Artificial Intelligence Executive Oversight Group AI Lexicon February 2026

The Shared Artificial Intelligence (AI) Lexicon was created by the AI Lexicon workstream of the Artificial Intelligence Executive Oversight Group (AIEOG). The AIEOG is a public-private partnership formed by the U.S. Department of the Treasury (Treasury), in collaboration with the Financial and Banking Information Infrastructure Committee (FBIIC) and the Financial Services Sector Coordinating Council (FSSCC).

The Lexicon was developed to promote a shared vocabulary for communication and collaboration on AI-related matters in the financial sector. It compiles commonly used risk management and technical terms that have specific meanings in the context of AI use in financial services. Definitions are drawn from a range of standards, academic publications, and government resources, and were informed by input from financial institutions, AI service providers, and public-sector participants. As AI technology and practice evolve, the Lexicon may be updated.

The Lexicon is an optional tool and is not intended for use either in the legal interpretation of any regulations or regulatory oversight reports or supervisory statements of US financial regulators, international arrangements or agreements, or private contracts. The inclusion of any source, term, or definition is for transparency and does not imply endorsement of any product, service provider, or organization. The Lexicon reflects the practical experience and collaborative work of participants and does not represent the official views of Treasury or any participating organization.

Lexicon Terms

Adversarial AI – Techniques and attacks used to manipulate AI systems, causing them to make incorrect or unintended predictions or decisions. These techniques exploit vulnerabilities in AI models, often by subtly altering input data, training data, or model interactions to manipulate the AI system.[1]

Agentic AI – A category of AI systems capable of independently making decisions, interacting with their environment, and optimizing processes without direct human intervention.[2]

AI Agent – A system that autonomously perceives its environment, decides what to do, and takes actions to achieve its goals.[3]

AI drift/ decay – The tendency for an AI model's performance to degrade over time when deployed in a real-world setting with differing conditions from those present in training and testing.[4]

AI model/system exploitation – Adversarial actions that exploit vulnerabilities an AI model or system to force misperformance against its intended objectives, disrupt access to its outputs or functionality, or enable unauthorized access to restricted or proprietary information.[5]

---

[1] Adapted from, https://doi.org/10.6028/NIST.AI.100-2e2025 and https://www.nccoe.nist.gov/ai/adversarial-machine-learning

[2] Adapted from, https://doi.org/10.1016/j.array.2025.100399 and https://doi.org/10.3390/fi17090404

[3] Adapted from, https://doi.org/10.3390/fi17090404

[4] Adapted from ISO/IEC 12792:2025 and https://doi.org/10.6028/NIST.AI.100-1

[5] Adapted from, https://doi.org/10.6028/NIST.AI.100-2e2025

AI governance –The set of organizational policies, rules, frameworks, roles, and oversight processes that direct how AI is adopted, developed, deployed, and monitored within the organization, with the objective of ensuring AI-related risks are identified, managed, and monitored across the AI lifecycle.[6]

AI lifecycle – The set of phases an AI system goes through. These are plan and design, collect and process data, build and use model, verify and validate, deploy and use, and operate and monitor. These phases are often iterative, and not necessarily sequential.[7]

AI model – A component of an information system that implements AI technology and uses computational, statistical, or machine-learning techniques to produce outputs from a given set of inputs.[8]

AI risk assessment – A risk-management process for identifying, estimating, and prioritizing risks arising from the operation and use of an AI system, incorporating threat and vulnerability analyses and considering mitigations provided by controls planned or in place.[9]

AI as a service (AIaaS) – Cloud-based systems providing on demand services to organizations and individuals to deploy, develop, train, and manage AI models.[10]

AI system – The term 'artificial intelligence system'

(A) means any data system, software, application, tool, or utility that operates in whole or in part using dynamic or static machine learning algorithms or other forms of artificial intelligence, whether

(i) the data system, software, application, tool, or utility is established primarily for the purpose of researching, developing, or implementing artificial intelligence technology; or

(ii) artificial intelligence capability is integrated into another system or agency business process, operational activity, or technology system; and

(B) does not include any common commercial product within which artificial intelligence is embedded, such as a word processor or map navigation system.[11]

AI use case – A specific scenario in which AI is designed, developed, procured, or used to achieve a particular objective, such as delivering a product or service, enhancing decision making, or providing a defined benefit.[12]

AI use case inventory – A maintained repository or listing of an organization's AI use cases, intended to support governance, transparency, and risk management by documenting where and how AI is designed, developed, procured, or used, and the purpose and outputs associated with those uses.[13]

---

[6] Adapted from, https://doi.org/10.6028/NIST.AI.100-1 and Governance of AI adoption in central banks
[7] Adapted from, https://doi.org/10.6028/NIST.AI.100-1 and OECD Framework for the Classification of AI systems (EN)
[8] https://doi.org/10.6028/NIST.SP.800-218A
[9] Adapted from risk assessment - Glossary | CSRC and https://doi.org/10.6028/NIST.AI.100-1
[10] https://doi.org/10.1007/s12599-021-00708-w
[11] 40 USC SUBTITLE III, CHAPTER 113, SUBCHAPTER I: DIRECTOR OF OFFICE OF MANAGEMENT AND BUDGET
[12] Adapted from, 2024-Guidance-for-AI-Use-Case-Inventories.pdf
[13] Adapted from, 2024-Guidance-for-AI-Use-Case-Inventories.pdf and Department of Justice | AI Inventory

Algorithm – A clearly specified mathematical process for computation; a set of rules that, if followed, will give a prescribed result.[14]

Algorithmic trading system – A system that fundamentally depends upon computerized algorithms, and the data and technological infrastructure through which they operate, to address various decisions and tasks associated with trading financial instruments.[15]

Anomaly detection system – A system for identifying the occurrence of a condition that deviates from expectations based on requirements specifications, design documents, user documents, or standards, or from someone's perceptions or experiences.[16]

Artificial general intelligence (AGI) – The currently hypothetical level of AI capability that is able to understand or learn an intellectual task as human being can. It is an AI system that can perform across diverse cognitive domains with versatility and proficiency, rather than being limited to a narrow task or domain.[17]

Artificial intelligence (AI) – 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.[18]

Benchmarking – An alternative prediction or approach used to compare a model's inputs and outputs to estimates from alternative internal or external data or models.[19]

Bias – A systematic distortion, as opposed to random error, that reduces the representativeness or accuracy of an AI system's outputs or performance for its intended purposes and operating conditions. Bias may be introduced inadvertently or purposely, and may also emerge as the AI system is used in an application; this could arise when the data used to develop or operate the system are not representative of the intended population or operating conditions. Common sources/subcategories or bias include statistical/computational, systemic, and human bias (not exhaustive).[20]

Black box – The nature of some AI techniques whereby the inferential operations are complex, hidden, or otherwise opaque to their developers and end users in terms of providing an understanding of how classifications, recommendations, or actions are generated and what overall performance will be.[21]

Capability evaluation – A comprehensive assessment of an AI model's or system's overall capabilities, including both planned capabilities and unplanned, emerging, or malicious capabilities. Unlike specific task-focused evaluations this evaluation seeks to understand the full range of an AI's capabilities. This includes evaluating how an AI might adapt or evolve beyond its initial training, identifying both

---

[14] https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-107r1.pdf
[15] https://www.sec.gov/marketstructure/research/hft_lit_review_march_2014.pdf
[16] Adapted from NIST SP 800-160v1r1 and NIST SP 800-37 Rev. 2
[17] Adapted from https://doi.org/10.48550/arXiv.2510.18212 and DOI:10.2478/jagi-2014-0001
[18] 15 USC 9401: Definitions
[19] Model Risk Management, Comptroller's Handbook
[20] Adapted from, https://doi.org/10.6028/NIST.SP.1270
[21] nscai_appendix_a_technical_glossary.9d423270fab8.pdf

beneficial emergent behaviors and potential risks that could arise from its autonomous operation or interaction with complex environments.[22]

Computer vision – The digital process of perceiving and learning visual tasks in order to interpret and understand the world through cameras and sensors.[23]

Data lineage – The history of processing of a data element, which may include point-to-point data flows and the data actions performed upon the data element.[24]

Data poisoning – An attack that corrupts and contaminates training data to compromise an AL system's performance.[25]

Data quality/validity – The usefulness, accuracy, and correctness of data for its application.[26]

Deep learning – A machine learning implementation technique that uses large quantities of data, or feedback from interactions with a simulation or an environment, as training sets for a network with multiple hidden layers, called a deep neural network, often employing an iterative optimization technique called gradient descent, to tune large numbers of parameters that describe weights given to connections among units.[27]

Deepfake – AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful.[28]

Deterministic (algorithm / model) – An algorithm/model that, given the same inputs, always produces the same outputs.[29]

Diffusion models – A type of generative AI model that produces output to match a prompt by iteratively refining noise. These types of models require substantial computational resources and processing time.[30]

Documentation – The collection of records that describe an AI system's purpose and intended use, key design choices, training and operational data characteristics and provenance, testing and evaluation results, limitations, and version history, maintained to support transparency, oversight, and accountability across the AI lifecycle.[31]

---

[22] Adapted from, https://arxiv.org/abs/2506.18213 [23] nscai_appendix_a_technical_glossary.9d423270fab8.pdf
[23] nscai_appendix_a_technical_glossary.9d423270fab8.pdf
[24] https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.01162020.pdf
[25] Adapted from, https://www.bis.org/fsi/publ/insights63.pdf
[26] NIST Big Data Interoperability Framework: Volume 4, Security and Privacy
[27] Adapted from, https://www.fsb.org/uploads/P011117.pdf
[28] EU Artificial Intelligence Act (Regulation (EU) 2024/1689), Article 3(60)
[29] Adapted from, https://csrc.nist.gov/glossary/term/deterministic_algorithm
[30] Adapted from, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-4.pdf, p.8
[31] Adapted from, AI System Documentation | National Telecommunications and Information Administration, GAO-21-519SP, ARTIFICIAL INTELLIGENCE: An Accountability Framework for Federal Agencies and Other Entities, and Model Risk Management, Comptroller's Handbook

Explainability – Property of an AI system that enables a given human audience to comprehend the reasons for the system's behavior; the ability to understand an AI system's output and decision given certain inputs.[32]

Federated learning – A method of training AI models across multiple devices or organizations without sharing underlying data. This machine learning architecture helps preserve privacy while enabling collaborative machine learning.[33]

Foundation models – Large machine learning models trained on vast amounts of raw and unlabeled data through unsupervised learning that can be adapted and applied to versatile downstream tasks. Large language models are common subsets of foundation models and underpin many generative AI applications in the financial sector.[34]

General purpose AI – AI designed for use across a broad array of tasks across many different applications rather than for a specific domain.[35]

Generative Adversarial Networks (GANs) – A machine learning framework in which two neural networks contest with each other in the form of a zero-sum game, where one agent's gain is another agent's loss. A GAN learns to generate new data with the same statistics as the training set.[36]

Generative AI – The class of AI that emulate the structure and characteristics of input data in order to generate derived synthetic content. This can include images, videos, audio, text, and other digital content.[37]

Guardrails – Layered safeguards to prevent access to bad information and behavior in an AI system. These may encompass policies, technical controls, and monitoring mechanisms, and may exist at the data, model, application, and infrastructure levels. These safeguards aim to ensure generative AI systems behave ethically, safely, and within organizational or regulatory boundaries by filtering training data, aligning model behavior, and enforcing post-deployment controls.[38]

Hallucination – A phenomenon when AI produces output that is erroneous or flawed but is still in the form of a convincing narrative or presentation. Generative AI can still produce flawed information even if underlying data is free of defects.[39]

Human biases – These biases reflect systematic errors in human thought based on a limited number of heuristic principles and predicting values to simpler judgmental operations…These biases are

---

[32] Adapted from, Artificial-Intelligence-in-Financial-Services.pdf and ISO/IEC TS 6254:2025(en), Information technology — Artificial intelligence — Objectives and approaches for explainability and interpretability of machine learning (ML) models and artificial intelligence (AI) systems

[33] Adapted from, https://www.nist.gov/blogs/cybersecurity-insights/protecting-trained-models-privacy-preserving-federated-learning and https://www.nist.gov/blogs/cybersecurity-insights/protecting-trained-models-privacy-preserving-federated-learning

[34] Adapted from, foundation model - Glossary | CSRC and https://www.bis.org/fsi/publ/insights63.pdf

[35] Adapted from, https://www.bis.org/fsi/publ/insights63.pdf

[36] https://doi.org/10.6028/NIST.AI.100-2e2025

[37] USCODE-2023-title15-chap119-sec9401.pdf

[38] Adapted from, https://doi.org/10.48550/arXiv.2512.10100

[39] Adapted from, https://home.treasury.gov/system/files/261/FSOC2023AnnualReport.pdf and https://home.treasury.gov/system/files/261/FSOC2024AnnualReport.pdf

omnipresent in the institutional, group, and individual decision-making processes across the AI lifecycle, and in the use of AI applications once deployed.[40]

Human-in-the-Loop (HITL) – A risk-control approach for AI where a human is integrated within the AI's decision-making process.[41]

Interpretability – Transparency into the inner workings of AI output in the context of their designed functional purposes, which helps users gain deeper insights into the functionality and trustworthiness of the system and its outputs.[42]

Large language model – A subset of machine learning that uses algorithms trained on large amounts of data through self-supervised machine learning to recognize patterns and respond to user requests in natural language.[43]

Machine learning – An AI learning method that enables computational systems to learn patterns, make predictions, and optimize decisions from large amounts of data without being explicitly programmed for each task. Machine learning encompasses supervised, unsupervised, and reinforcement learning paradigms, serving as the technical foundation for data-driven intelligence and automation.[44]

Model integrity – The process of protecting a model against improper information modification or destruction and ensuring information non-repudiation and authenticity.[45]

Model risk – The potential for adverse consequences from decisions based on incorrect or misused model outputs and reports. Model risk can be from individual models and be in the aggregate. Aggregate model risk is affected by interaction and dependencies among models; reliance on common assumptions, data, or methodologies; and any other factors that could adversely affect several models and their outputs.[46]

Multi-modal model – A model that processes and relates information from multiple data modalities, such as text, images, audio, and sensor data, among others.[47]

Natural language processing – The ability of a machine to process, analyze, and mimic human language, either spoken or written.[48]**Output validation** – Systematic process of verifying and confirming that AI system outputs meet specified requirements, accuracy standards, and quality criteria before being used for downstream processes.[49]

Override – Output or input that is ignored, altered, rejected, or reversed.[50]

---

[40] Adapted from, https://doi.org/10.6028/NIST.SP.1270
[41] Adapted from, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf
[42] Adapted from, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf
[43] https://home.treasury.gov/system/files/261/FSOC2024AnnualReport.pdf
[44] Adapted from, https://www.fsb.org/uploads/P011117.pdf and https://home.treasury.gov/system/files/261/FSOC2023AnnualReport.pdf
[45] Adapted from, https://www.nccoe.nist.gov/publication/1800-25/VolA/index.html
[46] Model Risk Management, Comptroller's Handbook
[47] Adapted from, https://doi.org/10.6028/NIST.AI.100-2e2025
[48] nscai_appendix_a_technical_glossary.9d423270fab8.pdf
[49] Adapted from, AI Test, Evaluation, Validation and Verification (TEVV) | NIST and Model Risk Management, Comptroller's Handbook
[50] Model Risk Management, Comptroller's Handbook

Performance monitoring – Ongoing activities that confirm an AI system is implemented appropriately, used as intended, and continues to perform as intended over time.[51]

Performance threshold – A particular value or range of values of a performance measure or diagnostic that determines the acceptance or rejection of a model's performance.[52]

Predictive analytics – A discipline within AI that leverages historical data, statistical algorithms, and machine learning techniques to identity patterns and forecast future outcomes, behaviors, or events. This discipline is distinguished by emphasis on forward-looking insights rather than descriptive analysis.[53]

Prompt – Natural language text describing the task that an AI should perform.[54]

Prompt injection – An attack on an AI system that exploits how an application combines untrusted input with a prompt written by a higher-trust party, such as the application designer, so the system follows the untrusted instructions.[55]

Reinforcement learning – A type of machine learning in which a model learns to optimize its behavior according to a reward function by interacting with and receiving feedback from an environment.[56]

Representation learning – Also known as feature learning, a set of techniques for automatically detecting feature patterns, replaces manual feature engineering.[57]

Responsible AI – Conscientious design, deployment, and governance of AI systems aligned with ethical principles, societal values, and legal requirements.[58]

Retrieval augmented generation (RAG) – A type of generative AI system in which a model is paired with a separate information retrieval system (or "knowledge base"). Based on a user query, the RAG system identifies relevant information within the knowledge base and provides it to the generative AI model in context for the model to use in formu-lating its response. RAG systems allow the internal knowledge of a generative AI model to be modified without the need for retraining.[59]

Service level agreement (SLA) – Contractually binding terms, often incorporated into a broader services contract, between a service provider and a customer that specify the services to be delivered and the measurable performance and service-quality commitments, such as availability and response times. SLAs also typically define each party's responsibilities and provisions for monitoring/reporting, issue resolution, and remedies if service levels are not met.[60]

---

[51] Adapted from, https://doi.org/10.6028/NIST.AI.100-1 and Model Risk Management, Comptroller's Handbook
[52] Model Risk Management, Comptroller's Handbook
[53] Adapted from, https://doi.org/10.6028/NIST.SP.1270
[54] Managing Artificial Intelligence-Specific Cybersecurity Risks in the Financial Services Sector
[55] Adapted from, https://doi.org/10.6028/NIST.AI.100-2e2025
[56] https://doi.org/10.6028/NIST.AI.100-2e2025
[57] Adapted from, ai-rmf-rfi-0010-attachment3.pdf and https://doi.org/10.48550/arXiv.1206.5538
[58]Adapted from, https://doi.org/10.6028/NIST.AI.100-1
[59] retrieval-augmented generation - Glossary | CSRC
[60]Adapted from, Bank Technology Bulletin: Tools to Manage Technology Providers' Performance Risk: Service Level Agreements

Service Provider Concentration (Financial Institution) – The extent to which a financial institution relies on a service provider, directly or indirectly, to support the financial institution's activities, particularly critical activities.[61]

Service Provider Concentration (Financial sector) – The extent to which financial institutions rely on a service provider, directly or indirectly, to support financial institutions' activities, particularly critical activities.[62]

Service Provider Concentration Risk (Financial Institution) – The potential for disruption or degradation at a service provider(s) to threaten the ability of a financial institution to continue performing the financial institution's activities, particularly critical activities, or cause the financial institution to suffer significant adverse effects.[63]

Service Provider Concentration Risk (Financial Sector) – The potential for disruption or degradation at a service provider(s) to threaten the ability of financial institutions to continue performing their activities, particularly critical activities, or cause the financial institutions to suffer significant adverse effects, with the potential for systemic impact to the financial sector.[64]

Supervised learning – A process for training algorithms by example. The training data consists of inputs paired with the correct outputs. During training, the algorithm will search for patterns in the data that correlate with the desired outputs and learn to predict the correct output for newly presented input data over iterative training and model updates.[65]

Structured data – Data that is divided into standardized pieces that are identifiable and accessible by both humans and computers.[66]

Synthetic data – Data that has been generated using a purpose built mathematical model or algorithm, that is statistically realistic but artificial, that can be used for activities like model development and training.[67]

Synthetic identity – The use of a combination of real and fake personally identifiable information (PII) to fabricate a person or entity.[68]

Text / word embedding – A numerical vector representation of text that machine learning and artificial intelligence systems use to work with meaning in text, such as comparing similarity between pieces of text.[69]

Third-party AI risk – Risk that arises when an organization relies on another entity to develop, provide, host, operate, or support AI systems or key AI components such as models, data, and related infrastructure."[70]

---

[61] Shared-Cloud-Lexicon.pdf
[62] Shared-Cloud-Lexicon.pdf
[63] Shared-Cloud-Lexicon.pdf
[64] Shared-Cloud-Lexicon.pdf
[65] nscai_appendix_a_technical_glossary.9d423270fab8.pdf
[66] SEC.gov | About Structured Data
[67] Adapted from, https://www.fca.org.uk/publication/corporate/report-using-synthetic-data-in-financial-services.pdf and NIST SP 800-188 3pd (third public draft), De-Identifying Government Data Sets
[68] Financial Trend Analysis, January 2024
[69] Adapted from, DOE/ID-Number
[70] Adapted from, https://doi.org/10.6028/NIST.AI.100-1 and Interagency Guidance on Third-Party Relationships

Traditional AI – Traditional AI, also referred to as symbolic or rule-based AI, is a subset of AI that focuses on performing discreet, preset tasks using predetermined algorithms and rules. These AI applications are designed to excel in a single activity or a restricted set of tasks, such as playing chess, diagnosing diseases, or translating languages.[71]

Training data – A subset of input data samples used to train a machine learning model.[72]

Unstructured data – Data that does not have a predefined data model or is not organized in a predefined way. This may also include data that is more free form, such as multimedia files, images, sound files, or unstructured text. Unstructured data does not necessarily follow any format or hierarchical sequence, nor does it follow any relational rules.[73]

Unsupervised learning – A learning strategy that consists in observing and analyzing different entities and determining that some of their subsets can be grouped into certain classes, without any correctness test being performed on acquired knowledge through feedback from external knowledge sources.[74]

Validation – Confirmation, through objective evidence, that an AI system or model meets requirements for a specific intended use or application and achieves its intended use in its intended operational environment.[75]

Version control: Systematic practice of tracking, managing, and documenting changes to AI assets through their development and deployment lifecycle.[76]

---

[71] Adapted from, https://www.uschamber.com/co/run/technology/traditional-ai-vs-generative-ai and https://doi.org/10.1038/d41586-025-03856-1

[72] ISO/IEC DIS 22989(en), Information technology — Artificial intelligence — Artificial intelligence concepts and terminology

[73] Adapted from https://doi.org/10.6028/NIST.SP.1500-1r2 and Glossary: Unstructured Data | resources.data.gov

[74] https://www.iso.org/obp/ui/es/#iso:std:iso-iec:2382:-31:ed-1:v1:en

[75] Adapted from, https://www.iso.org/standard/45481.html

[76] Adapted from, https://doi.org/10.6028/NIST.AI.100-1