Level 2: Applied | Digital Librarian

Module 08

Metadata & cataloging

Metadata work is labor intensive, detail oriented, and increasingly supported by AI tools. This module covers the real state of AI in cataloging: where it saves hours, where it introduces errors, and how to design quality controlled workflows.

By Yulia Brusova

20 min read · Reviewed June 2026

AI for Academic Libraries · ai-in-academic-libraries.vercel.app

WHAT YOU'LL BE ABLE TO DO

1. Apply AI to a batch metadata cleaning task in your current workflow
2. Evaluate at least one AI assisted metadata tool relevant to your collection
3. Identify where AI assisted subject tagging adds value and where human review is required
4. Draft a workflow for AI assisted record enhancement in your system
5. Understand the limitations and bias risks in automated classification

> *Cataloging and metadata work is, in certain respects, the library function most naturally suited to AI assistance: it involves pattern recognition, classification decisions, and normalization tasks that recur at scale across large and growing collections. At the same time, it is the function where errors are most consequential over time, because a poor subject heading or a mislabeled record degrades discoverability in ways that are difficult to detect and costly to correct at scale. In order to integrate AI into cataloging and metadata workflows responsibly, a digital librarian needs to understand both where AI accelerates the work without sacrificing quality, and where professional judgment remains the only reliable quality control available.*

Where AI adds genuine value in metadata work

The most reliable applications of AI in metadata work share a characteristic familiar from research support and instruction contexts: AI is most useful when working with content the professional has already assembled, rather than generating metadata from training data alone. AI that receives an item's title, abstract, table of contents, and existing partial record and generates candidate subject headings or a descriptive summary is operating from provided content, which carries lower hallucination risk and higher practical reliability. AI asked to produce authoritative catalog records from a title string alone is generating from training data patterns, with the attendant risk of confident errors.

Batch record cleaning and normalization is the area of most immediate practical value for most cataloging workflows. Collections accumulated over decades carry inconsistencies: date formats that vary across cataloging eras, name authority entries appearing in multiple forms, controlled vocabulary terms that have been superseded, encoding errors affecting display and indexing, and MARC field inconsistencies that cause problems in ILS export and

discovery. AI given a clear normalization ruleset and a batch of records can identify inconsistencies, flag records requiring attention, and apply standardized corrections at a scale that would require hundreds of hours of manual review. For example, a digital librarian managing a historical collection with inconsistent birth and death date formatting across several thousand records can provide Claude with a CSV export, a description of the target format, and a few examples, and receive a corrected output for review. Such a task requires no professional classification judgment from AI, only consistent rule application, which is among the most reliable AI capabilities.

Subject heading suggestion from provided item content is a second area of consistent value. For items with inadequate description (digitized photograph collections where records contain only title and date, or dissertation repositories where records carry no subject headings at all), AI can generate candidate subject headings based on provided content. For example, providing Claude with a dissertation abstract and asking it to suggest LCSH compliant subject headings produces a candidate list that a cataloger reviews, accepts, rejects, or modifies, rather than constructing the list from scratch. Such suggestions require professional review against the actual LCSH vocabulary and item content, but they provide a starting point that substantially reduces per record processing time.

Description generation for digital collections addresses one of the most common digitization bottlenecks: items have been technically processed, but minimal records without descriptive text limit their discoverability. AI can draft catalog descriptions from provided metadata, OCR text, or transcription content, which catalogers review and refine. For example, an archivist working with digitized correspondence can provide AI with letter transcriptions and ask it to draft three to five sentence catalog descriptions capturing sender, recipient, date, and primary subject. Such drafts accelerate the description workflow while leaving the substantive accuracy review in professional hands: the appropriate division of labor between AI efficiency and human expertise.

Batch record cleaning and normalization: a practical workflow

The workflow for AI assisted batch record cleaning is straightforward enough to implement immediately, but successful execution depends on the quality of the instructions provided to the AI rather than on the model's autonomous judgment. Such clarity is the professional contribution that makes AI batch processing reliable: the normalization rules come from the cataloger, not from the model.

The workflow proceeds in four stages. First, identify the specific normalization problem to address: inconsistent date formatting, name authority variants, outdated controlled

vocabulary terms, or missing MARC fields of a specified type. Being precise about the scope prevents AI from making changes beyond the intended correction. For example, a task specified as "normalize all 100 field entries to match AACR2 format, no other changes" is more reliable than "improve the name authority entries"; the first gives AI a specific rule to apply, while the second invites judgment it is not qualified to exercise.

Second, prepare a representative sample of records in a format AI can process, typically a CSV export with relevant fields or MARC records converted to a readable format. Provide AI with three to five examples of the current format alongside the target format, making the transformation rule explicit rather than inferred. For example: "The 100 field currently reads 'Smith, John (1945-).' The target format is 'Smith, John, 1945-' (note the removal of parentheses). Apply this transformation to all 100 fields in the attached records and flag any record where the date information is ambiguous or missing." Such explicit instruction produces consistent output that the cataloger can verify efficiently.

Third, process the output on a sample before committing to the full batch. Running AI corrections on fifty records before applying them to five thousand is not overcautious; it is the quality control step that catches misapplication of rules before the error is compounded across the entire collection. For example, a normalization rule that works correctly for most name authority formats may fail on compound surnames, names with prefixes, or non Western name structures that follow different ordering conventions. Catching such failures in a sample prevents them from propagating through the full collection. Such a pilot step should be standard practice for any AI batch processing task, regardless of how clearly the rules were specified.

Fourth, document the transformation rules applied as part of the collection's maintenance record. Such documentation is valuable both for quality audit purposes (if a systematic error is discovered later, knowing what rule was applied identifies which records to check) and for reproducibility, since the same normalization approach can be applied to new accessions without reconstructing the workflow from scratch. Indeed, a well documented normalization workflow becomes a reusable institutional resource that builds in value over time.

AI assisted subject heading suggestion: workflow and professional oversight

AI assisted subject heading suggestion is among the most directly labor saving applications in cataloging, and among the most professionally significant for quality control. The efficiency gains are real: reducing per record cataloging time by providing candidate

headings rather than requiring the cataloger to construct them from scratch is a measurable workflow improvement, particularly for large backlogs or high volume digitization projects. The professional responsibilities are equally real: candidate headings require review against the actual controlled vocabulary, the item's content, and the library's local practice.

The practical workflow for subject heading suggestion follows the information in principle. Provide AI with as much item content as is available: title, abstract or summary, table of contents, any existing partial metadata, and a description of the collection context. For example: "Suggest LCSH subject headings for an academic dissertation with the following abstract: [paste abstract]. The library uses LC subject headings without local modifications. Suggest between three and five headings, ordered from most to least specific, and note which headings you are less confident about." Such a prompt produces a candidate set from which the cataloger selects, adjusts, and authorizes against the actual LCSH authority file.

The most important quality control dimension is the verification step: every AI suggested subject heading must be confirmed against the LCSH authority file before being applied. AI suggestions will sometimes use outdated terminology, produce seemingly plausible headings that do not exist in the controlled vocabulary, or apply headings that are real but that the specific item's content does not fully warrant. For example, AI might suggest a broad geographic or demographic heading for an item that would be better served by a more specific term, not a fabrication but a professional judgment error that requires the cataloger's expertise to identify. Additionally, AI suggestions for items in specialized, technical, or interdisciplinary subject areas carry higher imprecision risk, because the relevant controlled vocabulary may be less well represented in the model's training data than mainstream disciplinary literature.

Furthermore, subject heading suggestion for items about communities and experiences that have been historically underrepresented or misrepresented in LCSH requires heightened professional attention. AI trained on existing catalog records inherits the vocabulary of those records, including terminology that contemporary reparative cataloging practice recognizes as inadequate or harmful. This dimension of AI assisted subject cataloging is examined more fully in the section on bias and representation, and it is the reason that "AI suggests" can never be treated as equivalent to "cataloger authorizes" in any responsible workflow.

Description generation for digital and archival collections

Description generation is the AI application with the clearest return on investment in digital collections work, because the bottleneck it addresses, namely the gap between digitized items that have been technically processed and items described thoroughly enough to be

discoverable, is a persistent challenge for nearly every digital collections program. Staff time for original description work is limited; the number of items requiring description is not. AI's ability to draft descriptions from provided content shifts the work from origination to review, a substantial efficiency gain that makes backlogs tractable without sacrificing the professional oversight that description quality requires.

The most reliable approach is to provide AI with every relevant piece of content available for the item: digitized text if OCR is available, existing partial record fields, physical description, provenance information, and collection context. For example, for digitized historical correspondence, a prompt might read: "Write a 75-word catalog description for the following letter based on the provided transcription: [paste transcription]. The description should identify the sender, recipient, date, primary subject, and any significant details that would help a researcher determine relevance. Write in past tense, third person, plain professional language." Such a prompt produces a description the archivist reviews for accuracy and completeness rather than drafting from scratch; the professional's time is spent on verification and refinement rather than on initial origination.

The professional review step is where AI drafted descriptions most commonly require correction. AI may infer geographic settings incorrectly from sparse metadata, misidentify persons or institutions from context clues in the text, assign dates with more precision than the evidence warrants, or describe images in ways that reflect training data biases about what subjects and settings look like. For example, a photograph labeled only "Community gathering, 1930s" may receive an AI description that assumes a specific ethnic or regional community based on visual pattern matching from training data; such an assumption may be factually incorrect and that carries representational significance. Such errors require professional correction, and they are the reason that description review must be substantive rather than cursory.

For archival finding aids, AI description generation serves a specific and well defined function: converting container level entries with minimal description into more fully described series and box summaries that improve discoverability without requiring item level cataloging for every folder. For example, a large archival collection where series level descriptions are adequate but box level descriptions are absent can be approached with AI drafting box descriptions from the folder titles within each box: a pattern based task that AI handles reliably when the folder level data is provided as context. Such an approach does not replace the archivist's professional interpretation of the collection's significance; it extends that interpretation to a level of granularity that would otherwise require far more staff time than is available. Furthermore, AI drafted finding aid descriptions can be reviewed at a pace that

matches staff availability, which makes the approach sustainable for ongoing digitization programs rather than only for one time retrospective projects.

Bias and representation in automated classification

Automated classification systems inherit the biases of their training data. In library cataloging, this is not a theoretical concern; it is a direct professional issue with consequences for collection discoverability and community representation that practicing catalogers encounter in specific, identifiable ways. AI systems trained on existing catalog records will replicate and in some cases amplify the biases embedded in those records, and librarians deploying AI assisted classification need to understand this mechanism well enough to design workflows that catch and correct for it.

The Library of Congress Subject Headings vocabulary has been the subject of sustained professional critique for decades, with scholarship and community advocacy documenting systematic inadequacy in the representation of Indigenous communities, LGBTQ+ people and experiences, communities of color, non Western cultural and intellectual traditions, and many other historically marginalized groups. The problems are not limited to obviously offensive historical terms; they include structural issues such as the use of dominant culture framing as the unmarked default, the absence of subject headings that reflect community self identification, and the granularity imbalance that provides detailed vocabulary for Western European topics while providing much coarser vocabulary for topics involving non Western cultures and communities. For example, LCSH has historically provided far more granular terms for the political and cultural histories of Western European nations than for comparable topics in African, Asian, or Indigenous contexts: a disparity that reflects the collections the headings were originally designed to serve, not the relative significance or complexity of those subjects.

AI trained on existing LCSH governed catalog records reproduces this vocabulary. A system that suggests subject headings from existing catalog patterns will suggest the terms that appear most frequently in training data, which means it will suggest inadequate terms for materials about underrepresented communities just as existing catalog records use inadequate terms for those materials. Furthermore, because AI weights high frequency training examples, the terms most commonly applied to materials about marginalized communities, including terms that reparative cataloging practitioners are actively working to replace, are precisely the terms AI will most confidently suggest. Such a system does not amplify bias in the sense of inventing new harms; it entrenches existing harm by applying it efficiently and at scale.

The professional responsibilities that follow from this understanding are clear. AI subject heading suggestions require human review for all materials, not as a formality, but as a substantive quality control step. For materials about historically underrepresented or marginalized communities, review should be conducted by catalogers with specific knowledge of reparative cataloging principles and current community driven vocabulary initiatives such as the Cataloging Lab and relevant SACO funnel projects. Additionally, the ARL Guiding Principles for AI explicitly call for librarians to understand and raise awareness of AI bias; in cataloging contexts, this means both applying that awareness in daily practice and contributing to institutional and professional conversations about how AI assisted cataloging workflows are designed and governed. There is no doubt that the efficiency gains from AI assisted subject tagging are only professionally defensible when the quality control layer is substantive, documented, and applied with particular care to the collections and communities that existing cataloging systems have historically served least well.

Tools in the AI metadata space: an honest evaluation

The tools available for AI assisted metadata work fall into two broad categories: integrated features within existing library systems that catalogers access through their established platforms, and general purpose AI tools applied to metadata tasks by catalogers who have developed their own workflows. Understanding what each category offers, and what it requires from the cataloger to use responsibly, which guides the practical tool decisions that digital librarians face in a landscape that is changing more rapidly than professional documentation can track.

Integrated library system AI features carry the most institutional significance because they operate within the systems the library already manages and are governed by existing vendor relationships and data agreements. Ex Libris has been integrating AI features into Alma and Primo across multiple release cycles, including AI assisted metadata enrichment, subject heading suggestions drawn from the WorldCat knowledge base, and authority control matching. For example, the Alma catalog module has included piloted features that suggest subject headings based on existing record content by matching against WorldCat holdings patterns. Such integrated features are worth investigating through the library's vendor account manager and the Ex Libris developer network for current availability and configuration options, as the release cadence is faster than formal documentation cycles. Similarly, OCLC has invested in AI applications for WorldShare record quality, including duplicate detection, subject heading suggestions, and automated quality scoring. Librarians whose institutions use WorldShare Management Services should review the OCLC AI feature documentation regularly, as capabilities have been expanding through the same period.

AERIE (AI Enhanced Record Improvement) is a tool designed specifically for academic library catalog record enhancement that has received discussion in professional library technology forums and LTI assessments. Its application to batch record improvement and subject heading suggestion addresses the same cataloging bottlenecks covered in this module. As with any specialized tool in this space, current availability, institutional pricing, and feature scope should be verified through direct vendor contact and current community assessments rather than relying on documentation that may be outdated, a caution that applies across the entire AI metadata tool landscape.

General purpose AI tools such as Claude and ChatGPT are a practical option for smaller scale normalization and description tasks that do not require specialized cataloging system integration. For example, a cataloger working on a batch of digital object records that need descriptive summaries can develop a reliable prompt workflow using Claude without any specialized tool, producing draft descriptions that are reviewed through the existing quality control process. Such an approach is most appropriate for rule based, content grounded tasks, not for those requiring authoritative MARC validation or ILS system integration. In order to evaluate any tool in this space, librarians should apply the framework from Module 03: understand the business model, review the privacy terms (patron search data entering external AI systems is a specific concern for catalog integrated tools), check for community assessments from LTI and LITA, and pilot with low stakes records before applying to production collections.

OCLC's December 2025 cataloging AI: classification, subject headings, and provenance

On December 8, 2025, OCLC announced a substantial expansion of AI assisted cataloging within WorldShare Record Manager and Connexion, the two interfaces most catalogers use daily for original and copy cataloging against WorldCat. The new features suggest Dewey Decimal Classification and Library of Congress Classification numbers, along with Library of Congress Subject Headings, drawn from the patterns present across WorldCat's aggregate holdings data. For example, a cataloger working a title with minimal existing copy can receive a suggested DDC number and a set of candidate LCSH terms generated from how WorldCat's vast body of existing records has classified and described similar titles, rather than from a general purpose language model's training data. Such grounding in WorldCat's own holdings is the same retrieval based design this module has favored throughout: the suggestion is anchored in records the cataloging community has already produced, not generated from an unrelated corpus. As with the subject heading suggestion workflow

described earlier, the cataloger retains full accept or reject control over every suggestion; nothing is applied to a record automatically.

In pilot use, catalogers described the new features as a "safety net": a second set of eyes that surfaces a classification number or heading the cataloger might not have considered, particularly for titles in unfamiliar subject areas, without removing the cataloger from the decision. Pilot participants also reported saving up to twenty minutes per title. Such a figure is worth taking seriously as a signal of where the time savings are concentrated, namely the search and comparison work that previously required manually checking similar records, but it should be treated as a vendor and pilot reported estimate rather than an independently audited benchmark. In order to apply this figure responsibly, a library piloting these features should measure its own time savings against its own baseline, in the same way this module has recommended for any AI assisted cataloging workflow: establish a calibration period, measure actual per title time across a representative sample, and let that measurement, not the vendor's reported figure, inform staffing and workflow decisions.

OCLC paired this functional expansion with an update to its own documentation. Bibliographic Formats and Standards §3.5 now specifies how to record provenance for AI generated metadata: how a record notes that a classification number, a subject heading, or another element originated as an AI suggestion that a cataloger reviewed and accepted. Such a provenance note is a small addition to a record, but it is professionally significant. It means that a future cataloger, or a future audit, can identify which elements of a record passed through an AI assisted step, which is exactly the kind of documentation this module has argued for at the workflow level, now formalized at the level of the standard itself. For a library designing its own AI cataloging workflow, BFAS §3.5 is worth reading directly, because it provides a model for how local documentation practices, such as the quality audit protocols discussed earlier in this module, can align with a national standard rather than inventing a parallel local convention.

Separately, OCLC has also been testing AI at the collection level rather than the record level. On February 11, 2025, OCLC ran an AI assisted WorldCat deduplication test targeting roughly five hundred thousand record pairs of print English language books: pairs of records that may describe the same physical item but exist as separate entries in WorldCat for reasons ranging from cataloging era differences to minor transcription variation. Deduplication at this scale is a different kind of task from the per record suggestions described above; it is a pattern matching problem across an enormous record set, the kind of task this module has consistently identified as well suited to AI precisely because it does not require generating new bibliographic content, only comparing existing records against each other and flagging likely matches for review. The professional decision about whether two records actually

describe the same item, and which record should be retained or merged, remains a cataloger's judgment; what AI changes is the scale at which candidate pairs can be surfaced for that judgment to be applied.

None of these developments changes the underlying professional relationship between cataloger and AI suggestion that this module has described throughout. A classification number drawn from WorldCat patterns, a subject heading flagged as AI suggested under BFAS §3.5, and a deduplication candidate pair surfaced from half a million records are all, in the end, suggestions: inputs to a decision that a cataloger makes, not a decision made on the cataloger's behalf. The twenty minute per title figure, if it holds up under a library's own measurement, represents time returned to catalogers for exactly this kind of judgment, applied to the titles, collections, and communities that a national aggregate dataset cannot know as well as the cataloger holding the item does.

Designing a quality controlled AI cataloging workflow

A responsible AI assisted cataloging workflow is not the existing cataloging workflow with AI added at the beginning. It is a deliberately designed process that specifies what AI will do, what human review will verify, what quality standards apply, and how the workflow will be audited and adjusted over time. Such design takes more upfront thought than informal AI adoption, and it produces substantially more reliable and professionally defensible results.

The workflow design begins with task specification: identifying precisely which cataloging tasks will be AI assisted, and which will remain entirely in professional hands. Not every cataloging task benefits equally from AI assistance, and some (original cataloging of complex, specialized, or culturally sensitive materials; classification decisions for items in emerging subject areas; authority work requiring interpretation of identity and community naming) require professional judgment that AI cannot reliably provide. For example, a workflow that uses AI for batch normalization of date and name authority fields while keeping original subject heading assignment entirely in professional hands is more defensible than one that uses AI for all MARC field generation, because the former concentrates AI assistance where rule following is sufficient and preserves professional control where judgment is required.

The review architecture is the most critical design decision. Determining what percentage of AI output will be reviewed, at what depth, and by whom: these decisions determine whether the efficiency gain from AI assistance is real or illusory. A workflow that processes one thousand records with AI and reviews ten percent produces genuine efficiency only if the ninety percent not reviewed are reliably correct. Establishing that baseline reliability requires

an initial period of higher review rates, specifically reviewing fifty to one hundred percent of output across a range of record types, to measure AI accuracy on the specific tasks the workflow handles. Such calibration is not optional; it is the professional foundation that makes reduced review rates defensible and the efficiency claim honest.

Quality auditing is the ongoing component that distinguishes sustainable AI workflows from initial experiments. After AI processed records go live in the catalog or digital collection, a sample should be reviewed at regular intervals to measure error rates, identify systematic problems, and adjust the workflow accordingly. For example, if a quarterly audit reveals that AI suggested subject headings for a specific collection type are consistently inaccurate at a higher rate than the overall workflow baseline, that collection type should move to a higher review category until the error source is understood and addressed. Such auditing is consistent with ACRL subcompetency 4.1; applying AI for task efficiency requires that the efficiency is real, which means the output must be accurate enough to serve the collection's users.

In order to communicate the workflow's design and quality standards to administrators and colleagues, it is useful to document the workflow in a brief written protocol that specifies the tasks included, the review percentages applied, the quality thresholds that trigger workflow adjustment, and the audit schedule. Such documentation supports the professional accountability principle that human professionals are responsible for AI assisted outputs, and provides the institutional record needed to evaluate the workflow's performance and defend its use if questions arise about record quality.

PRACTITIONER NOTE

A compelling case for AI assisted cataloging often comes from date normalization: legacy records with inconsistent date formats accumulated across cataloging eras. A collection where birth and death dates in the 100 field appear in four different formats across approximately eight thousand records represents a known problem that gets deferred for years because the manual correction time is prohibitive. Processing those records through Claude with a precise normalization rule and a set of examples, then reviewing the output on a sample of two hundred records before committing to the full batch, and correcting the outliers where the rule has not applied cleanly, can reduce a project estimated at several weeks to two days of work. Such an experience does not reduce professional care in cataloging; if anything, seeing how precisely the rules need to be specified to produce clean output increases attention to what professional specification actually requires. In order to identify similar opportunities, auditing the

catalog for the most common consistency problems and asking whether each one has a rule precise enough to be delegated to AI is the practical starting point. If the rule can be written clearly, the task is a candidate.

KEY TAKEAWAYS

- AI is most reliable in metadata work when applying explicit rules to provided content; batch normalization, description drafting from supplied text, and subject heading suggestions from provided abstracts all carry substantially lower risk than AI generating records from sparse inputs alone.
- Batch record cleaning (normalizing date formats, name authority variants, and controlled vocabulary) is the highest value, lowest risk AI application for most cataloging workflows; the rules come from the cataloger, not from the model.
- AI suggested subject headings require professional verification against the actual LCSH authority file before application; AI inherits the vocabulary of its training data, including outdated, superseded, and biased terminology.
- Automated classification systems replicate and can amplify the representational biases of existing catalog records; materials about historically marginalized communities require heightened human review regardless of workflow efficiency pressures.
- Integrated library system AI features (Alma, WorldShare) and specialized tools like AERIE should be evaluated through current vendor documentation and LTI assessments; general purpose AI is appropriate for smaller scale, rule based normalization without ILS integration.
- OCLC's December 2025 AI features in WorldShare Record Manager and Connexion suggest DDC, LCC, and LCSH from WorldCat data with full cataloger accept/reject control; BFAS §3.5 now specifies how to record AI provenance, and a February 2025 test applied AI deduplication to roughly 500,000 WorldCat record pairs. Reported time savings of up to twenty minutes per title are vendor and pilot reported, not independently audited.
- A defensible AI cataloging workflow specifies task scope, establishes initial review rates calibrated to measured accuracy, and includes regular quality audits that can trigger workflow adjustment when error rates exceed acceptable thresholds.

References

APA 7TH EDITION

1. Association of College and Research Libraries. (2025, October). *AI competencies for academic library workers*. American Library Association. https://www.ala.org/acrl/standards/ai
2. OCLC. (2025, December). *AI-assisted cataloging in WorldShare Record Manager and Connexion*. https://www.oclc.org/en/worldshare-record-manager.html

   Confirm exact release note URL when updating the site.

ACRL AI Competencies covered

Use & Application

Analysis & Evaluation

Sub-competencies: 4.1, 3.4, 4.4 · ACRL AI Competencies (2025)