Level 1: Foundations | All Librarians

Module 01

What AI actually is

AI tools feel like magic until something goes wrong. Understanding how they actually work, at a nontechnical level, changes how you use them. This module gives you the mental model you need to work with AI effectively and critically.

By Yulia Brusova

20 min read · Reviewed July 2026

© 2026 · Licensed under CC BY-NC-SA 4.0

WHAT YOU'LL BE ABLE TO DO

1. Explain in plain language what a large language model is and how it generates text
2. Distinguish AI from search engines and databases in terms of how they retrieve and construct information
3. Explain why AI generates false information, and apply at least two verification practices that reduce this risk
4. Identify at least three things AI does reliably and three things it does not
5. Explain why AI output is probabilistic rather than deterministic

> *Librarians who begin using AI often approach it the way they approach a search engine: type in a question, expect a correct answer. This approach does not serve them well, not because the tool itself is inadequate, but because it rests on a misunderstanding of what AI is actually doing. In order to use AI effectively in library work, one must first develop a foundational understanding of how it functions, not at the level of computer science but at the level of professional practice. Once that understanding is in place, practice changes entirely, and these tools become genuinely explainable to colleagues and patrons in terms that actually help them.*

AI is a prediction machine, not a knowledge database

When a librarian types a question into ChatGPT or Claude, the model does not search a database or retrieve stored facts. Instead, it predicts what text should come next, based on statistical patterns learned from billions of documents during a training process. For example, if a librarian asks about the history of interlibrary loan, the model draws on patterns from everything written about that topic in its training data, including journal articles, library blogs, textbooks, and websites, and constructs a response by predicting what a plausible and informative answer would look like. The model is an extraordinarily sophisticated pattern matcher, but it is not retrieving stored facts in the way a database retrieves records.

To understand how this works at a slightly deeper level, it is useful to know how AI processes language. Models do not read words; they read tokens, which are fragments of text roughly corresponding to three quarters of a word on average. For example, the word "cataloging" might be a single token, while "interlibrary" might be processed as two. The model converts every token in a prompt into numerical representations, processes those numbers through

layers of mathematical operations called a transformer architecture, and then generates a response token by token, each one selected based on the statistical probability of what should follow given everything that preceded it. Such a process produces remarkably fluent text because fluency, grammatical coherence, and topic relevance are precisely what the model learned to maximize during training.

Such a distinction matters considerably for library practice. A database returns records; a search engine returns links; AI generates text that sounds plausible based on what it has learned. Plausible is not the same as accurate, and this difference has significant implications for how librarians integrate these tools into their professional work. Understanding the mechanism, prediction rather than retrieval, is the single most important conceptual foundation for using AI well.

How AI learns: training data and its implications

Before a model can predict text fluently, it must learn from an enormous quantity of text. This process, called pretraining, involves exposing the model to hundreds of billions of words from books, websites, academic papers, code repositories, and other text sources, then adjusting the model's internal parameters until it becomes very good at predicting what comes next in any given sequence. For example, GPT-4 was trained on text collected from the internet through early 2023, including large web crawls, books, Wikipedia, and code from public repositories. Claude models are trained on a similar range of sources with different emphasis. Such training datasets are assembled by AI companies and are generally not disclosed in complete detail.

Several implications of this training process matter directly for library practice. First, training data has a cutoff date, the point after which no new information was incorporated. For example, a model trained through early 2024 has no knowledge of research published, events that occurred, or databases that changed after that date. This means that any question requiring current information will produce outdated or fabricated responses unless the user provides that information directly in the prompt. Additionally, training data reflects whatever was written on the internet, which means it reflects the biases, assumptions, and errors present in that broader written record.

Second, training data for major AI models included a substantial quantity of copyrighted material, specifically books, articles, and other published works, collected without explicit permission from rights holders. This has produced significant legal challenges; the New York Times filed suit against OpenAI in late 2023, and multiple authors have brought collective action suits over training data use. For librarians working in institutions with active intellectual

property policies, this is not merely background context. It is information relevant to professional decisions about which AI tools to recommend and how to discuss AI with faculty and students engaged in original research.

Furthermore, after pretraining, models undergo a process called fine tuning and reinforcement learning from human feedback, in which human reviewers rate model outputs and those ratings are used to adjust the model toward more helpful, accurate, and seemingly safe responses. Such refinement is why modern models tend to decline harmful requests and acknowledge uncertainty, but it does not make them more accurate about facts. It makes them better at presenting information in ways that seem trustworthy, which can paradoxically increase the risk of accepting seemingly plausible errors without verification.

The environmental cost of what AI does

The training and prediction processes described above are not free, and a complete picture of what AI actually is has to include what it costs the physical world to run. Training a single large model consumes an enormous quantity of electricity, and every query a patron sends to a model afterward draws additional energy from a data center that must be powered and cooled around the clock. For example, researchers estimating the water footprint of AI have found that the cooling and electricity behind a modest exchange with a large model can consume on the order of a bottle of fresh water, because data centers use vast amounts of water to keep their servers from overheating (Li et al., 2023). Such costs are invisible at the keyboard, which is exactly why they are easy to overlook, and a librarian teaching AI literacy is well positioned to make them visible rather than let them stay hidden.

Three costs are worth naming specifically. The first is energy and the carbon that accompanies it: the International Energy Agency projects that electricity demand from data centers, driven substantially by AI, is rising steeply, and the carbon consequence depends heavily on whether that electricity comes from clean or fossil sources. The second is water, the cooling cost just described, which is a particular concern where data centers sit in already water-stressed regions. The third is electronic waste, because the specialized hardware that runs AI has a short useful life and is replaced rapidly, and researchers have projected that generative AI could add substantially to the world's e-waste stream over the coming years (Wang et al., 2024). Such costs are real, cumulative, and unevenly distributed, falling hardest on the communities where the infrastructure is built rather than on the people sending the prompts.

A practical implication follows directly, and it is one librarians can act on. The largest general-purpose models are the most expensive to run in every one of these dimensions, and many

library tasks do not require them. For example, summarizing a document a librarian pastes in, or drafting a routine email, can often be done well by a smaller, task-specific model that consumes a fraction of the energy of a frontier model, and choosing the smaller tool when it is sufficient is both an environmental decision and a professional one. Such a choice reflects the sufficiency principle that the American Library Association's AI guidance names among its sustainability recommendations, and matching the tool to the task rather than reaching for the largest model by default is exactly the kind of judgment that remains with the librarian who understands both the work and its cost.

Why AI says things that are not true

Hallucination, the term used when AI generates false information with apparent confidence, is not a bug that will eventually be fixed. It is a structural consequence of how language models work. For example, if a librarian asks Claude to provide three peer reviewed sources supporting a particular argument about digital preservation, the model may generate three citations that look entirely plausible, with realistic author names, recognizable journal titles, and credible publication years, that do not actually exist. The model was not attempting to deceive; it was doing what it always does: predicting what a plausible response would look like. A seemingly convincing citation is, statistically, a plausible set of tokens given a question about sources. Such fabrications are particularly dangerous in library contexts precisely because they look authoritative.

It is evident that the model does not know what it does not know. Unlike a human expert who can say "I am not certain about that; let me check," the language model has no mechanism for distinguishing between information it learned reliably and text it is generating probabilistically. Both are produced by the same prediction process. The model may add disclaimers such as "I should note that I am not certain about this," but these disclaimers are also generated probabilistically. They appear when the model has learned that disclaimers are appropriate in certain contexts, not because the model has verified the accuracy of its output.

In order to protect library patrons and maintain the professional credibility that libraries depend on, librarians must verify any factual claims, statistics, and citations produced by AI before incorporating them into professional work. There is no exception to this rule, regardless of how confident or well sourced the response appears. Additionally, it is useful to tell patrons directly: AI is not a research tool; it is a drafting and thinking tool, and anything it tells you about facts requires verification with a real source. Such framing helps patrons

understand why the librarian's role in source evaluation remains indispensable, rather than diminished, by the presence of AI.

Dakan and Feller offer an analogy useful for patron instruction: hallucination is like a friend who recounts a story with absolute confidence while getting the details entirely wrong (Dakan & Feller, AI Fluency: Framework & Foundations, Anthropic Academy, 2025). Such a framing is useful precisely because it removes any implication of deception - the model is not lying; it is doing what it always does, predicting the next plausible word. The result is indistinguishable in tone from accurate output, which is exactly why verification is not an optional precaution but a non-negotiable professional step, regardless of how authoritative a response appears.

Why the same question gets different answers

AI responses are probabilistic in nature. Each time a question is posed, the model samples from a distribution of probable next tokens. It does not deterministically select the single "best" response but rather draws from a range of plausible options weighted by probability. For example, a reference librarian who asks Claude to draft a database instruction email on Monday may receive a response with different emphasis, different examples, and different phrasing than the same prompt produces on Friday. Indeed, this variability can occur even within a single conversation when a prompt is resubmitted without alteration.

Such unpredictability has direct implications for library practice. One cannot treat AI output as a stable, citable source, nor assume that because AI produced a particular answer once, it will produce the same answer again. Furthermore, the same prompt submitted to different AI tools will produce different outputs, because each model has different training data, different architectural choices, and different fine tuning, all of which affect how it weights plausible responses.

It is useful to think of AI as a well read colleague who may explain the same topic differently each time one asks, and whose responses must therefore be evaluated on their own terms rather than assumed consistent. This framing also helps explain to patrons why they cannot simply accept AI output as authoritative: not because AI is always wrong, but because the same question does not reliably produce the same answer, and there is no way to know in advance which response is more accurate without independent verification.

What AI does well and what it does not

There is no doubt that AI performs certain tasks reliably and others poorly, and understanding this distinction is essential for integrating these tools into library workflows effectively. Conflating these two categories by assuming that AI is either uniformly capable or uniformly unreliable produces professional errors in both directions.

AI demonstrates consistent strength in the following areas:

- **Drafting and editing:** emails, lesson plans, LibGuides, patron facing text, policy documents. For example, a library instruction email that would take thirty minutes to draft can often be generated and refined to a usable state in under ten.
- **Summarizing:** long documents, dense reports, or complex policy texts in accessible language for patrons unfamiliar with the subject matter.
- **Generating options and variations:** for example, five different ways to explain a database search to a first year student versus a graduate researcher, or three alternative framings of a research question.
- **Explaining complex topics** in plain language for patrons unfamiliar with scholarly conventions, library terminology, or database structures.
- **Brainstorming and ideation** at the planning stage of a workshop series, a library assessment, or an instruction program.
- **Working with text the user provides:** editing, restructuring, and improving documents that the librarian pastes directly into the prompt. This is the highest reliability mode of AI use and should be the default for most library workflows.

However, AI regularly fails in ways that carry significant risk for library practice:

- **Specific facts, dates, and statistics:** the model will fabricate these confidently and without acknowledgment.
- **Current events after its training cutoff date:** the model has no knowledge of research published, policies enacted, or events that occurred after training ended.
- **Precise citations:** AI generates seemingly credible references that often do not exist, as discussed above.
- **Anything requiring verified, authoritative retrieval:** library catalogs, database records, institutional policies, and accreditation requirements must be retrieved from authoritative sources.

- **Recognizing the limits of its own knowledge:** the model does not know what it does not know, which means it will often answer questions it should decline.

Such patterns make the professional role of the librarian indispensable. AI functions as a drafting and thinking partner, not as a reference source, and this distinction must inform every decision about how and when to use it.

The three types of AI you will encounter

It is useful to distinguish three broad categories of AI, because they function differently and raise different professional considerations for library practice.

Generative AI creates new content: text, images, audio, video, and code. ChatGPT, Claude, Gemini, and Perplexity all fall into this category. This is what most librarians are currently experimenting with, and it is the primary focus of Levels 1 and 2 of this curriculum. For example, when a librarian uses ChatGPT to draft a LibGuide introduction or Claude to summarize a lengthy accreditation report, that is generative AI. Such tools are powerful for drafting, editing, and ideation, but they require the verification practices discussed above because they generate rather than retrieve.

Predictive AI does not create new content but makes recommendations and classifications based on patterns. For example, the "you may also like" systems embedded in discovery layers and integrated library systems represent predictive AI that libraries have used for years, often without describing it as AI at all. Spam filters, recommendation engines in library discovery systems, and automated metadata enrichment tools all fall into this category. Such systems are familiar, if not always recognized as AI, and they raise different questions than generative AI does, primarily about algorithmic bias, data quality, and the transparency of automated decisions.

Agentic AI takes autonomous actions: it does not simply respond to prompts but executes multistep tasks with limited human intervention, browsing the web, writing and running code, sending emails, and interacting with other software systems on behalf of the user. This is a newer and rapidly evolving category, covered in depth in Module 15. Additionally, understanding how agentic AI differs from generative AI is becoming increasingly important for librarians involved in workflow automation, systems integration, and institutional AI policy decisions, since agentic systems raise considerably higher stakes for oversight and accountability than tools that simply generate text.

Two terms you'll keep encountering: RAG and agentic AI

Two terms appear constantly in current discussions of library AI, and ACRL's AI Competencies for Academic Library Workers (October 2025) names both explicitly within its Knowledge and Understanding competency category, which makes them foundational vocabulary for this entire curriculum.

The first term is retrieval-augmented generation, commonly abbreviated RAG. In a RAG system, the model does not generate a response purely from what it learned during training, the prediction process described earlier in this module. In order to ground its output in something more reliable than training data alone, it first retrieves relevant documents or passages from a defined knowledge base, and then generates its response based on those retrieved sources. Such grounding is what distinguishes a RAG system from pure prediction: the output is anchored to documents that actually exist, not only to patterns learned during training. For example, Ex Libris's Primo Research Assistant works this way: it retrieves candidate records from its discovery index before generating a response, which is why its answers come with citations to records a patron can actually open. Module 03 covers several library specific tools built on this principle.

The second term is agentic AI, introduced just above. ACRL defines agentic systems as those that "set goals, plan tasks, and act with minimal guidance," a meaningful step beyond a chatbot that only responds to one prompt at a time. Module 15 explores agentic AI in depth; for now, the working distinction worth holding onto is this one: a chatbot answers, an agent acts.

Three ways librarians will engage with AI

A practical distinction from Dakan and Feller's AI Fluency: Framework & Foundations course is the separation of AI engagement into three modes (Dakan & Feller, Anthropic Academy, 2025). The first is **automation**: the task is defined, the AI executes it. For example, a librarian might ask the AI to draft a patron-facing email about new database access, summarize a long policy document for staff, or generate subject heading options for a record under review. Such tasks work well when the outcome is clearly defined and the professional knows exactly what output is needed. The second mode is **augmentation**, in which the librarian works through a problem together with the AI rather than handing the task off entirely. For example, when developing a new information literacy workshop, the AI can serve as a thinking partner - testing different framings, pushing back on assumptions, or generating contrasting approaches before a direction is chosen. The AI does not do the work; it makes the work better. The third mode is **agency**, in which an AI system acts within parameters

established in advance - a workflow that routes incoming research requests, or a chatbot that handles common after-hours reference questions before a librarian reviews them. Such configurations require the most careful setup and the clearest understanding of what the AI may and may not do independently. Recognizing which mode a given task calls for is itself a professional judgment that no AI can make on the librarian's behalf.

How AI differs from search engines and databases

Librarians are among the professionals best positioned to understand why AI is not a search engine, because they already understand what search engines are and how they differ from databases. For example, when a student asks "Can I just use Google?" about library databases, librarians explain clearly why they cannot for certain research purposes. The same professional clarity is valuable when explaining to patrons and colleagues why AI is not Google, either.

The fundamental difference is as follows. A **database** stores records and retrieves them in response to queries. When a librarian searches JSTOR for articles on interlibrary loan, JSTOR returns actual articles that exist, were published, and are accessible in full text. The database is an index pointing to real documents. A **search engine** crawls and indexes the web and returns links to pages that exist at the time of crawling. The results are not always accurate, current, or authoritative, but they point to real, externally verifiable pages. An **AI language model** generates text by predicting probable continuations. It does not retrieve records or links; it produces new text. There is no document behind the response. Such a distinction is not intuitive for users who have spent their entire lives seeking information through retrieval systems.

In order to help patrons and colleagues understand this distinction clearly, one framing that proves consistently useful is this: a database is a library stacks, organized and retrievable; a search engine is a map of the stacks, helpful for navigation; an AI language model is a knowledgeable colleague who has read a great deal and can discuss what they have read, but who may misremember, conflate, and occasionally confabulate, and whose memory ends at a specific date. Such a colleague is valuable for certain purposes and wholly unreliable for others. Knowing which is which is the essential professional judgment.

Furthermore, this comparison clarifies the appropriate role of AI in library instruction. Librarians who teach information literacy should not simply add AI to existing instruction about databases and search engines as though it were another tool in the same category. It represents a fundamentally different kind of information system and requires a

correspondingly different evaluative framework, one that this curriculum, and the ACRL AI Competencies it is built on, is designed to provide.

A word on hype and skepticism

Interestingly enough, it is useful to hold two seemingly contradictory positions simultaneously: AI tools are genuinely useful for library work at this moment, and they are also overhyped in ways that create real professional risks. There is no doubt that neither uncritical enthusiasm nor reflexive skepticism serves librarians well in this environment. Both positions, taken alone, prevent the kind of calibrated professional judgment that good practice requires.

The claims made for AI in library contexts range from reasonable to extravagant. On the reasonable end: AI does help with drafting, editing, and explaining. On the extravagant end: AI will replace reference librarians, revolutionize cataloging overnight, or solve information literacy problems that decades of library instruction have not. For example, a 2024 Clarivate survey of academic librarians found that the majority had experimented with AI tools in their professional work, but the same survey found significant disagreement about which applications were genuinely valuable and which were being adopted because of institutional pressure rather than demonstrated benefit. Such variation is typical of early adoption periods and argues for measured, evidence informed practice rather than wholesale enthusiasm or wholesale resistance.

The ACRL AI Competencies framework identifies skepticism as a guiding mindset alongside curiosity, and that pairing is deliberate. We are meant to explore and question at the same time. Such dual orientation is precisely what librarians have always brought to information evaluation: the capacity to engage with new sources while maintaining the critical apparatus that distinguishes useful information from misleading information. AI does not require a new professional disposition, but it does require applying an existing one to a new domain.

One framing that proves consistently useful when introducing AI to skeptical colleagues is thinking of it as an amplifier rather than a replacement. The professional value AI delivers depends directly on what the practitioner brings to the interaction: the domain knowledge, the institutional context, the understanding of what a patron actually needs. A prompt written by a librarian who knows her collection, her student population, and her professional obligations produces a categorically different result than the same general prompt written by someone without that knowledge. Such a framing corrects the assumption underlying both enthusiastic overclaiming and reflexive dismissal, namely that AI is doing something

independent of the professional using it. It is not. The quality of AI output in professional library work is a function of the expertise the librarian brings to the conversation.

In order to develop this calibrated stance, treating AI as a capable but unreliable research assistant is the most useful frame. One would use a capable assistant. One would also verify their work. One would not send the assistant's output directly to a patron without review. This frame guides practice more reliably than either enthusiasm or resistance alone, and it maps cleanly onto the professional standards that library workers already hold.

PRACTITIONER NOTE

A common misconception among students is that AI tools like ChatGPT are searching the internet when they generate a response. This misunderstanding leads to two predictable errors: treating AI output as a search result rather than generated text, and assuming that AI has access to current information. An effective way to address this in instruction sessions is to demonstrate the distinction directly: showing the same question posed to Google, to a library database, and to ChatGPT, and asking students to identify what is different about the third response. Such a demonstration takes approximately five minutes and consistently produces a measurable shift in how students evaluate and use AI output for the rest of the session.

KEY TAKEAWAYS

- AI generates text by predicting probable next tokens. It is not retrieving stored facts from a database.
- Hallucinations are structural, not bugs: specific claims, statistics, and citations always require independent verification.
- AI output is probabilistic: the same question can produce different answers in different sessions.
- Training data has a cutoff date; AI has no knowledge of events, publications, or policy changes after that point.
- AI carries a real environmental cost in energy, carbon, water, and e-waste; choosing a smaller task-specific model when it is sufficient is both an environmental and a professional decision.

- AI is reliable for drafting, summarizing, and brainstorming; unreliable for citations, current facts, and verified retrieval.
- The most important professional frame: think of AI as a capable but unreliable assistant whose work always needs review.

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. Clarivate. (2024). *Pulse of the library 2024.* https://doi.org/10.14322/pulse.of.the.library.2024
3. Dakan, R., & Feller, A. (2025). *AI fluency: Framework & foundations* [Online course]. Anthropic Academy. https://anthropic.skilljar.com/ai-fluency-framework-foundations
4. *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.
5. Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). *Making AI less "thirsty": Uncovering and addressing the secret water footprint of AI models.* arXiv:2304.03271. https://arxiv.org/abs/2304.03271
6. Wang, P., Zhang, L. Y., Tzachor, A., & Chen, W.-Q. (2024). E-waste challenges of generative artificial intelligence. *Nature Computational Science, 4,* 818-823. https://doi.org/10.1038/s43588-024-00712-6
7. International Energy Agency. (2025). *Energy and AI.* IEA. https://www.iea.org/reports/energy-and-ai

ACRL AI Competencies covered

Knowledge & Understanding

Sub-competencies: 2.1, 2.3, 3.1 · ACRL AI Competencies (2025)