From search expert to AI steward: a practical framework for health librarians

Anonymous

Use of generative artificial intelligence

Generative artificial intelligence was used to support background searching, development of the task framework, and drafting and editing of this manuscript.

Introduction: beyond the question of replacement

Health librarians are already working inside an artificial intelligence (AI) transition. Sometimes the technology is obvious, as in a conversation with a general-purpose large language model (LLM). More often it is embedded in a discovery platform, a systematic review tool, an office suite, a design program or a knowledge service. The useful professional question is therefore no longer whether health librarians will encounter AI. It is how to decide where AI adds value, what information can safely be used, how outputs should be checked, and which decisions must remain human.

This matters more in health services than in most settings. An efficient but inaccurate output may shape a literature search, a policy review, a research project, an educational resource or a clinical question. Recent correspondence in this journal defended the continuing necessity of medical libraries against the assertion that they are no longer needed in a digital age (Anderson et al., 2026). The 2026 ALIA/Health Libraries Australia (HLA) competencies make the same direction of travel explicit: now expanded to nine competency areas, they name generative AI, large language models, responsible AI use, digital health, information governance and emerging technologies as part of ordinary health-library practice (ALIA HLA, 2026).

The evidence base is developing quickly but unevenly. A 2025 scoping review located only eleven studies describing AI use in health and medical libraries, spanning event planning, content enhancement, literature searching, training promotion and evidence synthesis, and concluded that high-stakes activities, particularly systematic searching, require continued human oversight (Sen, 2025). An early integrative review mapped ChatGPT's potential for reference support and personalised assistance in medical libraries alongside its risks around misinformation (Lund et al., 2024), and a parallel literature argues that realising any of it depends on unglamorous prerequisites: training, infrastructure and ethical standards adapted to local context (Orubebe et al., 2024). Three years ago this journal published an argument that machine learning would change everything and change nothing (Velli, 2023). The intervening period suggests both halves were right. The tools have changed enormously; the professional obligations have not. What is still missing is

not another catalogue of products, but a repeatable method for deciding whether and how to use them.

Start with the task, the data and the consequence
AI products change monthly. The questions governing responsible use are comparatively stable, and three of them will orient most decisions.

Task: is the work primarily drafting, transforming, classifying, retrieving, predicting or deciding?

Data: does the input contain personal, sensitive, confidential, copyrighted, culturally restricted or otherwise protected information?

Consequence: what happens if the output is incomplete, biased, outdated or wrong, and how readily can a knowledgeable human detect the error?
This shifts attention from brand names to fitness for purpose. A public chatbot, a source-grounded research assistant, a machine-learning screening system and an enterprise productivity assistant may all be called "AI", yet they differ in data controls, evidence base, transparency and characteristic failure mode. Retrieval augmentation can tie a response to supplied or retrieved documents, which helps, but a citation is not a guarantee. The source may be unsuitable, and the generated sentence may not faithfully represent it.

Where AI can enhance health-library work
Table 1 maps work described in the HLA competencies to useful forms of AI assistance. The third column is the essential part: enhancement occurs only where the librarian retains the controls that protect quality, safety and trust.

Health-library task	Potential AI assistance	Librarian control
Reference interview and question formulation	Turn notes into PICO/PEO elements; identify ambiguity; propose clarifying questions	Confirm intent, scope, urgency and intended use with the requester
Literature search development	Generate candidate synonyms, spelling variants and related concepts; suggest subject headings; translate a draft between interfaces	Verify every term and field; test against seed papers; inspect syntax and results; document the final strategy
Evidence summaries and current awareness	Summarise supplied sources; cluster results; draft	Read the sources; verify every claim and citation; disclose limitations, dates and search coverage

Health-library task	Potential AI assistance	Librarian control
	structured briefs or alert annotations	
Evidence synthesis	Support deduplication, prioritisation, screening, extraction and classification	Use a protocol and calibration set; monitor false negatives; preserve audit trails; retain human responsibility for inclusion and interpretation
Knowledge and collection management	Suggest metadata, tags and taxonomy mappings; analyse usage data; identify patterns in feedback	Apply standards; check bias and accessibility; respect licensing, copyright, retention and data-governance requirements
Teaching and information literacy	Draft scenarios, quizzes, lesson plans and alternative versions for different audiences	Validate content and links; align with learning outcomes; test accessibility; teach verification, not prompting alone
Consumer health information	Draft plain-language or translated versions of approved information	Use authoritative source text; apply health-literacy checks; obtain clinical and cultural review; never treat machine translation as final
Research and scholarly communication	Edit prose; structure protocols; suggest extraction fields; assist with code, tables and visualisation	Protect unpublished material; reproduce analyses; check attribution; follow journal, funder and institutional disclosure policies
Library operations and events	Draft schedules, run sheets, promotional copy and speaker briefs; summarise registrations and feedback	Check names, dates, accessibility and inclusive language; keep registrant data in approved systems
Leadership, management and advocacy	Draft reports and business cases; analyse non-sensitive service data; summarise consultations	Use approved data and tools; validate metrics; retain accountability for recommendations and decisions

Table 1. A task-to-AI map for health-library practice. Product examples are intentionally omitted because capabilities, licensing and governance change rapidly.

Search assistance is useful; search substitution is not

Literature searching illustrates the distinction sharply. LLMs are useful thinking partners because they generate terminology quickly, expose alternative phrasings, propose subject headings and reformat syntax between interfaces. Structured

prompts built on familiar frameworks such as PICO can make that interaction more consistent (Robinson et al., 2025). These uses are generative and developmental: they help the librarian draft.

They are not the same as executing a reproducible comprehensive search. Using a completed systematic review on Peyronie disease as a benchmark, Gwon et al. (2024) found that only 7 of 1,287 records identified through ChatGPT were directly relevant, compared with 19 of 48 from Bing AI against a human benchmark of 24 studies, and concluded that this approach was not sufficiently accurate for real-time systematic-review searching. Related work has found that LLM-generated Boolean queries vary substantially in quality and reproducibility across models, prompts and repeat runs (Wang et al., 2023; Staudinger et al., 2024), and a critical analysis of ChatGPT-developed search strategies in health reached similar conclusions about the need for expert revision (Guimarães et al., 2024).

Purpose-built AI search tools are not exempt. Research Information Services at Canada's Drug Agency compared three AI or automation tools for information retrieval, Lens.org, SpiderCite and Microsoft Copilot, against the agency's customary retrieval practice across seven completed projects, combining a literature review, a retrospective comparative analysis and a focus group with information specialists. Performance was inconsistent across retrieval tasks, and the specialists reported that a chatbot asked to add field codes or subject headings to a strategy often failed to apply them correctly (Featherstone et al., 2025).

A librarian can therefore use an LLM to propose terms, but must still verify vocabulary, select databases and grey-literature sources, apply valid syntax, test retrieval against known relevant studies, examine precision and document the process.

Three published examples worth borrowing

Three practice reports in a single 2025 issue of the Journal of the Medical Library Association show how differently these judgements play out across real services. At Stanford School of Medicine's Lane Medical Library, librarians applied generative AI across the event-planning lifecycle, managing everything from small workshops to larger conferences, and reported improved efficiency and staff time released for higher-level work (Huddleston & Cuddy, 2025). This is genuinely low-risk work: the data are non-sensitive, the cost of error is an awkward flyer, and mistakes are visible immediately.

Collection development proved harder. Portillo and Carson (2025) evaluated four generative models over six months using two prompts. Asking for recent eBook titles in specific health sciences fields returned inconsistent results and inaccuracies; asking the models to identify subject gaps in an existing collection worked considerably

better, yielding useful analysis and accurate Library of Congress call numbers. Their conclusion generalises well: LLMs are not yet reliable as primary tools where they must supply the facts, but are useful where they analyse material the librarian already holds.

Patron-facing guides sit between the two. Jones (2025) used an LLM to restructure a dentistry LibGuide previously organised by resource format and to draft page summaries, then recorded a 131% increase in guide access over the following four months against the same period a year earlier. The comparison is uncontrolled and the author notes no other changes were made, encouraging rather than conclusive, but exactly the kind of local evaluation the profession needs more of.

In practice: keeping the human in the loop

Systematic reviews are transparent summaries of research addressing a defined question, structured to reduce bias and support reliable conclusions. The search is the foundation of that transparency, and the PRESS checklist gives librarians and information specialists an evidence-based instrument for assessing it, with peer review recommended at two points: before the searches are run, and before publication (McGowan et al., 2016).

Much AI development in this space assumes infrastructure most health librarians do not have. Model Context Protocol connections, retrieval-augmented pipelines and highly agentic platforms can assemble impressive discovery workflows, but they also introduce opacity, the RAISE recommendations note that the retrieval methods of agentic platforms are frequently not transparent, and that LLM tools do not currently replace established search-development methods (Thomas et al., 2026). Cost, licensing, security assessment and technical capacity remain real barriers to adoption (Siemens et al., 2025).

Perhaps we could take the opposite approach in testing and deliberately narrow the variables: no Model Context Protocol connections, no agentic orchestration, no enterprise-scale context windows. Just the ordinary, widely available models a health librarian can reach today, applied to protocol and search-strategy development with a documented human decision recorded at each step. The question is not whether a model can produce a plausible strategy, it can. The question is whether the underlying models are reasonably reliable.

Grounded summaries still require source checking

Evidence summaries are a second promising, high-consequence area. Blasingame et al. (2025) compared an internally managed GPT-4 tool with medical librarians' gold-standard evidence syntheses. Of 216 questions, the tool's response was rated correct for 180 (83.3%) and partially correct for 35 (16.2%). Yet in a randomly selected subset of 66 answers containing 162 references, only 60 (37%) could be confirmed to exist.

The result is both encouraging and cautionary. LLMs can accelerate first-pass organisation and drafting, particularly when working from a defined source set, but librarians must verify citations and confirm that the evidence actually supports each statement. A polished answer is not evidence of a reliable method.

Specialist AI can support reproducible workflows

General-purpose LLMs are only one part of the landscape. Specialist tools apply machine learning to deduplication, prioritisation, screening or extraction, ASReview for active-learning-assisted screening, Rayyan for review workflows, priority screening in EPPI-Reviewer, and the Australian-developed Evidence Review Accelerator (TERA) suite. Classifiers of this kind have been shown to reduce screening workload with minimal risk of missing studies when properly validated (Thomas et al., 2021).

A 2026 systematic review of AI-assisted reviews in health confirms where the technology has actually landed: current tools are used predominantly for title and abstract screening, formal guidance from health technology assessment bodies remains limited, and the barriers to adoption are data quality, limited technical expertise, infrastructure constraints and regulatory uncertainty (Abogunrin et al., 2026).

These systems can remove genuinely repetitive work, but their outputs should be evaluated against an agreed protocol. A relevant record missed by an automated system costs more than the time saved, so teams should calibrate tools, sample excluded material where appropriate, monitor performance and record how automation influenced decisions. Cochrane, the Campbell Collaboration, JBI and the Collaboration for Environmental Evidence have taken a consistent position: AI and automation may be used provided authors can demonstrate that methodological rigour and integrity are not compromised, use is reported transparently, and human oversight is retained. The synthesist remains accountable for the synthesis (Flemyng et al., 2025).

Teaching AI literacy is core health-information literacy

Health librarians can use AI to develop teaching scenarios, quizzes, handouts and alternative explanations for different audiences. More importantly, they can teach clinicians, researchers and students how AI changes information behaviour. A useful session moves past prompt tips to compare a general chatbot, a database search and a source-grounded system; trace generated claims back to sources; identify fabricated or mismatched references; discuss coverage and currency; and decide what information must never be entered.

This is a natural extension of teaching source evaluation and critical appraisal, and it positions the library as a partner in organisational AI literacy rather than a late-stage

checker of problems created elsewhere. A 2026 assessment of AI and LLM integration in medical libraries reaches the same point from the other direction, identifying information retrieval and review support as the principal application domains while naming data privacy risk and staff AI literacy gaps as the limiting factors (Zulmi et al., 2026). It also requires deliberate professional development on our own part (O'Keefe et al., 2026), which is exactly what shared structures such as the HLA Artificial Intelligence Community of Practice exist to support.

Cultural authority cannot be automated

The HLA competencies give Aboriginal and Torres Strait Islander health context its own domain, emphasising respectful relationships, Indigenous authorship, Indigenous data sovereignty and culturally appropriate services (ALIA HLA, 2026). AI systems may reproduce the gaps and biases of dominant published literature, or flatten distinct communities into generic language. They are not cultural authorities. Any use involving Aboriginal and Torres Strait Islander knowledge, data, language or community-facing material must be guided by relevant protocols, community relationships and appropriate human review. The same principle applies to other underrepresented groups: apparent fluency does not establish cultural safety.

A green–amber–red decision model

A traffic-light model helps teams move from abstract principle to daily decision.

Level	Typical conditions	Examples	Required response
GREEN, assist	Public or synthetic information; low consequence; readily checked	Brainstorm headings; reformat text; create dummy examples; draft a generic lesson plan	Normal professional review
AMBER, assist with controls	Professional, research or organisational work where errors matter but can be detected	Develop search concepts; summarise supplied articles; prioritise screening; draft a service report; analyse de-identified data	Approved tool; minimum necessary data; expert verification; documentation and disclosure
RED, do not delegate	Prohibited data, or a decision whose error could cause serious harm or unfairness	Entering patient or staff information into an unapproved public tool; unsupervised clinical advice; accepting invented references; letting AI	Stop, redesign the task, or use an authorised governed process with accountable human decision-makers

Level	Typical conditions	Examples	Required response
		make final inclusion, procurement, employment or policy decisions	

Table 2. A proportional risk model. Local legislation, agency policy and approved-product lists always take precedence.

Governance for Australian health libraries

For Australian organisations, privacy and information governance are not optional additions to prompting technique. The Office of the Australian Information Commissioner advises that privacy obligations apply both to personal information entered into an AI system and to output containing personal information, recommends privacy-by-design, due diligence, privacy impact assessment and human oversight, and states that as a matter of best practice personal information, particularly sensitive information, should not be entered into publicly available generative AI tools (OAIC, 2024).

Queensland Government guidance is similarly clear on the point most relevant to daily practice: staff are responsible for understanding the classification of the information they hold, and should not share, input or upload information into generative AI products their agency has not approved (Queensland Government, 2025). Although published as a guideline rather than a mandatory standard, it sits alongside agency-specific policies and approved-product arrangements, including government-provided assistants such as QChat, which take precedence in any given workplace. Health librarians should not assume that a personal account, a free product or an apparently private chat is suitable for work information. Product terms, retention settings, training use, access controls, data location, copyright arrangements and audit capability all require organisational assessment.

Libraries can contribute to that assessment. Expertise in source authority, licensing, privacy, metadata, information lifecycles, accessibility and user education makes health librarians valuable members of AI governance, procurement and evaluation groups. The emerging role is not "expert prompter". It is AI steward: a professional who connects a useful capability to an appropriate task while protecting evidence quality, rights, context and accountability.

Five rules for practical implementation

1. **Use approved tools and the minimum necessary data.** If the classification, licence or privacy status is uncertain, pause and seek advice.
2. **Require inspectable evidence for factual work.** Prefer workflows that retain source links, quoted passages, versions and audit trails, then actually check them.

3. **Keep consequential judgement human.** AI may suggest, rank or draft; the accountable librarian or authorised professional makes the decision.
4. **Document and disclose material use.** Record the product and version, date, task, inputs, validation method, limitations and human role.
5. **Evaluate outcomes, not enthusiasm.** Compare time, error, recall, usability, equity and client impact against the previous workflow, and stop using a tool that is not fit for purpose.

Conclusion

AI can help health librarians work faster, communicate more clearly, explore terminology, organise information, support review workflows and extend teaching. It can also produce fluent error, conceal gaps in coverage, amplify bias and expose protected information. The difference between enhancement and degradation is not the presence of AI; it is the quality of the professional practice surrounding it.

Health librarians need not choose between resistance and uncritical adoption. A task-based approach asks what work is being assisted, what data are involved, what the consequences of error would be, and which controls preserve trust. On that model AI does not diminish health librarianship. It makes the profession's existing strengths, verification, transparency, contextual judgement, information governance and education, matter more.

References

Abogunrin, S., Liu, Y., & Zerbini, C. H. (2026). A systematic literature review (SLR) on the adoption of artificial intelligence-assisted SLRs: Implications for health technology assessments. *International Journal of Technology Assessment in Health Care, 42*(1), e29. https://doi.org/10.1017/S0266462326103535

Anderson, A., Siemensma, G., & Smith, A. (2026). Re: Albert et al.: Don't close medical libraries: That's where you find librarian partners to advance medicine and science. *Journal of Health Information and Libraries Australasia, 6*(1). https://doi.org/10.55999/johila.v6i1.222

Australian Library and Information Association Health Libraries Australia. (2026). *ALIA/Health Libraries Australia (HLA) competencies for health librarians and health library technicians*. https://hla.alia.org.au/wp-content/uploads/2026/06/Health-Libraries-Australia-HLA-Competencies-for-Health-Librarians-and-Health-Library-Technicians-proof3.pdf

Blasingame, M. N., Koonce, T. Y., Williams, A. M., Giuse, D. A., Su, J., Krump, P. A., & Giuse, N. B. (2025). Evaluating a large language model's ability to answer clinicians' requests for evidence summaries. *Journal of the Medical Library Association, 113*(1), 65–77. https://doi.org/10.5195/jmla.2025.1985

Flemyng, E., Noel-Storr, A., Macura, B., Gartlehner, G., Thomas, J., Meerpohl, J. J., Jordan, Z., Minx, J., Eisele-Metzger, A., Hamel, C., Jemioło, P., Porritt, K., & Grainger, M. (2025). Position statement on artificial intelligence (AI) use in evidence synthesis across Cochrane, the Campbell Collaboration, JBI and the Collaboration for Environmental Evidence 2025. *Cochrane Database of Systematic Reviews*, (10). https://doi.org/10.1002/14651858.ED000178

Featherstone, R., Walter, M., MacDougall, D., Morenz, E., Bailey, S., Butcher, R., Ford, C., Loshak, H., & Kaunelis, D. (2025). Artificial intelligence search tools for evidence synthesis: Comparative analysis and implementation recommendations. *Cochrane Evidence Synthesis and Methods, 3*(5), e70045. https://doi.org/10.1002/cesm.70045

Guimarães, N. S., Joviano-Santos, J. V., Reis, M. G., & Chaves, R. R. M. (2024). Development of search strategies for systematic reviews in health using ChatGPT: A critical analysis. *Journal of Translational Medicine, 22*(1), 1. https://doi.org/10.1186/s12967-023-04371-5

Gwon, Y. N., Kim, J. H., Chung, H. S., Jung, E. J., Chun, J., Lee, S., & Shim, S. R. (2024). The use of generative AI for scientific literature searches for systematic reviews: ChatGPT and Microsoft Bing AI performance evaluation. *JMIR Medical Informatics, 12*, e51187. https://doi.org/10.2196/51187

Huddleston, B., & Cuddy, C. (2025). Leveraging AI tools for streamlined library event planning: A case study from Lane Medical Library. *Journal of the Medical Library Association, 113*(1), 88–89. https://doi.org/10.5195/jmla.2025.2087

Jones, E. P. (2025). Use of large language model (LLM) to enhance content and structure of a school of dentistry LibGuide. *Journal of the Medical Library Association, 113*(1), 96–97. https://doi.org/10.5195/jmla.2025.2084

Lund, B. D., Khan, D., & Yuvaraj, M. (2024). ChatGPT in medical libraries, possibilities and future directions: An integrative review. *Health Information & Libraries Journal, 41*(1), 4–15. https://doi.org/10.1111/hir.12518

McGowan, J., Sampson, M., Salzwedel, D. M., Cogo, E., Foerster, V., & Lefebvre, C. (2016). PRESS peer review of electronic search strategies: 2015 guideline statement. *Journal of Clinical Epidemiology, 75*, 40–46. https://doi.org/10.1016/j.jclinepi.2016.01.021

Office of the Australian Information Commissioner. (2024, October 21). *Guidance on privacy and the use of commercially available AI products*. https://www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-

government-agencies/guidance-on-privacy-and-the-use-of-commercially-available-ai-products

O'Keefe, H., Eastaugh, C. H., Wallace, S. A., & Beyer, F. R. (2026). Information specialist roles in the era of large language models: Prompting continued professional development. *Campbell Systematic Reviews, 22*(2). https://doi.org/10.1177/18911803261449731

Orubebe, E. D., Ijaja, E. M., Ogwula, J. A., & Oladokun, B. D. (2024). Transforming medical libraries: Opportunities, challenges, and strategies for integrating artificial intelligence. *Asian Journal of Information Science and Technology, 14*(2), 82–87. https://doi.org/10.70112/ajist-2024.14.2.4298

Portillo, I., & Carson, D. (2025). Making the most of artificial intelligence and large language models to support collection development in health sciences libraries. *Journal of the Medical Library Association, 113*(1), 92–93. https://doi.org/10.5195/jmla.2025.2079

Queensland Government. (2025, August 7). *Use of generative AI in Queensland Government*. https://www.forgov.qld.gov.au/information-technology/queensland-government-enterprise-architecture-qgea/qgea-directions-and-guidance/qgea-policies-standards-and-guidelines/use-of-generative-ai-in-queensland-government

Robinson, K., Bontekoe, K., & Muellenbach, J. (2025). Integrating PICO principles into generative artificial intelligence prompt engineering to enhance information retrieval for medical librarians. *Journal of the Medical Library Association, 113*(2), 184–188. https://doi.org/10.5195/jmla.2025.2022

Sen, S. (2025). AI and generative AI in health and medical libraries: A scoping review of present use and emerging potential. *Journal of EAHIL, 21*(2), 22–26. https://doi.org/10.32384/jeahil21675

Siemens, W., von Elm, E., Binder, H., Böhringer, D., Eisele-Metzger, A., Gartlehner, G., Hanegraaf, P., Metzendorf, M.-I., Mosselman, J.-J., Nowak, A., Qureshi, R., Thomas, J., Waffenschmidt, S., Labonté, V., & Meerpohl, J. J. (2025). Opportunities, challenges and risks of using artificial intelligence for evidence synthesis. *BMJ Evidence-Based Medicine, 30*(6), 381–384. https://doi.org/10.1136/bmjebm-2024-113320

Staudinger, M., Kusa, W., Piroi, F., Lipani, A., & Hanbury, A. (2024). A reproducibility and generalizability study of large language models for query generation. In *Proceedings of the 2024 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region* (pp. 186–196). https://doi.org/10.1145/3673791.3698432

Thomas, J., Hair, K., Noel-Storr, A., Flemyng, E., Moy, W., Marshall, I. J., & Hajji, R. (2026). *Responsible use of AI in evidence synthesis (RAISE): Recommendations for practice* (Version 3, updated 13 March 2026). Open Science Framework. https://doi.org/10.17605/OSF.IO/FWAUD

Thomas, J., McDonald, S., Noel-Storr, A., Shemilt, I., Elliott, J., Mavergames, C., & Marshall, I. J. (2021). Machine learning reduced workload with minimal risk of missing studies: Development and evaluation of a randomized controlled trial classifier for Cochrane reviews. *Journal of Clinical Epidemiology, 133*, 140–151. https://doi.org/10.1016/j.jclinepi.2020.11.003

Velli, G. (2023). ChatGPT and AI hysteria: Why machine learning will change everything, and change nothing. *Journal of Health Information and Libraries Australasia, 4*(2), 8–18.

Wang, S., Scells, H., Koopman, B., & Zuccon, G. (2023). Can ChatGPT write a good Boolean query for systematic review literature search? In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1426–1436).

Zulmi, M. R., Yuadi, I., Khusna, T. N., & Jazuli, A. (2026). Artificial intelligence and LLM integration in medical libraries: A systematic assessment of evolving workflows. *IP Indian Journal of Library Science and Information Technology, 11*(2), 306–313. https://doi.org/10.18231/j.ijlsit.17323.1781448341