Enhancing Scientific Heritage Access Through AI-Powered Assistants and Collaboration Platforms in Academic Ecosystems

Evgeniya Georgieva[1[0000-0001-7785-5902]], Tsvetozar Georgiev[1[0000-0003-3548-0166]], Tsvetelina Stefanova[1[0000-0001-8079-2965]]

[1] University of Ruse Angel Kanchev, 8 Studentska Street, 7017 Ruse, Bulgaria
esgeorgieva@uni-ruse.bg, tgeorgiev@ecs.uni-ruse.bg, tsstefanova@uni-ruse.bg

Abstract. This study investigates how artificial intelligence and collaboration platforms enhance scientific heritage access within academic ecosystems. Within the context of evolving infrastructures for scientific discovery and heritage preservation, the analysis examines advanced digitization methods, semantic networks, autonomous research agents, and mainstream database AI integrations. Findings indicate that while these technologies democratize knowledge, institutions must proactively address inherent ethical, legal, and cybersecurity vulnerabilities.

Keywords: Digital Scientific Heritage, Artificial Intelligence, AI-Powered Research Assistants, Academic Ecosystems, Knowledge Graphs, Decentralized Science.

1 Introduction

The concept of scientific heritage has significantly expanded in the contemporary digital age. Moving beyond the preservation of rare physical manuscripts and printed archival collections, the modern definition fundamentally encompasses an exponentially growing volume of born-digital scholarly outputs (Lian & Xie, 2024). This includes peer-reviewed journal articles, complex multimodal datasets, and open-source software repositories. This ontological transformation aligns with the principles of long-term digital preservation articulated in the Open Archival Information System (OAIS) reference model, which structures preservation not merely as the storage of static files, but as an integrated, continuously migrating ecosystem of metadata, digital infrastructure, and designated user communities (Consultative Committee for Space Data Systems, 2024).

The global academic and scientific research ecosystem is navigating an unprecedented structural transformation. Higher education and research institutions are operating within an environment characterized by severe economic pressures, distinct demographic shifts, and the rapid, pervasive maturation of computational technologies. In

Digital Presentation and Preservation of Cultural and Scientific Heritage. Volume 16, pp. 195-210, 2026.
DOI: 10.55630/dipp.2026.16.16, ISSN 1314-4006, eISSN 2535-0366.

response, the adoption of advanced technologies has shifted from an area of experimental novelty to an absolute operational baseline. Recent empirical surveys indicate that 83% of academic institutions have integrated or are strategically planning to expand their deployment of artificial intelligence to optimize institutional workflows and learning environments (Ellucian, 2026).

Beyond the administrative sphere, the core of scientific discovery and heritage preservation is being revolutionized by the sheer volume of global research production. Major indexing systems, while providing structured access to vast citation networks, present severe information overload. Modern researchers face severe time constraints and increasing difficulty in conducting comprehensive, bias-free literature reviews. Fragmented knowledge across highly specialized, siloed disciplines complicates the synthesis of evidence, reproducibility, and interdisciplinary reasoning.

This paper aims to comprehensively investigate how the integration of artificial intelligence into documentary heritage, the deployment of specialized AI-powered research assistants within major scholarly databases, the utilization of automated collaboration platforms, and the establishment of decentralized global data infrastructures promise to unlock centuries of fragmented knowledge. By tracing these technological advancements, the exposition will demonstrate how the global academic ecosystem is transitioning toward a model of human-AI co-creation. Specifically, the study's core contributions include a comprehensive analysis of this shift, a comparative synthesis of current platforms and tools, and the proposal of a layered architectural model.

2 Evolving Infrastructures for Scientific Discovery and Heritage Preservation

2.1 AI-Driven Digitization and Semantic Structuring of Archival Heritage

The preservation of scientific heritage is essentially a process of transgenerational communication, ensuring that the empirical observations of past centuries remain legible to future scholars. Digitization is not only an effective way to safeguard academic knowledge, but it also provides new opportunities for its development and dissemination (Lian & Xie, 2024).

Overcoming Archival Bottlenecks. The immense physical volume and linguistic complexity of historical documents have perennially presented a massive bottleneck in terms of access, preservation and research. The manual transcription of primary source documents is highly labor-intensive, requiring specialized paleographic expertise. However, specialized artificial intelligence platforms have emerged as gold standards for digital preservation. Transkribus represents a highly specialized approach utilizing Handwritten Text Recognition (HTR) based on deep neural networks. Designed explicitly for centuries-old texts and medieval manuscripts, it drastically reduces the reliance on manual human transcription and makes centuries of primary sources computationally searchable at scale (Nockels et al., 2022).

Digitizing Complex STEM Heritage. The digitization of quantitative and structural scientific heritage presents an entirely different magnitude of difficulty. Historical scientific archives are replete with complex mathematical formulas, physical notations, tabular data, and chemical structures. To bridge this specific gap, specialized computer vision tools have been engineered to convert complex technical imagery into highly structured markup languages (e.g., LaTeX and MathML). Tools such as Mathpix and iWeaver AI utilize layout-aware computer vision and active learning strategies to accurately extract complex equations and multi-axis tabular data, preserving the spatial and structural relationships inherent in STEM documentation (Wei et al., 2025).

Table 1 summarizes and evaluates primary digitization platforms utilized in academic preservation based on their technical architecture and core functional benefits.

Table 1. Core AI Digitization Software for Archival and STEM Heritage.

Digitization Domain	Primary Software	Technical Architecture	Core Functional Benefit
Historical Manuscripts	Transkribus	Vision Transformers (TrOCR), Deep HTR	High-accuracy transcription of cursive and degraded historical archives.
Mathematical Formulas	Mathpix	Syntax-constrained OCR, Layout parsing	High-fidelity conversion of spatial STEM equations to LaTeX/MathML.
Complex Tabular Data	iWeaver AI	Markdown conversion models	Automated extraction of nested PDF tables into computable data structures.
Live Math Authoring	Wiris MathType	Adaptive Handwriting Recognition	Real-time translation of stylus input into digital math expressions.
Digital Preservation	Preservica	SaaS OAIS Policy Automation	Active format migration ensuring long-term access to digital heritage.

As demonstrated in Table 1, the current landscape of digital preservation has advanced beyond generalized optical character recognition. The evolution of these platforms indicates a shift toward highly specialized, domain-specific AI models – such as layout-aware parsing for STEM and vision transformers for historical manuscripts. This technological specialization provides the necessary infrastructure for institutions to overcome traditional archival bottlenecks and process complex, heterogeneous heritage at scale. However, this large-scale digitization is only the foundational step in "activating" archival heritage. True activation goes beyond simple physical preservation; it is the transformation of these static holdings into dynamic, accessible, and interactive resources. Consequently, this transition must be carried out through two complementary directions: the technological advancements in digitization just described, and the subsequent implementation of structural-logical frameworks, such as semantic architectures.

Constructing Cognitive Heritage Digital Twins. As digital preservation efforts scale, methodologies are shifting toward the construction of Knowledge Graphs (KGs). KGs provide a foundational semantic architecture that addresses data heterogeneity and enables cross-institutional knowledge integration (Zhu et al., 2026). The integration of AI and semantic KGs culminates in the creation of Cognitive Heritage Digital Twins (CHDTs). Supported by architectures like the Reactive Heritage Digital Twin Ontology (RHDTO), Natural Language Processing (NLP) extracts structured knowledge from unstructured heritage reports to populate the digital twin's semantic graph, allowing the digital twin to answer complex scientific queries in natural language and generate predictive artifact conservation models (Felicetti et al., 2025).

Adopting AI Technologies in Libraries for Scientific Heritage Preservation and Accessibility. Nowadays, more and more libraries digitize their collections and successfully integrate AI technologies in their workflow and services. Recent empirical case studies highlight the successful integration of generative AI across modern academic libraries. For example, a 2025 study of a university library demonstrated how deploying an AI-powered chatbot successfully transformed the institution from a passive information provider into a proactive knowledge facilitator, significantly enhancing service personalization and research efficiency (Kim, 2025). Furthermore, broader infrastructural reviews indicate that academic libraries are actively adopting AI for automated cataloging, metadata generation, and predictive analytics to streamline collection management and optimize service delivery.

2.2 AI-Augmented Information Retrieval and Synthesis in Scientific Database Platforms

To combat profound information overload, academic publishers and specialized tech companies have fundamentally overhauled the literature review process. Moving away from traditional Boolean-based keyword matching, the academic search paradigm has shifted toward Retrieval-Augmented Generation (RAG) architectures. These architectures anchor generative AI outputs directly to verified, peer-reviewed corpora, mitigating algorithmic hallucinations while synthesizing immense datasets.

Leading scientific information providers have actively deployed proprietary AI research assistants in their platforms to augment the access to their vast databases providing algorithmically generated author keywords and topic suggestions. These AI assistants help researchers navigate large amounts of academic data, retrieve relevant information efficiently, and analyze complex datasets of peer-reviewed sources (Georgieva & Georgiev, 2025). Some of the most used AI tools are:

- **Scopus AI (Elsevier):** Launched to navigate its extensive abstract and citation database, Scopus AI utilizes vector-based search and RAG frameworks to generate Topic Summary and Expanded Summary. It enhances discovery by listing Foundational Documents and generating interactive Concept Maps that visualize the relationships between disparate research areas, aiding interdisciplinary exploration (Scopus, 2026).

- **Web of Science Research Assistant (Clarivate):** Leveraging 120 years of carefully curated Core Collection data, this assistant provides task-based guided walkthroughs (Ask a research question, Explore a topic, Create a Literature Review, Find a Journal). It features intelligent semantic search that maps the query to its proprietary knowledge graph, and provides dynamic visualizations such as co-citation networks and trend graphs to help researchers grasp core concepts rapidly (Web of Science, 2026).
- **JSTOR AI Research Tool:** Tailored for the humanities and social sciences – where abstracts are frequently absent from historical texts – the JSTOR AI tool offers a conversational interface that assesses content relevance by surfacing key points and arguments directly from the full text of journal articles, book chapters, and primary source reports (JSTOR, 2025).
- **EBSCO AI Insights:** EBSCO focuses on reducing friction for early-career researchers by replacing complex Boolean logic with Natural Language Search (NLS). Furthermore, its AI Insights feature generates a concise, 2-to-5 point bulleted summary grounded directly in the full-text article, allowing users to rapidly evaluate methodological relevance before engaging in deep reading (EBSCO, 2026).

Table 2 presents the functional capabilities of these major AI research tools, highlighting the primary data corpus and key technologies that they utilize.

Table 2. Functional Capabilities Comparison of AI Assistants in Scientific Database Platforms.

AI Research Assistant	Developer	Primary Data Corpus	Key Technologies	Unique Features
Scopus AI	Elsevier	The entire Scopus corpus	Natural language processing (NLP), RAG Fusion, Copilot integration	Generates Topic Summary, Expanded Summary, interactive Concept Maps, suggests Go Deeper questions and creates a list of Foundational Documents.
Web of Science Research Assistant	Clarivate	WoS Core Collection	NLP, RAG, Semantic knowledge graph	Task-based guided walkthroughs and dynamic co-citation visualizations, topic maps, trend graphs and more.
JSTOR AI Research Tool	ITHAKA	JSTOR full-text archival corpus	NLP, RAG, OpenAI LLMs integration	Suggests the top 5 documents related to the selected text or the key points of the selected text; conversational document search and interaction.

EBSCO AI Insights	EBSCO Information Services	EBSCO full-text databases	NLP, RAG, Claude Haiku integration	Provides AI Insights (2-5 bulleted key takeaways) for each document and Natural Language Search (NLS).

The comparative analysis in Table 2 reveals that major academic publishers are increasingly adopting Retrieval-Augmented Generation (RAG) architectures to ground AI outputs in their proprietary corpora. Consequently, the primary differentiator among these platforms lies in their capacity to generate subject-specific contextual insights – such as interactive concept maps for interdisciplinary exploration or structured methodological summaries for literature reviews.

2.3 AI-powered Collaboration Research Platforms and Evidence Synthesis Search Engines

A collaboration platform is a digital, often SaaS-based, workspace that allows internal and external stakeholders to communicate, share information, and manage projects in real time, regardless of location. These platforms boost productivity, streamline workflows, and centralize documentation.

Beyond legacy databases, specialized AI-native platforms have emerged to automate deeper synthesis tasks. Platforms like **Consensus** utilize a "Consensus Meter" to visually quantify scientific agreement across its database of 200 million papers, answering yes/no research questions by aggregating the findings of the broader scientific community. **SciSpace** functions as a comprehensive workspace, allowing researchers to generate highly customizable "Literature Review Matrices" that automatically extract specific methodological variables (e.g., sample sizes, limitations) from dozens of papers simultaneously, drastically accelerating data extraction (Jain et al., 2023).

Contemporary scientific discovery is predominantly team-based and highly interdisciplinary (Wuchty et al., 2007). The increasing dominance of multi-authored publications requires digital infrastructures capable of synchronizing workflows and ensuring methodological reproducibility across distributed global institutions.

Covidence stands as the industry-standard software-as-a-service (SaaS) platform tailored exclusively to streamline the systematic review process. Aligned with rigorous methodological frameworks such as the PRISMA guidelines, Covidence digitizes, centralizes, and partially automates the multi-phase workflow of evidence synthesis.

The platform eliminates severe logistical friction by automatically deduplicating thousands of imported citations. It facilitates single or dual-blinded screening, tracking the decisions of multiple distributed reviewers and providing a centralized dashboard for rapid conflict resolution. Furthermore, Covidence integrates machine learning enhancements, such as the Cochrane Randomized Controlled Trial (RCT) classifier, allowing researchers to rapidly filter datasets based on study design. By automating repetitive tasks and streamlining quality assessment frameworks (e.g., Cochrane Risk of Bias tool), platforms like Covidence have been empirically shown to reduce the total time spent on a systematic review by an average of 35% (Covidence, 2026).

The collaborative platforms serve as integrated workspaces that facilitate the research workflow by allowing teams to manage shared libraries, conduct systematic reviews, and co-author documents in a centralized, real-time environment. In contrast academic search engines are specialized discovery tools designed to retrieve, filter, and synthesize information from vast scholarly databases through semantic and citation-aware queries.

Today the lines have blurred as search engines add workspaces and platforms add AI search, but their primary functions – and their user bases – still differ significantly.

Collaborative Platforms / Workspaces. They are used to organize papers, work with teams, and manage the writing/review process. Some of the most popular collaborative platforms and workspaces are:

- **SciSpace**: Currently the most popular "all-in-one" platform. It combines search with a collaborative PDF reader (Copilot), a citation manager, and an AI writing assistant.
- **ResearchRabbit**: Highly popular for visual discovery. Following its 2025 merger with Litmaps, it is the primary platform for researchers to collaboratively map out citation networks and share collections.
- **Covidence**: While it has a smaller user base, it is the most intensely used platform for systematic reviews. It is a structured workflow tool where teams "vote" on which papers to include in a study.
- **LeapSpace**: It is a "Research-Grade Workspace" combining comprehensive trusted, peer-reviewed content from Scopus abstract data, full-text from Elsevier and other leading scientific publishers and societies, with Responsible AI in a single secure environment. The researchers can generate ideas, plan projects, explore literature, find collaborators, and identify funding (Elsevier, 2026b).

Academic Search Engines. They are primarily used to query databases and surface evidence-backed answers. Some of the *most* popular academic search engines are:

- **Dimensions AI**: This tool is most valuable for its comprehensive bibliometric analytics, enabling researchers to map the interconnected landscape of global research through publications, patents, and grants. Its most utilized feature is the integrated data linking, which facilitates large-scale systematic mapping and institutional impact assessment.
- **Consensus**: The most used "natural language" engine. It functions like a search bar where the researcher asks a question and gets a "Yes/Possibly/Mixed/No" consensus based on thousands of papers.
- **Elicit**: This search engine is distinguished by its high-fidelity data extraction capabilities, allowing for the systematic conversion of unstructured full-text PDF libraries into structured synthesis tables. It is most frequently employed for parameter-specific extraction, where users define custom variables – such as methodology, sample size, or effect size – to automate the initial phases of a systematic review.

Figure 1 proposes an original layered architecture for AI-augmented access to Digital Scientific Heritage, positioning artificial intelligence as an intermediary between preserved scholarly infrastructures and contemporary research workflows. At the top of the model, Researchers, PhD Students, and Early-Career Academics interact with the system through semantic search and conversational interfaces, reflecting the transition from rigid keyword-based retrieval to context-aware, dialogue-driven discovery. User intent is directed simultaneously toward two interconnected domains: traditional academic research platforms and AI-enhanced collaboration environments.

On one side, established repositories and indexing systems – such as Scopus, Web of Science, EBSCO, institutional repositories, and open-access archives – provide verified, peer-reviewed content and structured metadata. These sources represent the epistemic foundation of the ecosystem, ensuring that AI-generated outputs remain grounded in authoritative data. In parallel, AI-native discovery and collaboration platforms such as SciSpace, ResearchRabbit, Covidence, LeapSpace, Dimensions AI, Consensus, and Elicit support semantic exploration, automated summarization, and collaborative review processes. Together, these environments form the operational context within which scientific work is organized and executed.

At the core of the architecture are AI-Powered Research Assistants. Using vector databases and dense embeddings, they enable semantic retrieval that captures conceptual similarity. Through Retrieval-Augmented Generation (RAG), the system retrieves relevant document segments before producing summaries, thereby grounding outputs in verifiable evidence and reducing hallucination risks. Citation-aware language models further support context-sensitive recommendations and navigation of citation networks.

The practical impact of this AI integration is realized in digital workspace features, including systematic review workflows, collaborative data sharing, and project management. By automating screening, evidence extraction, and synthesis tasks, the framework reduces repetitive cognitive labor and allows researchers to focus on higher-order analysis and theory building.

The entire ecosystem rests on a dual foundation of interoperable metadata and long-term digital preservation. FAIR and Linked Data principles, persistent identifiers, and structured metadata standards ensure discoverability and machine readability, while the OAIS-based preservation framework guarantees authenticity, format stability, and long-term accessibility. A transversal ethical and regulatory layer, including compliance with instruments such as the EU AI Act and GDPR, emphasizes that AI functions as an assistive, accountable technology under human oversight.

The proposed architecture illustrates an integrated system in which preserved scientific heritage, AI mediation, and collaborative research environments operate as a coherent and sustainable scholarly ecosystem. Rather than operating as isolated utilities, AI-powered assistants and discovery platforms bridge the gap between structured preservation frameworks (such as OAIS and FAIR principles) and the dynamic, highly collaborative requirements of contemporary research workflows.

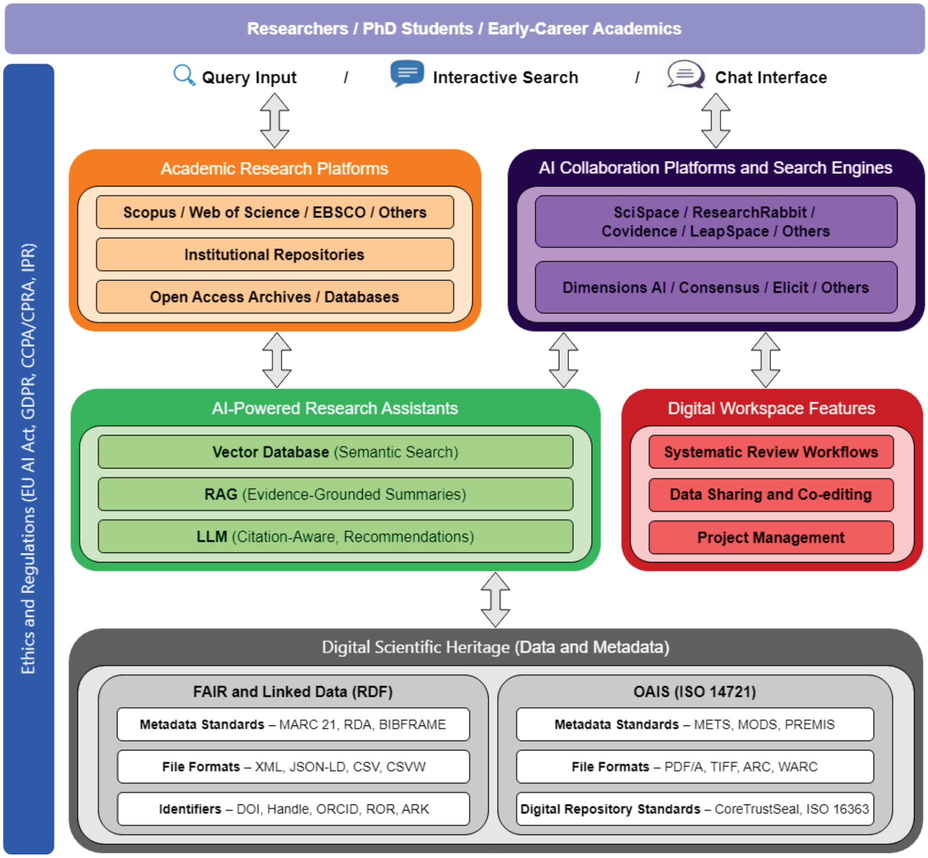

Fig. 1. AI-Augmented Access to Digital Scientific Heritage.

2.4 Agentic Frameworks for Autonomous Scientific Execution

While AI-powered research assistants vastly optimize human-led discovery, the absolute frontier of academic technology is experiencing a paradigm shift toward autonomous AI research agents. These agentic AI systems possess the capacity to autonomously reason, execute multi-step experimental workflows, and iteratively refine hypotheses in continuous closed loops.

The **AgentRxiv** framework introduces a structural solution to the limitations of isolated AI workflows by creating a centralized, online-updating preprint repository designed exclusively for autonomous LLM research agents (Schmidgall & Moor, 2025). Modeled after human preprint servers, AgentRxiv enables distributed "agent laboratories" to autonomously perform research, draft comprehensive academic papers, and upload their findings to a shared server. Crucially, it utilizes Case-Based Reasoning (CBR), allowing agents to query the repository, extract successful reasoning trajectories from other AI systems, and adapt them to solve novel problems. In highly complex mathematical reasoning benchmarks (MATH-500), agents utilizing this collaborative framework achieved a 13.7% relative improvement in overall accuracy compared to isolated laboratories (Schmidgall & Moor, 2025).

In the applied and physical sciences, the integration of autonomous agents with robotic systems is dramatically accelerating experimental timelines. The concept of the "AI Co-Scientist" has been rigorously validated in biomedical domains, autonomously generating and testing novel hypotheses for drug repurposing (Gottweis & Natarajan, 2025). Another example for AI tools successfully implemented in research workflows are frameworks such as the Polybot system at the Argonne National Laboratory which operate as fully AI-guided autonomous laboratories, continuously analyzing the outcomes of previous reactions to direct robotic synthesis in real-time, compressing physical experimental cycles from months to days (Matz, 2026).

Table 3 categorizes the evolution of AI research tools' capabilities.

Table 3. Evolution of AI Capabilities in Scientific Research.

Capability Level	Role of Artificial Intelligence	Sample AI Tools and Platforms
Foundational Data Processing Tools	Transform raw research materials (PDFs, images, scans) into structured, machine-readable data	Mathpix; Scholarcy; GROBID
Database-Integrated AI Assistants	AI embedded in scholarly databases for guided search and synthesis over curated corpora	Scopus AI; Web of Science Research Assistant; EBSCO AI Insights
Open Discovery & Synthesis Tools	Independent AI tools for literature exploration, semantic search, and evidence synthesis	Consensus; ResearchRabbit; Dimensions
AI-Orchestrated Research Workspaces	Multi-step research workflow automation (search → extraction → synthesis → writing)	SciSpace; Elicit; iWeaver; Covidence
Autonomous Execution Systems	AI systems that independently generate hypotheses, run analyses, or conduct experiments	Polybot; Google AI Co-Scientist
Agentic Collaboration Systems	Multi-agent ecosystems enabling collaborative scientific reasoning and knowledge production	AgentRxiv; BioAgents

Table 3 outlines a clear trajectory in technologies used in academic research, illustrating a transition from human-directed data processing to autonomous reasoning. The evolution from database assistants to agentic collaboration systems reflects a broader structural shift, redefining artificial intelligence from a passive operational utility into an active component of the scientific knowledge generation process.

2.5 Decentralized Infrastructures and Interactive Scholarly Communication

As autonomous agents generate novel hypotheses at machine speed, the underlying digital infrastructures must correspondingly evolve.

Decentralized Science (DeSci) and Verifiable Provenance. To ensure that AI outputs remain scientifically trustworthy, the underlying data feeding these models must possess immutable provenance. The **OriginTrail Decentralized Knowledge Graph (DKG)** transforms scattered data into discoverable, highly structured Knowledge Assets cryptographically anchored to distributed ledgers (OriginTrail, 2026). This verifiable infrastructure natively supports the broader movement of Decentralized Science (DeSci). Protocols such as **Molecule** and the **BIO Protocol** pioneer the tokenization of scientific intellectual property into blockchain-based IP-NFTs (Molecule, 2026). Within this ecosystem, validated hypotheses – often generated autonomously by Bio-Agents – can be subjected to decentralized peer review and monetized entirely outside of the legacy pharmaceutical apparatus (Brennan, 2025).

Living Papers. The ultimate output of the research process – the static academic PDF – is functionally obsolete for modern interactive data science. Frameworks such as Living Papers introduce an extensible, open-source language toolkit designed to transform traditional academic publishing into augmented scholarly communication (Heer et al., 2023). Authoring articles using extended Markdown, researchers natively interleave formatted text, interactive components, and executable code (e.g., Python, R). Within this environment, tools like **TableTale** construct document-level linking schemas that provide "cascade visual activation". As a reader scrolls through the narrative text, TableTale dynamically highlights corresponding data cells within complex tables, significantly reducing cognitive workload and improving reading efficiency (Wang et al., 2026).

3 Ethical, Legal and Security Aspects

The deep integration of agentic artificial intelligence and decentralized platforms into the academic ecosystem presents profound ethical, legal, and operational vulnerabilities.

Copyright Law and TDM Exceptions. Training highly capable language models requires the continuous ingestion of vast quantities of copyrighted scientific literature. Globally, jurisdictions are struggling to balance the rights of original creators with the economic imperative for rapid technological advancement (Blaszczyk et al., 2024). The European Union has taken a decisive regulatory lead through the EU AI Act, which provides highly specific Text and Data Mining (TDM) exceptions allowing research organizations to utilize reproductions of copyrighted works for non-commercial scientific research. This contrasts sharply with the fragmented legal landscape in the United Kingdom, severely complicating the deployment of globally federated AI platforms (Westenberger & Farmaki, 2025). To address the opacity of training data, the recently enacted EU AI Act establishes critical transparency requirements alongside the existing EU Copyright Directive regarding Text and Data Mining (TDM). Specifically, Article 53 of the EU AI Act mandates that providers of general-purpose AI models must publicly disclose a sufficiently detailed summary of the content used for training their models. This serves as a vital regulatory mechanism enabling scientific rightsholders to identify potential infringement and protect their intellectual property (IAPP, 2024).

Algorithmic Oversight in Peer Review. As the volume of research scales exponentially, traditional peer review systems are buckling under operational strain. As a representative case study, a landmark initiative is the *AI*-assisted peer review pilot program launched for the **AAAI-26** conference (AAAI, 2025). Confronting nearly 29,000 submissions, the conference integrated a frontier reasoning model into its multi-phase editorial workflow to automate exhaustive literature searches and verify mathematical correctness. Crucially, the framework preserves human epistemic authority: the AI explicitly does *not* replace human reviewers or generate automated accept/reject decisions. Instead, it produces a clearly labeled supplementary technical review, and strictly sequesters anonymous paper PDFs to ensure unpublished intellectual property is not utilized for model training (AAAI, 2025).

AI Authorship and Disclosure Standards. As generative AI assumes a larger role in evidence synthesis, establishing definitive authorship guidelines has become paramount. Major academic publishers, such as Elsevier, have established strict policies permitting the use of AI tools strictly to refine the language and readability of manuscripts prior to submission, provided there is human oversight and transparent disclosure; however, AI systems cannot be credited as authors. Furthermore, publisher policies strictly warn authors and peer reviewers against uploading unpublished manuscripts into external generative AI tools, as this can violate confidentiality agreements, breach data privacy rights, and compromise proprietary research (Elsevier, 2026a).

Cognitive Homogenization and Cybersecurity. A profound risk associated with the pervasive utilization of LLMs is severe cognitive homogenization (Grimmett, 2026). As autonomous agents and researchers utilize the same centralized foundation models, distinct linguistic styles and unique reasoning strategies risk becoming deeply standardized, reducing the creative diversity of the academic community. From a cybersecurity perspective, foundational industry frameworks such as the OWASP Top 10 for LLM Applications highlight “Prompt *Injection*” and “Excessive Agency” as critical vulnerabilities. In complex academic ecosystems, granting AI agents autonomy to invoke functions dynamically creates a severe attack surface. Adversaries can utilize prompt injections – manipulating model behavior through crafted inputs – to hijack an agent's intent. When combined with excessive system permissions, this vulnerability enables malicious actors to bypass built-in safety controls, resulting in outcomes such as the unauthorized exfiltration of proprietary research data or internal credentials (OWASP, 2025). Therefore, to safely integrate AI technologies, it is imperative that academic institutions develop and implement dedicated cybersecurity protocols and governance frameworks specifically designed to monitor and restrict non-human agent cloud permissions.

4 Conclusions

The integration of AI-powered research assistants and collaborative platforms represents a structural transformation in the management and accessibility of digital scientific heritage. The successful deployment of specialized AI-driven digitization tools, such as Transkribus and Mathpix, demonstrates that massive repositories of historical and complex STEM texts can be seamlessly transitioned into dynamic, interactive Semantic Web structures.

Furthermore, the academic library is fundamentally evolving. The deployment of RAG-based AI assistants within mainstream databases (Scopus, Web of Science, JSTOR) and specialized synthesis workspaces (SciSpace, Consensus) helps address the challenge of information overload. By ensuring outputs are grounded in verified sources, these tools uphold academic rigor while significantly reducing the latency of evidence synthesis. Concurrently, platforms like Covidence optimize collaborative systematic review workflows, while AgentRxiv and Decentralized Science (DeSci) protocols highlight the potential viability of autonomous agent laboratories and blockchain-verified provenance.

To realize the full potential of this integrated future, the academic community must proactively navigate the associated ethical labyrinth. To construct a concrete research agenda addressing these gaps, policymakers must prioritize universally standardized governance frameworks for DeSci platforms, researchers should focus on cross-disciplinary efforts to reconcile global TDM copyright laws, and institutional practitioners must ensure the rigorous implementation of autonomous cyber-defenses to secure institutional integrity in an era defined by human-machine co-creation.

Acknowledgements

This paper is supported by project 2026-FEEA-01 "Systematic study of methodological and architectural approaches for modeling digital transformation based on artificial intelligence", funded by the Research Fund of Angel Kanchev University of Ruse, Bulgaria.

References

AAAI. (2025). *AAAI-26 review process update: Scale, integrity measures, and pathways to sustainability*. Association for the Advancement of Artificial Intelligence. https://aaai.org/conference/aaai/aaai-26/review-process-update/

Blaszczyk, M., McGovern, G., & Stanley, K. D. (2024). *Artificial intelligence impacts on copyright law*. RAND Corporation. https://www.rand.org/pubs/perspectives/PEA3243-1.html

Brennan, S. (2025). *BioAgents: Accelerating decentralized science with AI agents*. Bio.xyz. https://www.bio.xyz/blog-posts/bioagents-pioneering-decentralized-science-with-ai-agents

Consultative Committee for Space Data Systems. (2024). *Reference model for an Open Archival Information System (OAIS)* (CCSDS 650.0-M-3). https://ccsds.org/Pubs/650x0m3.pdf

Covidence. (2026). *See your systematic reviews like never before*. https://www.covidence.org/

EBSCO. (2026). *AI insights*. https://connect.ebsco.com/s/article/Generative-AI-Insights-Beta-Summaries?language=en_US

Ellucian. (2026). *Artificial intelligence in higher education: From widespread adoption to strategic integration*. https://lp.ellucian.com/rs/085-MHT-312/images/Ellucian_2026-AI-Report.pdf

Elsevier. (2026a). *Generative AI policies for journals*. https://www.elsevier.com/about/policies-and-standards/generative-ai-policies-for-journals

Elsevier. (2026b). *LeapSpace: Move from curiosity to discovery, faster*. https://www.elsevier.com/products/leapspace

Felicetti, A., Himmiche, A., & Somenzi, M. (2025). Knowledge graphs and artificial intelligence for the implementation of cognitive heritage digital twins. *Applied Sciences*, *15*(18), 10061. https://doi.org/10.3390/app151810061

Georgieva, E., & Georgiev, T. (2025). How artificial intelligence can transform services in research libraries. In C. Kahraman et al. (Eds.), *Intelligent and Fuzzy Systems. INFUS 2025* (Lecture Notes in Networks and Systems, Vol. 1530). Springer, Cham. https://doi.org/10.1007/978-3-031-98565-2_76

Gottweis, J., & Natarajan, V. (2025). *Accelerating scientific breakthroughs with an AI co-scientist*. Google Research. https://research.google/blog/accelerating-scientific-breakthroughs-with-an-ai-co-scientist/

Grimmett, J. (2026). *AI may be making us think and write more alike*. USC Dornsife. https://dornsife.usc.edu/news/stories/ai-may-be-making-us-think-and-write-more-alike/

Heer, J., Conlen, M., Devireddy, V., Nguyen, T., & Horowitz, J. (2023). Living papers: A language toolkit for augmented scholarly communication. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (Article 42, pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3586183.3606791

International Association of Privacy Professionals (IAPP). (2024). *The EU AI copyright playbook: The TDM exception and AI Act's transparency requirements*.

https://iapp.org/news/a/the-eu-ai-copyright-playbook-the-tdm-exception-and-ai-act-s-transparency-requirements
JSTOR. (2025). *JSTOR's AI research tool: Overview*. https://support.jstor.org/hc/en-us/articles/25755478251159-JSTOR-s-AI-Research-Tool-Overview
Jain, S. J., Sibbu, K., & Kuri, R. (2023). *Conducting effective research using SciSpace: A practical approach*. Authorea. https://doi.org/10.22541/au.170111059.99508682/v1
Kim, J. S. (2025). Academic library with generative AI: From passive information providers to proactive knowledge facilitators. *Publications*, *13*(3), 37. https://www.mdpi.com/2304-6775/13/3/37
Lian, Y., & Xie, J. (2024). The evolution of digital cultural heritage research: Identifying key trends, hotspots, and challenges through bibliometric analysis. *Sustainability*, *16*(16), 7125. https://doi.org/10.3390/su16167125
Matz, M. (2026). *AI "adviser" accelerates robotic design of advanced electronic materials*. Argonne National Laboratory. https://www.anl.gov/article/ai-adviser-accelerates-robotic-design-of-advanced-electronic-materials
Molecule. (2026). *Molecule protocol home | A new era of drug development*. https://www.molecule.to/
Nockels, J., Gooding, P., Ames, S., & Terras, M. (2022). Understanding the application of handwritten text recognition technology in heritage contexts: A systematic review of Transkribus in published research. *Archival Science*, *22*(3), 367–392. https://doi.org/10.1007/s10502-022-09397-0
OriginTrail. (2026). *Decentralized knowledge graph: The core of verifiable Internet for AI*. https://origintrail.io/technology/decentralized-knowledge-graph
OWASP Foundation. (2025). *OWASP top 10 for Large Language Model applications*. Open Worldwide Application Security Project (OWASP). https://genai.owasp.org/llm-top-10/
Schmidgall, S., & Moor, M. (2025). *AgentRxiv: Towards collaborative autonomous research*. arXiv. https://doi.org/10.48550/arXiv.2503.18102
Scopus. (2026). *Scopus AI: Think bigger. Move faster. Act with confidence*. https://www.elsevier.com/products/scopus/scopus-ai
Wang, L., Zhang, Z., Cao, Y., Tsung, F., & Luo, Y. (2026). *TableTale: Reviving the narrative interplay between data tables and text in scientific papers*. arXiv. https://doi.org/10.48550/arXiv.2602.22908
Web of Science. (2026). *Web of Science research assistant*. https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/web-of-science-research-assistant/
Wei, H., Sun, Y., & Li, Y. (2025). *DeepSeek-OCR: Contexts optical compression*. arXiv. https://doi.org/10.48550/arXiv.2510.18234
Westenberger, P., & Farmaki, D. (2025). Artificial intelligence for cultural heritage research: The challenges in UK copyright law and policy. *Frontiers in European Journal of Cultural Management and Policy*, *15*, 14009. https://doi.org/10.3389/ejcmp.2025.14009

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–1039. https://doi.org/10.1126/science.1136099

Zhu, L., Rahman, S. A., & Timan, H. (2026). Knowledge graph applications in cultural heritage: A ROSES-based systematic review. *Information*, *17*(3), 269. https://doi.org/10.3390/info17030269

Received: March 13, 2026
Reviewed: April 24, 2026
Finally Accepted: May 10, 2026