

Source-Grounded and Human-Verified: An Archival Framework for AI-Assisted Assessment Design

Oct 7, 2026 · @Jami

Abstract

Generative AI has made writing assessment items fast and cheap; it has not made them trustworthy. This paper argues that archival practice already holds the controls AI-assisted assessment needs: provenance for every item, consistent description, and deliberate appraisal of what enters and stays in a collection.

I propose a six-stage framework that treats a question bank as an archival collection: blueprint, source mapping, AI-assisted drafting, human verification, structured description, and coverage and diagnostic review. It draws on my 2025 build of a study application for the Certified Archivist exam, a bank of more than 1,000 source-cited questions developed with ChatGPT in roughly one month.

The problem: cheap items, uncertain quality

A language model can draft a hundred plausible multiple-choice questions in minutes. Plausible is not the same as correct, and a question bank fails quietly: a wrong answer key or a fabricated citation looks exactly like a right one until a learner is misled by it.

Four risks recur in AI-drafted assessment:

- **Accuracy.** Answer keys and explanations can be wrong while sounding authoritative.
- **Provenance.** Models can attribute a concept to the wrong source or invent a page reference.
- **Coverage.** Volume hides imbalance. A thousand questions can still over-represent the easiest readings and leave whole domains sparsely covered.
- **Alignment.** Items drift from what the credential actually measures toward what is easy to generate, such as definitions instead of judgment.

These are not new problems in assessment. What is new is the speed: AI lets one person produce a bank in weeks that would once have taken a committee years, so the controls have to scale with it.

An archival lens on assessment

Archivists spend their careers deciding what to keep, documenting where it came from, and describing it so others can find and trust it. Each of those habits maps directly onto a question bank.

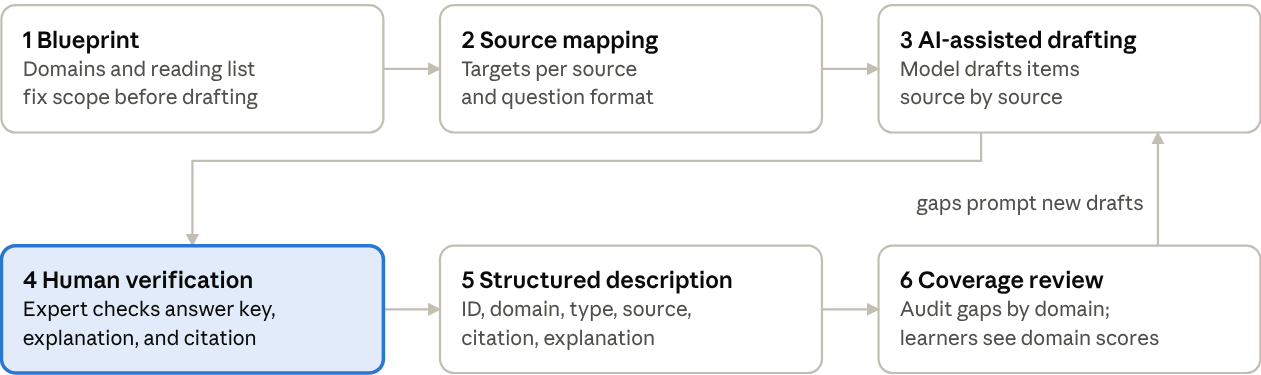
Archival principle	In a question bank
Provenance	Every item records the source it was drawn from, down to the page, and whether a person or a model drafted it
Description standards	Every item carries the same fields and controlled values, so it can be filtered, audited, and reused
Appraisal	Items are evaluated before they are accepted, not after a learner finds the error
Reappraisal and deaccessioning	Items are retired when sources are revised or the reading list changes
Collection policy	A blueprint defines what the bank should cover before anything is drafted

Seen this way, an AI model is a prolific donor: useful, fast, and in need of appraisal before anything it offers is accessioned.

The framework

The framework puts the model in one stage only, drafting, and surrounds it with human-owned stages before and after. Scope and sources are fixed before any item is drafted, and no item is described or released until a subject expert has verified it.

Every item passes human verification before it is described and released



AI-assisted assessment framework · 6 stages, 1 feedback loop

Items that fail verification go back to drafting rather than forward. Coverage review closes the loop: gaps it finds become new drafting targets, so the bank grows where it is thin rather than where generation is easiest.

Evidence from practice

In spring 2025, I built a study application for the Certified Archivist exam with ChatGPT as a drafting and coding partner. Dated project files show how each stage of the framework played out.

Stage	What I did	When (2025)
Blueprint	Adopted the ACA Study Guide's eight domains and the ACA Reading List as the scope	April 30
Source mapping	Set question targets per source and format, for example 50 items from Harvey's <i>Preservation Management Handbook</i>	May 2
Drafting	Drafted multiple-choice items source by source with ChatGPT; later added true/false and matching	April 30 to May 7
Verification	Ran a separate verified-answers pass across all eight domains	May 5

Stage	What I did	When (2025)
Description	Moved items from Word to CSV to JSON with ID, domain, type, source, citation, and explanation fields	May 3 to 7
Coverage and diagnostics	Built a source-by-domain coverage matrix; app reported per-domain scores	May 6 to 28

The build also tested the framework’s durability. When I deleted the coded app after platform errors, the question data, design specification, and coverage matrix survived intact because they were stored separately from the code. The content was the collection; the app was only an access tool.

Risks and ethical limits

The framework reduces risk; it does not remove it.

- **Exam security.** Certifying bodies keep their item banks confidential, and the ACA requires candidates to agree not to disclose exam content ([ACA Guide to the Exam](#)). Practice items must be written from published sources, never from recalled exam questions.
- **Copyright.** Feeding copyrighted books into AI tools and quoting them in items raises questions that call for paraphrase, short citations, and attention to each tool’s data terms.
- **Verification capacity.** Human review is the bottleneck. A bank that grows faster than a qualified person can verify it should not grow.
- **Format mismatch.** Practice formats should serve learning, but candidates must also rehearse the real exam’s format. The ACA exam is 100 multiple-choice questions in three hours.
- **Bias in sources and models.** A model trained on dominant perspectives can flatten contested topics. Domains such as cultural competency need especially careful human authorship.
- **Conflicts of interest.** Credential holders who serve their certifying body and also sell preparation materials should disclose that role and follow the body’s policies.

Recommendations

For practitioners building AI-assisted study tools:

1. Write the blueprint and the item schema before drafting a single question.
2. Record provenance on every item: source, citation, and whether a model drafted it.
3. Make verification a separate pass with its own record, not a step folded into drafting.
4. Track coverage by source and domain, not just total item count.
5. Store question data apart from application code, under version control, with off-platform backups.

For certifying bodies and professional associations:

1. Publish guidance on acceptable AI use in third-party preparation materials.
2. Make blueprints and reading lists machine-readable, so preparation tools align with them accurately.
3. Encourage disclosure of AI involvement and verification practices in commercial study products.

Conclusion

AI changes who can build serious assessment tools and how fast. Whether those tools deserve trust depends on controls archivists already know: provenance, description, and appraisal. Treating a question bank as a collection, and the model as a donor whose contributions must be appraised, turns speed into an advantage instead of a liability.

About the author

Jami Murphy, CA, is a project archivist at Auburn Avenue Research Library and serves on the Society of American Archivists Artificial Intelligence Task Force (2026–2029). She holds an MLIS and Graduate Certificate in Archival Studies from Louisiana State University and an MBA from Western Governors University.

Sources

- Academy of Certified Archivists, [Guide to the Exam](#)
- Academy of Certified Archivists, [Exam Development](#)

- Author's project files, April 30 to May 29, 2025: question-target summaries, verified answers, interface design notes, and source coverage matrix