www.eahil.eu

Journal of EAHIL

European Association for Health Information and Libraries

Vol. 22 no.2

June 2026

ISSN L-1841-0715

The production of this journal
was made possible by the generous support of:

ACS Publications
TDNet

Journal of the European Association for Health Information and Libraries

Vol. 22 No. 2 June 2026

CONTENTS

EDITORIAL

Artificial intelligence in libraries: literacy, policy and evaluation

David Ožura
Special Library of Oncology, Ljubljana, Slovenia
dozura@onko-i.si

Dear colleagues,

For the third year running, the *Journal of EAHIL* devotes its June number to artificial intelligence. The series began in June 2024, when Michelle Wake introduced a thematic issue (Vol. 20, No. 2) gathering six articles on the challenges, limitations and opportunities of AI for libraries and the wider profession. A year later, in June 2025 (Vol. 21, No. 2), I had the pleasure of presenting a second such issue, with contributions from Finland, the Netherlands, Italy and the United Kingdom. These examined practitioners' attitudes towards AI, its impact on literature searching and health information retrieval, and the AI literacy and training that information professionals increasingly need, while consistently emphasising human oversight, the critical evaluation of AI outputs, and ethical concerns such as bias, transparency and data privacy. What began as a timely response to a fast-moving technology has, in effect, become a regular feature of the journal, one the present issue is glad to continue.

Our readers' keen interest in the topic is plain from the figures: in the journal's bibliometric report for 2025, compiled by Rebecca Wojturska, seven of the ten most downloaded *JEAHIL* articles were about AI. They were led by Shampa Sen's scoping review of AI in health and medical libraries (1,785 downloads), with Veronica Parisi and Anthea Sutton's evidence summary on ChatGPT for systematic literature searches, Andrew Cox's article on AI and health information literacy, Tuulevi Ovaska's survey of Finnish health librarians' views on AI, and Sjoukje van der Werf's workshop perspective on AI in literature searching also among them. And in both 2024 and 2025, the most downloaded issue of the year was the June AI thematic one. AI is clearly a subject our medical library community wants to read about, reflect upon and, increasingly, act upon.

Yet, for all its current prominence, AI is no newcomer. EAHIL and its journal have been following the technologies behind it, and reflecting on their consequences for our work, for far longer than the generative-AI era might suggest. As early as the 2015 EAHIL Workshop in Edinburgh, a hands-on session explored the use of text mining and machine learning, including automatic classification trained on human screening decisions, to assist with citation screening for systematic reviews; the discussion was described in this journal that same year by Claire Stansfield, Alison O'Mara-Eves and James Thomas in their report Reducing systematic review workload using text mining: opportunities and pitfalls (Vol. 11, No. 3). Years before ChatGPT, then, health librarians were already weighing the benefits and risks of automating evidence synthesis. Claire

Stansfield, a contributor to that early workshop, returns as an author in the present issue, an indication that the profession's engagement with AI is long-standing rather than recent.

Taken together, the four contributions gathered in this June thematic issue, from the United Kingdom and Poland, show a profession that, having first defined and debated what AI is, now turns to the practical work of using it well: building literacy among staff and students, guiding institutional policy, and bringing rigorous, critical and ethical evaluation to the AI tools it adopts.

In Empowering our communities, Angela Young (University College London) describes how her department, Library, Culture, Collections and Open Science, is supporting AI literacy across its community, pairing guidance, training and staff development with an emerging framework for evaluating AI tools, all underpinned by a departmental AI strategy.

Anna Richards, Ishbel Leggat and Robert O'Brien (University of Edinburgh) apply Gibbs' Reflective Cycle in Development of AI literacy support within a small academic library team, reflecting candidly on three years of building such support and weighing early successes (strong engagement and a heavily used LibGuide) against the persistent challenges of keeping materials current, securing recognition of AI literacy as part of the library's professional remit, and sustaining the work within existing workloads.

In Policies of selected Polish universities regarding the use of AI and the potential role of libraries in their implementation, Iwona Kosowska (Jagiellonian University Medical College, Kraków) examines what these institutional policies, set against five international comparators, permit and require, particularly around the disclosure of AI use and the preservation of academic integrity, and argues that libraries are well placed to provide information, raise awareness and deliver AI literacy training as the policies are put into practice.

Finally, in Reflections on evaluating the utility of an LLM for keywording health research, Claire Stansfield and Ailbhe Finnerty Mutlu (EPPI Centre, University College London) assess how well a large language model can apply predefined keywords to health-research records, in a living repository of HPV-vaccine research and in two registers of health-promotion effectiveness, distilling their experience into five practical considerations: cyclical prompt development, data quality and availability, performance benchmarks and expectations, task complexity, and the workflows and tools that make such evaluation possible.

I am grateful to the authors for sharing their experience so generously, and I hope you find as much here to reflect on as I have.

Empowering our communities: Library support for AI literacy at UCL

Angela Young
UCL Library Services, Library, Culture, Collections & Open Science (LCCOS), University College London (UCL), London, United Kingdom

Abstract
As generative artificial intelligence (GenAI) continues to reshape higher education, libraries must adapt their services to support evolving user needs and practices. This article outlines how LCCOS (Library, Culture, Collections and Open Science) at University College London (UCL) is supporting AI literacy across its community. It describes the development of guidance and training for users, LCCOS staff development initiatives designed to build confidence and capability in the use of AI, and emerging approaches to evaluating AI tools to inform both library service decision-making and individual use. These activities are underpinned by a departmental AI strategy, which enables a coordinated and sustainable approach aligned with broader institutional priorities.

Key words: *artificial intelligence; generative artificial intelligence; information literacy; academia; education.*

Introduction

Initial academic skills support for generative AI (GenAI) at University College London (UCL) focused on its use in assessment, including the library's development of institutional guidance on acknowledgement and referencing (1). The library has since continued to collaborate with other UCL support services to ensure a coordinated approach, with each service developing support in its area of responsibility. In this follow-up article, we outline the developing strategic approach by UCL LCCOS (Library, Culture, Collections and Open Science) to engaging with and supporting the use of artificial intelligence (AI).

Strategy development

As GenAI is still new and is rapidly evolving, the need to draw on expertise across LCCOS was recognised early on and addressed through a staff special interest group formed in early 2023. The initial focus of the LCCOS AI Group was to monitor developments in the use of GenAI in higher education and its impact on library services and cultural heritage, while considering approaches to and priorities for adapting support and service provision. While the group's initial response to GenAI was necessarily reactive, the need for a more strategic approach was recognised.

A subgroup scoped existing strategic models (2-5) and made recommendations to inform the development of a strategy. The LCCOS AI strategy, formalised in September 2025, presents a vision in which LCCOS will "harness the potential of AI to drive innovation and engagement across our spaces, services and collections" (6) with a commitment to effective and responsible use, and to empowering "staff and communities to develop the skills and confidence needed to engage with these technologies thoughtfully and ethically" (6).

Five foundational principles underpin the strategy, including utilising AI for enhancement and innovation of services, operations and skills while recognising the value of continued human expertise and input. It advocates for responsible and ethical use, and a culture of experimentation and learning, collaboration and collective engagement.

The strategy identifies broad themes from which tangible outputs will be developed and implemented through a new committee structure, comprising an oversight group and working groups.

While the strategy was in development, initiatives and activities which addressed clear priorities continued to be progressed.

Address for correspondence: Angela Young, UCL Library Services, Library, Culture, Collections & Open Science (LCCOS), University College London (UCL), Gower St, London WC1E 6BT, United Kingdom. E-mail: angela.young@ucl.ac.uk

Journal of EAHIL 2026; Vol. 22 (2): 4-8 DOI:10.32384/jeahil22717

Support for AI literacy

One of the strategic themes, "Empowering our communities", builds on ongoing support for AI literacy. Initial library guidance (7) covered referencing and acknowledging the use of GenAI, and using GenAI as a source of information. It has expanded to include guidance on summarisation, copyright, and using AI for systematic reviews, as well as signposting selected AI tools for supporting the library research process, and it continues to be developed.

Library support focuses on principles for effective and responsible AI use, including critical engagement with inputs and outputs, and using AI to enhance learning and reflection. In parallel with guidance development, a student workshop was developed and initially delivered as part of UCL's ExtendED Learning programme, which offers "students a wide range of free learning opportunities to enhance skills and networks" (8). The programme is aligned with UCL's Pillars of Employability (9), which encourage students to develop skills to support their future and includes AI literacy as a foundational skill. The workshop is now incorporated into the library's regular programme of skills sessions, with its content complementing a session facilitated by UCL's Academic Communication Centre on critical use of Generative AI tools in the writing process.

Through practical activities and discussion, the "Critical thinking in a GenAI world: Summaries and sources of information" workshop explores GenAI as a summarisation tool and source of information for academic work and beyond. The copyright and intellectual property implications of uploading materials to GenAI tools regularly prompt more discussion than workshop time allows. To address this, a complementary session is now offered, which explores copyright considerations in the use of AI in depth.

Increasingly academic departments and other UCL groups request library sessions on the use of GenAI. The "Critical thinking in a GenAI world" workshop provides a useful model, easily adaptable to specific disciplines or use cases, such as a recent adaptation to address using GenAI to engage with and understand scientific articles.

Enabling change

The LCCOS AI Strategy recognises that to respond to the "transformative potential of AI", staff need to be equipped and supported, including ensuring all "staff have the opportunity to achieve baseline AI literacy" (6), and more specialist skills where required.

The LCCOS AI Group recognised that effective use of GenAI is shaped not only by task requirements but also by individual working practices, and developing effective usage is therefore an individual journey, best achieved through active engagement, experimentation, and reflection. The group observed that engagement with GenAI varied considerably across the department. Some colleagues readily engage independently, experimenting and identifying effective applications. Others struggle to achieve useful outcomes or are uncertain where to begin, and some express concern about potential implications of AI for their roles. The group identified the need for a structured, supportive environment in which colleagues could experiment and share experiences to build confidence, capability and understanding, supporting the strategic aim of staff baseline AI literacy.

In summer 2025, the group facilitated two workshops for LCCOS staff, one in-person and the other online, in which colleagues were introduced to a selection of AI tools through a range of activities. Some activities were directly related to the work context, such as simulating a role-play of a difficult conversation with a student at a library enquiry desk, while others encouraged engagement through fun, such as creating a "choose your own adventure" story. The sessions ended with reflection on how these uses of AI tools might be adapted to other work-related purposes and contexts.

More specialist skills in AI are required by library colleagues who deliver support for AI literacy across the UCL community in relation to the library research process. Dedicated subject specialist librarians support specific academic departments, so to ensure sustainable support for AI literacy development, all colleagues with a remit for teaching library skills must be confident and able to facilitate AI literacy sessions.

A barrier to library staff delivering AI literacy teaching is a personal perception of their own lack of expertise. As fundamentals of AI literacy align with the skills, attributes and behaviours inherent in information literacy, however, library staff are well placed to adapt their existing practices to provide AI literacy support and teaching. To address self-perceptions of their lack of confidence, library trainers at UCL are encouraged to

manage expectations in the workshops they deliver through transparency around the limitations of their technical expertise, and the content of sessions is designed around sharing of experiences and knowledge, and critical discussion around findings in the activities, with library teachers facilitating and drawing together key learning points, rather than solely imparting their knowledge as a perceived expert.

In supporting library staff to gain confidence in delivering teaching on AI, similar principles are adopted as for developing their broader information literacy teaching competencies, through gradual immersion, exposure and contribution. Colleagues attend sessions delivered by more experienced staff or, in the case of AI, early adopters, before being involved in delivery themselves, and subsequently progressing to independent design and delivery. To support shared practice, colleagues are encouraged to contribute materials from their teaching sessions to a central resource outlining session aims, coverage, delivery dates, and links for reuse or adaptation.

AI tools evaluation

To date, AI literacy support provided by the library has excluded recommendation or use of specific AI tools owing to:

- the rapidly evolving landscape, with new tools emerging and existing tools developing at pace;
- concern around endorsing specified tools without thorough evaluation;
- equality concerns regarding access to paid versions of tools.

In workshops, practical activities utilise Microsoft Copilot, to which UCL has an institutional subscription with added data protection. Other tools may occasionally be mentioned or signposted, or feature in discussion.

Increasingly the library receives enquiries relating to access to specific AI tools. The library does not have additional resource for AI tools, so any institutional subscriptions would require funding from existing budgets, which would impact on funds available for other electronic resources. Decisions therefore require careful consideration and thorough evaluation of tools.

The library is developing a framework for staff to evaluate AI tools, to inform acquisition decisions and respond to recommendations or enquiries about access to AI tools. The framework includes a screening stage, in which the purpose and function of the tool determine whether it is in scope for the library, and a staged evaluation process which progresses to greater depth where the evaluator determines it is worth continuing with the evaluation.

The framework was informed by existing examples developed at other institutions (10-15) and includes evaluation criteria grouped under 4 themes of functionality; usability; security, ethical and legal; and practicalities.

To support and empower students and researchers to make their own assessments around their individual choice of AI tools, a checklist (16) was developed in tandem with the library staff framework, with corresponding themes and prompting questions tailored towards individuals. It acknowledges that while consideration of all sections of the checklist ensures a balanced, thorough evaluation of any tool, the priority and relevance of the points to consider will vary depending on an individual's needs and the task at hand. It therefore does not incorporate a scoring system and instead supports critical decision-making.

Changing behaviours and future considerations

Rapid technological development requires support services to adapt to ensure ongoing effective support for users to navigate an evolving information landscape. Through interaction with and feedback from users in training sessions and examination of library enquiries we can gain understanding of current and changing behaviours and needs. Data from confidence polls taken at the start of training sessions delivered by UCL library staff has shown rapid increases in attendees' confidence in using AI tools for academic work, which reflect findings of broader studies (17). Such insights inform review and updating of sessions ahead of each instance of delivery to ensure user expectations and needs are addressed at appropriate levels.

Ongoing observation and analysis are needed to enable support services to adapt. User experience projects and surveys, such as a current survey by UCL Digital Skills and wider surveys (17, 18) help inform future AI skills support at UCL, but ongoing user experience work would further facilitate adaptation of services to support changing user behaviours and needs.

The development of AI in search and discovery tools is likely to lead to significant change in library skills support, as information-seeking behaviours of students and researchers evolve. Established search techniques may cease to be the most effective methodologies for finding information, and library staff will need to acquire new specialist knowledge and adapt their approaches.

In acknowledgment of UCL's 200-year anniversary, LCCOS colleagues are reflecting on the past while looking to the future. A workshop on the future of roles in libraries and cultural heritage will explore how roles are changing and anticipate how staff skillsets must evolve, with the impact of technological advances a key theme.

At an institutional level, UCL recently consulted with staff across academic and professional services departments to inform its ongoing approach to AI in education. Implementation of the LCCOS AI strategy will respond to these priorities while establishing a foundation that positions the department to adapt and contribute to a sustainable, institution-wide approach.

Submitted on invitation.
Accepted on 11 May 2026.

REFERENCES

1. Young A, Chandler J, Norris C, Prevatt-Goldstein A. Generative AI and academic skills support at UCL: an institutional approach. JEAHIL. 2024;20(2):16-9. Available from: https://doi.org/10.32384/jeahil20620
2. Pirgova-Morgan L. Looking towards a brighter future: the potentiality of AI and digital transformations to library spaces. Leeds: University of Leeds; 2023. Available from: https://library.leeds.ac.uk/downloads/download/196/artificial-intelligence-ai-in-libraries
3. Library AI Statement. Sheffield: University of Sheffield Library; 2025 [cited 2026 May 01]. Available from: https://sheffield.ac.uk/library/about/library-ai-statement
4. Association of Research Libraries. Research Libraries Guiding Principles for Artificial Intelligence. 2024 April. Available from: https://www.arl.org/resources/research-libraries-guiding-principles-for-artificial-intelligence/
5. Cox A. Developing a library strategic response to Artificial Intelligence. eLucidate [Internet]. 2024;16(4). Available from: https://doi.org/10.29173/elucidate847.
6. Moyle M. LCCOS AI Strategy (unpublished paper). UCL LCCOS (Library, Collections, Culture and Open Science); 2025.
7. UCL Library Skills. Generative AI and library skills. London: UCL Library Skills; 2026 [cited 2026 May 01]. Available from: https://library-guides.ucl.ac.uk/generative-ai
8. University College London (UCL). ExtendED Learning. London: University College London (UCL); 2026 [cited 2026 May 01]. Available from: https://www.ucl.ac.uk/study/current-students/academic-support-and-skills/extended-learning
9. Careers U. Developing your employability skills at UCL. London: UCL Careers; 2026 [cited 2026 May 01]. Available from: https://www.ucl.ac.uk/study/careers/preparing-your-application/developing-your-employability-skills-ucl
10. Bassett MA. SECURE. GenAI Use Framework for Staff. Australia: Charles Sturt University; 2025 [cited 2026 May 01]. Available from: https://secureframework.ai/
11. Mackie K, Aspenlieder E. Rubric for evaluating AI tools: Fundamental criteria. MacPherson Institute for Leadership, Innovation and Excellence in Teaching, McMaster University; 2024. Available from: https://ecampusontario.pressbooks.pub/mcmasterpracticalaiguide/chapter/evaluating-ai-tools/
12. University of Birmingham Libraries. Responsible AI tool selection for researchers. Birmingham: University of Birmingham; 2026 [cited 2026 May 01]. Available from: https://intranet.birmingham.ac.uk/student/libraries/copyright/researchers/responsible-ai-tool-selection.aspx
13. Purdue University. How to evaluate AI tools. West Lafayette, IN: Purdue University; 2026 [cited 2026 May 01]. Available from: https://www.purdue.edu/ai/how-to-evaluate-ai-tools/

14. CDA-AMC. Development of an evaluation instrument on artificial intelligence Search tools for evidence synthesis. Canadian Journal of Health Technologies [Internet]. 2024;4(10). Available from: https://doi.org/10.51731/cjht.2024.1004.
15. Michalak R, Tzoc E, Lewis JD. Practical considerations for adopting generative AI tools in academic libraries. Journal of Library Administration. 2025;65(5):596-615. Available from: https://doi.org/10.1080/01930826.2025.2506151.
16. AI tools evaluation checklist. [Internet]. London: UCL Library Skills; 2026 [cited 2026 May 01]. Available from: https://library-guides.ucl.ac.uk/generative-ai/ai-tools-evaluation-checklist
17. HEPI (Higher Education Policy Institute). Student generative artificial intelligence survey 2026. 2026. Available from: http://hepi.ac.uk/reports/student-generative-ai-survey-2026/
18. StudentXGenAI. Student experiences of generative artificial intelligence in UK universities. 2025 [cited 2026 May 01]. Available from: https://www.studentxgenai.co.uk/

Development of AI literacy support within a small academic library team: Reflections from the University of Edinburgh

Anna Richards, Ishbel Leggat and Robert O'Brien
University of Edinburgh, Edinburgh, Scotland, United Kingdom

Abstract
This paper explores the experiences of one small team of academic librarians in developing AI literacy support for their university. Gibbs' Reflective Cycle is used to encourage deeper consideration of successes as well as areas that still need developing. Early achievements include good engagement from staff and students, development of our own confidence and understanding of the technology, and acknowledgment of our experience and knowledge within the wider library. Areas that require ongoing improvement include promotion of our support within the wider university, ensuring material is kept up to date in a rapidly changing field, and managing the impact of this work on our workloads. This reflective report highlights the value of using reflective models to encourage strategic consideration of our work and how to ensure any improvements we make are appropriately targeted.

Key words: *information literacy; generative artificial intelligence; artificial intelligence.*

Introduction

When ChatGPT launched in November 2022 it quickly became clear that it could have a significant impact on users'search behaviour and on education more generally. As Academic Support Librarians (ASLs) supporting students in their information literacy skills we knew we needed to understand this new technology and develop new resources to support staff and students. This article provides a reflective report on how AI literacy expertise and support was developed within one small library team and utilizes Gibbs' Reflective Cycle to enable greater understanding of our experience (1).

Reflective practice is well suited to reviewing professional work, particularly projects or responses to a new area of work that is changing rapidly. We have taken a continuous improvement approach to our work on AI literacy, whereby we use feedback, our increasing knowledge, and changes in the environment (both AI developments and changes in approach within higher education) to make small but continual improvements to our material and the support we offer (2). To ensure these changes are true improvements, we need to reflect on what has gone well and what hasn't in our previous work. Using a reflective practice model ensures we examine our work from multiple perspectives systematically. Gibbs' six-stage cycle proved particularly valuable in helping us distinguish between surface-level problems (like declining workshop attendance) and their underlying causes (such as students' overconfidence in their existing AI skills). This structured approach revealed actionable insights we might have missed in an unstructured team discussion. As Miller, Ford and Yang (3) highlight, in order for reflection to truly improve the work we do, we need to approach reflection with intention and so we decided to use a particular model to guide our reflection.

Gibbs' Reflective Cycle encourages practitioners to analyse any experience from six different angles, moving sequentially through them all: description, feelings, evaluation, analysis, conclusion and action (1). Gibbs' model was designed for use within teacher education and so felt particularly valuable to our work teaching AI literacy (*Figure 1*) (3).

Address for correspondence: Anna Richards, University of Edinburgh, Old College, South Bridge, Edinburgh EH8 9YL, Scotland, United Kingdom. E-mail: a.richards@ed.ac.uk

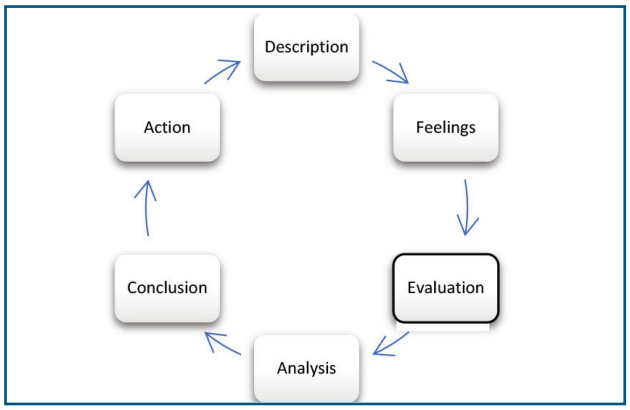

Fig. 1. *Gibb's reflective cycle.*

Gibb's reflective cycle

Description

The ASL AI Group was formed in May 2023, including the Associate Director, Library Academic Support, Elizabeth Williams, and ASLs Anna Richards, Ishbel Leggat, and Robert O'Brien. A further colleague, Christine Love-Rodgers, joined in 2025. One of our first tasks was to understand what AI literacy meant for us and our students.

Lo (4, para. 3) says that AI literacy is: "the ability to understand, use, and think critically about AI technologies and their impact on society, ethics, and everyday life". He argues that AI literacy in academic libraries includes:

- technical knowledge of how systems work;
- awareness of ethical issues;
- critical engagement with AI output use;
- practical knowledge of how to use tools;
- understanding of the societal impacts of AI.

We recognised that our existing expertise positioned us well to apply this conception of AI literacy to information literacy, in particular:

- understanding the ethical concerns around generative AI (GenAI) development and use, both generally and in an academic context;
- helping staff and students to use GenAI tools as part of the literature and information searching process, such as using it for pre-search preparation (deciding on a research question, designing a search strategy etc.), and using generative AI technologies for searching for academic literature.

Our approach has been one of continuous development from the start. Our first online lecture for students was only 30 minutes long and focused on ethical and responsible use, as well as discussing what GenAI is. We then extended this to 60 minutes long, before creating three, hour-long sessions. The first of these covered ethical and responsible use concerns, the second use of GenAI for pre-search preparation and the final one use of GenAI for literature searching. These workshops were supported by a LibGuide (Using Generative AI Tools in Academic Work), which mirrors their content. We have since delivered further sessions aimed at dissertation students and integrated content on GenAI into embedded teaching within the curriculum. We have also created shareable slides to help ASL colleagues embed GenAI content into their own in-curriculum sessions.

Feelings

Our initial feelings about GenAI were a combination of concern, uncertainty, and curiosity about this new technology. We knew that it could change how people searched and used information, as well as potentially impacting our careers and professional work (5). However, we were uncertain about the full extent of any impact this new technology would have – other technologies have had less impact than initially thought (e.g., Second Life (6)). This uncertainty, combined with our lack of knowledge about GenAI at this point, meant that we initially felt intimidated by the task of teaching ourselves whilst also developing material to inform our colleagues and support our students.

As we learnt more, our confidence in our ability to inform others grew. However, GenAI technology has moved very quickly – ChatGPT, for example, has released 11 models since November 2022 (not including the mini and nano versions of each model (7)) and new tools are being developed very rapidly. This has led to us feeling overwhelmed with the task of keeping up to date and also frustrated at the constantly shifting landscape. It has sometimes felt that as soon as we feel confident in what we are teaching students, the technology changes or we learn some new theory that changes how we perceive GenAI.

We have also felt a tension when it comes to the ethical implications of this technology. There are many librar-

ians and researchers who are critical of GenAI and the implications for information literacy and education (e.g., Shah & Bender (8), Williamson (9)). Some librarians argue that we should not encourage their use in any way (Slater (10)). However, we feel we have a duty to help our students use them in a responsible way, so they can make informed decisions about whether to use them or not, and if they do use them, to use them well.

Evaluation

The creation of the LibGuide proved advantageous in offering a framework around which to develop workshop sessions and to provide a useful support resource. Since August 2025 the LibGuide has been viewed 7441 times, more than any of our other guides. However, with the changing nature of AI development, it has been difficult to keep our support materials up to date. We also saw strong interest in live sessions delivered; we saw 280 staff and students in academic year 2024/25 and received 100% positive feedback. However, the numbers registering for sessions has declined; for instance, in 2024/25 we had 121 people register for our autumn session on using GenAI for literature searching whilst in 2025 we only had 49 registrations.

In addition to our activities around designing and delivering AI sessions, the ASL AI Group has become a useful reservoir of collective AI knowledge and a point of contact for ASL and academic colleagues on AI-related matters. The experience has been helpful as a springboard to collaborate on AI-related projects, both within the University of Edinburgh and externally, and to promote our work via conference presentation and experience sharing. It has also developed as a useful working group and as an advocacy group within the Library and University Collections Team, contributing to broader discussions on AI development within the University's Information Services Group and beyond.

An ongoing challenge – which may be familiar to colleagues in other institutions – is that academics do not yet widely recognise AI literacy, particularly the use of AI-powered tools for information searching, as falling within librarians' professional expertise. This creates a barrier to embedding our support within curricula and engaging with faculty on AI-related pedagogical decisions. We need to actively work to shift this perception, demonstrating that critical evaluation of AI search tools is a natural extension of information literacy and firmly within our professional remit.

Analysis

The aspects of our work that succeeded early on – strong LibGuide views, positive feedback from sessions, and requests for input from AI policy groups in certain schools and from academics planning their teaching – drew largely on competencies already embedded in our practice. As Lo (4) observes, academic librarians are well placed to play a key role in supporting AI literacy: libraries have long championed information literacy and extending this to AI contexts where librarians' expertise in critical evaluation, bias detection, and ethical judgments about academic information tools proves highly relevant. Our existing skills and knowledge enabled us to act quickly in creating a guide and teaching content on AI use. We were also aided by our focus on a relatively bounded application of AI's use in literature searching. This reduced the need for cross-disciplinary agreement and enabled us to respond quickly to emerging AI trends in higher education, helping the Library make an early contribution to AI literacy at the University.

The challenges we encountered reveal deeper, systemic issues beyond workload alone. Keeping the LibGuide current highlights a mismatch between the slow pace and ongoing learning demands of updating it and a field where tools, capabilities, and expectations evolve frequently. This echoes Huang, Cox, and Cox's (11) finding that UK academic library strategies rarely mention AI explicitly – leaving individual teams to monitor fast-moving developments themselves, without sector-wide structures to support timely knowledge sharing and professional skills development. Declining attendance, despite evident need, may reflect students' perceived proficiency masking shallower literacy: Freeman's (12) HEPI/Kortext survey shows 92% of UK undergraduates now reporting AI tool use (up from 66% the prior year), yet only 36% reporting receiving institutional AI skills training.

The difficulty we have experienced in gaining traction within schools is also hard to explain but may be accounted for by Cox's (13) jurisdictional analysis, in which librarians' claim to emerging technology work is often contested – or simply unrecognised – by other professional groups, including academics. On reflection, the

issue may be less about the visibility of our work and more about a perception, across the University, that AI literacy does not fall within the Library's remit.

Conclusions and future actions

The final stages in Gibbs' reflective cycle are conclusions and an action plan. Ideally, you reflect on what you would do differently should a similar situation arise in the future (1). However, as this work is ongoing, we have reflected on how our experiences will inform future actions within the same project.

One of the most important conclusions we have come to is that keeping up with the fast-changing nature of AI requires planning for, rather than hoping to update materials on an ad-hoc basis. Whilst our focus on AI in literature searching has helped manage the workload, even keeping our small number of "deliverables" up to date has proven difficult. To counter this, we have already committed to re-working the LibGuide before the start of the 2026-2027 academic year. Once it has been updated, we will need to plan for how we can keep it up to date within our current workloads; this may mean only being able to commit to an annual review, even if tools change mid-year.

We also need to plan for how to transition some of our work to "business as usual". As AI has gained traction within higher education and society, it is no longer sufficient to provide support on AI information literacy in standalone, optional sessions. Rather, we need to consider how to ensure that all ASLs are confident in teaching the core skills and knowledge students need. Although we already provide reusable slides, and some librarians will be integrating this content into existing sessions, coverage across all disciplines is not uniform. We will therefore organise an ASL-focused session to discuss what is being taught by other ASLs and what support we can offer to them. Increasing the coverage of AI information literacy within in-curriculum sessions may also mitigate declining attendance at our optional sessions.

One final conclusion is that we should not assume that other people will recognise the links between AI literacy and the Library's areas of expertise. Whilst we could immediately see that aspects of GenAI were highly relevant to our work as librarians, this may not be apparent to academic colleagues. We therefore need to work alongside our ASL colleagues to increase promotion of our work in the wider University. This will not only increase knowledge of the support we can offer but also help us by increasing our knowledge of disciplinary differences in the approach to GenAI.

Discussion

Our hope in sharing this reflective report is to provide an insight into the approach taken by one small team within a large academic library. We chose Gibbs' Reflective Cycle due to its development within teacher education but also because it provides a clear structure for reflection. Whilst we have updated our material since it was initially developed, we have not collectively conducted intentional consideration of what has gone well or not, and what actions we need to take to address any issues. Writing this article has allowed us to do that and has encouraged us to consider how we can be more strategic in addressing the problems we have identified.

When our group initially formed, our first task was to inform ourselves about the capabilities and concerns around GenAI and then to support our students. Our reflection on the work of the past three years has shown that we have been successful in this; our live sessions and LibGuide have good engagement and we are a recognised point of contact for other members of staff in the library. However, conducting this reflection has also helped us to consider areas that still need improvement, and to consider what actions we might want to take next. This includes work to increase the visibility of our work outside of the library, including increasing our connections with faculty staff, and considering how we manage the need to keep materials current when the technology changes very rapidly. We also need to focus on working with our colleagues in the wider ASL team in order to increase the amount of AI literacy support they can provide as part of their own teaching.

The most profound observation we have taken from this process of reflection has been how difficult it is to integrate this work into our existing workload. A project should have an end point, but this work is ongoing and has indeed increased in complexity and volume. Moreover, "project work" should only comprise a small percentage of our workload. We are now at the point where integrating AI literacy into our work is no longer a project but a significant and time-consuming expan-

sion of our remit. Considering that burnout in academic librarians has been linked to role overload, serious consideration needs to be given to how our roles can change to integrate AI literacy in a sustainable way (14).

Submitted on invitation.
Accepted on 21May 2026.

.

REFERENCES

1. University of Edinburgh. Gibbs' Reflective Cycle. [Internet]. 2024 [cited 2026 April 22]. Available from: https://reflection.ed.ac.uk/reflectors-toolkit/reflecting-on-experience/gibbs-reflective-cycle
2. ASQ. Continuous improvement. [Internet]. 2026 [cited 2026 May 8]. Available from: https://asq.org/quality-resources/continuous-improvement?srsltid=AfmBOop8qkwnGHIwiU8Dkv-phYw-4fxMlRR6KZ0PDDgWK47Bw1hWJjUPx.
3. Miller JM, Ford SF, Yang A. Elevation through reflection: closing the circle to improve librarianship. Journal of the Medical Library Association: JMLA. 2020;108(3):353-63.
4. Lo LS. AI literacy: A guide for academic libraries. College & Research Libraries News. 2025;86(3):120-2.
5. IFLA. Developing a library strategic response to artificial intelligence. [Internet]. IFLA; 2023 [cited 2026 April 22]. 0Available from: https://www.ifla.org/g/ai/developing-a-library-strategic-response-to-artificial-intelligence/.
6. Lysiak L. 21st century innovations: Librarians, trend-watching, and the warning signs of fads. Pennsylvania Libraries: Research & Practice. 2020;8(2):130-7.
7. OpenAI. ChatGPT – Release Notes. [Internet]. 2026 [cited 2026 April 22]. Available from: https://help.openai.com/en/articles/6825453-chatgpt-release-notes.
8. Shah C, Bender EM. Situating search. Paper presented at: Proceedings of the 2022 Conference on human information interaction and retrieval; 2022 March 14-18; Regensburg, Germany: ACM, New York, NY, USA; 2022. p. 221-32.
9. Williamson B. Degenerative AI in education. 2023 June 30 [cited 2026 April 22]. In: Code acts in education [blog on the Internet]. Edinburgh: 2013-. Available from: https://codeactsineducation.wordpress.com/2023/06/30/degenerative-ai-in-education/.
10. Slater K. Against AI: Critical refusal in the library. Library Trends. 2025;73(4):588-608.
11. Huang Y, Cox AM, Cox J. Artificial intelligence in academic library strategy in the United Kingdom and the Mainland of China. The Journal of Academic Librarianship [Internet]. 2023 [cited 2026 April 22]; 49(6):102772.
12. Freeman J. Student generative AI survey 2025. Higher Education Policy Institute/Kortext; 2025 [cited 2026 April 22]. Available from: https://www.hepi.ac.uk/reports/student-generative-ai-survey-2025/.
13. Cox A. How artificial intelligence might change academic library work: Applying the competencies literature and the theory of the professions. Journal of the Association for Information Science and Technology. 2023;74(3):367-80.
14. De Prosperis M. Redesigning academic librarian roles for a post-crisis world: A role theory perspective on workload equity and occupational stress. The Journal of Academic Librarianship. 2026;52(2):103222.

Policies of selected Polish universities regarding the use of AI and the potential role of libraries in their implementation

Iwona Kosowska
Jagiellonian University Medical College, Medical Library, Kraków, Poland

Abstract
The development of generative artificial intelligence in recent years has led some Polish universities to work on certain policies regarding the use of artificial intelligence (AI) in academic environment. The analysis of these documents indicates which practices are permitted and which are prohibited. These practices are defined for students and lecturers, especially in the use of AI in written works (especially diploma theses), teaching and research work. A comparison was also made with 5 selected foreign universities. Generally, the policies adopted at universities do not prohibit the use of AI tools. There are some differences in the level of detail of policies. The most significant differences concern the methods of citing or reporting the use of AI. The role of libraries in this context could be to provide information about the potential benefits and risks of using AI and to raise awareness, especially in the context of ethical use of AI and plagiarism.

Key words: *artificial intelligence; AI policies; universities; libraries; medical libraries.*

Introduction

The enormous development and popularization of generative artificial intelligence (AI) in recent years has sparked a discussion about the opportunities and threats of using AI-based tools in higher education. Despite the widespread use of AI, many ethical and legal issues remain unresolved. In fact, there are still more questions than answers. Concerns about maintaining the appropriate quality of education and research continue. As a result, some universities started working on certain solutions regarding the use of AI, which took the form of official orders, recommendations, positions or guidelines for students and lecturers.

Methods

In November 2024, I conducted a review of selected existing documents of this type published in Poland. Based on "The Ranking of Academic Universities 2024" list published by Perspektywy Publishing House in 2024, the top 50 best universities in Poland (public and private) were selected. In total, there were 54 universities (5 universities took joint 50th place) (1). Additionally, in order to deepen the analysis with a broader perspective of medical universities, the list of universities was expanded to include 3 member universities of the Conference of Rectors of Academic Medical Schools (KRAUM), which were outside the top 50 of the Perspektywy Ranking (2). Investigating these documents showed the most common practices of selected Polish universities regarding the use of AI-based tools in teaching and scientific work. I compared the results of this study with AI guidelines at five universities from Great Britain, the Netherlands, Norway, Latvia and Czechia to perform a comparative analysis. I especially wanted to outline the potential role of libraries in implementing these policies. The results of this study were presented in the form of a presentation during the EAHIL 2025 Workshop, Lodz, Poland, from 9-13 June 2025.

Results

Of the 57 universities analyzed in November 2024, 40 do not have published this type of regulations, while 17 do. Access to 2 was possible only after logging in (Nicolaus Copernicus University in Toruń, Academy

Address for correspondence: Iwona Kosowska, Jagiellonian University Medical College, Medical Library, 7 Medyczna Str., 30-688 Kraków, Poland. E-mail: iwona.kosowska@uj.edu.pl

Journal of EAHIL 2026; Vol. 22 (2): 14-17 DOI: 10.32384/jeahil22718

of Physical Education in Katowice), which is why 15 universities were ultimately analyzed. These were: University of Warsaw, Jagiellonian University in Kraków, Warsaw University of Technology, Adam Mickiewicz University, Poznań, SGH Warsaw School of Economics, University of Gdańsk, Kozminski University, WSB University, University of Lodz, Medical University of Silesia, Poznań University of Economics and Business, Cardinal Stefan Wyszynski University in Warsaw, Wroclaw University of Economics and Business, Maria Curie-Sklodowska University and University of Rzeszów.

Of the 16 universities providing medical education that are members of The Conference of Rectors of Academic Medical Schools, only 4 have adopted AI policies. These are: Jagiellonian University in Kraków, University of Rzeszów, Medical University of Silesia and Nicolaus Copernicus University in Toruń (available only after logging in).

All the Polish policies, which I have examined, allow the use of AI. During the analysis of the content of the recommendations, I can distinguish the main areas on which the regulations focus.

Guidelines for the use of AI in thesis and other written works

The most common issue that the policies address are guidelines for the use of AI in thesis and other written works. The focus on written work is not surprising, given that AI easily generates texts. However, the problem is that it does not avoid providing false information and footnotes that look credible at first glance. In all the recommendations analysed, the use of AI in writing papers is not prohibited, but the decision about whether to allow students to use this tool or not is left to individual lecturers. First of all, it is emphasized that despite the use of artificial intelligence tools, the originality of thought must be preserved: the main theses, arguments and conclusions must come from the student. The lecturer, together with the student, determines the purpose and scope of using AI at work. The supervisor is mainly responsible for "the correct" use of artificial intelligence tools by the student. At the same time, the recommendations outline in very general terms the acceptable uses of AI. The examples mentioned include among others: translating the text into any foreign language, creating text plans, abstracts and summaries, reviewing sources, editing texts (changing style, correcting errors), helping with programming, generating graphics based on the author's hand-drawn drawings, creating diagrams, charts, figures, infographics according to detailed guidelines formulated by the author. As we see, examples of permitted uses of artificial intelligence tools do not include the generation of texts or text fragments. At the same time, the guidelines do not contain any clear statements prohibiting the generation of text. The exception is the guideline of the SGH Warsaw School of Economics. The guidelines directly prohibit: "generating text (paragraphs, chapters), generating the first version of the text and then editing it on one's own, expanding the existing content (e.g., using prompts such as: "add two sentences to the paragraph"); all substantive content in a written assignment, including both the author's own text and content cited from other authors, should be the result of the author's in-depth reflection" (3).

A very important issue raised by the guidelines is the way of reporting the use of artificial intelligence in written works. The policies are consistent that it is necessary to indicate the use of AI systems, because AI-generated content is not the author's work. The most common is to report the use of AI in the introduction or in the methods section. It should be reported: what kind of system was used, what was the purpose of using AI, what kind of method was used. Some policies provide more detailed guidance. It is required to post the prompts, which students wrote and the output, which they received. An example could be the regulation in force at the Adam Mickiewicz University in Poznań: "Fragments created using AI should be clearly marked in the work by providing the commands (prompts) used and/or the received output data (e.g. footnotes, appendix to the work)" (4).

Additionally, the Warsaw School of Economics points out that it is not allowed citing large-scale language models (LLM) or other generative AI tools as a source of information. It is worth emphasizing that this recommendation does not prohibit the use of AI in general, but as a source of information. This means that it can be a supporting tool, e.g. improving the style of the text (5). It is recommended to continuously monitor the student's progress in preparing the thesis to ensure that it is a student's work.

Guidelines for the use of AI in teaching and research

Another area on which recommendations focus is teaching and conducting scientific research.

The use of artificial intelligence in teaching is intended to make learning more attractive and engaging for students. Some examples of permitted uses of AI in teaching are: generating ideas in brainstorming, generating educational materials that engage students (generating questions, quizzes, scenarios). AI-powered test preparation and grading criteria are also permitted. AI can also be used to automatically grade short answers (6).

The use of artificial intelligence in research is a topic that is less frequently addressed by the policies. Examples such as using artificial intelligence to identify research trends, theories, research problems, research gaps, research methods were mentioned. Another area in which the use of AI tools is permitted is literature review and searching in collection of articles (e.g. using AI to extract key information from large collections). Scientists can also use AI to make research simulations (e.g. using AI to simulate different experimental scenarios and searching for the optimal) (7).

A short look at the different European universities

In my work in November 2024, I also looked at 5 foreign universities and their approach to the use of artificial intelligence. I wanted to highlight the most interesting AI-related issues from universities in The Netherlands, Norway, Latvia, Great Britain and Czechia.

Unlike Polish universities, University of Amsterdam policy initially in November 2024 doesn't allow students and lecturers to use AI tools. This was due to concerns about privacy and data security (8). However, the situation changed and currently the university has developed its own tool called UvA AI Chat. The general principle of using this tool can be summarized as follows: You can use UvA AI chat to support the learning process, but not to do the work. It was created with data security in mind, as the input data is not subsequently used to train the AI model (9).

A similar solution was previously implemented by the University of Oslo. One of the approved AI tools is GPT UiO. It is a privacy friendly chat for students and teachers of the University of Oslo. The main advantage of this tool is that it allows the use of OpenAI's ChatGPT within the requirements sets for privacy. The data entered will never be used to further train the model (10).

The University of Oxford followed a similar path in the 2025/2026 academic year. Students have the opportunity to use ChatGPT Edu, which works similarly to Chat GPT but ensures appropriate privacy and security standards (11).

University of Latvia guidelines differ in how AI usage is reported. It is recommended to provide a reference to use AI in the work. It is possible to cite artificial intelligence according to popular citation styles. You can refer to the AI tools you use using APA-style references, Harvard University-style reference and MLA-style both in the text and in the full citation (12).

Similarly at Charles University in Prague (Czechia) students should declare the use of AI tools in their work using guidelines (for example, those provided by Elsevier publishing). Charles University, which doesn't prohibit the use of AI, pays attention to creating an atmosphere of mutual trust in the academic environment, but warns off providing AI tools with information that includes the university's intellectual property (13).

Conclusions and discussion

In my opinion, the adoption of such regulations helps reduce information chaos, promotes transparency and helps avoid inappropriate practices. Unfortunately, the main problem with the policies is that most of them are very general, regardless of whether they were created at specialized universities or not. The frequently repeated recommendation is the need to provide training and knowledge about AI. Although they do not directly point to libraries as being responsible for these tasks, I think libraries can have a huge information and training role to play and in many libraries such activities are beginning to be undertaken. However, this requires that librarians themselves have such knowledge. Firstly, together with the academic staff we can provide information on both the risks and benefits of using AI. It is important to raise awareness, especially in the context of ethical use of AI and plagiarism. Definitely we should promote a critical approach to AI-generated content and recommend double checking. Our task can also be to inform about the methods of reporting

the use of AI. The role of library and librarians may be identifying useful AI-based tools and organized AI-literacy training. At the same time, we should popularize using various information resources provided free of charge by libraries. We can also consider providing access to paid AI tools.

Submitted on invitation.
Accepted on 11 May 2026.

.

REFERENCES

1. University Ranking 2024. [Internet] [cited 2026 Apr 12]. Available from: https://2024.ranking.perspektywy.org/ranking/university-ranking/
2. Członkowie zwyczajni KRAUM [Internet] [cited 2026 Apr 26]
   Available from: https://kraum.org.pl/?page_id=2270
3. SGH Warsaw School of Economics. Warsaw. Appendix to Rector's Regulation No. 5 of the SGH Warsaw School of Economics of 05 February 2024. Principles of using artificial intelligence in the preparation of written assignments at the SGH Warsaw School of Economics [Internet] [cited 2026 Apr 26].
   Available from: https://bap.sgh.waw.pl/d/8351/5/
4. Adam Mickiewicz University, Poznań. Poznan. Zarządzenie nr 426/2023/2024 Rektora Uniwersytetu im. Adama Mickiewicza w Poznaniu z dnia 20 marca 2024 roku w sprawie zasad korzystania z systemów sztucznej inteligencji w procesie kształcenia studentów [Internet] [cited 2026 Apr 26]. Available from: https://amu.edu.pl/__data/assets/pdf_file/0026/536372/ZR-426-2023-2024.pdf
5. SGH Warsaw School of Economics. Warsaw. Appendix to Rector's Regulation No. 5 of the SGH Warsaw School of Economics of 05 February 2024. Principles of using artificial intelligence in the preparation of written assignments at the SGH Warsaw School of Economics [Internet] [cited 2026 Apr 26].
   Available from: https://bap.sgh.waw.pl/d/8351/5/
6. Cardinal Stefan Wyszynski University in Warsaw. Warsaw. ZARZĄDZENIE Nr 70/2024 Rektora Uniwersytetu Kardynała Stefana Wyszyńskiego w Warszawie z dnia 8 listopada 2024 r. w sprawie zasad wykorzystywania narzędzi opartych na sztucznej inteligencji w dydaktyce i nauce [Internet] [cited 2026 May 03]. Available from: https://monitor.uksw.edu.pl/docs/8047
7. Adam Mickiewicz University, Poznań. Poznan. Zarządzenie nr 426/2023/2024 Rektora Uniwersytetu im. Adama Mickiewicza w Poznaniu z dnia 20 marca 2024 roku w sprawie zasad korzystania z systemów sztucznej inteligencji w procesie kształcenia studentów [Internet] [cited 2026 Apr 26]. Available from: https://amu.edu.pl/__data/assets/pdf_file/0026/536372/ZR-426-2023-2024.pdf
8. University of Amsterdam. Amsterdam. AI tools and your studies [Internet] [cited 2026 Apr 26]. Available from: https://student.uva.nl/en/information/ai-tools-and-your-studies
9. University of Amsterdam. Amsterdam. Background to UvA AI Chat [Internet] [cited 2026 Apr 26]. Available from: https://student.uva.nl/en/information/background-to-uva-ai-chat
10. University of Oslo. Oslo. How to use AI as a teacher [Internet] [cited 2026 Apr 26].
    Available from: https://www.uio.no/english/services/ai/teacher/index.html
11. University of Oxford. Oxford. ChatGPT Edu available to all students [Internet] [cited 2026 May 03]. Available from: https://www.ox.ac.uk/students/news/2025-09-24-chatgpt-edu-available-all-students
12. University of Latvia. Riga. Artificial intelligence in UL [Internet] [cited 2026 Apr 26]. Available from: https://www.lu.lv/en/studies/study-process/artificial-intelligence-in-ul/#c148098
13. Charles University. Prague. Artificial intelligence (AI). Recommendations for the academic members of staff of Charles University. Prepared by the working group for Distance Learning. Version 1 – June 2023 [Internet] [cited 2026 Apr 26]. Available from: https://ai.cuni.cz/AI-12-version1-ai_elearning_en.pdf

Reflections on evaluating the utility of an LLM for keywording health research

Claire Stansfield and Ailbhe Finnerty Mutlu
Evidence for Policy and Practice Information and Co-ordinating (EPPI) Centre,
UCL Social Research Institute, University College London (UCL), United Kingdom

Abstract
Generative artificial intelligence shows promise for rapidly annotating collections of research at scale. We reflected on our experiences of evaluating the ability of a Large Language Model (LLM) to apply predefined keywords to records of health research in the context of an evidence repository on vaccine research and research registers of health promotion effectiveness. Five aspects of evaluation helped us articulate key considerations and challenges across the use cases: 1) cyclical prompt development, 2) data availability and quality, 3) performance benchmarks and expectations, 4) task complexity and perspective and 5) workflows and tools.

Key words: *Large Language Models; databases as topic; evaluation studies as topic.*

Introduction

Our context is specialist evidence repositories and research registers that are compiled for purposes of research discovery, such as investigating the research landscape and informing evidence synthesis. They may contain hundreds or thousands of research records, and be annotated with predefined keywords (also called labels, codes) describing broad characteristics of the research, such as study design, topic or thematic focus, or detailed aspects such as who delivered an intervention, country it was delivered, dosage etc. These keywords are often bespoke to the context of the repository and might be assigned from title and abstracts or full-texts. Using an LLM to apply keywording is particularly appealing for achieving rapid keywording at scale in regularly updated (or "living") systems. However, there are challenges in evaluating acceptable performance for accuracy and reliability with a view to implementation. We have separate experiences of using this approach for two use cases and highlight our key reflections here. For a broader perspective, the living documents produced within Responsible Use of AI in Evidence Synthesis (RAISE) is an important reference point. RAISE 2 focuses building and evaluating AI evidence synthesis tools (1).

We begin by outlining the use cases and the approach taken to apply an LLM, followed by reflections across five aspects of evaluation: 1) cyclical prompt development, 2) data availability and quality, 3) performance benchmarks and expectations, 4) task complexity and perspective and 5) workflows and tools.

Context

One author (AFM) is evaluating the utility of an LLM to apply keywords based on full text, to a living evidence repository of Human Papillomavirus (HPV) vaccine research (2, 3). The repository contains research articles on HPV vaccine delivery in low and middle income countries and forms part of an evidence bank that will be updated through regular searches. The bibliographic records have been keyworded according to a taxonomy of HPV-relevant terms. These keywords enable use of the repository to identify relevant research and research gaps to inform research and practice. The first iteration of the HPV taxonomy had around 230 keywords, including 127 individual country names, these keywords ranged from thematic focus, delivery of vaccines to descriptions of the populations of interest. The taxonomy is being further developed by a team of experts in the HPV vaccine community

Address for correspondence: Claire Stansfield, Evidence for Policy and Practice Information and Co-ordinating (EPPI) Centre, UCL Social Research Institute, University College London (UCL), London, United Kingdom.
E-mail: c.stansfield@ucl.ac.uk.

Journal of EAHIL 2026; Vol. 22 (2): 18-23 DOI: 10.32384/jeahil22724

to add more detailed keywords. The initial keywording was done by single human reviewers and pairs of humans independently keyworded a percentage of records (10-20%) that was done by an LLM.

Another author (CS) is evaluating keywords based on titles and abstracts in two registers of health promotion effectiveness, Trials Register of Promoting Health Interventions (TRoPHI) and Database of promoting health effectiveness reviews (DoPHER) (4). The TRoPHI register contains over 25,000 records of research using controlled trials, and descriptive keywords to describe the content related to thematic focus, study design, population groups, and geographical region. Thematic focus forms the largest keyword set comprising 32 keywords on particular areas for health promotion (e.g. injury, mental health, cancer). DoPHER contains over 10,000 reviews of effectiveness and currently has no keywording in the public version. We previously described our planned steps for full automation (5).

Each use case had a bespoke set of keywords that were defined (prior to automation) to support consistent and reliable annotation of multiple keywords to a research record by humans. Each keyword was converted into an LLM prompt in the form of a command or question to guide the LLM to produce a desired action or answer (6). The cases here focus on prompts requiring the LLM to provide a true or false response, so that if an answer is true, the keyword is annotated to the record about that research. The prompts generally followed a similar structure, detailing the role of the LLM and the criteria to be met. One prompt was used per

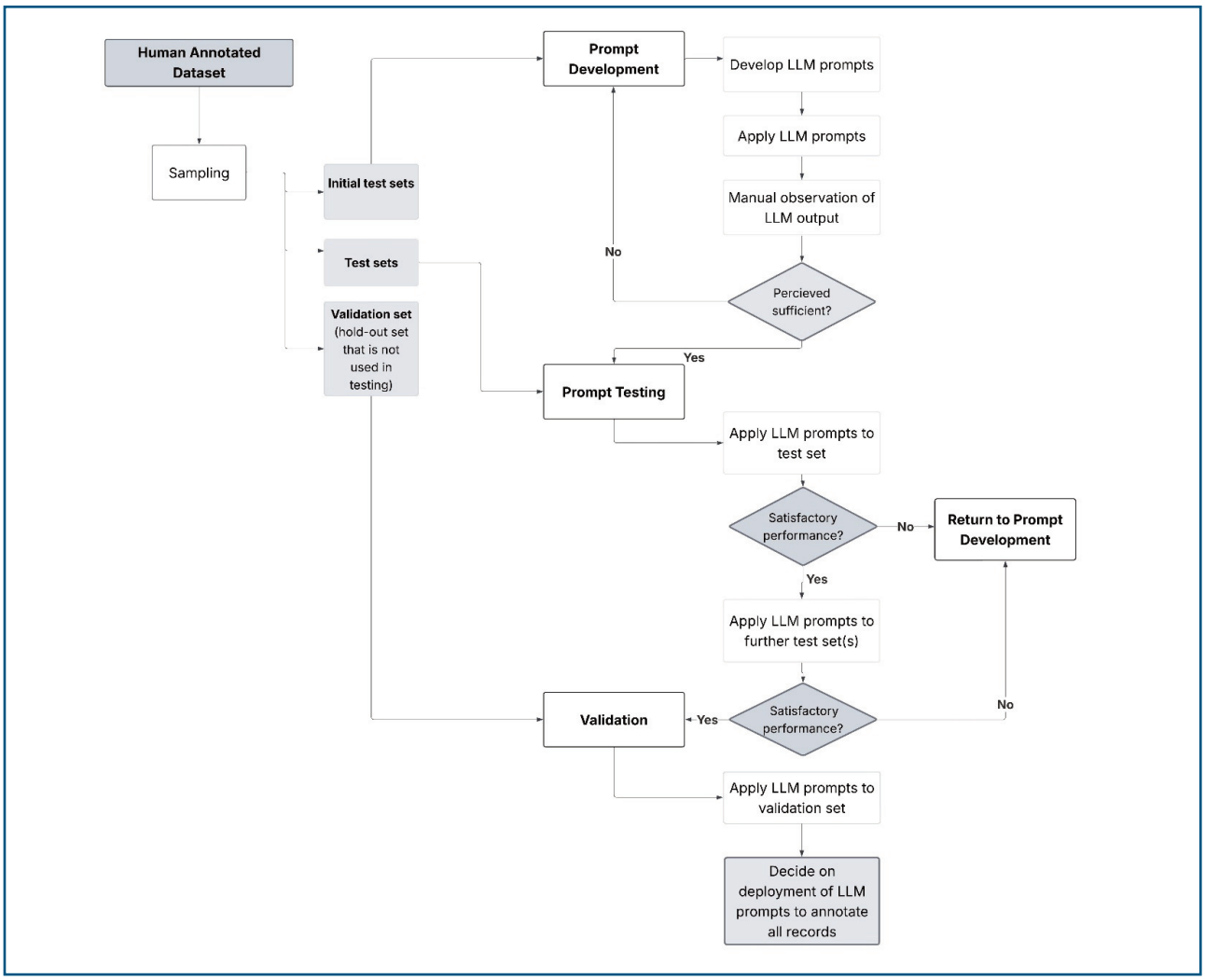

Fig. 1. *The cyclical process of prompt development and evaluation.*

keyword. For example, in TRoPHI, over 90 question prompts are applied to one title and abstract record.
Mostly, our evaluations focused on the reliability of an individual keyword being annotated to a single research record (title/abstract or full-text article), rather than the collective set of keywords annotated to a single record.
We used the GPT 4.1 LLM within EPPI-Reviewer, using parameters fixed within this interface (7, 8), as this interface is developed within our research centre. Our reflections here relate to evaluations generally rather than on the performance of a specific LLM model.

Cyclical prompt development and evaluation

In both cases, we began by iteratively writing prompts and comparing performance of the LLM with prior human annotations on relatively small samples of records. Most records had multiple keywords assigned to them, it was only possible to manually check reasons for discrepancies over relatively small samples, as it involved considerable time checking the text that the keywords were applied to (either full text or title/abstract). Once the prompts were perceived as sufficient from this first step, they were evaluated on further samples to obtain measures of reliability compared with humans, followed by further investigations, and cycles of revision and testing of prompts. Strategies to improve prompts included: sense checking with another human, asking an LLM to edit prompts, and trial and error observations of what seemed to work on small samples. An overview of the process is shown in *Figure 1.*
Repeated cycles of prompt development and testing proved challenging as it required having a representative sample of the keywords. It was important to evaluate new prompts on new, previously unseen, sample records that had not informed prompt refinements and these samples needed to represent all the keywords being tested. Ideally these samples should be sufficiently large but depended on the data or the resources available to manually annotate new data.
An ongoing challenge is deciding how much to iterate prompts or when to stop. Identifying content within a sample can be useful to improve a prompt (e.g. setting out certain ways authors describe an intervention or population) but care is needed to avoid being too specific in order to anticipate other relevant scenarios that may apply to that keyword. Stopping requires accepting limitations of the possibility of the prompt at the current time or abandoning prompts that perform less well. Another aspect is that performance differs between LLM models and if tested, can add further cycles within the prompt development and evaluation process.

Data quality and availability

Ensuring there is sufficient volume of good quality data of records with keyword annotations with which to compare to the LLM generated data is important. However, this is not always available for various reasons. For example, having all data annotated independently by multiple humans is ideal, though requires resources to generate. If this data is not available there should be at least a proportion of records that are annotated by more than one human, and the reliability of these annotations between humans should be calculated. There also needs to be sufficient representation of all the keywords in the data including those that are rarely used. Producing truly random samples of records is not always possible due to factors such as small sample sizes or uneven distributions of keywords across samples.
In the case of TRoPHI, random and stratified samples of dual keywording by humans were generated (using approximately 700 records), and pairwise comparisons of humans and LLM data were made. However, when 200 records from this set became the only validation samples (that had not been used to inform prompt refinement), some keywords were underrepresented owing the smaller size of set. These sets were compared with larger samples of single human and LLM data (approximately 3,000) to observe performance at a larger scale. However, we recognise that these volumes of data are not routinely available, including our other use cases.
Furthermore, it is recognised that there are often inconsistencies between humans when applying multiple keywords for this type of task (9, 10). The quality and structure of the text itself can vary. The interpretation of text and interpretation of the keywords can vary between humans. Humans may also have varying level of expertise with the keywording task. There are studies which have identified that the human is influenced by an AI decision if they are presented with AI judgements for the same task (9). Sometimes the human an-

notators may also have used text beyond that being used by the LLM, such as a journal name or author keywords. There may also be inconsistencies applied by the same human over time, or changes in interpretation of how keywords are applied over time, or human fatigue may be another factor in identifying all keywords. For example, in the HPV repository, when there were many relevant keywords relating to a full-text document, humans missed one or two which were identified by the LLM.

Performance benchmarks and expectations

RAISE 2 (1) includes a taxonomy of common performance metrics in evidence synthesis. In our use cases, the benchmarks used to evaluate performance of LLM needed to reflect variability in the accuracy and consistency in the comparison data that was available. Inter-rater reliability (IRR) was compared to evaluate the reliability of keywording instructions, rather than directly comparing the accuracy of the LLM with human data. This is consistent with the domains of subject classification (10), content analysis (12), where pre-existing human annotations of multiple keywords are not a sufficient gold standard. We used IRR to assess the performance of records keyworded by human and the LLM, or where available, between two humans and two LLMs. The IRR is calculated from the rates of agreement and disagreement of the application of each keyword and evaluates whether different observers of the same records yield the same data within accepted levels of error (12). An advantage is that this allows comparison of consistency, and comparison over large sets of records. Disadvantages are that it might not detect consistent errors unless checking individual instances of disagreement or it does not work well for 'rarely' used keywords that are not sufficiently represented in evaluation. IRR metrics vary on the level to which they adjust for randomness in their agreements and disagreements, and the number of people or entities the IRR is being measured between (11).

Across our evaluations Krippendorff's alpha and Fleiss Kappa were used based on pair-wise comparisons (e.g. human-human, human-LLM, LLM-LLM). Evaluating IRR provided an overview of performance at scale, and comparison across different pairs of annotators. However, we are aware that sample size of some (under) represented keywords could affect some results. Furthermore, this metric alone does not provide a complete picture, as it does not highlight consistent omissions or show nuances of how keywords are applied. We found it useful to review how keywording was applied in individual cases, though such investigations were targeted owing to the time involved. A further challenge is the level of performance to accept. Acceptable performance of annotation tasks is ill-defined, and expectations may vary regarding use cases. The type of task and the real-world implications inform the level of acceptable reliability and tasks which can have a more serious consequence require a higher level of reliability (11). For example, if the evidence that is aligned to a set of keywords is intended to be relied upon to inform policy and decision making (without using further approaches for discovery), then a high standard is required as the consequences are a hindrance to discovery and there is potential for altering the findings of a systematic review (this is highlighted in RAISE 2 (1)). On the other hand, if the use case is more for exploratory discovery, rapid mapping, or if the limitations of keywording are communicated then performance standards are more arbitrary. For example, if keywording of a record acts as a signpost it is less critical if it is liberally applied (as researchers can filter out records manually following their own checks), and also having other options available (such as free-text searches) to discover useful research not keyworded.

Task complexity and perspective

The complexity of the task is another consideration. Identifying records which can be considered edge cases of records, within the register or repository and where the keywording encompasses a complex area, can be helpful to incorporate as examples of edge cases into training material given to the human annotators and the LLM prompts in order to improve reliability across a wide variety of records. However, there are some concepts which are complicated for an LLM to infer from a piece of text and refinement of prompts might not sufficiently improve performance. We found this in some cases where human agreement was low (such as assignment to certain topic areas), which indicates a high level of task difficulty to infer from text. Furthermore, humans and LLMs address tasks in different ways and types of difficulties differ. For some keywords, we found the LLM prompt needed to have clearer instructions, and contain more context about a

keyword, compared with human. Having a keyword category for 'unclear' was useful to isolate records that the LLM had not been able to annotate.

We also needed to be mindful of being influenced in favour of wanting an LLM to perform well in order to improve efficiencies, and mitigating this with appropriate testing. We observed when using information from the title/abstract that some keywords were more suited to describe primary research than reviews, as the abstract of the primary research contained specific information on a population or context compared with a systematic review covering multiple populations or contexts. This highlighted the importance of testing prompts that we wanted to re-use within their context of use.

Workflows and tools

A key facilitator to evaluation is having efficient workflows and tools to conduct the evaluations. It was important to have a process to manage different versions of the test LLM prompts, collate decisions of the human and LLM model about each keyword for each research record, and calculate performance metrics for each sample. It took time to develop suitable templates and processes; for example, converting outputs of LLM and human decisions, and calculating inter-rater reliability were done through developing bespoke Excel templates for each project. However, we are aware of a shiny app being developed that would improve this process. As we progressed more tools have become available to support evaluations within the interface we used, such as a tool to compare the performance of different LLMs, and multiple iterations of LLMs, across small samples (13). Other development underway includes an opensource data extraction and evaluation toolkit that will allow easy iteration of instructions for keywording and data extraction as well as their evaluation through comparison of human annotated data and a human-in-the-loop evaluation process.

Conclusion

Implementing LLM predefined keywording currently requires considerable development and testing time and may be an efficient method for automating some types of annotations for evidence repositories and research registers at scale. The five areas reflected upon here helped us articulate some considerations and challenges of evaluation and implementation. It also provides a structure for further conversations beyond individual use cases.

Funding

AFM has funding from Wellcome Trust (313586/Z/24/Z) and CS has funding from National Institute for Health and Care Research Policy Review Facility (NIHR200701).

Acknowledgements

A number of colleagues from EPPI Centre, UCL and from the London-York PRP Evidence Review Facility have participated in discussions that have informed this work.

Submitted on invitation.
Accepted on 11 May 2026.

REFERENCES

1. Thomas J, Hair K, Noel-Storr, A. et al. Responsible use of AI in evidence SynthEsis (RAISE 2026): building and evaluating AI evidence synthesis tools (version 3; updated 13 March 2026). In: Open Science Framework [https://osf.io/fwaud/files/wbf4y], Washington DC: Center for Open Science. DOI 10.17605/OSF.IO/FWAUD
2. Eni TE, Bond M, Finnerty Mutlu A, Tifuh OB, Wehtuogenyi TR, Leopold L, Lontum A, Okwen TT, Mbinkar BN, Cholong BJ, Okwen PM, Yong NB, Jeppesen BT. Living evidence repository and evidence maps on human papillomavirus vaccines delivery. 2025. EPPI Visualiser database. https://eppi.ioe.ac.uk/eppi-vis/login/open?webdbid=923
3. HPV living evidence partnership. Available from: https://aliveevidence.org/hpv-living-evidence/
4. EPPI Centre. Databases. Available from: https://eppi.ioe.ac.uk/cms/databases
5. Stansfield C, Thomas J. Applying automation to maintain research registers in health promotion. J Eur Assoc Health Info Libr. 2024;20(2):26-9.

6. Homiar A, Thomas J, Ostinelli EG, Kennett J, Friedrich C, Cuijpers P, et al. Development and evaluation of prompts for a large language model to screen titles and abstracts in a living systematic review. BMJ Mental Health. 2025;28:e301762.
7. Thomas J, Graziosi S, Brunton J, Ghouze Z, O'-Driscoll P, Bond M, Koryakina A. EPPI-Reviewer: advanced software for systematic reviews, maps and evidence synthesis. EPPI Centre, UCL Social Research Institute, University College London; 2023.
8. EPPI-Reviewer. Automated screening and data extraction using Large Language Models. Available from: https://eppi.ioe.ac.uk/cms/er4/Help/Automated-coding-using-LLMs
9. Gartlehner G, Kahwati L, Nussbaumer-Streit B, et al. From promise to practice: challenges and pitfalls in the evaluation of large language model for data extraction in evidence synthesis. BMJ Evidence-Based Medicine. 2025;30:385-9.
10. Golub K, Soergel D, Buchanan G, Tudhope D, Lykke M, Hiom D. A framework for evaluating automatic indexing or classification in the context of retrieval. Journal of the Association for Information Science and Technology. 2016;67(1):3-16.
11. Krippendorff K. Reliability. In: Content analysis: An introduction to its methodology. (4. ed.). SAGE Publications, Inc; 2019. 277-360.
12. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. Communication Methods and Measures. 2007;1(1):77-89.
doi: 10.1080/19312450709336664
13. EPPI-Reviewer. Latest changes (16/04/2026 - V6.18.0.0).
Available from:
https://eppi.ioe.ac.uk/cms/er4/Help/Version-History-Announcements/Latest-Changes-16-04-2026-V61800

Letter from the President

Francesca Gualtieri

Library and Archives Rottapharm Biotech SpA, Monza, Italy
francesca.gualtieri@rottapharmbiotech.com

Dear EAHIL Members, Colleagues

As I write, we are preparing to attend the EAHIL 2026 Congress in Antalya (Turkey), from 9–13 June 2026. By the time you read this letter, the event will have concluded, and I would like to extend my sincere gratitude to Antalya Bilim University and Marmara University for their kind hosting of this congress, to the Scientific Committee, the EAHIL Board, our sponsors, the organizing company Sanatkar Reklam Agency, and to everyone who has contributed to making this event possible.

The 2026 theme was "Bridges to New Horizons: Innovations in Health Information and Medical Libraries", a very inspiring theme – as we look ahead, our association is approaching a new challenge in exploring different ways of working. We are entering a transformative era, driven by advancements that will fundamentally redefine our profession. Artificial Intelligence is set to change our lives and our jobs as profoundly as the internet did decades ago.

While this technological shift will bring uncertainty in some of us, it also presents an extraordinary opportunity. Librarians will serve as the essential bridge between human expertise and machine intelligence, guiding how knowledge is discovered and utilized.

This will be our next great opportunity. As information professionals, we will guide our students, researchers, and clinicians in navigating this complex new landscape. Our role will be crucial in teaching them how to use these tools with deep critical appraisal.

Our recent experience in Antalya has given us the true power of connection. It reminded us of the immense value of sharing news and ideas with one another. This network allows us to receive wise, practical contributions from colleagues who have already faced these new digital tools and can guide us with their firsthand experience.

To turn these challenges into opportunities, we need your active voice. I invite all of you to share your ideas, insights, and solutions. Please get involved and collaborate with our Special Interest Groups (SIGs) through their activities throughout the year. Your contribution is vital to shaping our collective future.

EAHIL is committed to supporting you through this transition. Together, we will turn this challenge into a powerful evolution for medical libraries worldwide.

In 2027 EAHIL will be led by a new President, and I'd like to wish him to be a visionary leader to navigate the future – he comes from Lodz that means boat as you can remember – with wisdom and dedication, and build at the same time a safe port for the association, we will help him into this goal. I will give you more detailed information on the recent EAHIL election in my next Letter in September.

Thank you for your continued dedication and vision.

Warm regards,

Francesca Gualtieri

EAHIL President

Annarita Barbaro
Istituto Superiore di Sanità, Rome, Italy
annarita.barbaro@iss.it

Dear colleagues,

in this issue's "Publications and New Products" column, I am pleased to present a selection of recent publications and initiatives that have caught my attention and that I hope you will find interesting. The topics explored reflect several of the key challenges and developments shaping our daily work, including artificial intelligence, research integrity, and Open Science. Among the recommended readings, I would particularly like to highlight the article on hyper-prolific authors—an important research integrity issue that, in my view, still does not receive the attention it deserves. Enjoy!

FROM THE WEB

- **An analysis of the paper mills market: BuyTheBy**
  Reese Richardson, Spencer Hong, and Anna Abalkina have created a dataset called "BuyTheBy" (available on Zenodo: https://zenodo.org/records/19684278), which systematically analyses the paper mill market. Paper mills are fraudulent companies that sell scientific articles or authorship positions on papers that are ready to be, or in the process of being, published. Paper mills have existed for years and have become especially widespread in countries where academic careers depend heavily on the number of published works, offering shortcuts to researchers (or aspiring researchers) who want to obtain publications quickly. BuyTheBy collects more than 18,000 advertisements from seven paper mills operating in different countries (including India, Russia, and Ukraine), gathered between 2020 and 2026. Analysis of the dataset shows that prices vary widely: from about $56 to over $5,600 to be listed as the first author. According to the dataset's creators—although they affirm that this point still requires further investigation—some advertisements appear to be linked to published articles. Combating the phenomenon seems difficult, as this market evolves rapidly, also thanks to the use of artificial intelligence. For further reading related to the subject: A preprint on the dataset, published on arXiv: BuyTheBy: A dataset of 18,710 text-based paper mill advertisements with 51,812 timestamped prices; a preprint on paper mills in conferences, published on arXiv: Opening Pandora's box: Paper mills in conference proceedings

- **Living guidelines on the responsible use of generative AI in research**
  The European Research Area Forum has published an update of the ERA Living Guidelines on the responsible use of Generative AI in research with the goal of helping the scientific community to use generative AI in a responsible manner. While non-binding, they should be considered as a supporting tool for researchers, research organisations and research funding bodies, including the ones applying to the European Framework Programme for Research and Innovation. These Guidelines consider key principles on research integrity as well as already existing frameworks for the use of AI in general and in research specifically. Given the dynamic nature of this technology and its surrounding policy landscape, these guidelines will be updated on a regular basis to ensure that they remain a useful resource for researchers and organisations.

Bengaluru Roadmap and Action Plan on Diamond Open Access

The Bengaluru Roadmap and Action Plan on Diamond Open Access is the outcome document of the 3rd Global Summit on Diamond Open Access, organised in Bengaluru, India, from 2 to 6 February 2026. The Roadmap affirms that Diamond Open Access is a necessary structural transformation of scholarly communication, which requires the alignment of policy, funding and evaluation systems, and the necessary governance to enable Diamond Open Access to function as the primary model for disseminating publicly funded knowledge. The Roadmap advances a coordinated global vision across six priority action areas: integrating Diamond Open Access into national policy and legal frameworks; redirecting publishing expenditure toward community-governed infrastructures; reforming research evaluation systems so that they recognise Diamond Open Access publications, multilingual scholarship, and community-governed venues; strengthening shared infrastructure and interoperability; recognising and supporting the human labour behind scholarly communication; and promoting linguistic diversity and marginalised knowledge systems.

- **Is Subscribe to Open Good for Libraries?**

  The article Is Subscribe to Open Good for Libraries? discusses the strengths and weaknesses of the Subscribe to Open (S2O) model for academic libraries. In S2O, journals become openly accessible if enough libraries continue their subscriptions, allowing publishers to fund open access without charging authors APCs (Article Processing Charges). The authors, three librarians from different research libraries, argue that S2O has several advantages for libraries: it supports equitable open access, avoids shifting costs to researchers, and helps maintain sustainable scholarly publishing. However, they also point out challenges, such as uncertainty around long-term funding, the risk of "free riders" (institutions benefiting without subscribing), and the difficulty libraries face in evaluating the real value of supporting S2O initiatives.

- **DORA course on responsible research assessment**

  The Declaration on Research Assessment (DORA) responsible research assessment (RRA) course aims to empower anyone to challenge existing assumptions around research quality and impact, by introducing them to the concept of responsible research assessment, its foundational principles, and how RRA aims to improve evaluation practices. The course is free, self-paced, and designed to pick and choose which lessons are most relevant and useful. Each lesson includes a glossary of key terms and annotated bibliography. The course is designed for anyone working in research, research administration or leadership, research funding or policy, scholarly publishing or communication, or metrics providers who seek to drive positive change in the research system.

- **The Economic Benefits of Open Science**

  PLOS commissioned Technopolis to provide an independent study on the economic benefits of Open Science (OS) to clarify where economic value is created in an OS ecosystem, what conditions enable it, and where structural barriers and cost pressures remain. This report, The economic benefits of Open Science, examines the economic implications of OS and what it would mean to move towards a research ecosystem in which research elements, including data, code, software, workflows, methods, and publications, are openly shared and valued. The study focuses on three key areas: understanding the impacts of migrating to an OS ecosystem in which all types of research outputs are valued; exploring the costs of such a transition for the scholarly communication ecosystem; and evaluating the variety of economic impacts arising from an OS transition. To ensure the findings reflect the breadth of OS beyond publications alone, the study includes five illustrative case studies spanning digital collections, computational workflows, open-source software, an open large language model, and open training resources.

READING SUGGESTIONS

- Gibney E. Open-source AI tool beats giant LLMs in literature reviews — and gets citations right. Nature News, 04 February 2026
- Almeida JM, Antelmi A, Gonçalves MA et al. The complex ecosystem of hyperprolific authors. Scientometrics 131, 1847–1895 (2026)
- Knowledge Exchange, Fraser, K., Pinfield, S., & Chiarelli, A. (2026). Charting New Paths: The Promise of Alternative Publishing Practices. Zenodo.
- Naddaf M, Quill E. Hallucinated citations are polluting the scientific literature. What can be done? Nature News, 01 April 2026

SOME FORTHCOMING EVENTS

DCMI 2026 Conference
August 3 – 7, Seoul, South Korea
DCMI 2026, the twenty-fourth International Conference on Dublin Core and Metadata Applications, invites researchers, practitioners, and experts from diverse domains to explore the dynamic landscape of metadata in the theme of "Meaning-Driven AI: Using Metadata to Align Systems with Human Values". More info at this link

IFLA 2026 World Library and Information Congress
August 10-13, Busan, South Korea
The theme of the Congress, "Libraries Powering Transformation," reflects the vision of the IFLA Strategy 2024–2029: "Sustainable futures for all through knowledge and information." For registration and more information click here.

PUBMET. 13th Conference on Scholarly Communication in the Context of Open Science
September 9-11, Zadar, Croatia
The main Conference topic is "Communities in Open Scholarly Communication: Governance, Responsibility, and Shared Infrastructures". The PUBMET 2026 proposes a reflection on how Open Science works in practice, including questions of responsibility, ownership, governance, inclusion, and sustainability, as well as the gap between Open Science policies and everyday research practices. More info here

Peer review week
September 14-18 everywhere
This year's theme is "Peer Review Capacity: Volume, Speed, and Quality" addressing the sustainability of peer review under increased pressure, ensuring robust, fair, and high-quality assessments. To be updated on all the initiatives click here
OASPA annual conference
September 21-23, Zagreb, Croatia
For registration and more information click here

19th IFLA Interlending and Document Supply Conference
September 28-20 Rome, Italy
The conference focus is "Knowledge Without Walls," targeting international delegates to discuss library operations, technology, and resource sharing. For registration and more information click here

Please feel free to contact me (annarita.barbaro@iss.it) if you have any further suggestion about initiatives or events you would like to promote

Advertisements
To advertise in this journal, please contact **eahil-secr@lists.eahil.eu**

Instructions to Authors
Instructions to Authors are available online at **www.eahil.eu**. For further information please contact Federica Napolitani, Chief Editor of *JEAHIL* **federica.napolitani@iss.it**

Editorial layout and pagination: **De Vittoria srl**, Rome, Italy