



## UvA-DARE (Digital Academic Repository)

### The Librarian Skillset and the Needs of Artificial Intelligence

Harper, Corey; Groth, Paul

**DOI**

[10.1080/01639374.2025.2539787](https://doi.org/10.1080/01639374.2025.2539787)

**Publication date**

2025

**Document Version**

Final published version

**Published in**

Cataloging & Classification Quarterly

**License**

CC BY

[Link to publication](#)

**Citation for published version (APA):**

Harper, C., & Groth, P. (2025). The Librarian Skillset and the Needs of Artificial Intelligence. *Cataloging & Classification Quarterly*, 63(6-7), 546-565.  
<https://doi.org/10.1080/01639374.2025.2539787>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

# The Librarian Skillset and the Needs of Artificial Intelligence

Corey Harper<sup>a,b</sup> and Paul Groth<sup>b</sup>

<sup>a</sup>Elsevier Discovery Lab, New York, NY, USA; <sup>b</sup>University of Amsterdam, Amsterdam, Netherlands

## ABSTRACT

New roles in AI are emerging throughout Industry, Academia, and Government. From responsible AI management to AI systems engineering, these roles require new skills, many of which align closely with librarian competencies. Roles in AI where librarian skills are especially relevant include organizing information, facilitating documentation, assessing societal impacts, and working with communities of both users and developers. In this paper, we provide mappings from ALA-defined librarian competencies and Library and Information Studies curricula to an inventory of AI competencies. We provide an analysis of five areas of alignment, as in correspondence, including responsible AI and social justice and data engineering and evaluation.

## ARTICLE HISTORY

Received April 2025  
Revised July 2025  
Accepted July 2025

## KEYWORDS

Artificial intelligence; library science education; skills and competencies; responsible AI; social justice

## Introduction

In the Spring of 2022, Huggingface, the Artificial Intelligence (AI) and Machine Learning community and platform, posted a job listing for a Machine Learning Librarian.<sup>1</sup> The post notes that Huggingface sees “an increasing need for curating documentation and general knowledge about all of the many models, datasets, and application contexts for this technology.” They are recognizing the vital nature of librarian skillsets – skill in information organization, knowledge discovery, and cataloging – in helping organize and provide access to the rapidly growing resources used in AI.<sup>2</sup> Since this posting, a growing number of jobs have appeared for AI Librarians, though the bulk of these are in academic institutions and are focused on developing and supporting AI driven applications for scholarly research.<sup>3</sup> Recently, organizational consulting firm Korn Ferry has posited that librarians will increasingly be candidates for various corporate roles helping reframe thinking about AI and create value.<sup>4</sup> The Artificial Intelligence Skills Alliance (ARISA) “AI Skills Needs Analysis”

**CONTACT** Corey Harper  [c.harper@elsevier.com](mailto:c.harper@elsevier.com)  Elsevier Discovery Lab, New York, NY, USA.

© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

is a EU co-funded review of AI skills and related roles.<sup>5</sup> It lists a large variety of both technical and management and support roles including NLP engineer, prompt engineer, AI strategist, AI ethics officer, and AI quality controller. In this paper, we identify skills and capabilities of librarians to support AI based on a review of the literature and on a review of emerging AI practices. The result is a mapping of existing librarian skillsets to needs in developing AI systems. We use the American Library Association Core Competencies (ALACCs) as a reference set of library skills and produce a mapping of these competencies to skills needed for developing AI systems.<sup>6</sup> We supplement the ALACCs with additional skills from the Professional Competencies for Reference and User Services Librarians (RUSA), from the Core Competencies for Cataloging and Metadata Professional Librarians (CCCMPL), and from the curricula of several leading Library and Information Science Graduate Programs.<sup>7</sup> We also identify places to further emphasize or add to skillset development. We focus on how librarians, and particularly catalogers, can apply their competencies to AI, but not on whether the library community should be engaged in AI development. The latter discussion, while interesting and important, is out of scope for this paper.

The rest of this paper is organized as follows: we first review related work. We then proceed to describe our methodology and mapping of competencies. Based on this mapping of competencies, we explore in the analysis and discussion section AI development needs in five areas: information retrieval, model and data cataloging, knowledge engineering, responsible AI and social justice, and data engineering and evaluation, looking at each through the lens of librarian and cataloger skills.

## **Related work**

Little has been published on the role of the librarian in improving and developing AI systems. Most of the literature on AI and libraries has investigated the impact of AI on academic libraries, the value of AI for library users, or the implementation of ChatBots in academic libraries.<sup>8</sup> Additionally, many papers delve into the applications of AI in library operations, including indexing and abstracting, recommendations, and summarization.<sup>9</sup>

There has also been much written in the library literature about the ethical implications of AI. Mannheimer et al. write about responsible AI practice in libraries and archives and lay out some challenges that libraries will face in ensuring new AI based services are built with ethics in mind.<sup>10</sup> Responsible AI involves paying close attention to ethical implications of AI application. Examples include assessing the impact of an AI application on individuals, considering societal impacts, and reviewing applications

for any implicit or inherent biases. Inherent bias is common in AI systems due to the reliance on training data, which unintentionally encodes many human biases. This can lead to harmful or unfair AI outputs and should be actively mitigated by AI developers. Significantly, Mary Lee Kennedy reflects on the role of libraries in advocating for Responsible AI and Ethics more broadly, by actively engaging “in informing the principles and practices of AI institutionally, in public policy, and in the research and learning community.”<sup>11</sup> The ALACCs address social justice directly with competencies like “Recognize, challenge, and change practices, services, and programs that have traditionally replicated dominant systems and marginalized others.”<sup>12</sup> Librarians are well positioned to play a role in assessing training and evaluation data for inherent bias, recommending approaches to mitigating such bias, and identifying gaps in data and algorithms that could have negative impacts on under-represented and marginalized groups.

Several resources have been published in recent years that could potentially be used as sources for an inventory of AI Competencies. The AI Alliance, a community of AI technology developers and researchers formed by IBM and Meta in 2023 and comprising over 50 member organizations from academia, government, and industry, published a “Guide to Essential Competencies for AI.”<sup>13</sup> This guide provides specific skills and knowledge needed across a series of AI Competency areas as well as more general enabling soft skills. “The Transformational Opportunity of AI on ICT Jobs”, a report of the AI-Enabled Information and Communication Technology Workforce Consortium, documents new skills, skills that are increasing in importance, and deprecating skills.<sup>14</sup> This is analyzed from the perspective of existing Information & Communication Technology job groups, such as Data Scientist, Business and Management, Cybersecurity, and Software engineering. It does not cover the capabilities needed for new roles in AI. The US Office of Personnel Management issued a memorandum for Heads of Executive Departments and Agencies in April 2024 that lays out AI competencies needed at each grade level of US Government jobs.<sup>15</sup> They address general AI competencies and technical competencies. While their list is fairly exhaustive and contains similar items to those found in the AI Alliance Guide, they do not provide detailed skill and knowledge components. ARISA followed the discussion of AI skills and rolls in the “AI Skills Needs Analysis” with a detailed Core Curriculum mapping learning objectives in a wide array of skill areas to various technically focused EU job roles, including Machine Learning Engineer and Data Scientist.<sup>16</sup>

### **Methodology: Mapping competencies**

We develop a mapping between library skills and core competencies and the emerging needs of AI development. As a representation of AI competencies, we draw on the AI Alliance “Guide to Essential Competencies

for AI”.<sup>17</sup> We select this guide because it centers Responsible AI while also providing significant detail of the skills needed while extending beyond purely technical engineering and data science roles.

They identify five primary competency areas:

- Responsible Use discusses the ethical use of AI and mitigation of negative impacts,
- Identifying Data Limitations discusses understanding the limitations of AI systems and includes competencies around evaluating AI output,
- Data Analysis focuses on data selection and management,
- Machine Learning covers development of machine learning algorithms and selection of training data, and
- AI Logic discusses more traditional, logic-based symbolic AI systems.

Within each of these five areas, they identify knowledge and skills associated with four competency levels: Fluency, Proficiency, Expertise, and Mastery. They additionally identify “durable competencies” in areas including “personal agency”, problem solving, communication, and project management, as well as a set of foundational competencies in areas like Machine Learning, Algorithms, and Natural Language Processing. In addition to these AI competency areas, they also define six “Durable Competencies”, defined as “Cross-cutting competencies that may be needed to facilitate learning and application of AI competencies”, and eleven “Foundations”, defined as “Terms and concepts that are prerequisites to the acquisition of specific AI competencies.”

We take their five high level competency areas, durable competencies, and foundational competencies and map them to core competencies drawn from the American Library Association Core Competencies (ALACC), the Reference and User Services Association’s Professional Competencies for Reference and User Services Librarians (RUSA), and the Core Competencies for Cataloging and Metadata Professional Librarians (CCCMPL) from the ALA CORE (formerly published by the Association for Library Collections and Technical Services).<sup>18</sup> Where relevant, we map AI Alliance competencies to individual items drawn from the set of three ALA competency documents. These mappings are shown in [Table 1](#) in the results section. Our mapping covers three of the five AI Competency Areas, five of the six Durable Competencies, and three of the eleven foundations. These represent the set that have clear analogs in the set of Library Competency documents we reviewed. We exclude the AI Competency Areas around AI Logic (Symbolic AI) and Machine Learning, the Project Management durable competency, and Foundations such as Deep Learning, Supervised Learning, and Algorithms, because they do not map to the Library

**Table 1.** Mapping of AI Alliance’s five AI Competency categories, Foundational competencies, and Durable competencies to library competencies drawn from ALA’s Core Competences of Librarianship (ALACC), Core Competencies for Cataloging and Metadata Professional Librarians (CCCMPL), Professional Competencies for Reference and User Services Librarians (RUSA).

High-Level Category	Competencies & References
<b>Responsible Use</b>	<ul style="list-style-type: none"> <li>• ALACC 1A: Employ the ethics, values, and foundational principles of the library profession.</li> <li>• ALACC 1G: Understand the legal framework in which libraries operate, including laws relating to copyright and fair use, privacy, freedom of expression, equal rights (e.g., the Americans with Disabilities Act), open access, and intellectual property.</li> <li>• ALACC 7B: Recognize the ethical and appropriate application of key research methods, techniques, and designs in the field, including the generation, analysis, evaluation, and presentation of data, and the utilization of research tools.</li> </ul>
<b>Identifying Data Limitations</b>	<ul style="list-style-type: none"> <li>• ALACC 2A. Consider the issues related to the lifecycle of recorded knowledge and information, from creation through various stages of use, in relationship to material format and genre.</li> </ul>
<b>Data Analysis</b>	<ul style="list-style-type: none"> <li>• RUSA 5A: Accesses relevant and accurate recorded knowledge and information.</li> <li>• RUSA 5B: Evaluates, collects, retrieves, and synthesizes information from diverse sources.</li> <li>• CCCMLP: Understands principles behind controlled vocabularies, particularly international, domain, and community-centric structures.</li> <li>• CCCMLP: Understands principles of classification and information organization.</li> <li>• CCCMLP: Understands different data structures (flat, hierarchical, graph, relational, semantic, etc.).</li> <li>• CCCMLP: Indexing and database structure.</li> </ul>
<b>Durable Competencies</b>	
<b>Flexibility</b>	<ul style="list-style-type: none"> <li>• CCCMLP: Initiative &amp; adaptability – Flexible Thinking.</li> <li>• CCCMLP: Initiative &amp; adaptability – Ability to handle ambiguity and changing priorities.</li> </ul>
<b>Growth Mindset</b>	<ul style="list-style-type: none"> <li>• ALACC 4E. Employ the concepts behind, issues relating to, and methods of principled, transformational, and change management leadership, in addition to other leadership philosophies.</li> <li>• CCCMLP: Initiative &amp; adaptability – Commitment to continuous learning.</li> <li>• RUSA 5G: Leads a Life as a Learner</li> </ul>
<b>Personal Agency Creative Problem Solving</b>	<ul style="list-style-type: none"> <li>• CCCMLP: Initiative &amp; adaptability – Independence in decision making.</li> <li>• ALACC 1I. Use techniques to identify, codify, and analyze complex problems and create appropriate and collaborative solutions within library environments.</li> <li>• CCCMLP: Problem solving: Implements project management techniques and tools; collaborates with colleagues to solve problems; evaluates problem solving results and applies knowledge to future tasks</li> </ul>
<b>Communication and Collaboration Foundations (General)</b>	<ul style="list-style-type: none"> <li>• ALACC 1J: Demonstrate effective verbal and written communication techniques, including video, live chat, and email.</li> <li>• ALACC 9A. Identify appropriate technologies and uses that support access to and delivery of library services and resources.</li> <li>• ALACC 9B. Understand and navigate ethical and cultural considerations and impacts on library practices and community members when applying technology to library services and resources.</li> <li>• ALACC 9C. Conduct regular evaluation of existing and emerging technologies and their impact on library services and resources in terms of accessibility, practicality, sustainability, and effectiveness.</li> </ul>
<b>Artificial Intelligence Training Data</b>	<ul style="list-style-type: none"> <li>• CCCMLP: Understands the impact, limitations, and possible usage of artificial intelligence (AI) applications and large language models (LLM)</li> <li>• ALACC 4C. Implement the concepts behind, and methods for, assessment and evaluation of library services and their outcomes.</li> <li>• ALACC 5A: Understand the principles, systems, trends, and goals involved in the organization and representation of recorded knowledge and information.</li> <li>• ALACC 5C. Maintain the systems of cataloging, collection, metadata, indexing, and classification standards and structures, and implement methods used to apply, create, and discover recorded knowledge and information, and the weaknesses and strengths of these systems.</li> <li>• ALACC 6A: Employ techniques used to discover, retrieve, evaluate, and synthesize information from diverse sources for use by varying user populations and information environments.</li> <li>• ALACC 9C. Conduct regular evaluation of existing and emerging technologies and their impact on library services and resources in terms of accessibility, practicality, sustainability, and effectiveness.</li> </ul>

(Continued)

**Table 1.** Continued.

High-Level Category	Competencies & References
<b>Bias</b>	<ul style="list-style-type: none"> <li>• ALACC 5D: Recognize the ways that cultural biases impact and influence the collection and description of recorded knowledge and information.</li> <li>• ALACC 8A: Understand one's own cultural identity including positionality related to power, privilege, and oppression and how that influences the ways they interact with the community and among decision makers.</li> <li>• ALACC 8B: Recognize, challenge, and change practices, services, and programs that have traditionally replicated dominant systems and marginalized others.</li> <li>• ALACC 9B: Understand and navigate ethical and cultural considerations and impacts on library practices and community members when applying technology to library services and resources.</li> <li>• CCCMLP: Ethical and transparent data.</li> <li>• RUSA 5F: Assesses and responds to diversity in user needs, user communities, and user preferences.</li> </ul>

Competencies analyzed. Although the Durable Competencies are not specific to libraries, we include them in our mapping because analogous entries are included in the set of ALA competencies.

Additionally, we review the curricula of the top three ALA accredited Library and Information Science programs in the United States per 2021 US News and World Reports rankings:<sup>19</sup>

- University of Illinois Urbana-Champaign (UIUC)<sup>20</sup>
- University of North Carolina–Chapel Hill (UNC)<sup>21</sup>
- University of Washington<sup>22</sup>

For each program, we review the full course listings, including descriptions, but not syllabi. For the purpose of the mapping, we focus on UIUC and UNC because the selection of elective courses offered as part of the library science curriculum is closely aligned with the emerging needs of AI. The University of Washington also has many relevant classes, including an entire AI specialization, but it is part of their Master of Science in Information Management Curriculum rather than the Master of Library and Information Science. This is contrasted with UIUC and UNC, which are both iSchools, and while they include a number of courses related to traditional library skills, they also cover data science and information science courses *as part of the Library and Information Science curriculum*. Our mappings cover both traditional and iSchool curricula, but do not fully represent all subject areas: there are many additional courses listed that have less relevance to the AI domain. We review the required courses and all elective courses – including review of course descriptions and occasionally syllabi when they are available – and we map specific “Knowledge” components from each of the AI Alliance’s five AI Competency areas at each of the four competency levels to course offerings from these two programs.<sup>23</sup> Table 2 in the Results Section contains the mappings from courses to AI Competency Knowledge. Eight out of 78 courses listed for

**Table 2.** Mapping of library courses from University of North Carolina (UNC) and University of Illinois Urbana Champaign (UIUC) to AI Alliance Competency knowledge.

Course Title	AI Alliance Competency Knowledge
<b>UIUC IS 407 Introduction to Data Science</b>	<ul style="list-style-type: none"> <li>• DA2: Understanding of basic statistical and mathematical concepts, including means, correlations, and others, that are relevant to data analysis.</li> <li>• DA2: Understanding of data characteristics and differences between unstructured and structured data.</li> <li>• DA3: Understanding of the strengths and weaknesses of machine learning libraries appropriate for a given task, mitigating data limitations.</li> <li>• DA3: Understanding of approaches and techniques to evaluate outcomes and assess models, including computational performance and accuracy.</li> <li>• DA3: Knowledge of data augmentation techniques and handling imbalanced data sets.</li> <li>• ML1: Understanding of the types of machine learning.</li> <li>• ML3: Understanding of classification and clustering, methods by which data is organized and labeled.</li> <li>• ML3: Understanding of differences between supervised, unsupervised, and reinforcement machine learning.</li> </ul>
<b>UIUC IS 537 Theory &amp; Practice of Data Cleaning</b>	<ul style="list-style-type: none"> <li>• IDL3: Understanding of how to address and mitigate data limitations in model design, data selection, and evaluation.</li> <li>• DA3: Knowledge of strategies for data collection, cleaning, and preprocessing to ensure high-quality inputs for training models.</li> </ul>
<b>UIUC IS 455 Database Design and Prototyping</b>	<ul style="list-style-type: none"> <li>• AIL3: Awareness of methods for structuring and organizing information in ways that can be used by machines.</li> </ul>
<b>UIUC IS 584 Privacy in the Internet Age</b>	<ul style="list-style-type: none"> <li>• RU1: "Understanding of the role of personal data in AI systems, and awareness of basic methods of personal data privacy protection."</li> <li>• RU1: "Awareness of methods of data and privacy protection required in one's work environment."</li> <li>• RU3: "Awareness of the legal and regulatory requirements of data protection, consequences of breaches, and actions required in response to data breaches."</li> <li>• RU4: "Awareness of the management strategies of teams, projects, and organizations to maximize data privacy and protection."</li> <li>• ML3: Understanding of Retrieval-Augmented Generation.</li> </ul>
<b>UIUC IS 530 Collection Development</b>	<ul style="list-style-type: none"> <li>• ML3: Understanding of Retrieval-Augmented Generation.</li> </ul>
<b>UNC INLS 509: Information Retrieval</b>	<ul style="list-style-type: none"> <li>• IDL1: Awareness of how human biases impact AI training data, leading to potentially inaccurate or incomplete results from those models.</li> <li>• IDL1: Understand the implications of misinterpreted or inaccurate results from AI systems on one's role and organization.</li> <li>• IDL2: Understanding of how to check or evaluate AI responses for accuracy, including where one can go to find alternative sources of information.</li> <li>• IDL4: Understanding of how data limitations in AI tools and products may create negative implications for individuals, enterprises, and societies.</li> </ul>
<b>UNC INLS 690: Data Criticism</b>	<ul style="list-style-type: none"> <li>• IDL1: "Awareness of the potential for misleading results and "hallucinations" in AI systems."</li> <li>• IDL1: "Understand the implications of misinterpreted or inaccurate results from AI systems on one's role and organization."</li> <li>• IDL2: Understanding of how to check or evaluate AI responses for accuracy, including where one can go to find alternative sources of information.</li> <li>• IDL4: Understanding of opportunities and venues to advocate for tools, policies, and campaigns that enable end-users to recognize AI data limitations.</li> </ul>
<b>UNC INLS 690: Misinformation and Society</b>	<ul style="list-style-type: none"> <li>• ML3: Understanding of natural language processing, which uses machine learning to enable computers to understand and communicate with human language.</li> </ul>
<b>UNC INLS 512: Applications of Natural Language Processing</b>	<ul style="list-style-type: none"> <li>• DA1: Understanding of basic data visualization tools and approaches, including simple charts and graphs.</li> <li>• DA3: Understanding of data visualization and interpretation approaches, enabling effective insights gathering and model transparency.</li> <li>• ML1: Understanding of basic coding concepts.</li> </ul>
<b>UNC INLS 541: Information Visualization</b>	<ul style="list-style-type: none"> <li>• DA1: Understanding of basic data visualization tools and approaches, including simple charts and graphs.</li> <li>• DA3: Understanding of data visualization and interpretation approaches, enabling effective insights gathering and model transparency.</li> <li>• ML1: Understanding of basic coding concepts.</li> </ul>

*(Continued)*

**Table 2.** Continued.

Course Title	AI Alliance Competency Knowledge
<b>UNC INLS 560: Programming for Information Professionals</b>	<ul style="list-style-type: none"> <li>• ML1: Understanding of basics of low-code and no-code development solutions, approaches that reduce the need for manual programming.</li> <li>• ML2: Basic understanding of programming languages for AI Development.</li> </ul>
<b>UNC INLS 490: The Idea of AI</b>	<ul style="list-style-type: none"> <li>• RU1: "Awareness that applications, products, and tools use AI, and the ways and reasons AI may be used within those products."</li> <li>• RU2: "Understanding how, where, and for what purpose AI tools and products can be deployed to complement and/or benefit one's purposes, and to meet one's needs."</li> <li>• RU4: "Understanding the potential future opportunities and limitations of AI methods and tools."</li> </ul>
<b>UNC INLS 690: Philosophy and Ethics of AI</b>	<ul style="list-style-type: none"> <li>• RU2: "Awareness of the implications and consequences of AI use, including legal, reputational, personal, and professional risks."</li> <li>• RU3: "Awareness of how the use of AI may complement or disagree with standards, norms, and requirements for ethics applicable to one's country, field, and/or organization."</li> <li>• RU4: "Awareness of how project and program structures, including personnel, incentives, and internal corporate cultures, may contribute to or mitigate unethical use of AI tools and products."</li> </ul>

Competency categories: DA: Data analysis, ML: Machine Learning, RU: Responsible use, IDL=Identifying data Limitations, AL=AI logic. Numbers indicate AI Alliance fluency levels.

UNC appear in our table, along with five out of the 122 courses listed for UIUC.

## Results

In mapping the ALA and RUSA competencies to the high-level AI Alliance categories (Table 1), the most frequently mapped categories are Responsible Use and Identifying Data Limitations, whereas the CCCMLP competencies more often map to Data Analysis and to the Durable Competencies. The curricula (Table 2) of both UIUC and UNC also are very strong in areas related to Responsible Use of AI, Identifying Data Limitations, Data Analysis, and Machine Learning. Generally, there is less focus on AI Logic, which is more likely to be situated with Computer Science departments and other programs that would focus more on the previous generation of symbolic AI.

The competency mapping in Table 1 is meant to show areas of the set derived from the ALA documents, and do not include competencies that are not represented in both sets of sources. Additionally, Table 2 covers other areas of overlap between Librarian and AI skills that are not represented in the ALA Competencies, but that can be seen in iSchool Curricula. The Analysis and Discussion that follows is organized along the dimensions of *information retrieval, model and data cataloging, knowledge engineering, responsible AI and social justice, and data engineering and evaluation*. While these are not directly derived from our mapping tables, we focus our discussion on these areas because they help to frame our discussion while highlighting areas such as Responsible AI and

Information Retrieval that we see as critically important of AI and deeply embedded in the library identity.

## **Analysis and discussion**

Using the above mappings, we now discuss how librarian competencies intersect with key areas of AI systems development and deployment.

### ***Information retrieval***

AI application development frequently relies heavily on search.<sup>24</sup> Initial work on Retrieval Augmented Generation (RAG) techniques focused on embedding models and vector databases for text similarity-based retrieval.<sup>25</sup> The idea behind RAG is that, rather than relying exclusively on the memorized information contained in the parameters of AI models, you can retrieve trustworthy, vetted chunks of “context” based on a user query, and include that context when prompting AI systems. Increasingly, application developers are finding vector-based search to provide sub-optimal retrieval and are looking to hybrid techniques that combine traditional lexical and taxonomy-based search with vector search. Vector search involves calculating embeddings, numerical representations of words in a vector space, for both content and user queries, and then using similarity metrics such as cosine similarity and dot product to find pieces of content that closely match a user’s query. Vector databases are often a significant part of the information retrieval layer of RAG systems.

Given long-standing traditions in metadata design, search, and information retrieval, librarians are again well-positioned to assist in the development of robust search frameworks to help feed data to AI systems, grounding them and mitigating hallucination. Metadata fields in search systems can be used to filter data, scoping searches to specific subsets of corpora, much like faceted navigation has been used in retrieval systems pioneered by the library community. Creation of relevant metadata to support these kinds of filter and navigation operations is a role that is closely aligned with the librarian skill set. Additionally, the ALACCs and RUSA competencies focus heavily on collection development, a skill which can be applied to designing and structuring the corpora used in RAG systems. These skills are especially valuable in building domain specific and even task specific RAG systems, where the development of the corpus must provide appropriate coverage of the specific use cases and questions that users are likely to ask.

The collection development, taxonomy development, and classification work that goes into building good search systems and selecting the corpora to be indexed relies on heavy domain and subject matter

expertise. Subject specialization has long been a hallmark of the library profession, with subject specialists providing reference services and building collections for specific academic departments. These professionals often have additional advanced degrees in their subject matter areas, and many of the top Library Science programs, including both UIUC and UNC, offer dual degree programs to help develop library skills and domain expertise in parallel.<sup>26</sup>

### ***Model and data cataloging***

There are other gaps in the AI training and evaluation data ecosystem that librarian expertise can help to fill. Emerging work on Dataset Nutrition Labels, Model Cards, the CLeAR Documentation Framework, and Croissant all lay out metadata and documentation requirements for ensuring that datasets, machine learning models, and AI systems are fair, accountable, transparent, and ethical.<sup>27</sup> Model Cards are designed to accompany machine learning models and describe information about a model's training data, possible biases, risks, and intended use of the model. CLeAR extends this beyond models, providing a framework for "Comparable, Legible, Actionable, and Robust (CLeAR) Documentation" of larger AI systems, while Croissant provides a data format for Machine Learning datasets that incorporates metadata and documentation directly into the dataset. Interestingly, the Croissant model builds on W3C standards, such as the Data Cataloging Vocabulary (DCAT), that are commonly part of librarian and information science curricula.<sup>28</sup> And in general metadata skills are an important skillset for librarians, e.g., "ALACC 5C. Maintain the systems of cataloging, collection, metadata, indexing, and classification standards and structures, ...."<sup>29</sup>

Additionally, finding relevant domain specific datasets both for tuning and evaluating AI systems is increasingly difficult and time consuming.<sup>30</sup> Data is scattered across many different repositories, referenced in conference publications, and tracked on leaderboards. There are no comprehensive repositories of robust and meaningful evaluation data, making the identification of appropriate datasets challenging. Librarians in both Corporate and Academic institutions would help identify, catalog, and provide access to relevant domain specific data, which relate to the ALACCs in section 5, concerning "Organization of Recorded Knowledge and Information."<sup>31</sup> As AI systems begin to call APIs and use other tools while helping users undertake complex tasks, API endpoints, tools, and other modules that will be employed by these systems will also benefit from the description and access that the library community is so adept at providing.<sup>32</sup>

### **Knowledge engineering**

An additional, and closely related, area of promising research is in Knowledge Engineering. This includes taxonomy and ontology development, which can assist in the cataloging and metadata creation of the wealth of resources discussed in the previous section. Taxonomies are also useful in helping to understand emerging technologies, explain the way these technologies are related to each other, how they are similar, and how they are different, and reach consensus around the vocabularies used when talking about technologies. Many survey papers covering large categories of Machine Learning or AI topics include taxonomies and ontologies of the set of techniques covered, often with hierarchies grouping techniques together based on shared characteristics. For example, Lialin et al. provide a taxonomy of Parameter-Efficient Fine-Tuning methods; Mao et al. offer a taxonomy of Low-Rank Approximation (LoRA) methods for LLMs; and Xu et al. present a detailed taxonomy of Knowledge Distillation methods for LLMs.<sup>33</sup> All three of these examples build their taxonomy around the plethora of papers they cite for their review, use the taxonomy to categorize the techniques described in these papers, and structure the rest of their survey around their taxonomy.

Another aspect of Knowledge Engineering is research into Knowledge Graphs (KGs). Knowledge graphs are formal representations of entities and relationships frequently used in knowledge management and application development.<sup>34</sup> KGs are closely related to, and build on, earlier W3C Semantic Web and Linked Open Data work that librarians have historically engaged in.<sup>35</sup> Recent research has evaluated both using AI to create and augment KGs and using KGs to further ground AI applications.<sup>36</sup> Detailed entities and relationships are hard to discover and interpret when working with unstructured text, which can be seen as a shortcoming when dealing with complex knowledge. Knowledge graphs can help provide relational structure to improve AI systems. The librarian skillset includes competencies related to building and populating knowledge bases and knowledge graphs and developing the taxonomies and ontologies that support knowledge engineering. The librarian's discipline specific domain knowledge and subject matter expertise, taken together with collection management, cataloging, and classification skills, can help develop KGs for AI systems in specific domains and for specific professions. Library competencies such as the CCCMLP's focus on "principles of identity management and authority control" and "principles behind controlled vocabularies" and courses such as UIUC's *Database Design and Prototyping* and *Bibliographic Metadata* and UNC's *Organization of Information* focus on these skills.<sup>37</sup>

## **Responsible AI and social justice**

The ALA competencies highlight the importance of ethics and social justice, which aligns well with a growing industry focus on the responsible use of AI, Machine Learning, and Data Science. At the time of this writing, a search for “Responsible AI” on LinkedIn Jobs yields 111 results.<sup>38</sup>

Responsible AI is the first competency category identified in the AI Alliance report. Similarly, Stanford’s 2024 AI Index Report has a chapter devoted to Responsible AI.<sup>39</sup> This chapter focuses on evaluation, privacy and data governance, transparency and explainability, security, and fairness. The fairness aspect of this directly aligns with Librarian values around social justice as expressed in the ALA competencies, which in turn aligns closely with the AI Alliance Foundational Competency concerning Bias. Understanding the impact of AI applications on individuals and groups of people, especially disenfranchised and underserved populations, is an essential part of the responsible deployment of AI applications. Asking questions like:

- Who might be harmed by the rollout of a new application or the productization of a project?
- Could an application I’m developing be subject to accidental or intentional misuse?
- What potential biases exist in the training data being used by a new machine learning model or AI application?
- How interpretable, transparent, and understandable are the outcomes of my data science project or AI application?

These types of questions are an increasingly important part of responsible AI roles, and the importance librarian training places on ethics and social justice position librarians to undertake these kinds of assessments. Courses like *The Idea of AI* and *Philosophy and Ethics of AI* at UNC indicate the intention to foster the skills and perspectives necessary to guide projects through these challenging considerations.

Responsible use of AI also includes legal implications, questions of fair use, licensing of data and of machine learning models, privacy laws, and more. These are all things that land squarely in the ALA Core Competencies Gateway Knowledge category. Example questions that a trained librarian might be well positioned to help answer include:

- When training a machine learning model – what data should one use from the open web?
- Is the LLaMA foundation model from Facebook really open source? What are the implications of the various different licenses used by the different versions of the model?<sup>40</sup>

- What are the privacy implications of using OpenAI APIs? How do these compare to the privacy implications of using the DeepSeek APIs?<sup>41</sup> Would it be safer to use a locally hosted version of DeepSeek, or one from a provider like Amazon Web Services (AWS) or Microsoft Azure?
- Is this particular AI research allowable under the EU's AI act?

While specific providers (e.g., Azure and AWS) or models (e.g., DeepSeek) are not explicitly mentioned in these curricula, the curricula provide a framework to train and prepare professionals to answer these kinds of questions. These questions are applicable whether in corporate, academic, or government environments. Relevant training can be found throughout the academic Library Science Curriculum in courses like *Privacy in the Internet Age*, *Ethics & Policy for Data Science*, and *Copyright for Information Professions* at UIUC or *Data Governance* at UNC.

### **Data engineering and evaluation**

Another broad category of AI development work is around data management, data engineering, and data preparation. These areas are covered in part by the AI Alliance competencies under *Identifying Data Limitations* as well as the Foundational Competency in *Training Data* but potentially have much broader applications than are mentioned in the AI Alliance documentation.<sup>42</sup> Closely connected to the Responsible AI topic is a set of questions concerning *Data Limitations*, including suitability of data, challenges with data sources, biases in training data, and evaluation of AI outputs. Again, these map well to library functions and library skills. ALA's Core Competencies 2A and 2B look at "issues related to the lifecycle of recorded knowledge and information" and the "concepts, issues, and methods of collection management", while multiple RUSA competencies focus on providing access to information and on evaluation.<sup>43</sup> Understanding *Data Limitations* and working with *Training Data* both depend on knowledge of storing and maintaining recorded knowledge. The management of such data, which requires understanding of data structures and formats, is closely connected to traditional library skills. Courses such as *Theory & Practice of Data Cleaning* at UIUC and *Data Criticism and Misinformation and Society* at UNC will prepare students to understand how to assess data quality and evaluate the output of AI systems for inaccuracies, hallucinations, and misinformation. Being able to trace errors back to training data or application prompts will help AI developers curate better training data sets and build more trustworthy systems.

This is closely related to the data preparation, analysis, and model evaluation that is undertaken as part of data science. Over the course of

the last decade, more and more Library Science programs and Information Schools have increased their Data Science offerings. While we are primarily focused on the more traditional librarian skillset and Library Science curricula in this paper, the mappings presented do include Data Science electives that are available to Library Science students. Courses like *Introduction to Data Science* at UIUC and *Applications of Natural Language Processing, Information Visualization, and Programming for Information Professionals* at UNC align closely to the Data Analysis and Machine Learning competencies from the AI Alliance. While these skillsets do not appear in the various ALA competency documents, many librarians are developing these skills over the course of their academic training. Khan discusses the role of data science learning objectives in LIS education.<sup>44</sup> While the exploration is focused on research data librarians, Khan also notes that many traditional library roles will increasingly need data science skills. Zhang et al. and Urs & Minhaj provide much more detail on the development of Data Science curricula in the context Information Science education.<sup>45</sup>

### ***New roles for librarians and implications for library education***

AI is an incredibly fast-moving area, and new needs are emerging constantly. Some of these new needs would benefit greatly from the librarian skillset. Two examples of this are AI Auditing and Agentic AI. *Harvard Business Review* defines an AI, or algorithmic audit as “a way of answering a question about an algorithmic system by analyzing data on how users and systems interact” ... “to understand whether a system is working as expected or is harming users.”<sup>46</sup> Agentic AI refers to trends in AI development that give more autonomy, problem solving, and decision-making responsibility to AI algorithms and specifically to large language models.

The ALA competencies focus on Social Justice and the courses seen in both the degree programs we analyzed will help prepare students for roles in Responsible AI. A related practice, auditing of AI systems, is becoming an increasingly significant part of Responsible AI.<sup>47</sup> Audits can be performed by either internal or external auditors.<sup>48</sup> Librarian competencies in areas like documentation, organization of information, and data governance can support AI auditing. The skills developed in courses like UNC’s *Data Governance* align well with the skills taught in AI auditing courses.<sup>49</sup>

In 2024 and 2025, more attention has been paid to Agentic AI, where more and more autonomy is given to Large Language Models and AI to engage in complex problem solving with less guidance and human intervention.<sup>50</sup> While this is in many ways exciting, it is also met with some trepidation, as many express ethical and safety concerns about increased

autonomy and even suggest that fully autonomous systems should not be developed.<sup>51</sup> The risks have become even more pronounced as increasingly independent “reasoning” models such as OpenAI’s o1 have demonstrated poor alignment with both their programmers and users, some even showing capacity to deceive their users.<sup>52</sup> One approach to mitigating these risks is through “Human in the Loop” Agentic AI systems, where there are opportunities for end-users as well as other participants in a system’s execution to advise and direct models, perhaps changing query plans or updating workflows.<sup>53</sup> As multi-agent systems become more complex, including multiple humans, perhaps with different subject matter expertise, is a potentially valuable solution. As noted previously, many Library and Information Science programs offer dual degrees so that potential librarians can develop subject matter expertise. This subject matter expertise, along with competencies in areas such as information verification, misinformation, resource assessment, and providing reference services, perfectly position librarians to play an active role in these human-in-the-loop multi-agent systems.

### **Limitations and future work**

This current exploration is a qualitative analysis of the correspondence between the AI Competencies developed by the AI Alliance and the librarian skillset, as represented by ALA Competency documents and LIS Curricula. The methodology of this study is a manual mapping. One possible direction of future work is to apply this mapping methodology to competencies in other information center domains such as Information Management and Human-Computer Interaction. The Korn Ferry report cited in the introduction notes that insurance actuaries have similar skill sets to librarians around organizing and analyzing information.<sup>54</sup> This is an indication that a similar analysis could be conducted for domains well outside of the scope of traditional information systems and computer science fields.

An additional area of research could expand on the *Implications for Library Education*, conducting a more detailed gap analysis between the AI competencies identified and iSchool Curricula. This could help inform recommendations around how to evolve iSchool Curricula to support new AI roles such as Prompt Engineering, Responsible AI, and AI Auditing. Additional examples of such roles could be pulled from job descriptions as well as the discussion of AI roles listed in the EU ARISE materials.

Finally, additional work could be done to augment this qualitative analysis with a more quantitative study. An automated analysis of a larger set of curricula, including more syllabi and job descriptions harvested from LinkedIn or Indeed could provide statistical insights into the

correspondence between AI Roles and iSchool Curricula. This could also have the added benefit of helping to prioritize recommendations about iSchool curriculum updates based on statistical analysis of the frequency with which skills are mentioned in job descriptions.

## Conclusion

In the coming years, we expect that indexing and abstracting practices are going to evolve to better meet information retrieval needs of AI systems. New metadata standards and formats, new indexing techniques, and new taxonomies and ontologies are going to be critical in the development of robust Artificial Intelligence systems. Librarians have an opportunity to play a critical role in defining these new standards, techniques, and vocabularies. The LIS Education System can help to train professionals for new roles in AI in a pivot that will resemble the shift toward Information Science, which trained professionals for roles in Information Architecture and Information Retrieval outside of library organizations. As we have shown throughout this analysis, librarians already have a significant set of competencies that are essential in the successful, responsible, and ethical development of AI systems.

## Notes

1. "Machine Learning Librarian," Startup Jobs, last modified May 9, 2022, <https://startup.jobs/machine-learning-librarian-hugging-face-3262782> (accessed July 14, 2025).
2. Andrew Cox, "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* 74, no. 3 (2023): 367–80, <https://doi.org/10.1002/asi.24635> (accessed July 14, 2025).
3. "Artificial Intelligence Librarian," Indeed, <https://www.indeed.com/viewjob?jk=d7e207c145aeb466&from=shareddesktop> (accessed July 14, 2025); "AI Librarian," code4lib jobs, last modified October 2, 2024, <https://jobs.code4lib.org/jobs/62538-ai-librarian>, (accessed July 14, 2025).
4. "Wanted for AI: Librarians," This Week In Leadership, Korn Ferry, archived December 20, 2024 at the Wayback Machine, <http://web.archive.org/web/20241220143346/https://www.kornferry.com/insights/this-week-in-leadership/artificial-intelligence-talent-boardroom-briefings> (accessed July 14, 2025).
5. "AI Skills Needs Analysis: AN INSIGHT INTO THE AI ROLES AND SKILLS NEEDED FOR EUROPE, Draft version," ARISA, May 31, 2023, [https://aiskills.eu/wp-content/uploads/2023/06/ARISA\\_AI-Skills-Needs-Analysis\\_DRAFT.pdf](https://aiskills.eu/wp-content/uploads/2023/06/ARISA_AI-Skills-Needs-Analysis_DRAFT.pdf) (accessed July 14, 2025).
6. "ALA's Core Competences of Librarianship," American Library Association, January 28, 2023. <https://www.ala.org/educationcareers/careers/corecomp/corecompetences>, (accessed December 20, 2024).
7. "Professional Competencies for Reference and User Services Librarians," American Library Association, September 29, 2008, <https://www.ala.org/rusa/resources/guidelines/professional> (accessed March 30, 2025); "Core Competencies for Cataloging and

- Metadata Professional Librarians,” American Library Association, December 13, 2023, <https://alair.ala.org/items/64a98b5b-c5f5-4bb8-a6cd-b06833b0c89c> (accessed March 30, 2025).
8. Khurram Shahzad, Shakeel Ahmad Khan, and Abid Iqbal, “Effects of artificial intelligence on university libraries: an SLR of cite score and IF journals’ articles from 2018 to 2023,” *Global Knowledge, Memory and Communication* (2024), <https://doi.org/10.1108/GKMC-12-2023-0498>, (accessed July 17, 2025); Andrew M. Cox, Stephen Pinfield, and Sophie Rutter, “The intelligent library: Thought leaders’ views on the likely impact of artificial intelligence on academic libraries,” *Library Hi Tech* 37, no. 3 (2019): 418–35. <https://doi.org/10.1108/LHT-08-2018-0105> (accessed July 14, 2025); James Oluwaseyi Hodonu-Wusu, “The rise of artificial intelligence in libraries: the ethical and equitable methodologies, and prospects for empowering library users,” *AI Ethics* 5 (2024): 755–65, <https://doi.org/10.1007/s43681-024-00432-7> (accessed July 14, 2025); Rumeng Yan, Xin Zhao, and Suvodeep Mazumdar, “Chatbots in Libraries: A Systematic Literature Review,” *Education for Information* 39, no. 4 (2023): 431–49. <https://content.iospress.com/articles/education-for-information/efi230045> (accessed July 14, 2025).
  9. Adetoun A. Oyelude, “AI and libraries: trends and projections,” *Library Hi Tech News* 38, no. 10 (2021): 1–4. <https://doi.org/10.1108/LHTN-10-2021-0079> (accessed July 14, 2025).
  10. Sara Mannheimer, Natalie Bond, Scott W. H. Young, Hannah Scates Kettler, Addison Marcus, Sally K. Slipher, Jason A. Clark, Yasmeen Shorish, Doralyn Rossmann, and Bonnie Sheehey, “Responsible AI Practice in Libraries and Archives: A Review of the Literature,” *Information Technology and Libraries* 43, no. 3 (2024) <https://doi.org/10.5860/ital.v43i3.17245> (accessed July 14, 2025).
  11. Mary Lee Kennedy, “What Do Artificial Intelligence (AI) and Ethics of AI Mean in the Context of Research Libraries?” *Research Library Issues* 299 (2019): 3–13. <https://publications.arl.org/18nm1db/> (accessed July 14, 2025).
  12. American Library Association, “ALA’s Core Competences of Librarianship,” 9.
  13. “Guide to Essential Competencies for AI,” AI Alliance, December, 2024. <https://thealliance.ai/docs/guide-to-essential-competencies-for-ai.pdf>, (accessed March 30, 2025).
  14. “The Transformational Opportunity of AI on ICT Jobs”. AI-Enabled Information and Communication Technology Workforce Consortium (2024) <https://www.cisco.com/c/dam/m/ai-enabled-ict-workforce-consortium/report.pdf>, (accessed July 15, 2025).
  15. United States Office of Personnel Management, “Skills-Based Hiring Guidance and Competency Model for Artificial Intelligence Work,” memorandum from Kiran A. Ahuja to Heads of Executive Departments and Agencies (Washington, DC, April 29, 2024), <https://www.opm.gov/policy-data-oversight/hiring-information/skills-based-hiring/ai-competency-model.pdf>, (accessed July 15, 2025).
  16. ARISA, “AI Skills Needs Analysis”; “D4.2 Core Curricula,” ARISA, 2024, [https://aiskills.eu/wp-content/uploads/2024/10/D4.2\\_Core-Curricula.pdf](https://aiskills.eu/wp-content/uploads/2024/10/D4.2_Core-Curricula.pdf), (accessed July 15, 2025).
  17. AI Alliance, “Guide to Essential Competencies for AI.”
  18. American Library Association, “ALA’s Core Competences of Librarianship;” American Library Association, “Professional Competencies for Reference and User Services Librarians;” American Library Association, “Core Competencies for Cataloging and Metadata Professional Librarians.”
  19. “Best Library and Information Studies Programs,” *US News and World Report*, <https://www.usnews.com/best-graduate-schools/top-library-information-science-programs/library-information-science-rankings>, (accessed March 30, 2025).
  20. “Plan Your Program,” School of Information Sciences, University of Illinois Urbana-Champaign, <https://ischool.illinois.edu/academics/graduate/ms-library-and-information-science/plan-your-program>, (accessed March 30, 2025).

21. “The SILS Curriculum: A Guide for MSLS and MSIS Students,” School of Information and Library Science, University of North Carolina, [https://sils.unc.edu/wp-content/uploads/sites/1369/2024/02/SILS-Curriculum-Student-Guide-Revised-Spring2024\\_Casey-Rawson.pdf](https://sils.unc.edu/wp-content/uploads/sites/1369/2024/02/SILS-Curriculum-Student-Guide-Revised-Spring2024_Casey-Rawson.pdf), (accessed March 30, 2025).
22. “Master of Library and Information Science: MLIS Academics,” Information School, University of Washington, <https://ischool.uw.edu/programs/mlis>, (accessed March 30, 2025).
23. AI Alliance, “Guide to Essential Competencies for AI.”
24. Wenqi Fan, Yujuan Ding, Liangbo Ning, Shijie Wang, Hengyun Li, Dawei Yin, Tat-Seng Chua, and Qing Li, “A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models,” in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24)*, (2024), 6491–501 <https://doi.org/10.1145/3637528.3671470> (accessed March 30, 2025).
25. Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, and Haofen Wang, “Retrieval-Augmented Generation for Large Language Models: A Survey,” arXiv, 2024, <https://arxiv.org/abs/2312.10997>, (accessed July 15, 2025).
26. “MSLIS Joint Degrees,” School of Information Sciences, University of Illinois Urbana-Champaign, <https://ischool.illinois.edu/academics/graduate/ms-library-and-information-science/mslis-joint-degrees>, (accessed July 16, 2025); “Dual Master’s Degrees,” School of Information and Library Science, University of North Carolina, <https://sils.unc.edu/dual-masters-degrees/>, (accessed July 16, 2025).
27. Kasia S. Chmielinski, Sarah Newman, Matt Taylor, Josh Joseph, Kemi Thomas, Jessica Yurkofsky, and Yue Chelsea Qiu, “The Dataset Nutrition Label (2nd Gen): Leveraging Context to Mitigate Harms in Artificial Intelligence,” arXiv, 2022, <https://arxiv.org/abs/2201.03954>, (accessed July 16, 2025); Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru, “Model Cards for Model Reporting,” in *FAT\* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency* (2019), 220–9, <https://doi.org/10.1145/3287560.3287596>, (accessed July 16, 2025); Mubashara Akhtar, Omar Benjelloun, Costanza Conforti, Luca Foschini, Pieter Gijssbers, Joan Giner-Miguel, Sujata Goswami, Nitisha Jain, Michalis Karamousadakis, Satyapriya Krishna et al., “Croissant: A metadata format for ml-ready datasets,” *Advances in Neural Information Processing Systems* 37 (2024): 82133–48; Shorenstein Center on Media, Politics and Public Policy, The CLEAR Documentation Framework for AI Transparency: Recommendations for Practitioners and Policymakers (United States: Shorenstein Center on Media, Politics and Public Policy, 2024), <https://coilink.org/20.500.12592/2ngf7xs>, (accessed July 16, 2025).
28. “Data Catalog Vocabulary (DCAT) - Version 3,” World Wide Web Consortium, last modified August 22, 2024, <https://www.w3.org/TR/vocab-dcat-3/>, (accessed July 16, 2025).
29. American Library Association, “ALA’s Core Competences of Librarianship,” 7.
30. Steven Euijong Whang, Yuji Roh, Hwanjun Song and Jae-Gil Lee, “Data collection and quality challenges in deep learning: A data-centric ai perspective,” *The VLDB Journal* 32, no. 4, (2023): 791–813; Aiswarya Raj Munappy, Jan Bosch, Helena Holmström Olsson, Anders Arpteg, and Björn Brinne, “Data Management for Production Quality Deep Learning Models: Challenges and Solutions,” *Journal of Systems and Software* 191 (2022): 111359; Adriane Chapman, Elena Simperl, Laura Koesten, George Konstantinidis, Luis-Daniel Ibáñez, Emilia Kacprzak, and Paul Groth, “Dataset search: a survey,” *The VLDB Journal* 29, no. 1, (2019): 251–72.
31. American Library Association, “ALA’s Core Competences of Librarianship,” 6–7.

32. Timo Schick, Jane Dwivedi-Yu, Roberto Dessi, Roberta Raileanu, Maria Lomeli, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom, “Toolformer: Language Models Can Teach Themselves to Use Tools,” arXiv (2023). <https://arxiv.org/abs/2302.04761>, (accessed July 16, 2025).
33. Vladislav Lialin, Vijeta Deshpande, Xiaowei Yao, and Anna Rumshisky, “Scaling Down to Scale Up: A Guide to Parameter-Efficient Fine-Tuning,” arXiv (2023). <https://arxiv.org/abs/2303.15647>, (accessed July 16, 2025); Yuren Mao, Yuhang Ge, Yijiang Fan, Wenyi Xu, Yu Mi, Zhonghao Hu, and Yunjun Gao, “A Survey on LoRA of Large Language Models,” arXiv, 2023, <https://arxiv.org/abs/2407.11046>, (accessed July 16, 2025); Xiaohan Xu, Ming Li, Chongyang Tao, Tao Shen, Reynold Cheng, Jinyang Li, Can Xu, Dacheng Tao, and Tianyi Zhou, “A Survey on Knowledge Distillation of Large Language Models,” arXiv (2024). <https://arxiv.org/abs/2402.13116>, (accessed July 16, 2025).
34. Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu, “A Survey on Knowledge Graphs: Representation, Acquisition, and Applications,” *IEEE Transactions on Neural Networks and Learning Systems* 33, no. 2 (2022): 494–514, <http://dx.doi.org/10.1109/TNNLS.2021.3070843>, (accessed July 16, 2025).
35. Gillian Byrne and Lisa Goddard, “The Strongest Link: Libraries and Linked Data.” *D-Lib Magazine*. 16, no. 11/12, (November/December 2010), <http://www.dlib.org/dlib/november10/byrne/11byrne.html>, (accessed July 16, 2025).
36. Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu, “Unifying Large Language Models and Knowledge Graphs: A Roadmap.” *IEEE Transactions on Knowledge and Data Engineering* 36, no. 7, (July 2024): 3580–99, <https://doi.org/10.1109/tkde.2024.3352100> (accessed July 16, 2025); Bradley P. Allen, Lise Stork, and Paul Groth, “Knowledge Engineering Using Large Language Models,” in *Special Issue on Trends in Graph Data and Knowledge, Transactions on Graph Data and Knowledge (TGDK)* 1, no. 1 (2023): 3:1-3:19, <https://doi.org/10.4230/TGDK.1.1.3>, (accessed July 16, 2025).
37. American Library Association, “Core Competencies for Cataloging and Metadata Professional Librarians,” 5.
38. “responsible AI’ in Worldwide,” LinkedIn, [https://www.linkedin.com/jobs/search/?geoId=92000000&keywords=responsible AI](https://www.linkedin.com/jobs/search/?geoId=92000000&keywords=responsible%20AI), (accessed July 16, 2025).
39. Nestor Maslej, Loredana Fattorini, Raymond Perrault, Vanessa Parli, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika et al., “The AI Index 2024 Annual Report,” AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2024. [https://hai-production.s3.amazonaws.com/files/hai\\_ai-index-report-2024-smaller2.pdf](https://hai-production.s3.amazonaws.com/files/hai_ai-index-report-2024-smaller2.pdf) (accessed July 16, 2025).
40. Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample, “Llama: Open and efficient foundation language models.” arXiv (2023), <https://arxiv.org/abs/2302.13971>, (accessed July 16, 2025).
41. DeepSeek-AI, Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang et al., “DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning,” arXiv, 2025, <https://arxiv.org/abs/2501.12948>, (accessed on July 15, 2025).
42. AI Alliance, “Guide to Essential Competencies for AI,” 24-28.
43. American Library Association, “Core Competencies for Cataloging and Metadata Professional Librarians,” 5; American Library Association, “Professional Competencies for Reference and User Services Librarians.”

44. Hammad Rauf Khan, “Exploring Data Science Learning Objectives in LIS Education,” *Proceedings of the Association for Library and Information Science Education Annual Conference: ALISE*, 2020, 317-325, <https://www.ideals.illinois.edu/items/116457>, (accessed July 16, 2025).
45. Yin Zhang, Dan Wu, Loni Hagen, Il-Yeol Song, Javed Mostafa, Sam Oh, Theresa Anderson, Chirag Shah, Bradley Wade Bishop, Frank Hopfgartner et al., “Data science curriculum in the iField,” *Journal of the Association for Information Science and Technology* 74 no. 6, (202): 641–62, <https://doi.org/10.1002/asi.24701>, (accessed July 16, 2025); Shalini R. Urs, Mohamed Minhaj, “Evolution of data science and its education in iSchools: An impressionistic study using curriculum analysis,” *Journal of the Association for Information Science and Technology* 74, no. 6 (2023): 606–22, <https://doi.org/10.1002/asi.24649>, (accessed July 16, 2025).
46. Luca Belli, “What Leaders Need to Know About Auditing AI,” *Harvard Business Review*, March 29, 2025. <https://hbr.org/2025/03/what-leaders-need-to-know-about-auditing-ai>, (accessed July 16, 2025).
47. Yueqi Li and Sanjay Goel, “Making It Possible for the Auditing of AI: A Systematic Review of AI Audits and AI Auditability,” *Information Systems Frontiers* (2024), <https://doi.org/10.1007/s10796-024-10508-8>, (accessed July 16, 2025).
48. Gregory Falco, Ben Shneiderman, Julia Badger, Ryan Carrier, Anton Dahbura, David Danks, Martin Eling, Alwyn Goodloe, Jerry Gupta, Christopher Hart et al., “Governing AI safety through independent audits,” *Nature Machine Intelligence* 3 (2021): 566–571. <https://www.nature.com/articles/s42256-021-00370-7>, (accessed July 16, 2025).
49. “The world’s first advanced AI audit certification,” ISACA, <https://www.isaca.org/credentialing/aaia>, (accessed July 16, 2025).
50. Deepak Bhaskar Acharya, Karthigeyan Kuppan, and B. Divya, “Agentic AI: Autonomous Intelligence for Complex Goals—A Comprehensive Survey,” *IEEE Access* 13 (2025): 18912–36, <https://ieeexplore.ieee.org/document/10849561>, (accessed July 16, 2025).
51. Margaret Mitchell, Avijit Ghosh, Alexandra Sasha Luccioni, and Giada Pistilli, “Fully Autonomous AI Agents Should Not be Developed,” arXiv (2025). <https://arxiv.org/abs/2502.02649>, (accessed July 16, 2025).
52. Alexander Meinke, Bronson Schoen, Jérémy Scheurer, Mikita Balesni, Rusheb Shah, and Marius Hobbhahn, “Frontier Models are Capable of In-context Scheming,” arXiv (2025). <https://arxiv.org/abs/2412.04984>, (accessed July 16, 2025).
53. Eduardo Mosqueira-Rey, Elena Hernández-Pereira, David Alonso-Ríos, José Bobes-Bascarán and Ángel Fernández-Leal, “Human-in-the-loop machine learning: a state of the art,” *Artificial Intelligence Review* 56 (2023): 3005–54, <https://doi.org/10.1007/s10462-022-10246-w>, (accessed July 16, 2025); Uwe M. Borghoff, Paolo Bottoni, and Remo Pareschi, “Human-Artificial Interaction in the Age of Agentic AI: A System-Theoretical Approach,” arXiv (2025). <https://arxiv.org/abs/2502.14000>, (accessed July 16, 2025).
54. Korn Ferry, “Wanted for AI: Librarians”.

## Acknowledgements

The authors would like to thank Elsevier’s Discovery Lab for support of this research.

## Disclosure statement

No potential conflict of interest was reported by the author(s).