

AI in Assessment Glossary

A Special Publication from ATP

Authored by the AI Subcommittee of the ATP Committee on Technology-Based Assessment

Updated August 2025

This glossary was created by the ATP Subcommittee on Artificial Intelligence (AI) in Assessment, which is a subcommittee of the ATP Committee for Technology-Based Assessment. The glossary provides a non-exhaustive list of terms related to AI to aid assessment professionals in their understanding of AI tools and systems.

The glossary complements the ITC/ATP Guidelines for Technology-Based Assessment. As such, many of the terms draw directly from the glossary in the TBA Guidelines. The addition of newer terms broadens the scope to include specific AI tools and concepts.

As the landscape and systems of AI remain constantly changing, the need to revise and update these terms is a certainty. Thus, ATP may revise and republish this glossary as needed.

The ATP wishes to thank Sean Gasperson of PSI who led the team who produced this glossary, working with Andriani Papapericleous (British Council) and Leslie Thomas (Kryterion). Reviewers included Andrew Hinkle (University of Minnesota), Jim Hussey (ACT), Zach Irwin (Kryterion), Ron Lancaster (ITS), and Pat Ward (ITS). The team worked under the leadership of AI Subcommittee co-chairs Kimberly Swygert of NBME and John Kleeman of Learnosity.

Glossary

*Definition from Guidelines for Technology-Based Assessment

**Definition from Organisation for Economic Co-Operation and Development

***Definition from EU AI Act

***Adaptive Instructional System:** An artificially intelligent, computer-based system that guides learning experiences by tailoring instruction and recommendations based on the goals, needs, preferences, and interests of each individual learner or team in the context of domain learning objectives. See **Computer Adaptive Test** and **Smart Content**.

Adversarial Attacks (or Adversarial Examples): Inputs intentionally crafted to cause AI models to make errors or to exfiltrate confidential or proprietary information.

***Algorithm:** A software program to handle a variety of computational tasks. Simple algorithms, more commonly referred to as “automation software”, merely automate traditionally manual tasks by following established rules or routines and are not usually considered to be Artificial Intelligence (AI) (e.g., automated scoring, some forms of automated item generation), whereas more complex algorithms usually involve the use of AI.

Algorithmic Bias (or AI Bias): Systematic and repeatable errors in AI outputs resulting from biased data, design assumptions, or model training processes.

****AI Actors:** Individuals who play an active role in the AI system lifecycle, including organizations and individuals that deploy or operate AI.

****AI Knowledge:** The skills and resources, such as data, code, algorithms, models, research, know-how, training programs, governance, processes, and best practices required to understand and participate in the AI system lifecycle, including managing risks.

*** AI Systems:** Software and/or hardware systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge or learning by processing the information derived from this data, and deciding the best action(s) to take to achieve the given goal. AI systems can be used in technology-based assessments to assist humans in making test administration and scoring decisions or to make automated decisions in place of humans.

****AI System Lifecycle:** The lifecycle typically involves several phases to: plan and design; collect and process data; build model(s) and/or adapt existing model(s) to specific tasks; test, evaluate, verify and validate; make available for use/deploy; operate and monitor; and retire/decommission. These phases often take place in an iterative manner and are not necessarily sequential. The decision to retire an AI system from operation may occur at any point during the operation and monitoring phase.

***Assessment Data:** Data collected from learner or examinee interactions and aggregation of data collected from learner or examinee interactions that contribute to the evidence base required to make an inference of attainment of knowledge, skills, and attributes of interest.

Augmented Intelligence: AI systems designed to assist rather than replace human decision-making.

***Automated Item Generation (AIG):** Templating, cloning, applying automated software, or in other settings, applying AI cognitive modeling processes, to generate a set of items from a single input or scenario created by an item writer. Some forms of AIG do not incorporate AI, while emerging methods have begun to utilize AI.

***Automated Scoring:** A technological method that involves response matching or text/natural language processing to review and evaluate text responses in a reproducible way that matches defined scoring rubrics and is in agreement with human raters. Many forms of automated scoring merely use non-AI software to perform a manual task; however, in some settings, AI-based scoring systems are used and are continuing to evolve.

Bias Mitigation: Methods or processes employed to identify, reduce, and monitor algorithmic bias in AI systems. Common strategies include adjusting training data, using fairness metrics, or implementing ongoing performance audits.

***Big Data:** A volume, variety, veracity, and velocity of data that can be used by an AI system to train on, learn from, or reason against.

Chatbot: A program designed to simulate conversations with human users.

Cognitive Diagnostic Models (CDM): Models that identify specific knowledge and skills of test-takers based on their responses.

Computational Linguistics: The use of AI to process, understand, or synthesize language.

***Computer-Adaptive Test (CAT):** A form of automated testing where the test taker receives successive items, or sets of items, which are selected in relation to the test taker's responses to previous items, in consideration of psychometric and content information. See **Adaptive Instructional System** and **Smart Content**.

Contextual Bandits: A type of reinforcement learning used in AI to personalize assessments by selecting optimal questions based on real-time performance.

Controller: (See **Data Controller**)

Convolutional Neural Networks (CNN): A network architecture that uses deep learning techniques to learn directly from data. CNN can be used for image recognition and reclassification.

Data Augmentation: AI techniques to increase the size and diversity of a dataset, which can enhance the training of machine learning models.

***Data Controller (or Controller):** An organization that, alone or jointly with others, determines the purposes and means of the processing of personal data. Typical examples of data controllers might be bodies that set certification exams, employers who test personnel or job candidates, or

educational institutions testing students to make admission decisions or to score tests on course work.

***Data Forensics:** In the field of assessment, data forensics pertains to the application of statistical methods to detect anomalies in test taker response patterns and test data to identify potentially serious test irregularities (e.g., cheating, proxy testing, content theft, and infringement of intellectual property rights).

***Data Governance:** The exercise of authority and control (planning, monitoring, and enforcement) over the management of data assets. Policies and best practices that ensure data is managed properly.

***Data Lake:** A system that acquires data from multiple sources in an enterprise in its original form and may also have internal, modeled forms of this same data for various purposes. The data may be any type of information, ranging from structured to completely unstructured data. A Data Lake is expected to be able to derive relevant meanings and insights from scored information using various analysis and machine learning algorithms.

Data Mining: Identifying useful patterns from large datasets.

***Data Processor (or Processor):** An organization that processes personal data on behalf of a controller. Typical examples of processors might be services companies that provide assessment or analytic services, test publishers that provide tests for an employer to use, or proctoring companies.

Deep Learning: Advanced machine learning involving multi-layered neural networks.

Deepfake: An AI-generated image, video, or voice meant to appear authentic, often used for deception or impersonation.

Dynamic Difficulty Adjustment (DDA): AI systems that adjust the difficulty of test questions in real time (dynamically) based on test-taker performance.

Educational Data Mining (EDM): The use of AI to analyze educational data to improve learning outcomes and assessment effectiveness.

Ethical AI: A framework of principles, policies, and practices intended to ensure AI systems operate fairly, transparently, and reduced bias.

Explainable AI (XAI): An attribute of AI systems that are designed in a way that make their decision-making processes easily explainable, sometimes referred to as their “explainability.”

Feedback Loop: In a machine learning (ML) system, the incorporation of user responses to continuously improve tools and predictions.

Gamification: The use of game-like elements in assessments.

*****General Purpose AI Model:** 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.

*****General Purpose AI System:** An AI system which is based on a general-purpose AI model, that has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems.

Generative AI: AI systems that generate new content, ranging from text to images to computer code, based on input data or learning objectives.

Generative Pre-trained Transformer (GPT): A type of AI model used to generate natural language responses or other content.

Graph Neural Networks (GNN): AI models that accept and analyze data presented in the form of graphs.

Guardrails: Mechanisms, procedures, or guidelines designed to ensure AI systems operate within desired boundaries.

Hallucination: Inaccurate information that is generated by AI-powered systems that is presented as fact.

Human-in-the-loop (HITL): AI systems that incorporate and empower human oversight or intervention.

Hybrid AI: Combining multiple AI approaches.

Implicit Feedback: Data collected by AI systems without direct input from users.

***Interoperability:** The ability of systems or software to exchange and make use of information.

***InterpretML:** An open-source library of machine learning explainers and interpretability techniques for explaining an AI model in an interpretable manner.

Jailbreaking: The process of bypassing restrictions in AI tools or systems.

Knowledge Tracing: Modeling a test taker's knowledge over time using AI to predict future performance and adapt learning approaches and resources accordingly. See **Learning Analytics**.

Large Language Model (LLM): A program trained on large amounts of data that processes information and generates human-like responses as text or voice.

Latent Semantic Analysis (LSA): A technique to analyze relationships between words and concepts in texts, often used in automated essay scoring.

Learning Analytics: The use of AI to analyze learning behaviors and outcomes to inform instructional and assessment decisions. See **Knowledge Tracing**.

***Machine Learning (ML):** A form of artificial intelligence that makes predictions from data. ML entails the use and development of computer systems that are able to learn and adapt without following explicit instructions by using algorithms and statistical models to analyze and draw inferences from patterns in data.

***Metadata:** Data about data. More formally, characterization of the structure, content, and quality of data, including source and lineage and the definition and intended uses of entities and data elements.

Model (AI Model): A program trained on sets of data to autonomously operate by learning patterns and making decisions.

Model Drift: The degradation in the performance of AI models over time due to external disruptions or failures to update source data.

***Natural Language Processing:** A branch of artificial intelligence, linguistics and computer science in which computer software is used to analyze and “understand” written and spoken human language.

Natural User Interface (NUI): User interfaces that interact with human gestures, speech, or touch.

Neural Networks: Machine learning algorithms inspired by and modeled after the human brain.

Neurofeedback: AI systems that track brain activity during assessments or other activities to understand cognitive engagement and performance.

Non-Determinism: Allowing AI systems to produce various outputs from the same input in an AI system, rather than restricting the response to a single output.

***Personal Data (or Personal Information, or PI):** Any information relating to an identified or identifiable natural person (also sometimes referred to as a data subject or test taker). Among the examples found in various national privacy laws and regulations, PI would include the person’s name, address, IP address, national identification number/social security number, payment (card) information, and even some types of pseudonymized identifiers that are capable of being re-identified.

Personalization Engine: AI systems that recommend specific assessments or learning paths tailored to a test taker's abilities and goals.

***Processing:** Any operation performed on personal data, including but not limited to collection, recording, organization, structuring, storage, retrieval, using, transmitting, disseminating, or making the data available, as well as restricting, erasing, or destroying the data.

Proctoring AI: AI systems that monitor test takers during online exams to enforce test taker compliance to established testing rules, using tools such as facial recognition, screen monitoring, or eye tracking.

Prompt Engineering: The practice of designing and refining text prompts to elicit desired outputs from generative AI models (e.g., LLMs).

***Recommender System:** A type of AI machine learning system designed to leverage content and person-specific metadata to predict or provide personalized recommendations. In a consumer-oriented context, recommendations can be products or services, often relevant to online search-related behaviors. In a technology-based assessment context, the recommendations can be of items and learning content. The purpose is to leverage far more metadata than traditional

computer-adaptive assessments, leading to greater depth and diversity of reported information, ultimately extending performance results to instruction and learning.

Reinforcement Learning: A type of machine learning in which AI learns from interacting with its environment, and the positive or negative reinforcements that environment provides. Often used in adaptive testing and game-based learning assessments.

Robustness: The ability of AI systems to maintain performance across different conditions or data variations in assessments.

Sentiment Detection: AI that identifies positive or negative sentiment often through voice or facial recognition data.

Smart Content: AI-generated educational materials, including assessment items, that adapt to test taker needs. See **Adaptive Instructional System** and **Computer Adaptive Test**.

Speech Recognition: AI technology that processes spoken language, Examples of use in assessment include oral assessments and language proficiency tests.

SVM (Support Vector Machine): A type of machine learning model used in AI systems to classify data.

Synthetic Data: Artificially generated data that imitates the statistical properties of real datasets.

Training (of an AI Model): The process of teaching an AI model to operate by adjusting parameters, inputs, or prompts.

Transfer Learning: A machine learning technique in which a model developed for one task is adapted to another.

Transparency: The degree to which the details of an AI system, such as decision rules and models, are made available to stakeholders.

Voice Biometrics: AI used to identify people based on their voice patterns. In assessment, it can be used to authenticate test takers.

Wearable Technology: Devices such as smartwatches, headsets, or eyeglasses that can be used to collect and measure data about the user, including behaviors and performance, during assessments.

Zero-Shot Learning: An AI approach through which a model can solve tasks for which it has not been explicitly trained.

LEGAL NOTE

This document is published by the Association of Test Publishers, 601 Pennsylvania Ave., N.W. South Building, Suite 900 Washington, DC 2000.

Definitions which are in the Guidelines for Technology-Based Assessment (i.e. those marked with a single asterisk) are included in this document with permission from the International Test Commission and are copyright ©2022 by the International Test Commission and the Association of Test Publishers. All rights reserved.

Definitions that are not marked with an asterisk are new for this document and are copyright © Association of Test Publishers 2025. All rights reserved. Copyright is not claimed in the definitions with double asterisks.

The document should be regarded as providing general information about AI but does not contain legal advice for any individual organization's specific circumstances. It is written in good faith and does not represent ATP policy, nor does it represent the views of the member organizations whose employees have contributed to it. This document does not constitute legal advice; readers should consult their legal experts for legal advice on the matters addressed in this document.