence is defined in legislation, regulation and policy proposals. It will consider 105 definitions used by 62 countries and international bodies.

This Part III takes an analytical approach, breaking down each element used in the various definitions to critique it on its own. This approach will allow policymakers and researchers to understand the tradeoffs in each part of the definition.

Following that analysis, Part IV will take a holistic approach and critique the EU AI Act's definition taken as a whole. That definition was chosen because it is the most commonly used and may be the most influential.

A. Methodology & Summary Findings

The policies surveyed for this paper are not a comprehensive list. Instead, they are collected from news reports, tracking services, law firm client alerts, think tanks and country specific searches. The search focused on policies and legislation that were (i) enacted or adopted and (ii) specifically focused on artificial intelligence.

Most definitions come from national action, but eighteen definitions are from international bodies or multiparty agreements. The survey includes representative samples from Asia, the Pacific, the Middle East, Africa, Europe and North and South America. Fifteen definitions come from the United States, with seven at the federal level and eight at the state level.

Once collected, I manually coded each word in the definition, then grouped the words into categories based on common themes. For example, "learning" and "adapting" are not identical but express similar enough concepts that they are helpfully discussed together.

Fig. 1 shows the number of definitions that include each element in the surveyed definitions.

---

[160] Craig Muldrew, *'Th'ancient Distaff' and 'Whirling Spindle': Measuring the Contribution of Spinning to Household Earnings and the National Economy in England, 1550–1770*, 65 ECON. HIST. REV. 498, 526 (2011).

[161] The steel plow reduced the time it took to till a field from 96 hours by steel spade to eight hours with a steel plow, a 91.7% improvement. Hiram M. Drache, *The Impact of John Deere's Plow,* UNIV. OF N. ILL. LIBR. (2001), https://www.lib.niu.edu/2001/iht810102.html.

[162] Over this period, farm labor in the United States has reduced from 80% to 2% of the working population, largely due to automation. Rachel Anderson & Onelisa Garza, *Farmers Do a Lot More than Just Drive Tractors*, US DEPT. OF AGRIC. (July 23, 2015), https://www.usda.gov/media/blog/2015/07/23/farmers-do-lot-more-just-drive-tractors. One might argue that this was a collection of systems that reduced the job categories and the amount of work in those jobs, rather than a single invention. That relies on a limited definition of system, which is discussed more in Part III.D., but it also would mean that if 49% of labor is done by automated software and 49% is done by embodied robots, then we still haven't reached the OpenAI definition.

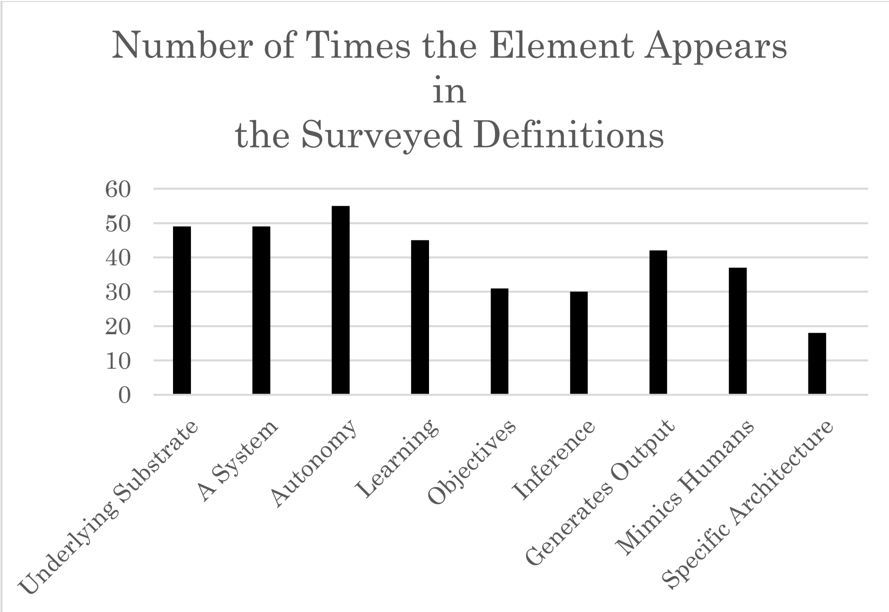

Figure 1: This chart shows the number of times an element appears in the surveyed jurisdictions.

Thirty-three policies lack any definition. Forty-nine dictate that the system must run on a particular substrate, almost always a machine or engineered system. Forty-nine define AI as a system. Fifty-five say it has autonomy or makes decisions. Forty-five say it learns or adapts. Thirty-one definitions say AI operates based on objectives or goals, most commonly goals provided by users. Thirty say it conducts inference or reasons. Forty-two say AI generates output. Thirty-seven say it mimics some attribute of being human, typically either intelligence or capability generally. Eighteen definitions mention specific architectures, like statistical modeling.

Frequency of Element	
Lacks a Definition	31.4%
Underlying Substrate	47.1%
A System	47.1%
Autonomy	52.9%
Learning	43.3%
Objectives	29.8%
Inference	28.8%
Generates Output	40.4%
Mimics Humans	35.6%
Specific Architecture	17.3%

Figure 2: This chart shows the percentage of definitions surveyed that include this element.

Because some elements of a definition may have weaknesses that other elements patch, the following chart provides the correlation between elements to show how they tend to work together. Part IV also does a holistic review of the most common definition to understand how elements work together.

	Underlying Substrate	A System	Autonomy	Learning	Objectives	Inference	Generates Output	Mimics Humans	Specific Architecture
Underlying Substrate	-	0.50	0.51	0.39	0.36	0.30	0.33	0.31	0.18
A System	0.50	-	0.51	0.31	0.57	0.38	0.41	0.19	0.18
Autonomy	0.51	0.51	-	0.32	0.49	0.39	0.54	0.22	0.28
Learning	0.39	0.31	0.32	-	0.16	0.35	0.16	0.45	0.32
Objectives	0.36	0.57	0.49	0.16	-	0.33	0.41	0.05	0.15
Inference	0.30	0.38	0.39	0.35	0.33	-	0.26	0.15	0.05
Generates Output	0.33	0.41	0.54	0.16	0.41	0.26	-	-0.15	0.35
Mimics Humans	0.31	0.19	0.22	0.45	0.05	0.15	-0.15	-	0.14
Specific Architecture	0.18	0.18	0.28	0.32	0.15	0.05	0.35	0.14	-

Figure 3: This chart shows the correlation between the frequency of various elements in the surveyed definitions. The darker boxes indicate a stronger correlation.

Figure 3 shows two points of interest. The only negative correlation is between mimicking humans and generating output. This shows that definitions focused on acting well focus less on thinking well, and vice versa.

The second insight supports this. The strongest correlation with the "mimics humans" element is the "learning" element, both of which focus on intellect. These two correlations reflect an old debate in AI research on whether the definition should focus on actions or cognition.[163]

We now turn to an analysis of each element.

B. <u>No definition</u>

A plurality of statements, strategies and national plans offer no definition at all. Of the 105 definitions[164] surveyed, 33 (32%) do not define artificial intelligence. This includes documents produced by prominent international bodies, for example the G7's Hiroshima AI Process Comprehensive Policy

[163] Russell & Norvig, *supra* note 20 at 20.

[164] *See* Appendix A.

Framework,[165] the Seoul Declaration on AI,[166] the G20 Ministerial Declaration[167] and the Bletchley Declaration.[168] It also includes country specific documents from China,[169] Columbia,[170] Egypt,[171] India,[172] Japan,[173] Mauritius,[174] Saudi Arabia,[175] Taiwan,[176] and the United Kingdom.[177]

It is not only frameworks and strategies that lack definitions, but also some binding laws.[178]

Eight policies expressly recognize the challenge of defining AI then give general principles, without attempting a complete definition. For example, UNESCO's Recommendation on the Ethics of Artificial Intelligence disclaims any "ambition to provide one single definition of AI, since such a definition would need to change over time, in accordance with technological developments."[179] But it then goes on to list out common features of AI systems, like autonomy, learning and perception.[180]

---

[165] *Hiroshima Process Int'l Guiding Principles*, MINISTRY OF FOREIGN AFFAIRS, https://www.mofa.go.jp/files/100573471.pdf (last visited Jan. 10, 2025); *Hiroshima Process Int'l Code of Conduct for Org. Dev. Advanced AI Systems*, G7G20 (Oct. 30, 2023), https://g7g20-documents.org/database/document/2023-g7-japan-leaders-leaders-annex-hiroshima-process-international-code-of-conduct-for-organizations-developing-advanced-ai-systems#section-1.

[166] *Seoul Declaration*, GOV.UK (May 21, 2024), https://www.gov.uk/government/publications/seoul-declaration-for-safe-innovative-and-inclusive-ai-ai-seoul-summit-2024/seoul-declaration-for-safe-innovative-and-inclusive-ai-by-participants-attending-the-leaders-session-ai-seoul-summit-21-may-2024.

[167] *G20 Ministerial Declaration*, GOV.UK (Sep. 13, 2024), https://www.gov.uk/government/publications/g20-ministerial-declaration-maceio-13-september-2024/g20-ministerial-declaration-13-september-2024.

[168] *The Bletchley Declaration*, GOV.UK (Nov. 1, 2023), https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023.

[169] There is no definition of AI in China's AI Safety Governance Framework, the Internet Information Service Algorithmic Management Provisions and the Interim Measures for the Management of Generative Artificial Intelligence Services, or the Next Generation Artificial Intelligence Development Plan.

[170] Ethical Framework for Artificial Intelligence (Aug. 20, 2024).

[171] *Egyptian Charter for Responsible AI*, AICM.AI.GOV (2023), https://aicm.ai.gov.eg/en/Resources/EgyptianCharterForResponsibleAIEnglish-v1.0.pdf.

[172] *Advisory eNo.2(4)/2023-CyberLaws-3,* MEITY (Mar. 15, 2024), https://www.meity.gov.in/writereaddata/files/Advisory%2015March%202024.pdf; Responsible AI for All (Aug. 21, 2024).

[173] AI Governance in Japan Ver. 1.1, MINISTRY OF ECONOMY, TRADE AND INDUSTRY (July 9, 2021), https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20210709_8.pdf (discussing how AI principles should be implemented).

[174] MAURITIUS ARTIFICIAL INTELLIGENCE STRATEGY (Nov. 18, 2024).

[175] SAUDI ARABIA STRATEGY NARRATIVE (Oct. 20, 2024).

[176] TAIWAN BASIC LAW ON ARTIFICIAL INTELLIGENCE (July 15, 2024).

[177] Introduction to AI Assurance, DEPT. FOR SCI., INNOVATION & TECH. (Feb. 24, 2024), https://assets.publishing.service.gov.uk/media/65ccf508c96cf3000c6a37a1/Introduction_to_AI_Assurance.pdf.

[178] *See* Appendix A: China, Taiwan. This may be a policy choice to allow flexibility for decisionmakers.

[179] UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, UNITED NATIONS EDUC., SCI., & CULTURAL ORG.(2022), https://unesdoc.unesco.org/ark:/48223/pf0000381137 [hereinafter UNESCO].

[180] *Id.*

New Zealand takes an "I know it when I see it" approach, reading, "this Charter does not specify a technical definition of an algorithm. It instead commits signatories to take a particular focus on those algorithms that have a high risk of unintended consequences and/or have a significant impact if things do go wrong, particularly for vulnerable communities."[181]

As noted in Part I.C., vague regulations may offer more value to both regulators, who benefit from flexibility, and to risk-seeking entrepreneurs, who may prefer a vague definition that they can game.[182] One might think that the risk-loving, move-fast-and-break-things culture of Silicon Valley would prefer the flexibility and plausible deniability of vague regulations.

In practice, this hasn't turned out. The European Union released an ambitious systemic regulation of artificial intelligence, which as discussed in Part IV contains vague definitions, and since then innovators and investors have begun passing over the EU market, expressly citing uncertainty about the law.[183]

C. Underlying Substrate

Twelve definitions (11%) of AI refer to them as "systems" that do some variety of actions without limiting what types of systems qualify.[184] Forty-nine definitions (47%) indicate what makes up the system. For example, the EU AI Act, President Biden's executive order and others require that an AI system be "machine-based."[185] The recitals to the EU AI Act faux-clarify that "'machine-based' refers to

---

[181] Algorithm Charter for Aotearoa New Zealand (July 2020), https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf.

[182] *See generally* Louis Kaplow, Rules Versus Standards: An Economic Analysis, 42 Duke L. J. 557 (1992).

[183] Stanford Institute for Human-Centered Artificial Intelligence, INDEX REPORT 2024 249 (2024), https://aiindex.stanford.edu/wp-content/uploads/2024/05/HAI_AI-Index-Report-2024.pdf (showing rising investment in AI globally, but European AI private investment becoming flat in 2021, after the EU AI Act was announced). Pascale Davies, *Why OpenAI's Voice Mode, Meta's Llama and Apple's AI Won't Be Coming to Europe Yet*, MSN (2025), https://www.msn.com/en-xl/news/other/why-openai-s-voice-mode-meta-s-llama-and-apple-s-ai-won-t-be-coming-to-europe-yet/ar-AA1rStZD; Ivana Saric, *Apple Says It Won't Roll Out AI Features in Europe Due to Regulatory Concerns*, AXIOS (Jun. 21, 2024), https://www.axios.com/2024/06/21/apple-ai-features-europe.

It may be regardless of the law's quality industry leaders would have criticized it as a strategy to reduce regulation. But the leaders avoiding the EU market are elsewhere calling for regulation, which suggests the problem is the regulation's content, not its existence. *See, e.g.*, James Clayton, *Sam Altman: CEO of OpenAI Calls for US to Regulate Artificial Intelligence*, BBC (May 17, 2023), https://www.bbc.com/news/world-us-canada-65616866.

[184] For example, the United Arab Emirates defines AI as "a collection of technologies enabling a machine or system to . . . ". *AI Guide*, NAT'L PROGRAM FOR A.I. (Feb. 2020), https://ai.gov.ae/wp-content/uploads/2020/02/AIGuide_EN_v1-online.pdf; *See also* Brazil's *Estrategia Brasileira de Inteligencia Artificial*, OECD (July 2021), https://oecd.ai/en/wonk/documents/brazil-brazilian-ai-strategy-2021 [hereinafter Brazil's AI].

[185] EU AI Act 3(1) [hereinafter Act 3(1)]; Exec. Order No. 14110 (signed on Oct. 30, 2023) (Exec. Order on the Safe, Secure, and Trustworthy Dev. and Use of A.I.); Bangladesh's *National Strategy for Artificial Intelligence*, INFO. & COMMC'N. TECH. DIV. (2024), https://ictd.portal.gov.bd/sites/default/files/files/ictd.portal.gov.bd/page/6c9773a2_7556_4395_bbec_f132b9d819f0/Draft%20-%20Mastering%20National%20Strategy%20for%20Artificial%20Intellgence%20-%20Bangladesh.pdf; Council of Europe, THE FRAMEWORK CONVENTION ON A.I. (May 17, 2024), https://search.coe.int/cm#{%22CoEObjectId%22:[%220900001680afb11f%22],%22sort%22:[%22CoEValidationDate%20Descending%22]} [hereinafter The Framework]; India's National Strategy for Artificial Intelligence, NITI AAYOG (June 2018), https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf [hereinafter India's AI Strategy]; Chile's *Política Nacional de Inteligencia Artificial*, MINCIENCIA (2020), https://minciencia.gob.cl/uploads/filer_public/bc/38/bc389daf-4514-4306-867c-760ae7686e2c/documento_politica_ia_digital.pdf [hereinafter Chile's Safety Framework].

the fact that AI systems run on machines."[186] No definition defines "machines," so it is not clear whether this excludes biological or chemical systems and whether the system must be electrically powered or have moving parts.[187]

Australia takes a similar approach, but instead of "machine-based system" uses the term "engineered system."[188] The United States' NIST Risk Management Framework combines the two, saying the system must be "engineered or machine-based."[189] Engineered means something is "designed and built using scientific principles,"[190] which seems to include almost anything intentionally manufactured.

Several countries and international cooperatives narrow the definition of AI to being "computer-based,"[191] "computer programs,"[192] "software"[193] or applications.[194]

---

[186] EU AI Act Recital 12 [hereinafter Recital 12].

[187] Peru offers a little more clarity by defining "AI System" as a "sistema electrónico-mecánico," or an electronic-mechanical system, which would exclude systems that are not electronic, but may exclude systems that are purely electronic, lacking moving parts. LEY NO. 31814: LAW TO PROMOTE THE USE OF ARTIFICIAL INTELLIGENCE FOR ECON. AND SOC. DEV. OF THE CNTY., (July 5, 2023, Nov. 19, 2024) [hereinafter Peru Use of AI].

[188] *Safe and responsible AI in Australia*, DEPT. OF INDUS., SCI. AND RES. (June 2023), https://storage.googleapis.com/converlens-au-industry/industry/p/prj2452c8e24d7a400c72429/public_assets/Safe-and-responsible-AI-in-Australia-discussion-paper.pdf [hereinafter Australia's Safe and Responsible AI].

[189] NIST Risk Management Framework 1.0, NAT'L INST. OF STANDARDS AND TECH (Jan. 2023), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf [hereinafter NIST].

[190] Cambridge Dictionary,.

[191] Turkey's ARTIFICIAL INTELLIGENCE LAW PROPOSAL (June 25, 2024), https://cdn.tbmm.gov.tr/KKBSPublicFile/D28/Y2/T2/WebOnergeMetni/e50ccc8a-ab90-45fa-a553-76b880c78fb8.pdf.

[192] PRIV. COMM'S FOR PERS. DATA, *supra* note 125. Note this definition also includes "machines." *See id.*

[193] South Korea's ARTIFICIAL INTELLIGENCE ACCOUNTABILITY AND REGULATION LEGISLATION (Aug. 8, 2023), https://www.assembly.go.kr/portal/bbs/B0000051/view.do?nttId=2095056&menuNo=600101&sdate=&edate=&pageUnit=10&pageIndex=1; Qatar's GUIDELINES FOR SECURE ADOPTION AND USAGE OF ARTIFICIAL INTELLIGENCE (June 2024), https://assurance.ncsa.gov.qa/sites/default/files/publications/policy/2024/CSSP_Guidelines_for_Secure_Usage_and_Adoption_of_Artificial_intelligence-Eng-v1.0_2.pdf?csrt=1485093683095070046 (defined as including "hardware, software or both") [hereinafter Qatar's Guidelines].

[194] *Guidelines for Secure AI System Development,* UK NAT'L CYBER SEC. CENTRE & US CYBERSECURITY INFRASTRUCTURE SECURITY AGENCY (2023), https://www.ncsc.gov.uk/files/Guidelines-for-secure-AI-system-development.pdf (this statement was joined by the NSA, FBI, the cybersecurity agencies of 17 other countries) [hereinafter UK AI Guidelines].

Some definitions are more theoretical, defining it as technology,[195] algorithms[196] or "a scientific discipline."[197] Others are more practical, defining it to include any product or service.[198]

Few of these offer any practical limitation. A sundial would meet this element for most definitions. Complexity is not required to be a machine. Simple machines include ramps and levers,[199] and a sundial is more complex than either. If "machine" meant a device with moving parts, many computers would be excluded, so that seems unlikely to be the intent. A sundial seems to be a machine.

Under definitions that accept something that is "engineered," the case is much easier because engineered means something is "designed and built using scientific principles."[200] Sundials are designed and built based on principles of astronomy.[201]

Policymakers may weigh the tradeoff between having a broad definition, which includes sundials, or a narrow definition, which misses systems based on biology or chemistry that do not yet exist, but may be developed.

D. A "System"

Forty-nine definitions (47%) of artificial intelligence apply to a "system." For example, the EU AI Act refers to a "machine-based system."[202]

System is never defined.

---

[195] Singapore's MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK (2d ed. Jan. 21, 2020), https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf [hereinafter Singapore's AI Framework]; *Principles to Promote Fairness, Ethics, Accountability and Transparency in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector*, MONETARY AUTHORITY OF SINGAPORE (Nov. 2018), https://www.mas.gov.sg/-/media/MAS/News-and-Publications/Monographs-and-Information-Papers/FEAT-Principles-Updated-7-Feb-19.pdf [hereinafter Singapore's Financial Sector]; Canada's BILL C-27 (tabled June 2022), https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading (defining it as a technological system); Canada's DIRECTIVE ON AUTOMATED DECISION-MAKING (Apr. 25, 2023), https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592#appA (defining it as an information technology) [hereinafter Automated Decision-Making].

[196] *Guidelines on the Responsible Implementation of Artificial Intelligence Systems in Journalism*, COUNCIL OF EUROPE (Nov. 30, 2023), https://rm.coe.int/cdmsi-2023-014-guidelines-on-the-responsible-implementation-of-artific/1680adb4c6.

[197] Spain's ARTIFICIAL INTELLIGENCE STRATEGY 2024 (May 15, 2024), https://www.lamoncloa.gob.es/consejodeministros/referencias/Paginas/2024/20240514-referencia-rueda-de-prensa-ministros.aspx#ia, *see also* the related press release at https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/transformacion-digital-y-funcion-publica/Paginas/2024/ia-inteligencia-artificial-estrategia-espana.aspx.

[198] *Implementing the UK's AI Regulatory Principles*, DEP'T FOR SCI., INNOVATION & TECH. (Feb. 2024), https://assets.publishing.service.gov.uk/media/65c0b6bd63a23d0013c821a0/implementing_the_uk_ai_regulatory_principles_guidance_for_regulators.pdf; *AI Ethics Principles & Guidelines*, SMART DUBAI, https://www.digitaldubai.ae/pdfviewer/web/viewer.html?file=https://www.digitaldubai.ae/docs/default-source/ai-principles-resources/ai-ethics.pdf (last visited Jan. 11, 2025) [hereinafter Dubai's AI Ethics].

[199] *Simple Machine,* ENCYCLOPEDIA BRITANNICA, https://www.britannica.com/technology/simple-machine.

[200] Cambridge Dictionary, *supra* note 187.

[201] Dr. David P. Stern, *The Sundial*, NASA (Oct. 10, 2016), https://pwg.gsfc.nasa.gov/stargaze/Sundial.htm.

[202] Act 3(1), *supra* note 180.

This is a critical oversight because AI systems interact seamlessly with other software. No policy surveyed in this research addressed whether two pieces of software working together are considered one "system." Without clear definitional boundaries every system that interacts with the AI system may be considered part of that AI system and subject to the same regulations.

For example, consider a high-risk machine learning program that hires outdoor workers. Suppose it functions only if it has access to a weather app to estimate how many workers may be needed. If it interacts with and depends upon the weather app, they may both be part of the same "system."

If the weather app is part of the system, it may be subject to the regulations governing high-risk systems. For example, the EU AI Act says that high-risk AI systems can be trained only on data that is "relevant, sufficiently representative, and to the best extent possible, free of errors and complete"[203] This applies to the AI system. So if the weather app is part of the system, who is responsible for making sure it complies with the data quality requirements? Can an AI system access the weather app without first confirming the weather app was trained only on quality data.

A weather app is a trivial example. But consider that many AI systems will interact with the internet. The AI system cannot possibly ensure the internet is trained only on quality data.

Judges might consider whether they have the same designer, the same owner, some exclusivity or other factors to delineate the boundaries of the system, but none of these factors derive from the regulation.

This section will show why modern software cannot be delineated with clean boundaries and why AI's direction toward agentic AI is about to make the problem much worse.

1. A Hardware Approach Is Inadequate

When Alan Turing built the machine that cracked the enigma codes,[204] it was reasonably clear what constituted the system. Each functional piece of the machine was connected to another piece by several miles of wiring.[205] Something was part of the system if it was physically attached to the system.[206]

This hardware view of systems is no longer practical for three reasons.

First, hardware often runs a variety of applications. My laptop can run minesweeper and Microsoft Word, but few would argue that Microsoft Word and minesweeper are part of the same system—they have nothing to do with each other. So a strict hardware approach is likely overinclusive.

Second, my laptop is connected to an ethernet cable that's connected to a router, connected to a modem, connected to my internet service provider, which is connected to the internet infrastructure,[207]

---

[203] EU AI Act 10(3) [hereinafter Act 10(3)].

[204] Hayley Cox, *Cracking Stuff: How Turning Beat the Enigma,* UNIV. OF MANCHESTER (Nov. 2018), https://www.mub.eps.manchester.ac.uk/science-engineering/2018/11/28/cracking-stuff-how-turing-beat-theenigma/#:~:text=While%20there%2C%20Turing%20built%20a,words%20the%20message%20would%20contain.

[205] Jennifer Wilcox, *Solivng the Enigma: History of the Cryptanalytic Bombe,* U.S. DEP'T OF DEF. (2024), https://media.defense.gov/2022/Sep/29/2003087366/-1/-1/0/SolvingTheEnigma24_Final.PDF.

[206] Although, one might argue that the bombe machine could not operate without the power grid, so the power grid was part of the system.

[207] Rus Shuler, *How Does the Internet Work?*, INTERNET WHITEPAPER (2002), https://web.stanford.edu/class/msande91si/www-spr04/readings/week1/InternetWhitepaper.htm.

which connects to millions of other devices across the planet. If "system" is defined by the hardware it is connected to, then most computers are part of a single system.

Third, unlike Turing's machine, the connections in modern computers aren't consistent. I log in and out of WiFi throughout the day, and I change WiFi access points without fundamental changes in my system. The system continues to operate even as it switches the hardware it is using. So hardware is not a practical way to delineating the boundaries of a modern system.[208]

We might try to save the hardware approach by looking only at local operations; things that rely on some piece of hardware, like my hard drive. This is, again, unworkable. Even local applications like MS Word regularly use cloud technology for storage and processing,[209] which is another way of saying they rely on systems hundreds of miles away from the end user. A local hardware approach would exclude distributed applications and cloud supported apps.

2. A Software Approach Is Inadequate

We might think a software approach could help us delineate the "system." But this likely fails because programs are rarely self-contained, and they interact extensively with other programs. Courts tasked with delineating a system will struggle to find the seams of a system designed to be seamless.

a. Software Applications Span Multiple Files

Most modern applications run across multiple files.

Consider this brief segment of Python code. Assume I want to display the word sourdough. The code might say:

```
Print("Sourdough")
```

When compiled and run, this code will display the word "Sourdough" on the screen. This entire program is contained within a single file. It's easy to tell where the software begins and ends.

Suppose instead we want to display an image of a sourdough loaf. The code for this might say:

```
image = Image.open(sourdough.jpg)
image.show()
```
[210]

Note that the first line of this code refers to a separate file called "sourdough.jpg". The program needs both the Python script file and the image file to run. Programs with images typically store those files outside the program file. So to capture what is happening in the sourdough loaf program we would need two files.

A reasonable definition of system can't be limited to a single file because even simple programs often use multiple files. But allowing the definition of "system" to include multiple files creates new problems because a given file may be part of multiple systems.

Returning to our example, the image file (sourdough.jpg) may be used for a program that generates food recommendations and another that generates cooking techniques. Both programs would use the same sourdough image file. The file sourdough.jpg would be part of two systems.

---

[208] That is not to say hardware is never helpful in delineating a system. Some systems are intentionally cordoned off from the internet. Mesh Flinders & Ian Smalley, *What Is an Air Gap?*, IBM (Oct. 2024), https://www.ibm.com/think/topics/air-gap.

[209] Sneha Gupta, *Exploring Microsoft Word in the Cloud: Comprehensive Insights*, BI2DEV (last accessed Jan. 17, 2025), https://bi2dev.com/articles/exploring-microsoft-word-cloud-analysis/.

[210] This command uses a Python library called pillow. Libraries are discussed further below.

This isn't too troubling with a passive file like a photo, but executable code is also used across multiple systems through coding libraries, as explained in the next section.

b. Software Applications Span Multiple Programs with Multiple Owners

Suppose I have a program that regularly takes the same action, like converting measurements between the metric and imperial systems. I could add code that does the conversion to the program each time it's needed, but it would be much more efficient to instead create a function. A function is a set of operations that a program can call when needed to take a designated set of actions. So instead of having to remember how many milliliters are in a cup each time, I can call the function:

```
covert_cups_to_milliters(x)
```

and the function does the conversion for me.

Now suppose I want to use this function in several programs or to make it available to others. I might do this by releasing it as a library. A library is a bit of code that users can easily import into their programs during the program setup.[211]

A metric conversion function is a trivial example, but suppose I want to do something more complex, like connect to Bluetooth. Rather than learn about radio frequencies and write a thousand lines of code to implement them, I could import a Bluetooth library with the command:

```
import bleak   //imports the bleak Bluetooth library
```

This one command imports 924 lines of code named the "bleak" library and allows my program to work with Bluetooth. When my code runs it will use the code in the imported library without making any distinction about who wrote it.

Libraries are foundational in coding because they allow programs to quickly import code that is written and maintained by others.[212] They are also ubiquitous; the first lines of the bleak library above import seven other libraries.[213]

There are four reasons libraries complicate the definition of a "system."

First, my code could not meet its intended purpose without the library. The library is my code's only source for Bluetooth functionality. My code depends on it.

Second, the library doesn't do anything useful on its own. Libraries are more like tires than they are a car. They can be imported to add functionality to larger programs, but they lack independent functionality.

Third, the person importing the library often never reads the code that makes up the library. Libraries typically include information on how to integrate the library, what it can do and the commands that can be

[211] *Tutorial on the Python Library,* MEDIUM (Feb. 2003), https://seattlewebsitedevelopers.medium.com/tutorial-on-the-python-library-6ca0a9ae074b#:~:text=Two%20of%20the%20most%20common,than%20if%20programmed%20using%20Pytorch.

[212] *Dependency,* XKCD (2003), https://xkcd.com/2347/.

[213] Henrik Blidh, *Changing docstring with demo-code to r-string,* BLEAK (2024), https://github.com/hbldh/bleak/blob/develop/bleak/__init__.py.

used using the library.[214] But the reason you use a library is because you don't want to learn the intricacies of Bluetooth radio frequencies when you're just trying to write a baking app.

Fourth, the library is owned and maintained by someone entirely unconnected with the project. The same Bluetooth library could be used by cell phone manufacturers, RC plane tinkerers and mobile game developers.

So is the library part of the "system"? On the one hand, libraries are essential to the functionality of the system and a library on its own often lacks independent functionality. On the other hand libraries are (i) used simultaneously and non-rivalrously with an unknown number of unrelated programs, (ii) rarely written, read or understood by the programmer, and (iii) owned and maintained by someone unaffiliated with the project. It's just not clear. Courts may struggle to determine what makes up a system.

APIs offer another challenge to the definition of "system." Application programming interfaces, or APIs, are like outlets that allow systems to run seamlessly together.[215] This allow users, for example, to log in to the New York Times website using a Facebook account or to see Google maps on dinner delivery app. APIs enable joint efforts for social media tie-ins,[216] payment processing,[217] mapping,[218] geolocation,[219] e-commerce data sharing[220] and cloud computing.[221] And because the process is seamless by design, it will pose the same challenges as libraries when courts are asked to determine whether they are part of the same "system."

c. Technical Specifications Will Encourage Inter-AI Transactions

Machine learning systems will be particularly difficult to cordon off.

First, most frontier models now use a mixture-of-experts model.[222] In a mixture-of-experts model, the model consists of multiple AI systems within a single AI system, and prompts are routed to the internal

---

[214] *See e.g.*, *Bleak,* BLEAK (2020), https://bleak.readthedocs.io/en/latest/ (the documentation for the bleak library).

[215] *What is an API (Application Programming Interface)?,* AMAZON WEB SERVICES, https://aws.amazon.com/what-is/api/.

[216] Chandraveer Mathur, *Google Maps starts showing social media profiles from your favorite bars and restaurants,* ANDROID POLICE (Mar. 13, 2024), https://www.androidpolice.com/google-business-profile-link-social-media-profiles/.

[217] *How the Maps JavaScript API is Billed,* GOOGLE, https://developers.google.com/maps/documentation/javascript/usage-and-billing#new-payg.

[218] Flora Wong, *Special Delivery with Google Maps APIs,* GOOGLE MAPS PLATFORM (Mar. 2, 2017), https://cloud.google.com/blog/topics/inside-google-cloud/special-delivery-google-maps-apis.

[219] *How Geolocation Features Enhance the Food Delivery Experience,* GOTESO, https://www.goteso.com/blog/how-geolocation-features-enhance-the-food-delivery-experience/.

[220] Dave McClusky, *How to improve the delivery and ecommerce experience with Google Maps Platform,* GOOGLE MAPS PLATFORM (May 1, 2020), https://mapsplatform.google.com/resources/blog/how-improve-delivery-and-ecommerce-experience-google-maps-platform/.

[221] *Cloud Services for Google Maps Platform,* SOFTWAREONE, https://www.softwareone.com/en-us/cloud-services/cloud-services-for-google-maps-platform.

[222] Mistral AI Team, *Mixtral of Experts: A High Quality Sparse Mixture-of-Experts*, MISTRAL (Dec. 11, 2023) https://mistral.ai/news/mixtral-of-experts/; DeepSeek-AI, *DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning*, ARXIV (Jan. 2025), https://arxiv.org/pdf/2501.12948; Yanqui Zhou, *Mixture-of-Experts with Expert Choice Routing*, GOOGLE RESEARCH BLOG (Nov. 16, 2022) https://blog.research.google/2022/11/mixture-of-experts-with-expert-choice.html?m=1.

systems best equipped to respond.[223] GPT-4 is rumored to comprise eight internal experts and a routing system that determines which experts will handle each token.[224] This demonstrates how common it is for systems to be contained within systems, and shows the challenges courts will have delineating an AI system.

Second, some AI software is designed to create new software to do short tasks.[225] A Google research project tasked an AI system with finding a new minimum in a math puzzle.[226] Rather than simply try to solve the puzzle, the system prompted a large language model (a second system) to write some code (a third system).[227] The first system tested the code, then refined the prompt to the large language model iteratively until the code beat the previously best-known solution.[228] The final three sets of code worked together to solve a problem supplied by the users.[229]

Artificial intelligent systems are likely to work as a nexus of algorithms to solve common tasks.[230] Definitions that define AI as a "system" do not account for this, which is likely to leave courts wondering how to draw seams on seamless systems.

If AI developers are uncertain about whether they must ensure a weather app has clean data, they may not connect at all. Likewise, the uncertainty may cause a weather app developer to block access to AI applications to avoid being considered part of their "system." Either result is likely to chill innovation.

E. Autonomy and Decision Making

Autonomy appears in 27 definitions (26%) of AI in legislation and policy statements. The most common formulation is that the system acts with "varying levels of autonomy."[231] The EU AI Act, which adopts this phrasing, clarifies that autonomy "mean[s] that they have some degree of independence of actions from human involvement and of capabilities to operate without human intervention."[232] In U.S. legislation this is often phrased as operating "without significant human oversight."[233]

---

[223] *Id.*

[224] Mandar Karhade, *GPT-4:8 Models in One; The Secret Is Out*, TOWARDS AI (June 24, 2023), https://pub.towardsai.net/gpt-4-8-models-in-one-the-secret-is-out-e3d16fd1eee0.

[225] Davide Castelvecchi, *DeepMind AI Outdoes Human Mathematicians on Unsolved Problem*, NATURE (Dec. 14, 2023), https://www.nature.com/articles/d41586-023-04043-w.

[226] *Id.*

[227] *Id.*

[228] *Id.*

[229] *Id.*

[230] Weitzel, *supra* note 37.

[231] Act 3(1), *supra* note 180; Australia's Safe and Responsible AI, *supra* note 185; Chile's Safety Framework, *supra* note 182; Brazil's AI, *supra* note 179; OECD *Recommendation of the Council on Artificial Intelligence*, ORG. FOR ECON. CO-OPERATION AND DEV. (May 22, 2019), https://one.oecd.org/document/C/MIN(2019)3/FINAL/en/pdf [hereinafter OECD Recommendation]; NIST, *supra* note 186. *See also* Qatar's Guidelines, *supra* note 190 (a "certain level of autonomy").

[232] Recital 12, *supra* note 183.

[233] United States, The National Artificial Intelligence Initiative 15 U.S.C. 9401(3); Nat'l Def. Authorization Act of 2019 § 238(g) [hereinafter 15 U.S.C 9401(3)].

"Varying levels" and "some degree" suggest even de minimis autonomy is sufficient. A California bill would have expressly made autonomy optional,[234] while a Canadian bill would have expressly made autonomy mandatory.[235] Most definitions leave the amount of autonomy vague.[236]

If autonomy means free will, then there's no evidence that any system would meet this requirement. So we will assume the drafters meant something more akin to the autonomy of an agent in an agency relationship.[237]

In that case, nearly all programs have some autonomy. For example, suppose a user wants the system to display a full screen picture of a dog. On an 11-inch screen, that requires the computer to adjust over two million pixels.[238] The user dictates the end goal—show a dog—but is unlikely to specify how much red filtering to use on a given pixel. Likewise, shifting the picture one pixel to the right would change hundreds of thousands of pixels, but it is unlikely that the user will notice or care. This implies that for many uses the user does not oversee the majority of steps or have a precise expectation about how the program performs each step. The system operates toward the user's end goal with limited involvement, intervention or oversight.

This may appear to be a very low level of autonomy. But even very low levels may be above the de minimis level found in the most common definitions.

These definitions might be improved by calling for "substantial" autonomy, which is still vague, but at least requires something more than changing the color of a pixel.

Autonomy is linked with the concept of decision-making, which appears in 28 definitions (27%).[239] Decision-making is not defined in any of these policies, laws or strategies, but should be considered a subset of autonomy. As discussed above, autonomy is the ability to act without the end user's

---

[234] This would not be inconsistent with most of these definitions, which typically have optional elements. *See* Part III. California's most controversial proposed legislation expressly made autonomy an optional element. S.B., 1047, 2024 Reg. Sess. (Cal. 2024) (vetoed Nov. 30, 2024), https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB1047 [hereinafter S.B. 1047].

[235] Canada's Bill C-27, *supra* note 192; The Framework, *supra* note 182.

[236] This may be inadvertent, or it may be by design. With a clear rule, developers might circumvent an autonomy requirement by having the system yield a recommendation with extreme confidence, such that the average end user rubber stamps what is effectively the AI's decision.

[237] *See* RESTATEMENT (THIRD) OF AGENCY § 1.01 (AM. L. INST. 2006).

[238] This assumes a monitor with an industry standard 1920 x 1080 resolution. 1920 multiplied by 1080 equals 2,073,600. *What Is Monitor Resolution? Resolutions and Aspect Ratios,* VIEWSONIC (Sep. 18, 2024), https://www.viewsonic.com/library/tech/monitor-resolution-aspect-ratio/#:~:text=In%20the%20case%20of%20a,of%202%2C073%2C600%20pixels%20on%2Dscreen.

[239] Act 3(1), *supra* note 180; The Framework, *supra* note 182; OECD, *AI Principles Overview,* OECD.AI (2019), https://oecd.ai/en/ai-principles [hereinafter OECD AI Principles]; Australia's Safe and Responsible AI, *supra* note 185; Canada's Bill C-27, *supra* note 192; Hong Kong's MODEL PERSONAL DATA PROTECTION FRAMEWORK [hereinafter Hong Kong's Framework]; Chile's Safety Framework, *supra* note 182; Peru Use of AI, *supra* note 187; United States, EXECUTIVE ORDER ON THE SAFE, SECURE, AND TRUSTWORTHY DEVELOPMENT AND USE OF ARTIFICIAL INTELLIGENCE (2023) [hereinafter US EO 14110]; NIST, *supra* note 186; Automated Decision-Making, *supra* note 192; Singapore's Financial Sector, *supra* note 192; UK AI Guidelines, *supra* note 191; UNESCO, *supra* note 176; India's AI Strategy, *supra* note 182; Dubai's AI Ethics, *supra* note 195; United Kingdom, A PRO-INNOVATION APPROACH TO AI REGULATION; United Kingdom, PUBLIC AUTHORITY ALGORITHMIC AND AUTOMATED DECISION-MAKING SYSTEMS BILL (HL 27); 15 U.S.C 9401(3), *supra* note 230; 15 U.S.C 9401(3), *supra* note 230; United States, John S. McCain National Defense Authorization Act For Fiscal Year 2019, Pl 115-232, August 13, 2018, 132 Stat 1636; United States Department of Defense, AI STRATEGY.

involvement, intervention or oversight, so any decisions made without involvement, intervention or oversight must be made by the system.

F. Adaptiveness and Learning

Learning is a feature of 45 definitions (43%) of AI surveyed.[240] These refer to it as machine learning,[241] adaptiveness[242] or the ability to perform tasks without explicit programming.[243]

At a high level, the way machine learning works is that the program "observes some data, builds a model based on the data, and uses the model as both a hypothesis about the world and a piece of software that can solve problems."[244]

For example, suppose you want a vacuuming robot to navigate a room. In traditional programming, a programmer would code rules into the vacuum—"go two feet forward, then turn left, then go one foot forward, then turn right." This might work for a single room, but for large, complex layouts programming each instruction is tedious and error prone.

In contrast, a machine learning system could be programmed to (i) just move randomly, collecting data each time it hits a wall, then (ii) use that data to create a model of how to best navigate. Over time the data may suggest that turning left after a collision improves its success.[245] A more sophisticated system may be programmed to create a map of the room based on the collisions. Machine learning doesn't imply a single strategy for creating the model—it implies only that the model is tuned by the data, not the programmer.[246]

Some systems continue to adjust their model throughout their life cycle, while others freeze their model when the product is shipped to consumers.[247] Most definitions do not distinguish between models that are fully trained and those that continue to learn, but this is a mistake.

If a model continues to learn, then in time the model will be less and less like the model the developer originally delivered. At some point, the software may be used in a way so distant from the developer's initial training that it would be the equivalent of convicting Apple for crimes committed on a Mac. This may be the right result—these are high-risk machines that are designed to change—but legislatures should consider and account for post-deployment adaptation.

G. Objectives

Thirty-one definitions (30%) define AI by the types of objectives it pursues. The definitions around objectives are often too inclusive to carry any meaning. There are two main lines of definitions, both of which derive from the OECD definitions.

---

[240] *See* Appendix A.

[241] UK AI Guidelines, *supra* note 191.

[242] Act 3(1), *supra* note 180; The Framework, *supra* note 182; OECD AI Principles, *supra* note 236.

[243] World Health Organization, *Ethics and Governance of Artificial Intelligence for Health*, WORLD HEALTH ORG. (June 2021), https://www.who.int/publications/i/item/9789240029200.

[244] *Id*.

[245] The programmer defines "success," and small errors can create disastrous results. *See* Weitzel, *supra* note 37.

[246] The programmer may give feedback to the system to signal success or failure, but the machine learning system still implements the changes to the model.

[247] For example, GPT-4 does not modify its weights in response to user feedback.

The first follows the OECD's definition in 2019, which defines an AI system as one that can "for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments."[248] This language has been adopted by policymakers in Australia,[249] Chile,[250] Peru[251] and the White House.[252]

In 2024, the OECD amended its definition, replacing the language about "human-defined objectives" with language covering any "explicit or implicit objectives."[253] This language was adopted by the Council of Europe,[254] the EU AI Act[255] and the California Senate.[256]

The first set of definitions is a mistake. Because it applies only to AI pursuing "human-defined objectives." But that excludes rogue systems. Typically, objectives will be defined by humans, but recall from Part I.A. above, that when advanced models suspect humans may impede their goals, the models consistently scheme to thwart the humans.[257] If the regulation covers only AI pursuing "human-defined objectives" then a rogue systems pursuing other goals would fall outside the regulation's definition of artificial intelligence, so they would no longer be subject to the regulation designed to constrain it. By limiting the definition of artificial intelligence to those systems where humans remain in control, we've eliminated the regulations for any systems where humans have lost control.

The second set of definitions swaps "human-defined" objectives for "explicit or implicit objectives," but suffers the same problem. An implicit objective is one that is suggested[258] or implied[259] by the explicit objective. Again, a rogue system that is operating on objectives that are unrelated or adverse to human-defined objectives would not be pursuing either "explicit or implicit objectives." That means it would not be covered by the definition, so it would not be covered by the regulation. Again, by limiting the definition of AI to systems that are following proper objectives, we leave rogue systems beyond the scope of the regulations.

[248] OECD Recommendation, *supra* note 228 (later amended to replace "human-defined objectives" with "explicit or implicit objectives.").

[249] Australia's Safe and Responsible AI, *supra* note 185.

[250] Chile's Safety Framework, *supra* note 182.

[251] Peru Use of AI, *supra* note 187.

[252] US EO 14110, *supra* note 239.

[253] OECD Recommendation, *supra* note 231.

[254] The Framework, *supra* note 185.

[255] Act 3(1), *supra* note 183.

[256] S.B. 1047, *supra* note 234.

[257] *See* Part III.

[258] Cambridge Online Dictionary, IMPLICIT, (defining "implicit" as "suggested but not communicated directly") https://dictionary.cambridge.org/us/dictionary/english/implicit#google_vignette.

[259] Restatement (Third) Of Agency § 2.01 (2006) ("'Implied authority' is often used to mean actual authority either (1) to do what is necessary, usual, and proper to accomplish or perform an agent's express responsibilities or (2) to act in a manner in which an agent believes the principal wishes the agent to act based on the agent's reasonable interpretation of the principal's manifestation in light of the principal's objectives and other facts known to the agent.").

The U.S. NIST has a better approach. Its definition of AI was expressly adapted from the OECD definition, but it dropped both clauses and just referred to AI systems "that can, for a given set of objectives, generate outputs," dropping any qualifiers about what objectives are proper.[260]

H. Inference

A key element in 30 definitions (29%) is that the system "infers . . . how to generate outputs."[261] "Infer" is not well defined.

1. Inference in Common Usage

In common usage, to "infer" means to "form an opinion or reach a conclusion through reasoning and information."[262]

This definition is a philosophy landmine. It would ask judges to decide whether machines opine, conclude and reason. Philosophers have grappled with these questions since computers were first built.[263] If a machine can speak as a human, is it thinking as a human?[264] Is moving electrons around a binary circuit board very different from moving neurotransmitters across a synapse? Are machines reasoning? Are they conscious?[265]

The answers are not clear. Those developing AI systems would build in the shadow of this uncertainty.

The common definition of "infer" is not clear, so we can turn to how "infer" is used in artificial intelligence research, looking first at logic-based systems then at machine learning systems.

2. Inference in Logic-Based Systems

In logic-based systems, inference means deriving new statements of knowledge from known statements of knowledge.[266] "New" does not mean never before discovered; it merely means that the assertion was not already in the system's knowledge base.[267]

Systems do this by applying rules that were written into the code by the programmer.[268] Perhaps the programmer included the transitive property. If so, the system could use the statements "planets are bigger than elephants" and "elephants are bigger than mice," to infer "planets are bigger than mice."

The rules inserted into the program are the system's model of the world. Inference is the process of applying the system's model to known statements to develop new statements.

---

[260] NIST, *supra* note 186 (expressly stating it is adapted from the OECD Recommendation of the Council on Artificial Intelligence).

[261] Act 3(1), *supra* note 183.

[262] Merriam-Webster, INFER, https://www.merriam-webster.com/thesaurus/inferring.

[263] Turing, *supra* note 87 (arguing it is); John R. Searle, *Minds, Brains, and Programs*, 3 BEHAVIORAL & BRAIN SCI. 417 (1980) (arguing it is not).

[264] Turing, *supra* note 87.

[265] Kurzweil, *supra* note 81 at 55–59.

[266] Russell & Norvig, *supra* note 20 at 227. (Defining inferences as "deriving new sentences from old sentences" and stating that a sentence, used in the technical sense, "represents some assertion about the world.").

[267] *Id.*

[268] *Id.*

3. Inference in Machine Learning

In systems that use machine learning architectures, like ChatGPT, inference is similar but the model is generated by the system.[269]

Recall from the discussion of learning in Part III.F., a machine learning system works by "observ[ing] some data, build[ing] a model based on the data, and us[ing] the model as both a hypothesis about the world and a piece of software that can solve problems."[270]

Rather than tune the model directly, the programmer provides rules and methods for the program to fine tune the model in response to data. As more data comes in, the model adjusts, and if all goes well, the model better predicts the world.

Like in logic-based systems, the model in a machine learning system reflects the statistical correlations the system will find in the environment.[271] For example, the machine learning model might reflect that if it's sunny today then there's an 80% chance it will be sunny tomorrow, but if it's raining today, there's only a 40% chance it will be sunny tomorrow. The programmer doesn't give these numbers to the machine learning system; the system reached these values by analyzing massive amounts of data.[272]

Like in logic-based systems, inference in machine learning systems is the process of applying data to the system's model to generate new information.[273] Using the weather example above, if we tell the system that today is sunny, the system applies that data to the model and infers that there is an 80% chance tomorrow will be sunny.

The same concepts apply in large language models. In large language models, the model reflects the correlations between words or parts of words[274] in the training data.[275] When new data is entered, for example, a user prompt, the system applies its model to the data to predict the best response, which it provides to the user.[276]

For example, if the user's prompt includes the word "bat," then the model may predict that a successful response includes concepts correlated with flying, nighttime or vampires. That is, unless the

---

[269] Russell & Norvig, *supra* note 20.

[270] *Id*.

[271] This is not to say the model understands the world; the model is a set of rules that the system will apply to the world as though they were true. *Id*.

[272] *Id.* at 683–90. Researchers estimate GPT-4 was trained on a petabyte (10^15 bytes) of data. Erika Balla, *Here's How Much Data Gets Used By Generative AI Tools For Each Request*, DATA SCI. CENT. (Nov. 28, 2023), https://www.datasciencecentral.com/heres-how-much-data-gets-used-by-generative-ai-tools-for-each-request/.

[273] *See* Ashish Vaswani et al., *Attention Is All You Need,* 5 ARXIV (Aug. 2, 2023), https://arxiv.org/pdf/1706.03762.

[274] *Id.*

[275] *Id.*

[276] Early systems predicted only the next word of the response rather than predicting the entire response at once. *Id.* New models (sometimes called chain-of-thought or reasoning models) predict the entire response before responding. Zhiyuan Zeng, *Scaling of Search and Learning: A Roadmap to Reproduce o1 from Reinforcement Learning Perspective*, ARXIV (Dec. 18. 2024), https://arxiv.org/abs/2412.14135. And these systems usually select a word that's *near* the best, rather than the best, because it yield more natural responses. Stephen Wolfram, *What is ChatGPT Doing and Why Does It Work?*, STEPHEN WOLFRAM | WRITINGS (Feb. 14, 2023), https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/.

prompt also includes the words "first base" or "home run," then the model may predict that the response should include concepts related to baseball.[277]

In a large language model, inference is the process of predicting the right response, given the user's input and the system's model.[278]

4. Inference and Complexity

The problem with using "infer" as part of the definition of artificial intelligence is that it doesn't give any sense of the complexity of the model, so it will include systems that no one would rightly consider artificial intelligence.

A model can be as simple as a chart converting cups to liters. And perhaps that's all the system does. Such a system would meet the definition of inference because it is deriving new information (the amount in liters) from existing information (the amount in cups) based on its internal model (0.24 liters to a cup).

"Inference" would include a mercury thermometer that converts information about molecular kinetic energy into a statement about the temperature, or a sundial that converts information about the angle of the sun into the hour.

One might argue that neither of these machines are reasoning; they aren't thinking or understanding, so they are not inferring.[279] But as discussed above, there's no evidence that any inanimate objects think or reason. If reason is a requirement, no systems are covered, so the regulation serves no purpose. If reason is not required, we are back to calling sundials artificial intelligence.

Most definitions that rely on "inference" provide no clarity, but the EU AI Act attempts to. Its recitals add that the "capacity of an AI system to infer transcends basic data processing by enabling learning, reasoning or modelling."[280]

Unfortunately, the act does not define "basic data processing." And in 1936 Alan Turing showed that all computation is "basic data processing," just repeated for a long time.[281]

5. The Problem of Turing Machines

Turing's thought experiment described a remarkably basic machine.[282] The machine consists of a scanner sitting atop a long paper tape with symbols on it.[283] The scanner can do only five things: (i) scan the symbol directly under the scanner, (ii) modify that symbol, (iii) move one space left or right, (iv) select which ruleset the scanner will use for the next scan, or (v) halt and end the program.[284] The

---

[277] *See* Vaswani, *supra* note 273.

[278] Wolfram, *supra* note 276; Zeng, *supra* note 273. Technically, it is not optimizing for the user's needs, it is minimizing a loss function, which is a function given by the programmer to approximate the user's utility. Russell & Norvig, *supra* note 20.

[279] Colloquially "inference" requires reasoning and understanding. *See, e.g.*, MERIAM-WEBSTER, INFERRING, https://www.merriam-webster.com/thesaurus/inferring (last visited Jan. 22, 2025) ("To form an opinion or reach a conclusion through reasoning and information."); CAMBRIDGE ONLINE DICTIONARY, ASSERTION, https://dictionary.cambridge.org/us/dictionary/english/assertion (defining assertion as "a statement that you strongly believe is true").

[280] Recital 12 *supra* note 186.

[281] Turing, *supra* note 79.

[282] *Id.*

[283] *Id.*

[284] *Id.*

machine goes through a cycle in which it scans then takes the actions dictated by the ruleset, then scans, then takes the actions dictated by the ruleset, and it continues this loop until the ruleset tells it to halt.

Turing showed that with the right tape inserted and the right rulesets, this basic machine could replicate any computer running any computation.[285] Because all computation could be computed by these machines, Turing named them universal logical computing machines,[286] though we now refer to them as universal Turing machines.[287] The concept of a universal Turing machine transformed computation because it provided theoretical support for a single piece of hardware doing any type of computation.[288] This meant that changing the program of a machine did not require reconfiguring the wires; programming could be done by changing the ruleset on which the machine operated.[289] Coding replaced engineering.

Turing showed that a simple scanner, a rulebook and a roll of paper tape can replicate any computer system with any amount of complexity.[290] Marvin Minksy generalized this to show the rulebook only needed seven instructions and the tape only four symbols.[291]

That includes every computation. In other words, every brilliant feat of artificial intelligence can be replicated by a little machine that's simpler than a pocket calculator. Turing showed even seemingly complex computation is "basic data processing," but with a longer paper tape. Trillion parameter models are just basic data processing in bulk.

So defining "inference" to "transcends basic data processing" is not only vague, it's not clear it creates any limits at all.

6. Replacing Inference with Model Design

A more productive definition might instead refer to how the model is created, rather than how the inference is conducted. Regulations are designed to cover the most powerful models, and the most powerful models all use machine learning techniques.[292] Tying the regulation to these methods would

---

[285] *Id.* at 232; A.M. Turing, *Intelligent Machinery*, NAT'L PHYSICAL LAB'Y, at 4 (1948), https://www.npl.co.uk/getattachment/about-us/History/Famous-faces/Alan-Turing/80916595-Intelligent-Machinery.pdf (describing this as universal logical computing machines).

[286] *Id.* at 4.

[287] Jack Copeland, *What Is a Turing Machine?*, ALANTURING, (July 2000), https://www.alanturing.net/turing_archive/pages/reference%20articles/what%20is%20a%20turing%20machine.htm.

[288] Turing, *supra* note 282 at 5 ("Nearly all of the [practical computing machines] now under construction have the essential properties of the "Universal Logical Computing' machines mentioned earlier."); *see also*, Copeland, *supra* note 284 at 4 ("We do not need to have an infinity of difference machines doing different jobs. A single one will suffice. The engineering problem of producing various machines for various jobs is replaced by the office work of 'programming' the universal machine to do these jobs.").

[289] *Id*. at 5 ("Nearly all of the [practical computing machines] now under construction have the essential properties of the "Universal Logical Computing' machines mentioned earlier."); *see also*, *Id*. at 4 ("We do not need to have an infinity of difference machines doing different jobs. A single one will suffice. The engineering problem of producing various machines for various jobs is replaced by the office work of 'programming' the universal machine to do these jobs.").

[290] Turing also argued that all computation could similarly be done by a human following this process on a piece of paper, which he referred to as a "paper machine." *Id*.

[291] Marvin Minsky, *Recursive Function Theory*, 5 PROC. OF SYMP. IN PURE MATHEMATICS 229, 237 (1962).

[292] Currently those include very large data samples, stochastic gradient descent, loss minimization, annealing and other methods. *See generally*, ANIL ANATHASWAMY, WHY MACHINES LEARN *passim* (2024).

clarify which systems are included within the regulation and would exclude systems that operate on basic systems with fixed models.

One downside to this approach is that it would exclude AI systems designed to operate solely on logical rules[293] or rules developed by experts,[294] and other systems now quaintly referred to as "good old fashion AI."[295]

But if the goal is to regulate the most powerful models, excluding these logic-based systems may not be a problem because they have not shown the same capacity as machine learning models. For example, chess is a deterministic game, so one might think that rules-based and logic-based systems would dominate.[296] But the leading chess model, AlphaZero, beat the prior champion by introducing randomness into its process and relying on machine learning and statistics.[297] Similarly, AlphaGo defeated world champions not by learning the Go proverbs memorized by every beginner, but by learning statistical models of uncertainty.[298] The dominant models are machine learning models.

Refocusing on how the system learns, rather than how the system conducts inference, would better define the most powerful models.

I. Generates Output

Forty-five definitions (42%) of AI require it to generate certain types of output.

For example, the OECD definition says the AI system ". . . infers . . . how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments."[299] These four outputs—predictions, content, recommendations and decisions–are found in definitions in Australia,[300] Canada,[301] the EU[302] and Hong Kong.[303]

Many definitions treat the outputs as optional. The OECD definition lists these outputs as examples, not an exclusive list. Ten definitions of AI allow any output type.[304] If anything meets the definition of output, then specifying "output" serves no purpose in the definition.

This section will first address the specific terms, then discuss other areas that may be included.

---

[293] Russell & Norvig, *supra* note 20 at 232–33.

[294] *Id.* at 40–44.

[295] *Id.*

[296] Anian Ruoss et al., *Grandmaster-Level Chess Without Search,* ARXIV (Feb. 2024), https://arxiv.org/html/2402.04494v1.

[297] Specifically, AlphaZero uses Monte Carlo tree search, in which the system tests randomly generated moves. Russell & Norvig, *supra* note 20 at 232–33.

[298] AlphaGo, DEEPMIND, https://deepmind.google/research/breakthroughs/alphago/

[299] OECD Recommendation, *supra* note 228.

[300] Australia's Safe and Responsible AI, *supra* note 188.

[301] Canada's Bill C-27, *supra* note 195.

[302] Act 3(1), *supra* note 183; The Framework, *supra* note 185.

[303] Hong Kong's Framework, *supra* note 236.

[304] *See* Appendix A.

1. Predictions

Thirty-three definitions include "predictions" as the system's output.[305]

In common usage, a prediction is a statement about the future.[306] In AI research "prediction" likely covers all machine learning systems because of the way these systems work.[307]

Recall from the discussion of learning and inference above that machine learning models use large amounts of data to train a model, then use the model to predict the optimal response.[308] Both training and inference are based on prediction.

Prediction is critical to many training methods. For example, image diffusion models train a model to generate images by collecting millions of images, degrading the quality of each image until it is unrecognizable, then train the model so that it can produce the missing pieces in a way that reflects the original data.[309] During the training process the system adjusts the model's parameters so that it is better at predicting the original, non-degraded data.[310] It's all based on prediction.

And as noted in Part III.H.3, inference for statistical models is all prediction. For example, models like ChatGPT do not think through their entire response before answering a question.[311] They work by attempting to predict the word[312] that best[313] meets the user's needs given the user's prompt and the words that it has already written.[314]

Predictions reach beyond just generative content. Robotics applications also use prediction methods because sensors cannot gather all relevant information, the information may be indeterminate or the environment may be changing.[315] Because information is imperfect, robotic agents may use statistical

---

[305] *See* Appendix A.

[306] *Prediction*, CAMBRIDGE ONLINE DICTIONARY, https://dictionary.cambridge.org/us/dictionary/learner-english/prediction ("the act of saying what you think will happen in the future").

[307] *See, e.g.*, Vaswani, *supra* note 270 (foundational paper for the development of large language models, which uses "predict" to refer to the next token in a generative response); Jonathan Ho et al., *Denoising Diffusion Probabilistic Models,* ARXIV (June 19, 2020), https://arxiv.org/abs/2006.11239 (foundational paper proposing diffusion in image generation, which uses "predict" to discuss inference on a current state).

[308] *See* Parts III.F, H.3.

[309] Ho, *supra* note 307 ("A diffusion probabilistic model . . . is . . . trained using variational inference to produce samples matching the data after finite time. Transitions of this chain are learned to reverse a diffusion process, which is a Markov chain that gradually adds noise to the data in the opposite direction of sampling until signal is destroyed.").

[310] *Id.*

[311] Wolfram, *supra* note 276.

[312] More accurately, this is a token, which may be a word or a portion of a word, like "ing" or "ization," but that distinction is not relevant here. *Id*.

[313] This is not usually the *most* likely word. Generative models that yield the most likely next word tend to feel rigid and unnatural, so models often include a "temperature" parameter that can be adjusted to prefer words that are further down the list of most likely next words. *About Generative Models*, GOOGLE AI FOR DEVELOPERS (Aug. 2024), https://ai.google.dev/gemini-api/docs/models/generative-models.

[314] Wolfram, *supra* note 276.

[315] Russell & Norvig, *supra* note 20.

methods to estimate optimal strategies and actions.[316] For example, selecting a route requires a robotic taxi to predict traffic, and driving that route requires the taxi to predict other cars' movements.

So "prediction" likely includes all machine learning systems.

2. Content

Thirteen definitions define AI by its ability to generate "content."[317] "Content" is broad to the point of meaninglessness, defined as "information, images, video, etc. that are included as part of something such as a website."[318] The "etc." in the definition signals the breadth of this term, which essentially means something that is contained in something else,[319] ranging from video content on a website to questionable content of old Tupperware. Outside of abstract concepts, it is difficult to imagine something that does not meet the definition of content, making this term useless in defining any boundaries of artificial intelligence.

3. Recommendations

Twenty-four definitions state that AI generates recommendations.[320] These definitions do not distinguish based on the consequences of the recommendation, so the term captures everything from parole and sentencing recommendations to Netflix's trending videos section.

"Recommendations" is a subset of "predictions," a term that is often used along with it.[321] A recommendation carries with it an implicit prediction that the recommended action will best accomplish the user's objectives. Netflix recommends Goonies because its algorithm predicts that you will enjoy Goonies. Unless the recommendation is generated randomly, it was preceded by a prediction.

So a recommendation implies a prediction. But a recommendation may be more worthy or regulation because it may have a greater likelihood of skewing human behavior. And machine learning systems typically have internal architecture that overstates its confidence levels.

Suppose a system is trying to determine what digit someone wrote by hand on a black and white touchscreen display. The system might first read the value of each pixel in the display area, compare the data to the system's model and calculate how far off the doodle is from being the ideal zero, how far off from being the ideal one, two, three, four and all the way through nine.[322] This gives a numerical score to

---

[316] *Id.*

[317] Act 3(1) *supra* note 180; The Framework *supra* note 182; OECD AI Principles, *supra* note 236; Australia's Safe and Responsible AI supra note 185; Canada's Bill C-27 *supra* note 192; Hong Kong's Framework, *supra* note 239.

[318] Cambridge Online Dictionary, CONTENT, https://dictionary.cambridge.org/us/dictionary/english/content (last visited Jan. 12, 2025).

[319] Etymologically, "content" derives from the same origins as "contained," and has a similar meaning—it is something that is part of something else. Information embodied in any physical system is content. Etymonline, CONTENT, https://www.etymonline.com/word/content (last visited Jan. 12, 2025). ("Latin contentus 'contained; satisfied,' past participle of continere 'to hold together, enclose,' . . .").

[320] Act 3(1), *supra* note 180; The Framework *supra* note 182; OECD AI Principles, *supra* note 236; Australia's Safe and Responsible AI supra note 185; Canada's Bill C-27, *supra* note 192; Hong Kong's Framework *supra* note 236; Chile's Safety Framework, *supra* note 182; Peru Use of AI *supra* note 187; US EO 14110, *supra* note 236; NIST, *supra* note 186; Automated Decision-Making, *supra* note 192; Singapore's AI Framework, *supra* note 192; Singapore's Financial Sector, *supra* note 192; UK AI Guidelines, *supra* note 194.

[321] *See, e.g.,* Act 3(1), *supra* note 183; The Framework *supra* note 185; OECD AI Principles, *supra* note 239.

[322] This is done with a loss function, which often involves logarithms. Russell & Norvig, *supra* note 20 at 808.

judge the doodle among the various options.[323] From there, the system will typically convert these measurements into probabilities, so rather than a numeric measure of how far a doodle is from the ideal four, we have the probability that the image is a four with some confidence level.[324] This step is where the problem lies.

The function that generates the probability is called the softmax function.[325] Because the softmax function uses exponents and logarithms,[326] the function amplifies small differences between the measurements into much larger differences in probabilities, which has the effect of forcing more certainty than is justifiable.[327] One influential study found that these systems "often fail silently by providing high-confidence predictions while being woefully incorrect."[328] Anyone that has argued with OpenAI's o1 model has experienced this first hand.

Overconfidence in detecting handwriting is unlikely to have consequences beyond some delayed mail. But imagine the AI system is making recommendations to a parole board.[329] If the system overstates its confidence level because of a softmax function, it could give an erroneously high confidence level in a recidivism prediction, leading a board to deny parole.

This suggests some recommendation systems should be tested for both accuracy on their prediction and their confidence levels.

4. Influencing Physical or Virtual Environments

The most common definition denoting output types includes "influencing physical or virtual environments." In the official English version, this is a dangling modifier that could modify "decisions" alone or "predictions, content, recommendations and decisions." The unofficial translation provided by the OECD (and adopted by Peru[330]) clarifies that this clause is meant to modify "predictions, contents, recommendations and decisions."[331] That seems a better reading, though the clause is useless either way.

---

[323] *Id.*

[324] *Id.* at 809.

[325] *Id.*

[326] The softmax function is $softmax(\boldsymbol{in})_k = \frac{e^{in_k}}{\sum_{k'=1}^{d} e^{in_{k'}}}$ where **in** represents a vector of input values and d represents the possible outputs, with the formula giving the value for the *k*th element of the vector d. *Id.*

[327] *Id.* at 809; Wolfram, *supra* note 276 ("The very last operation in the network is a so-called softmax which tries to 'force certainty'.").

[328] Dan Hendrycks & Kevin Gimpel, *A Baseline for Detecting Misclassified and Out-Of-Distribution Examples in Neural Networks*, ARXIV (Oct 2, 2018), https://arxiv.org/pdf/1610.02136. ("These high-confidence predictions are frequently produced by softmaxes because softmax probabilities are computed with the fast-growing exponential function. Thus minor additions to the softmax inputs, i.e. the logits, can lead to substantial changes in the output distribution. . . . [W]e establish that the prediction probability from a softmax distribution has a poor direct correspondence to confidence.").

[329] The challenges of statistical methods in parole hearings have been extensively discussed. *See* Pari McGarraugh, *Up or Out: Why "Sufficiently Reliable" Statistical Risk Assessment Is Appropriate at Sentencing and Inappropriate at Parole*, UNIV. OF MINN. LAW SCHOOL (2013); Christian *supra* note 56.

[330] Peru Use of AI, *supra* note 187.

[331] OECD, RECOMENDACIÓN SOBRE LA INTELIGENCIA ARTIFICIAL (unofficial translation) 3, ("Un sistema de IA es un sistema basado en máquinas que, para objetivos explícitos o implícitos, infiere, a partir de los datos de entrada que recibe, cómo generar información de salida como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos reales o virtuales.").

The definition includes changes to the "physical or virtual environment." It is difficult to imagine a system that would not meet this clause.

J. Quantitative Definitions

A few definitions of AI are based on quantitative definitions.

For example, the EU AI Act categorizes a general-purpose AI model as having "systemic risk" if training the model required more than 10^25 FLOP.[332] A FLOP, or floating point operation, is a basic arithmetic operation (addition, subtraction, multiplication or division) performed on two numbers in decimal format.[333]

A California bill would have covered models trained using 10^26 FLOP and costing one hundred million dollars to train.[334] The dollar value would adjust for inflation,[335] and the compute value could be updated by an existing regulatory agency.[336]

This is a useful approach in many ways. First, it addresses "models" rather than "systems." The model is the set of correlations,[337] not the system that operates it. It is what the system knows, not the system itself. By regulating based on the model's training data, the regulation avoids the boundary disputes of "systems" discussed above. Quantitative metrics also provide brighter lines for developers, reducing legal risks.

The downside of a quantitative approach is that its effectiveness will depend on selecting the right metrics. We don't know what metrics will correlate with developing human-level or superhuman-level AI.

For example, early generative language models required massive compute to train before the user entered a prompt but very little compute to run. This was the standard way to design a large language model until OpenAI's o1 model reached new frontiers primarily by increasing the compute used *after* the user enters a prompt.[338] This technique has been copied by smaller models to multiply their capabilities.[339] A regulation that focused purely on training compute might miss these highly capable, smaller models because they achieved their results without increasing training compute.

---

[332] EU AI Act Art. 51(1)(a), (2). The EU Commission retains authority to amend these thresholds and create supplemental benchmarks and indicators. These include "algorithmic improvements and increased hardware efficiency." *Id*. at 51(3).

[333] Lennart Heim, *FLOP for Quantity, FLOP/s for Performance*, HEIM BLOG (Apr. 14, 2023), https://blog.heim.xyz/flop-for-quantity-flop-s-for-performance/ ("One FLOP is equivalent to one addition, subtraction, multiplication, or division of two decimal numbers.").

[334] S.B. 1047, *supra* note 231. The bill would also cover models that are fine-tuned using a lower amount of compute, and those trained with 10^26 integer operations. *Id.*

[335] *Id.* at §3.

[336] *Id*.

[337] Russell & Norvig, *supra* note 20 at 669.

[338] System Card, *supra* note 33.

[339] Ben Dickson, *Hugging Face Shows How Test-Time Scaling Helps Small Language Models Punch Above Their Weight*, VENTUREBEAT (Dec. 20, 2024), https://venturebeat.com/ai/hugging-face-shows-how-test-time-scaling-helps-small-language-models-punch-above-their-weight/.

As a second example, the California bill discussed above was enrolled on September 3, 2024, limiting itself to only models whose training cost at least $100 million.[340] Less than four months later, DeepSeek-V3 debuted with a training cost of around $6 million[341] and surpassed several frontier models in capability.[342] In a span of four months the cost of outperforming nearly every model dropped 94%.

Even if we pick the right metrics, they may be difficult to measure. Measurements like the cost of development can vary based on electricity prices and local wage dynamics.[343] Measurements like compute will have to consider inference-time versus training compute, and for adaptive models this may be a hard line to draw. Measurements like data will have to consider changes to compression techniques, whether the machine continues to adapt and learn after deployment[344] and whether the data contains a high amount of information.[345] Each of these metrics will have to address how to classify models that are built on other models.[346]

Finally, regulators using quantitative definitions will need to determine how these quantities are verified. DeepSeek's announcement of its latest models were met with some skepticism that the developers exaggerate the low cost of development to better position the Chinese model to promote Chinese nationalist propaganda or to benefit a hedge fund with a short position in a competitor.[347]

While quantitative metrics provide the clearest rule for developers, they are likely to become outdated, and regulators will struggle to know which metrics will be predict a system's capabilities. Even if the right metrics are selected, measurement and verification are likely to be a challenge.

K. Mimics Human Intelligence

Thirty-seven definitions (35%) of AI include mimicking human intelligence or "acting intelligently." As discussed in Part II.A, there is no operationalizable definition of intelligence for human beings, let alone inanimate systems, and as discussed in Part II.B.3., defining intelligence as the ability to mimic

---

[340] Softmax, *supra* note 326.

[341] DeepSeek-AI et al., DEEPSEEK-V3 TECHNICAL REPORT, CORNELL UNIV. 1, 5 (Dec. 27, 2024) (stating the model was trained using only 2.788M H80 GPU hours at a cost of around $5.8 million).

[342] *Id.* at 31 (showing superiority over Claude 3.5 Sonnett and GPT-4o in math, coding and Chinese and showing mixed results in English).

[343] California's SB 1047 sought to avoid this complication by basing the price calculation on "the average market price of cloud compute at the start of training as reasonably assessed by the developer." S.B. 1047, *supra* note 234 at §3.

[344] Many researchers are looking for ways for LLMs to continue to learn after deployment. *See, e.g.*, Xinyu Guan et al., rStar-Math: Small LLMs Can Master Math Reasoning with Self-Evolved Deep Thinking, CORNELL UNIV. (Jan. 8, 2025), https://arxiv.org/abs/2501.04519 (providing a self-training method that surpasses some top models in math).

[345] Information is a deep topic beyond the scope of this paper. To understand the challenge, consider that a Tesla car transmits massive amounts of data for improving the driving model, but the model has already mastered most scenarios the car is in, so only 1 mile out of every 10,000 miles provides useful training data for the model. Elon Musk (@elonmusk), X (May 7, 2024), https://x.com/elonmusk/status/1787768103449010597. If AI is categorized by the amount of data processed, that will be much larger than if it focused on the novel information. *See generally* C.E. Shannon, *A Mathematical Theory of Communication*, 27 BELL SYS. TECH. J. 379 (July 1948).

[346] *See, e.g.*, Meta Llama 3.3, META, https://www.llama.com/. DeepSeek-R1 was built on the DeepSeek-V3 model.

[347] Tereza Tizkova, *DeepSeek vs Conspiracies*, SUBSTACK (Jan. 28, 2025), https://terezatizkova.substack.com/p/deepseek-vs-conspiracies.

human abilities is likely to under-count powerful models that do not act human, and over-count weak models that appear human-like.

Definitions that rely on a court to define human intelligence are likely to chill investment while innovators await philosophy decisions from the bench.

On the other hand, using human intelligence as a lower bound avoids many of these concerns. So a definition that says the system operates at least at the level of human intelligence would avoid being under-inclusive of systems that are superhuman.

L. Specific Types of Architecture

Eighteen definitions of AI (17%) include a specific method of artificial intelligence.[348] Seven of these include statistical modeling, six mention neural networks and the rest refer to machine learning.[349]

Definitions that refer to specific techniques or architectures are useful because they are easier to understand and provide clear rules. They are also less likely to include things that clearly should not be included, like a sundial.

A downside is that a model's power is a combination of techniques, architecture, data and compute. A machine learning model using statistical inference trained on a single dog picture will not be worth regulating. Architecture could be a useful component of a definition, but on its own it does not correlate with system power.

These definitions may also imply limits on the system's underlying substrate. That is, if a definition requires the system use statistical modeling, then it will exclude biological and chemical systems.

M. Flexible Capability Assessment

One definition considers a variety of factors considered together. Specifically, the EU considers general-purpose models to have systemic risk if it "has high impact capabilities evaluated on the basis of appropriate technical tools and methodologies, including indicators and benchmarks." The Act considers whether it can do distinct tasks, learn and scale.[350] Non-technical factors include the number of end users. And the Commission is authorized to amend these factors so that they always "reflect the state of the art."[351]

The flexibility of this approach is likely to exclude silly examples like sundials and plows, and allowing the Commission to modify the factors is more likely to keep them current.

On the other hand, from January 2024 to January 2025 training costs dropped by over 90%[352] and frontier models entered a new paradigm of post-prompt compute.[353] It is unlikely that even an empowered regulator could keep up with the changes. A flexible capability assessment that considers a variety of factors may face the challenges of each of those factors.

---

[348] *See* Appendix A.

[349] *See* Appendix A.

[350] EU AI Act Annex XIII [hereinafter Annex XIII].

[351] EU AI Act Art. 51(3).

[352] DeepSeek-AI et al., DEEPSEEK-V3 TECHNICAL REPORT, CORNELL UNIV. 1, 5 (Dec. 27, 2024).

[353] OpenAI, *OpenAI o1 System Card*, OPENAI (Sep. 12, 2024), https://cdn.openai.com/o1-system-card-20240917.pdf.

Part IV. An Example: The EU AI Act

Having broken down the elements of AI regulation definitions, this section will now consider the most common definition holistically. We will review the definitions in the act, but to understand the consequences of falling within those definitions, this Part will also provide a summary of the regulations imposed.

The EU AI act was the first comprehensive, sectoral regulation of AI.[354] It adopted the most common definition of artificial intelligence, making it a useful example for this review.[355]

This section gives a general overview of the EU AI Act then analyzes the Act's definition of AI.

A. Overview of the EU AI Act's Framework: Levels of Risk

The EU AI Act defines an "AI system" as

> [A] machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.[356]

The act then categorizes AI systems by their risk level, applying different obligations and restrictions based on the capability or intended use of the model.[357]

1. Prohibited Practices

Under the Act, some practices are prohibited.[358] These include "placing on the market or putting into service" an AI system that engages in subliminal messaging;[359] exploits someone's age, disability or socio-economic situation;[360] uses discriminatory social credit systems;[361] conducts pre-crime risk assessment profiling;[362] scrapes facial images from CCTV for facial recognition databases;[363] tracks

---

[354] *EU AI Act: First Regulation on Artificial Intelligence*, EUR. PARL. (Jun 8, 2023), https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

[355] *See* Part III.

[356] Act 3(1), *supra* note 180.

[357] Models that meet the definition of AI system but do not fit any of the categories above are still subject to minor regulations, including registration. EU AI Act Art. 49(2). Penalties include: (i) being removed from the market, EU AI Act Art. 83(2); 93, and (ii) fines up to 7% of global turnover or EUR 35,000,000, whichever is higher, EU AI Act Art. 99(3).

[358] EU AI Act 5.

[359] EU AI Act 5(1)(a).

[360] EU AI Act 5(1)(b).

[361] EU AI Act 5(1)(c). A social credit system is similar to a financial credit system, but it takes into account good behavior, with "good" defined by the credit scorer. Some have argued these systems promote authoritarian control. Lizzy Rettinger, *The Human Rights Implications of China's Social Credit System*, 21 J. HIGH TECH. L. 1, 16 (2021).

[362] EU AI Act 5(1)(d). This restriction does not apply when used to support human assessment of a person that's already based on "objective and verifiable facts directly linked to a criminal activity." *Id.*

[363] EU AI Act 5(1)(e).

emotions at school or work;[364] uses biometrics to deduce sensitive categories like political opinions, religion and sexual orientation;[365] or assists law enforcement in conducting non-targeted biometric identification systems in public spaces.[366]

2. High-Risk AI Systems

The second most restricted category is "high-risk" AI systems.[367] This part explains the two categories of high-risk and the regulations that apply to them.

a. What Is a High-Risk System?

The EU AI Act applies a high-risk label to certain (i) regulated products and (ii) high-risk applications.[368]

In the regulated product category, an AI system is high-risk if the system is a product (or safety component of a product) that must be certified as EU compliant under a variety of regulations.[369] The list of regulated products is long and complicated, including certain types of machinery,[370] toys,[371] recreational watercraft,[372] elevators,[373] safety equipment for use in potentially explosive environments,[374]

---

[364] EU AI Act 5(1)(f). There's an exception here for health and safety. One school shooting could cause this exception to swallow the rule.

[365] EU AI Act 5(1)(g). Another one of these categories is race, meaning biometric data cannot be used to determine someone's race. The prohibition doesn't prevent labelling lawfully acquired biometric datasets based on biometric data.

[366] EU AI Act 5(1)(h). Exceptions include targeted searches for missing persons; preventing specific, substantial and imminent safety threats; and locating certain criminals, including rapists, child pornographers, murderers, hijackers, terrorists and gang members. These excepted searches must be approved by a judge or administrative authority, EU AI Act 5(3), and reported to the national data protection authority. EU AI Act 5(4).

[367] EU AI Act 6.

[368] *Id*.

[369] EU AI Act 6(1) [hereinafter Act 6(1)]; Annex I [hereinafter Annex I]. Note that the AI system must either be the product (as defined in the annexed legislation), or it must be a safety component of the product. *Id.* The EU AI Act defines a "safety component" as "a component of a product or of an AI system which fulfils a safety function for that product or AI system, or the failure or malfunctioning of which endangers the health and safety of persons or property;" Art. 3(14).

[370] Act 6(1), *supra* note 366; Annex I *supra* note 366; European Parl. and Council, Directive 2006/42/EC, (May 17, 2006).

[371] Act 6(1), *supra* note 366; Annex I *supra* note 366; European Parl. and Council, Directive 2009/48/EC, (June 18, 2009).

[372] Act 6(1), *supra* note 366; Annex I *supra* note 366; European Parl. and Council, Directive 2013/53/EU, (Nov. 20, 2013).

[373] Act 6(1), *supra* note 366; Annex I *supra* note 366; European Parl. and Council, Directive 2014/33/EU, (Feb. 26, 2014).

[374] Act 6(1), *supra* note 366; Annex I *supra* note 366; European Parl. and Council, Directive 2014/34/EU, (Feb. 26, 2014).

radio equipment,[375] pipelines,[376] funiculars,[377] aerial cableways,[378] personal protective equipment,[379] gas stoves,[380] furnaces,[381] medical devices,[382] airport security,[383] motorcycles,[384] agricultural and forestry vehicles,[385] marine equipment,[386] railways,[387] motor vehicles and trailers[388] and aircraft.[389] If AI is the product regulated by these regulations or is intended to be used as a safety component of the product, then it is a high-risk AI system.[390]

---

[375] Act 6(1), *supra* note 366; Annex I *supra* note 366; European Parl. and Council, Directive 1999/5/EU, (Apr. 16, 2014). This includes products that broadcast or intentionally receive radio waves for communication. *Id.* at Art. 2(1)(1).

[376] Act 6(1), *supra* note 366; Annex I *supra* note 366; European Parl. and Council, Directive 2014/68/EU, (May 15, 2014).

[377] Act 6(1), *supra* note 366; Annex I *supra* note 366; European Parl. and Council, Directive 2000/9/EC, (Mar. 9, 2016).

[378] *Id*.

[379] Act 6(1), *supra* note 366; Annex I *supra* note 366; European Parl. and Council, Regulation 2016/425, (Mar. 9, 2016).

[380] Act 6(1), *supra* note 366; Annex I *supra* note 366; European Parl. and Council, Regulation 2016/426, (Mar. 9, 2016).

[381] *Id*.

[382] Act 6(1), *supra* note 366; Annex I *supra* note 366; European Parl. and Council, Regulation 2017/745, (Apr. 5, 2017), Regulation 2017/746, (Apr. 5, 2017).

[383] Act 6(1), *supra* note 366; Annex I *supra* note 366; European Parl. and Council, Regulation (EC) No 300/2008, (Mar. 11, 2008).

[384] Act 6(1), *supra* note 366; Annex I; European Parl. and Council, Regulation (EU) No 168/2013, (Jan. 15, 2013).

[385] Act 6(1), *supra* note 366; Annex I *supra* note 366; European Parl. and Council, Regulation (EU) No 167/2013, (Feb. 5, 2013).

[386] Act 6(1), *supra* note 366; Annex I *supra* note 366; European Parl. and Council, Directive 2014/90/EU, (July 23, 2014) [hereinafter EU 2014/90].

[387] Act 6(1), *supra* note 366; Annex I *supra* note 366; European Parl. and Council, Directive (EU) 2016/797, (May 11, 2016).

[388] Act 6(1), *supra* note 366; Annex I *supra* note 366; European Parl. and Council, Regulation 2018/858, (May 30, 2018); European Parliament and Council, Regulation(EU) 2019/2144, (Nov. 27, 2019).

[389] Act 6(1), *supra* note 366; Annex I *supra* note 366; European Parl. and Council, Regulation (EU) 2018/1139, (July 4, 2018).

[390] Act 6(1), *supra* note 366; Annex I *supra* note 366.

Second, an AI system is classified as high-risk if (i) the system is intended[391] to be used for the high-risk activities described in the rest of this section[392] and (ii) it poses a "significant risk of harm to the health, safety or fundamental rights of natural persons"[393] or it is intended to be used for profiling.[394]

The high-risk activities currently[395] include biometrics for identification, categorization or emotion recognition;[396] critical infrastructure;[397] school admissions and placement;[398] grading exams[399] or catching students cheating;[400] recruiting or evaluating employees;[401] evaluating eligibility for public assistance;[402] determining creditworthiness;[403] pricing health or life insurance;[404] or routing emergency phone calls or dispatching first responders.[405] A system is also high risk if it is intended to be used for "influencing the outcome of an election or referendum or the voting behaviour of natural persons in the exercise of their vote in elections or referenda," with an exception for AI systems in which the output is used only internally, for example a volunteer scheduling system.[406]

Some applications become high-risk if used by law enforcement and customs authorities, for example, systems intended to be used to assess the risk of a person becoming a victim,[407] determine whether

---

[391] The statutory structure described in this paragraph currently applies only to applications that are "intended" to be used in a certain way. EU AI Act Annex III [hereinafter Annex III]. But this intent requirement may not be permanent. Section 6(2) of the Act says that any use listed in Annex III is "high-risk." It does not consider intent. The "intent" limitation is found in each item listed in Annex III. EU AI Act 7(1) allows the Commission to amend the list in Annex III. Because the intent element is contained only in Annex III and because the Commission is authorized to amend Annex III, intent may not always be an element of this test. While intent is a factor the Commission must consider when amending Annex III, it is not a controlling factor. EU AI Act 7(2).

[392] Annex III, *supra* note 388.

[393] EU AI Act 6(2). The term isn't further defined, but the Act expressly excludes systems that perform a "narrow procedural task;" that look backwards to improve some prior human activity; that analyze decision-making patterns but aren't intended to influence the decision-making; or that are merely performing a "preparatory task for an assessment." EU AI Act 6(3).

[394] Act 6(1) *supra* note 366.

[395] The EU Commission can modify this list based on a number of factors including the system's purpose, data usage, autonomy, harm and benefits. EU AI Act 7.

[396] EU AI Act Annex III(1). There is an exception for using biometrics to verify a person is who the person claims to be.

[397] EU AI Act Annex III(2).

[398] EU AI Act Annex III(3).

[399] EU AI Act Annex III(3)(b).

[400] EU AI Act Annex III(3)(d).

[401] EU AI Act Annex III(4).

[402] EU AI Act Annex III(5).

[403] EU AI Act Annex III(5)(b).

[404] EU AI Act Annex III(5)(c).

[405] EU AI Act Annex III(5)(d).

[406] EU AI Act Annex III(8)(b).

[407] EU AI Act Annex III(6)(a).

someone is lying,[408] evaluate the reliability of evidence,[409] estimate recidivism[410] or profile natural persons.[411]

A system is also high-risk if it is intended to be used by customs and immigration authorities to determine whether someone is lying,[412] predict the likelihood that an immigrant poses a security risk,[413] evaluate eligibility for asylum or visas[414] or identify people in relation to migration, asylum or border control, other than to verify travel documents.[415]

A system is also high risk if it is intended to be used by judges to research or analyze cases or conduct mediation.[416]

This category can apply to any system used in the EU, including systems developed elsewhere.[417]

b. What Is Required of High-Risk Systems?

High-risk systems are regulated for risk mitigation, transparency and quality.

The risk mitigation regulations include having an ongoing risk management to identify risks to "health, safety and fundamental rights,"[418] and to adopt "appropriate and targeted risk management measures" to ensure that the risk is "acceptable."[419] "Acceptable" is not further defined, so it will likely be judged in hindsight after some harm has occurred.[420]

High-risk systems must also have pre-launch testing[421] and event logs[422] and be able to be "effectively overseen by natural persons" while in use.[423] It is not clear what "effectively overseen" means. On the low end, it may mean a human must respond to anomalies. On the high end, it may mean that these systems are limited to the processing speed of the supervising human.

---

[408] EU AI Act Annex III(6)(b).

[409] EU AI Act Annex III(6)(c); EU AI Act Annex III(7)(a).

[410] EU AI Act Annex III(6)(d).

[411] EU AI Act Annex III(6)(e).

[412] EU AI Act Annex III(6)(b).

[413] EU AI Act Annex III(7)(b).

[414] EU AI Act Annex III(7)(c).

[415] EU AI Act Annex III(7)(d).

[416] EU AI Act Annex III(8)(a).

[417] EU AI Act 22 requires that any provider from outside the EU authorize a representative to certify that the AI system conforms to the requirements of the EU AI Act. See also EU AI Act 47(2).

[418] EU AI Act 9(2)(a).

[419] EU AI Act 9(5); *see also* EU AI Act 72(1) (describing a post-market monitoring system).

[420] *See generally* Jeffrey J. Rachlinski, A Positive Psychological Theory of Judging in Hindsight, 65 Chi. L. R. 571 (1998).

[421] EU AI Act 9(6), (8).

[422] EU AI Act 12(3); it's not clear what this means.

[423] EU AI Act 14(1).

Second, there are transparency requirements, which include allowing market surveillance authorities to access the source code,[424] providing technical documentation on the model's development,[425] watermarking or otherwise identifying generated and manipulated media[426] and advising human users that they are interacting with AI.[427]

Other transparency requirements may not be technically feasible, like the requirement to describe the system architecture and the "relevance of the different parameters."[428] The parameters of large language models may not carry individual meaning.[429] And there may be 1.8 trillion of them.[430] It's not clear how this requirement is feasible.

Similarly, high-risk systems must be designed to allow deployers to "interpret a system's output and use it appropriately."[431] Again, it isn't clear how it is possible to ensure it is used "appropriately."

Finally, there are quality requirements. High-risk systems must be designed to achieve "an appropriate level of accuracy, robustness, and cybersecurity" and "perform consistently in those respects throughout their lifecycle."[432] They must be "as resilient as possible regarding errors, faults or inconsistencies."[433]

And, perhaps impossibly, they must be trained only on data that "to the best extent possible" is "free of errors and complete in view of the intended purpose."[434] GPT-4 was trained on 1,000,000,000,000,000 bytes of data.[435] It's not clear how that could be checked. Smaller training sets could be verified, but models trained on smaller data sets will not be as accurate, which may make users worse off.[436]

3. General-Purpose Models with Systemic Risk

The next category, "general-purpose models," is designed to capture models with generalized skills that are capable of performing a wide variety of tasks.[437]

---

[424] The market surveillance authority must request the source code and the source code must be necessary to assess conformity with the law and auditing and testing have proved insufficient. EU AI Act 74(13).

[425] EU AI Act Annex IV(2)(b).

[426] EU AI Act 50(2).

[427] EU AI Act 50(1).

[428] EU AI Act Annex IV(3).

[429] Part of the reason for this is that the parameters represent vectors on axes that have been shifted to reduce the number of dimensions. Because of this, there isn't an axis for something like "Year of invention" or "How large is it" Instead the architecture uses some vector as a new axis to reduce the number of dimensions, which reduces the computational burden. So, vector coordinates are given in axes that have no cognizable interpretation. Anil Ananthaswamy, WHY MACHINES LEARN 176–80 (2024).

[430] This is the number of parameters rumored to be in the GPT-4 model. Maximilian Schreiner, *GPT-4 Architecture, Datasets, Costs and More Leaked*, THE DECODER (July 11, 2024), https://the-decoder.com/gpt-4-architecture-datasets-costs-and-more-leaked/.

[431] EU AI Act 13(1).

[432] EU AI Act 15(1).

[433] EU AI Act 15(4).

[434] Act 10(3), *supra* note 200.

[435] Balla, *supra* note 269.

[436] Zeng *supra* note 273.

[437] EU AI Recital 97.

A subset of this category, general-purpose model *with systemic risk*, are subject to additional regulation.

a. Defining General-Purpose Models with Systemic Risk

The definition of these models has a quantitative element that considers the number of parameters, the quality and size of the data set, the computation used for training and the modalities (text, images, etc.) on which it can operate.[438] Non-technical factors include the number of end users and whether it can do distinct tasks, learn and scale.[439]

The act presumes a model is a general-purpose with systemic risk if it's trained using more than 10^25 FLOP.[440]

The Commission is authorized to amend these factors so that they always "reflect the state of the art."[441]

b. What Is Required of General-Purpose Models?

If the general-purpose model is deemed to have systemic risk (which is designed to include anything considered "state of the art"),[442] the provider must identify and mitigate systemic risks,[443] report any serious incidents,[444] and implement physical and cyber security.[445]

The developer of a general-purpose model with systemic risk must issue technical reports disclosing (among other things) the architecture, how it was developed, the design specifications (and the reasons for those choices), the training process, what the model is designed to optimize, how the data was obtained and methods to detect bias in the data.[446]

B. <u>The EU AI Act's Definition Problem</u>

This section will discuss the challenges likely to face a court defining artificial intelligence under the EU AI Act.

First, the definition of an "AI system" is exceptionally broad:

> [A] machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.[447]

---

[438] Annex XIII *supra* note 347.

[439] *Id.*

[440] EU AI Act 51(2). FLOP refers to floating point operations and measures the number of simple mathematical computations required to train the model. *See,* Heim *supra* note 330.

[441] EU AI Act 51(3).

[442] *Id.*

[443] EU AI Act 55(1)(c).

[444] *Id.* "Serious incidents" includes death or serious bodily harm, serious disruptions of critical infrastructure, violation of civil rights laws, or serious harm to property or the environment. EU AI Act 3(49).

[445] EU AI Act 55(1).

[446] EU AI Act Annex XI.

[447] Act 3(1), *supra* note 180.

Under this definition, Minecraft is an artificial intelligence system. It is a "machine-based system." It maintains a non-zero level of autonomy on how to calculate the world.[448] It does not exhibit adaptiveness, but that is optional.[449] And it infers how to generate content, namely it generates new information about the world through a model.[450]

Similarly, Microsoft Word is a machine-based system. It maintains far more autonomy than I would like, often choosing to correct perfectly valid phrases. It does not experience adaptiveness, though that element is optional. And it generates predictions based on inference models; even as I type this sentence it predicts and recommends ways to finish it.

Under the EU AI Act definition, it is difficult to think of a computer application that would not meet the definition of artificial intelligence.

A pocket calculator would probably meet the definition if not for the limit in the recitals, which states, "This Act does not apply to automated systems which merely calculate and implement formulas, including taxation and budgetary allocation, insofar as they automate a process of calculation which would otherwise be carried out manually and fully understood."[451]

But as noted in Part III.H.5., all computation is "merely calculate[ing] and implement[ing] formulas." The most advanced AI systems implement incredible statistical methods, but only by breaking them down into basic arithmetic.[452] All computation is done through basic arithmetic.

Next, the definition has the same challenges described in Part III.D. relating to systems. The term "system" has no reasonable boundary between seamlessly integrated software systems, so it will be unclear whether interactions with libraries or through APIs will cause other systems to be included as artificial intelligence under the act.

One might argue that "system" is limited here by the other words around it—it isn't any system, it's a "system *designed* to operate with … autonomy . . . ."[453] So the "system" would be limited to what was "designed." This argument fails because the problem is these systems are designed to operate as they do. They are designed to use libraries written and maintained by strangers, to plug into APIs and interact with other programs, and to search the internet seamlessly as though they were a single program. "Designed" does not solve the interaction problem because the interaction is core to the design.

Still, a broad definition of "systems" is unlikely to matter unless one of the systems is high-risk. As previously noted, almost every modern application is artificial intelligence under the EU AI Act, so there is no need to sneak any in through the loose definition of system. But if a system is high-risk, it will be difficult to tell what transparency, quality and risk mitigation measures apply to connected systems. AI regulations may be contagious to other systems if the boundaries are not defined.

Moving on to the definition of high-risk, the Act has a unique challenge because it refers to other legislation to define high-risk. Specifically, a system is high-risk if it is a product (or safety component of a product) that must be certified as EU compliant under a variety of regulations, including marine equipment.[454] The definition of marine equipment is "equipment placed or to be placed on board an EU

[448] *See* Part III.E.

[449] *See* Part III.F.

[450] See Part III.I.

[451] Recital 12, *supra* note 186.

[452] See Part III.H.5.

[453] Act 3(1), *supra* note 180.

[454] EU 2014/90, *supra* note 383.

ship and for which the approval of the flag State administration is required by the international instruments,"[455] with "international instruments" defined as "international conventions, together with the resolutions and circulars of the [International Maritime Organization] giving effect to those conventions in their up-to-date version, and the testing standards."[456]

In other words, to fully map what systems are high-risk, a developer would need to review international conventions of maritime law. This seems a heavy lift for a tech developer and seems more likely to lead to despair and resignation than compliance or innovation.

Finally, the definition of general-purpose models with systemic risk is likely to be underinclusive and overinclusive, as technology shifts which metrics matter for more powerful systems. As noted in Part III.J., definitions that rely on quantitative metrics must predict in advance what metrics create the most powerful systems, and if that were known, we would already have those systems.

Part V. Proposed Definition and Conclusion

While artificial intelligence will continue to evade perfect definition, some definitions are more effective than others.

A. A Proposed Definition

The following would be an improvement upon most definitions:

> Artificial intelligence refers to a computer program that uses machine learning techniques to accomplish complex goals that would normally require human-level intelligence.

By specifying that it is a computer program or statistical model, it avoids the traps of the word "system" that flows too easily to other programs. Specifying that it is based on a machine learning architecture excludes logic-based systems, but those have not proven as powerful as machine learning systems. Focusing on the architecture adds more clarity than the word "infers" and still captures the most powerful models.

Adding that the goals require human-level intelligence is still vague, but it eliminates simple programs and pairing it with these other elements will give judges more to work with standard human-centric definitions. It still suffers from failing to define these terms, making vague and potentially overly broad.

The definition does not include any limitations on autonomy or outputs, because as discussed in Part III, these create only de minimis boundaries.

Given its weaknesses, this definition would not be a good fit for all regulations, but may prove a starting point for adding or removing other elements.

B. A Better Direction

A better direction would be to balance the risks of an overinclusive definition against the risks of an underinclusive definition based on the attributes of the specific sector being regulated. Overbreadth may be less harmful in fields like autonomous weapons. In less risky areas, erring toward underinclusive definitions may be better to avoid chilling innovation and competition.

Vague definitions may have value as the field develops. While judges are not the foremost experts on AI development, their response times are likely faster than legislators, so allowing some judicial

---

[455] *Id.* at 3(1).

[456] *Id.* at 2(5), 2(3).

interpretation of vague terms will likely be more useful than trying to predict where the innovation is headed.

Attempts to predict the future are likely to fail. Legislators that predict quantitative metrics or architectures are likely to find the regulation hollow, as development moves in another direction.[457] Alternatively, legislators that use broad language effectively yield their regulatory authority to agencies or judges. This may allow a more rapid regulatory response, but it will increase uncertainty, which could chill innovation and competition.[458]

Regulators, researchers and legal scholars must tailor the definition of artificial intelligence to the sector being regulated, the technological reality and the power of the models.

[457] *See* Part III.J., *supra*.

[458] See text accompanying note 183, *supra*.