

Module 04

Ethics, copyright & policy

Ethics are not separate from practical AI use; they are embedded in every decision about which tool to use, what to put into it, and what to do with what comes out. This module maps the ethical landscape you need to navigate in library work.

By Yulia Brusova

30 min read · Reviewed July 2026

WHAT YOU'LL BE ABLE TO DO

- 1 Explain what hallucinations are and apply at least two strategies to detect them
- 2 Identify patron privacy obligations that apply when using AI in reference work
- 3 Describe the current copyright landscape for AI generated content and what it means for library practice
- 4 Review or contribute to an institutional AI acceptable use policy
- 5 Apply the ARL 'No Human, No AI' principle to a library workflow decision
- 6 Locate your role within the ACRL AI Competencies framework and identify one subcompetency to develop further
- 7 Apply an ethics evaluation framework to a vendor AI product before recommending or adopting it
- 8 Identify at least two scenarios in which FERPA or HIPAA constraints apply to AI use in library work

A striking feature of the ACRL AI Competencies document is how much of it is about caution. Not fear, but caution. There's a difference. This module is about developing the professional judgment to use AI in ways that serve your patrons and uphold your values.

ALA's six core values: the anchor for library AI

The most useful way to organize everything in this module, and much of what follows in the rest of the curriculum, is around the framework the American Library Association built for exactly this purpose. ALA's Guidance on the Use of Artificial Intelligence in Libraries grounds every recommendation in six core library values, and holding those six in mind gives a librarian a durable structure for reasoning about any AI decision, long after the specific tools discussed in this curriculum have been replaced by newer ones. The six values are Public Good, Intellectual Freedom, Privacy, Sustainability, Diversity, Equity, Inclusion, and Access, and Labor. For example, when a vendor offers a new AI discovery feature, the six values generate the questions worth asking almost automatically: does it serve a documented public-service purpose, does it preserve transparent and non-personalized paths to information, does it protect patron privacy, what is its environmental cost, does it widen or narrow equitable access, and what does it mean for the people who do the work. Such a

framework does not make the decision, but it ensures the decision is made against the profession's stated commitments rather than against vendor enthusiasm alone.

The single principle from this guidance that most directly governs day-to-day practice is its treatment of AI output. ALA states that libraries should treat AI outputs as "drafts requiring human review, not authoritative answers," and that "human expertise remains necessary for nuance, originality, contextual judgment, accountability, and trust." This is the sentence to internalize before any other in this module, because it settles in advance the question that causes the most professional trouble: what standing does an AI response have. The answer is that it has the standing of a draft. For example, a chatbot's answer to a reference question, a tool's summary of an article, and a model's suggested subject headings are all starting points that a librarian composes with and verifies, never finished products that a librarian forwards. Such a stance is not skepticism for its own sake; it is the recognition that accountability for anything the library provides rests with a human professional, which is the same principle the ARL formulation "No Human, No AI," discussed later in this module, expresses from a different angle.

These values also carry a commitment that librarians sometimes find surprising: that adopting AI is not always the right choice. The guidance is explicit that libraries should use AI only where it serves "clear user-centered purposes," not simply because a tool is available, and that patrons must always retain "a clear path to human assistance." For example, a library that keeps a non-AI, human-staffed route to every service it offers is not failing to modernize; it is honoring the Public Good and Labor values that the guidance places at the center of responsible practice. In order to use this framework well, a librarian does not need to memorize the guidance in detail. She needs to hold the six values as standing questions and to treat AI output as a draft, and the rest of this module fills in the specific obligations, around privacy, copyright, bias, and policy, that each value entails.

Hallucinations: the most misunderstood AI problem

Hallucination, the term for when AI confidently states something false, is not a bug in the traditional sense. It is a direct consequence of how large language models work. The model is always predicting the most plausible next token, and sometimes plausible is wrong. Module 01 covers the mechanism in depth; the professional implication for ethics practice is that hallucination is structural and persistent, not a feature that model updates will eventually eliminate. The question is not whether AI hallucinates but how to work with it responsibly given that it does.

In library contexts, the hallucination types that carry the most professional risk are the ones that look most authoritative: invented citations with plausible journal names, volume numbers, and page ranges; statistics that sound reasonable but cannot be traced to a primary source; incorrect dates for events or policies; wrong author attributions; and legal or policy characterizations that are outdated or simply wrong. Such errors are not random; they tend to occur precisely where the model has learned that seemingly authoritative specificity is expected. A hallucinated citation does not announce itself; it is formatted correctly and plausible in content. Recognizing this pattern is the first step in the professional practice of verification that Module 05 covers in full.

What hallucination is not is a reason to dismiss AI entirely. It is a reason to verify, always, and proportionally to the stakes and claim type of the output. The same critical evaluation librarians apply to any information source applies here, with the added professional knowledge of which claim types carry the highest risk in AI generated text.

Patron privacy: our professional obligation comes first

Library confidentiality is foundational to professional practice. The ALA Code of Ethics commits librarians to protecting patron privacy, and that commitment does not have a technology exception. AI tools do not automatically honor patron confidentiality; that requires deliberate professional choices about what information enters a prompt and which tools are used for which tasks.

The risks are specific. Consumer tier AI tools may log inputs and use them for model training, which means a patron's research question, entered verbatim into a free tier tool, may be processed by the vendor's servers, retained, and potentially reviewed by company employees or incorporated into future training data, depending on the tool's terms of service. Such a practice would constitute a meaningful breach of patron privacy in most library contexts, particularly for inquiries that touch on sensitive circumstances: health information, legal questions, immigration status, financial difficulty, crisis situations. Inputting any patron information that could identify the individual into an unapproved commercial AI tool is not a minor procedural lapse; it is a professional conduct matter.

In order to use AI ethically in reference and patron services, the practical guidance is straightforward. Strip all identifying information before putting any patron inquiry into AI. For example, instead of pasting a patron's verbatim question with their name and institution, describe the research need anonymously: "A nursing student is looking for peer reviewed articles on postpartum depression for a capstone project." Use institutional AI agreements that include contractual data privacy protections where they are available, rather than

consumer tier tools with opt out checkboxes. Follow your institution's IT and legal guidance on which tools are approved for which data types. And when the situation involves information sensitive enough that you are uncertain, handle the interaction without AI; the professional judgment to decline a tool in a specific context is itself an expression of the competency this module addresses.

ALA's AI guidance turns this professional norm into a concrete prohibition, and it is worth stating as a flat rule rather than a nuanced principle, because a rule is what staff can actually follow under time pressure at a busy desk.

Never enter any of the following into an AI tool your institution has not formally approved for that category of data:

- Patron names, contact information, barcodes, or any other personally identifiable information
- Borrowing, reading, or search histories
- Reference and research questions that can be traced back to an individual
- Patron account information, and staff or personnel records
- Any prompt containing identifiable details about a real person

The value of stating it this way is that it removes the moment of judgment that leads to breaches. A librarian who has internalized the rule does not weigh whether a particular patron detail is sensitive enough to withhold; she withholds all of it by default and adds only the anonymized description the task actually requires. Such a default is the practical form of the ALA Code of Ethics commitment to patron privacy, and it is the single habit that most reliably keeps AI use inside the profession's confidentiality obligations.

Copyright and AI generated content

The copyright landscape for AI generated content is unsettled in ways that will likely require years of litigation and legislation to resolve, but several positions are clear enough to guide current professional practice. In order to navigate this landscape responsibly, librarians need a working understanding of two distinct copyright questions: who, if anyone, holds copyright in AI generated outputs, and what copyright obligations arise from the training data that AI models were built on.

On the first question, the U.S. Copyright Office has been consistent: copyright requires human authorship, and content generated by AI without significant human creative input is not copyrightable. For library work, this means that AI drafted communications, LibGuide content, or instruction materials that were generated by AI and lightly edited carry at most partial copyright protection for the human authored elements. Such a position matters when libraries produce content intended for publication, grant applications, or official institutional documents; the authorship question is less critical for patron facing materials where accuracy matters more than attribution, but it is professionally relevant for any content the library publishes in its own name.

On the training data question, the situation is actively litigated. Major AI models were trained on copyrighted materials, specifically books, articles, and other published works, collected without explicit license from rights holders. The New York Times filed suit against OpenAI in late 2023, multiple authors have brought collective action suits, and the outcomes of these cases will shape how AI generated outputs can be used commercially. For current library practice, the implications are still developing. Being transparent about AI use when it is relevant to publication or attribution, following emerging institutional disclosure policies, and documenting AI assisted workflows for content produced under grant funding or for official publication are all defensible professional practices while the legal landscape continues to evolve. The ARL Guiding Principles (2024) address this directly: libraries have a professional stake in preserving copyright flexibility for nonprofit research and educational uses, which means the training data copyright debate is not merely a legal matter for AI companies to resolve but a professional advocacy issue for the library field.

The ARL Guiding Principles, and 'No Human, No AI'

The Association of Research Libraries published seven guiding principles for AI in April 2024. They're worth reading in full. The principle that proves most consistently clarifying for library practice is Principle 4:

"No Human, No AI": consequential decisions must not be delegated entirely to AI systems. Human involvement is essential at critical decision making junctures.

For library practice, this principle is clarifying. AI can help you draft, research, brainstorm, and organize. But decisions that affect patrons, including what information to provide, how to support someone in crisis, whether to escalate a situation, and how to evaluate a collection, require human professional judgment.

This isn't about AI being untrustworthy. It's about recognizing that accountability requires a human. When something goes wrong, and sometimes it will; there needs to be a professional who made a decision and can answer for it.

Diligence as a professional practice

Dakan and Feller frame responsible AI use as a competency they call diligence - shifting the question from what AI can do to what professionals are responsible for when they use it (Dakan & Feller, *AI Fluency: Framework & Foundations*, Anthropic Academy, 2025). The framework organizes this into three layers that map directly onto library practice. The first is **creation diligence**: before beginning any AI-assisted task, the questions to ask are which system is being used, how it handles the data being entered, and whether institutional policies permit this particular use. For librarians handling patron data or confidential institutional information, this is not a one-time question but a habitual check that should precede each new use case. The second layer is **transparency diligence**: who needs to know that AI was involved in this work, and how explicitly? The standard is not what is technically required but what maintains the trust of the people affected. For example, a committee report drafted with significant AI assistance calls for a different disclosure than a patron email lightly proofed by the AI. Such decisions require judgment, not formulas. The third layer is **deployment diligence**: when AI-assisted work is shared - with a patron, a colleague, or the public - the professional who deployed it is accountable for its accuracy and appropriateness. The AI is not. In order to meet the same standards that would apply to fully human-authored work, facts must be verified, bias checked, and AI output treated as a draft requiring professional sign-off before it goes anywhere.

Institutional AI policies: reading and contributing

Your institution either has an AI policy, is developing one, or should be. Either way, librarians have a professional role to play, not merely as policy followers but as the professionals best equipped to understand what responsible information use requires.

Reading a policy: When reviewing an institutional AI policy, librarians should look for the following elements. First, which tools are explicitly approved or prohibited, and whether the approval list is current, given how rapidly the tool landscape changes. Second, what categories of data may and may not be input into AI systems, and whether patron related data is specifically addressed. Third, whether there are disclosure requirements for professional communications or publications that used AI assistance, and how those requirements apply to library produced content such as LibGuides, research handouts, and

instruction materials. Fourth, who holds institutional responsibility for AI related decisions and incidents, and whether the library has a designated point of contact for AI policy questions.

Contributing to policy: Librarians have professional expertise that is directly relevant to AI policy development, including information ethics, intellectual property, patron privacy, and information literacy, and are rarely at the table when institutional policies are drafted. For example, an institutional AI policy written without library input may fail to address the specific circumstances of reference interactions, patron data protection, or the use of AI in instruction settings. In order to ensure that library concerns are represented, librarians should proactively request inclusion in AI policy working groups, submit written comments when policies are circulated for review, and position the library as the natural institutional home for information governance questions that AI use raises.

ACRL subcompetency 2.5 specifically asks us to "understand AI related policies and regulations." At the professional level, this means both following them and helping to shape them. Module 11 covers the advocacy dimension of this work in considerably more depth, including how to make the case for library leadership on AI governance to administrators who may not have considered the library's role.

The ACRL AI Competencies framework: your professional roadmap

In October 2025, ACRL published the AI Competencies for Academic Library Professionals, the first framework from the national association to define what professional AI competency looks like for librarians. For example, this is the document that provides the structure for this entire curriculum: each module is mapped to specific subcompetencies, and the three levels of this learning pathway correspond to the framework's progression from foundational knowledge through applied practice to advanced leadership.

The ACRL framework is organized into five competency domains. The Knowledge domain addresses understanding what AI is, how it works, and its limitations, the territory covered in Modules 01 through this one. The Application domain addresses using AI tools effectively for library tasks, including prompting, tool selection, and workflow integration, as covered in Modules 05 through 10. The Ethics domain addresses professional responsibilities around privacy, bias, intellectual property, and accountability, which is the core of this module. The Analysis domain addresses critical evaluation of AI output, AI systems, and AI policy, woven throughout the applied modules. The Leadership domain addresses advocacy, institutional

strategy, and the library's role in shaping AI use across the broader institution, which is the focus of Modules 12 through 18.

For a practicing librarian, the most useful function of the ACRL framework is not as a credential checklist but as a professional self assessment tool. For example, a reference librarian who feels confident in Knowledge and Application competencies but is uncertain about how to approach the Leadership domain now has a published professional standard to anchor her development planning, and can point to that standard when making a case for professional development time or funding. Such alignment between individual professional development and an ACRL endorsed framework is significant in institutional conversations about the value of library AI training. The full ACRL framework document is freely available at acrl.ala.org; it is a document worth reading in its entirety rather than relying on summaries, including this one.

FERPA, HIPAA, and sensitive institutional data in AI workflows

The patron privacy obligations covered earlier in this module become substantially more complex when specific federal privacy laws apply. Two are particularly relevant for academic library practice: FERPA (the Family Educational Rights and Privacy Act) and, in health sciences library and hospital library contexts, HIPAA (the Health Insurance Portability and Accountability Act). In order to work with AI responsibly in these contexts, librarians need a practical understanding of what these frameworks require, not at the level of legal expertise, but at the level of professional judgment about what information should and should not enter an AI workflow.

FERPA protects education records of students who are enrolled at institutions receiving federal funding, which includes virtually every college and university in the United States. Education records include grades, transcripts, financial aid information, disciplinary records, and any document that contains personally identifiable information about a specific student. For example, if a librarian receives a request from a faculty member that includes a student's name alongside details about the student's research challenges, academic standing, or accommodation status, that communication may contain FERPA protected information. Inputting such information into a commercial AI tool, even a paid institutional plan, without explicit institutional authorization for that data category would constitute a FERPA concern, not merely a professional privacy preference. The determination of which AI tools are FERPA compliant for which data types belongs to the institution's compliance office, not the individual librarian.

HIPAA applies in health sciences library contexts, including medical school libraries, hospital libraries, and any library serving clinical environments where patient information may be referenced. HIPAA's minimum necessary standard requires that only the information necessary for a specific purpose be accessed or disclosed, which has direct implications for AI prompts. For example, a clinical librarian assisting a nurse with a literature search on a patient's condition should not include any patient identifiers or case specific clinical details in an AI prompt, regardless of how helpful such context might seem. Such information is protected health information (PHI) under HIPAA, and its entry into a commercial AI tool would constitute a potential breach even when the intent is purely professional.

In order to navigate these requirements without institutional legal training, librarians should establish a simple default: when uncertain whether information is subject to FERPA or HIPAA, do not include it in an AI prompt. Additionally, librarians should consult their institution's privacy officer or general counsel for guidance on which AI tools have been assessed for compliance with applicable federal privacy laws, and should document that consultation, since having sought guidance is itself evidence of good faith professional conduct.

Evaluating vendor AI tools through an ethics lens

Library vendors are integrating AI into the systems librarians manage daily, including catalog discovery, research databases, interlibrary loan platforms, digital repository software, and learning management system integrations. For example, a discovery layer with an AI powered "research assistant" feature that suggests sources and summarizes article abstracts is now available from multiple major vendors, and librarians are increasingly expected to evaluate and configure these features without clear guidance from their institutions. In order to fulfill this responsibility with professional rigor, librarians need an ethics evaluation framework that can be applied to vendor AI features, not just general purpose tools.

Several questions are essential to this evaluation. First: where does the AI feature process data, and does that processing happen on the vendor's servers, on third party AI provider servers, or within the institution's infrastructure? The answer affects what data governance agreements apply and whether patron interaction data leaves the institution's control. For example, a discovery layer that sends patron search queries to an external AI provider for summary generation may be processing patron search behavior in ways that are not covered by the library's existing agreement with the discovery layer vendor. Second: does the vendor's AI feature have a training data arrangement that raises intellectual property concerns? Several vendors have faced questions about whether their AI tools were trained

on licensed library content, which could implicate the terms of the institution's database agreements.

Third: what happens when the AI feature produces incorrect information and a patron relies on it? Vendor marketing rarely addresses error liability directly, but this is a question that should be asked before deployment, and the answer should inform how the feature is presented to patrons. Fourth: does the vendor provide transparency about which AI model or models power the feature, what training data was used, and how the outputs are grounded or verified? Such transparency is not universal and its absence should be treated as a flag, not a norm. Additionally, librarians should consult LITA/CORE, LTI Choice360, and Library Technology Reports for community sourced assessments of specific vendor AI features, since independent professional evaluation is more reliable than vendor documentation in a rapidly evolving market. The ARL statement on principles for vendor AI transparency (2024) is also a useful framework for structuring vendor conversations and contract negotiations.

New policy anchors: the ALA guidance debate, copyright consensus, and the research library outlook

Three policy documents have entered the landscape since the rest of this module was framed, and each speaks directly to the foundations laid above: the privacy framework, the copyright landscape, and the strategic horizon for research libraries.

The first is ALA's own. Following a public comment period in the spring of 2026, ALA adopted its "Guidance on the Use of Artificial Intelligence in Libraries" at the June 2026 Annual Conference, and it is now the association's stated framework rather than a draft; its six core values anchor the opening section of this module. Two of its commitments are worth committing to memory, because they make explicit principles this module has already been building toward. The guidance states that "libraries will not adopt AI tools that fail to protect individuals' privacy, personal data, and security," which takes the patron privacy commitment covered earlier in this module and reframes it as a procurement standard rather than a best practice. It also requires that libraries obtain "algorithmic transparency" from AI vendors, which extends the vendor evaluation questions raised above into a formal professional expectation: a vendor's unwillingness to explain how its AI feature was trained and how its outputs are grounded is not merely a yellow flag but a violation of an emerging professional norm. For example, a librarian negotiating a discovery layer contract that bundles in a new AI summarization feature can point to this language as evidence that the profession expects vendor transparency as a baseline condition of doing business with libraries.

The second development resolves, at least for now, one corner of the copyright uncertainty discussed earlier in this module. In October 2025, the Library Copyright Alliance, which speaks for ALA, ACRL, and ARL together, issued version 4.0 of its "Statement on Copyright and Generative Artificial Intelligence." Its position is direct: existing U.S. copyright law, including the fair use doctrine libraries have relied on for decades, is adequate to address generative AI, and no new legislation is required in order to resolve the training data and output questions raised earlier in this module. Such a position does not settle the pending litigation; courts will still decide individual cases on their merits. What it does give librarians is a clear professional reference point when colleagues, faculty, or administrators ask whether new "AI law" is coming that will change how libraries operate. The field's collective position, through LCA, is that the legal tools already exist and what is needed is correct application of fair use principles to new technology, not a new statute.

The third document widens the lens considerably. In March 2026, ARL and the Coalition for Networked Information published "Strategic Implications of AI Futures for Research Libraries," a report that moves past day to day policy questions in order to ask what research libraries need to anticipate over the next five to ten years as AI reshapes scholarly communication, research workflows, and the role of the library within the research enterprise. For example, the report considers how AI assisted literature review tools may change what researchers expect from library research support, and how AI's growing role in data curation and research data management may shift staffing and skill requirements within research libraries. This report belongs on the reading list of any librarian whose subcompetency development plan, drawing on the ACRL framework discussed earlier in this module, includes the Leadership domain; it is squarely the kind of strategic document the later Leadership modules will draw on.

None of these three documents changes the core judgment this module has argued for from the start. A draft guidance document, however carefully worded, does not evaluate a specific vendor contract; a copyright statement, however authoritative, does not decide whether a specific use of a specific tool is fair use in a specific library's circumstances; and a strategic report, however well researched, does not tell a specific library what to prioritize this year. Reading these documents, understanding their implications, and translating them into decisions that fit a particular institution's mission, resources, and risk tolerance remains a professional judgment that no policy document, however good, can make for us.

PRACTITIONER NOTE

Librarians are frequently asked to help faculty develop AI policies for their syllabi, and increasingly to help institutions review draft AI acceptable use policies before they go to senior leadership. The productive approach in both contexts is the same: start with the ARL principles and the ACRL framework, identify which subcompetencies the policy needs to address, and then adapt the language to the specific audience and context. Librarians are natural policy translators: they understand both the information landscape and the professional obligations embedded in it. Making the ACRL framework and the ARL Guiding Principles standard references in any policy conversation the library enters, whether as a reviewer, a contributor, or an advocate, gives librarians professional standing in discussions that too often happen without them.

KEY TAKEAWAYS

- ALA's AI guidance is built on six core values (Public Good, Intellectual Freedom, Privacy, Sustainability, DEIA, and Labor) and directs libraries to treat AI output as drafts requiring human review, not authoritative answers.
- Never enter patron PII, borrowing or reading histories, reference interactions, or patron and staff records into an AI tool your institution has not formally approved for that data; withhold all patron detail by default and add only anonymized context.
- Patron privacy obligations extend to AI prompts: strip all identifying information before including any patron context, regardless of the tool.
- The ARL principle 'No Human, No AI' is the clearest professional guide: consequential decisions affecting patrons require human judgment.
- AI generated content is generally not copyrightable in the U.S.; training data copyright is actively litigated and professionally relevant.
- FERPA and HIPAA constrain which patron and student data may enter AI workflows; consult your compliance office, not the vendor's marketing.
- The ACRL AI Competencies framework (2025) maps professional development across five domains: Knowledge, Application, Ethics, Analysis, and Leadership.
- Librarians have standing in institutional AI policy development; information ethics, patron privacy, and information literacy are core library expertise.
- ALA's draft 'Guidance on the Use of Artificial Intelligence in Libraries' (April 2026) is pending Council action as of this writing, but its privacy and vendor transparency principles are worth tracking now; the Library Copyright Alliance (October 2025) holds

that existing copyright law is adequate for generative AI, and ARL/CNI's March 2026 report maps the strategic AI outlook for research libraries.

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. American Library Association. (2021). *ALA code of ethics*. <https://www.ala.org/tools/ethics>
3. Association of College and Research Libraries. (2025, October). *AI competencies for academic library workers*. American Library Association. <https://www.ala.org/acrl/standards/ai>
4. Association of Research Libraries. (2024). *Research libraries guiding principles for artificial intelligence*. <https://www.arl.org/resources/research-libraries-guiding-principles-for-artificial-intelligence/>
5. *The New York Times Company v. Microsoft Corporation*, No. 1:23-cv-11195 (S.D.N.Y. filed Dec. 27, 2023).
Legal citation follows Bluebook convention.
6. U.S. Copyright Office. (2023, March 16). *Copyright registration guidance: Works containing material generated by artificial intelligence*. 88 Fed. Reg. 16190. https://www.copyright.gov/ai/ai_policy_guidance.pdf
7. U.S. Copyright Office. (2025, January 29). *Copyright and artificial intelligence, Part 2: Copyrightability*. <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf>

ACRL AI Competencies covered

Ethical Considerations

Knowledge & Understanding

Sub-competencies: 1.1, 1.2, 1.3, 1.4, 1.5, 2.2, 2.5 · ACRL AI Competencies (2025)