Level 2: Applied | All Librarians

Module 11

AI for collections & vendor evaluation

Vendors are selling academic libraries AI products faster than libraries can evaluate them. This module covers where AI actually fits in collection development and acquisitions, how to test vendor AI search tools against your own queries, and how to decide what is worth the budget.

By Yulia Brusova

25 min read · Reviewed July 2026

WHAT YOU'LL BE ABLE TO DO

1. Distinguish where AI genuinely assists collection development from where vendor marketing overstates it
2. Identify the AI features already bundled into the platforms your library subscribes to
3. Test a vendor AI search product against your own reference queries rather than vendor demonstrations
4. Distinguish bundled AI features from paid add-ons when evaluating cost and value
5. Apply a published evaluation rubric to a vendor AI product against your specific student population
6. Decide whether to adopt, pilot, or decline a vendor AI feature using a defensible framework

> *Collection development and vendor evaluation is the part of library work where AI has arrived most aggressively and where the gap between what is marketed and what is delivered is widest. Every major content vendor now sells an AI product, and the sales conversations have moved well ahead of the independent evaluation that should inform them. A selector or electronic resources librarian sitting through a vendor demonstration is shown a polished interface answering a curated question well, and is rarely shown the same tool failing a harder question or returning five sources when the patron needed forty. The most useful orientation for a librarian approaching AI in this space is to separate three things that vendors deliberately blur: the back-office automation that genuinely works, the patron-facing AI search products that are widely available but still limited, and the marketing language that attaches the word AI to analytics and metadata tools that are not generative AI at all. This module is built to help you tell those apart, test the claims against your own collection and your own patrons, and decide what deserves a place in a budget that is already under pressure.*

Where AI fits in collection development: selection, deselection, and the new problem of AI itself

AI in collection development is real, but it is mostly decision support rather than decision making, and most of what vendors label AI in this space is machine learning and analytics

rather than generative AI. The distinction matters because the professional actions appropriate to each are different. Ex Libris Rialto, the marketplace and selection platform now under Clarivate, embeds AI-driven recommendations that refine search results for selectors based on past purchases and patron engagement, and Clarivate has announced that the Collecto collection analytics suite is being folded into Rialto to support deselection, retention, and space optimization across the Alma community. For example, a selector using such a platform receives ranked recommendations and usage-based deselection candidates rather than a finished selection list, which means the tool narrows the field of attention but does not make the collection decision. Such a division of labor is the realistic shape of AI in selection today, and much of the more ambitious language in vendor materials describes roadmap features rather than capabilities a selector can use this month. In order to evaluate any selection or analytics tool honestly, the librarian should ask the vendor to distinguish what is shipping now from what is planned, because the marketing rarely makes that distinction unprompted.

Weeding and deselection follow the same pattern. Data-driven deselection predates generative AI by years through usage statistics, retention candidates, and overlap analysis, and the current direction layers machine learning onto that established data rather than replacing the librarian's judgment. For example, a machine learning model trained on a library's prior weeding decisions can pre-screen candidates so that a librarian reviews a focused list rather than the entire collection, but the model is reproducing patterns in past decisions, not exercising the collection knowledge that justified them. Such a tool is genuinely useful for managing scale, and it is also exactly the kind of tool whose recommendations must be read as a starting point for review rather than an instruction to discard. Demand-driven and evidence-based acquisition are well integrated into these platforms as well, though the 2025 disruption to perpetual ebook purchasing discussed later in this module directly affects how those workflows can be configured.

The scholarly evidence supports a cautious posture. Portillo and Carson, evaluating general large language models for collection development in the Journal of the Medical Library Association in January 2025, concluded that the models are not yet reliable as primary tools for collection development due to inaccuracies and hallucinations, but can serve as supplementary tools for analyzing subject coverage and identifying gaps. This indicates that a general model is appropriate for gap analysis against an existing collection, where the librarian can verify every suggestion, and inappropriate as a source of titles to buy, where a hallucinated citation can enter an order list unchecked.

A newer wrinkle is that AI has become a collection development problem in its own right. AI-generated books have begun entering library collections through aggregator platforms,

documented by 404 Media in February 2025 and discussed at IFLA the same year, where machine-generated titles appear in catalogs without ever passing through selection, a particular hazard in pay-per-use lending models. Such titles are precisely the kind of low-quality material that selection exists to keep out, and they enter through automated supply chains that bypass the selector entirely. In order to address this, libraries are updating collection development policies to name AI-generated content explicitly. What none of these tools and none of these problems change is the core of the work: deciding what a specific community of users actually needs, and standing behind that decision, remains a judgment that belongs to the librarian who knows the collection and the patrons it serves.

AI for acquisitions and licensing operations: less mature than the marketing suggests

Acquisitions and licensing is where the gap between expectation and reality is largest, because the infrastructure that tracks open access agreements and the tools that review licenses are mostly automation and metadata matching rather than generative AI. Open access agreement tracking runs on the ESAC Registry, the cOAlition S Journal Checker Tool, and OA.Report from the nonprofit OA.Works, and these are built on structured data and automated analysis rather than language models. For example, a librarian confirming whether an author qualifies for a transformative agreement consults a registry lookup and a checker tool, not a generative assistant, and the reliability of that answer depends on metadata quality rather than on any AI summarization. Such a distinction matters because a librarian who expects an AI assistant to interpret an agreement will be disappointed, while a librarian who understands these as data tools will use them correctly. The most explicitly AI-enabled product in this adjacent space, Copyright Clearance Center's OA Intelligence, uses affiliation-matching technology, but it is publisher-facing for modeling open access offers rather than a library acquisitions tool.

AI-assisted license and contract review, the application most librarians ask about first, barely exists as a library-specific product. The clearest documented case is a 2025 pilot by an assistant director at the San Diego Law Library, published on the AALL CRIV blog, who trialed Spellbook, a general legal-AI Word add-in, on electronic resource license terms and asked it to flag problematic clauses for a public library. For example, the tool flagged ambiguous authorized-user language that matched issues the library had already identified and negotiated, which both validated the tool's usefulness and revealed its limits, since it surfaced a known problem rather than an unknown one. Such a result is encouraging and modest at the same time: the librarian concluded there is a definite possibility for future application in reviewing contracts, but that human review and human context will always be

warranted. This suggests that a librarian experimenting with a general AI tool on license terms today should treat it as a second reader that may catch what fatigue misses, never as a substitute for reading the agreement.

The more consequential licensing conversation in 2024 through 2026 is not about AI summarizing agreements but about AI clauses inside the agreements themselves. The ICOLC statement on AI in licensing, the Association of Research Libraries' electronic resource licensing guidance, and consortial work on library-friendly AI terms all concern how text-and-data-mining and AI-use rights are negotiated into electronic resource licenses, and whether uploading licensed content into a public AI tool violates the license. For example, a librarian who pastes a licensed article into a general chatbot to summarize it may be transmitting licensed content to a third party in violation of the agreement, which is a licensing question the AI tool will never raise on its own. Such risks make the human reading of the contract more important rather than less, and the professional judgment about what the institution may lawfully do with licensed content under an AI workflow remains squarely with the librarian negotiating and administering the license.

Evaluating vendor AI products: the same architecture, the same limits, tested against your own queries

The patron-facing AI search products that vendors are selling most aggressively share a common architecture, and understanding it is the fastest way to evaluate any of them. Nearly all of these tools use retrieval-augmented generation, which means a natural-language question is converted into a search, a small set of results is retrieved from a proprietary index, and a language model summarizes those results with citations. This architecture grounds answers in licensed content and reduces hallucination, but it does not eliminate it, and it introduces a characteristic limitation: the summary reflects only the handful of sources the system retrieved, not the full literature on the question. For example, a tool that retrieves and summarizes five sources will produce a confident, well-written answer whether or not those five are the five a subject expert would have chosen. Such confidence in the prose is precisely what makes independent testing necessary, because the fluency of the output is unrelated to the completeness of the underlying search.

The products differ mainly in what they search and how they are priced. Ex Libris Primo Research Assistant, released to production in September 2024, is bundled at no additional cost for Primo customers, converts the question to a Boolean query, retrieves results from the Central Discovery Index, and summarizes the top five with inline citations; it searches metadata and abstracts rather than full text and can surface items the institution does not

own. Elsevier's Scopus AI draws only on Scopus-indexed peer-reviewed content and offers summaries, expanded summaries, and concept maps, and it is a paid add-on with undisclosed pricing. EBSCO's AI Insights and Natural Language Search, launched in 2025 across EBSCOhost and EBSCO Discovery Service, generate key-point summaries and translate conversational queries into Boolean. JSTOR's AI tool, out of beta and available to participating institutions by mid-2025, is document-scoped, helping a user assess, summarize, and question a single selected item. Clarivate's Web of Science Research Assistant, a paid add-on launched in September 2024, is grounded only in the Web of Science Core Collection. For example, the same reference question will return different answers across these tools not because one model is smarter but because each searches a different index with a different scope, which is the single most important fact to convey to a patron who assumes an AI answer is comprehensive. Independent reviewers, including Choice, the Katina reviewer, and university LibGuide authors, consistently rate these tools as useful starting points that require critical evaluation rather than authoritative search replacements.

Not all vendor AI in this space is patron-facing, and the back-office applications are often the more mature. OCLC has deployed AI for WorldCat de-duplication and, in December 2025, AI metadata enrichment in WorldShare Record Manager and Connexion that suggests classification and subject headings with catalogers retaining accept-or-reject control, and Ex Libris offers a comparable metadata assistant in Alma. For example, the OCLC de-duplication work was trained on tens of thousands of human-labeled duplicate pairs before it was run at scale, which is the kind of grounding that distinguishes a genuinely useful production tool from a demonstration. In order to evaluate any of these products honestly, the most reliable method is the one independent reviewers use: run your own real reference queries through the tool, compare the result against a non-AI search of the same index, and read the cited sources rather than trusting the summary. Such testing on your own questions, with your own collection and your own patrons in mind, is the evaluation no vendor demonstration can substitute for, and the professional judgment about whether a tool serves your users belongs to the librarian who runs that test, not to the vendor who designed the demonstration.

Cost, budget, and value: bundled versus add-on, and the lesson of the Clarivate ebook controversy

The central financial question with vendor AI is whether a feature is bundled into a subscription you already pay for or sold as a separate add-on, because that distinction determines both the cost and the leverage you have. Primo Research Assistant and JSTOR's tool are bundled at no additional charge for existing subscribers, while Scopus AI and Web of

Science Research Assistant are paid add-ons with pricing that is not publicly disclosed and is negotiated case by case. For example, a library evaluating a bundled tool is deciding only whether to turn a feature on, while a library evaluating an add-on is committing new money against a budget that is already strained, and the evaluation rigor should scale accordingly. Such opacity in add-on pricing is a recurring pattern in this market, and it is compounded by the fact that none of these tools currently provide COUNTER usage statistics, which means the traditional cost-per-use analysis that libraries rely on to justify renewals cannot yet be performed on AI add-ons.

The budget context makes this more pointed. Clarivate's Pulse of the Library 2025 report, based on more than two thousand librarians across over a hundred countries, found that sixty-seven percent of libraries are exploring or implementing AI tools, up from sixty-three percent the prior year, while budget constraints had become the primary barrier to AI adoption, cited by sixty-two percent of respondents and surpassing lack of expertise, which had ranked first the year before. This indicates that interest is rising and money is tightening at the same time, which is exactly the condition under which an opaque, COUNTER-less add-on deserves the most scrutiny. Within most institutions these content-linked AI tools fall on the library materials and electronic resources budget rather than on central IT, which means the library bears the cost directly and should weigh it against the books, databases, and staff time the same money would otherwise buy.

The Clarivate ebook controversy of 2025 is the cautionary tale to teach alongside the numbers. In February 2025 Clarivate announced it would phase out one-time perpetual ebook and print purchases along with demand-driven and evidence-based acquisition in favor of subscription access, framing the shift partly around AI-powered discovery, and the international backlash was immediate, including an ICOLC statement and a competing vendor publicly reaffirming perpetual and demand-driven options. For example, within weeks Clarivate apologized for the lack of consultation and extended perpetual-purchase ability across its platforms through June 2026, a reversal that demonstrates how quickly a vendor's strategic positioning can shift and how much leverage a coordinated library response can have. Such an episode is a teaching case in vendor lock-in, the fragility of ownership, and the way AI-era platform strategies intersect with budget and collection control. Formal dollar-figure return-on-investment analyses for AI search add-ons are essentially absent from the literature, which means the value judgment cannot be outsourced to a published benchmark; deciding whether a given AI product is worth its cost against everything else the budget must cover remains a judgment that belongs to the librarian accountable for that budget.

Patron-facing and privacy considerations specific to vendor AI search tools

Vendor AI search tools introduce privacy considerations that differ from those of the underlying databases, and the vendor commitments, while broadly reassuring, vary enough that they must be read clause by clause rather than trusted by reputation. Across Primo Research Assistant, Scopus AI, and Web of Science Research Assistant, vendors claim that user queries are processed in private or walled environments and are not used to train public models, and several state that inputs are used only for the current session and not stored. For example, Clarivate states that Web of Science Research Assistant data is not used to train large language models directly or indirectly and that only the user's query, not publisher content or library-owned materials, is passed to the model, while JSTOR retains de-identified conversation logs and caps the retention of data sent to its model providers at thirty days. Such commitments differ in their specifics, and the differences are exactly where a privacy review must focus: retention windows, whether queries feed analytics, and whether the tool requires a personal login.

The login requirement is the consideration most specific to these tools and most easily overlooked. Primo Research Assistant and JSTOR's tool require user authentication, which means a patron's queries can in principle be associated with an identified individual, a different privacy posture than an anonymous database search. For example, a student researching a sensitive health or legal question through an authenticated AI assistant is generating a query log tied to identity in a way that the same search through an anonymous catalog interface would not, which is a meaningful distinction for reader privacy. Such identification questions sit directly against the library profession's long-standing commitment to the confidentiality of patron research, expressed in the Library Digital Privacy Pledge and the American Library Association's intellectual freedom guidance. Additionally, independent commentators have urged libraries to remain deeply skeptical of AI summarization tools that strip citations from their original context, even while cautiously adopting natural-language discovery, because a summary that detaches a claim from its source undermines the patron's ability to evaluate it.

The honest state of this area is that no published, tool-specific privacy audit by a named library privacy officer yet exists for these products, which means libraries are currently relying on vendor self-description rather than independent verification. This is a genuine gap and an area to monitor. In order to act responsibly in the meantime, a library should read each product's privacy policy against its own patron-privacy standards before enabling the tool, communicate clearly to patrons when an AI feature logs identified queries, and treat the

vendor's assurances as claims to be verified rather than facts to be assumed. Such verification, and the decision about whether a tool's privacy posture is acceptable for a specific community of users, is a professional responsibility that cannot be delegated to the vendor whose product is under review.

ALA's vendor data-review checklist: the questions to ask before you buy

The privacy considerations above become actionable when they are turned into a fixed set of questions asked of every vendor, every time, and ALA's AI guidance provides exactly that list. Rather than relying on whatever a vendor volunteers in a demonstration, which is reliably the flattering subset, the guidance directs libraries to establish, as a condition of procurement, whether and how an AI product handles patron and staff data. The value of a standing checklist is that it does not depend on a librarian remembering the right concern in the moment; it makes the concern part of the process, and a vendor's difficulty answering any single question is itself a finding worth recording.

The guidance asks libraries to identify, before adopting an AI feature, the answers to the following:

- Whether the AI features are enabled by default, and whether they can be turned off
- Whether the system collects prompts, searches, account activity, or reference interactions
- Whether that data is used for model training or product improvement
- Where the data is stored, and which subprocessors or third parties have access to it
- How long the data is retained, and what mechanism exists to delete it
- What audit rights and exit rights the library has, including what happens to the data if the contract ends

Each of these maps to a concrete risk rather than an abstract preference. For example, the default-on question matters because a feature that is enabled by default has already begun processing patron data by the time a library thinks to evaluate it, which inverts the "opt-in" model the guidance elsewhere insists on. The training question matters because a tool that improves its model on patron searches is extracting patron behavior for the vendor's benefit, a concern this curriculum's labor and privacy discussions both address. The retention and deletion question matters because a vendor that cannot say how long it keeps patron queries, or cannot delete them on request, cannot support the library's own confidentiality

obligations no matter how reassuring its marketing language is. And the exit-rights question matters because data governance does not end when a contract does; a library that has not secured the return or destruction of patron data at contract termination has left that data in the vendor's hands indefinitely.

The practical way to use this checklist is to put it in writing into the procurement process itself, so that these questions are asked during evaluation and their answers are documented alongside the pricing and feature comparisons that libraries already record. For example, a request for information or a vendor questionnaire that includes these six items produces a written record that can be compared across competing products and revisited when a contract comes up for renewal, which is far more durable than a verbal assurance given in a sales meeting and forgotten by the time it matters. Additionally, the guidance frames a vendor's unwillingness or inability to answer as informative in its own right: a vendor who cannot describe its data handling in these terms is either not examining its own product closely or is choosing not to disclose, and either reading is a reason for caution. What the checklist cannot do is make the adoption decision. Weighing a tool's documented data practices against its value for a specific patron community, and deciding whether the residual risk is acceptable, remains the professional judgment of the librarians accountable for both the data and the service.

A decision framework: adopt, pilot, or decline a vendor AI feature

The volume of vendor AI offerings makes a repeatable decision framework more valuable than any single product evaluation, because the products will change but the questions a librarian must ask about them will not. Two published frameworks anchor this well. The Choice rubric for evaluating generative AI resources, from November 2024, screens any tool on four quick questions: transparency, meaning whether citations are traceable to real sources; efficiency, meaning whether the tool augments rather than reinvents an existing workflow; critical thinking, meaning whether it preserves rather than eliminates the user's analysis; and privacy, meaning whether the policy is acceptable and personal data stays out. For example, applying these four screens to a bundled discovery assistant takes a single testing session and produces a defensible yes-or-no on each dimension, which is far more useful to administration than a general impression. Such a rubric is especially apt for a community college, because it explicitly asks the evaluator to weigh a tool against a specific student population rather than against an abstract ideal user. The ACRL AI Competencies for Academic Library Workers, approved in October 2025, provide the complementary

professional backbone, organized around ethics, knowledge, evaluation, and application, and deliberately product-agnostic so that they can be adapted as tools change.

The practical decision resolves into three outcomes, and naming them explicitly prevents the common failure of drifting into adoption by default. A library should adopt a feature when it is bundled, passes the rubric, and addresses a real workflow need, since the cost of turning on a tested bundled tool is low and the benefit is concrete. A library should pilot a feature, rather than adopt or decline it outright, when the tool is promising but unproven for the local context or when it is a paid add-on whose value is uncertain. For example, a pilot following the methodology that Empire State University and a University of California San Diego team have used runs the same set of real reference queries through the AI tool and through a non-AI search, scores both with the Choice rubric, and produces local evidence rather than relying on the vendor's claims or another institution's experience. Such a pilot converts an opaque purchasing decision into a measured one, and it generates exactly the documentation administration needs to approve or reject the spend. A library should decline a feature when it fails the rubric on a dimension that matters, when its privacy posture is unacceptable for the patron community, or when its cost cannot be justified against everything else the budget must cover.

The final element of the framework is continuous monitoring rather than a one-time verdict, because everything in this space dates quickly. The Ithaka S+R Generative AI Product Tracker, public and continuously updated since 2024, maintains a living table of these products with their purchasing models, features, and limitations, and it is more reliable as a current reference than any feature list a librarian could memorize. For example, a selector who bookmarks the tracker and re-checks it each term will catch a bundled tool moving to paid pricing, or a new privacy disclosure, far sooner than one relying on vendor outreach. Such ongoing attention should be paired with a verify-before-teaching habit, re-confirming pricing, privacy clauses, and feature scope each term, since a claim that was accurate last semester may not be accurate now. In order to keep this from becoming overwhelming, the discipline is to evaluate only the tools your library actually encounters, apply the same four-screen rubric every time, and document the decision. Such a disciplined, repeatable process is what allows a library to keep pace with a flood of vendor AI products, and the judgment about which products earn a place in the collection and the budget remains, at every step, the responsibility of the librarians who know their patrons and their resources best.

PRACTITIONER NOTE

The moment this stopped being abstract for me was a vendor demonstration of an AI research assistant, where the sales representative typed a clean question about a well-documented topic and the tool produced a confident, nicely cited summary in seconds. It looked impressive in the room. Afterward I ran the questions my students actually bring to the desk, the messy ones about half-remembered assignment topics and narrow local subjects, and the tool returned five sources where the student needed a starting point that acknowledged its own gaps, and it surfaced articles we do not even subscribe to. Neither result was wrong, exactly, but neither matched what a first-generation student writing a first research paper actually needs. Running it through the Choice rubric, transparency and privacy were the screens that mattered most for my population, and the tool was stronger on the demonstration question than on any question a real student had ever asked me. That is the gap the demonstration is designed to hide, and the only way to see it is to test the tool on your own patrons rather than the vendor's example.

KEY TAKEAWAYS

- Most of what vendors call AI in collection development is machine learning and analytics rather than generative AI, and even the genuine tools provide decision support rather than making the collection decision, which remains the librarian's judgment.
- Acquisitions and licensing AI is less mature than the marketing suggests: open access agreement tracking is automation rather than generative AI, and AI license review is at the pilot stage, useful as a second reader but never a substitute for reading the contract.
- Vendor AI search products share a retrieval-augmented-generation architecture and a common limitation: a fluent summary of only the few sources retrieved, which is why the products must be tested on your own real reference queries rather than judged by vendor demonstrations.
- The central cost question is bundled versus paid add-on; add-on pricing is opaque and undisclosed, none of these tools yet provide COUNTER statistics, and the Clarivate 2025 ebook controversy shows how fast vendor strategy and pricing can shift.
- Vendor privacy commitments vary and must be read clause by clause, with particular attention to tools that require an authenticated login and therefore tie research queries to an identified patron, and no independent privacy audit of these products yet exists.

- ALA's guidance provides a concrete vendor data-review checklist to build into procurement: whether AI is on by default and can be disabled, whether it collects prompts and reference interactions, whether data trains the model, where data is stored and who can access it, retention and deletion mechanisms, and audit and exit rights; a vendor's inability to answer is itself a finding.
- A repeatable decision framework, anchored in the Choice four-screen rubric and the ACRL AI Competencies, resolves each product into adopt, pilot, or decline, paired with continuous monitoring through the Ithaka S+R tracker and a verify-before-teaching habit.

References

APA 7TH EDITION

1. American Library Association. (2026). *Guidance on the use of artificial intelligence in libraries.* https://www.ala.org/tools/standards-and-guidelines/guidance-use-artificial-intelligence-libraries
2. Association of College and Research Libraries. (2025, October). *AI competencies for academic library workers.* American Library Association. https://www.ala.org/acrl/standards/ai
3. Choice. (2024, November). *Evaluating generative AI resources: Separating the tools from the toys.* Association of College and Research Libraries. https://www.choice360.org/tools/evaluating-generative-ai-resources-separating-the-tools-from-the-toys/
   Includes a downloadable transparency/efficiency/critical-thinking/privacy rubric. Confirm exact URL when updating the site.
4. Clarivate. (2025). *Pulse of the library 2025.* https://clarivate.com/pulse-of-the-library/
5. Ithaka S+R. (2024–2025). *Generative AI product tracker.* https://sr.ithaka.org/our-work/generative-ai-product-tracker/
   Continuously updated; confirm current URL when updating the site.
6. Maiberg, E. (2025, February 4). AI-generated slop is already in your public library. *404 Media.* https://www.404media.co/ai-generated-slop-is-already-in-your-public-library/
7. Portillo, A., & Carson, P. (2025, January). Evaluating large language models for collection development. *Journal of the Medical Library Association, 113*(1). https://doi.org/10.5195/jmla.2025.2079

ACRL AI Competencies covered

Analysis & Evaluation | Use & Application | Ethical Considerations

Sub-competencies: 3.1, 3.2, 4.2, 1.4 · ACRL AI Competencies (2025)