Bioinformatics Advances, 2026, **6**(1), vbag102
https://doi.org/10.1093/bioadv/vbag102
Perspective | Databases and Ontologies

OXFORD

AI in biocuration: challenges, opportunities, and a roadmap for sustainable integration

Valerio Arnaboldi[1], Lynn M. Schriml[2], Matt Jeffryes[3], Pengyuan Li[4], Susan Bello[5], Paul Sternberg[1], Carol Bult[5], Charles E. Cook[6,*]

[1]Caltech, Biology and Bioengineering, Pasadena, CA United States
[2]University of Maryland School of Medicine, Epidemiology and Public Health, Baltimore, MD United States
[3]EMBL-European Bioinformatics Institute, Wellcome Genome Campus, Cambridge, United Kingdom
[4]IBM Almaden Research Center, Almaden Research Center, San Jose, CA United States
[5]Jackson Laboratory, Mouse Genome Informatics Team, Bar Harbor, ME United States
[6]Global Biodata Coalition, EMBL-European Bioinformatics Institute, Wellcome Genome Campus, Cambridge, United Kingdom
*Corresponding author. Global Biodata Coalition, EMBL-European Bioinformatics Institute, Wellcome Genome Campus, Cambridge, UK.
E-mail: ccook@globalbiodata.org
Associate Editor: Michael Gromiha

Abstract

Motivation: **Biocuration is the integration of biological information databases for the enhancement of research. Curation of these databases is challenged by the exponential growth of scientific data and literature. Integration of machine learning and artificial intelligence methods into biocuration workflows may help address this challenge. We report on the discussions, ideas, and recommendations gathered from a workshop "AI and biodata resources: implications for sustainability and best practices in biocuration" at the 18th Annual International Biocuration Conference 2025.**

Results: **Participants agreed that while AI offers transformative potential for efficiency and expanded curatorial capacity, its integration faces substantial hurdles. Key challenges revolve around data and model quality. Reproducibility issues and a lack of open, domain-specific training datasets further compound these problems. Broader concerns include inconsistent data standards, underdeveloped ontologies, unstructured data, legacy systems, and underfunded teams. Despite these issues, several successful AI applications were identified, including tools for literature summarization and workflow assistance. Participants emphasized the need for a refined model of human-AI collaboration requiring clear data provenance and transparency, new skills, and avoiding over-reliance on AI-generated data. The workshop ultimately recommended concerted efforts in infrastructure development, standardization, training, and quality assurance to guide the community toward effective human-AI collaboration that maintains scientific rigor.**

1 Introduction

1.1 Biocuration

Biocuration is the process of organizing, and enhancing biological data and information relevant to biology into databases or biodata resources. This organization of information occurs at many different points in the research cycle. Researchers submitting data to repository databases must curate those data to ensure they are in the required format and include relevant metadata, and experts within those repositories must ensure that submitted data complies with submission standards and to ensure readiness for further analyses. Biocuration also involves mining the scientific literature, and archival databases, to extract information that adds value and knowledge to, and is included in, other databases and biodata resources. The participants of the workshop were primarily involved in literature-based biocuration, which is therefore the focus of much of the discussion below. However, the discussion, and conclusions below, are relevant for all aspects of biocuration.

For the past 30+ years, biocurators have typically undertaken curation by extracting relevant information from published papers or manuscripts and entering them into a database to enhance those resources (Howe et al. 2011, Burge et al. 2012, Skrzypek and Nash 2015, International Society for Biocuration 2018, Holinski et al. 2020). However, the exponential increase in scientific knowledge, and publications describing that knowledge, have exceeded the capacity of the field to curate all publications. In response, biocurators and database managers have been actively experimenting with machine learning and artificial intelligence methods to: 1) identify which publications should be curated; 2) undertake curation by extracting new information

Received: 23 December 2025. **Revised:** 5 March 2026. **Accepted:** 21 March 2026

from those publications; 3) explore how machine learning/artificial intelligence (ML/AI) can augment and enhance the curation workflow; and 4) understand how ML/AI utilization will impact the role of biocurators (König et al. 2018, Nielsen et al. 2021, Raciti et al. 2025).

Professional biocurators provide essential analytical assessment and knowledge synergy. They ensure that data submitted to archival biodata resources conforms to relevant standards and includes as much associated metadata as possible, and they integrate knowledge from the wealth of published biological data into highly utilized biological data resources. The next great challenge, addressed here, is how to maximize the utility of ML/AI for biocuration without losing the critical insights provided by human curators, and how do we address the concern, within the biocuration profession, of being replaced by ML/AI.

In a rapidly evolving landscape, where concepts and definitions are often blurred, we aim to provide clarity and a common reference point for discussion in the biocuration community. To this aim, this article reports thoughts, ideas and recommendations collected during the workshop "AI and biodata resources: implications for sustainability and best practices in biocuration," at the International Society for Biocuration (ISB; https://www.biocuration.org) Biocuration Conference in 2025 (https://www.stowers.org/events/biocuration2025), and builds on discussions from a previous workshop at the Biocuration conference in 2023 "Gaining perspective towards enhancing the intersection of biocuration and machine learning" (https://biocuration2023.github.io/workshops#ml). In addition to reporting the discussions at the workshop, we are offering, in this manuscript, a biocuration perspective on AI and ML.

The workshop revealed that while AI integration in biocuration faces substantial hurdles including model reliability, infrastructure limitations, and the need for human validation, it offers transformative potential for enhancing efficiency and expanding curatorial capacity when implemented thoughtfully. Participants developed actionable recommendations spanning infrastructure development, standardization efforts, training initiatives, and quality assurance frameworks to guide the community toward effective human-AI collaboration in biological knowledge curation. Here, we present these findings and recommendations to provide the biocuration community with practical guidance for navigating the challenges and opportunities of AI adoption while maintaining the scientific rigor and quality standards essential to biological knowledgebases.

1.2 Artificial intelligence and machine learning

Artificial Intelligence (AI) refers to the study and design of systems (software or machines) that can perform tasks normally requiring human intelligence, such as perception, reasoning, learning, problem-solving, and language understanding. More formally, as defined by Russell and Norvig 2020, AI is about building agents that perceive their environment and take actions to maximize their chances of achieving goals. In the context of biocuration, AI has been used to automate or semi-automate various curation tasks that require specialized knowledge, such as identifying relevant articles for curation (also known as "triaging"), associating articles with lists of biological entities (referred to as "indexing") and distilling knowledge from articles into annotations (information extraction, summarization, and standardization).

Over the years, different methods have been proposed to automate biocuration. Early solutions included simple rule based (or query) systems, such as keyword based search filters for triaging (Dowell et al. 2009, Van Auken et al. 2012). Thanks to Machine Learning (ML), a core subfield of AI, biocuration tasks have been enriched with algorithms to automatically learn from data, identify patterns, and make predictions or decisions without being explicitly programmed with rules for every scenario. Early examples of ML usage in biocuration are Support Vector Machines (SVM) and Neural Networks (NN) for triaging and for biological data annotation predictions (Fang et al. 2012, Hirschman et al. 2012, König et al. 2018,). Further advances in Natural Language Processing (NLP), an AI discipline that enables computers to understand and process human language, helped improve methods to automatically extract entities (Named Entity Recognition; NER) and relationships (Relationship Extraction; RE) from unstructured text, enabling highly automated biocuration workflows (Dowell et al. 2009, Wei et al. 2024). Figure 1 depicts a concise representation of the relationships between these fields.

As deep learning emerged as a subfield of ML where large neural networks are trained on large and complex data, models like Convolutional Neural Networks (CNN; Li et al. 2022) and Long Short-Term Memory (LSTM) neural networks (Hochreiter and Schmidhuber 1997) began to outperform earlier methods in biological text mining tasks such as NER and RE. A major breakthrough came with the introduction of pretrained language models such as BERT (Bidirectional Encoder Representations from Transformers; Devlin et al. 2019), which leveraged large-scale corpora and transfer learning to achieve strong results across diverse natural language processing tasks. Building on this, the development of generative models such as GPT-2 and GPT-3 introduced powerful capabilities in text generation and few-shot learning using decoder-only Transformer architectures (Brown et al. 2020). More recently, large language models (LLMs; Fralick et al. 2023, Telenti et al. 2024, Sarumi and Heider 2024) have evolved into foundation models, such as PaLM (Pathways Language Model; Chowdhery et al. 2023) and LLaMA (Touvron et al. 2023), characterized by massive parameter scales, improved training efficiency, and the ability to generalize across tasks without task-specific fine-tuning. These advances have culminated in state-of-the-art performance across reasoning, generation, and multilingual tasks, marking a new era of scalable, general-purpose AI.

The advent of pretrained language models like BioBERT (Lee et al. 2020) and PubMedBERT (Gu et al. 2022), which were adapted from BERT and trained on large-scale biological corpora, brought substantial improvements across a range of biological NLP tasks including NER, document classification, and question answering. The rapid adoption of these models was also facilitated by centralized repositories such as Hugging Face (https://huggingface.co) and the availability of standardized libraries that lowered the barrier to reuse, fine-tune and share ML models. However, these models primarily focused on generic language understanding. To support generation tasks and broader reasoning capabilities in the biomedical field,

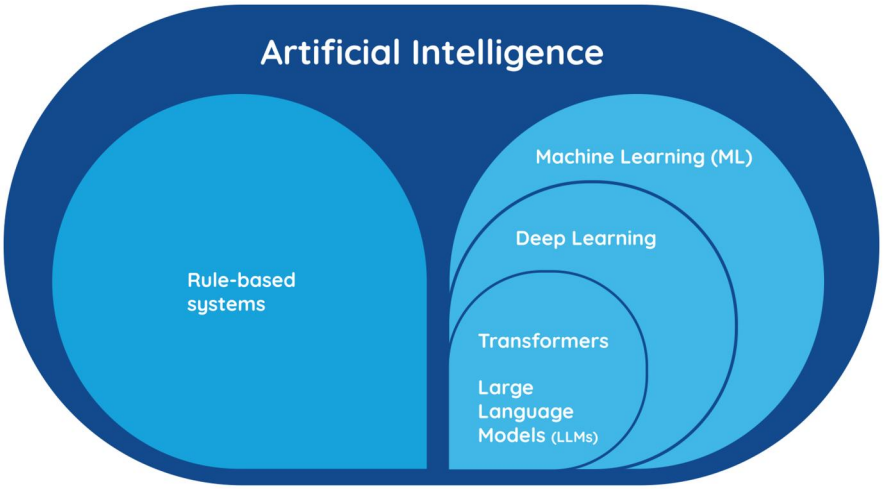

Figure 1 Conceptual Hierarchy of Artificial Intelligence and its Subfields. This figure illustrates the nested relationship between Artificial Intelligence (AI), Machine Learning (ML), Deep Learning, Transformers, and Large Language Models (LLMs). Rule-based systems are also included as a parallel, non-learning-based AI approach.

domain-specific generative models like BioGPT (Luo et al. 2022) and vision-language models such as BioMedGPT (Luo et al. 2026) were introduced. These models were pre-trained on biological literature to enable tasks such as relation extraction (e.g., protein-protein or drug-drug interactions), summarization, and biological question answering. More recently, large-scale foundation models such as Me-LLaMA (Xie et al. 2025) and BioMedGPT-10B (Luo et al. 2026) have been developed to integrate multimodal biological data (e.g., molecular structures and protein sequences) with natural language, expanding the potential of LLMs in drug discovery, clinical decision support, and biological knowledge curation. New specialized biocuration pipelines that integrate LLMs in various processing steps have been recently proposed in the literature (Kahsay et al. 2025, Poretsky et al. 2025, Islamaj et al. 2025). Moreover, publicly available biocuration tools for identifying relevant biological information from publications such as Pubtator have recently been improved with AI (Wei et al. 2024).

Recent techniques have improved the ability of LLMs to access biological data from the literature and from curated databases. In particular, Retrieval Augmented Generation (RAG) allows LLMs to access knowledge outside their training data and retrieve relevant information by searching for similarities between user prompt and sections of text documents (Zheng et al. 2025). Custom RAG LLM-powered chatbots have been successful in aiding both professional biocurators and biological database users in curating and finding relevant biological information (Poretsky et al. 2025). As one of the latest trends in AI, Agentic AI has strong potential in biocuration by autonomously orchestrating complex workflows across diverse biomedical domains (Toro et al. 2024). Through combining reasoning, retrieval, and code execution, Agentic AI can mine literature, tools (e.g., through the Model Context Protocol standard—MCP; https://modelcontextprotocol.io/), and databases at scale, reducing repetitive tasks and freeing curators to focus on higher-level scientific insights (Huang et al. 2025). Agent swarms promise to further scale concurrent execution of large numbers of AI agents for increased throughput. However, despite these promising developments, most current LLM-based systems are not yet fully aligned with the practical requirements and workflows of the biocuration community that demands high precision, transparency (i.e., knowledge provenance), and integration with existing curation pipelines. A major challenge in biocuration is the tendency of LLMs to generate factually inaccurate claims—often referred to as 'hallucinations' (Caspi and Karp 2024). Equally critical is the issue of data provenance, as curators must be able to reliably trace information back to verified references and ground-truth resources.

2 Workshop on AI in biocuration

The integration of ML and AI into workflows for curation of biodata resources represents a paradigm shift in how biological data are extracted from the scientific literature, curated, and shared with users. The recent advances and proliferation of LLMs, as noted above, are accelerating this transformation, bringing both opportunities and challenges. There is a pressing need for a shared understanding of what these tools can and cannot realistically deliver, and how these capabilities enhance, or detract from, human curation and management of biodata resources. In particular, will these new tools improve efficiency, leading to more sustainability, or will the costs exceed gains in efficiency?

We (Global Biodata Coalition; Alliance of Genome Resources; Human Disease Ontology Knowledgebase) (Anderson 2017, Schriml et al. 2022, Bult and Sternberg 2023; https://zenodo.org/records/7468719) organized a workshop, "AI and biodata resources: implications for sustainability and best practices in biocuration," at Biocuration 2025 in April 2025 that aimed to map the current landscape of ML and AI applications in biocuration, identify available tools, and discuss the successes and challenges faced while incorporating these technologies into the management of biodata resources, with particular emphasis on sustainability, cost, and transparency of their usage. Participants of the workshop included primarily professional

biocurators, alongside academic researchers and faculty, representatives from government and publicly funded agencies, commercial biotechnology and bioinformatics companies, scientific publishers, data-infrastructure consortia, and a small number of students and early-career scientists. This diversity broadened the discussion beyond routine curation to encompass research, sustainability, infrastructure, industry, and publishing perspectives. Here we summarize the conclusions from discussions among the participants of the workshop, and describe both successes and challenges experienced by the workshop participants during curation of their data resources. Our objective was to gain and then report an understanding of how current ML and AI models can support curation of biodata resources and to provide guidance on how to assess the cost/benefit of using these tools.

2.1 Overview of the workshop

The workshop was held on Sunday 6 April 2025 from 10:30–12:30 and 13:30–17:00, one day prior to the formal opening of Biocuration 2025. Both morning and afternoon sessions attracted 40 in-person (including six organizers) and 12 online participants. The workshop began with a 15-minute introduction that included a short poll of the participants (using Slido: https://www.slido.com) asking three questions intended to collect basic characteristics and opinions of the group and to act as an icebreaker and facilitate discussion. All questions permitted multiple responses from the same respondent.

1) What is your role? (Free text)
2) Are you using, building, or testing AI tools? (Multiple choice: Building AI curation tools, Using off-the-shelf tools like ChatGPT in curation, Using bespoke curation tools in curation, Testing off-the-shelf tools like ChatGPT in curation)
3) General feelings on using AI (one word)? (Free text).

A majority of respondents described themselves as curators. Nearly half stated a scientific role. Almost two thirds of respondents reported having tested off the shelf tools like ChatGPT in curation, with just over half actively using them. Respondents were invited to express their general feelings in a single word, or Emoji. Ambiguous responses were common, with ‘cautious’ being the single most common response. Categorising responses into ‘positive’, ‘negative’ or ‘ambiguous’; slightly more positive responses were provided than negative, though respondents could, and were, encouraged to give multiple responses if they wished. As this exercise was intended primarily to facilitate discussion, responses were not collected in a systematic or rigorous manner. Following the introduction, participants were assigned to six in-person groups and one online group to engage in a World Café discussion (Elliott et al. 2005). Each group was tasked with discussing three organiser-assigned questions, one from each of three topic areas (Table 1), during the allotted 75-minutes. Each group was asked to select a rapporteur, with each rapporteur presenting a short summary of their group’s discussion during the last 30 minutes of the morning session.

Following lunch, the afternoon session began with two panel discussions. The subject of the first panel (75 minutes) was “AI impact on biocuration.” Five panelists were asked before the workshop to prepare for discussion one of the six topics during the session (Tables 2 and 3). The panel discussion began with a short three-minute introduction from each panelist describing their experience working with AI or ML in biocuration. Following this, the floor was opened to all workshop participants to raise questions and discussion points. The second panel, also 75 minutes, discussed “AI usage and tool development.” Each of the four panelists was asked to prepare for discussion a series of topics related to AI tool development and application then, as

Table 1 Topic areas, questions, and group assignments for world café groups.

Sustainability	
Big picture: will AI allow for more efficient curation?	Group 1, Group 6
Will AI enable more curation, at a cost savings?	Group 2, online
Have you experienced or do you foresee AI impacting biocurator(s) employment?	Group 3
Do you see AI impacting the visibility of biocuration (positively or negatively)?	Group 4
Do you see your curation role changes (positively or negatively), with the usage of AI?	Group 5
Cost effectiveness of methods	
Where does AI save you benefit/time?	Group 1, Group 4, online
Has ML/AI been cost-efficient?	Group 2, Group 5
Have you experienced or do you foresee AI impacting biocurator(s) employment?	Group 3
Are there concerns regarding the validity of AI-generated data	Group 6
AI/ML usage	
Have you used AI/ML for biocuration? • Generic vs bespoke AI tools • Has AI/ML worked for your projects? • Tipping point, in deciding to use a particular model/tool	Group 1, Group 6
Are you building new tools based on AI/ML?	Group 2, online
Have you been asked to verify/validate machine generated results? Gr	Group 3
What are the primary tasks you have applied AI/ML to, such as classification, named entity detection, or summarization?	Group 4
AI/ML usage concerns (e.g., accuracy, provenance, opacity of commercial models)?	Group 5

Table 2 Panelists.

AI Impact on Biocuration panelists	
Robert Allaway	Sage Bionetworks
Kimberly Van Auken	Wormbase, Alliance of Genome Resources
Chris Hunter	GigaScience Press
Rachel Lyne	Wellcome Sanger Institute
Madhura Vipra	Medvolt AI
AI Usage and Tool Development panelists	
Andrew Green	EMBL-European Bioinformatics Institute
Benjamin Gyori	Northeastern University
Daniela Raciti	Wormbase, Alliance of Genome Resources
Melanie Vollmer	EMBL-European Bioinformatics Institute

Table 3 Discussion topics for afternoon panleists.

Panel 1 AI Impact on Biocuration
Community curation and AI/ML
Capacity building via AI (reduce costs, increase efficiencies, enable new capabilities)
Positive or negative impact on biocuration (improve work, impact on funding, workforce)
Practical implementation and management of AI/ML within working groups
Best practices in explaining innovation based on AI/ML in grants (justification and inclusion of AI experts)
How is the biocurator role changing? Current and future impact of AI/ML
Panel 1 AI Usage and Tool Development
AI adoption examples, Biocuration applications
Off the shelf use—ChatGPT—using in their daily tasks
AI FAIRness and usage transparency
AI Trustworthiness metrics
How to create a benchmark for using models for the biocuration community
Criteria for producing AI-ready datasets (biocuration role)
Where do you source AI-ready datasets
How can we make biocuration datasets more accessible to LLMs (structured information vs textual information)
Allow/limit access to curated information by LLM bots

with the first panel, each panelist presented a brief three-minute introduction and the floor was opened to workshop participants to start discussion (Tables 2 and 3). After the panel discussion the workshop ended with brief concluding remarks from the workshop organizers.

3 AI applications in biocuration

Participants reported successful integration of AI tools in a variety of biocuration tasks. Some participants used AI as part of their literature processing, to classify the subject matter of articles, or to estimate their relevance to the resource they work on. Others used LLM-based tools to directly extract and summarise knowledge.

The LitSumm tool was developed to summarise literature on noncoding RNAs for the RNAcentral resource (Green et al. 2025). This approach retrieves relevant articles from the literature, extracts the sentences discussing the target RNA, and then prompts an LLM with the sentences and appropriate instructions. The output is then filtered, and in some cases, the LLM is invoked again to correct substandard summaries.

The Alliance of Genome resources has developed GeneScribe, which uses an LLM to produce a summary of a gene's function, expression, involvement in biological processes and other available information derived from Gene Ontology annotations (Li et al. 2025). Reactome is developing a tool to assist curators, using RAG to produce a summary of biological pathways, which can then be used as the basis for a curator to complete the entry in the knowledgebase. The LLM is instructed to include references and quotations, so that the curator can verify the accuracy of statements (Gillespie et al. 2025).

Participants also reported the use of AI tools to assist in their curation workflows, as an assistant for repetitive processes. The CurateGPT suite of utilities (Caufield et al. 2024) assists with literature search, and with knowledgebase search, and summarises information found. It can also generate new ontology terms, or find existing matches. It is able to access a variety of important biomedical resources, and integrate the information within them using RAG. Either local or remote LLM endpoints can be used in CurateGPT. Participants also reported more ad hoc use of generic tools, like ChatGPT. In contrast to prior machine learning applications to biocuration, no specialist knowledge is required to begin using LLM chatbots: they are able to follow natural language instructions and output information in formats described by the user, allowing them to be integrated into existing workflows without modification.

Beyond these specific tools, workshop participants described a range of successful implementations across diverse organizational contexts and curation domains. Research institutions reported focusing on developing custom tools for specific domains, often through collaborative partnerships with computer science departments. Database organizations have been implementing AI primarily for literature processing, automated annotation, and data validation workflows, while commercial entities are developing AI-powered curation tools and services designed for broader community adoption. Individual practitioners shared experiences using existing AI tools—including ChatGPT (https://chatgpt.com), Microsoft Copilot(https://copilot.microsoft.com), Elicit (https://elicit.com), and ChatPDF (https://www.chatpdf.com)—for daily curation tasks, highlighting the accessibility of modern LLM-based approaches.

Use cases for data extraction and validation extended well beyond literature summarization. Participants reported successful applications in information extraction from supplementary materials, protein structure validation through literature analysis, gene function prediction and annotation, and synonym identification for ontology mapping. Tools like CurateGPT, DRAGON-AI (Toro et al. 2024) and Ontomate (Liu et al. 2015) have proven particularly valuable for ontology-related tasks, as well as for enhancing human-in-the-loop curation (Tsaneva and Sabou 2024). For workflow enhancement, teams have integrated machine learning tools into existing curation software to enable automated quality checking and error detection, metadata standardization and organization, and support for community curation pipelines. The GO curation support for microRNAs exemplifies how AI can assist with highly specialized annotation tasks. Practical utilities like pdftotext

(https://pypi.org/project/pdftotext/) for Linux command-line extraction, combined with LLMs for literature recommendations and processing, demonstrate that effective AI integration often combines sophisticated LLMs with simpler, targeted tools to create robust curation workflows

4 Results from the workshop

The discussions at the workshop brought together diverse perspectives from panelists and attendees working at the intersection of computational methods and biological data curation. Below, we provide a graphical summary of the key challenges, opportunities, and possible solutions for applying AI to biocuration that participants shared and developed during the sessions, then discuss these topics in detail in the sections below. These insights reflect both practical lessons learned in applying AI/ML to real-world biocuration tasks and broader reflections on the evolving role of these technologies in the community (Figure 2).

4.1 Key challenges

4.1.1 Data and model quality

The biocuration community faces significant challenges with AI-generated outputs that can appear confident while containing critical errors. AI tools frequently produce "hallucinations"—in the context of biocuration these include incorrect ontology term identifiers, false gene-function statements, and other factually inaccurate claims that pose serious risks in applications where scientific accuracy is paramount. This concern is compounded by the well-documented tendency of LLMs to make confident but incorrect assertions (Lin et al. 2024, Jiang et al. 2024). Performance deteriorates markedly when AI systems encounter poorly structured inputs (Sundaram and Musen 2025). Curators routinely report that inconsistent terminology, missing metadata, and badly formatted publications confuse AI models, requiring additional preprocessing and curation effort that workshop participants note can negate potential efficiency gains, but agree that AI can assist by treating metadata cleaning, standardization, and gap-filling as part of a preprocessing stage. At the same time, improving metadata quality at the source is critical. Publishers and repositories have a responsibility to ensure that scientific outputs are accompanied by correct and comprehensive metadata, thereby reducing downstream burdens on curators and making AI applications more robust. In practice, progress will require a combination of both strategies: upstream improvements in publishing practices and downstream AI-assisted tools that help bridge the inevitable gaps in real-world data.

The scope of collecting data for training and fine-tuning ML models presents another critical limitation. Many biocuration datasets lack explicit negative examples or rich contextual detail, which constrains model robustness and can lead to biased or over-confident predictions within narrow domains (González et al. 2025). When domain-specific training data are unavailable, models must rely on broader, potentially less relevant datasets for inference, introducing uncertainty into specialized biocuration tasks (D'Amour et al. 2022). Human validation remains indispensable across all AI applications in biocuration (Boman 2023, Chen et al. 2024, Niyonkuru et al. 2025). Even sophisticated high-throughput AI pipelines demonstrate frequent errors, including incorrect associations, identification of non-target entities, and various forms of hallucination. This reality means that manual review continues to be essential for correcting AI output, potentially negating some of the anticipated efficiency gains and, in the short term, may actually increase the manual validation burden on biocuration groups.

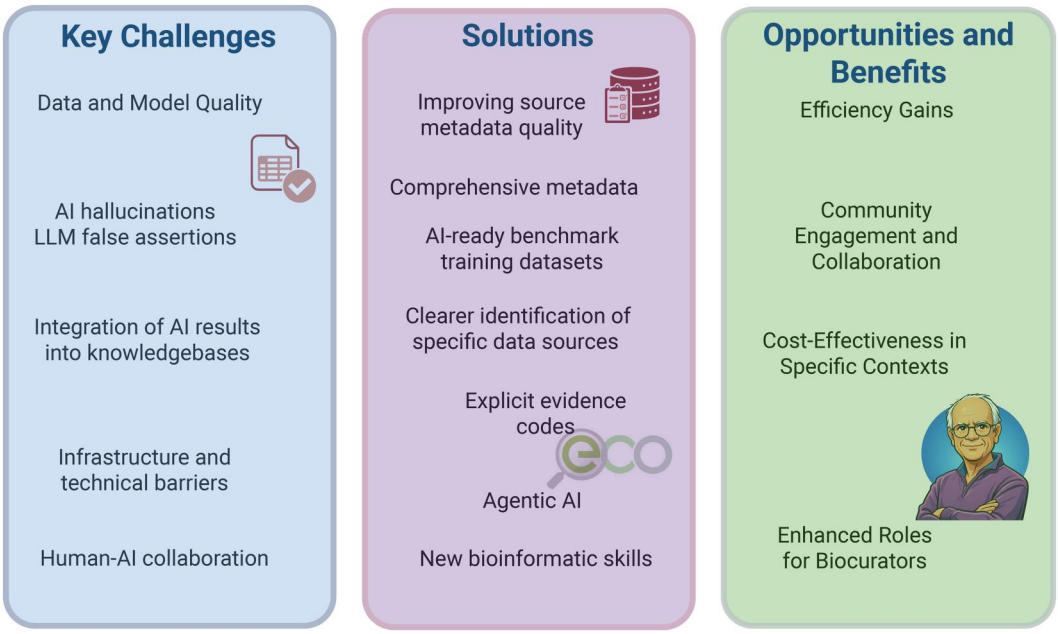

Figure 2 Challenges, opportunities, and solutions for AI in biocuration. Summarizes the key challenges to using AI for biocuration, suggested solutions to those challenges, and the opportunities and benefits of using AI for biocuration. In the third panel we represent all biocurators with an image of Amos Bairoch, who was a pioneer of biocuration (Bateman 2026).

Reproducibility issues plague many AI implementations, with model outputs varying significantly based on different prompts, model versions and tasks (Wang and Wang 2025). This variability creates inconsistent results unless prompts and parameters are meticulously controlled, making it difficult to establish reliable workflows and creating substantial challenges for collaborative efforts across institutions. This is in part inherent to the models: modern generative and large-language models are stochastic by design and rely on sampling methods, which introduce randomness in output even for identical inputs. Deploying models locally (or self-hosted) so that one has full control over all inference parameters (seed, decoding method, hardware, environment) and running multiple inference runs for the same prompt and then averaging, majority voting or otherwise aggregating the results are some of the proposed solutions to mitigate the issue (Wang and Wang, 2025). Real-world scenarios often present nuanced situations that do not align neatly with traditional binary classifications. False positives and false negatives are often assumed to be straightforward binary classifications, but real-world scenarios can be more nuanced, such as when AI correctly identifies a target element that exists but proves irrelevant to the specific context, creating results that do not clearly fall into either traditional category (Borlund 2003).

The biocuration community has expressed a clear preference for open models, favoring locally hosted, open-source AI systems that offer greater transparency and control. This preference reflects widespread caution toward opaque proprietary systems and aligns with broader scientific values of transparency and reproducibility. However, security concerns arise when using LLMs with data containing sensitive information about research study participants. While open-source models may help mitigate these issues, they currently lag behind closed-source solutions in performance (Zhang et al. 2024).

Integration of AI results into knowledgebases raises quality concerns, as AI-generated content may have lower quality than manually curated data, but the mixing of sources can obscure data provenance. This highlights the need for clearer identification of specific data sources and more explicit evidence codes to improve provenance tracking in knowledgebases. Specifically, The current ECO codes (Nadendla et al. 2022) for machine predicted annotations are not sufficient to describe data origins for LLM-generated data. Possible solutions proposed by the workshop participants include expanding ECO codes and adopting different types of labelling for ML-derived information.

4.1.2 Lack of standards

Biocuration is inherently complex as it involves extracting, interpreting, and standardizing biological knowledge from large volumes of heterogeneous scientific literature. This process is further complicated by ambiguities in experimental data, evolving ontologies, and inconsistent reporting standards, all of which challenge the maintenance of accuracy (Groß et al. 2012). This complexity makes harmonization across different domains challenging, and renders full curation by AI models difficult and error-prone, particularly when systems are not fully customized to specific biocuration subfields or organisms (Li et al. 2024).

Inconsistent reporting formats in the scientific literature create substantial barriers to AI implementation. Publications frequently omit standardized identifiers or employ non-uniform formatting, such as free-text methods sections or text and tables included as images, which severely hinders machine readability (Atkinson et al. 2024, Mahadevkar et al. 2024). When standard metadata or experimental details are missing, automated parsing becomes unreliable, forcing curators to invest additional effort in data extraction processes.

The biocuration field suffers from a scarcity of benchmark datasets and lacks widely adopted gold-standard corpora for AI-supported biocuration. This absence makes it difficult to evaluate or compare tools effectively, impeding progress and making it challenging for teams to select appropriate tools for their specific needs (Chatr-Aryamontri et al. 2011, Hirschman et al. 2012). The development of structured screening checklists could help identify well-described papers suitable for AI assistance, but such systematic pre-screening approaches remain underdeveloped (Leitner et al. 2010). Similarly, the creation of "competency questions" or similar frameworks could significantly improve the efficiency of AI-assisted curation workflows (Leitner et al. 2010, Hirschman et al. 2012).

Incomplete ontologies present ongoing challenges, particularly in specialized domains where comprehensive controlled vocabularies are lacking (Nijsten et al. 2016, Solovieva et al. 2018). For instance, glycans have almost no machine-readable function terms, and reporting standards in that field remain minimal (Solovieva et al. 2018, Ibrahim et al. 2025). When standard ontologies do not exist, AI systems may be forced to guess or generate term identifiers, introducing potential errors into curated datasets (Nijsten et al. 2016). Even where ontologies exist, their inconsistent application leads to variable data interpretation and missing reporting standards, resulting in widely variable annotations (Nijsten et al. 2016, Ibrahim et al. 2025). In practice, this means AI tools may invent or misuse identifiers when correct terms are not clearly documented.

4.1.3 Infrastructure and technical barriers

Unstructured data pipelines represent a major bottleneck in AI adoption for biocuration. Curators frequently contend with articles in standard file formats like PDF (Portable Document Format), and websites that lack Application Programming Interfaces (APIs), requiring significant effort to build custom pipelines for extracting usable text from these sources. The technical complexity of these extraction processes exceeds the capabilities of typical curation teams, creating a fundamental skills gap that limits implementation. This challenge is not unique to biocuration; across industries, unstructured data management is increasingly recognized as a strategic enabler, yet it remains a significant hurdle for AI integration (Mahadevkar et al. 2024).

In recent years, significant efforts have been made to make scientific content openly available. With the recent implementation of the Nelson memo, even more research outputs are expected to become open access, which will increase the amount of freely accessible content (https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf). However, a substantial portion of relevant content still remains behind paywalls and is not freely accessible. Some knowledgebases have implemented ad hoc solutions to retrieve non-open access content through institutional subscriptions, but unifying and streamlining access for AI systems remains a major challenge. Furthermore, content is often available in multiple

languages (for example, as of 30 September 2025, 1.52% of scientific articles included in the Alliance of Genome Resources literature database are not in English; Alliance of Genome Resources Consortium 2024), and non-English publications pose additional challenges for AI due to the limited availability of high-quality training data in those languages. Querying PubMed for "Artificial Intelligence" (30 September 2025), identifies 324,857 publications, of which 318,958 (98%) available in English. Utilizing PubMed filters, there are 2,560 in Chinese, 1265 in German, 580 in French, and 442 in Spanish. (https://pubmed.ncbi.nlm.nih.gov/? term=%27Artificial±intelligence).

These limitations create two distinct types of gaps for AI-assisted biocuration. First, there may be gaps in the curated knowledgebases themselves, which can be partially mitigated by negotiating access and adding derived data while respecting copyright. Second, and more critically, gaps in the literature accessible for training and fine-tuning AI models limit the comprehensiveness of automated curation, potentially leaving systematic blind spots in what AI can identify and extract (Ibrahim et al. 2025). On the other hand, not every human curation team has experts in every research area, so knowledge gaps are also inherent in manual curation, highlighting a shared limitation between AI and human curators (Ibrahim et al. 2025).

Resource constraints significantly restrict AI adoption across the biocuration community. Teams tend to be small and underfunded, making the upfront costs of developing and maintaining AI workflows, including compute resources, software engineering, and cloud costs, prohibitive. Limited personnel and budgets make even prototype systems difficult to sustain over time, creating a barrier to long-term AI integration. Domain adaptation challenges emerge when attempting to apply general AI tools to specific biocuration tasks. Off-the-shelf NLP tools typically underperform on specialized biological tasks without extensive customization (Wang et al. 2025). Experts at the workshop noted that integrating general models often encounters a "reusability wall," requiring models to be retrained or extensively tuned for each subdomain, multiplying development costs and complexity (Sapkota et al. 2025). Agentic AI offers a promising solution to this issue by enabling autonomous, purpose-built agents that can adapt to specific domains and execute tasks efficiently, reducing the need for extensive retraining (Karunanayake et al. 2025).

Legacy systems pose additional integration challenges for many biocuration groups that rely on older codebases and limited compute infrastructure. Integrating new AI components into these established pipelines often requires substantial refactoring and creates ongoing maintenance challenges that may exceed current technical capabilities within teams. The increased use of crawling bots to collect training data for LLMs has placed significant strain on biocuration databases and resources, with some platforms reporting outages or dramatically higher operational costs due to uncontrolled scraping traffic (Kwon 2025). This creates a technical and budgetary barrier for biocuration groups, who must balance the goal of maintaining open data accessibility and discoverability with the financial burden of scaling infrastructure to handle automated harvesting. At the same time, the lack of standardized, AI-ready datasets means that much of the collected content is inefficiently used, reinforcing the need for curated, structured, and accessible formats that can support both human and machine use (Wilkinson et al. 2016, Sansone et al. 2019). Thus, the challenge is twofold: mitigating the immediate technical risks posed by aggressive data scraping while also investing in long-term solutions that deliver reliable, FAIR-compliant datasets optimized for AI-assisted biocuration workflows.

All in all, the discussion at the workshop highlighted that managing AI tools presents new challenges that require skills not traditionally found in biocuration teams. Keeping track of available tools, providing unified access for research groups, and monitoring costs requires new management competencies that can be particularly challenging for those unfamiliar with AI technologies. The infrastructure management burden represents an additional barrier to adoption.

4.1.4 Human-AI collaboration

Trust and transparency issues arise from "black box" AI outputs that lack clear provenance, though recent developments, such as Google's Gemini Deep Research, demonstrate that newer models can provide detailed citations and explanations for their outputs, potentially addressing some transparency concerns. Nevertheless, the opacity of many AI systems continues to create hesitation among curators who need to understand the basis for automated recommendations (Budhkar et al. 2025).

Effective AI utilization requires new skills that biocuration professionals must acquire. Prompt-engineering expertise and understanding of model behavior have become crucial competencies, with crafting appropriate queries proving time-consuming and requiring specialized knowledge (Knoth et al. 2024). Training curators in these new skills has become essential for successful human-AI collaboration, but formal training programs combining curation expertise with AI literacy remain scarce. The risk of over-reliance on AI presents a significant concern for maintaining curation quality. Undue trust in AI outputs, combined with the potential need for extensive manual review to identify errors, may actually increase workload rather than reduce it (Boyacı et al. 2024). There is growing concern that AI mistakes could propagate into databases and become extremely difficult to identify and correct over time, potentially compromising the integrity of biological knowledgebases (Chen et al. 2024).

Training and sustainability challenges reflect the rapid pace of AI development. While there is a growing call for formal education (see for example the "AI and new coming trends in bioinformatics training" program at the 2025 edition of the Global Bioinformatics Education Summit), adoption of robust checklists (Chen et al. 2024) and validation frameworks, very few established training programs currently exist that adequately combine traditional curation expertise with modern AI literacy. Validation frameworks are particularly important because they provide structured approaches for assessing reproducibility, generalizability, and fairness of AI models before their outputs are integrated into biological databases. Beyond generic benchmarking, frameworks in healthcare and bioinformatics emphasize transparent reporting, continuous monitoring, and stakeholder involvement to ensure that automated outputs align with domain-specific standards (Amann et al. 2020).

Another important consideration is how specific AI training for biocurators needs to be: while domain-specific modules on

curation workflows are essential, the general principles of how AI models function, ranging from conceptual understanding of model behavior to hands-on exercises such as calculating attention mechanisms in simplified settings, can be drawn from a wide array of existing machine learning courses. This layered approach would allow biocuration professionals to build upon general AI literacy while tailoring training to the unique demands of knowledgebase construction and maintenance. At the same time, the environmental and social costs of AI adoption raise additional questions about whether widespread implementation represents a worthwhile investment, particularly given concerns that replacing entry-level curatorial positions may prove more expensive in the long run.

4.2 Opportunities and benefits

4.2.1 Efficiency gains

Despite the substantial challenges, workshop participants identified remarkable opportunities for efficiency improvements through AI integration. Panel discussions revealed dramatic potential time savings, and recent studies in biocuration and related workflows have reported significant reductions in screening and annotation effort (Perlman-Arrow et al. 2023, Niyonkuru et al. 2025). These efficiency gains are particularly pronounced in several key areas such as literature screening and ontology mapping. Notably, initial literature screening and publication filtering are areas where AI can provide substantially better results than current custom solutions (Perlman-Arrow et al. 2023) while demonstrating superior adaptability to emerging methodologies and evolving publication trends. AI systems excel at metadata interpretation and organization, offering capabilities that extend to data extraction and harmonization from diverse and complex sources, including PDF tables from supplemental files, Excel spreadsheets, images, and videos.

Additional areas showing significant promise include: synonym identification and term mapping (normalization) (Niyonkuru et al. 2025, Wiegers et al. 2025); gene function summarization (Green et al. 2025) and function prediction (Joachimiak et al. 2024); knowledge graph building and ontology development (Toro et al. 2024); and hypothesis generation (O'Brien et al. 2024). Recent improvements in RAG models (Lewis et al. 2020) and other LLM enhancements such as Agentic AI show particular potential for reducing spurious results while maintaining high throughput (Zheng et al. 2025, Huang et al. 2025). AI tools also demonstrate value in identifying knowledge gaps, handling helpdesk tasks, and recognizing which processes can be automated effectively.

4.2.2 Enhanced roles for biocurators

The integration of AI into biocuration workflows is not merely replacing human biocurators but is transforming and expanding their roles, creating new opportunities for expertise and oversight. Rather than eliminating positions, AI is facilitating the evolution of biocuration into a more dynamic and collaborative field.

Validation specialists are becoming essential: these are experts who review and correct AI outputs, ensuring accuracy and maintaining quality standards. Prompt engineers are developing as professionals who design and optimize AI interactions, crafting queries and instructions that maximize system performance for specific biocuration tasks. AI-human workflow coordinators are emerging as specialists who design and manage hybrid curation pipelines, optimizing the integration of automated and manual processes. Quality assurance analysts represent another growing role, focusing on developing and implementing validation frameworks that ensure AI-assisted curation meets scientific standards, assuring that AI predictions are technically sound by following best practices and checklists during development and deployment, while also that the predictions are scientifically sound based on raw input and content. Critical thinking will be core to content quality assurance. Curators' domain-specific expertise allows them to cross-examine statements produced by an algorithm to judge its validity and likely truth. These roles collectively address the reality that while AI can handle increasingly tedious work, the need for human oversight and expertise remains critical.

Workshop participants expressed optimism that AI integration may actually increase employment opportunities for biocurators as the demand for validation and oversight of AI-generated annotations grows. The shift toward less tedious work for curators may eliminate some entry-level positions, but given that most research groups operate with limited funding, this transformation may enable projects to achieve greater effectiveness with existing staff levels. In particular, and converse to the initial concerns of AI replacing humans, workshop participants reported that the role of biocurator will probably neither disappear nor diminish, but instead is predicted to increase in demand. The role is evolving, and AI will be leveraged by biocurators to handle the growing volume of data, which exceeds what curators can manage manually and creates a greater need for biocuration expertise. This evolution requires increased collaboration between curators and AI experts, with curators needing to develop skills in instructing and managing AI systems. Integration of AI results with additional context, such as highlighting source text and presenting it alongside results, appears to improve both transparency and usability, as demonstrated by tools like ChatPDF (https://www.chatpdf.com) and sentence classification systems (Raciti et al. 2025).

4.2.3 Cost-effectiveness in specific contexts

While initial implementation costs can be substantial, AI tools demonstrate clear cost-effectiveness for well-defined applications. High-volume, routine curation tasks show particular promise for cost savings, as do initial screening and filtering operations where AI can help human curators process large quantities of literature more efficiently (Poretsky et al. 2025, Wiegers et al. 2025). The cost of AI adoption in biocuration has three main components: data preparation, system design and deployment, and model inference. Preparing data, which involves cleaning, and harmonizing data, defining annotation schemas, and creating gold-standard datasets, represents a large investment, as it requires substantial human expertise. Similarly, designing and deploying curation pipelines based on AI requires considerable development effort and specialized biocuration expertise. Inference costs, by contrast, are usually lower but can rise when models are accessed through commercial APIs, where usage fees scale with volume.

One way to reduce inference costs while improving data security is the use of offline, open-source models–e.g., Llama (Touvron et al. 2023), Gemma (https://blog.google/technology/

developers/gemma-open-models), Mistral (https://huggingface.co/mistralai), DeepSeek (https://huggingface.co/deepseek-ai), and Qwen (https://huggingface.co/Qwen). Deployed in quantized form through platforms like Hugging Face or Ollama, these models can run within fully closed systems. Such air-gapped environments eliminate vendor data sharing and are particularly valuable for sensitive curational workflows, while also lowering long-term expenses compared with recurring API access. Standardized annotation processes benefit significantly from AI automation, particularly when dealing with large-scale literature processing requirements. An additional benefit that emerged from the workshop discussions is the potential for AI to identify results that human curators might miss, thereby improving the comprehensiveness of curation efforts while potentially reducing costs. The cost-effectiveness equation becomes particularly favorable when AI tools can be applied to repetitive tasks that require consistent application of established rules or criteria, freeing human experts to focus on complex interpretation and quality assurance activities that require domain expertise and nuanced judgment.

4.2.4 Community engagement and collaboration

The workshop highlighted growing community engagement around AI integration in biocuration, with several positive trends emerging across the field. Open tool development and sharing initiatives are fostering collaboration and reducing duplicated effort across institutions. Collaborative learning resources, exemplified by efforts such as the Alliance of Genome Resources AI working group, are helping teams share knowledge and best practices. Cross-institutional partnerships are forming to tackle common challenges and pool resources for AI development projects. These collaborations often involve multidisciplinary team formation that combines expertise from biology, computer science, and data science, creating more robust approaches to AI integration in biocuration workflows. Computer science groups often have ready access to models, quantization and deployment tooling, and algorithmic expertise while biocurators bring domain knowledge, curation criteria, and data sources. Hybrid teams can jointly reduce total costs and improve quality by combining their strengths. The community's commitment to open science principles appears to be extending into AI tool development, with many groups prioritizing transparent, reproducible approaches that can benefit the broader biocuration community. This collaborative spirit may prove essential for overcoming the technical and resource challenges that individual institutions face when implementing AI solutions independently.

5 Workshop community recommendations and action items

Following the workshop, as a group, we developed a list of consensus recommendations and associated action items identified through the presentations, discussions and breakout groups. These topics were then organized topically into six categories, presented here to guide future discussions, outline our roles as biocurators and to provide practical next steps for utilizing AI in biocuration.

5.1 Infrastructure and collaboration

Shared repositories for curated data: participants recommended using shared infrastructure or building new solutions when necessary (e.g., common databases or annotation platforms) where AI-extracted results and human-curated annotations can be stored together, reviewed, and improved collaboratively. Community agreed metadata standards will underpin the repositories to make the curated data sets easily findable, and machine-readable formats will be crucial to use those data to their full potential. Datasets could be used for different applications, and metadata standards could also help define data requirements for different use cases (e.g., "well-curated" data for one application may not be detailed enough for another).

Federated or distributed models and data: suggestions emerged to support multiple groups sharing partial datasets and models/pipelines, connected through standards rather than centralization. This approach respects domain specificity while enabling broader reuse.

Shared annotation tools: the community called for "curation-as-a-service" platforms where domain experts and AI developers could co-annotate documents or validate outputs together, for example TeamTat (Islamaj et al. 2020); Hypothesis (https://web.hypothes.is/);

Markup (Dobbie et al. 2021); and other similar tools (Neves and Ševa 2021). This would facilitate tool testing, training data generation, and feedback collection.

Model registries: participants proposed creating a public registry of models used in biocuration workflows, including details on training data, tuning, domains, and validation methods, either on generic AI/ML platforms like Hugging Face or specialized repositories like BioModels (https://www.ebi.ac.uk/biomodels). This would improve transparency and encourage responsible reuse and ensure accessibility in the long term.

Shared prompt libraries: many endorsed developing a communal repository of prompts, workflows, and use-case templates so that biocurators can build on each other's best practices and avoid reinventing the wheel.

Modular pipelines: there was encouragement to build more modular, interoperable AI pipelines that can be customized to individual projects but still plug into broader infrastructures or standards.

5.2 Evaluation and validation

Gold-standard corpora: multiple participants emphasized the need for curated, domain-specific gold-standard datasets, and corpora, that can be used to evaluate AI systems in biocuration tasks. These should be openly available and maintained by the community. Collaborative opportunities with these types of initiatives, for example, the Bridge2AI's Voice Dataset (https://b2ai-voice.org/the-b2ai-voice-dataset/), hosting and defining AI ready datasets.

Formal benchmarking tasks: there was broad support for defining standard evaluation tasks (e.g., gene-function extraction, entity linking) with clear metrics and test sets, allowing fair comparison across models and workflows. The focus should be on what makes a good benchmarking dataset, most likely defined by well thought-out metadata. Then the specific

application for which data are useful can come from any domain; e.g., chemical, functional, structural, or gene-disease. Defining guidelines and a testing pipeline to assess user-submitted datasets on whether they meet the criteria to make them benchmarking datasets will, in the long run, expand the availability of benchmarking sets.

Model performance tracking: participants proposed a dashboard or leaderboard showing model accuracy on key biocuration benchmarks, updated over time. This would help curators choose appropriate tools and hold developers accountable.

Error cataloging: some suggested building shared repositories of common model errors (e.g., types of hallucinations, misclassifications) to guide tool improvement and training.

Standardized reporting metrics: the community identified the need for consistent metrics to measure AI implementation success and impact across different organizations and use cases.

5.3 Ontologies and standards

Co-development of ontologies: there was consensus that ontology developers should work closely with biocurators and AI developers to ensure terms reflect real-world usage and are formatted for machine readability. This will also enable identification of cross-over points between ontologies, in particular for concepts and entities that can assume multiple roles in different contexts, resulting in more comprehensive and semantic coverage.

Community workshops: participants called for regular domain-specific workshops (like those held for glycobiology and microbial data) to assess ontology gaps and propose targeted updates.

Minimal information standards: support emerged for reviving or expanding "minimum information" checklists (e.g., MIAME, MINSEQE, MIAPPE, MIxS; see: https://www.researchobject.org/initiative/mim/) to help authors report key data elements needed for machine-assisted extraction.

Structured paper formats: some proposed advocating for more structured publication formats (e.g., tagged methods sections, machine-readable tables) to make AI parsing easier, potentially in collaboration with journals. Since PDF conversion and text extraction are needed for most AI/ML applications, having machine-readable files available alongside PDFs from publishers would save time and computational resources for the biocuration community at large.

Standardized output formats: tools should output results in standardized, interoperable formats (e.g., BioC, JSON-LD, RDF) to enable integration with existing databases and ontologies.

Schema.org implementation: participants recommended implementing schema.org metadata and adopting metadata standards like Croissant (https://research.google/blog/croissant-a-metadata-format-for-ml-ready-datasets/), as well as proper robot.txt crawling rules to manage AI bot access while maintaining accessibility. This would help with the creation of standardized models and benchmarking sets and would simplify access to models, data and tools, benefitting services based on protocols for LLMs such as MCP (Model Context Protocol).

5.4 Training and capacity building

Prompt engineering training: participants emphasized the need for formal training on how to effectively interact with LLMs, including prompt design and model behavior troubleshooting.

Biocuration curriculum: several participants called for establishing accredited or recognized training programs for biocurators, especially ones incorporating AI and NLP components. The community noted that these skills do not necessarily require formal university education, as online courses and resources are becoming increasingly available.

Mentorship networks: suggestions included building mentorship structures to help new curators learn from more experienced peers, particularly in rapidly evolving areas like AI integration.

AI tool literacy: broader education initiatives were recommended to improve comfort and literacy with AI tools across the biocuration workforce—not just technical training, but understanding strengths, limitations, and use cases.

Curation career support: calls to recognize curation as a formal career path, with stable funding, training, and job opportunities, especially in institutions and funding agencies.

Community seminar series: participants proposed organizing ongoing seminars where community members can share their experiences and knowledge about implementing AI tools in biocuration.

5.5 Community organization and policy

Working groups: multiple discussions highlighted the need for ongoing cross-community working groups to maintain shared benchmarks, coordinate ontology development, and advise on standards.

Shared funding proposals: some recommended forming consortia to apply for joint funding (e.g., for tools, corpora, infrastructure) towards solving targeted biological challenges, curating data for a dedicated application, as single groups rarely have sufficient resources. Cross-departmental applications comprised of computational science groups along with those hosting the data.

Engagement with publishers: participants proposed engaging with scientific publishers and journals to promote reporting standards, encourage structured formats, and incentivize machine-readable submissions. This includes requiring disclosure of AI usage in manuscript submissions, as some publishers have already begun implementing. Guidelines from the biocuration community can inform publishers and their reviewers on emerging trends and best practices.

Institutional advocacy: support emerged for coordinated advocacy to funders and academic institutions to raise the profile of biocuration and ensure AI integration is done responsibly.

Ethics and governance: some discussions addressed the need for ethical guidelines and governance frameworks to prevent misuse of AI tools and ensure curated outputs remain reliable and transparent.

Clear labeling systems: the community recommended implementing clear disclosure systems to distinguish between AI-generated and manually curated content in databases and publications. Usage guidance for current and potentially expanded rules defined by the Evidence and Conclusion Ontology (ECO)

(e.g., rule based: ECO : 0000256; ECO : 0000259; deep learning: ECO : 0008006) could further this implementation (Nadendla et al. 2022).

5.6 Technical considerations and best practices

Data Management and Access: Workshop discussions revealed significant challenges in making data resources available to AI tools while managing system impact. Participants reported issues with ill-behaved bots ignoring crawling rules and overwhelming servers. Recommended solutions include:

- Implementing better licensing frameworks
- Providing clear sitemaps and crawling guidelines on DB homepage and in robots.txt.
- Using tools like AWS WAF for bot management
- Converting structured database data into general English text format suitable for language model training
- Provisioning of AI ready datasets on platforms like Hugging Face and following metadata standards like Croissant.

5.7 Model selection and implementation

The community identified key considerations for AI tool selection:

Resource Requirements: LLMs are computationally intensive and costly to deploy, yet they often yield superior performance on complex tasks compared to traditional methods. Recent developments in agentic applications enable a single base model to be shared across multiple use cases, with each instance assuming a distinct agentic role through prompt engineering rather than model retraining or fine-tuning. This approach enhances flexibility while reducing the computational and financial burden associated with maintaining multiple specialized models. At the same time, there is a growing trend toward developing smaller, domain-specific language models that are highly optimized for particular tasks. Such targeted models may, in many cases, offer a more efficient and practical alternative to large, general-purpose foundation models.
Task-Specific Optimization: the choice between LLMs and simpler methods depends on the specific task and available resources. Not all curation tasks require the most sophisticated AI approaches.
Feedback Mechanisms: implementing systems to incorporate user corrections and improve AI outputs over time is crucial for long-term success.
Documentation Requirements: teams should document curation processes and decision-making steps to create better training data for AI models.

5.8 Quality control and validation

Participants emphasized the critical importance of maintaining quality standards:

- Implementing confidence scoring for AI predictions
- Developing systematic validation workflows
- Creating feedback loops for continuous improvement
- Balancing false positive and false negative trade-offs based on use case requirements

Ensuring data integrity through proposed stages including: exploration; validation; integration; and monitoring. It is imperative to clearly mark ML/AI generated data and to not mix the curated and the AI data within resources and datasets, as exemplified by the AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/) which hosts the 3D structure of a protein predicted through DeepMind from the protein's amino acid sequence.

6 Conclusion: future directions and sustainability

The workshop highlighted the expanding role of AI in biocuration; identified areas of concern regarding data quality and reproducibility; and focused considerations of the future of AI in biocuration. Discussions illuminated the importance of cross-disciplinary collaborations for developing, testing and implementing AI tooling in the biocuration workflow. Significantly, there was unanimous support for community-level sharing of AI usage experiences, developing AI-ready datasets and taking the next step towards democratizing access to AI tools, such as google colab and AI classifier builder. Utilisation of AI for developing code was a common theme whereas the number and types of tasks which could be augmented by AI, e.g., addressing curation bottlenecks and enhancing data quality, continue to be explored and expanded upon. The overall sentiment at the conclusion of the meeting was positive, encouraging our community to explore and share other opportunities to utilise AI in biocuration.

6.1 Environmental and economic considerations

Workshop participants acknowledged the need to consider environmental costs when implementing AI and ML processes. This includes evaluating the carbon footprint of large-scale AI operations and implementing sustainable computing practices. Economic sustainability remains a key concern, with participants noting that while AI can provide significant efficiency gains, the startup costs and ongoing maintenance requirements must be carefully balanced against available resources.

6.2 Evolving roles and skills

The integration of AI in biocuration is fundamentally changing the skill requirements and career paths in the field. It became evident that the use of AI in biocuration is creating new opportunities for future biocurators skills develop in, for example: AI tool evaluation and selection; prompt engineering and optimization; quality validation and error detection; cross-disciplinary collaboration; and technology assessment and implementation.

6.3 Community building and knowledge sharing

The workshop highlighted the importance of community-driven approaches to AI adoption. Key initiatives for the future include: centralized resource sharing through ISB platforms; regular knowledge-sharing events and workshops; collaborative tool development and testing; and cross-institutional partnerships and consortia.

Acknowledgements

We thank the International Society for Biocuration (https://www.biocuration.org/) for the opportunity to host this workshop in 2025, the workshop attendees for their thoughtful discussion and Marcela Karey Tello-Ruiz, Melanie Vollmar, Gerard Lazo, Claudia M. Sanchez-Beato Johnson, Leonore Reisner, Kimberly Van Auken, Daniela Raciti, Raja Mazumder, Jade Hotchkiss for their insightful comments that improved this manuscript.

Author contributions

Valerio Arnaboldi (Conceptualization [Equal], Writing—original draft [Equal], Writing—review & editing [Equal]), Lynn Marie Schriml(Conceptualization [Equal], Writing—original draft [Equal], Writing—review & editing [Equal]), Matt Jeffryes (Conceptualization [Equal], Writing—original draft [Equal], Writing—review & editing [Equal]), Pengyuan Li (Conceptualization [Equal], Writing—original draft [Equal], Writing—review & editing [Equal]), Susan M Bello (Funding acquisition [Equal], Writing—review & editing [Equal]), Paul Sternberg (Funding acquisition [Equal], Writing—review & editing [Equal]), and Carol Bult (Funding acquisition [Equal], Writing—review & editing [Equal]), Charles E Cook (Conceptualization [Equal], Funding acquisition [Equal], Writing—original draft [Equal], Writing—review & editing [Equal])

Conflicts of interest

None declared.

Funding

This work was supported by: the National Institutes of Health–National Human Genome Research Institute (NHGRI) [grant number 1U24HG012557-01]; ARISE1 Fellowships from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreements [agreement number 945405; the National Institutes of Health–National Human Genome Research Institute (NHGRI) [grant numbers U24HG010859, U24HG002223].

Data availability

No data to make available

References

Alliance of Genome Resources Consortium. Updates to the alliance of genome resources Central infrastructure. *Genetics* 2024;**227**: iyae049. https://doi.org/10.1093/genetics/iyae049

Amann J, Blasimme A, Vayena E, *et al.* Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak* 2020;**20**:310. https://doi.org/10.1186/s12911-020-01332-6

Anderson WP, Global Life Science Data Resources Working Group. Data management: a global coalition to sustain core data. *Nature* 2017;**543**:179. https://doi.org/10.1038/543179a

Atkinson CF *et al.* AI-pocalypse now: automating the systematic literature review process. *Comput Biol Med* 2024;**123129**.

Bateman A. Amos bairoch (1957–2025): pioneer of bioinformatics and founder of Swiss-Prot. *Bioinformatics Adv* 2026;**6**: vbag009. https://doi.org/10.1093/bioadv/vbag009.

Boman M. Human-curated validation of machine learning algorithms for health data. *Digital Soc* 2023;**2**:46. https://doi.org/10.1007/s44206-023-00076-w

Borlund P. The concept of relevance in information retrieval. *J Am Soc Inf Sci Technol* 2003;**54**:913–25. https://doi.org/10.1002/asi.10286

Boyacı T, Canyakmaz C, de Véricourt F. Human and machine: the impact of machine input on decision making under cognitive limitations. *Manage Sci* 2024;**70**:1258–75. https://doi.org/10.1287/mnsc.2023.4744

Budhkar A, Song Q, Su J *et al.* Demystifying the black box: a survey on explainable artificial intelligence (XAI) in bioinformatics. *Comput Struct Biotechnol J* 2025;**27**:346–59. https://doi.org/10.1016/j.csbj.2024.12.027

Bult CJ, Sternberg PW. The alliance of genome resources: transforming comparative genomics. *Mamm Genome* 2023;**34**: 531–44. doi: https://doi.org/10.1007/s00335-023-10015-2

Burge S, Attwood TK, Bateman A *et al.* Biocurators and biocuration: surveying the 21st century challenges. *Database (Oxford)*. 2012; **2012**:bar059. https://doi.org/10.1093/database/bar059

Brown TB, Mann B, Ryder N *et al.* Language models are Few-Shot learners. *Adv Neural Inf Process Syst* 2020;**33**:1877–901.

Caspi R, Karp PD. An evaluation of ChatGPT and bard (gemini) in the context of biological knowledge retrieval. *Access Microbiol* 2024;**6**:000790.v3. https://doi.org/10.1099/acmi.0.000790.v3

Caufield H, Kroll C, O'Neil ST29 Oct *et al.* 2024. CurateGPT: A flexible language-model assisted biocuration tool. arXiv [preprint] arXiv : 2411.00046. https://doi.org/10.48550/arXiv.2411.00046

Chatr-Aryamontri A, Winter A, Perfetto L *et al.* Benchmarking of the 2010 BioCreative challenge III text-mining competition by the BioGRID and MINT interaction databases. *BMC Bioinformatics* 2011;**12 Suppl 8**:S8. https://doi.org/10.1186/1471-2105-12-S8-S8

Chen M, Wang Y, Wang Q *et al.* Impact of human and artificial intelligence collaboration on workload reduction in medical image interpretation. *NPJ Digit Med* 2024;**7**:349. https://doi.org/10.1038/s41746-024-01328-w

Chowdhery A, Narang S, Devlin J *et al.* PaLM: scaling language modeling with pathways. *J Mach Learn Res* 2023;**24**:1–113.

D'Amour A, Heller K, Moldovan D *et al.* Underspecification presents challenges for credibility in modern machine learning. *J Mach Learn Res* 2022;**23**:1–42. https://doi.org/10.5555/3586589.3586815

Devlin J, Chang M-W, Lee K *et al.* 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, p. 4171–4186. Minneapolis, Minnesota. Association for Computational Linguistics.

Dobbie S, Strafford H, Pickrell WO *et al.* Markup: a web-based annotation tool powered by active learning. *Front Digit Health* 2021;**3**: 598916. https://doi.org/10.3389/fdgth.2021.598916

Dowell KG, McAndrews-Hill MS, Hill DP, Drabkin HJ, Blake JA. Integrating text mining into the MGI biocuration workflow. *Database (Oxford).* 2009;**2009**:bap019. https://doi.org/10.1093/database/bap019Epub 2009 Nov 21.

Elliott J, Heesterbeek S, Lukensmeyer CJ *et al.* 2005. Steyaert, Stef; Lisoir, Hervé (eds.). *Participatory methods toolkit: a practitioner's manual*. [Brussels]: King Baudouin Foundation/Flemish Institute for Science and Technology Assessment. pp. 185ff. ISBN 978-90-5130-506-7.

Fang R, Schindelman G, Van Auken K *et al.* Automatic categorization of diverse experimental information in the bioscience literature. *BMC Bioinformatics* 2012;**13**:16. https://doi.org/10.1186/1471-2105-13-16

Fralick M, Sacks CA, Muller D *et al.* Large language models. *NEJM Evid* 2023;**2**:EVIDstat2300128. https://doi.org/10.1056/EVIDstat2300128

Gillespie M, Li NT, Beavers D *et al.* AI-augmented human biocuration in reactome [version 1; not peer reviewed]. *F1000Res* 2025; **14**:378. https://doi.org/10.7490/f1000research.1120139.1.

González M, Durán RE, Seeger M *et al.* Negative dataset selection impacts machine learning-based predictors for multiple bacterial species promoters. *Bioinformatics* 2025;**41**:btaf135. https://doi.org/10.1093/bioinformatics/btaf135

Green A, Ribas CE, Ontiveros-Palacios N *et al.* LitSumm: large language models for literature summarization of noncoding RNAs. *Database (Oxford)* 2025;**2025**;baaf006. https://doi.org/10.1093/database/baaf006

Groß A, Hartung M, Prüfer K *et al.* Impact of ontology evolution on functional analyses. *Bioinformatics* 2012;**28**:2671–7. https://doi.org/10.1093/bioinformatics/bts490

Gu Y, Tinn R, Cheng H *et al.* Domain-specific language model pretraining for biomedical natural language processing. *ACM Trans Comput Healthcare* 2022;**3**:1–23. https://doi.org/10.1145/3458754

Hirschman L, Burns GA, Krallinger M, Arighi C, Cohen KB, Valencia A, Wu CH, Chatr-Aryamontri A, Dowell KG, Huala E, Lourenço A, Nash R, Veuthey AL, Wiegers T, Winter AG. Text mining for the biocuration workflow. *Database (Oxford).* 2012; **2012**:bas020. https://doi.org/10.1093/database/bas020

Hochreiter S, Schmidhuber J. Long Short-Term memory. *Neural Comput* 1997;**9**:1735–80. https://doi.org/https://doi.org/10.1162/neco.1997.9.8.1735.

Holinski A, Burke ML, Morgan SL, McQuilton P, Palagi PM. Biocuration - mapping resources and needs. *F1000Res*. 2020;**9**: ELIXIR-1094. https://doi.org/10.12688/f1000research.25413.2

Howe DG, Frazer K, Fashena D *et al.* Data extraction, transformation, and dissemination through ZFIN. *Methods Cell Biol* 2011; **104**:311–25. https://doi.org/10.1016/B978-0-12-374814-0.00017-3

Huang K, Zhang S, Wang H *et al.* Biomni: A General-Purpose Biomedical AI Agent. bioRxiv [Preprint]. 2025. Jun 2: 2025.05.30.656746. https://doi.org/10.1101/2025.05.30.656746

Ibrahim H, Liu F, Mengal K *et al.* Causal evidence of racial and institutional biases in accessing paywalled articles and scientific data. arXiv.[preprint] 2025. Sep 10; 2509.08299. Available from: https://arxiv.org/abs/2509.08299.

International Society for Biocuration. Biocuration: distilling data into knowledge. *PLoS Biol* 2018;**16**:e2002846. https://doi.org/10.1371/journal.pbio.2002846

Islamaj R, Kwon D, Kim S *et al.* TeamTat: a collaborative text annotation tool. *Nucleic Acids Res* 2020;**48**:W5–W11. https://doi.org/10.1093/nar/gkaa333

Islamaj R, Wei CH, Lai PT *et al.* Assessing artificial intelligence (AI) implementation for assisting gene linking (at the national library of medicine). *JAMIA Open* 2025;**8**:ooae129. https://doi.org/10.1093/jamiaopen/ooae129

Jiang C, Qi B, Hong X *et al.* On large language models' hallucination with regard to known facts. *Proc Conf North Am Chapter Assoc Comput Linguist Hum Lang Technol* 2024;**1**:1041–53. https://doi.org/10.18653/v1/2024.naacl-long.60

Joachimiak MP, Caufield JH, Harris NL *et al.* Gene Set Summarization Using Large Language Models. ArXiv [preprint]. 2024. Jul 4: arXiv : 2305.13338v3.

Kahsay R, Bhuiyan U, Au CCH *et al.* GlycoSiteMiner: an ML/AI-assisted literature mining-based pipeline for extracting glycosylation sites from PubMed abstracts. *Glycobiology* 2025;**35**:cwaf030. https://doi.org/10.1093/glycob/cwaf030

Karunanayake N, Silva M, Tanaka H. Next-generation agentic AI for transforming healthcare. *Inform Health* 2025;**2**:73–83. https://doi.org/10.1016/j.slsci.2025.100014

Knoth N, Tolzin A, Janson A *et al.* AI literacy and its implications for prompt engineering strategies. *Comput Educ Artif Intel* 2024;**6**:100225. https://doi.org/10.1016/j.caeai.2024.100225

König C, Shaim I, Vellido A *et al.* Using machine learning tools for protein database biocuration assistance. *Sci Rep* 2018;**8**: 10148. https://doi.org/10.1038/s41598-018-28330-z

Kwon D. Web-scraping AI bots cause disruption for scientific databases and journals. *Nature* 2025;**642**:281–2. https://doi.org/10.1038/d41586-025-01661-4

Lee J, Yoon W, Kim S *et al.* BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 2020;**36**:1234–40. https://doi.org/10.1093/bioinformatics/btz682

Leitner F, Mardis SA, Krallinger M *et al.* An overview of BioCreative II.5. *IEEE/ACM Trans Comput Biol Bioinform* 2010;**7**: 385–99. https://doi.org/10.1109/tcbb.2010.61

Lewis P, Perez E, Piktus A *et al.* Retrieval-Augmented generation for Knowledge-Intensive NLP tasks. *Adv Neural Inf Process Syst* 2020; **33**:9459–74. https://doi.org/10.5555/3495724.3496517

Li Q, Hu Z, Wang Y *et al.* Progress and opportunities of foundation models in bioinformatics. *Brief Bioinform* 2024; **25**:bbae548.

Li P, Kishore R, Arnaboldi V *et al.* GeneScribe: leveraging large language models for gene summary generation [version 1; not peer reviewed]. *F1000Res* 2025;**14**:429. https://doi.org/10.7490/f1000research.1120159.1

Li Z, Liu F, Yang W *et al.* A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Trans Neural Netw Learn Syst* 2022;**33**:6999–7019. https://doi.org/10.1109/TNNLS.2021.3084827

Lin Z, Guan S, Zhang W *et al.* Towards trustworthy LLMs: a review on debiasing and dehallucinating in large language models. *Artif Intell Rev* 2024;**57**:243. https://doi.org/10.1007/s10462-024-10896-y.

Liu W, Laulederkind SJ, Hayman GT *et al.* OntoMate: a text-mining tool aiding curation at the rat genome database. *Database (Oxford)* 2015;**2015**:bau129. doi : https://doi.org/10.1093/database/bau129

Luo R, Sun L, Xia Y *et al.* BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Brief Bioinform* 2022;**23**:bbac409. https://doi.org/10.1093/bib/bbac409

Luo Y, Zhang J, Fan S *et al.* BioMedGPT: an open multimodal large language model for BioMedicine. *IEEE J Biomed Health Inform* 2026;**30**;981–92. https://doi.org/10.1109/JBHI.2024.3505955

Mahadevkar SV, Patil S, Kotecha K *et al.* Exploring AI-driven approaches for unstructured document analysis and future horizons. *J Big Data* 2024;**11**:21.

Nadendla S, Jackson R, Munro J *et al.* ECO: the evidence and conclusion ontology, an update for 2022. *Nucleic Acids Res* 2022;**50**:D1515–21. https://doi.org/10.1093/nar/gkab1025

Neves M, Ševa J. An extensive review of tools for manual annotation of documents. *Brief Bioinform* 2021;**22**:146–63. https://doi.org/10.1093/bib/bbz130

Nielsen SS, Ostaszewski M, McGee F *et al.* Machine learning to support the presentation of complex pathway graphs. *IEEE/ACM Trans Comput Biol Bioinform* 2021;**18**:1130–41. https://doi.org/10.1109/TCBB.2019.2938501

Nijsten H, Geyer H, Geyer R. Ontology-based annotation of glycan structures: challenges and opportunities. *Bioinformatics* 2016;**32**:3585–91. https://doi.org/10.1093/bioinformatics/btw380

Niyonkuru E, Caufield JH, Carmody LC *et al.* Leveraging generative AI to assist biocuration of medical actions for rare diseases. *Bioinformatics Adv* 2025;**5**:vbaf141. https://doi.org/10.1093/bioadv/vbaf141

O'Brien T, Stremmel J, Pio-Lopez L *et al.* Machine learning for hypothesis generation in biology and medicine: exploring the latent space of neuroscience and developmental bioelectricity. *Digital Discov* 2024;**3**:249–63. https://doi.org/10.1039/D3DD00185G

Perlman-Arrow S, Loo N, Bobrovitz N *et al.* A real-world evaluation of the implementation of NLP technology in abstract screening of a systematic review. *Res Synth Methods* 2023;**14**:608–21. https://doi.org/10.1002/jrsm.1636

Poretsky E, Blake VC, Andorf CM *et al.* Assessing the performance of generative artificial intelligence in retrieving information against manually curated genetic and genomic data. *Database (Oxford)* 2025;**2025**;baaf011. https://doi.org/10.1093/database/baaf011

Raciti D, Van Auken KM, Arnaboldi V *et al.* Characterization and automated classification of sentences in the biomedical literature: a case study for biocuration of gene expression and protein kinase activity. *Database (Oxford)* 2025;**2025**:baaf063. https://doi.org/10.1093/database/baaf063

Russell SJ, Norvig P. Artificial intelligence: a modern approach. 4th ed. Pearson; 2020.

Sansone SA, McQuilton P, Rocca-Serra P *et al.* FAIRsharing as a community approach to standards, repositories and policies. *Nat Biotechnol*. 2019. **37**(4):358–367. https://doi.org/10.1038/s41587-019-0080-8

Sapkota R, Nguyen T, Patel S. AI agents vs. Agentic AI: a conceptual taxonomy. *Artif Intell* 2025;**250**:671–88. https://doi.org/10.1016/j.artint.2025.103045

Sarumi OA, Heider D. Large language models and their applications in bioinformatics. *Comput Struct Biotechnol J* 2024;**23**:3498–505. https://doi.org/10.1016/j.csbj.2024.09.031

Schriml LM, Munro JB, Schor M *et al.* The human disease ontology 2022 update. *Nucleic Acids Res* 2022;**50**:D1255–61. https://doi.org/10.1093/nar/gkab1063

Skrzypek MS, Nash RS. Biocuration at the saccharomyces genome database. *Genesis* 2015;**53**:450–7. https://doi.org/10.1002/dvg.22862

Solovieva E, Shikanai T, Fujita N *et al.* GGDonto ontology as a knowledge-base for genetic diseases and disorders of glycan metabolism and their causative genes. *J Biomed Semantics* 2018;**9**:14. https://doi.org/10.1186/s13326-018-0182-0

Sundaram SS, Musen MA. Toward Total Recall: Enhancing FAIRness through AI-Driven Metadata Standardization. *GigaScience* 2026; giag019. https://doi.org/10.1093/gigascience/giag019

Telenti A, Auli M, Hie BL *et al.* Large language models for science and medicine. *Eur J Clin Invest* 2024;**54**:e14183. https://doi.org/10.1111/eci.14183

Toro S, Anagnostopoulos AV, Bello SM *et al.* Dynamic retrieval augmented generation of ontologies using artificial intelligence (DRAGON-AI). *J Biomed Semantics* 2024;**15**:19. https://doi.org/10.1186/s13326-024-00320-3

Touvron H, Lavril T, Izacard G, *et al.* LLaMA: open and efficient foundation language models. arXiv preprint2023. preprint: not peer reviewed. arXiv : 2302.13971. https://doi.org/10.48550/arXiv.2302.13971

Tsaneva S, Sabou M. Enhancing human-in-the-loop ontology curation results through task design. *J Data and Information Quality* 2024;2024;**16**:1–25. https://doi.org/10.1145/3626960.

Van Auken K, Fey P, Berardini TZ, *et al.* Text mining in the biocuration workflow: applications for literature curation at WormBase, dictyBase and TAIR. *Database (Oxford)* 2012;**2012**:bas040. https://doi.org/10.1093/database/bas040.

Wang JJ, Wang VX. Assessing consistency and reproducibility in the outputs of large language models: evidence across diverse finance and accounting tasks. arXiv [Preprint]. 2025. Mar 21; arXiv : 2503.16974. https://doi.org/10.48550/arXiv.2503.16974

Wang Z, Wang H, Danek B *et al.* A perspective for adapting generalist AI to specialized medical AI applications and their challenges. *NPJ Digit Med* 2025;**8**:429. https://doi.org/10.1038/s41746-025-01789-7

Wei CH, Allot A, Lai PT *et al.* PubTator 3.0: an AI-powered literature resource for unlocking biomedical knowledge. *Nucleic Acids Res* 2024;**52**:W540–46. https://doi.org/10.1093/nar/gkae235

Wiegers TC, Davis AP, Wiegers J, Sciaky D, Barkalow F, Wyatt B, Strong M, McMorran R, Abrar S, Mattingly CJ. Integrating AI-powered text mining from PubTator into the manual curation workflow at the comparative toxicogenomics database. *Database (Oxford)*. 2025;**2025**:baaf013. https://doi.org/10.1093/database/baaf013

Wilkinson MD, Dumontier M, Aalbersberg IJ *et al.* The FAIR guiding principles for scientific data management and

stewardship. *Sci Data* 2016;**3**:160018. https://doi.org/10.1038/sdata.2016.18

Xie Q, Chen Q, Chen A *et al.* Medical foundation large language models for comprehensive text analysis and beyond. *NPJ Digit Med* 2025;**8**:141. https://doi.org/10.1038/s41746-025-01533-1

Zhang G, Jin Q, Zhou Y *et al.* Closing the gap between open source and commercial large language models for medical evidence summarization. *NPJ Digit Med* 2024;**7**:239. https://doi.org/10.1038/s41746-024-01239-w

Zheng Y, Yan Y, Chen S *et al.* Integrating retrieval-augmented generation for enhanced personalized physician recommendations in web-based medical services: model development study. *Front Public Health* 2025;**13**:1501408. https://doi.org/10.3389/fpubh.2025.1501408