

# **A Primer on Generative AI in Assessment**

## ***A Special Publication from ATP***



*Image generated with Bing Copilot*

Authored by the International Privacy Subcommittee of the ATP Security Committee

January 2025

**Contents**

**Executive Summary ..... 3**

**Introduction ..... 4**

**Understanding AI and GenAI..... 5**

**Testing-Specific Considerations for Use of GenAI ..... 8**

    Test Design and Development: ..... 8

    Eligibility, Registration and Accommodations ..... 8

    Assessment Delivery ..... 8

    Scoring ..... 9

    Reporting ..... 9

    Customer Service and Tech Support ..... 9

    Business operations, sales and marketing .....10

**Common Concerns with GenAI and Considerations for Testing Programs .....10**

**What is the future? .....13**

**Conclusion.....14**

**Acknowledgements.....14**

## **LEGAL NOTE**

This document is published by the Association of Test Publishers. This document is copyright © Association of Test Publishers 2025. 601 Pennsylvania Ave., N.W. South Building, Suite 900 Washington, DC 2000. All rights reserved. Copyright is not claimed on the Copilot-generated image on the cover page.

The document should be regarded as providing general information about AI but not contain legal advice for any individual organization's specific circumstances. It is written in good faith but 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 own legal experts for legal advice on the matters addressed in this document.

## **Executive Summary**

This Primer introduces and discusses Generative AI and its potential applications in the testing industry. Generative AI, a subfield of AI based on the use of large language model AI (“LLMs”), focuses on creating new content based on learned patterns from existing data. It holds broad potential for application in the testing industry, including for test design and development, test taker eligibility, registration, and identification, and test delivery, scoring, reporting, and customer service. However, such systems equally pose risks to ethics, privacy, safety, security, and human autonomy. This Primer explains how Generative AI works and how testing organizations may choose to interact with such systems. It highlights both the potential benefits (such as increased efficiency and personalization) and technical risks (such as hallucinations and bias) associated with the use of Generative AI tools, as well as risks of operational challenges and protection of test integrity.

Looking to the future, this Primer suggests that Generative AI has the potential to revolutionize the testing industry by enabling more personalized learning and testing experiences, streamlining operations, and even redefining traditional assessment paradigms. It also stresses the need for robust legal and ethical frameworks to guide the responsible use of AI in all forms of testing, from employment to education, to certification/training.

## **Introduction**

This Primer introduces Generative AI<sup>1</sup> within the testing and assessment industry. The goal of this Primer is to provide readers looking for an introduction to the subject with an understanding of Generative AI technology and its uses in the industry. Two other Primers are planned in this series: one to address legal aspects of AI and Generative AI and the other to address security and governance practices.<sup>2</sup>

Artificial intelligence (AI) has been used in various forms for decades, including within the testing industry. Generative AI, however, is a subset of AI that recently became publicly available for commercial use in 2022 with the launch of ChatGPT followed by other similar LLM products. Generative AI has seen rapid advancements since 2022, and today, a wide number of different Generative AI systems have become available to individuals and organizations.

---

<sup>1</sup> Other ATP documents may refer to variants of this name, and some legislation uses different terms, for example “General Purpose AI” or “AI System.”

<sup>2</sup> ATP has published several other documents on AI and assessment including the 2024 “Creating Responsible and Ethical AI Policies for Assessment Organizations”, the 2022 [Artificial Intelligence Principles](#), and the 2021 [Artificial Intelligence and the Testing Industry: A Primer](#).

For the testing industry, Generative AI presents transformative possibilities as well as serious challenges and risks. Set forth below is a Primer on Generative AI, potential applications for professionals working with assessments, and concerns and considerations that should be evaluated prior to implementation of such technology.

## **Understanding AI and GenAI**

Artificial Intelligence is the capacity of computers to exhibit or simulate human intelligence.<sup>3</sup> According to the Organization for Economic Cooperation and Development (“OECD”), 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.”<sup>4</sup>

“Generative AI” (Gen AI) is a subfield of AI that focuses on creating new content, such as text, images, software source code, audio, or video, based on learned patterns and rules from existing data. This Primer focuses on Generative AI.

### **How Generative AI Works**

The most widely used Generative AI models generate text using systems called large language models. Some Generative AI models are “multimodal”, meaning they can receive input in a number of different modes and generate output in a number of different modes including some or all of text, graphics, audio, music, videos, or other media. There are large numbers of publicly available Generative AI systems including ChatGPT, Claude, DALL-E, Gemini, Llama, Midjourney, Mistral, and many others. Some are available for basic use at no charge while others require subscription.

Publicly-available Generative AI models are typically trained on a significant amount of data, often taken from the Internet. Training data may be comprised of data from books, articles, computer code repositories, websites, libraries of images and videos. The training and computing process for these models involves specialist skillsets, specific equipment, and vast amounts of data.

In addition to publicly-available Generative AI models, two other models may be created. A closed Generative AI model may be created by an enterprise with significant amounts of private data and typically focuses on a particular domain. These closed models are often, but not always, trained on data from only a single owner to protect the user’s content and limit risks of infringement. A “walled garden” Generative AI model is a hybrid; it is typically built within a publicly-available

---

<sup>3</sup> See Oxford English Dictionary definition of artificial intelligence at [www.oed.com](http://www.oed.com), accessed 26 November 2024.

<sup>4</sup> OECD (2024), “Explanatory memorandum on the updated OECD definition of an AI system”, *OECD Artificial Intelligence Papers*, No. 8, OECD Publishing, Paris, <https://doi.org/10.1787/623da898-en>.

model but contractually and operationally the model is limited as to how data inputs and outputs are used for training. To some extent, any of these Generative AI models can be fine-tuned or augmented to make them more useful in particular domains by adding rules or providing extra content.

### How to Use Generative AI

The simplest way to use Generative AI is to give it a “prompt,” which is a question, text, or other material to stimulate the model. Prompts can be provided via plain language in either English or another non-English language and can also be provided using other inputs, such as images or sound. The Generative AI model produces an output in response to the prompt. The clarity and quality of the prompt are important to obtain relevant and accurate output. The user can typically continue to prompt the model by submitting additional input to further refine the output.

Communication with Generative AI can come from keyboard, audio, or image inputs. The output similarly can be textual, audio, static images, or video. When Generative AI is embedded in other programs, it is usually called via Application Program Interfaces (APIs) where a computer program interacts with the user. For example, an item authoring system might present a user interface asking the user what kind of question to create, and then use an API call to the Generative AI model to generate a possible item.

The outputs from Generative AI can be surprisingly capable and, in some cases, comparable to the output humans might write or create. The process for creating the output is different, however. While humans create output based on the biological cognition of the human brain, Generative AI uses statistical probabilities of patterns it has identified in its enormous training data set to respond to a prompt. Because the data sets are so large and the patterns may be impossible for a human to discern, it can be difficult to explain how or why the model responds in a particular way and with the specific resulting output.

### Beneficial Uses of Generative AI

The world is still working out the ways in which Generative AI can be used, but there is consensus that businesses and society at large will likely experience significant changes as a result of Generative AI. At a high level, the benefits of Generative AI could include increased efficiency, more personalization, better objectivity, idea recognition, and pattern recognition. Generative AI can also be used to adjust text or other content by applying the style of one piece of content to other material. Additionally, because it is trained on normal patterns in data, it can be good at finding anomalies.

There are thousands of use cases for Generative AI, such as:

- Generating text: writers, whether working in marketing or other sectors or areas, use Generative AI to help draft articles or to edit articles.
- Improving search: Search engines like Google and Bing use Generative AI to improve the results and informational quality of web searches.
- Generating code: Software developers use Generative AI to help generate code.
- Translation: Generative AI can be used to translate material from one language to another.
- Creating “art” and entertainment: There is significant activity in using Generative AI for artistic purposes, including in images, films and music.
- Education: There is also much activity in using Generative AI to help provide more personalized education and learning.
- Customer support: Organizations are experimenting with using Generative AI to provide chat support and other kinds of customer service and technical support.

### Cautions Concerning Generative AI

Because it produces material which is statistically likely, Generative AI can make mistakes that are often called “hallucinations.” For example, the model may get its facts wrong or make up something entirely new and inaccurate. At the present time, Generative AI models also tend to be weak at mathematics and numerical analysis. It is therefore critical for a human to check all output and verify the output independently, particularly if it is used in sensitive or important contexts.

In addition, because it is trained on material that often comes from the Internet, Generative AI models may generate output that reflects the same bias as contained in its training data. For example, there is a risk that it can perpetuate social biases against females or against some demographic groups. Because of the risks of hallucinations and bias, the most effective use cases for Generative AI are ones where it is used to create material and a human then reviews the output.

There is also a risk that Generative AI can be used for malicious purposes. For example, a user can submit a prompt to plan an illegal activity. In addition, Generative AI can be used to create “deepfake” media. Deepfakes are false representations of an individual generated using visual and audio inputs to the model. In those instances, Generative AI can be used to create synthetic images, animations, videos, or sound bites of individuals. Already, several deepfake instances have been documented of celebrity or political figures saying things they did not say or of non-consensual pornography.

## **Testing-Specific Considerations for Use of GenAI**

There are use cases for AI in all parts of the assessment lifecycle. This Primer highlights a few examples of use cases, but many more are emerging.<sup>5</sup>

### **Test Design and Development:**

- **Blueprint generation:** Generative AI can be used to aid the creation of blueprints. For example, it can summarize important points in other documents that can be used to identify potential domains for test questions.
- **Item writing:** Given a blueprint or other direction, Generative AI can write items that match a specification and are then reviewed and refined by expert authors. Because of the risks of hallucination and bias, it is important that automatically generated items undergo fairness and bias reviews.
- **Item tagging and management of item banks:** Generative AI can assign tags to items, which is helpful in the management of item banks.
- **Graphic and video generation:** Images and videos can provide a richer testing experience, but production and licensing costs may render them infeasible. Generative AI is able to produce images and videos efficiently and more affordably.
- **Chatbot item type:** Many Generative AI models are designed to function as chatbots and may be suited for chatbot question types. Both written and oral chatbots are possible, as multimodal Generative AI models integrate speech recognition, language models and text-to-speech.

### **Eligibility, Registration and Accommodations**

- **Data review:** Generative AI can facilitate the triaging of registration forms and accommodation requests by summarizing or making recommendations that are subsequently reviewed by humans.
- **Data completion:** By integrating AI-driven techniques such as predictive filling, NLP, error detection and correction suggestions, organizations can streamline their registration processes, ensuring they collect necessary, accurate, and complete data efficiently.
- **Initial approval:** Generative AI can be used in auto-approving some of these accommodation requests through smart workflows, document verification, integration with registration systems, and so on. Some form of human oversight is necessary.

### **Assessment Delivery**

- **Chatbot delivery:** Generative AI can give short, generic answers to questions a candidate might have regarding exam registration, scheduling, retesting attempt rules, documentation and similar issues, saving administrative time.

---

<sup>5</sup> In most of these use cases, human involvement will be required to help ensure accuracy and mitigate against bias, and to comply with responsible AI governance principles and applicable laws.



- Proctor assistance: Generative AI can be used to improve test takers' interaction with proctors, by translating instructions or even enriching them with policy information from a knowledge base.
- Language support: Generative AI can be used to translate instructions for test takers, making it easier for test takers who are not fluent in the main language used.
- GenAI prompting / tool: Generative AI could be provided as a tool that test takers can use during the exam, as it is increasingly becoming a more integral part of working lives.
- Exam security: Generative AI can be used to help ensure the integrity of an exam, such as by identifying anomalies or by detecting plagiarism.
- Personalization and feedback for test takers: Generative AI can be used to create exams and experiences that are more adaptive and personalized to test takers.

### **Scoring**

- Transcripts of audio: Generative AI can be used to generate transcripts of audio for speaking-based items.
- Unstructured response scoring: Generative AI can be used for scoring open-ended questions, depending upon the stakes and nature of the question.

### **Reporting**

- Feedback for test takers: Generative AI can be used to analyze test takers' answers, and provide feedback on areas of improvement, explaining the correct answer.
- Feedback for educators / trainers: Generative AI can be used by educators to summarize feedback given to test takers or to highlight for educators areas for which additional instruction is needed.
- Predictive analysis: Generative AI uses predictive analytics to provide early feedback to test takers by analyzing data from past performances and learning behaviors. It can build predictive models to offer real-time and pre-test feedback, suggesting study focuses and potential performance outcomes.

### **Customer Service and Tech Support**

- Chatbots for immediate support: Policies, Knowledge Bases, and Frequently Asked Questions can be fed into Generative AI models for the purpose of customer support.
- Data monitoring and tracking to identify issues faster: Generative AI models can be used for a triage of customer support tickets, helping determine whether escalation to human agents is necessary or providing high level summaries of trends.
- Language support: Generative AI can be used to translate interactions with customer support, aiding those who speak different languages.
- System check: Generative AI can be used to enhance security and integrity of the process by checking software compatibility checks, system specification analysis, network and Wi-Fi speed test, pre-defined security protocol if any, and so on.

### **Business operations, sales and marketing**

- Content generation: Generative AI can be used to generate content for emails, blog posts, and articles. The generated content can be personalized at a high-volume scale.
- Idea generation: Generative AI can be used in ideation phases, as it is able to suggest several ideas from an initial prompt. It is useful to address writer's block.
- Marketing asset generation (graphics, etc.): Generative AI can produce marketing assets efficiently and affordably.
- Meeting transcriptions: Generative AI can generate transcripts for meetings automatically and summarize them.
- Communication channel search: Generative AI can search within the communication channels of an organization, and answer questions regarding projects, policies, or past decisions.
- Measuring sentiment and tone: Generative AI can be used to measure sentiment and tone of interactions with public internet forums, social media, etc., which may be useful for monitoring user or customer perceptions of an exam, test, or a testing organization.

As discussed further below, to protect against the risks of hallucinations and bias, human involvement is critical and necessary for many of these use cases.

### **Common Concerns with GenAI and Considerations for Testing Programs**

While Generative AI holds the potential to transform how individuals learn and work, those same uses often involve significant risks that require management. Any testing organization seeking to use Generative AI should be aware of the limits of the technology and prepared to incorporate appropriate processes and capabilities to mitigate those risks. Testing programs should also be aware that Generative AI can be used to undermine the integrity of the testing process and security of testing assets. In this section some of the risks and operational considerations of Generative AI are discussed.

**Output Concerns:** Generative AI outputs can be incredibly creative and thought-provoking; at the same time, this output can also raise significant concerns, including inaccurate, biased, and manipulative responses. Some common concerns are described below.

- Hallucinations: Hallucinations are false or misleading outputs created by Generative AI. The result is textual, visual, and other outputs that can be convincingly stated and well-crafted, yet which are not accurate. Numerous instances of AI hallucinations have been made public, from Generative AI producing fake images based on real events to Generative AI citing caselaw that never existed. Before relying on Generative AI output, human users should carefully review and fact check the information provided.

- **Bias:** Another significant concern with Generative AI is the extent to which the output can reflect societal biases and stereotypes. This problem may arise because the training data is biased, or the model's algorithms incorporate or amplify bias. For example, if an AI hiring model is trained using application data primarily from male workers, the AI may wrongly conclude that one of the key traits required in future hires is that the applicant is male. Even if a data set is initially unbiased, bias can creep into the system over time. Before incorporating an AI model, users should understand how the model was trained, and plan for human-in-the-loop review. There needs to be initial and frequent monitoring of the model in use, especially when significant upgrades are made.
- **Non-determinism:** Even when given the same input, Generative AI can produce different outputs. This variability – also referred to as non-determinism - makes it difficult to know what outputs to trust and can lead to confusion for users. Users should bear in mind that Generative AI produces probabilistic responses, but not necessarily accurate responses. Different Generative AI models will also usually produce different outputs.
- **Jailbreaking:** One common concern with GenAI in the testing industry is jailbreaking. Jailbreaking occurs when users attempt to bypass the AI's built-in safeguards, potentially leading to the generation of harmful or inappropriate content. In an assessment context, this could allow test takers to manipulate the AI into providing unauthorized assistance, generating dangerous or unethical instructions, or even producing biased or disparaging responses. Such vulnerabilities could compromise the integrity of exams, validity of scores, and fairness of the evaluation process, posing significant risks to the industry.
- **Emotional and Psychological considerations:** Generative AI outputs can be quite persuasive and convincing, even if those outputs are hallucinations or biased. Generative AI chatbots, for example, are designed to simulate conversation with another human and can result in a level of trust in the outputs that can be manipulated or abused. In some cases, Generative AI outputs can cause both emotional and psychological harm to users, especially those who are not aware of the limitations of the technologies.

**Operational Concerns:** Generative AI also raises operational concerns for entities seeking to incorporate it into their business activities. For example, it is important to review and understand the contractual terms of use and to understand how the model will use any data input into the model and any output generated by the users' activities. It is not uncommon for terms to provide that any data put into the model and any output will be retained and used to train the model. In the case of personal information and sensitive or private data, this may result in loss of control of such information and lack of compliance with applicable data protection laws and individual rights. There are also significant issues around intellectual property and copyright. These issues will be explored more in the upcoming Legal AI/GAI primer in this series. Users should determine in advance whether the model being used is an open model, closed model, or a hybrid.

- Open models can be used by anyone and trained on the data input into the model.
- Closed models tend to be proprietary and more limited in scope, with the data input being held under strict control.

- Hybrid models, which are often used by professional users and testing vendors, blend the ability to call upon open-source data as in the open model, but operate within constraints that can, among other things, limit whether the data will become training data and can provide more control over the output for fine-tuning.

Other operational concerns include the cost of implementation, as new tools can require significant changes to data collection and storage capabilities as well as platform acquisition, implementation, and maintenance costs. This, in turn, may raise staffing concerns. Use of Generative AI may require the organization to upskill or hire new talent. In addition, the concern of Generative AI replacing humans or making jobs less enjoyable may impact staff morale. At a minimum, organizations should anticipate the need for thoughtful change management tactics as they incorporate Generative AI into operational procedures. Further, Generative AI models require additional computing power, which raises concerns that the energy needed to power Generative AI increases the burden on the environment.

**Overview of Legal Considerations and Exam Integrity Concerns:** Generative AI is highlighting gaps and forcing changes in the legal and regulatory landscape. For example, the issue of intellectual property protection and ownership for Generative AI created output is unclear from a legal perspective. Another important consideration is that of data privacy and IT security. As noted earlier, Generative AI models were built using vast amounts of data. Some of this data might be personal data that was acquired without individual consent or other legal permission. As use of these models continues to grow, the risk of loss and misuse of personal data grows. Further, from an information security perspective, as more coding relies upon Generative AI produced code, it raises increased concerns about cyber security risk. To protect private and confidential information, organizations implementing Generative AI should put in place an AI Governance Framework tailored to their program. Such governance will better enable an organization to manage its risks and fulfill its obligations concerning personal data while still leveraging the benefits of Generative AI.<sup>6</sup>

Additionally, some jurisdictions are regulating Generative AI based on usage in certain sectors. For example, due to concerns of bias and discrimination, some jurisdictions have limited use of Generative AI (and other kinds of AI) in automated decision-making in education or in employment settings and also imposed higher standards on users of the tool and requirements for genuine human involvement. This can include requiring bias audits or enhanced notice to learners and applicants. Failure to comply with regulations may result in escalating fines that can become financially significant.<sup>7</sup>

---

<sup>6</sup> The Legal and Security/Governance primers will also cover AI Governance.

<sup>7</sup> The Legal primer will address EU, U.S. federal and state laws and regulations in this area.

Finally, as will be further discussed in this series, Generative AI can also pose a security risk to the integrity of the testing process. The unauthorized use of Generative AI can be used to compose short answers or essay responses. Further, Generative AI can be used to create image or voice deepfakes that can deceive observers and to engage in unauthorized behavior, such as use of a proxy test taker or to hide access to unauthorized materials. Testing programs should be aware of the risks that third parties using AI can pose to their testing programs and incorporate tools to deter and detect these types of efforts.

### **What is the future?**

Generative AI has potential to transform testing and the test taker experience, so long as its uses are balanced with reasonable efforts to manage the risks. From supporting innovations in how tests are created, delivered, and used, to enabling more efficient and secure test operations, delivery, and credential management, the potential uses of Generative AI models range across the activities of education and assessment organizations. By enabling further personalization of learning and assessment, Generative AI can in principle adapt in real-time to the needs and abilities of each individual being assessed, potentially offering a more customized and effective assessment journey. This could improve learner engagement and make education and assessment more personalized while also providing enhanced operational efficiencies and greater security during the assessment event. In short, these advancements have the potential to impact learners and workers across their lifetime. For education and training, these advancements may offer more personalized, accessible, and accurate assessments and improved learning. In the work force, Generative AI may lead to a dynamic evolution in how individuals are certified and credentials are managed and recognized across various sectors.

At the same time, however, integrating Generative AI poses significant risks and raises concerns regarding ethical and responsible use. We are only beginning to understand and research the impact of Generative AI on humans. The potential for hallucination and bias is well-documented and the damage Generative AI can do when providing inaccurate or biased output is significant. It is unclear at this time if these risks can be fully mitigated against. As such, considerable, meaningful, and appropriately-timed human oversight is warranted. Further, governmental regulation is expected to continue to grow and legal issues related to bias, ownership, infringement, data privacy, cybersecurity, and liability for harm will likely continue to be litigated. Because of these concerns, integrating Generative AI in the assessment industry necessitates robust industry frameworks to guide its safe and ethical use and protect individual personal data and rights. Internally, organizations will need to address issues related to data privacy, cybersecurity, the potential for bias in AI algorithms, and the accuracy of AI-generated recommendations and decisions. Establishing clear guidelines will be crucial for maintaining the integrity and reliability of testing processes and ensuring assessment outcomes are valid and defensible.

## **Conclusion**

Well-balanced uses of Generative AI provide significant opportunities for the testing industry. The possibilities are exciting, and the potential for increased productivity, improved accessibility, and enhanced pedagogical approaches appears boundless. That said this is still for the most part potential and not actuality. Many of these benefits are as yet unproven, and regulators and the public are likely to expect heightened responsibility for ensuring accuracy and fairness by organizations employing Generative AI models in the education or employment space. Appropriate care must be put into the development and integration of such technology. Concerns such as safety, bias, data privacy, cybersecurity, and responsible management should be addressed before organizations integrate Generative AI models into critical operations.

## **Acknowledgements**

The ATP wishes to thank (in alphabetical order) André Kenji Horie (Duolingo), John Kleeman (Learnosity), Trushant Mehta (OpenEyes Technologies), and Rachel Schoenig (Cornerstone Strategies) for writing this Primer under the leadership of the International Privacy Subcommittee Co-Chairs Donna McPartland (ConleyRose, P.C.) and Gary Behrens (FifthTheory, LLC).