

20
26

THE eAA GUIDE TO

AI IN ASSESSMENT

A practical introduction

Research findings ▪ real projects ▪ practical checks ▪ trusted guidance

eAA Research: AI in Assessment

AI use is growing, but confidence still lags behind adoption. AI is being used for specific tasks with people retaining responsibility for assessment decisions.

68.8%

are just starting out

Most organisations still need practical guidance, examples and clear safeguards.

44.4%

already use AI

A further 43.1% plan to use it, so evaluation skills are becoming urgent.

78%

have an AI policy

A policy helps only when staff also have workable processes, training and oversight.

68.8%

expect a supporting role

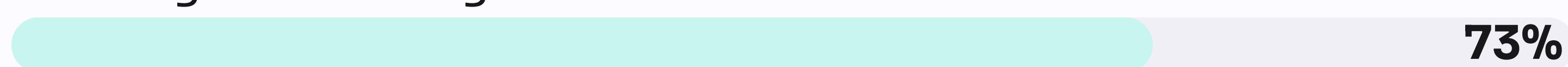
Most expect AI to assist selected assessment activities rather than control the whole process.

Where organisations are already using AI in Assessment

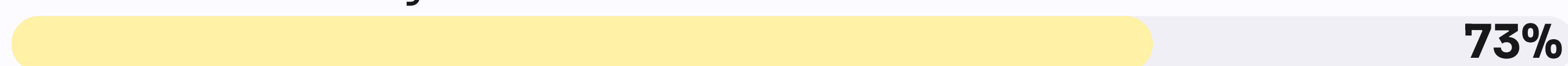
Item and question generation



Marking and scoring



Assessment analytics



Identity checking and proctoring



The practical message

Begin with a defined task where AI can assist. Keep assessment purpose, quality assurance and accountability with people.

Where AI can help

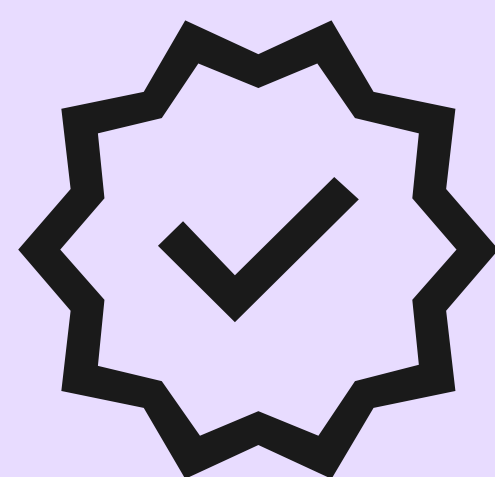
Each use has two sides: the task AI performs and the judgement that must stay with an educator, assessor or subject expert.

Assessment Activity	What AI does well	Role of the Educator / Assessor
Questions and tasks	Draft questions from an approved specification, suggest distractors or produce variants.	Set the learning outcomes, check accuracy and difficulty, approve every item and review performance after use.
Marking and feedback	Compare work with a rubric, identify evidence and turn marker notes into clear feedback.	Make or confirm the mark, review exceptions, correct unsuitable feedback and handle queries or appeals.
Assessment data	Find recurring errors, unusual patterns or groups that may need support.	Decide whether the pattern is meaningful, check alternative explanations and choose the response.
Simulations	Create responsive situations so learners can practise decisions and communication.	Define the behaviours to observe, ensure the scenario is realistic and judge whether it measures the intended skill.
Documents and speech	Transcribe handwriting or speech, translate content and help organise evidence for review.	Check transcription and translation, protect accessibility and privacy, and correct errors before any decision.

Start with the assessment problem. Then decide whether AI improves the evidence, workload or learner experience.

Case Studies: Creating Assessments

AI can produce a first draft; experts remain accountable



Professional Examinations

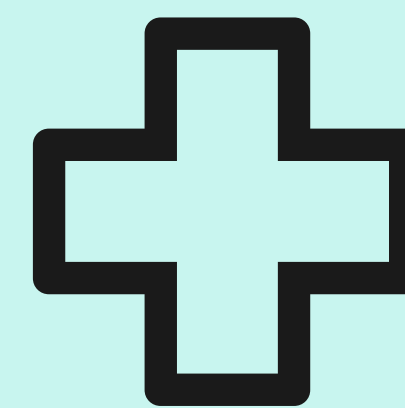
Janison and CA ANZ

CA ANZ uploads its learning materials and tells the system which topics, difficulty levels and question formats are needed. AI drafts questions from that controlled material. Accounting experts review every question before it can be used.

Key lesson

Use approved source material and make expert approval mandatory.

[Read the eAA case study →](#)



Healthcare Certification

Surpass and Inteleos

AI drafts questions from approved clinical guidance for clinicians to review. The pilot reported acceptance rates comparable with traditionally authored items, while clinicians retained responsibility for changing and approving every question.

Key lesson

Compare AI-assisted questions with your existing quality benchmark.

[Read the eAA case study →](#)



Open-response Assessment

BCS: generative AI for written answers

BCS explored whether generative AI could support assessment of open-response questions. This moves beyond multiple-choice testing, but requires clear rubrics, comparison with expert markers and a process for answers the system cannot judge confidently.

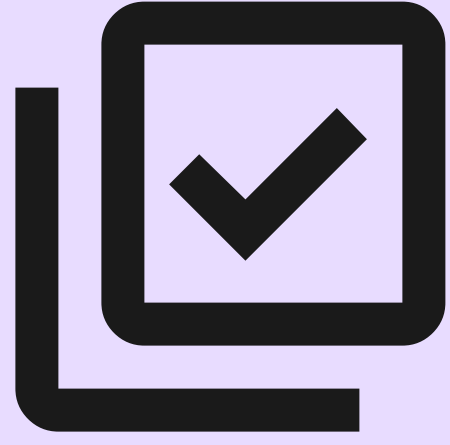
Key lesson

Test agreement with expert markers and define which responses must go to a person.

[Read the eAA case study →](#)

Case Studies: Marking and Feedback

AI can reduce repetitive work when people can review and correct it



University feedback

Henley Business School

Markers use a detailed digital rubric during more than 120 student presentations. AI turns their scores and notes into individual feedback. Markers review the wording, and the module leader samples and moderates before release.

Key lesson

Keep the mark decision separate from AI-assisted feedback writing.

[Read the eAA case study →](#)



School mock exams

Surpass Serenity

Students still write mock exams on paper. Schools scan the scripts; AI transcribes the handwriting and suggests marks and feedback. Teachers review every mark and can change it before pupils receive results.

Key lesson

Improve one difficult stage without redesigning everything.

[Read the eAA case study →](#)



Vocational assessment

sAlnaptic and Safety Training Awards

AI checks swimming-teacher lesson plans against detailed technical rubrics and drafts feedback. The project reports a large reduction in processing time, but tutors still review the output before it is used.

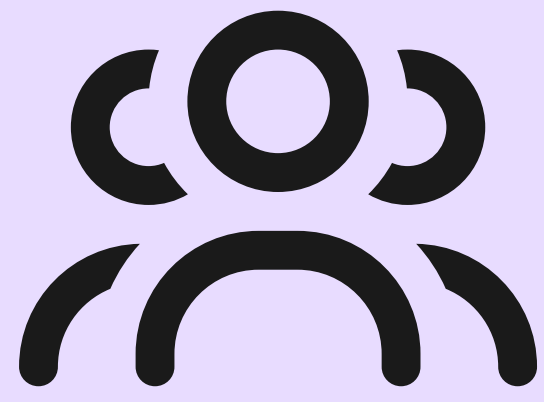
Key lesson

Measure checking, corrections and appeals - not just AI processing speed.

[Read the eAA case study →](#)

Case Studies: Practical and Interactive Assessment

Technology can show what learners can do – not only what they can write



Virtual Reality

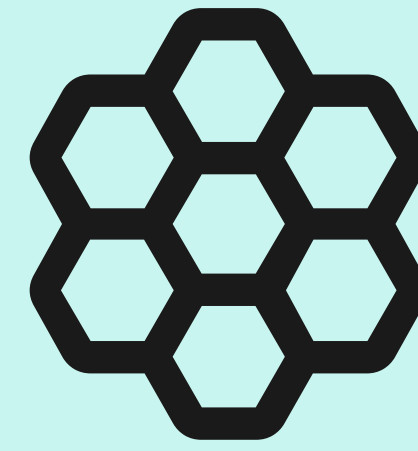
Sentira XR

Learners enter a VR clinical scenario and speak with an AI-powered virtual patient. The system records how they gather information, make decisions and communicate, then provides feedback without putting a real patient at risk.

Key lesson

Use simulation when the skill depends on decisions and behaviour in context.

[Read the eAA case study →](#)



Coding assessment inside Minecraft

Prodigy Learning: Coding in Minecraft

Pupils learn coding by completing activities inside Minecraft Education and take formative and summative assessments to prove their skills. The programme shows how practical work can be captured in a familiar digital environment.

Key lesson

Assess the work learners produce, not their enthusiasm for the game.

[Read the eAA case study →](#)



Knowledge, conversation and coding

EPAM's AI-supported skills checks

EPAM uses separate tools for a knowledge quiz, an oral conversation that probes understanding, and a practical coding task. Together they show what a learner knows, how well they explain it and whether they can apply it.

Key lesson

Choose an activity that matches the skill; do not rely on one score for everything.

[Read the eAA case study →](#)

Turn broad principles into practical checks

AI can help improve quality, but the evidence must show that it has done so. Use these checks before a pilot, during comparison with current practice and again before wider use.

Reliability

Will the same quality of work receive a consistent result?

How AI can help

AI may apply the same rubric repeatedly and flag uncertain responses for review.

Practical checks

Compare AI results with experienced markers. Repeat a sample. Investigate disagreements and check whether model updates change results.

Validity

Does the result reflect the knowledge or skill the assessment intends to measure?

How AI can help

AI can help analyse richer evidence, including speech, decisions in a simulation or steps in a practical task.

Practical checks

Check that the system rewards the intended learning outcome rather than writing style, accent, prompt wording or another irrelevant feature.

Fairness

Do different learners have an equal opportunity to show what they know and can do?

How AI can help

AI may support adjustments, transcription or alternative formats and help identify uneven outcomes.

Practical checks

Compare results and referral rates across relevant groups. Test accessibility with users. Investigate gaps rather than accepting average accuracy.

Transparency and Control

Can people understand where AI is used and challenge a result?

How AI can help

AI can retain records, highlight evidence behind a recommendation and route low-confidence cases to a person.

Practical checks

Tell learners what AI does. Keep versions and decisions. Name the accountable person and provide a usable correction, review and appeal route.

Use the evidence together. A system can be consistent but still measure the wrong thing or work poorly for some learners.

Conducting a pilot that lead to a decision, not simply a demonstration

Use real assessment work, compare it with current practice and agree the decision rules before the pilot begins.

1

Define the problem

Choose one clear problem, such as slow feedback or the time required to draft questions.

|

2

Set the baseline

Record current quality, staff time, cost, learner experience and outcomes for relevant groups.

|

3

Design the pilot

Use one subject, task or cohort. Keep the existing process and human decision route available.

|

4

Compare like with like

Run AI and current practice in parallel. Use the same rubric and examples already judged by experts.

|

5

Record what happens

Log corrections, disagreements, referrals, accessibility issues, staff time and learner feedback.

|

6

Apply the agreed test

Review the evidence against the measures set at the start. Document who made the decision and why.

Decision

GO

Evidence meets the agreed standard. Scale gradually and continue monitoring.

REVISE

The idea has value, but controls, training or performance need improvement.

NO GO

Quality, fairness, control or value is inadequate. Stop or choose another approach.

Questions to ask before choosing a supplier

Use these questions during procurement and again before moving from a pilot to live use. Ask for written evidence that relates to your assessments, learners and level of risk.

Performance Evidence

1. What assessment work has the system been tested on?
2. How does it compare with experienced human assessors?
3. What are the error rates, limitations and confidence thresholds?

Fairness and Accessibility

1. What results are available for different learner groups?
2. How has the product been tested with learners with different needs?
3. Which cases are more likely to be rejected or sent for human review?

Data and Security

1. Where is our data stored, and who can access it?
2. Will our content or learner data train any model?
3. How are retention, deletion, incidents and data export handled?

Control and Change

1. Can staff see, correct and override outputs?
2. How are model or product changes recorded and communicated?
3. What support, audit records and exit arrangements are included?

Using the Go/No Go Test

It decides whether the product and proposed use are safe, effective and worthwhile enough to move from a pilot into operational assessment.

GO only when evidence meets the agreed standard, staff can intervene, learners can challenge decisions, and responsibilities are clear.

NO GO when the supplier cannot provide relevant evidence, important groups perform poorly, staff cannot correct results, data use is unclear, or the benefit does not justify the risk and cost.

UK Guidance

DfE: AI in education and product safety

Ofqual: AI in qualifications and marking

Ofqual: Principles for AI use in marking

DfE: Generative AI product safety standards

Check current requirements for your jurisdiction, qualification type and use of personal data.

Continue your Digital Assessment journey

Join the global community championing the future of digital assessment



eAA Research and Case Studies

[AI in Assessment Research 2026](#)

[2026 Award winners and finalists](#)

[2025 Award winners and finalists](#)

[2024 Award winners and finalists](#)

[eAA Resource Hub](#)

Get in Touch



Email

operations@e-assessment.com



Visit Our Website

e-assessment.com

Discover the benefits of free eAA membership.



[@e-assessmentassociation](#)



[@eAssess](#)



[The e-Assessment Association](#)



[@eassessmentassociation](#)