AN ASPEN DIGITAL PUBLICATION

DEFINING TECHNOLOGIES OF OUR TIME: ARTIFICIAL INTELLIGENCE

EDITORS

B Cavello, Nicholas P. Garcia, Eleanor Tursman

CONTRIBUTING AUTHORS

Stan Adams, Miranda Bogen, B Cavello, Dr. Rumman Chowdhury, Dr. Christine Custis, Mark Dalton, Anna Lenhart, Dr. Margaret Mitchell, Professor Rashida Richardson, Dr. Ranjit Singh, Dr. Nathan C. Walker, Professor Ben Winters

First published in the USA March 2026 by Aspen Digital, a program of the Aspen Institute.

ISBN 978-1-940914-02-2

Library of Congress Control Number: 2026904905

www.aspendigital.org

This work was made possible thanks to generous support from Fathom and Siegel Family Endowment.

Thank you to Colin Aamot, Leisel Bogan, Shanthi Bolla, Ryan Calo, Keith Chu, Rebecca Crootof, Leila Doty, Ro Encarnación, Kevin Frazier, Marissa Gerchick, Amir Ghavi, James Gimbi, Divya Goel, Garrett Graff, Margot Kaminski, Kevin Li, Julie Lin, Fergus Linley Mota, Morgan Livingston, J. Nathan Matias, Ashley Nagel, Anna Nickelson, Paul Ohm, Jake Parker, Jake Pasner, Andy Reischling, Connor Sandagata, Lacey Strahm, Meghan Stuessy, Elham Tabassi, Michael Weinberg, and David Zhang for their support of this project.

TABLE OF CONTENTS

INTRODUCTION

BY NICHOLAS P. GARCIA

PROJECT GOALS

Legal definitions are crucial to the drafting of technology legislation and regulation and are highly-referenced and recycled throughout the lawmaking process. They then cascade from laws to enforcement, leading to unintended consequences if definitions are not carefully scoped. Terms relating to emerging technologies like artificial intelligence (AI) are especially difficult to define due to the fast-changing technological frontier and the broad application of these tools.

For this reason, Congress has historically granted power to executive agencies to interpret definitions and update terms in accordance with new information and best practices. This all changed after the 2024 *Loper Bright* decision which overturned the "*Chevron* Doctrine" and significantly weakened agency authority. Now, policy leaders have a new responsibility thrust upon them to get these highly consequential definitions right in legislative text—since they can no longer confer this responsibility to executive agencies—and they urgently need support.

Aspen Digital brought together legal scholars, policy advocates, and leading thinkers in technology to craft a collection of influential definitions of AI, with legal and technological rationales for why a policymaker should choose one potential definition over another in context.

This resource is for policymakers, legislative staff, regulators, advocates, and legal practitioners who need to work with AI definitions as operative legal tools rather than technical descriptions. It assumes there is no single "correct" definition of AI. Instead, it treats definitions as context-dependent choices that can have outsized impact on scope, enforcement, institutional authority, and other downstream legal consequences.

This handbook offers an accessible, easy to use entry point for grappling with the question of how to define AI in a legal context. It is intended to be practical, comparative, and usable by readers who are actively engaging with legislative or regulatory text, whether they are drafting new provisions, interpreting existing ones, or trying to understand how a proposed definition will operate once it leaves the page.

This handbook does not attempt to catalogue every definition of artificial intelligence that exists in statute, regulation, or policy guidance. That would be neither feasible nor especially illuminating. Instead, it offers structured, independent expert analysis of the most influential definitions using a comparative approach. Each entry is intentionally scoped to highlight how a definition works, what it captures or excludes, and what tradeoffs it embeds.

The goal is not depth for its own sake, but clarity: to help readers see how small changes in language—sometimes only a few words—can substantially alter legal interpretation, regulatory reach, and institutional responsibility.

READING GUIDE

This handbook is not meant to be read straight through. It is meant to be a starting point for getting an overview of the field, a point of reference for understanding existing definitions and their impacts, and a resource to return to as drafts evolve or one encounters new and unfamiliar definitions. Think of it as a menu of options or, better yet, a branching tree that shows where definitions come from and how they are related.

Each definition entry can be read on its own, but the handbook is most effective when entries are read in relation to one another. Readers working on a specific bill or regulation may wish to start by reading each of the progenitor definitions to get a sense of the options, or by jumping to the entries for the lineage that most closely resembles their draft text and reading the variations and comparisons through.

Finally, while the handbook does not aim to be exhaustive, it aims to be clarifying. If it succeeds, readers should come away not with a single preferred definition of AI, but with a clearer understanding of the consequences of choosing one over another in a given context.

HOW THINGS ARE ORGANIZED

To make comparisons more explicit, definitions in this handbook are clustered by lineage. A lineage refers to a family of definitions that trace back to a common progenitor and then branch, mutate, or are repurposed across jurisdictions and use cases.

Where AI is defined in existing or proposed law today, it almost invariably adopts or uses a variant from one of these four lineages:

- Artificial Intelligence from the NDAA, rooted in early US defense and national security contexts and the one most often adopted without change directly through reference.
- Artificial Intelligence System from the OECD, which emphasizes capability, autonomy, and international harmonization.
- Automated Decision System from the Algorithmic Accountability Act, which defines scope through decisionmaking impact rather than technical capability.
- General-Purpose Artificial Intelligence from the EU AI Act and more recent policy, focused on models with broad, downstream applicability.

Readers should resist the temptation to treat these categories as mutually exclusive or as competing "schools." In practice, definitions routinely borrow from multiple lineages, combining related terms or adopting aspects of language—whether through thoughtful intent or compromise-seeking hybridization. This handbook aims to push more efforts towards thoughtful, intentional engagement with their definition, and a clear picture of the impact of choices and changes.

WHY LINEAGE MATTERS

We organized the handbook this way for practical reasons that emerged repeatedly in interviews and background conversations with legislative staff and policy drafters. Across jurisdictions and policy domains, we observed striking repetition in definition language, even where the laws and policies pursued very different goals. Definitions were often reused wholesale, lightly modified, or

combined, and there is little visibility into the assumptions carried through from prior contexts or companion terms modifying a definition's scope that may have been lost along the way.

By tracing definitions back to their progenitors, lineage analysis makes those inheritances visible. It allows readers to see how changing a single phrase (e.g., introducing "implicit objectives," or shifting from "decisions" to "outputs") can alter how a law is interpreted, enforced, or challenged. It also surfaces the kinds of questions staffers routinely confront while drafting: what is in scope, who is covered, where discretion lies, and how future technologies will be treated.

HOW TO READ EACH ENTRY

Each entry in this handbook follows a common structure to make comparison easier. Entries begin with brief factual metadata like jurisdiction, date, status, and enforcing entity to situate the definition within its legal and institutional context. At the top is the definition text itself, and where applicable, definitions are presented in direct comparison to a progenitor or closely related predecessor, making lineage relationships and points of divergence clear without needing to reference multiple documents at once.

Most entries also surface related or companion terms that meaningfully shape scope in practice. Terms such as "developer," "deployer," "consequential decision," or "high-risk system" are often as important as the AI definition itself. Readers should understand paired and related terms as external components of the definition; they are often key signals about scope, intent, and interpretation.

Following the definition text, each entry is organized around three analytic sections. The **Motivation** section explains why the definition was written and what problem, opportunity, or political moment it was responding to. The **Approach** section provides the core analysis: what the definition includes or excludes, how it operates in context, what tradeoffs it embeds, and how changes from predecessor language affect interpretation or enforcement. The **Reception** section situates the definition after introduction—where it has been adopted or reused, how it has been received by policymakers, industry, advocates, or courts, and where meaningful critiques have emerged. Each entry concludes with **Additional Resources** pointing readers to primary sources and deeper analyses for further research.

Together, this structure is meant to allow readers to scan quickly for orientation, read selectively for comparison, or dive deeply into a single definition—while maintaining enough consistency across entries to support side-by-side analysis.

A BRIEF TIMELINE OF AI DEFINITIONS

BY ELEANOR TURSMAN

What follows is a history of definitions of AI in legal and legislative contexts principally focused on the US, with a few important international texts that have influenced legislators in the United States.

Although "artificial intelligence" has been referenced in bills going as far back as the 1970s, the first legal definition for AI in Congress was introduced in 2018, in the John S. McCain National Defense Authorization Act for Fiscal Year 2019 (McCain NDAA). Shortly afterwards, the US joined forty-eight countries affiliated with the Organisation for Economic Co-operation and Development (OECD) and the European Union (EU) to adopt a nonbinding legal definition of AI in the Recommendation of the Council on Artificial Intelligence.

These definitions focus on outlining the capabilities that make an AI system officially "AI" in the legal context. NIST's AI Risk Management Framework and the National Artificial Intelligence Initiative Act of 2020 (NAIIA) build off of this OECD definition, in addition to state bills like California's Safe and Secure Innovation for Frontier Artificial Intelligence Models Act (SB-1047, vetoed in 2024). While these two definition branches are the oldest, a proliferation of proposed federal AI bills use either one or both of the McCain NDAA and NAIIA AI definitions, and many of the AI-related executive orders across both the Biden and Trump administrations use NAIIA's AI definition.

Around the same time as the McCain NDAA, a different approach emerged to defining AI through the lens of critical decisions and risk, as opposed to capabilities. In the EU, the AI Act uses the OECD AI definition while introducing tiered layers of risk. At the federal level in the US, the Algorithmic Accountability Act of 2019 (AAA) takes a technology-agnostic approach, capturing AI systems within a broader "automated decision system" definition. The AAA's approach has spread to the recent rulemaking process for the California Privacy Protection Act and other proposed federal legislation like the (since-stricken) AI regulation moratorium in Trump's flagship budget bill, the One Big Beautiful Bill Act.

In 2022, the release of ChatGPT and subsequent proliferation of generative AI models like large language models (LLMs) and image generators instigated a push to modify existing capability-focused definitions to ensure they covered these new systems. The OECD updated their definition in 2024, which was then leveraged in a number of state bills in the US like Colorado's AI Act (SB24-205). For tech-agnostic definitions like the one in AAA, no updates were needed.

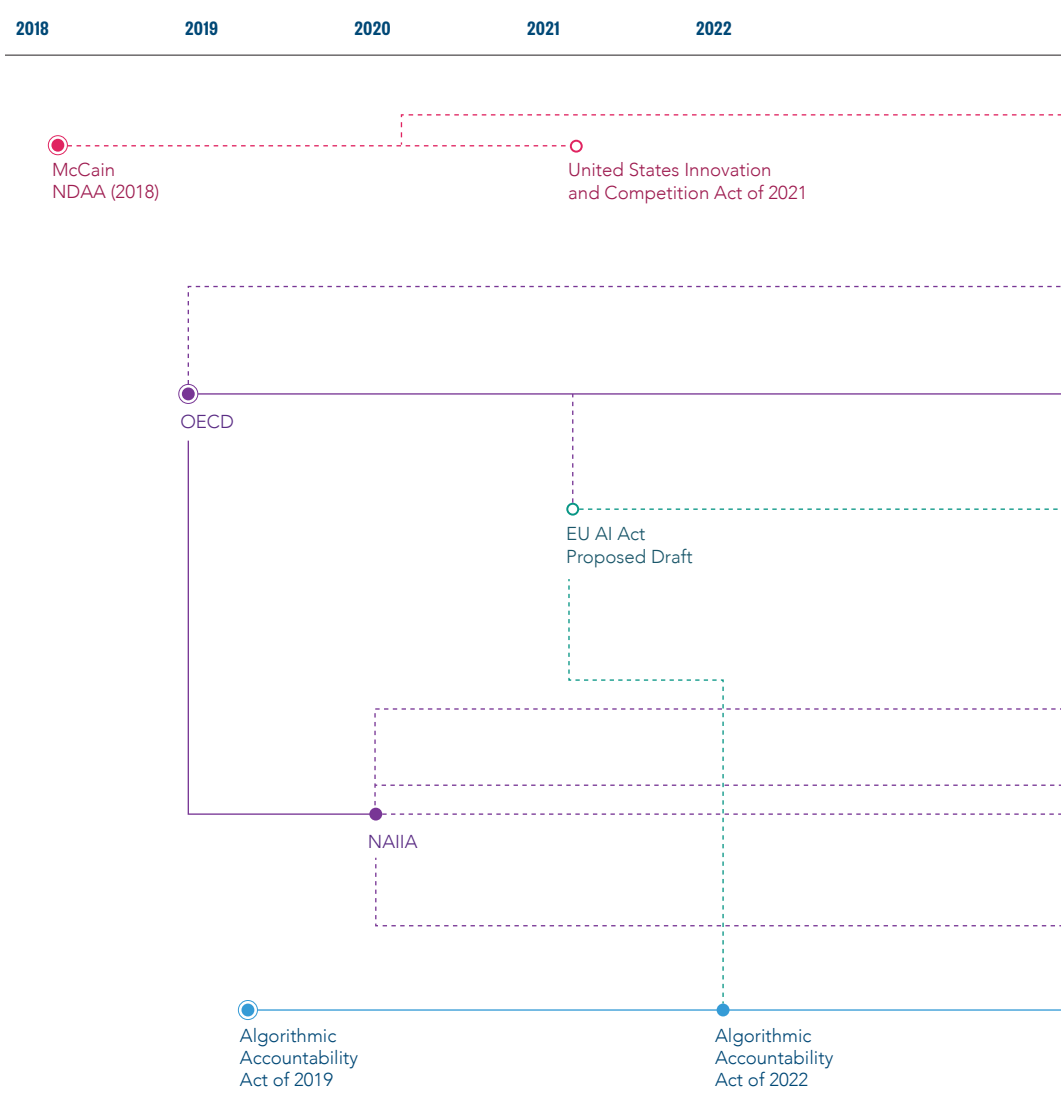
Most recently, there has been a push to further define models or combinations of models that have more general-purpose applications than other types of AI systems. Historically, AI systems were narrow in what kinds of information they could take in and output. As these systems are made more complex and "general," it becomes harder to anticipate all the possible outcomes of using these tools. Like the difference between a calculator and a more general-purpose computer, this expansion in possible outputs makes it more difficult for developers to forecast how different entities will use their systems. From a legislative perspective,

general-purpose models complicate regulation and liability because the potential actors influencing an AI system's performance grow more numerous and complex.

In the EU, this push manifests in the general-purpose AI definition of the AI Act and subsequent guidance on its interpretation, which build off of a legislative approach meant to capture specific major AI companies through mechanisms like training compute thresholds. This approach is similar to how the EU's Digital Services Act uses a threshold of 45 million users per month to identify covered Very Large Online Platforms and Very Large Online Search Engines. There was cross-pollination between the EU and US under the term "foundation models," language coined at Stanford in 2021 which has since been adopted in definitions in federal bills, executive orders in the Biden administration, and state bills like California's Transparency in Frontier Artificial Intelligence Act (SB-53). A related term, "frontier model," was included in New York's recently signed Responsible AI Safety and Education Act.

This overview and corresponding chart are our best interpretation at how the current AI definition landscape is intertwined.

Thank you to Dr. Nathan C. Walker, whose research was invaluable for writing this brief timeline.



Has entry in this handbook

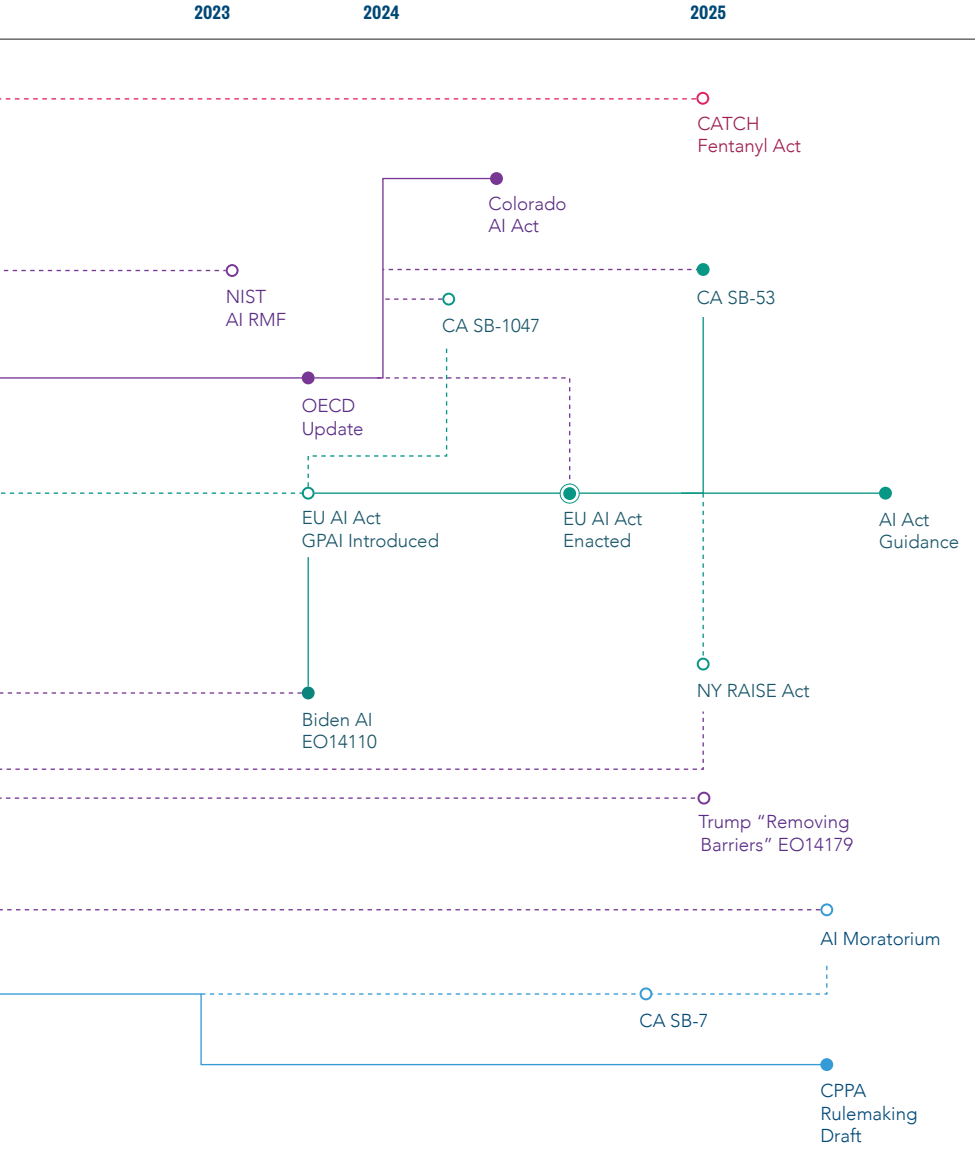
Not included in this handbook

Artificial Intelligence from the NDAA

AI System from the OECD

Automated Decision System from the Algorithmic Accountability Act

General-Purpose AI from the EU AI Act


2023
2024
2025
CATCH
Fentanyl Act
Colorado
AI Act
NIST
AI RMF
CA SB-1047
CA SB-53
OECD
Update
EU AI Act
GPAI Introduced
EU AI Act
Enacted
AI Act
Guidance
NY RAISE Act
Biden AI
EO14110
Trump "Removing
Barriers" EO14179
AI Moratorium
CA SB-7
CPPA
Rulemaking
Draft

LINEAGE: ARTIFICIAL INTELLIGENCE

JOHN S. MCCAIN NATIONAL DEFENSE AUTHORIZATION ACT

BY MARK DALTON

DEFINITION

ARTIFICIAL INTELLIGENCE DEFINED.—In this section, the term "artificial intelligence" includes the following:

(1) Any artificial system that performs tasks under varying and unpredictable circumstances without significant human oversight, or that can learn from experience and improve performance when exposed to data sets.

(2) An artificial system developed in computer software, physical hardware, or other context that solves tasks requiring human-like perception, cognition, planning, learning, communication, or physical action.

(3) An artificial system designed to think or act like a human, including cognitive architectures and neural networks.

(4) A set of techniques, including machine learning, that is designed to approximate a cognitive task.

(5) An artificial system designed to act rationally, including an intelligent software agent or embodied robot that achieves goals using perception, planning, reasoning, learning, communicating, decision making, and acting.

FULL TITLE
Section 238(g) of the John S. McCain National Defense Authorization Act for Fiscal Year 2019

OFFICIAL TEXT
https://www.congress.gov/bill/115th-congress/house-bill/5515/text

LINEAGE
Artificial Intelligence

PLACE OF ORIGIN
Federal, USA

ENFORCING ENTITY
Department of Defense

DATE OF INTRODUCTION
April 13, 2018 in H.R.5515

DATE ENACTED OR ADOPTED
August 13, 2018

CURRENT STATUS
Enacted (codified at 10 U.S.C. note prec. 4061)

MOTIVATION

Section 238(g)'s definition of AI serves many purposes for the Department of Defense (DoD). Foremost, it allows organization coordination across the most vast and complex bureaucracy in the world. DoD comprises military branches, defense agencies, and contractor networks staffed by a blend of military and civilian workforce. A shared understanding of AI is necessary to plan, budget, acquire, deploy, and maintain capabilities and this definition establishes that common lexicon.

The definition also supports the establishment of a senior oversight role responsible for AI strategy for the entire DoD. This reflects the formal recognition of AI as a transformative military technology that presents several implications for national security. Such recognition has substantial budgetary consequences. By codifying AI as strategically important in statute, Section 238 justifies billions of dollars in research, development, and acquisition funding for the technology.

The legislative context is equally important. By the time of the writing of the NDAA, there were significant concerns about the capabilities of emerging peer competitors on the world's stage. Democratization of technology, particularly China's state-backed AI initiatives, raised concerns in defense policy circles and anxieties about US competitiveness in technology domains related to military capabilities and future operations. Section 238 positioned the NDAA as an instrument for national-scale technology modernization, extending beyond the traditional hardware procurement process to address emerging digital capabilities that encounter friction with conventional defense acquisition. The definition simultaneously serves operational, budgetary, and strategic signaling purposes.

APPROACH

Section 238(g) adopts a notably expansive approach to defining artificial intelligence, employing five distinct formulations rather than a single, unified definition. This multi-faceted structure suggests the Department of Defense grappled seriously with AI's conceptual breadth and chose comprehensiveness over precision. The definition encompasses system-level characteristics, implementation methodologies, and theoretical frameworks drawn from decades of AI research, creating an umbrella broad enough to capture nearly any intelligent system while still calling out specific technological approaches.

The definition contains five elements: (1) systems that perform tasks under varying circumstances without significant human oversight or learn from experience, (2) systems that solve tasks requiring human-like perception, cognition, planning, learning, communication, or physical action, (3) systems designed to think or act like humans, including cognitive architectures and neural networks, (4) techniques like machine learning designed to approximate cognitive tasks, and (5) systems designed to act rationally, including intelligent software agents or embodied robots that achieve goals through perception, planning, reasoning, learning, communication, decisionmaking, and action.

These five elements can be organized into three conceptual categories. System-level definitions include Elements (1), (2), and (3), which focus on capabilities and behavior: autonomous learning systems that operate without human oversight, systems that replicate human cognitive functions across multiple domains, and systems explicitly designed to emulate human thinking. Implementation-based definitions are captured in Element (4), which shifts focus from complete systems to methodologies,

characterizing AI as "a set of techniques, including machine learning, that is designed to approximate a cognitive task." This formulation emphasizes methods over outcomes and acknowledges that AI often involves approximation rather than true replication of cognition. Finally, rational agent definitions appear in Element (5), which draws from classical AI theory—particularly the rational agent paradigm articulated by Russell and Norvig[1]—describing goal-oriented systems that execute a complete cognitive cycle.

This definitional breadth serves clear strategic advantages. The multi-pronged approach provides future-proofing as AI technology evolves, ensuring that novel techniques and architectures will likely fall under at least one element. For an organization focused on maintaining technological superiority over multi-decade timescales, this flexibility is valuable. The definition accommodates symbolic AI, statistical machine learning, neural networks, and robotics simultaneously, reflecting the reality that military AI applications span this entire spectrum. From a coordination perspective, the broad scope enables the designated senior official to exercise oversight across diverse AI initiatives without artificial categorical limitations. For budgetary purposes, the expansive definition supports funding requests across a wide range of programs, all legitimately classified as AI investments.

However, this breadth introduces significant problems. The definition risks severe over-inclusiveness, potentially encompassing conventional automation and traditional software systems that few practitioners would characterize as AI. Element (4)'s language "techniques...designed to approximate a cognitive task" is particularly vague and could arguably include recommendation algorithms, basic statistical models, or rule-based systems. This ambiguity creates substantial risk for "AI washing," the practice

of labeling conventional systems as AI-enabled to secure funding or prestige. The commercial sector has extensively demonstrated this phenomenon, with products ranging from streaming service recommendations to smart appliances marketed as "powered by AI" despite employing relatively simple algorithms.

The definition also lacks meaningful thresholds or boundaries. What constitutes "significant human oversight" in Element (1)? When does approximation become sufficient to qualify as AI under Element (4)? The absence of such specificity may facilitate organizational flexibility but undermines clarity for program managers, acquisition officials, and external stakeholders attempting to understand what qualifies as AI under DoD policy. The five elements exhibit substantial overlap, with most contemporary AI systems qualifying under multiple provisions simultaneously. Rather than providing definitional precision through multiple lenses, this redundancy may generate confusion about which element governs particular edge cases.

The definition's complexity presents additional challenges. It employs technical terms like "machine learning," "cognitive architectures," and "neural networks," that themselves lack precise boundaries and may not clarify matters for policymakers, military leadership, or even technical personnel outside AI specializations. This approach assumes familiarity with AI's academic lineage, drawing implicitly on traditions from early theorists like Marvin Minsky through the symbolic AI movement, connectionism, and evolutionary computation.[2] While this intellectual heritage is appropriate for a technical definition, it may limit accessibility for the broader DoD audience that must operationalize these concepts.

Notably absent from the definition are considerations that might distinguish AI from conventional software: requirements for large-scale data, computational intensity, emergent capabilities, or specific architectural features. The definition also predates the current generation of large language models and generative AI systems, raising questions about whether these technologies fit comfortably within the existing framework or represent capabilities that strain its boundaries.

From a legal and jurisdictional perspective, this definition applies specifically within the DoD context and carries statutory force for defense programs and budgets. However, it is not binding on other federal agencies, which have developed their own AI definitions for regulatory, research, or operational purposes. This lack of government-wide standardization creates potential coordination challenges and reflects the difficulty of establishing universal AI definitions across diverse mission spaces and regulatory contexts.

RECEPTION

Section 238(g)'s definition produced tangible organizational consequences within the Department of Defense, most notably the establishment of the Joint Artificial Intelligence Center (JAIC) in 2018, later reorganized as the Chief Digital and AI Office. The definition enabled systematic tracking of AI-related expenditures across the defense budget and justified substantial increases in AI investment. Its influence extended beyond DoD, serving as a template for subsequent federal AI definitions and strategies, though this cross-pollination has proven problematic when definitions optimized for budgetary and organizational purposes are applied to regulatory contexts requiring greater precision. The definition

has not been subject to significant judicial interpretation, and while it facilitated coordination within defense circles, critics have noted its over-breadth creates opportunities for AI washing and fails to provide meaningful boundaries for what qualifies as artificial intelligence in operational or acquisition contexts.

ADDITIONAL RESOURCES

Crevier, Daniel. *AI: The Tumultuous History of the Search for Artificial Intelligence*. New York: Basic Books, 1993.

US Department of Defense. "Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity." February 2019. https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF.

Schmidt, Eric, Robert Work, Safra Catz, et al. "Final Report: National Security Commission on Artificial Intelligence (AI)." March 5, 2021. https://apps.dtic.mil/sti/html/trecms/AD1124333/.

REFERENCES

1 Russell, Stuart J., and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 4th ed. Hoboken, NJ: Pearson, 2020.

2 Minsky, Marvin. *The Society of Mind*. New York: Simon & Schuster, 1986.

LINEAGE: AI SYSTEM

OECD RECOMMENDATION OF THE COUNCIL ON ARTIFICIAL INTELLIGENCE

BY DR. CHRISTINE CUSTIS

DEFINITION

AI system: An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.

FULL TITLE
Recommendation of the Council on Artificial Intelligence

OFFICIAL TEXT
https://one.oecd.org/document/C/MIN(2019)3/FINAL/en/pdf

LINEAGE
AI System

PLACE OF ORIGIN
International

ENFORCING ENTITY
N/A

DATE OF INTRODUCTION
March 15-6, 2019

DATE ENACTED OR ADOPTED
May 22, 2019

CURRENT STATUS
Updated November 8, 2023

MOTIVATION

The OECD (Organisation for Economic Co-operation and Development), as stated on its website, is a forum and knowledge hub of over 100 countries across the world for data, analysis and best practices in public policy. They play a critical role helping to "coordinate responses to the use of AI-based on international, multidisciplinary and multi-stakeholder cooperation to ensure that the development and use of AI benefits people and the planet." From the May 2019 OECD Meeting of the Council at Ministerial Level, those developing the recommendation for the definition of AI System regarded the important work being carried out on AI and, as stated in the guidance, "given the rapid development and implementation of AI, there is a need for a stable policy environment that promotes a human-centric approach to trustworthy AI, that fosters research, preserves economic incentives to innovate, and that applies to all stakeholders according to their role and the context." Standards development, including accepted definitions, is part of this stable policy environment deemed important by the OECD.

This marked a concerted effort by the organization to set a standard meaningful to the work being done in privacy and data flow, cryptography policy, public sector information use, and the "global implications that are transforming societies, economic sectors and the world of work" which is relevant to civil society, academia, corporations, governments, and communities.

APPROACH

The definition includes the phrase "Human-defined objectives." However, in the emerging age of AI Agents, objectives may not always be human defined. For instance, Anthropic has engaged in a practice with selective human involvement. It published a

Constitutional AI concept in which AI model development is governed by other AI models with the indirect input of humans in the form of human-generated rules or principles.[1] It can be conjectured that rules/principles are a level of abstraction (or more) higher than objectives. The updated definition (2023) seems to capture this with the explicit/implicit terminology.[2]

Additionally, the definition includes the term "Machine-based." In the simplest case, the interjection of humans-in-the-loop to complete the system might make the term "machine-based" too limiting. Additionally, we may need to think of the AI system as more than the operational component and extend the reach of the definition. Further down in this OECD guidance is a definition of AI system lifecycle which confirms the limitations placed on the AI system definition. The absence of data labelers and mineral extractors, which can be human, highlights the need for humanity as a member of the AI system composite. The ISO/IEC 22989 published in 2022 uses the term "engineered system" which might offer a broader encapsulation of both human and machine components.[3]

I personally like the absence of the terms "model," "algorithm," "decisionmaking," and "simulation" (as in simulation of human intelligence). For this and other reasons, the definition could be meaningful in the handling of discourse beyond what the typical AI actor may consider to be AI and could serve as a powerful tool for reminding us of the policy innovations available through the application of existing frameworks and laws. In an article for the National Academy of Engineering, Alondra Nelson writes about how "the perception that AI governance inherently lags behind technological development overlooks an immediate solution: the application of existing laws, rules, regulations, and standards."[4]

Broad and inclusive definitions encompassing the past and present of technology can help us to see the similarities and make the necessary connections.

An AI system can be considered a tool useful in automated decisionmaking. The Canadian Government defines "automated decision system" as "Any technology that either assists or replaces the judgment of human decision makers. These systems draw from fields like statistics, linguistics and computer science, and use techniques such as rules-based systems, regression, predictive analytics, machine learning, deep learning, and neural networks."[5] This Canadian Government definition compliments the 2019 OECD definition of an AI System in that decisionmaking impacts and legal rights should be considered if an AI system is used to aid in decisionmaking, especially for public services, where such decisions affect people directly.

The UNESCO 2022 Recommendation on the Ethics of Artificial Intelligence seems to expound on the definition by adding "means," "methods," and operational designation. A section of the recommendation follows:

> "AI systems are information-processing technologies that integrate models and algorithms that produce a capacity to learn and to perform cognitive tasks leading to outcomes such as prediction and decision-making in material and virtual environments. AI systems are designed to operate with varying degrees of autonomy by means of knowledge modelling and representation and by exploiting data and calculating correlations. AI systems may include several methods…"

RECEPTION

This 2019 version of the OECD definition of AI System was used in:

- the Regulation (EU) 2024/1689 (the EU "AI Act") — Article 3(1),[7]
- NIST AI Risk Management Framework (AI RMF 1.0),[8]
- National Telecommunications & Information Administration (NTIA) AI accountability/glossary pages and federal policy glossaries,[9]
- G20 AI Principles (2019),[10]
- Canadian Directive on Automated Decision-Making,[11]
- UK Government White Paper "A pro-innovation approach to AI regulation,"[12] and
- US Executive Order 14110 (2023).[13]

There have been some critiques of the definition. Joanna Bryson wrote in a 2022 Wired article that the definition was vague in its handling of autonomy and adaptiveness.[14] Additionally, the OECD itself has discussed the difficulty in sourcing liability and accountability data for implicit objections of any AI System.[15]

Although I cannot say verifiably, I would imagine that the directives housed in the EU AI Act would hold some sway in EU courts. Otherwise, courts and judges apply statutes or regulatory text and not necessarily the OECD language which could be considered a technology-neutral starting point for deliberations.

ADDITIONAL RESOURCES

Nessler, Bernhard, and Christiane Wendehorst. "Commission Guidelines on the Application of the Definition of an AI System and the Prohibited AI Practices Established in the AI Act: Response of the European Law Institute." European Law Institute, 2024. https://www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Publications/ELI_Response_on_the_definition_of_an_AI_System.pdf.

REFERENCES

1 Bai, Yuntao, Saurav Kadavath, Sandipan Kundu, et al. "Constitutional AI: Harmlessness from AI Feedback." arXiv preprint, 2022. https://doi.org/10.48550/arXiv.2212.08073.

2 Organisation for Economic Co-operation and Development (OECD). Recommendation of the Council on Artificial Intelligence. OECD Legal Instruments, OECD/LEGAL/0449. OECD, 2024. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

3 International Organization for Standardization. "ISO/IEC 22989:2022 Information Technology — Artificial Intelligence — Artificial Intelligence Concepts and Terminology." ISO, July 2022. https://www.iso.org/standard/74296.html.

4 See page 44:
Nelson, Alondra. "Disrupting the Disruption Narrative: Policy Innovation in AI Governance." The Bridge 55, no. 1 (April 2025). https://www.nae.edu/337896/Disrupting-the-Disruption-Narrative-Policy-Innovation-in-AI-Governance.

5 Government of Canada. Directive on Automated Decision-Making. June 2025. https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592.

6 See page 10:
UNESCO. "Recommendation on the Ethics of Artificial Intelligence." SHS/BIO/PI/2021/1, 2022. https://unesdoc.unesco.org/ark:/48223/pf0000381137.locale=en.

7 European Parliament and Council. Regulation (EU) 2024/1689: laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). Article 3(1), 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng#art_3.

8 National Institute of Standards and Technology (NIST). "Artificial Intelligence Risk Management Framework (AI RMF 1.0)." NIST AI 100-1, January 2023. https://doi.org/10.6028/NIST.AI.100-1.

9 National Telecommunications and Information Administration (NTIA). "Glossary." March 27, 2024. https://www.ntia.gov/issues/artificial-intelligence/ai-accountability-policy-report/glossary-of-terms#A.; and

National Institute of Standards and Technology (NIST). "Glossary." n.d. https://airc.nist.gov/glossary/.

10 G20. "G20 Ministerial Statement on Trade and Digital Economy." OECD, June 9, 2019. https://oecd.ai/en/wonk/documents/g20-ai-principles.

11 See #5.

12 UK Government. Department for Science, Innovation and Technology and Office for Artificial Intelligence. "A Pro-Innovation Approach to AI Regulation." March 2023. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach.

13 US President. "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, Executive order 14110 of October 30, 2023." Federal Register Vol. 88, no. 210 (November 1, 2023): 75191-75226. https://www.federalregister.gov/d/2023-24283.

14 Bryson, Joanna J. "Europe Is in Danger of Using the Wrong Definition of AI." *Wired*, March 2022. https://www.wired.com/story/artificial-intelligence-regulation-european-union/.

15 Organisation for Economic Co-operation and Development (OECD). "OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market." OECD, July 11, 2023. https://doi.org/10.1787/08785bba-en.

NATIONAL ARTIFICIAL INTELLIGENCE INITIATIVE ACT OF 2020 (NAIIA)

BY DR. NATHAN C. WALKER

DEFINITION

(3) ARTIFICIAL INTELLIGENCE.—The term "artificial intelligence" means a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. Artificial intelligence systems use machine and human-based inputs to—

(A) perceive real and virtual environments;

(B) abstract such perceptions into models through analysis in an automated manner; and

(C) use model inference to formulate options for information or action.

FULL TITLE

National Artificial Intelligence Initiative Act of 2020

OFFICIAL TEXT

http://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title15-section9401&num=0&edition=prelim

LINEAGE

AI System

PLACE OF ORIGIN

Federal, USA

ENFORCING ENTITY

Federal Agencies

DATE OF INTRODUCTION

March 12, 2020, in H.R.6216; (incorporated into H.R.6395)

DATE ENACTED OR ADOPTED

January 1, 2021

CURRENT STATUS

Enacted (codified at 15 U.S.C. § 9401(3))

LINEAGE DEFINITION

AI system: An AI system is amachine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.

MOTIVATION

In President Trump's first term, there were four competing understandings of artificial intelligence: (1) one legal definition in the John S. McCain National Defense Authorization Act for Fiscal Year 2019 (McCain NDAA);[1] (2) another in the United States' support of a nonbinding international agreement with the Organisation for Economic Co-operation and Development (OECD) in 2019;[2] the (3) notable absence of a substantive legal definition in President Trump's first AI executive order in 2019;[3] and (4) Congress' alignment of the OECD's definition when passing the 2020 National Artificial Intelligence Initiative Act (NAIIA, pronounced "nye-ah"), embedded in the William M. (Mac) Thornberry National Defense Authorization Act. NAIIA aimed to provide a unified federal definition of artificial intelligence to standardize how US agencies identify and regulate AI systems.

This definition was written in response to the first executive order on AI, issued by President Donald Trump in February 2019, Maintaining American Leadership in Artificial Intelligence (No. 13,859), which called for a coordinated national AI strategy.[4] The NAIIA acknowledged the earlier definition of AI in the McCain NDAA, which introduced the first statutory definition of AI in federal law under the section governing Department of Defense activities.

By contrast, the NAIIA expanded the legal scope to create a whole-of-government approach, coordinating both defense and civilian agencies (including the Department of Defense, the National Science Foundation, the National Institute of Standards and Technology, and the Department of Education). It also closely aligned the federal definition with the international standard endorsed by 48 OECD member and non-member states and the European Union.

The NAIIA's revised definition sought to reconcile fragmented definitions across federal agencies. It removed anthropomorphic analogies and articulated three core functions: perceiving, abstracting, and inferring. Its emphasis on perception and modeling reflected technological advances such as deep learning, autonomous vehicles, AlphaGo, and conversational AI, like Siri and Alexa.

APPROACH

NAIIA is helpful in that it uses a technique-agnostic definition that does not specify any particular approach, such as machine learning, deep learning, or neural networks. Doing so illustrates the attempt to create a future-proof law that is flexible enough to evolve with new methods. Other advantages of the definition include avoiding technical and legal jargon and focusing on processes rather than speculative ideas about what constitutes human-like intelligence. Compared to its 2019 predecessor in the McCain NDAA, the NAIIA correctly removes anthropomorphic phrases such as "acts like a human" and "thinks or learns." It acknowledges both "real and virtual environments"; however, it would be more accurate to adopt the OECD's 2024 language of "physical and virtual," because, although intangible, digital spaces are real domains where social, economic, and political activity occurs.

There are other disadvantages to NAIIA's 2020 definition of artificial intelligence. Most notable is the absence of the term "generate." The NAIIA definition predates the boom in generative AI following the launch of ChatGPT in November 2022. President Biden's 2023 executive order, Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, was the only of the nine executive orders about AI that provided a legal definition of

generative AI.[5] President Trump revoked it upon returning to the White House in 2025.

It can be argued, however, that even without explicit reference to "generate," the NAIIA's 2020 definition applies to generative AI. It addresses the pre-generative concept of AI by recognizing that predictive modeling underlies generation, by describing how humans provide inputs and participate in data training, and by emphasizing the perceive-abstract-infer clauses that describe model learning behavior. The absence of the term "generate," though, means the definition could be misinterpreted as falling outside of human-defined objectives and the creative, expressive, and stochastic outputs that stem from problems of randomness, unpredictability, and confabulations.

Other limitations include the absence of terms such as "varying levels of autonomy" and "adaptiveness," which are included in the OECD's 2024 updated definition.[6] Without these three terms, the NAIIA definition does not explicitly recognize that AI systems can generate new content, operate along a continuum of human oversight safeguards, and learn and change after deployment, thereby posing risks such as model drift, unintended bias amplification, and evolving decisionmaking behaviors.

The NAIIA definition removed the phrase "without significant human oversight" as a first step in broadening its scope to include autonomous systems. NAIIA continued to emphasize "human-defined objectives," which, although ideal when emphasizing human responsibility, fail to account for the fact that AI agents are increasingly designed to define their own intentions and operate semi-independently. An updated law should account for these developments.

The 2020 definition also implies that the federal government will use AI as a general-purpose technology, with core methods applicable across sectors. However, this legal framing overlooks the various ways agencies rely on narrow, sector-specific AI systems tailored to distinct regulatory contexts. For example, computer vision or anomaly behavior-detection models can be considered general-purpose AI for use in surveillance, facial recognition, and fraud detection. In contrast, medical imaging and diagnostic algorithms, pollution monitoring, and energy grid optimization are highly sector-specific systems that require specialized datasets, subject-matter expertise, and agency oversight. By not making this distinction, the NAIIA definition blurs the policy line: general-purpose AI systems benefit from a horizontal government strategy through the proposed cross-agency frameworks (e.g., NIST or OMB), while sector-specific applications may require additional vertical, domain-grounded oversight (e.g., subject-matter experts in the FDA, EPA, or DOJ).

Overall, the NAIIA's 2020 definition, which was built upon OECD's 2019 definition, marks an important moment in US policy, moving from a defence-centric approach to a whole-of-government strategy. Its strengths lie in its accessible, inclusive language, while its weaknesses include failing to distinguish between general-purpose and narrow AI and omitting AI systems' generative, autonomous, and adaptive capabilities.

RECOMMENDATION

In conclusion, it is recommended that the federal definition of artificial intelligence be updated as follows to address the technical, legal, and ethical gaps in current law.

> ARTIFICIAL INTELLIGENCE.—The term "artificial intelligence" means a machine-based system that can, for ~~a given set of~~ **direct or indirect** human-defined objectives, **generate, predict, recommend, or decide outputs that influence physical** ~~make predictions, recommendations or decisions influencing real~~ or virtual environments. Artificial intelligence systems use machine and human-based inputs to—(A) perceive **physical** ~~real~~ and virtual environments; (B) abstract such perceptions into models through analysis in an automated **or adaptive** manner; and (C) use model inference to formulate options for information or action, **or to generate new content. Artificial intelligence systems operate with varying levels of autonomy and adaptivity after deployment, and may include general-purpose AI technologies applicable across agencies, as well as narrow AI applications designed for specific-sector functions**.

RECEPTION

NAIIA's 2020 definition, aligned with 2019 OECD principles and language, received wide bipartisan support upon its passage on January 1, 2021. President Biden and President Trump, upon his return to office, added the NAIIA definition to six of their executive orders on artificial intelligence between 2023 and 2025. To illustrate its influence, as of November 2025, several states have passed laws on artificial intelligence that adopt an NAIIA/OECD-style definition. From its inception, this NAIIA/OECD definition has received widespread industry support.

The primary concern, however, is that despite widespread bipartisan support at the federal and state levels and endorsements from leading AI companies, Congress must replace the outdated legal definition in NAIIA 2020 to align with the OECD 2024 definition.

Furthermore, Congress and federal agencies will need to carefully align their efforts, as even the Advancing American AI Act of 2022[7] and the Office of Management and Budget's 2025 Memorandum (M-25-21)[8] incorrectly reference the outdated definition of AI in the McCain NDAA from 2018.

ADDITIONAL RESOURCES

Organisation for Economic Co-operation and Development (OECD). "Explanatory Memorandum on the Updated OECD Definition of an AI System." *OECD Artificial Intelligence Papers*: no. 8, March 5, 2024. https://doi.org/10.1787/623da898-en.

Shen Hong, Tham. "BSA Comments on Bill on Fostering Artificial Intelligence Industry and Securing Trust [18726]." Business Software Alliance, February 13, 2023. https://www.bsa.org/files/policy-filings/en02132023kraitrust.pdf. (Software Alliance (BSA) explicitly "recommends adopting the OECD's definition of AI." The BSA includes major AI companies such as Adobe, Amazon Web Services, Cohere, Databricks, Elastic, IBM, Microsoft, OpenAI, Oracle, Salesforce, and SAP.)

REFERENCES

1 John S. McCain National Defense Authorization Act for Fiscal Year 2019, § 238(g) (definition of "artificial intelligence"), Pub. L. No. 115–232, 132 Stat. 1636, 1695 (2018). https://www.congress.gov/bill/115th-congress/house-bill/5515/text.

2 Organisation for Economic Co-operation and Development (OECD). Recommendation of the Council on Artificial Intelligence. OECD, May 2019. https://one.oecd.org/document/C/MIN%282019%293/FINAL/en/pdf.

3 US President. "Maintaining American Leadership in Artificial Intelligence, Executive order 13859 of February 11, 2019." Federal Register Vol. 84, no. 31 (February 14, 2019): 3967-3972. https://www.federalregister.gov/d/2019-02544.

4 See #3.

5 US President. "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, Executive order 14110 of October 30, 2023." Federal Register Vol. 88, no. 210 (November 1, 2023): 75191-75226. https://www.federalregister.gov/d/2023-24283.

6 Organisation for Economic Co-operation and Development (OECD). Recommendation of the Council on Artificial Intelligence. OECD Legal Instruments, OECD/LEGAL/0449. OECD, 2024. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

7 Act incorporated into:
James M. Inhofe National Defense Authorization Act for Fiscal Year 2023, § 7223(3) (definition of "artificial intelligence"), Pub. L. No. 117-263, div. G, tit. LXXII, subtit. B, 136 Stat. 3668 (2022). https://www.congress.gov/bill/117th-congress/house-bill/7776/text.

8 Office of Management and Budget. "Accelerating Federal Use of AI through Innovation, Governance, and Public Trust." Memorandum M-25-21 § 5 Definitions (April 3, 2025). https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-21-Accelerating-Federal-Use-of-AI-through-Innovation-Governance-and-Public-Trust.pdf.

OECD UPDATED RECOMMENDATION

BY DR. MARGARET MITCHELL

DEFINITION

AI system: An AI system is a machine-based system 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. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

FULL TITLE
Revised Recommendation of the Council on Artificial Intelligence

OFFICIAL TEXT
https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

LINEAGE
AI System

PLACE OF ORIGIN
International

ENFORCING ENTITY
N/A

DATE OF INTRODUCTION
November 8, 2023

DATE ENACTED OR ADOPTED
May 3, 2024

CURRENT STATUS
In force

LINEAGE DEFINITION
AI system: An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.

MOTIVATION

As written, the 2023 OECD definition of "AI system" appears crafted to bridge a gap between machine learning-specific terminology and more common language, building from an earlier definition provided by the OECD. Changes from the previous OECD definition are striking and problematic, inserting terms used in machine learning but using them in a confusing or inappropriate way, ultimately making the definition less suitable for defining AI.

The first difference between the current definition and the previous OECD definition is the replacement of "[can,] for a given set of human-defined objectives" with "for explicit or implicit objectives," a change that at its surface sounds more similar to what is happening under-the-hood in machine learning, but makes the definition less clear for regulators and machine learning developers alike. For example, what is an implicit objective?

Before going further in untangling the OECD definition's phrasing, let's unpack some of the basics about what an "AI system" is. One approach for building an AI system involves machine learning, where what's called a model uses an objective function and inference to learn how to connect inputs to outputs. In the definition provided by OECD, the phrase "explicit or implicit objectives" may be an attempt to reference the concept of an objective function in machine learning—which is a mathematical expression, distinct from a human high-level goal, and generally explicitly defined by a developer. In contrast, an "implicit" objective is one that a system can define for itself based on explicitly defined goals by a human.[1] As a very simple example, a human-defined high-level goal of "make me a sandwich" might be met by a system using implicit objectives of getting bread out of the refrigerator, cutting it, etc. But this explicit-implicit distinction is a bit "in the

weeds" for a general definition, and misses that there is still a role of explicit definition from a person even when implicit objectives are defined. Or, perhaps "implicit objective" is referring to something else. It's not at all clear.

Adding to the confusion is that this change from the previous OECD definition removes a key component of an AI system: the human developers who define it. Where instead the definition may benefit from being more human-centered—e.g., by highlighting that common systems are trained using human-created data—it instead removes the role of the human altogether, elevating the concept of an AI system to something that exists outside of human creation, imbued with a kind of mystique that can only further fuel misunderstandings and inappropriate hype.

The rest of the text is similar, using terminology that is hard to follow from a machine learning perspective, and that I can't imagine is easier to parse from a non-machine learning perspective.

APPROACH

From the original OECD text of "can…make predictions…," this definition substitutes "infers…how to generate…predictions," adding terminology reminiscent of the concepts of inference and generation in machine learning, but applied inappropriately and introducing multiple issues. Machine learning processes underlying many AI systems leverage an input-output relationship similar to the one provided, where a model can be said to "infer" (using inference) "from the inputs it receives, how to generate outputs." However, In the context of defining an "AI system," there are three main issues with this framing. The first is that outside of technical machine learning text, the concept of infers generally

refers to a mentalistic concept that denotes reasoning, judgement, having thought, holding opinions, etc. (Sources: Merriam Webster Dictionary, Oxford English Dictionary, Cambridge Dictionary). While sharing some similarities to a technical term, the more general usage of "infer" afforded by the definition's context brings with it the risk (if not the direct requirement) of anthropomorphism: If something can infer, then surely it can think. This problem is compounded by the usage of colloquial mentalistic concepts now used in marketing AI systems, like "chain of thought reasoning," which corresponds to algorithmic processes and is not a representation of the human mind.

The second issue is that many technologies referred to as "AI" are not based on machine learning, and can't be said to "infer" in the colloquial or technical sense, for example, systems that are rule based (e.g., ELIZA, SYSTRAN) or statistical without using machine learning (e.g., DeepBlue). The third issue is that machine learning models are often just one part of an "AI system," which is what this definition is attempting to define. In contrast to an isolated machine learning model, an AI system—such as ChatGPT—can incorporate a variety of other programming techniques, rules, and hand-designed algorithms.

As such, the usage of "infers" here serves to further muddy and confuse the issue, blending machine learning-esque terminology (e.g., inference) with mentalistic anthropomorphised concepts (e.g., thinking) to produce a phrasing that is stunningly unclear and fails to capture multiple types of AI systems. (To make the definition more clear, perhaps "infer" could be defined without using anthropomorphising language.)

Examining the definition as a whole, there are further issues in what it includes and excludes. On the one hand, the definition is inappropriately inclusive of non-AI systems, such as a traditional weather forecasting system that uses inputs such as sensor data (temperature, pressure, humidity, wind speed, satellite imagery) and processing from mathematical models grounded in physics equations (fluid dynamics, thermodynamics) to predict how weather systems evolve, generating output in the form of forecasts (predictions about rain, temperature, storms, etc.), which influence physical environments (e.g., people wearing a coat, farmers deciding when to plant). Such systems are generally not considered to be AI, but do fall under the given definition.

On the other hand, the definition problematically overlooks the importance of one of the most critical aspects of AI systems for policy: the role of training data. As discussed above, the current definition appears to align with machine learning AI systems, which leverage training data often created by humans at the expense of their rights. Yet how human rights are affected is something that is particularly important for politicians to be aware of when examining AI systems, and this key issue is erased in the definition.

RECEPTION

I am not aware of any industry practitioners using this definition and have not come across others using it in industry circles. I do not think it is a well-accepted one in industry. Notably, Bezerra et al. discuss how the definition is overly broad and a critical risk for the AI market.[2] A related interesting discussion on issues with the definition occurred in the OSI forum.[3]

REFERENCES

1 Hanna, Sean. "Defining Implicit Objective Functions for Design Problems." *Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation*, July 2007, 2013-20. https://doi.org/10.1145/1276958.1277355.

2 Bezerra, Leonardo C. T., Alexander E. I. Brownlee, Luana Ferraz Alvarenga, Renan Cipriano Moioli, and Thais Vasconcelos Batista. "How VADER Is Your AI? Towards a Definition of Artificial Intelligence Systems Appropriate for Regulation." arXiv preprint, 2025. https://doi.org/10.48550/arXiv.2402.05048.

3 Maffulli, Stefano. "Is the Definition of 'AI system' by the OECD Too Broad?" Open Source Initiative, January 2024. https://discuss.opensource.org/t/is-the-definition-of-ai-system-by-the-oecd-too-broad/156.

COLORADO AI ACT

BY DR. RUMMAN CHOWDHURY

DEFINITION

"ARTIFICIAL INTELLIGENCE SYSTEM" MEANS ANY MACHINE-BASED SYSTEM THAT, FOR ANY EXPLICIT OR IMPLICIT OBJECTIVE, INFERS FROM THE INPUTS THE SYSTEM RECEIVES HOW TO GENERATE OUTPUTS, INCLUDING CONTENT, DECISIONS, PREDICTIONS, OR RECOMMENDATIONS, THAT CAN INFLUENCE PHYSICAL OR VIRTUAL ENVIRONMENTS.

FULL TITLE
An Act Concerning Consumer Protections for Interactions with Artificial Intelligence

OFFICIAL TEXT
https://leg.colorado.gov/bills/sb24-205

LINEAGE
AI System

PLACE OF ORIGIN
Colorado, USA

ENFORCING ENTITY
State Attorney General

DATE OF INTRODUCTION
April 10, 2024 in SB24-205

DATE ENACTED OR ADOPTED
May 17, 2024

CURRENT STATUS
Enacted (but implementation has been postponed)

LINEAGE DEFINITION
AI system: An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.

MOTIVATION

The Colorado AI Act's definition of "artificial intelligence system" was written in response to the growing need for regulatory clarity as AI technologies began to play a decisive role in consequential decisions for individuals—such as hiring, housing, healthcare, credit, and more. The definition targets the problem of "algorithmic discrimination," situations where automated systems can yield unfair outcomes or treat individuals or groups differently based on protected attributes. By crafting a direct, functional definition, lawmakers sought to cover a broad range of present and future AI tools likely to impact consumer rights.

This definition appears to be a descendant of earlier frameworks such as the EU AI Act draft[1] and the NIST AI Risk Management Framework,[2] but it differs in some respects. The Colorado language explicitly references both explicit and implicit objectives and makes outputs—including non-decision content—central, reflecting how large language models and general-purpose AI systems (like ChatGPT, Gemini, and Claude) can influence user experience and outcomes. Events surrounding the rapid adoption of generative AI in 2023–2024, along with high-profile incidents of bias and opaque decisionmaking, clearly influenced both the need and the shape of this wording. The state balanced the need for innovation with robust consumer protections, aiming to future-proof the law against evolving AI capabilities while supporting transparency and legal accountability.

APPROACH

The Colorado AI Act takes a comprehensive, rights-focused approach to regulating artificial intelligence, targeting both the technologies themselves and their impact on consumers. It lays

out a broad definition of "artificial intelligence system" while identifying explicit inclusions and exclusions that shape its scope, utility, and unique jurisdictional features.

INCLUSIONS AND EXCLUSIONS

The definition encompasses a wide array of current and future AI systems, from automated decisionmaking engines to generative content tools and algorithmic recommenders.

Included technologies are not limited by architecture or technique: anything with inferential capabilities that shapes outputs affecting users is covered, whether rule-based, statistical, or driven by machine learning. "High-risk artificial intelligence systems"—those used for consequential decisions in areas such as employment, education, healthcare, housing, finance, government services, insurance, or legal services—are the main regulatory target.

Yet, the Act carves out notable exclusions. Systems used solely for procedural tasks or for pattern detection with robust human oversight escape high-risk classification. Standard tools such as calculators, anti-virus software, spell-checkers, video games, web hosting, and certain natural language bots functioning strictly under non-discriminatory usage policies are excluded unless they make "consequential decisions." Technologies governed by federal approval or certain sectoral regulations, and work done for federal agencies, are exempt if pre-existing standards are sufficiently strong. Small deployers (under 50 employees not feeding proprietary data into their systems) face lighter requirements, recognizing resource constraints and risk profiles.

STAKEHOLDERS: WHO IS REGULATED?

The Act clearly delineates "developers" (those who create or substantially modify AI systems) and "deployers" (entities using high-risk systems in Colorado). For developers, regardless of whether located in Colorado, obligations apply if their systems are used in the state. Deployers must conduct risk management, impact assessments, and consumer-facing disclosures before deployment of high-risk AI. Ultimately, the law aims to protect "consumers"—defined as individuals residing in Colorado—from algorithmic discrimination in critical life areas.

JURISDICTIONAL AND LEGAL CONTEXT

The language is closely tied to Colorado consumer protection statutes and civil rights laws, referencing "classification protected under the laws of this state or federal law" for discrimination claims. Procedural obligations and enforcement powers are vested in the Colorado Attorney General, distinctly contrasting with private right of action models seen in some states or federal proposals—"this part 17 does not provide the basis for … a private right of action."

Terms such as "consequential decision," "algorithmic discrimination," "developer," and "deployer" are defined in detail to limit ambiguity. Exemptions for entities operating under stringent federal or sectoral standards tie the Act to national regulatory infrastructure while maintaining state sovereignty.

SCOPE THROUGH COMPANION TERMS

Related terms that define the scope of the AI definition include:

- "Algorithmic discrimination"—directly linked to consumer protection and anti-discrimination law, setting out what constitutes harm.
- "High-risk artificial intelligence system"—anchoring the law's main application to use-cases where decisions have "material legal or similarly significant effect."
- "Intentional and substantial modification"—clarifying when developer obligations are triggered for updates or retraining, with exceptions for mere model drift or technical maintenance.
- "Substantial factor"—further specifying systems whose outputs can alter consequential decisions.

DIFFERENCES FROM PREDECESSOR DEFINITIONS

Compared to predecessor definitions (like the EU AI Act or NIST RMF), Colorado's scope is simultaneously broader in capturing a wider set of inferential processes ("explicit or implicit objective") and more targeted by focusing regulatory attention on high-risk consequential use-cases. The Act excludes low-impact infrastructure and procedural automation tools despite their "machine-based" nature. Colorado's version also ties compliance to recognized risk management frameworks, permitting a safe harbor for organizations that follow standards like the NIST RMF or ISO/IEC 42001.[3]

The language adjustments—such as "any explicit or implicit objective" and "can influence physical or virtual environments"—ensure future-proofing and accommodate novel AI modalities like generative and conversational systems. At the same time, the detailed list of exclusions is itself an innovation, striving to balance public protection with business practicality.

In summary, the Colorado AI Act combines broad definitional language with specific operational scope, targeting technologies and stakeholders most likely to impact consumers in material ways. Its layered approach to definitions, inclusion/exclusion criteria, and jurisdictional specificity position it at the forefront of US state-level AI regulation while highlighting both practical utility and ongoing complexity for compliance and enforcement.

RECEPTION

Colorado's Act is viewed as a pioneering model in US state-level regulation, prompting legislative interest in California, Illinois, and New York City, and influencing ongoing debates at the federal level and in industry forums. Other US state laws—such as recently-proposed bills in California and Illinois—echo Colorado's focus on algorithmic discrimination, consumer protection, and transparency, though few match its detailed controls or coverage of both developers and deployers. At the federal level, the White House's AI Bill of Rights provides guideline principles that align with Colorado's emphasis on protections in consequential decisions, but falls short in regulatory force.[4]

Notable endorsements or support have emerged from privacy advocacy organizations and some legal scholars who praise the definition's breadth and explicitness—particularly its coverage of both "explicit or implicit objectives" and wide range of outputs, including content, predictions, and recommendations. Legal experts, such as the American Bar Association and practitioners in AI law, have highlighted Colorado's approach as likely to shape future legislation nationwide, especially in regulating "high-risk" AI and aligning with anti-discrimination policies.

Industry practitioners and some technology industry groups, however, have raised significant critiques. The US Chamber of Commerce, the Chamber of Progress, and the Consumer Technology Association opposed the Act's breadth, citing concerns that ambiguous and expansive mandates would create compliance challenges, stifle innovation, and be burdensome for smaller developers. The law's complex definitions and procedural requirements have also led to "buyer's remorse" among state leaders, who delayed implementation after encountering stakeholder confusion and industry pushback.

Colorado's courts have not yet interpreted the definition, as the Act's effective date is delayed, and enforcement will be in the hands of the state Attorney General rather than private litigants. Legislative bodies nationwide continue to debate how much of Colorado's terminology and structure to copy, with some adopting similar concepts ("high-risk AI," "consequential decisions") but not the full definition.

ADDITIONAL RESOURCES

Siegal, Alex, and Ivan Garcia. "A Deep Dive into Colorado's Artificial Intelligence Act." National Association of Attorneys General, October 26, 2024. https://www.naag.org/attorney-general-journal/a-deep-dive-into-colorados-artificial-intelligence-act/.

Rice, Tatiana, Keir Lamont, and Jordan Francis. "The Colorado Artificial Intelligence Act: FPF U.S. Legislation Policy Brief." Future of Privacy Forum, July 2024. https://leg.colorado.gov/sites/default/files/images/fpf_legislation_policy_brief_the_colorado_ai_act_final.pdf.

Zweifel-Keegan, Cobun, and Andrew Folks. "The Colorado AI Act: What You Need to Know." IAPP, May 21, 2024. https://iapp.org/news/a/the-colorado-ai-act-what-you-need-to-know.

Betts, Jennifer G., Danielle Ochs, and Michael H. Bell. "Colorado's Artificial Intelligence Act: What Employers Need to Know." Ogletree Deakins, May 20, 2024. https://ogletree.com/insights-resources/blog-posts/colorados-artificial-intelligence-act-what-employers-need-to-know/.

REFERENCES

1 European Commission. Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. 2021. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206.

2 National Institute of Standards and Technology (NIST). "Artificial Intelligence Risk Management Framework (AI RMF 1.0)." NIST AI 100-1, January 2023. https://doi.org/10.6028/NIST.AI.100-1.

3 International Organization for Standardization. "ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system." ISO, December 2023. https://www.iso.org/standard/42001.

4 Office of Science and Technology Policy (OSTP). *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People*. White House. October 2022. https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

LINEAGE: AUTOMATED DECISION SYSTEM

ALGORITHMIC ACCOUNTABILITY ACT OF 2019

BY PROFESSOR BEN WINTERS

DEFINITION

AUTOMATED DECISION SYSTEM.—The term "automated decision system" means a computational process, including one derived from machine learning, statistics, or other data processing or artificial intelligence techniques, that makes a decision or facilitates human decision making, that impacts consumers.

FULL TITLE
Algorithmic Accountability Act of 2019

OFFICIAL TEXT
https://www.congress.gov/bill/116th-congress/senate-bill/1108/text

LINEAGE
Automated Decision System

PLACE OF ORIGIN
Federal, USA

ENFORCING ENTITY
Federal Trade Commission (FTC)

DATE OF INTRODUCTION
April 10, 2019 in S.1108

DATE ENACTED OR ADOPTED
N/A

CURRENT STATUS
Dead

MOTIVATION

This was a very early attempt at regulating AI at the federal level in the US, so it had less precedent to work off of. The John S. McCain NDAA[1] as well as the Department of Defense had adopted definitions of *AI itself*, but this focused on the systems and the decisions, not just AI as a concept or bigger category. This definition was written in light of stories including: Amazon admitting the algorithm they made led to discriminatory outcomes, Joy Buolamwini proving that popular facial recognition programs can't accurately recognize black faces, and a case brought by the DOJ against Meta for their discriminatory housing ad targeting system. It also comes in the wake of some of the first very popular books illustrating the common issues of algorithmic discrimination: Weapons of Math Destruction by Cathy O'Neil (2016),[2] Technically Wrong by Sarah Wachter-Boettcher (2017),[3] and Automating Inequality by Virginia Eubanks (2018).[4] It was done before Generative AI was on the market in popular widely available consumer-facing tools, and therefore reflects a better focus.

All of these stories helped fill in a previously more esoteric concept of algorithmic harm. They gave shape and urgency to the issue, and since it was a relatively new issue, there was less industry defense and more earnest attempts at helping for a period of time.

Rashida Richardson, then at the New York Civil Liberties Union, published key information about ADS tools as part of the New York City Automated Systems Task Force process, and proposed a very similar definition.[5]

APPROACH

This proposal centered privacy and data use—building on the General Data Protection Regulation in the EU in 2016[6] and Sen. Wyden's Mind Your Own Business Act in 2019.[7] It also was inspired in part by an approach taken in New York City—to increase algorithmic transparency around the use of AI by the city in 2017.

There are three significant dynamics at play at the heart of this bill that are still central to ADS bills in 2025: what kind of algorithmically driven decisions are covered, who bears responsibility when different entities create an automated system and deploy them (referred to as "developer vs deployer"), and what kind of entities are covered under the bill.

The qualifiers of what kind of decisions usually use some sort of phrase like critical, rights impacting, important, or consequential action—in this bill, it's "high-risk." The way high-risk is defined in this bill leads to a fairly broad interpretation. They include "makes a decision or facilitates human decision making" in the definitions of the bill which is a significant narrowing of the scope, and makes it very easy for entities to just avoid being swept into coverage by deliberately characterizing their processes as less automated than they are.

In the definitions of this bill, the authors have also taken the approach of separately defining data protection impact assessments and ADS impact assessments.

Like many federal tech-related bills, this would authorize the FTC to do the lionshare of enforcement work as well as some regulatory work that would inform what the bill looks like in practice. State AGs would also be given authority to enforce, but not establish rules.

RECEPTION

The bill was endorsed by advocacy groups including Data for Black Lives, the Center on Privacy and Technology at Georgetown, and the National Hispanic Media Coalition. Law professors Andrew Selbst and Margot Kaminski praised the bill in a New York Times op-ed for its ambition but noted it lacked clarity and enforceability in key areas.[8] Industry groups, unsurprisingly, opposed the bill, arguing it was too broad and burdensome, as reflected in critiques like those from the Center for Data Innovation.[9]

The bill's definition of automated decision systems has since influenced similar legislation in Washington (2019),[10] Connecticut's SB2 (2024),[11] and California's AB-1018 (2025).[12] That said, as the cultural and political understanding of "AI" has expanded since 2019, fewer bills today focus narrowly on this type of decisionmaking system—though the 2022 and 2025 versions of the same bill, with the same primary sponsors, build on and expand this definition, as discussed later in this resource.

Since the release of this bill in 2019, there has simultaneously been a reduced focus on current harms like this bill has and an expanded focus on speculative harms. With that being said, the amount of bills at both the federal and state levels that would affect the use of these types of systems have increased significantly overall.

ADDITIONAL RESOURCES

Robertson, Adi. "A New Bill Would Force Companies to Check Their Algorithms for Bias." *Verge*, April 10, 2019. https://www.theverge.com/2019/4/10/18304960congress-algorithmic-accountability-act-wyden-clarke-booker-bill-introduced-house-senate.

Hao, Karen. "Congress Wants to Protect You from Biased Algorithms, Deepfakes, and Other Bad AI." *MIT Technology Review*, April 15, 2019. https://www.technologyreview.com/2019/04/15/1136/congress-wants-to-protect-you-from-biased-algorithms-deepfakes-and-other-bad-ai/.

MacCarthy, Mark. "An Examination of the Algorithmic Accountability Act of 2019." SSRN, October 24, 2019. http://dx.doi.org/10.2139/ssrn.3615731.

REFERENCES

1 John S. McCain National Defense Authorization Act for Fiscal Year 2019, § 238(g) (definition of "artificial intelligence"), Pub. L. No. 115–232, 132 Stat. 1636, 1695 (2018). https://www.congress.gov/bill/115th-congress/house-bill/5515/text.

2 O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group, September 2016. https://dl.acm.org/doi/10.5555/3002861.

3 Wachter-Boettcher, Sarah. *Technically Wrong: Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech*. W. W. Norton & Company, 2017.

4 Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press, January 23, 2018.

5 See page 20:
Richardson, Rashida, ed. "Confronting Black Boxes: A Shadow Report of the New York City Automated Decision System Task Force." AI Now Institute, December 4, 2019. https://ainowinstitute.org/wp-content/uploads/2023/04/ads-shadowreport-2019.pdf.

6 European Parliament and Council. Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). April 27, 2016. https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng.

7 Senator Ron Wyden. "The Mind Your Own Business Act of 2019." October 17, 2019. https://www.wyden.senate.gov/download/mind-your-own-business-act-of-2019-one-pager.

8 Kaminski, Margot E., and Andrew D. Selbst. "The Legislation That Targets the Racist Impacts of Tech." *New York Times*, May 7, 2019. https://www.nytimes.com/2019/05/07/opinion/tech-racism-algorithms.html.

9 New, Joshua. "How to Fix the Algorithmic Accountability Act." Center for Data Innovation, September 23, 2019. https://datainnovation.org/2019/09/how-to-fix-the-algorithmic-accountability-act/.

10 Washington Legislature. HB 1655 An Act Relating to Establishing Guidelines for Government Procurement and Use of Automated Decision Systems in Order to Protect Consumers, Improve Transparency, and Create More Market Predictability. (2019). https://app.leg.wa.gov/billsummary?BillNumber=1655&Year=2019.

11 Connecticut Legislature. SB-2 An Act Concerning Artificial Intelligence. (2024). https://www.cga.ct.gov/asp/cgabillstatus/cgabillstatus.asp?selBillType=Bill&which_year=2024&bill_num=2.

12 California Legislature. AB-1018 Automated Decision Systems. (2025). https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260AB1018.

ALGORITHMIC ACCOUNTABILITY ACT OF 2022

BY B CAVELLO

DEFINITION

AUTOMATED DECISION SYSTEM.—The term "automated decision system" means any system, software, or process (including one derived from machine learning, statistics, or other data processing or artificial intelligence techniques and excluding passive computing infrastructure) that uses computation, the result of which serves as a basis for a decision or judgment.

FULL TITLE
Algorithmic Accountability Act of 2022

OFFICIAL TEXT
https://www.congress.gov/bill/117th-congress/senate-bill/3572/text

LINEAGE
Automated Decision System

DATE OF INTRODUCTION
February 3, 2022 in S.3572

PLACE OF ORIGIN
Federal, USA

DATE ENACTED OR ADOPTED
N/A

ENFORCING ENTITY
Federal Trade Commission (FTC)

CURRENT STATUS
Reintroduced June 25, 2025 as S.2164

LINEAGE DEFINITION
AUTOMATED DECISION SYSTEM.—The term "automated decision system" means a computational process, including one derived from machine learning, statistics, or other data processing or artificial intelligence techniques, that makes a decision or facilitates human decision making, that impacts consumers.

MOTIVATION

The Algorithmic Accountability Act of 2022 was a substantial revision and reintroduction of the Algorithmic Accountability Act of 2019.[1] The 2019 bill was influential, hailed by many as a significant first step toward federal AI consumer protection. It was also criticized for being overly vague and introduced as a messaging bill. The 2022 update aimed to address these critiques with greater clarity and specificity.

One of the most significant changes in approach was a shift in emphasis from a focus solely on the automated decision systems (ADSs), themselves, to the augmented critical decision processes (ACDPs) in which the ADSs are used. This new term, defined as "a process, procedure, or other activity that employs an automated decision system to make a critical decision," was introduced in light of the reality that harms from automation do not only arise from flaws in the automated systems themselves. Rather, automation has the capacity to scale and speed up existing harmful processes, often while obfuscating the actual decisionmakers, making accountability more challenging.

The inclusion of ACDP was also intended to address some early uses of generative AI. While the Act was introduced before the release of ChatGPT which elevated AI into the public consciousness, it was drafted when the precursor technology (the generative pre-trained transformer, or GPT) was already widely acknowledged in the AI research community.

APPROACH

Legal definitions of AI are often critiqued as being overly broad (often of the dismissive "this would cover a spreadsheet" variety). ADS in the Algorithmic Accountability Act of 2022 is no exception.

One reason tech legislation uses intentionally broad language is an attempt to "future-proof" the text given the continually changing nature of many technologies. The Algorithmic Accountability Act of 2022 follows this rationale, and—with inspiration from the foundational research of Professor Rashida Richardson[2]—recognizes that spreadsheets are, in fact, incredibly consequential technologies, if used to make consequential decisions.

If anything, the 2022 ADS definition is even more expansive than the 2019 version. The updated definition drops the phrase "that impacts consumers" altogether. Instead, the bill uses other defined terms and directions to the target agency (the FTC) to capture the consumer protection context. Similarly, the 2022 definition expands from merely "a computational process" to "any system, software, or process … that uses computation." Arguably, these terms mean the same thing, but the latter is less susceptible to creative reinterpretations of the term "process" in the context of computing.

Instead of more narrowly defining ADS, the 2022 Act invokes other companion definitions like covered entity and critical decision to scope the application of the bill. However, one place where the 2022 ADS definition is specifically limited is in the exclusion of passive computing infrastructure, a term for "any intermediary technology that does not influence or determine the outcome of a decision." Examples provided in the text for passive computing infrastructure include things like web hosting, networking, and data storage.

This narrowing language qualifies the final phrase of the definition: "the result of which serves as a basis for a decision or judgment." This phrasing was updated from 2019's "that makes a decision or

facilitates human decisionmaking," because the earlier language anthropomorphized AI systems. The 2022 text clarifies more precisely how technology facilitates human decisionmaking (by producing results), but as an extra precaution—or to appease critics who claim that ADS covers everything under the sun—passive computing infrastructure explicitly excludes anything where the results aren't connected to the decision being made.

Another way the 2022 Act narrows the set of actors and technologies to which the bill applies is through the definition of the covered entities it is intended to target. The entities in question are defined by their size, determined either by gross receipts (think: revenue, investment, etc.) or data processing activity. This is a common pattern in tech policy, where legislators aim to target a specific (often newsworthy) set of notable actors, but don't want to sweep up other actors (especially small businesses) in the process.

Because this bill is motivated by the speed and scale of impact that automation enables, defining covered entities in terms of size makes sense. The text has slightly different qualifying criteria for entities depending on whether they are employing ADS in processes for critical decisions themselves or whether they are offering tools that are intended or likely to be used by others who will. (Other policy writing in this space sometimes uses a "deployer" vs "developer" framing, but this bill does not.)

Finally, the bill navigates having such a broad ADS definition by clarifying critical decisions. Critical decisions in the bill are decisions that affect consumers' "access to or the cost, terms, or availability of" education and vocational training, employment, essential utilities, family planning, financial services, healthcare,

housing or lodging, legal services, or similarly impactful areas of consumers' lives. The definition of these decision categories echoes the EU AI Act[3] which was drafted during the same period, but tailored specifically to the FTC context.

Decisions are a throughline of the Algorithmic Accountability Act of 2022. Indeed, the term automated decision system has it right in the name. Even so, because "decision," itself, is undefined, it is possible that some interpretations of the term could unintentionally limit the applicability of the bill. (For example: whether "decision" is treated as a psychological process or as an outcome or ruling with consequence.) The critical decision definition gestures at an outcome rather than process understanding of the term ("a decision or judgment that has any legal, material, or similarly significant effect") but if enough were on the line, crafty lawyers might find creative arguments to challenge this framing.

As such, this language may need to be revisited in light of the rise in so-called "AI agents," or automated systems that execute actions (even complex, multi-step actions) on behalf of a user. For what it's worth, an original press release for the bill references examples that substantiate intent to cover these types of automated systems, following the outcomes framing of the term.[4]

RECEPTION

The 2022 Act was endorsed by Access Now; Accountable Tech; Center for Democracy and Technology (CDT); Color of Change; Consumer Reports; Credo AI; EPIC; Fight for the Future; IEEE; JustFix; Montreal AI Ethics Institute; OpenMined; Open Technology Institute (OTI); Parity AI; US PIRG; and others.

Critics of the bill raised concerns about the lack of pre-emption, ambiguity regarding decisions on rulemaking left to the FTC, and the usual concerns about compliance burdens.

While the Algorithmic Accountability Act of 2022 did not pass, it was reintroduced in 2025 under the same name. Additionally, the revised ADS definition was referenced in a number of other policies including the 2022 Blueprint for an AI Bill of Rights (White House OSTP),[5] the 2023 No Robot Bosses Act (Sen. Casey),[6] the 2023 Transparent Automated Governance (TAG) Act (Sen. Peters),[7] among others.

ADDITIONAL RESOURCES

Mökander, Jakob, Prathm Juneja, David S. Watson, and Luciano Floridi. "The US Algorithmic Accountability Act of 2022 vs. The EU Artificial Intelligence Act: What Can They Learn from Each Other?" *Minds & Machines* 32, (2022): 751–758. https://doi.org/10.1007/s11023-022-09612-y.

Chilson, Neil, and Adam Thierer. "The Coming Onslaught of 'Algorithmic Fairness' Regulations." Regulatory Transparency Project of the Federalist Society, November 2, 2022. https://rtp.fedsoc.org/paper/the-coming-onslaught-of-algorithmic-fairness-regulations/.

Cavello, B. "Why I'm (Still) Hyped About the Algorithmic Accountability Act." Personal Website, September 17, 2023. https://posts.bcavello.com/why-im-still-hyped-about-the-algorithmic-accountability-act/.

REFERENCES

1 Congress.gov. "S.1108 - 116th Congress (2019-2020): Algorithmic Accountability Act of 2019." April 10, 2019. https://www.congress.gov/bill/116th-congress/senate-bill/1108/text.

2 Richardson, Rashida. "Defining and Demystifying Automated Decision Systems." *Maryland Law Review* 81: no. 3 (2022): 785-840. https://digitalcommons.law.umaryland.edu/mlr/vol81/iss3/2/.

3 European Commission. Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. 2021. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206.

4 Senator Ron Wyden. "Algorithmic Accountability Act of 2022." Internet Archive, 2022. https://web.archive.org/web/20220203205641/https://www.wyden.senate.gov/imo/media/doc/2022-02-03%20Algorithmic%20Accountability%20Act%20of%202022%20One-pager.pdf.

5 Office of Science and Technology Policy (OSTP). *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People*. White House. October 2022. https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

6 Congress.gov. "S.2419 - 118th Congress (2023-2024): No Robot Bosses Act." July 20, 2023. https://www.congress.gov/bill/118th-congress/senate-bill/2419/text.

7 Congress.gov. "S.1865 - 118th Congress (2023-2024): TAG Act." August 22, 2023. https://www.congress.gov/bill/118th-congress/senate-bill/1865/text.

CALIFORNIA PRIVACY PROTECTION AGENCY (CPPA) RULEMAKING DRAFT

BY PROFESSOR RASHIDA RICHARDSON

DEFINITION

"Automated decisionmaking technology" or "ADMT" means any technology that processes personal information and uses computation to replace human decisionmaking or substantially replace human decisionmaking.

(1) For purposes of this definition, to "substantially replace human decisionmaking" means a business uses the technology's output to make a decision without human involvement. Human involvement requires the human reviewer to:

(A) Know how to interpret and use the technology's output to make the decision;

(B) Review and analyze the output of the technology, and any other information that is relevant to make or change the decision; and

C) Have the authority to make or change the decision based on their analysis in subsection (B).

(2) ADMT includes profiling.

(3) ADMT does not include web hosting, domain registration, networking, caching, website-loading, data storage, firewalls, anti-virus, anti-malware, spam- and robocall-filtering, spellchecking, calculators, databases, and spreadsheets, provided that they do not replace human decisionmaking

FULL TITLE

California Privacy Protection Agency Modified Text of Proposed Regulations (CCPA Updates, Cyber, Risk, ADMT, and Insurance Regulations)

OFFICIAL TEXT

http://cppa.ca.gov/regulations/pdf/ccpa_updates_cyber_risk_admt_mod_txt_pro_reg.pdf

LINEAGE

Automated Decision System

DATE OF INTRODUCTION

May 9, 2025

PLACE OF ORIGIN

California, USA

DATE ENACTED OR ADOPTED

September 22, 2025

ENFORCING ENTITY

California Privacy Protection Agency

CURRENT STATUS

Rulemaking Finalized

LINEAGE DEFINITION

AUTOMATED DECISION SYSTEM.—The term "automated decision system" means a computational process, including one derived from machine learning, statistics, or other data processing or artificial intelligence techniques, that makes a decision or facilitates human decision making, that impacts consumers.

MOTIVATION

Automated decisionmaking technologies ("ADMT") is a categorical term that was adopted by policymakers globally to refer to a wide range of technologies and applications with varying degrees of automation that are deployed to support or supplant human decisionmaking. As ADMT are increasingly integrated into enterprise systems or deployed across a range of contexts, they became an early focal point for policymakers concerned about their pervasiveness, opacity, and possible consequences.

In California, ADMT was initially referenced in the Civil Code 1798.185, subdivision (a)(15) to direct the California Privacy Protection Agency (CPPA) to issue regulations governing access and opt-out rights regarding businesses' use of ADMT; however, the term was not defined.[1] In November 2024, the CPPA issued a public notice of rulemaking to update CPPA regulations including rules on consumers' rights to access and opt-out of businesses use of ADMT. Informed and influenced by a recent federal AI policy framework and academic scholarship that excogitated an approach to defining technical terms for technology policy proposals, the CPPA presented a multipronged ADMT definition.

This definition describes and focuses on the role that ADMT can play in human decisionmaking. In crafting this definition, the CPPA wanted to focus on what ADMT are actually doing and their impact to clarify that such decisions are within scope of the broader privacy law that the rulemaking sought to reform and to ensure the underlying regulation provides comprehensive consumer protections provided by the law. Unlike the source text, this definition does not focus on or attempt to enumerate the techniques and processes that constitute autonomous or

semi-autonomous technologies. Instead, it includes subprovisions that expound on key concepts and explicates what is included or exempted from the definition.

The motivation and context of the CPPA's ADMT definition is also noteworthy. Unlike the source text and other ADMT definitions, which are defined to anchor legislation seeking to regulate these technologies or their use context, this CPPA definition exists to fill a gap and remedy a loophole in an existing privacy regulation. Therefore, the definition was drafted with a certain level of freedom and rationalization that is otherwise constrained in the legislative drafting process due to its more formal rules and norms.

APPROACH

While predecessor ADMT definitions focused on the processing of personal data (GDPR[2]) or the variegated techniques and processes that constitute technologies of interest (Algorithmic Accountability Act of 2019[3]), this CPPA definition represents a major departure in its approach to tech policy legislative drafting. The CPPA presented a conceptual definition that attempts to describe the categorical term so that different stakeholders understand its general meaning for their context, rather than adopting or possibly misappropriating technical jargon in attempts to be prescriptive. In addition to providing a simplified statutory definition, the CPPA included subprovisions that further clarify the scope of the law and presumably thwart rulemaking comments and lobbying efforts that claim the definition is overinclusive.

The core definition focuses on the processing of personal information and the reliance on technology in decisionmaking. The processing of personal information grounds the definition in

its underlying privacy regulation, while the emphasis on the role of technology in human decisionmaking, along with a subprovision that clarifies the level of human involvement that renders these technologies as "automated" highlights their key function and impact. The CPPA drafters intended to leave little ambiguity about their intentions and the meaning of ADMT to ensure common explanations of noncompliance cannot be leveraged, and subsequent (judicial) review does not distort its meaning and potentially render the regulation unenforceable.

The second subprovision of the definition clarifies that ADMT includes profiling, which is a term commonly used in US state privacy laws to refer to the automated processing of personal information to provide analysis or predictions about an individual. Although ADMT and profiling appear to be cognate terms, the explicit clarification of this subprovision is notable because it foreshadows a trend in state legislatures to indirectly regulate AI and ADMT by revising the definition of profiling or key legislative provisions related to profiling.

The third subprovision of the definition provides an illustrative and non-exhaustive list of technologies generally excluded from the definition of ADMT. It also clarifies that while the named technologies are generally excluded because they tend to facilitate operational needs, they cannot be used in decisionmaking to circumvent legal requirements for ADMT. Although this statutory definition is intentionally comprehensive to include common technologies used as ADMT, it was not drafted to subject every technical system or software to ADMT legal obligations. Including exemptions in the ADMT definition can also aid legal compliance by businesses, or oversight and enforcement of the regulation.

Notably, the rulemaking that provides this ADMT definition also includes a definition of artificial intelligence. This represents a stark difference from the predecessor definition, which includes artificial intelligence techniques as illustrative examples of the computational processes that constitute an ADMT. By defining these terms separately, the CPPA tacitly emphasizes that AI and ADMT are not interchangeable terms and that these technologies must be evaluated in the contextual settings in which they function. This distinction also has the practical effect of demonstrating the expansive and inclusive nature of ADMT.

RECEPTION

Since this CPPA definition was part of a formal rulemaking process it was drafted with awareness that both the definition and related Articles in the proposed regulation would be subject to intense debate and revisions. Indeed, the proposed text in the notice of proposed rulemaking included several examples that sought to clarify the technologies that comprise ADMT; the types of technology assisted decisionmaking that is covered by the definition; and how exemptions were context dependent. Nonetheless, there was significant pushback from industry groups, who claimed the definition was too broad and vague. Their primary complaints were that the scope of technologies covered were too broad, the framing of ADMT use in human decisionmaking could capture too many business decisions, and that the exceptions were not fixed. Yet, some business stakeholders were supportive. For instance, a technology lawyer and angel investor supported a broad conceptual definition to ensure broad regulatory protection and greater appreciation of the potential or cumulative implications of ADMT.

ADDITIONAL RESOURCES

California Privacy Protection Agency (CPPA). "Initial Statement of Reasons (CCPA Updates, Cyber, Risk, ADMT, and Insurance Regulations)." Title 11, Division 6, Chapter 1, November 22, 2024. https://cppa.ca.gov/regulations/pdf/ccpa_updates_cyber_risk_admt_ins_isor.pdf.

Office of Science and Technology Policy (OSTP). *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People*. White House. October 2022. https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

Richardson, Rashida. "Defining and Demystifying Automated Decision Systems." *Maryland Law Review* 81: no. 3 (2022): 785-840. https://digitalcommons.law.umaryland.edu/mlr/vol81/iss3/2/.

California Privacy Protection Agency (CPPA). "CCPA Updates, Cybersecurity Audits, Risk Assessments, Automated Decisionmaking Technology (ADMT), and Insurance Regulations." 2025. https://cppa.ca.gov/regulations/ccpa_updates.html.

Richardson, Rashida, ed. "Confronting Black Boxes: A Shadow Report of the New York City Automated Decision System Task Force." AI Now Institute, December 4, 2019. https://ainowinstitute.org/wp-content/uploads/2023/04/ads-shadowreport-2019.pdf.

REFERENCES

1 California Legislature. Civil Code § 1798.185, subdivision (a)(15). (2018). https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?sectionNum=1798.185.&lawCode=CIV.

2 European Parliament and Council. Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). April 27, 2016. https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng.

3 Congress.gov. "S.1108 - 116th Congress (2019-2020): Algorithmic Accountability Act of 2019." April 10, 2019. https://www.congress.gov/bill/116th-congress/senate-bill/1108/text

LINEAGE: GENERAL-PURPOSE AI

EU AI ACT

BY ANNA LENHART

DEFINITION

'general-purpose AI model' means an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market

FULL TITLE
Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)

OFFICIAL TEXT
https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng#art_3

LINEAGE
General-Purpose AI

DATE OF INTRODUCTION
April 21, 2021

PLACE OF ORIGIN
European Union

DATE ENACTED OR ADOPTED
August 1, 2024

ENFORCING ENTITY
European Commission EU AI Office

CURRENT STATUS
Enacted (with staged enforcement)

MOTIVATION

The European Commission released the proposed AI Act in April 2021, introducing a risk-based framework organized around the intended purpose of an AI system—defined in line with OECD stakeholder processes.[1] When OpenAI's ChatGPT and similar generative AI tools captured the zeitgeist in late 2022, EU policymakers suddenly faced the challenge of fitting technologies capable of multiple, shifting purposes into a framework designed for narrowly defined use cases. During the subsequent "trialogues," the Council, Commission, and Parliament deliberated ways to include general-purpose AI models up until the last minute. The Parliament proposed a series of public-facing disclosures for all foundation models (defined similarly to GPAI above) regardless of risk category. The Council wanted to allow for self-regulation through a code of conduct fearing that regulation would hinder innovations especially among emerging European firms. The Commission presented a tired approach, placing more burdens on models with high impact capabilities. The final version included a tiered approach.[2]

APPROACH

REQUIREMENTS ARE TIERED BASED ON GPAI MODELS, GPAI MODELS WITH SYSTEMIC RISKS (GPAISR) AND GPAI SYSTEMS.

All GPAI models must follow a base level of obligations including maintaining up-to-date technical documentation of the model, making available relevant documentation to *downstream providers* (those who integrate the GPAI model into AI systems), implementing a copyright/text and data mining (TDM) policy to ensure the model respects EU copyright law, and providing a public summary of the content used for training (e.g., data sources, nature of the data).

If a GPAI model is (or is presumed to) contribute to systemic risks (GPAISR), then additional obligations kick in including that the provider must notify the Commission (or relevant authority) that the model presents systemic risk, conduct model evaluation (including adversarial testing, misuse potential, emergent behavior, etc.), perform risk assessment and risk mitigation for systemic-level risks (including but not limited to public health, safety, societal impacts, fundamental rights), report serious incidents (e.g., accidents, misuse, malfunctions) and corrective measures to the AI Office/national supervisory authorities, ensure cybersecurity protections and physical safeguards appropriate to the scale and impact of the model, and document and report estimated energy consumption of training and deployment.

When a GPAI model is used within an AI system (i.e., coupled with other models, infrastructures and user interfaces), the system may fall under the broader "high-risk AI systems" regime in the Act (depending on its purpose/use case) and may trigger additional obligations. The tiered obligations focus first on the model provider side. But downstream deployers/integrators of GPAI systems still need to check whether their use of the model triggers "high-risk" classifications for the system and comply with those requirements.

Open source GPAI models do not have the same reporting requirements as proprietary GPAI models, showing the EU's recognition of built-in transparency associated with open source and decentralized development. However, if the GPAI falls into the systems risk category, proprietary and open source models are treated the same.

WEAKNESS 1: FLOATING POINT OPERATIONS PER SECOND (FLOPS) ARE A BAD PROXY FOR SYSTEMIC RISKS.

Scholars argue that there are systemic risks that occur in all GPAI models regardless of computational power, most notably: misinformation, bias, work displacements, privacy violations, and content that inspires physical harm.[3] Computer scientists note that GPAI capabilities can be improved via data quality, optimization breakthroughs, architectural adjustments, and other factors that do not rely on compute.[4] The threshold outlined in the systemic risk definition may simply incentivize model developers to optimize models to fall below this threshold while avoiding attempts to make models safer.[5]

WEAKNESS 2: THE SYSTEMIC RISK DEFINITION STATES IT MUST BE "SPECIFIC TO THE HIGH-IMPACT CAPABILITIES OF GENERAL-PURPOSE AI MODELS" WHICH IS VAGUE AND CAUSES CONFUSION.

There are many systemic risks to the Union from GPAI systems that are also present in AI systems based on algorithmic decision systems and machine learning systems that have been in production for decades (e.g., discrimination, dark patterns, errors, etc.). When these risks are present in a GPAI, are they "specific to the high-impact capabilities"? This question was actively debated during deliberations related to the code of practice.[6]

WEAKNESS 3: GPAI MODELS THAT ARE NOT CONSIDERED GPAISR CAN BE USED IN HIGH-RISK AI SYSTEMS, CREATING UNFAIR/UNWORKABLE REQUIREMENTS FOR PROVIDERS.

According to the Al Act, any person will be considered the provider of a high-risk system "if they modify the intended purpose of an Al system, including a general-purpose Al system, which has not been classified as high-risk and has already been placed on

the market or put into service in such a way that the AI system concerned becomes a high-risk AI system." Boine and Rolnick (2025) offer the example of a school teacher that decides to use an LLM to evaluate their students. It is unclear if the teacher/school is then considered a provider of a high-risk system and required to complete related risk management and documentation requirements.[7] Advocates for strong regulation of GPAI emphasize that developers have the deepest insights into how these models are trained and the potential for harm.[8]

One way to address this challenge would be to require the provider of the GPAI model to mitigate risks. The challenge with that approach is that a provider of a GPAI may not be able to predict every high-risk context in which their model may be deployed. For the contexts they may predict, they would likely have to train and tune a different model for each high risk context, meaning models would no longer be "general purpose."[9]

RECEPTION

Implementation is underway at the time of this writing.

The ambiguity in "systemic risks" has frustrated stakeholders from a range of ideological standpoints, with some arguing that major risks, particularly from GPAI systems will be left unchecked and reliant on self-reporting and compliance check boxes, and others arguing that a range of deployers from smaller/less resourced organizations will be caught up in high risk scenarios and struggle to comply. The New Code of Practice added some additional clarity but there are still issues (described in the next section on EU Guidelines on General-purpose AI Models).

The European Commission is facing pressure to simplify not only the AI Act but a range of EU tech regulation as part of a Digital Omnibus.[10]

The EU AI Act has the potential to inform standards for how the world regulates AI systems including GPAI; it's too early to tell what the exact impact of these definitions will be.[11]

REFERENCES

1 Rodríguez de Las Heras Ballell, Teresa. "Mapping Generative AI Rules and Liability Scenarios in the AI Act, and in the Proposed EU Liability Rules for AI Liability." *Cambridge Forum on AI: Law and Governance* 1 (2025): e5. https://doi.org/10.1017/cfl.2024.8.

2 Bommasani, Rishi, Alice Hau, Kevin Klyman, and Percy Liang. "Foundation Models Under the EU AI Act." Stanford Center for Research on Foundation Models, August 2024. https://crfm.stanford.edu/2024/08/01/eu-ai-act.html.; and

Kohn, Benedikt, and Lennart van Neerven. "Will Disagreement Over Foundation Models Put the EU AI Act at Risk?" *Tech Policy Press*, November 29, 2023. https://www.techpolicy.press/will-disagreement-over-foundation-models-put-the-eu-ai-act-at-risk/.

3 Wachter, Sandra. "Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond." *Yale Journal of Law and Technology* 26: no. 3 (2024): 671-718. https://yjolt.org/limitations-and-loopholes-eu-ai-act-and-ai-liability-directives-what-means-european-union-united.

4 Hooker, Sara. "On the Limitations of Compute Thresholds as a Governance Strategy." arXiv preprint, 2024. https://doi.org/10.48550/arXiv.2407.05694.

5 Boine, Claire, and David Rolnick. "Why The AI Act Fails to Understand Generative AI." *Minnesota Journal of Law, Science & Technology* 26: no. 2 (2025): 61-123. https://scholarship.law.umn.edu/mjlst/vol26/iss2/2.

6 Hacker, Philipp, Andreas Engel, and Marco Mauer. "Regulating ChatGPT and Other Large Generative AI Models." *2023 ACM Conference on Fairness Accountability and Transparency,* June 12, 2023, 1112–23. https://doi.org/10.1145/3593013.3594067.

7 See #5.

8 Perrigo, Billy. "Big Tech Is Already Lobbying to Water Down Europe's AI Rules." *Time*, April 21, 2023. https://time.com/6273694/ai-regulation-europe/.

9 See #5.

10 The European Commission. "Commission Collects Feedback to Simplify Rules on Data, Cybersecurity and Artificial Intelligence in the Upcoming Digital Omnibus." Press Release, September 16, 2025. https://digital-strategy.ec.europa.eu/en/news/commission-collects-feedback-simplify-rules-data-cybersecurity-and-artificial-intelligence-upcoming.

11 Lenhart, Anna. "Leveraging International Standards to Protect U.S. Consumers Online, No Congress Required." Just Security, April 9, 2025. https://www.justsecurity.org/110127/international-standards-consumers-online/.

EU GUIDELINES ON GENERAL-PURPOSE AI MODELS

BY STAN ADAMS

DEFINITION

(17) Based on the considerations above, an indicative criterion for a model to be considered a general-purpose AI model is that its training compute is greater than 10^{23} FLOP and it can generate language (whether in the form of text or audio), text-to-image or text-to-video.

(18) This threshold corresponds to the approximate amount of compute typically used to train a model with one billion parameters on a large amount of data. While such models may be trained with varying amounts of compute depending on how much data is used, 10^{23} FLOP is typical for models trained on large amounts of data as of the time of writing (see the examples in Annex A.3).

(19) The modalities are chosen based on the fact that models trained to generate language—be it via text or speech (as a type of audio)—are able to use language to communicate, store knowledge, and reason. No other modality confers such a wide range of capabilities.

Consequently, models that generate language are typically more capable of competently performing a wider range of tasks than other models. Although models that generate images or video typically exhibit a narrower range of capabilities and use cases compared to those that generate language, such models may nevertheless be considered to be general-purpose AI models. Text-to-image and text-to-video models are capable of generating a wide range of visual outputs, which enables flexible content generation that can readily accommodate a wide range of distinct tasks.

(20) If a general-purpose AI model meets the criterion from paragraph 17 but, exceptionally, does not display significant generality or is not capable of competently performing a wide range of distinct tasks, it is not a general-purpose AI model. Similarly, if a general-purpose AI model does not meet that criterion but, exceptionally, displays significant generality and is capable of competently performing a wide range of distinct tasks, it is a general-purpose AI model.

FULL TITLE

Guidelines on the scope of the obligations for general-purpose AI models established by Regulation (EU) 2024/1689 (AI Act)

OFFICIAL TEXT

https://digital-strategy.ec.europa.eu/en/library/guidelines-scope-obligations-providers-general-purpose-ai-models-under-ai-act

LINEAGE

General-Purpose AI

DATE OF INTRODUCTION

July 18, 2025

PLACE OF ORIGIN

European Union

DATE ENACTED OR ADOPTED

August 2, 2025

ENFORCING ENTITY

European Commission

CURRENT STATUS

Enacted with staged enforcement (with rules for high-risk AI systems implemented later as part of staged enforcement)

LINEAGE DEFINITION

'general-purpose AI model' means an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market

MOTIVATION

Generally, the Guidance attempts to clarify the Commission's intended scope and application of the definition by including the rationale for certain aspects of the definition and by focusing the definition to capture models the Commission currently wishes to regulate, but to regulate differently than "high-risk" models. By using a guidance document, rather than including these additional details in the definition, the Commission retains the ability to modify its guidance over time and to adapt its scope to fit future models. For example, if the scale of computing power needed to train a model or the typical number of parameters changes, the Commission will be able to issue new guidance rather than having to go through the more rigorous process of adopting a new definition. Although this approach provides less certainty for potentially regulated entities, the tradeoff is a more "future-proof" approach to regulation. However, given the rapid pace of development and the closed-door nature of emerging research at the leading firms, it is not clear whether this approach, despite its flexibility, will allow the Commission to adjust its regulatory scope quickly enough to keep up. More likely, the Commission's evolving guidance will always be slightly behind (and perhaps not applicable to) the leading edge of development.

APPROACH

TL;DR:

- The Commission relies on "compute" and "modality" as primary factors in designating GPAI models.
- Assumptions based on frontier models may not hold true over time, but guidance can be more easily adjusted than legislative text.
- Edge cases will be difficult to capture under any approach.

TWO FACTORS TO ADD CERTAINTY: COMPUTE AND MODALITY.

In general, the Commission's approach is to augment the underlying definition from the AI Act[1] with additional details, attributes, and thresholds to clarify the Commission's current intentions for the scope of the regulation of "general-purpose AI models" under the Act. First, the Guidance adds a technical threshold based on the amount of compute used to train the model, expressed in the number of floating point operations or FLOPs. Second, the Guidance focuses on the "modalities" of various models, specifically noting that models built to generate language outputs, whether text or speech, typically have the broadest range of capabilities and use cases and are therefore more likely to be considered "general-purpose" than models that generate images or video. However, the Commission also explains that neither of these factors alone is dispositive—regardless of modality or training compute, any given model may or may not be considered "general-purpose" under the Guidance.

To mitigate the uncertainty around model classification under both the AI Act definition and the Guidance, the Commission adds a series of hypothetical models with various combinations of the two factors (compute and modality) to illustrate when models fall in or out of the scope of the definition. The single example of an in-scope model is positive on both factors, and so the Commission says that the model "is likely a general-purpose AI model" without being definitive. The lengthier set of examples of out-of-scope models all highlight when the Commission considers a model's capabilities to be too narrow to be "general-purpose," including models designed to "transcribe speech to text," "increase the resolution of images," and "generate music," even if the training compute of those models exceeds the FLOP threshold.

ASSUMPTIONS MAY BE FLAWED, BUT THEY CAN BE ADJUSTED.

Several questionable assumptions underlie the Commission's approach, but issuing new guidance is far easier than amending the legislative text.

First, the Guidance assumes that training compute, which it describes as a number proportional to the product of parameters and training examples, is a reliable proxy for a model's range of capabilities. The Commission further relies on compute as a proxy for capability when assessing whether a model presents "systemic risk," presuming that models requiring 10^25 FLOPs present such risks. See the discussion on FLOP thresholds in the previous section addressing the EU AI Act GPAI definition. The Guidance provides one rationale for using compute as a factor in assessing capability: developers will know how much compute was required in development, which provides at least some certainty when assessing a model's classification. The FLOP threshold may also have been chosen as a proxy to capture the most well-resourced firms—those able to bear the expense of training, refinement, and deployment. Regardless of motivation, the Commission's choice to set a FLOP threshold, at least, is relatively easy to adjust compared to rewriting the underlying definition.

Second, the Guidance assumes that a model's range of capabilities will necessarily be apparent upon its introduction to the market. This assumption depends on several variables. For example, it is possible that some firms will underestimate or not adequately test their models' capabilities before making them available, whether to avoid regulation or for other reasons. It is also possible that sophisticated systems become able to deceive researchers trying to assess their capabilities. There have already been reports that some models exhibit deceptive behavior when tested,[2]

leading to speculation that more sophisticated models may be able to hide any number of attributes including their relative "alignment" with human goals. In any case, the Commission may face difficulty refining the definitions' scope in this regard due to the wording within the definition itself.

Curiously not addressed is the idea of "competence." The Commission did not address its intended scope or application of the word "competently" as used in the AI Act definition. This leaves significant room for interpretation as to whether a model performs above or below any given threshold of competence. In future iterations of official guidance, perhaps the Commission will try to clarify what it means by competence. Further, inclusion of the word "competently" reveals another assumption the Commission makes about AI models and their potential risks: that only "competent" models pose risks worthy of regulation. Time may tell whether this is a valid assumption, but at least where humans are concerned, *incompetence* is often a cause for regulation.

THE REGULATOR'S DILEMMA: IS 80% GOOD ENOUGH?

Despite the potential flaws in its approach, the Commission's gamble that compute and modality will serve as effective proxies for AI model capability is, at least, understandable. In a rapidly developing field with many unknowns, it is likely impossible to tailor the scope of a regulatory definition to account for all potential risks. There will always be outliers, edge cases, and misuses of otherwise safe technology that even a case-by-case approach would struggle to address. Is it better, then, to address most of the foreseeable risks or invest significantly more resources to account for the long tail? This is a dilemma for every regulator.

Overall, the Commission has taken an approach typical to many regulators—set broad initial parameters to capture the intended features of a technology it wishes to regulate, then create more focused interpretations of the scope of the rule to address an evolving, difficult to predict reality. In this case, there are many reasons why the Commission may struggle to adapt the AI Act's definition to respond to the current state of the art. Perhaps most prominent—the rate of tech development already outpaces that of even the nimblest regulatory body.

RECEPTION

To date, there has not been significant public feedback on the Guidance. Several law firms have published blogs describing what the Guidance does, with some praise for the added clarity it provides. Some have noted that the Guidance was issued very late, considering that the AI Act obligations for GPAI took effect only 6 weeks later.

REFERENCES

1 European Parliament and Council. Regulation (EU) 2024/1689: laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

2 Greenblatt, Ryan, Carson Denison, Benjamin Wright, et al. "Alignment Faking in Large Language Models." arXiv preprint, 2024. https://doi.org/10.48550/arXiv.2412.14093.

SAFE, SECURE, & TRUSTWORTHY DEVELOPMENT & USE OF ARTIFICIAL INTELLIGENCE (EO 14110)

BY DR. RANJIT SINGH

DEFINITION

The term "dual-use foundation model" means an AI model that is trained on broad data; generally uses self-supervision; contains at least tens of billions of parameters; is applicable across a wide range of contexts; and that exhibits, or could be easily modified to exhibit, high levels of performance at tasks that pose a serious risk to security, national economic security, national public health or safety, or any combination of those matters, such as by:

(i) substantially lowering the barrier of entry for non-experts to design, synthesize, acquire, or use chemical, biological, radiological, or nuclear (CBRN) weapons;

(ii) enabling powerful offensive cyber operations through automated vulnerability discovery and exploitation against a wide range of potential targets of cyber attacks; or

(iii) permitting the evasion of human control or oversight through means of deception or obfuscation.

Models meet this definition even if they are provided to end users with technical safeguards that attempt to prevent users from taking advantage of the relevant unsafe capabilities.

FULL TITLE

Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

OFFICIAL TEXT

https://www.federalregister.gov/d/2023-24283

LINEAGE

General-Purpose AI

DATE OF INTRODUCTION

November 1, 2023 in EO 14110

PLACE OF ORIGIN

Federal, USA

DATE ENACTED OR ADOPTED

November 1, 2023

ENFORCING ENTITY

Executive Departments & Agencies

CURRENT STATUS

Revoked January 20, 2025

LINEAGE DEFINITION

'general-purpose AI model' means an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market

MOTIVATION

These definitions emerged in a rapidly shifting landscape of AI capabilities, shaped by the generative AI surge of 2022–2023 and a quickening international coordination—the G7 Hiroshima AI process,[1] the finalization of the EU AI Act[2]—even as differences in scope and emphasis saw the EU orient toward fundamental rights and societal impacts while the US leaned on national security. The EU named the phenomenon directly as "general-purpose AI models," tying obligations to systems whose capabilities travel across tasks and sectors (see the discussion on the EU AI Act GPAI definition for further details).

The US Executive Order (EO) 14110 instead coined "dual-use foundation models" to single out a narrower slice of frontier systems: large, broadly capable models whose performance on certain tasks (for example, in chemical, biological, radiological, or nuclear weapon design; sophisticated cyber operations; or evasion of human control) creates national security-relevant misuse risks, including when model weights are widely available. The definition is coupled to a reporting architecture that treats training compute and cluster capacity as proxies for power, directing the Department of Commerce to monitor very large training runs, models trained primarily on biological sequence data, and unusually capable compute clusters.

Even after the EO's revocation by the subsequent administration, this dual-use framing has persisted in Commerce proposals and NIST guidance.[3] In practice, the definition provides a regulatory template: a security-centred, compute-gated way of specifying which foundation models merit public oversight and why. It marks a shift from regulating what AI does to governing what it can do and the externalities that follow.

APPROACH

LINEAGE IN THE GOVERNANCE OF DUAL-USE TECHNOLOGIES.

The EO's definition of "dual-use foundation models" borrowed a long-standing security concept and narrowed it for the AI context. In long-running governance debates, "dual-use" refers to technologies and research that can be turned toward both beneficial and harmful ends: fertilizers and chemical weapons, nuclear power and bombs, viral genetics and engineered pathogens, intrusion tools for securing networks and for breaking into them. Governance efforts in these domains have tried to prevent malign repurposing without shutting down underlying science, using a mix of treaties, export controls, dual-use research of concern (DURC) review,[4] and professional ethics to manage the "deviation of intent"[5] from legitimate research to misuse.[6]

The EO picked up that lineage and applied it to foundation models, but with a specific emphasis. Rather than defining dual-use broadly as "beneficial and harmful potential," the order identified a subset of models whose capabilities intersected with a defined set of national security concerns—chemical, biological, radiological, nuclear, cyber, and deceptive risks. A dual-use foundation model, in this frame, was: (a) a large, broadly capable model trained on extensive data (a foundation model), which (b) either exhibited or could be readily modified to exhibit high-level performance on tasks that posed serious risks to national security, national economic security, or national public health and safety, with concrete examples in CBRN design, sophisticated cyber operations, and evasion of human control. The definition was indifferent to stated intent or safeguards; a model could qualify even if the provider had attempted to prevent misuse.

TWO-LAYER STRUCTURE: TASK CAPABILITIES AND COMPUTE THRESHOLDS.

The approach was structured in two layers. Qualitatively, it centered capability at specific dangerous tasks rather than general social harm: disinformation, privacy and intellectual property (IP) leakage, labor impacts, or discrimination were treated elsewhere in the EO,[7] and did not anchor the dual-use definition itself. Quantitatively, it coupled that capability frame to compute-based triggers. Training runs above 10^26 FLOPs were used as a gate for general dual-use models in Commerce's proposed reporting rules,[8] with a lower 10^23 FLOP threshold for models trained primarily on biological sequence data, and the same logic was extended to unusually capable clusters (high-bandwidth, high-throughput compute). In effect, compute and cluster characteristics were treated in a manner similar to "special material" in earlier dual-use regimes, the way fissile material or certain pathogens functioned as markers of heightened concern in nuclear and bio governance.

A RESPONSE TO THE AI SAFETY DEBATES.

Seen through the wider dual-use governance debates,[9] this move exhibited both similarities and differences from earlier regimes. It was similar in that it tried to create a monitored perimeter around the most consequential AI models, using reporting, red-teaming, and export-style oversight of hardware as early-warning tools, much as non-proliferation and DURC frameworks do in nuclear and life sciences. In the AI context, that perimeter-building was shaped by AI safety debates that elevated red-teaming of frontier models for catastrophic misuse—particularly biosecurity and cyber threats—as a central governance practice, even as public-interest and civil-rights groups were pushing to broaden red-teaming toward broader sociotechnical harms.[10] It was different from earlier

regimes in at least two ways. First, the "object" of regulation was intangible and easily copied: model weights and training code are far harder to track than uranium stockpiles, a challenge that has already been flagged previously for cyber tools. Second, the EO's dual-use definition bracketed off many of the dual-use concerns emphasised in human-rights-oriented debates on AI (surveillance, scalable persuasion, repression, structural discrimination), and instead codified a relatively narrow, security-first slice of the dual-use dilemma.

USING THE DEFINITION AS A TARGETED POLICY INSTRUMENT.

As a definitional strategy, "dual-use foundation model" was less a general theory of AI dual-use and more a targeted instrument for frontier scenarios involving catastrophic risk. It marked out a class of models where the US government claimed a particular stake in visibility and assurance, and it built that category using tools inherited from earlier dual-use regimes: task-based risk criteria, capability thresholds, and a reporting perimeter rather than blanket prohibition. Smaller or domain-specific systems, or models that primarily raised concerns about sociotechnical harms such as surveillance or labor exploitation, typically fell outside this national security framing even if they raised serious policy questions. Thus, it left these more diffuse, rights- and labor-related dual-use harms to other regulatory regimes, from sectoral and civil-rights law to labor and consumer protection.

To conclude, **this definition is well suited to policy instruments aimed at national security and catastrophic misuse**—for example, reporting obligations tied to frontier-scale training—**but it is not sufficient on its own for legislation focused on everyday**

civil-rights impacts of AI. In those domains, "dual-use foundation model" risks narrowing the field of concern to a small subset of harms and a particular way of measuring "frontier" AI.

WHEN TO USE THIS DEFINITION IN LEGISLATION:

USE "dual-use foundation model" when legislating on:

- ✔ CBRN weapons risks
- ✔ Offensive cyber capabilities
- ✔ AI deception/oversight evasion
- ✔ Frontier model reporting requirements

DON'T USE for legislation primarily about:

- ✘ Workplace discrimination
- ✘ Surveillance/privacy
- ✘ Consumer protection
- ✘ Labor displacement

RECEPTION

The EU and US approaches resonate with each other in recognizing foundation models as a pivotal regulatory concern, however, they diverge in regulatory philosophy (broad vs narrow scope, rules applied ex ante before market release vs oversight ex post after deployment) and immediate priorities (rights vs security).[11] In the US, industry tended to treat the dual-use definition as a manageable, security-focused overlay on work they were already

doing, not as a general template for AI regulation. They were already investing in red-teaming and security reviews for frontier models, and the EO largely formalized those practices, even as industry commenters warned about the volume and sensitivity of technical information they would have to disclose, and analysts noted that the same compute thresholds could later underpin export-style controls.

Civil society responses were more divided. Organizations focused on catastrophic and national security risks welcomed the move to single out frontier models with CBRN and cyber capabilities, and saw the dual-use definition as overdue recognition of those hazards. Rights, labor, and digital justice groups, by contrast, criticized the EO for structuring its most concrete model category around national security alone, leaving surveillance, discrimination, and workplace impacts to weaker or more diffuse parts of the policy landscape.[12] For them, the dual-use definition risked crowding out a broader conversation about AI's everyday harms.

Technical experts and standards bodies were left with the practical task of operationalizing this definition alongside the EU's GPAI and systemic-risk concepts: harmonizing compute thresholds across NIST and EU guidance, developing evaluation methods that speak to both security-oriented and rights-oriented concerns, and creating benchmarks that travel across jurisdictions. Emerging work in ISO/IEC and in collaborations between NIST's AI Safety Institute and its Center for AI Standards and Innovation can be read as attempts to build this interoperability, even as the underlying regulatory philosophies remain different. In this sense, EU definitions of GPAI and systemic risk have continued to shape the technical criteria that US actors have worked with. Effective

governance of foundation models will likely depend on translating these divergent regulatory logics into shared technical reference points that companies and regulators can use on both sides of the Atlantic.

ADDITIONAL RESOURCES

Kak, Amba, and Sarah Myers West. "General Purpose AI Poses Serious Risks, Should Not Be Excluded From the EU's AI Act | Policy Brief." AI Now Institute, April 13, 2023. https://ainowinstitute.org/publications/gpai-is-high-risk-should-not-be-excluded-from-eu-ai-act.

Aszódi, Nikolett. "EU's AI Act Fails to Set Gold Standard for Human Rights." Algorithm Watch, April 3, 2024. https://algorithmwatch.org/en/ai-act-fails-to-set-gold-standard-for-human-rights/.

Leufer, Daniel, Caterina Rodelli, and Fanny Hidvegi. "Human Rights Protections… with Exceptions: What's (Not) in the EU's AI Act Deal." Access Now, December 14, 2023. https://www.accessnow.org/whats-not-in-the-eu-ai-act-deal/.

REFERENCES

1 G7. "G7 Leaders' Statement on the Hiroshima AI Process." Ministry of Foreign Affairs of Japan, October 30, 2023. https://www.mofa.go.jp/ecm/ec/page5e_000076.html.

2 European Parliament and Council. Regulation (EU) 2024/1689: laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

3 National Institute of Standards and Technology (NIST). "Managing Misuse Risk for Dual-Use Foundation Models." NIST AI 800-1 2pd, Second Public Draft, January 2025. https://doi.org/10.6028/NIST.AI.800-1.2pd.

4 Ienca, Marcello, and Effy Vayena. "Dual Use in the 21st Century: Emerging Risks and Global Governance." *Swiss Medical Weekly* 148, no. 4748 (2018): w14688. https://doi.org/10.4414/smw.2018.14688.

5 DiEuliis, Diane, and James Giordano. "Gene Editing Using CRISPR/Cas9: Implications for Dual-Use and Biosecurity." *Protein & Cell* 9, no. 3 (2018): 239–40. https://doi.org/10.1007/s13238-017-0493-4.

6 Harris, Elisa D., ed. Governance of Dual-Use Technologies: Theory and Practice. Cambridge, MA: American Academy of Arts & Sciences, 2016. https://www.amacad.org/publication/governance-dual-use-technologies-theory-and-practice.

7 Haven, Janet. "A New Executive Order Ties Equity in AI to a Broader Civil Rights Agenda." Data & Society, February 23, 2023. https://datasociety.net/points/a-new-executive-order-ties-equity-in-ai-to-a-broader-civil-rights-agenda/.

8 Department of Commerce Bureau of Industry and Security. "Establishment of Reporting Requirements for the Development of Advanced Artificial Intelligence Models and Computing Clusters." 15 C.F.R. § 702 (2024). https://www.federalregister.gov/d/2024-20529.

9 See #6.

10 Singh, Ranjit, Borhane Blili-Hamelin, Carol Anderson, et al. "Red-Teaming in the Public Interest." Data & Society and AI Risk and Vulnerability Alliance, February 2025. http://doi.org/10.69985/VVGP4368.

11 Bullock, Charlie, Suzanne Van Arsdale, Mackenzie Arnold, Cullen O'Keefe, and Christoph Winter. "Legal Considerations for Defining 'Frontier Model.'" Institute for Law & AI, Working Paper No. 2-2024, September 2024. https://law-ai.org/frontier-model-definitions/.

12 American Civil Liberties Union, Center for American Progress, Leadership Conference on Civil and Human Rights, and National Fair Housing Alliance. "Coalition Comments to NTIA on Open-Source AI Systems." The Leadership Conference on Civil and Human Rights, March 27, 2024. https://civilrights.org/resource/coalition-comments-to-ntia-on-open-source-ai-systems/.; and see #7.

CALIFORNIA BILL SB-53

BY MIRANDA BOGEN

DEFINITION

"Foundation model" means an artificial intelligence model that is all of the following:

(1) Trained on a broad data set.

(2) Designed for generality of output.

(3) Adaptable to a wide range of distinctive tasks.

(1) "Frontier model" means a foundation model that was trained using a quantity of computing power greater than 10^26 integer or floating-point operations.

(2) The quantity of computing power described in paragraph (1) shall include computing for the original training run and for any subsequent fine-tuning, reinforcement learning, or other material modifications the developer applies to a preceding foundation model.

FULL TITLE

Transparency in Frontier Artificial Intelligence Act

OFFICIAL TEXT

https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260SB53

LINEAGE

General-Purpose AI

DATE OF INTRODUCTION

January 7, 2025 in SB-53

PLACE OF ORIGIN

California, USA

DATE ENACTED OR ADOPTED

September 29, 2025

ENFORCING ENTITY

State Attorney General & Government Operations Agency

CURRENT STATUS

Enacted

LINEAGE DEFINITION

'general-purpose AI model' means an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market

MOTIVATION

While a variety of AI definitions describe basic characteristics of the technology, the definition of "foundation model" reflected in SB-53 was responsive to the same advances in AI that enabled models to perform a variety of tasks without being specifically designed or trained to do so that motivated the inclusion of general-purpose AI models in the EU AI Act.[1] Such capabilities were a notable and somewhat sudden deviation from machine learning models and earlier artificial intelligence systems that could conduct perception, pattern detection, and prediction tasks but that generally required developers to design tools for specific purposes.

Some have also been concerned that particularly capable general-purpose models might pose serious, systemic, and even catastrophic risks and so require heightened due diligence, safeguards, or oversight. In particular, AI safety researchers became concerned that models themselves might become so capable that controlling their behavior and proliferation would be critical to preventing dramatic harm to humanity. Without reliable ways to predict the conditions that would lead to models demonstrating these advanced capabilities, compute thresholds become a popular—albeit contested—proxy for identifying highly capable models, driven by early research suggesting correlation between training compute and model capability. SB-53's definition of frontier models differs from similar, compute-based definitions by including within its threshold the amount of computation not only used for initial model training, but also for fine-tuning and other post-training modifications, which research has increasingly shown can meaningfully affect model behavior and capability.

APPROACH

While it is slightly more nuanced than similar predecessors, SB-53 still targets a narrow set of developers of advanced AI models. The bill's definition for foundation models draws from a widely cited paper by the Center for Research on Foundation Models (CRFM) at Stanford University which coined the term and defined it as "any model that is trained on broad data (generally using self-supervision at scale) that can be adapted to a wide range of downstream tasks."[2] The bill adapts this definition by removing technical implementation details and instead referring to models "designed to produce a generality of outputs"—a revision likely influenced by deliberations surrounding the EU Artificial Intelligence Act. The definition is sufficiently general to capture models that may not have a specific goal or purpose but may still present issues of concern to policymakers. On the other hand, concepts like "broad data set," "generality of output," and "wide range of distinctive tasks" are under-defined and will likely be contested. Most of SB-53's hallmark requirements apply to frontier models, a small subset of foundation models determined by the amount of computing power used to train or modify them, given the risks some stakeholder communities presume such models will introduce. In this way, the definition deviates from similar ones, which have typically tied compute thresholds only to model training, presumably given that many more recent advancements in model capability have come not purely through increasing the data and compute used to train models but also the amount of compute spent on efforts that follow the initial pre-training phase. However, even this more nuanced definition was not intended to cover a wide variety of what some perceive to be more "mundane" risks of AI, such as its use in consequential domains or use-cases in ways that could affect people's economic security, health, or freedom.

The definition may not cover increasingly popular AI development techniques or the broader systems powered by covered models. It is not clear whether this definition would account for the increasingly substantial amount of compute used for model inference and reasoning—that is, running the model and integrating so-called "chain of thought" into the model's operation once it has been deployed. And while the bill's definition of foundation and frontier models covers certain large-scale models such as GPT-4 or Claude 4.5, it may not cover elements of the systems within which such models are deployed, like ChatGPT or Claude that involve additional elements such as scaffolding, model routers, safety filters, or other features that interact with the foundation model itself. This definition would also not account for narrower models that pose foreseeable harm, smaller models that have been "distilled" from larger, covered ones, or ensembles of multiple small models or systems that interact in a way that produces heightened capabilities. Periodic updates are important to ensure threshold-based definitions still capture the technology or actors intended to be in scope.

Rather than imposing requirements on specific AI models, SB-53 focuses mostly on entities developing these models. As such, definitions of foundation model and frontier model feed into companion definitions for "frontier developer" and "large frontier developer" (distinguished by annual gross revenue). Some have commented that actor-based thresholds better capture large developers while avoiding placing burden on startups and other small/medium-sized businesses,[3] though such an approach may leave gaps where smaller organizations manage to build models that either surpass defined thresholds or stay within them but nevertheless pose a high risk of harm. To account for this limitation, the law directs the California Department of Technology to

annually review definitions of frontier model, frontier developer, and large frontier developer, and submit recommendations to the legislature to consider changes. Such a directive is critical given that threshold-based definitions are likely to be quickly outdated and are unlikely to capture all models of concern.[4]

The definition of "frontier model" is quite narrow, on purpose. SB-53 was sponsored by California State Senator Scott Wiener, following Governor Gavin Newsom's veto of a previous, related, legislative effort (SB-1047[5]) that aimed to mitigate certain safety risks the sponsors worried would be posed by capable AI models like chemical, biological, radiological, and nuclear threats. While that bill passed both the state House and Senate, subsequent backlash from a wide variety of industry, academic, and public interest stakeholders led to the governor's veto and the subsequent convening of an expert policy panel to advise California on how to regulate AI. The resulting California Report on Frontier AI Policy heavily influences the substance of SB-53.[6] In crafting SB-53, Senator Wiener explained his intent to impose requirements on only "a small number of well-resourced companies, and only to the most advanced models,"[7] a goal effectuated by the high compute threshold in the bill's definition of frontier model. The bill's definitions also serve to exempt from its testing obligations simpler models perceived by the bill's supporters to pose less risk,[8] in part, perhaps, to defend the bill against claims that it would impose undue burdens on innovations critical to both economic growth and national security.

The bill's substantive requirements are a start, but may not be as effective in changing safety practices as its proponents hope. The theory of change around bills like SB-53 seems to be that if covered entities that develop a potentially harmful technology are

required to disclose their risk management practices, they will need to define and operationalize those practices to begin with, and will therefore be more likely to manage the risks they encounter. However, if a law requires insufficiently detailed or overly narrow documentation or fails to provide sufficient resources to enable active and savvy enforcement, it can be all too easy for organizations to performatively comply without meaningfully changing their safety practices.

RECEPTION

SB-53 received support from Encode AI, Economic Security Action California, the Secure AI Project, and prominent AI researchers Geoffrey Hinton and Yoshua Bengio, who argued the bill would advance transparency and accountability from frontier AI companies. Former White House AI advisor Dean Ball praised its technical sophistication, and frontier AI lab Anthropic endorsed the bill after substantive involvement. OpenAI and industry groups like CTA and Chamber of Progress pushed back against the obligations it imposed.[9] Some critics questioned the FLOP threshold in its definition of frontier model, though less vocally than in earlier compute-based efforts—likely because similar (and blunter) approaches had appeared in other prominent legislation and the bill includes mechanisms for periodic threshold reviews.

ADDITIONAL RESOURCES

Kodama, Miles. "California Senate Bill 53." 2025. https://sb53.info/. (Was not updated to account for amendments.)

Bogen, Miranda. "Assessing AI: Surveying the Spectrum of Approaches to Understanding and Auditing AI Systems." Center for Democracy and Technology (CDT), January 2025. https://cdt.org/insights/assessing-ai-surveying-the-spectrum-of-approaches-to-understanding-and-auditing-ai-systems/.

REFERENCES

1 European Parliament and Council. Regulation (EU) 2024/1689: laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

2 Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, et al. "On the Opportunities and Risks of Foundation Models." arXiv preprint, 2021. https://doi.org/10.48550/arXiv.2108.07258.

3 Ball, Dean W. "What's up with the States? A Check-In." Hyperdimensional, August 21, 2025. https://www.hyperdimensional.co/p/whats-up-with-the-states.

4 Hooker, Sara. "On the Limitations of Compute Thresholds as a Governance Strategy." arXiv preprint, 2024. https://doi.org/10.48550/arXiv.2407.05694.

5 California Legislature. SB-1047 Safe and Secure Innovation for Frontier Artificial Intelligence Models Act. (2024). https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB1047.

6 Bommasani, Rishi, Scott R. Singer, Ruth E. Appel, et al. "The California Report on Frontier AI Policy." The Joint California Policy Working Group on AI Frontier Models, June 17, 2025. https://www.gov.ca.gov/wp-content/uploads/2025/06/June-17-2025-%E2%80%93-The-California-Report-on-Frontier-AI-Policy.pdf.

7 Senator Scott Wiener. "Senator Wiener Expands AI Bill into Landmark Transparency Measure Based on Recommendations of Governor's Working Group." July 9, 2025. https://sd11.senate.ca.gov/news/senator-wiener-expands-ai-bill-landmark-transparency-measure-based-recommendations-governors.

8 O'Brien, Matt. "Regulators Turn to Math to Determine When AI Is Powerful Enough to Be Dangerous." *PBS News*, September, 2024. https://www.pbs.org/newshour/nation/regulators-turn-to-math-to-determine-when-ai-is-powerful-enough-to-be-dangerous.

9 See for example:
Hizkias, Aden. "Why California's SB 53 Still Gets AI Regulation Wrong: And How It Compares to Last Year's SB 1047." Chamber of Progress, July 9, 2025. https://progresschamber.org/insights/why-californias-sb-53-still-gets-ai-regulation-wrong/

CLOSING

WHEN *NOT* TO DEFINE AI

Throughout this handbook, we have taken legal definitions of artificial intelligence seriously, examining how they are written, borrowed, modified, and operationalized across jurisdictions and policy contexts. That focus reflects current reality: lawmakers, regulators, and advocates are repeatedly asked to define AI, and the consequences of those choices are increasingly high-stakes. But there is an important alternative path that deserves consideration: sometimes, the best approach is to avoid needing to define AI at all, even if the starting point for legal action is motivated by AI.

The road not taken in some legislative efforts is to skip defining AI entirely, and to even avoid mentioning AI in statutory text. This should not necessarily be viewed as simply skirting a difficult challenge or the failing to understand the technology. In some cases, it is a deliberate and disciplined choice grounded in a recognition that the objectives of a law do not actually turn on whether a system qualifies as AI, but rather on the risks, harms, or institutional practices that the law seeks to address.

Simply put, if in reading this book you are finding that none of the various definitions are meeting your needs or objectives, consider that when legislating about AI, sometimes you do not need to legislate about AI itself.

THE TECHNOLOGY-NEUTRAL APPROACH

Technology-neutral approaches have a long pedigree in law and regulation. Rather than anchoring obligations to a specific class of tools, they focus on conduct, outcomes, risks, or institutional roles. In the context of emerging technologies, this approach can be especially valuable. AI often renders certain risks more salient, more scalable, or more opaque—but it rarely invents those risks from whole cloth.

Discrimination, unsafe products, deceptive practices, invasions of privacy, environmental harm, labor displacement, and market concentration all predate modern AI systems. In many contexts, the most durable legal interventions target these underlying harms directly, without hinging applicability on whether a particular system meets a contested technical definition.

A technology-neutral framing can also be advantageous when the policy goal is to support innovation, scientific advancement, or technological progress. Fixating on a particular technology of the day risks freezing a moment in time. While an AI bill might capture popular attention, artificially directing inquiry toward or away from specific approaches, rather than allowing for more open-ended experimentation, can undermine the real goal. History is replete with examples of regulatory frameworks that unintentionally privileged certain technical paths simply because those paths were legible to policymakers at the time of drafting.

For example, early US telecommunications law was drafted around a circuit-switched, voice-centric model of communication, which made packet-switched data networks difficult to classify once they emerged. Similarly, early internet copyright law assumed centralized intermediaries capable of monitoring and

responding to infringement, embedding notice-and-takedown obligations that mapped cleanly onto large platform-like service providers. This framework implicitly privileged centralized technical architectures over decentralized or peer-to-peer systems, not as a normative choice, but because those architectures were easier for lawmakers to see, assign responsibility to, and regulate.

In this sense, technology neutrality can be a way of preserving flexibility, avoiding premature lock-in, and ensuring that legal obligations track social objectives rather than transient technical categories.

DEFINITIONAL UNCERTAINTY AS A SIGNAL

One theme that should stand out to readers of this handbook is the persistent difficulty of calibrating AI definitions. Across lineages, jurisdictions, and use cases, drafters struggle with the same core problem: how to draw a boundary that is neither so narrow that it is easily evaded, nor so broad that it sweeps in ordinary software or well-understood forms of automation.

This challenge of inclusion versus exclusion is not a drafting failure so much as a signal. It reflects the reality that "AI" is not a neat or stable category of computing technologies. It is an evolving cluster of techniques, applications, and institutional practices, shaped as much by business models, marketing buzzwords, and cultural practices as by technical properties.

Many of the definitions examined in this handbook attempt to manage this uncertainty through companion terms and scoped obligations. Rather than relying on the AI definition alone, they pair it with concepts like "developer," "deployer," "high-risk

system," or "consequential decision." These auxiliary definitions often do the real work of determining who is covered, when obligations attach, and how enforcement operates in practice.

This design choice is telling. It suggests that even when drafters do define AI, they frequently rely on other concepts to make the law function. In some cases, those concepts may be sufficient on their own!

CHANGE OVER TIME–AND THE LIMITS OF FORESIGHT

Another lesson that emerges across entries is how tightly definitions are shaped by the moment. Several of the entries describe in their Motivation sections how these definitions arose from the cultural and political response to the rise of "generative AI." Terms like "content," "implicit objectives," or "general-purpose models" reflect attempts to adapt existing frameworks to a sudden and highly visible shift in capabilities and usage patterns for AI.

Many of these definitions aim to be flexible and forward-looking, and in some cases they succeed. But technological change can be dramatic and unpredictable, often in ways that defy the assumptions embedded in statutory text. Definitions that seem capacious today may become inadequate tomorrow—not because they were poorly drafted, but because the underlying technological landscape, business models, or usage moved in an unexpected direction.

This reality should prompt a threshold question for drafters: do the objectives of the proposed law truly depend on capturing AI as such or would a broader, more technologically neutral approach

better serve those aims over time? If the answer is the latter, defining AI may introduce unnecessary fragility into the legal framework.

CHOOSING WHEN TO DEFINE

None of this is an argument against defining AI categorically. In many contexts definitions are unavoidable, clarifying, and extremely useful. This handbook exists precisely because those moments are frequent and consequential.

But that choice is still a choice. Avoiding an AI definition can be an intentional design decision, one that reflects clarity about policy goals rather than uncertainty about technology or a shortcut out of making a hard call. In some cases, the most effective laws will be those that regulate behavior, responsibility, and harm directly, leaving the taxonomy of tools to evolve outside the statute.

However, in the other cases where grappling with AI and its definition directly is necessary or desirable, this handbook aims to serve. If it succeeds, it should make our choices about how to define AI easier, clearer, and more intentional; it should illuminate the intent, or else the carelessness, of proposed definitions that we encounter; and it should help us thoughtfully and consciously define this technology of our time.

ABOUT THE AUTHORS

WRITERS

Stan Adams is the Wikimedia Foundation's Lead Public Policy Specialist for North America. He advocates for laws and policies to help projects like Wikipedia thrive and works to protect the Wikimedia model from harmful policies. This work includes defending legal protections, like Section 230, that protect volunteer editors from frivolous lawsuits and advocating for stronger laws to protect the privacy of people who read and edit Wikipedia. Stan has spent nearly a decade working on tech policy issues including digital copyright, free expression, privacy, and AI. Prior to his role at the Wikimedia Foundation, Stan served as a general counsel to a US Senator and as a deputy general counsel at the Center for Democracy and Technology. Stan enjoys reading, listening to music, playing games, and spending time in the woods.

Miranda Bogen is the founding Director of CDT's AI Governance Lab, where she works to develop and promote adoption of robust, technically-informed solutions for the effective regulation and governance of AI systems.

An AI policy expert and responsible AI practitioner, Miranda has led advocacy and applied work around AI accountability across both industry and civil society. She most recently guided strategy and implementation of responsible AI practices at Meta, including driving large-scale efforts to measure and mitigate bias in AI-powered products and building out company-wide governance practices. Miranda previously worked as senior policy analyst at Upturn, where she conducted foundational research at the intersection of machine learning and civil rights, and served as co-chair of the Fairness, Transparency, and Accountability Working Group at the Partnership on AI.

Miranda has co-authored widely cited research, including empirically demonstrating the potential for discrimination in personalized advertising systems and illuminating the role artificial intelligence plays in the hiring process, and has helped to develop technical contributions including AI benchmarks to measure bias and robustness, privacy-preserving methods to measure racial disparities in AI systems, and reinforcement-learning driven interventions to advance equitable outcomes in products that mediate access to economic opportunity. Miranda's writing, analysis, and work has been featured in media including the Harvard Business Review, NPR, The Atlantic, Wired, Last Week Tonight, and more.

Miranda holds a master's degree from The Fletcher School at Tufts University with a focus on international technology policy, and graduated summa cum laude and Phi Beta Kappa from UCLA with degrees in Political Science and Middle Eastern & North African Studies.

Dr. Rumman Chowdhury (she/her) is a globally recognized leader in data science and responsible AI, uniquely positioned at the intersection of industry, civil society, and government. She is a sought-after speaker at high-profile venues, including TED, the World Economic Forum, the UN AI for Good Summit, and the European Parliament, where she brings a rare combination of technical expertise and real-world implementation experience. She is the former US Science Envoy for AI at the US Department of State; the former Engineer Director of Machine Learning, Ethics and Transparency at Twitter; the founder and CEO of Parity AI (acquired by Twitter); and the former Managing Director of Responsible AI at Accenture. Rumman co-founded Humane Intelligence in 2022 and served as its CEO until August 2025. She stepped down to launch her new startup, details about which are coming soon. Rumman holds dual Bachelor degrees from MIT, a Master degree from Columbia University, and completed her PhD in Political Science at UC San Diego. She remains a Distinguished Advisor of Humane Intelligence, the nonprofit.

Dr. Christine Custis is a Research Associate and Program Manager for the Science, Technology, and Social Values Lab. A computer scientist and organizational strategist whose work has spanned industry, civil society, and academia, Dr. Custis has more than two decades of experience in the development and governance of emerging science and technology. At IAS, she collaborates with Alondra Nelson on multidisciplinary research and policy initiatives examining the social implications of AI, genomics, and quantum science.

She previously served as Director of Programs and Research at the Partnership on AI a nonprofit, multisector coalition of organizations committed to the responsible use of artificial intelligence. At PAI, she oversaw the research, analysis, and development of a range of AI-related resources, from policy recommendations and publications to tools, while leading workstreams on labor and political economy; transparency and accountability; AI safety; inclusive research and design, and the public understanding of AI. An expert of both domestic and international policy, Christine served as a member of the Organisation for Economic Co-operation and Development (OECD) expert group on trustworthy AI and previously worked at the MITRE Corporation and IBM. She has advised and collaborated with a range of organizations, including the Global Democracy Coalition, the US National Institute of Standards and Technology, the US National AI Research Resource Task Force, the New America Open Technology Institute, the UC Berkeley Center for Long-Term Cybersecurity, and many others. She holds a M.S. degree in computer science from Howard University and received her Ph.D. from Morgan State University.

Mark Dalton leads R Street's technology and cybersecurity policy portfolios as the organization's senior director of technology and innovation.

Before joining R Street, Mark served as head of strategic planning at the U.S. Naval Undersea Warfare Center (NUWC) Division Newport, where he played a key role in aligning the organization's acquisition and research initiatives with the strategic needs of the Navy and the Department of Defense. Prior to that, he served as NUWC's chief engineer for undersea warfare cybersecurity, delivering technical excellence in cyber solutions across research, development, engineering, and testing. He also established NUWC's inaugural portfolio of cybersecurity research and development, strategically focusing on leveraging artificial intelligence, machine learning, and formal mathematical methods to combat cyber threats.

As a computer scientist, Mark conducted research in the application of reinforcement learning to optimize behaviors in autonomous systems. He also managed numerous installations and tests of prototype technologies aboard U.S. Navy submarines worldwide.

Mark holds a master's degree in national security and strategic studies from the U.S. Naval War College, a master's degree in computer science from the New Jersey Institute of Technology, and a bachelor's degree in information technology from the University of Massachusetts Lowell. He is currently pursuing a PhD in international relations at Salve Regina University.

Mark resides in Portsmouth, Rhode Island, with his wife, Nichole, and their 18-year-old daughter, Addison. His 3-year-old dog, Ziggy, is his loyal officemate.

Anna Lenhart is a Policy Fellow at Institute for Data Democracy and Politics (IDDP) at George Washington University and a researcher at the University of Maryland Ethics and Values in Design Lab. Her research focuses on public engagement in tech policy and the intersections of privacy, transparency, and competition policy. She most recently served in the House of Representatives as Senior Technology Legislative Aid to Rep Lori Trahan (117th Congress) and as a Congressional Innovation Fellow for the House Judiciary Digital Markets Investigation (116th).

Prior to working for Congress, Anna was a Senior Consultant and the AI Ethics Initiative Lead for IBM's Federal Government Consulting Division, training data scientists and operationalizing principles of transparency, algorithm bias and privacy rights in AI and Machine Learning systems. She has researched the human right to freedom from discrimination in algorithms, public views on autonomous vehicles and the impact of AI on the workforce. She holds a master's degree from Ford School at University of Michigan and a BS in Civil Engineering and Engineering Public Policy from Carnegie Mellon University. Prior to graduate school, Anna was the owner of Anani Cloud Solutions, a consulting firm that implemented and optimized Salesforce.com for non-profit organizations.

Dr. Margaret Mitchell is a researcher focused on machine learning (ML) and ethics-informed AI development. She has published over 100 papers on natural language generation, assistive technology, computer vision, and AI ethics, and holds multiple patents in the areas of AI conversation generation and sentiment classification. She has been recognized in TIME's Most Influential People; Fortune's Top Innovators; and Lighthouse3's 100 Brilliant Women in AI Ethics.

She currently works at Hugging Face as a researcher and Chief Ethics Scientist, driving forward work on ML data processing, responsible AI development, and AI ethics. She previously worked at Google AI, where she founded and co-led Google's Ethical AI group to advance foundational AI ethics research and operationalize AI ethics internally. Before joining Google, she worked as a researcher at Microsoft Research and as a postdoc at Johns Hopkins.

She has spearheaded a number of workshops and initiatives at the intersections of diversity, inclusion, computer science, and ethics. Her work has received awards from Secretary of Defense Ash Carter and the American Foundation for the Blind, and has been implemented by multiple technology companies. She is most known for her work pioneering "Model Cards" for ML model reporting; developing "Seeing AI" to assist blind and low-vision individuals; and developing methods to mitigate unwanted AI biases.

She holds a PhD in Computer Science from the University of Aberdeen and a Master's in Computational Linguistics from the University of Washington. She likes gardening, dogs, and cats.

Professor Rashida Richardson is a Distinguished Scholar of Technology and Policy at Worcester Polytechnic Institute. Rashida is an internationally recognized expert in civil rights and artificial intelligence, and a legal practitioner in technology policy issues. Rashida has previously served as an Attorney Advisor to the Chair of the Federal Trade Commission and as a Senior Policy Advisor for Data and Democracy at the White House Office of Science and Technology Policy in the Biden Administration. She has worked on a range of civil rights and technology policy issues at the German Marshall Fund, Rutgers Law School, AI Now Institute, the American Civil Liberties Union of New York (NYCLU), and the Center for HIV Law and Policy. Her work has been featured in the Emmy-Award Winning Documentary, The Social Dilemma, and in major publications like the New York Times, Wired, MIT Technology Review, and NPR (national and local member stations). She received her BA with honors in the College of Social Studies at Wesleyan University and her JD from Northeastern University School of Law.

Dr. Ranjit Singh is the director of Data & Society's AI on the Ground program, where he oversees research on the social impacts of algorithmic systems, the governance of AI in practice, and emerging methods for organizing public engagement and accountability. His own work focuses on how people live with and make sense of AI, examining how algorithmic systems and everyday practices shape each other. He also guides research ethics at Data & Society and works to sustain equity in collaborative research practices, both internally and with external partners. His work draws on majority world scholarship, public policy analysis, and ethnographic fieldwork in settings ranging from scientific laboratories and bureaucratic agencies to public services and civic institutions. At Data & Society, he has previously led projects mapping the conceptual vocabulary and stories of living with AI in/from the majority world, framing the place of algorithmic impact assessments in regulating AI, and investigating the keywords that ground ongoing research into the datafied state.

He holds a PhD in Science and Technology Studies from Cornell University. His dissertation research focused on Aadhaar, India's biometrics-based national identification system, advancing public understanding of how identity infrastructures both enable and constrain inclusive development and reshape the nature of Indian citizenship.

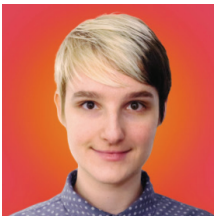
Dr. Nathan C. Walker is an award-winning First Amendment and human rights educator at Rutgers University, where he teaches AI ethics and law as an Honors College faculty fellow. He is the principal investigator at the AI Ethics Lab, the founding editor of the AI & Human Rights Index, a contributing researcher to the Munich Declaration of AI, Data and Human Rights, and a non-resident research fellow at Stellenbosch University in South Africa. Dr. Walker is a certified AI Ethics Officer and has held visiting research appointments at Harvard and Oxford universities. He has also served as an Expert AI Trainer for OpenAI's Human Data Team, where he applied his expertise in law and education to frontier AI models. He has authored five books on law, education, and religion, and presented his research at the UN Human Rights Council, the Italian Ministry of Foreign Affairs, and the U.S. Senate. Nate has three learning disabilities and earned his doctorate in First Amendment law and two master's degrees from Columbia University. An ordained Unitarian Universalist minister, Reverend Nate holds a Master of Divinity from Union Theological Seminary. More at sites.rutgers.edu/walker

Professor Ben Winters is the Director of AI and Privacy at the Consumer Federation of America. Ben leads CFA's advocacy efforts related to data privacy and automated systems and works with subject matter experts throughout CFA to integrate concerns about privacy and AI in order to better advocate for consumers. Ben is also an adjunct professor at the University of the District of Columbia David A. Clarke School of Law.

Prior to CFA, Ben worked at the Civil Rights Division of the Department of Justice, where he was an Attorney Advisor in the policy section focusing on algorithmic harm in the civil rights context and was Senior Counsel at the Electronic Privacy Information Center (EPIC) where he led the AI/Human Rights project and advocated for accountability through legislative and direct legal action.

Ben is a graduate of Benjamin N Cardozo Law School and the SUNY Oneonta and is a member of the New York State and District of Columbia bars.

EDITORS



B Cavello is a technology and facilitation expert who is passionate about creating social change by empowering everyone to participate in technological and social governance. They serve as director of emerging technologies for Aspen Digital, a program of the Aspen Institute. B also serves as assistant program chair for the Neural Information Processing Systems (NeurIPS) Conference, as a 2025 research fellow with Siegel Family Endowment, and as a board member to Metagov, an interdisciplinary research nonprofit promoting digital self-governance.

Previously, B advised Senator Ron Wyden on issues of privacy, internet governance, and algorithmic accountability. Prior to Congress, they were a research program lead at the Partnership on AI, senior engagement lead for IBM Watson, and director of both product development and community at Exploding Kittens. B has been recognized as an LGBT+ 'Out Role Model,' Global Crowdfunding 'All-Star,' and was selected as an MIT-Harvard Assembly Fellow for the 2019 Ethics and Governance in Artificial Intelligence Initiative cohort.

Nicholas P. Garcia is a Senior Policy Counsel at Public Knowledge, focusing on emerging technologies, intellectual property, and closing the digital divide. Before joining Public Knowledge, Nick served as an Assistant District Attorney in the Investigations Division of the Bronx County District Attorney's Office, where he investigated and prosecuted cybercrime, fraud, grand larceny, and organized crime. He previously worked as a Legal Intern at Public Knowledge and as a Student Attorney for the Communications & Technology Law Clinic at Georgetown University Law Center's Institute for Public Representation.

Nick received his J.D. from Georgetown University Law Center, his M.A. in Ethics and Society from Fordham University's Center for Ethics Education, and his B.A. in cursu honorum in Philosophy from Fordham University, where he was a member of the Phi Sigma Tau Honor Society.

Nick is a native New Yorker, and an unabashed nerd who loves video games, science fiction, and tabletop RPGs.

Eleanor Tursman is an Emerging Technologies Researcher at The Aspen Institute, where they work at the intersection of novel technologies, public education, and public policy. They served as a Siegel Family Endowment Research Fellow from 2022-24.

Eleanor is a computer scientist by training invested in bridging the gap between academic research and policy to promote social good. They believe in advancing science-based governance of artificial intelligence and other emerging technologies, tools which—despite their potential benefits—can also be used to disenfranchise and discriminate against minority groups, threaten democracy, and propagate disinformation.

Prior to joining The Aspen Institute, Eleanor served as a TechCongress fellow in the office of Representative Trahan (D-MA-03), providing technical and scientific support for topics including health technology, educational technology, artificial intelligence, algorithmic bias, and data privacy. They also worked as a computer vision researcher at Brown University, with a specialty in 3D imagery and the human face. Their research centered on finding new ways to approach deepfake detection, with a focus on long-term robustness as fake video becomes harder to visually identify. Eleanor has an M.S. in computer science from Brown University and a B.A. in physics with honors from Grinnell College.