The Journal of Prevention of Alzheimer's Disease 13 (2026) 100398

Contents lists available at ScienceDirect

The Journal of Prevention of Alzheimer's Disease

journal homepage: www.elsevier.com/locate/tjpad

Special Article

Towards an AI biomedical scientist: Accelerating discoveries in neurodegenerative disease

Kaleigh F. Roberts [a,aa,1], Eric C. Landsness [b,aa,1], Justin Reese [c,aa], Donald Elbert [d,aa], Gabrielle Strobel [e,aa], Elizabeth Wu [e,aa], Yixin Chen [f,aa], Albert Lai [f,g], Zachary B. Abrams [g,aa], Mingfang Zhu [f], Justin Melendez [b,z,aa], Srinivas Koutarapu [b,aa], Sihui Song [b,aa], Yun Chen [b,aa], Robert Lazar [h], Payam Barnaghi [i], John F. Crary [j], Sergio Pablo Sardi [k,aa], Marc D. Voss [k], Rajaraman Krishnan [k], Joel W. Schwartz [l], Ron Mallon [m,aa], Gustavo A. Jimenez-Maggiora [n,aa], Chenguang Wang [o,aa], Thomas Sandmann [p,aa], Niranjan Bose [q,x], Mukta Phatak [r,aa], Gayle Wittenberg [s], Yannis G. Kevrekidis [t,aa], Cassie S. Mitchell [u], Ludovico Mitchener [v], Towfique Raj [w], Luca Foschini [x,aa], Gregory J. Moore [y,aa], Randall J. Bateman [b,z,aa,*]

[a] Department of Pathology & Immunology, Washington University, St. Louis, MO, USA
[b] Department of Neurology, Washington University, St. Louis, MO, USA
[c] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[d] Department of Neurology, University of Washington, Seattle, WA, USA
[e] Alzforum Foundation Inc., MA, USA
[f] Department of Computer Science & Engineering, Washington University, St. Louis, MO, USA
[g] Institute for Informatics, Data Science and Biostatistics (I2DB), Washington University, St. Louis, MO, USA
[h] Booz Allen Hamilton
[i] Department of Brain Sciences, Imperial College London, London, England United Kingdom
[j] Department of Pathology, Nash Family Department of Neuroscience, Department of Artificial Intelligence & Human Health, Neuropathology Brain Bank & Research CoRE, Ronald M. Loeb Center for Alzheimer's Disease, Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, NY, NY, USA
[k] Sanofi, Cambridge, MA, USA
[l] Bristol Myers Squibb, Cambridge, MA, USA
[m] Department of Philosophy, Washington University, St. Louis, MO, USA
[n] Alzheimer's Therapeutic Research Institute, Keck School of Medicine of University of Southern California, San Diego, CA, USA
[o] Department of Computer Science and Engineering, University of California Santa Cruz, CA, USA
[p] Denali Therapeutics Inc., South San Francisco, CA, USA
[q] Alzheimer's Disease Data Initiative, USA
[r] The 10,000 Brains Project, USA
[s] Johnson & Johnson, USA
[t] Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, USA
[u] Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, USA
[v] FutureHouse, USA
[w] Department of Neuroscience, Icahn School of Medicine at Mount Sinai, NY, NY, USA
[x] Sage Bionetworks, USA
[y] Gates Ventures, USA
[z] Tracy Family SILQ Center, Washington University, St. Louis, MO, USA
[aa] Consortium for Biomedical Research and Artificial Intelligence in Neurodegeneration (C-brAIn)

ARTICLE INFO

Keywords:
Artificial intelligence
Alzheimer's disease

ABSTRACT

Despite major advances in Alzheimer's disease and related diseases (ADRD) research, the translation of discoveries into impactful clinical interventions remains slow. Overwhelming data complexity, fragmented

* Corresponding author.
E-mail address: batemanr@wustl.edu (R.J. Bateman).
[1] Equal contribution.

https://doi.org/10.1016/j.tjpad.2025.100398
Received 4 September 2025; Received in revised form 14 October 2025; Accepted 20 October 2025
2274-5807/© 2025 Published by Elsevier Masson SAS on behalf of SERDI Publisher. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Large language model
Biomedical AI
Multimodal datasets

knowledge, and prolonged research cycles hinder progress in understanding and treating neurodegenerative diseases. Artificial intelligence (AI) offers a promising path forward, particularly when developed as a scientist-in-the-loop system that collaborates with researchers throughout the scientific discovery process. This paper introduces the concept of an AI Biomedical Scientist, an intelligent platform designed to support literature synthesis, hypothesis generation, experimental design, and data interpretation. This platform aims to function as a holistic scientific partner, integrating diverse biomedical data and expert reasoning to accelerate discovery. We review commercial and academic efforts and introduce targeted Minimum Viable Products (MVPs) needed for general biomedical research lab utilization of AI, such as robust and accurate tools for literature and data analysis, negative data models, and virtual peer review, with a longer-term vision of foundation models trained directly on biomedical datasets. In AD and neurodegeneration research, such tools are anticipated to deliver efficiency gains ranging from modest improvements in specific research tasks to potential multi-fold accelerations in discovery workflows as systems mature and scale. This review examines the technical foundations, challenges, and anticipated impacts of AI and aims to inform and engage researchers in utilizing these systems to transform biomedical discovery, starting with AD and extending to other complex conditions.

1. Introduction: the need for innovation in Alzheimer's research

Alzheimer's disease (AD) is a common neurodegenerative disorder, affecting an estimated 50 million people worldwide and contributing substantially to human, social, and economic burdens at an immense scale. More than a century after its initial description by Alois Alzheimer, research has advanced significantly, leading to the development of highly accurate diagnostic biomarkers and modestly effective disease-modifying therapies [1–3]. However, for most patients, treatments that meaningfully alter disease progression or outcomes remain limited [4].

A key challenge in AD research is the biological complexity of the disease. Studies have demonstrated that AD pathology begins up to 25 years before symptom onset, initiating a prolonged and dynamic pathogenic process involving compensatory mechanisms, evolving cell states, and interacting molecular pathways [5]. The extended course of Alzheimer's disease, combined with the rapid growth and complexity of biomedical data, has outpaced the capacity of traditional research methods to effectively synthesize all information to generate maximally informed actionable insights. Although over two million biomedical papers are published annually, fewer than 0.1 % of discoveries have a direct impact on human health outcomes [6]. Translational barriers, including fragmented knowledge across disciplines, slow cycles of hypothesis testing, and challenges in integrating diverse data types, are particularly pronounced in neurodegenerative diseases [7]. Moreover, more than 97 % of drugs entering AD clinical trials do not achieve approval (data from https://www.alzforum.org/therapeutics), underscoring inefficiencies in current discovery pipelines [8].

The challenges of AD research are compounded by the vast and accelerating amounts of biomedical data generated, including published data and "dark data" hidden in inaccessible sources or unpublished findings, such as negative results [9–13]. The huge number of publications (currently $>$ 200,000 for AD based on PubMed query for "Alzheimer's disease") make it impossible for researchers to stay abreast of findings outside their specialization, further exacerbating siloed understanding across domains and hampering disruptive science that fundamentally shifts current understanding [14]. As a consequence, the timelines to train researchers have also substantially increased, leading to a relative shortage in the number of qualified scientists needed for the challenge [15].

Artificial intelligence (AI) has the potential to improve how researchers navigate and interpret complex biomedical data [16–20]. While expert scientists possess highly refined reasoning skills, they are fundamentally limited by cognitive bandwidth, i.e. the amount of literature, data modalities, and prior findings they can hold in mind and integrate at once. AI systems can help overcome this limitation by surfacing relevant knowledge from across vast datasets and literature corpora, organizing connections, and enabling hypothesis generation that is informed by a broader and more comprehensive information space than a human could synthesize alone [21,22]. In this way, AI acts as a contextual amplifier, allowing scientists to apply their expertise more effectively across the full scope of available evidence.

Despite this potential, many current AI applications in biomedicine remain narrowly focused on specific tasks such as literature mining, diagnostic image analysis, risk prediction, or natural language summarization. While valuable, these tools stop short of supporting the more integrative and iterative processes of scientific reasoning required for foundational discovery. There remains a need for AI systems that can assist with the cognitive and analytical tasks central to discovery, including synthesizing knowledge, generating hypotheses, designing experiments, and learning from new data. Toward this end, the Consortium for Biomedical Research and AI in Neurodegeneration (c-brAIn) has been launched to accelerate health impactful basic science discoveries through the use of AI tools.

In this paper, we describe the concept of an AI Biomedical Scientist, a collaborative, scientist-in-the-loop system intended to support researchers throughout the scientific process. We focus on AD as an initial use case, given its clinical relevance, extensive datasets, and established research infrastructure. We outline the technical foundations of this approach, the early-stage tools currently in development, and the potential for AI to enhance research efficiency and outcomes in neurodegenerative disease studies.

2. What is an AI biomedical scientist?

An AI Biomedical Scientist is designed to support researchers across the biomedical research process, including literature and data review, hypothesis generation, experimental design, and data interpretation. Unlike traditional AI tools focused on single tasks, the AI Biomedical Scientist is designed to iterate through the entire scientific pipeline, including identifying biological questions, synthesizing literature, generating hypotheses, designing experiments, analyzing data, and interpreting results, by integrating publications, biomedical data, critical reasoning, and domain expertise to accelerate discovery and generate insights that improve human health.

A defining feature of this approach is its scientist-in-the-loop design, in which scientific experts remain involved in developing, refining, and interpreting the system's outputs to ensure scientific rigor and contextual relevance [23]. Rather than aiming to replace researchers, the AI scientist is intended to augment human work by improving efficiency, reproducibility, and the capacity to quickly and meaningfully explore new scientific questions.

The AI scientist is designed to integrate data from multiple domains, including genomics, proteomics, lipidomics, metabolomics, imaging, electronic health records, and behavioral measures, which are often distributed across institutions and research silos. Scientists can conduct analyses using local or external data sources through natural language interfaces. The system automates analyses through machine learning and deep neural net approaches, accelerating the time from measurement to interpretation.

Together, these capabilities position the AI Biomedical Scientist as a

valuable tool for advancing research in complex areas such as AD, underscoring the need to understand the specific AI technologies that make such a system possible.

3. Core AI technologies

The AI Biomedical Scientist combines several artificial intelligence technologies to support various aspects of biomedical research. To better understand how these technologies contribute, it is helpful to clarify key terms that are sometimes used interchangeably but refer to distinct concepts (Fig. 1).

Artificial intelligence broadly describes computational systems that perform tasks typically requiring human intelligence, such as understanding language, recognizing patterns, or making decisions. Within AI, *machine learning* refers to algorithms that improve performance by learning from data rather than following explicitly programmed rules. *Deep learning* is a specialized subset of machine learning that relies on large *neural networks*, which are computational models inspired by the structure and function of neurons in the brain. These networks consist of layers of interconnected nodes ("neurons") that process input data through weighted connections, enabling the system to learn complex patterns. *Transformers* are a specific class of deep neural network architectures that excel at processing sequential data and have become central to many modern AI applications.

A key component of the AI Biomedical Scientist is the *large language model* (LLM) (Fig. 2). LLMs are transformer-based neural networks trained on extensive textual corpora, including scientific literature, to generate human language [24]. They can summarize information, answer questions, and suggest hypotheses. However, general-purpose LLMs, such as GPT-5, Gemini, Claude, and Grok4, while effective in conversational tasks, often lack the precision, domain-specific knowledge, and interpretability needed for rigorous biomedical research [25]. Additionally, these models can produce "hallucinations" (or more accurately, confabulations), plausible but incorrect information, which poses challenges for their use in scientific contexts, where accuracy and truth are paramount [26].

An alternative approach to improve domain relevance is the development of biological *specialist LLMs*, models pre-trained or fine-tuned specifically on biomedical text. Examples include BioBERT[27], BioGPT[28], BiomedLM[29], Med-Gemini[30], and Med-PaLM[31]. These models can offer enhanced vocabulary coverage, improved factual recall, and greater alignment with domain-specific language. However, specialist LLMs face several limitations. First, their knowledge is static and can quickly become outdated in fast-evolving fields. Second, they are often narrow in scope, performing well in specific biomedical subdomains but struggling when tasks require broader reasoning or interdisciplinary integration. Finally, they still retain the inherent limitations of LLMs, including susceptibility to hallucinations and opaque decision-making.

To address these limitations, the AI Biomedical Scientist incorporates *retrieval-augmented generation (RAG)* architectures. RAG systems enhance LLMs by connecting them to external databases or curated literature, helping ensure that generated responses are grounded in factual sources [32,33]. This design allows us to circumvent some of the limitations of static pretrained models by deferring knowledge retrieval to inference time, increasing flexibility, and reducing the need for frequent retraining. A further refinement of this approach is integrating *knowledge graphs* with RAG (e.g. GraphRAG), to link textual outputs directly to structured evidence nodes [34]. Knowledge graphs represent biomedical concepts and the relationships among them and offer a way to organize information and support reasoning that goes beyond simple keyword matching.

In addition, the proposed AI Biomedical Scientist leverages emerging concepts from *agentic AI* [35,36]. Unlike traditional models that only respond to queries, agentic AI systems are designed to autonomously plan, reason, and take actions toward defined goals. Agentic AI can iteratively break down complex research tasks, select appropriate tools (such as querying databases, running analyses, or simulating models), and adapt based on intermediate results. This agent-like behavior shifts AI from a reactive assistant toward a proactive collaborator capable of

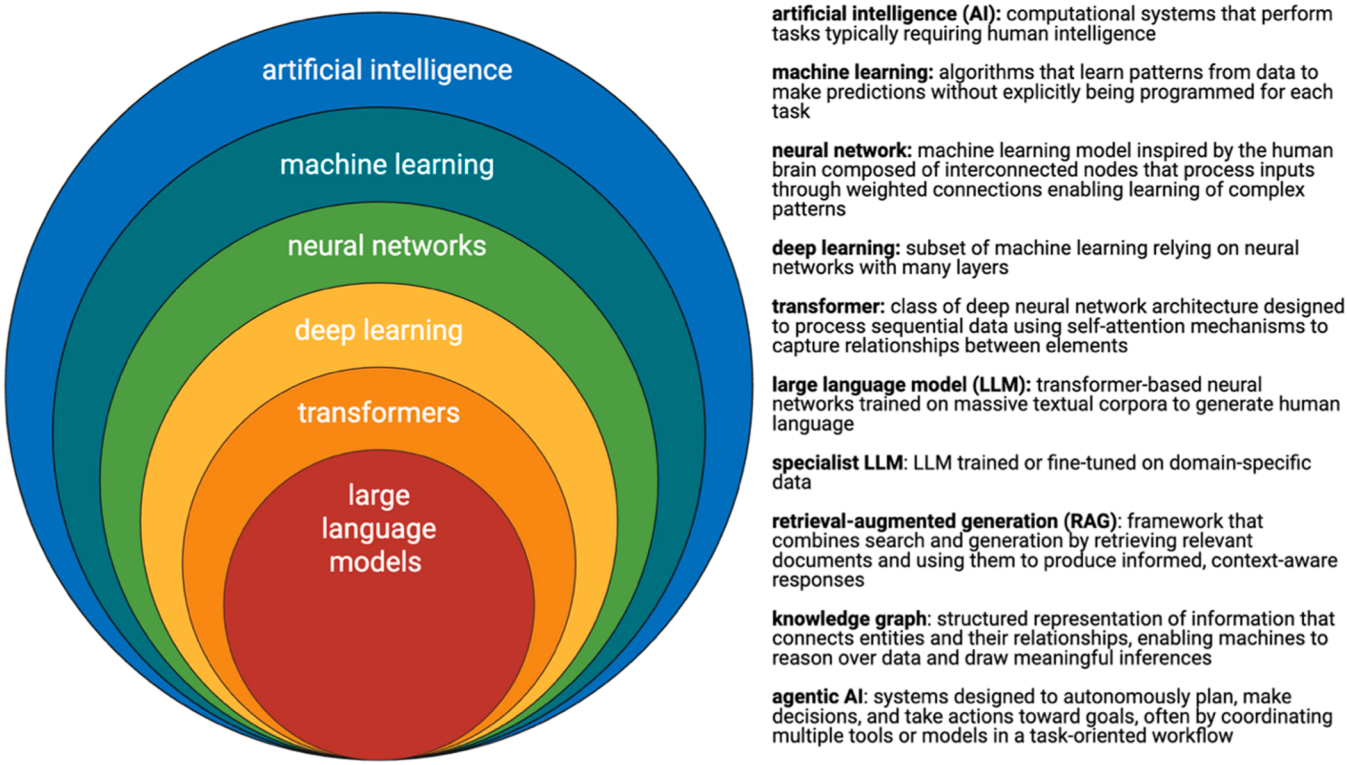


Fig. 1. Glossary of AI terms.

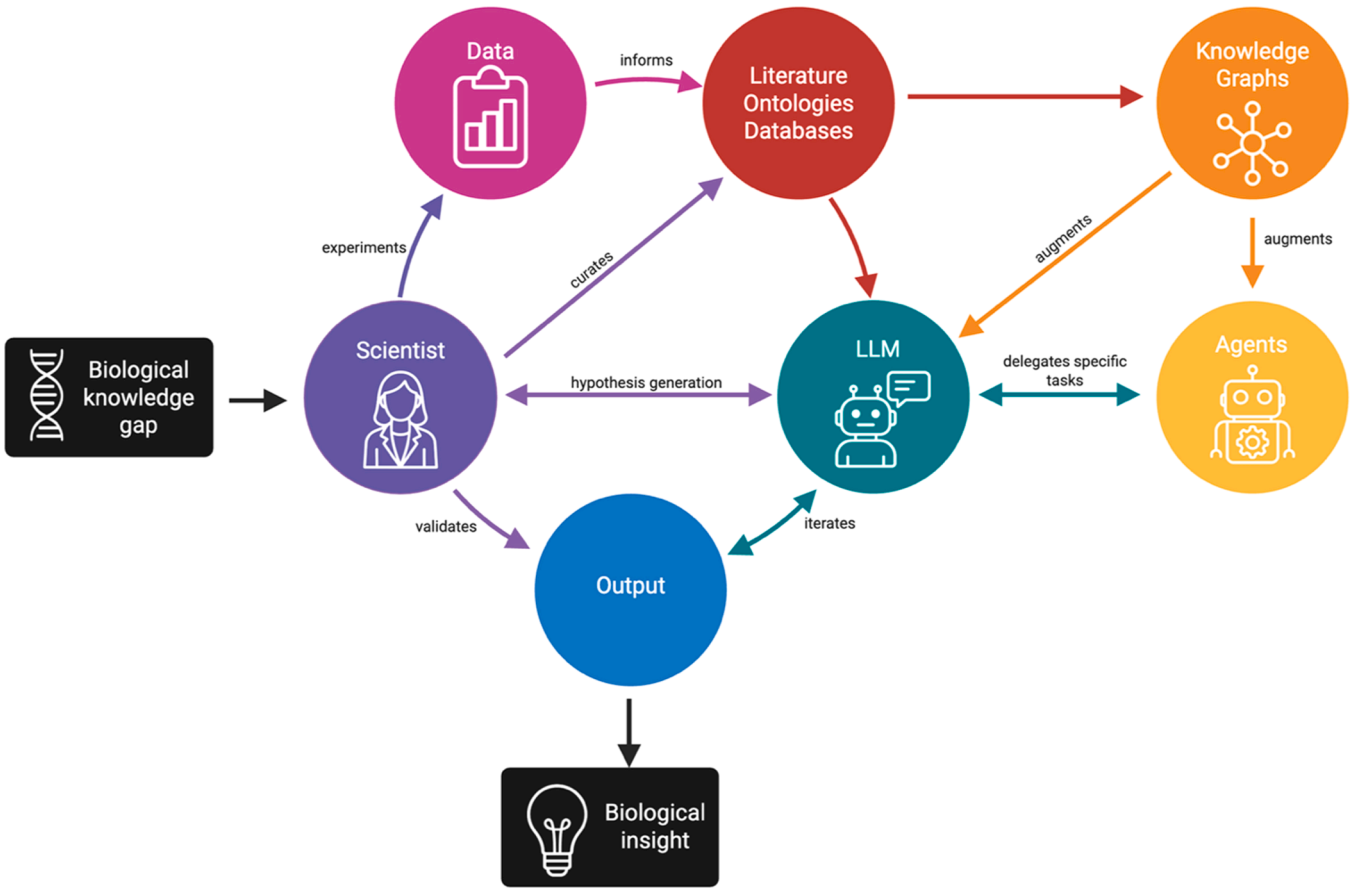


Fig. 2. Workflow diagram for AI Biomedical Scientist. A biological knowledge gap is first identified by a scientist, who conducts experiments to generate raw data. These data are then captured in published literature, ontologies, and databases curated by human experts. Organized knowledge frameworks derived from these sources can be transformed into knowledge graphs and incorporated into large language models (LLMs). Through an iterative process, human scientists interact with the LLM, which assigns specific tasks to specialized AI agents such as data analysis, hypothesis generation, literature retrieval, and critical review. This human-AI collaboration supports the identification of new biological insights, their contextualization within existing data and literature, and expert validation to ensure scientific relevance.

orchestrating multi-step workflows. However, while these systems introduce powerful automation capabilities, they are not intended to operate in isolation. A key design principle is maintaining an appropriate balance between automated task execution and human oversight. Researchers remain essential in setting goals, curating inputs, interpreting results, and determining when to trust or override AI-driven decisions.

Another envisioned capability of the AI Biomedical Scientist is *multimodal data integration*. Modern biomedical research generates diverse datasets across domains, scale, and time, including genomics, proteomics, and other biomolecular measures across atomic, molecular, cellular, tissue, and organ scales with human imaging, clinical, vocal, and behavioral measurements that are often stored in separate systems. The AI Biomedical Scientist is designed to utilize these heterogeneous data sources, helping researchers examine relationships across different domains and develop integrated models of disease processes.

Finally, the AI Biomedical Scientist's architecture is designed to operate in *federated environments*, allowing analyses to be performed across multiple institutions without requiring the sharing of raw data. In a federated approach, data remain securely within each institution while algorithms or models are shared and run locally, and only the aggregated results are combined. This can be especially important for working with proprietary pharmaceutical data, sensitive patient records, or other data subject to privacy regulations. Examples of successful federated environments in multiomic workflows include PPML-Omics[37], DataSHIELD[38] and OmicSHIELD[39]. Within the AD research community, the Alzheimer's Disease Data Initiative (ADDI) utilizes the Federated Data Sharing Appliance (FDSA), a secure data application that enables multiple organizations to securely share data without the need for centralization in a single repository [40]. Federated training of AI models is achieved by distributing global model parameters to local sites, training on private local data, and then returning updated model parameters to the centralized server [37,41–44]. Such methods help address privacy, sovereignty, and regulatory requirements while enabling collaboration at a scale needed for research in AD and other complex conditions.

Together, these technologies can create an assistive framework designed to complement scientific expertise. By combining natural language processing, structured knowledge representation, data integration, agentic AI, and customized foundational models, the AI Biomedical Scientist aims to accelerate how researchers generate hypotheses, design experiments, and analyze results. In our own development efforts, we have implemented and evaluated early prototypes of this system, specifically focusing on literature retrieval and synthesis using RAG architectures trained on curated Alzheimer's disease corpora [45], and also testing prototype multi-agentic systems. These systems represent the foundation for future more advanced capabilities such as multimodal integration and added-value agentic workflows.

4. Current AI tools and systems

The use of artificial intelligence to support scientific research is

gaining momentum in both commercial and academic settings. Several initiatives are exploring LLM-based systems designed to function as AI scientists, capable of parsing scientific literature, suggesting hypotheses, or assisting with experimental planning [46–49]. These efforts have emerged in response to the growing challenge that individual researchers face in keeping up with the rapidly expanding volume of scientific publications and data.

Examples of specialized AI tools for scientific applications include Google's Co-Scientist [50], FutureHouse's Robin [51], and AI2's Asta. These systems differ in their scope, underlying technologies, and the extent to which they incorporate expert scientific input (also discussed in the paper by Funk et al. in this special edition of JPAD)., [ref] While these systems represent significant progress and are at the current cutting edge of applying AI to research tasks, we believe significant additional validation is required to demonstrate their reliability and impact in real-world scientific research. A non-exhaustive list of AI biomedical scientific platforms is summarized in Table 1.

These initiatives represent important early progress in applying generative AI and agentic tools to biomedical research, and many offer capabilities that will likely inform and complement future systems. Our proposed AI Biomedical Scientist builds on this growing foundation, with a particular focus on addressing the specific challenges of AD and biomedical research. Rather than replacing or competing with existing platforms, it seeks to integrate and extend their capabilities within a disease-focused, scientist-guided framework. Key distinguishing features include its explicit design for AD and neurodegeneration, enabling deeper incorporation of domain-specific knowledge, and its emphasis on high-quality multimodal "dark" data fusion across genomic, proteomic, imaging, clinical, and behavioral domains that are otherwise not accessible. Importantly, the system is developed around a scientist-in-the-loop model, in which domain experts play an active role in curating inputs, interpreting outputs, and refining system behavior to ensure scientific rigor and practical relevance.

5. Why Alzheimer's disease is the ideal proving ground

AD is a particularly suitable area for developing and implementing an AI Biomedical Scientist [63]. Decades of both broad and deep research have produced enormous datasets, from molecular characterization, cellular and animal models, through human pathologic, imaging, genomic, proteomic, and clinical data from initiatives such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) [64], Dominantly Inherited Alzheimer Network (DIAN) [65], DIAN-TU [66], and national and international observational cohorts and clinical trials summarized in Table 2. These resources provide a strong basis for developing and validating AI models.

Beyond data availability, the field also benefits from structured knowledge platforms like Alzforum and data aggregators such as the AD Data Initiative (ADDI). These curated resources are valuable for training AI systems designed to work with the complexities of neurodegenerative biology.

Furthermore, the AD research ecosystem is characterized by strong cross-sector research groups in pharmaceutical and biotechnology companies, academic institutions, and patient advocacy groups, with thousands of researchers working to decipher the causes and pathophysiology of AD. This large and diverse group of researchers will enable scientist-in-the-loop training and integration and evaluation of AI-driven research tools.

Taken together, diverse datasets, structured knowledge platforms, and strong research collaboration across academia and industry make AD an ideal domain for developing AI tools. These resources create opportunities for practical systems that support researchers in addressing complex scientific questions, a goal that begins with developing focused initial solutions.

Table 1
Examples of AI biomedical scientific platforms currently in development including notable features and links for access.

System	Use	Access
Readily accessible via web user interface		
FutureHouse Robin [51]	Multi-agent system for literature search, hypothesis generation, experimental design, data analysis, figure generation, and experimental planning	Free partial web access to agents at platform.futurehouse.org, code available at https://github.com/Future-House/robin
Ai2 Asta	Multi agent system for literature search and summarization	Free, https://asta.allen.ai/chat
BenchSci Ascend	Multi agent system using proprietary multimodal LLMs supported by a knowledge graph and ontology knowledge base	https://knowledge.benchsci.com/home/platform-fundamentals, free access to Selector Tool for academics, most tools require paid subscription
AlzAssistant	Literature-based Q&A using PaperQA2, curated AD paper corpus, and Alzheimer's knowledge graph	Free, https://chat.alzassistant.org/
AlzheimerRAG[52]	Q&A using multimodal RAG pipeline with AD PubMed corpus	Free, https://tinyurl.com/AlzheimerRAG
Biomni[22]	Generalist agentic architecture that integrates LLM reasoning with retrieval-augmented planning and code-based execution, enabling complex biomedical workflows	Free, https://biomni.stanford.edu/ https://github.com/snap-stanford/biomni
DORA[53]	Multi-agent scientific exploration and draft outline research assistant for automated or semi-automated research studies and report generation	https://dora.insilico.com
Code available, but no web interface		
Sakana The AI Scientist[48]	Idea generation, computational experiment conduction, paper writing and review	https://github.com/SakanaAI/AI-Scientist/tree/main/ai_scientist
SemNet[54]	Literature-based discovery system enabling PubMed relationship literature mining	https://github.com/pathology-dynamics/semnet-2
The Virtual Lab[55]	LLM principal investigator agent guiding a team of LLM agents with different scientific backgrounds (e.g., a chemist agent, a computer scientist agent, a critic agent), with a human researcher providing high-level feedback	https://github.com/zou-group/virtual-lab
Data-to-paper[56]	Automation platforms that guides LLM agents starting with annotated data through hypothesis generation, data analysis, results interpretation, and manuscript preparation	https://github.com/Technion-Kishony-lab/data-to-paper
RBio by CZI[57]	Reasoning model combining virtual cell models with chat interface of LLMs to predict how cells will behave in experiments	https://github.com/czi-ai/rbio
X-Master[58]	Tool-augmented reasoning agent designed to emulate human researchers by interacting flexibly with external tools during its reasoning process	https://github.com/sjtu-sai-agents/X-Master

(*continued on next page*)

Table 1 (*continued*)

System	Use	Access
BioResearcher[59]	Modular multi-agent architecture integrating search, literature synthesis, experimental design, and programming	https://github.com/XMUDM/BioResearcher
BioDiscoveryAgent [60]	Agent for designing genetic perturbation experiments	https://github.com/snap-stanford/BioDiscoveryAgent
Not publicly available		
Google Co-Scientist [50]	Multi-agent system focused on literature search and iterative hypothesis refinement	Not publicly available, paid institutional access
Lila.ai	Platform announced by Flagship Pioneering to develop "superintelligence in science," integrating LLMs, reasoning systems, and autonomous laboratory platforms to accelerate discovery	Not publicly available
PROTEUS[61]	Fully automated scientific discovery system for hypothesis generation from raw proteomic data	Not publicly available
STELLA[62]	Multi-agent architecture that self-evolves reasoning strategies and discovers and integrates bioinformatic tools	Not publicly available

6. Phase 1: minimum viable products (MVPs) in development

A practical step toward applying the AI Biomedical Scientist in AD research is the development of targeted Minimum Viable Products (MVPs) that address specific challenges in the research process. These initial tools aim to test technical approaches and support AI-driven research amid large datasets, expanding experimental findings, and the complex biology of neurodegeneration and AD. One of the first areas of development is creating tools for literature search and synthesis. Early work across various domains suggests that traditional search is still better than AI-based search tools, however as AI continues to improve the advantage it gives in speed will become more relevant [95–97]. We have been developing AI literature search systems that integrate RAG architectures with LLMs trained on AD-focused scientific texts and connected to knowledge graphs [45]. The goal is to help researchers quickly identify, compare, and summarize relevant information from both published literature and internal datasets, making it easier to navigate the existing scientific knowledge. Importantly, domain-specific scientists are actively integrated into the developmental process to ensure accuracy and relevance of AI-generated responses. Additional emphasis on the curation of a trustworthy, high-quality corpus of training literature ensures a solid central knowledge foundation.

A second MVP focuses on addressing the challenge of negative and unpublished results in biomedical research. Negative findings are often absent from published literature, which can contribute to redundant research efforts and leave important areas of biological understanding unexplored [13]. The proposed "Negative Data Analyzer" would aim to process data from both published studies and non-public sources, including data held in federated environments such as pharmaceutical company datasets and unpublished results in labs. By integrating information from studies with non-significant or null results, this tool is intended to help reduce unnecessary duplication of experiments and identify conditions that influence biological mechanisms.

A third MVP, referred to as "Reviewer Three," is envisioned as a virtual scientific reviewer and research advisor trained specifically on a corpus of documents and paired reviews. Its purpose would be to provide feedback on grant applications, experimental designs, and manuscripts. Initial evaluations of LLMs as scientific reviewers have demonstrated substantial overlap between human and AI generated

Table 2
Examples of Multimodal Data Resources.

Resource Type	Examples	Types of Data	Utility/Relevance
AD Data repositories	AD Knowledge Portal[67], ADDI Repository[68], GNPC[69], IDA [70], NACC[71], NIAGADS[72]	Clinical, cognitive, imaging, biomarkers, genomic, transcriptomic, proteomic, metabolomic	International harmonized datasets over multiple studies
AD-specific observational cohorts	ADNI[64], ADRCs, ADSP[73], AIBL [74], BioFINDER [75], DELCODE [76], DIAN[65], EPAD[77], ROSMAP[78]	Clinical, cognitive, imaging, biomarkers, genomic, transcriptomic, proteomic, metabolomic	Imaging, CSF/ biomarkers, cognitive assessments, and multi-omic data across diverse AD cohorts
AD-specific interventional studies	A4/LEARN[79], AHEAD 3–45[80], APEX, DIAN-TU [66]	Clinical, cognitive, imaging, biomarkers, genomic, transcriptomic, proteomic, metabolomic	Highly phenotyped AD cohorts with therapeutic interventions
General population and longitudinal aging cohorts	100-plus Study [81], All of Us Research Program [82], BLSA[83], CAMCAN[84], FinnGen[85], HASD/ACS[86], Human Connectome Project[87], MCSA[88], RESILIENT[89], UK Biobank[90]	Genomics, imaging, clinical	Large-scale data integrating genetic profiles, imaging, and health records across the lifespan
Real-world clinical datasets	Optum Clinformatics, TriNetX[91]	Federated EHR networks, claims data, and clinical datasets	Real-world data capturing clinical heterogeneity and operational variability not present in curated or protocol driven datasets
Digital biomarker studies	mPower[92], RADAR-AD[93], TIHM[94]	Wearable sensor data	Digital monitoring of daily living to capture dynamic changes missed in episodic clinical assessments
Not publicly available datasets	Proprietary Pharma data, individual laboratory research data	Clinical trial results, experimental data, negative results	Often unpublished, but critical for hypothesis generation and reducing duplication of effort
Curated knowledge	Alzforum	Curated literature, structured knowledge	AD-focused commentary, news, and community consensus

A4: Anti-Amyloid Treatment in Asymptomatic Alzheimer's, ACTC: Alzheimer's Clinical Trials Consortium, ADDI: Alzheimer's Disease Data Initiative, ADNI: Alzheimer's Disease Neuroimaging Initiative, ADRC: Alzheimer's Disease Research Center, ADSP: Alzheimer's Disease Sequencing Project, AIBL: Australian Imaging, Biomarkers and Lifestyle Study, APEX: Alzheimer's Plasma Extension Study, BioFINDER: Biomarkers For Identifying Neurodegenerative Disorders Early and Reliably, BLSA: Baltimore Longitudinal Study of Aging, CAMCAN: Cambridge Centre for Ageing and Neuroscience, DELCODE: DZNE Longitudinal Cognitive Impairment and Dementia Study, DIAN: Dominantly Inherited Alzheimer Network, DIAN-TU: DIAN Trials Unit, EPAD: European Prevention of Alzheimer's Dementia, GNPC: Global Neurodegeneration Proteomics Consortium, HASD/ACS: Healthy Aging & Senile Dementia/The Adult Children Study, IDA: Image & Data Archive at LONI, LEARN: Longitudinal

Evaluation of Amyloid Risk and Neurodegeneration, MCSA: Mayo Clinic Study of Aging, mPower: Mobile Parkinson Disease Study, NACC: National Alzheimer's Coordinating Center, NIAGADS: National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site, RADAR-AD: Remote Assessment of Disease And Relapse – Alzheimer's Disease, ROSMAP: Religious Orders Study and Rush Memory and Aging Project, TIHM: Technology Integrated Health Management.

reviews and overall positive user perceptions of usefulness [98]. By simulating different review styles, from constructive mentoring to critical peer evaluation, this system is intended to help researchers refine hypotheses, strengthen experimental plans, and anticipate potential reviewer concerns.

Taken together, these MVPs represent a practical, additive approach to applying AI in AD research. Each is designed to address specific challenges faced by researchers and to serve as an early demonstration of how AI tools can integrate into scientific workflows. While these initial efforts are focused and targeted, they lay the foundation for more comprehensive systems in the future, where larger gains in efficiency and scalability may be possible through broader applications of AI technologies.

7. Phase 2: toward an integrated platform and foundation models for scientific discovery

While our initial focus is on developing targeted tools to address specific challenges in AD research, the longer-term vision extends beyond individual solutions. In Phase 2, the aim is twofold: to create a unified platform that integrates these capabilities into a cohesive system and to develop foundation models trained directly on biomedical data.

A key objective of Phase 2 is to combine the functions demonstrated in the initial MVPs into a single, integrated platform. Such a system is intended to help researchers navigate the entire scientific process more efficiently, providing interconnected tools for the iterative cycle of questioning, analysis, and discovery. The second major goal in Phase 2 involves developing foundation models specifically trained on large-scale biomedical data. Unlike general-purpose language models trained primarily on internet text, these biomedical foundation models would be developed using domain-specific datasets such as genomic and proteomic profiles, neuroimaging data, longitudinal clinical records, and results from both published and non-public studies. Unlike textual corpora, these raw biomedical data types are less prone to becoming outdated, offering a more durable substrate for foundational learning. Our strategy focuses on leveraging these resilient data sources to build more robust and broadly applicable models. These models are intended to capture complex patterns and relationships within the data, potentially enabling the generation of new hypotheses, predictions about disease mechanisms, or identification of biomarkers associated with disease progression and therapeutic response. Initial efforts have demonstrated that LLMs can generate novel and valid hypotheses even when tested on literature unrelated to the training data [99].

Overall, Phase 2 represents a transition from testing individual tools to building a unified, scalable AI system intended to support the entire biomedical research process. While data-related, methodological, and practical challenges remain, the potential for improved efficiency and deeper scientific insights underscores the importance of this next stage of development. Realizing this vision will require addressing these challenges, which are discussed in the following section.

8. Challenges and mitigation strategies

The application of AI in biomedical research, while promising, presents several important challenges that must be addressed to ensure scientific integrity and responsible use [100–102]. From a technical perspective, developing and deploying advanced AI models requires significant computational resources and specialized infrastructure, which can pose practical barriers for many current research groups.

Data privacy and security also remain critical concerns, particularly given the sensitive nature of biomedical information and the ethical and legal frameworks (e.g., HIPAA, GDPR) that govern its use [42,43,103,104]. In addition, intellectual property and copyright restrictions can limit the use of scientific publications and proprietary datasets for training AI models or deploying research tools, due to licensing agreements and data ownership concerns. A related concern is the risk of data duplication across repositories, which can lead to biased analyses, redundant processing, and inflated sample sizes. Strategies for mitigating data duplication include use of persistent universally unique identifiers (UUIDs) for participant-level tracking where available, and matching algorithms to identify likely duplicate records in datasets without direct identifiers [105,106].

Beyond infrastructure and governance, adoption within the scientific community presents its own set of challenges. Many scientists, especially those less familiar with AI methods, may have valid concerns about the reliability, transparency, effectiveness, and interpretability of current generative AI outputs, including the risk of generating hallucinations [26]. Addressing these concerns will require clear communication about both the capabilities and limitations of AI systems, along with careful validation and demonstration of their practical value in scientific contexts.

The rapid pace of advancement in AI further complicates adoption. New tools, models, and best practices evolve quickly, making it difficult for biomedical researchers to stay current. At the same time, biomedical data itself is constantly growing, posing logistical and organizational challenges for version control, reproducibility, and integration with existing systems. Ensuring that models remain both accurate and aligned with the latest knowledge will require adaptive infrastructure, modular system designs, and ongoing collaboration between AI experts and domain scientists.

A particularly important set of challenges centers on ethical concerns, specifically algorithmic bias, accountability, and potential misuse [103,101,107]. Like all data-driven tools, AI models are susceptible to biases embedded in their training data, which can result in unequal performance across populations or misleading conclusions when trained on limited or biased data. The types, quality, amount, and range of biological data will completely change what an AI system can identify and discover at all levels. Many of the strategies used to detect and mitigate human bias in science, such as disaggregated analyses, dataset balancing, and transparency in decision-making, can and should be adapted to AI development and evaluation. Establishing accountability frameworks is also essential. These may include audit trails, logging systems, and traceable outputs that allow researchers to understand how a model arrived at a conclusion, assess its reliability, and flag potential errors. Such infrastructure is especially important as AI becomes more integrated into workflows, where overreliance or automation bias may lead to uncritical acceptance of flawed results. Misuse can also take more subtle forms, such as reinforcing low-quality analyses or contributing to inefficiencies. These risks highlight the importance of a scientist-in-the-loop design, in which both users and developers share responsibility for evaluating outputs, selecting appropriate tools, and maintaining high standards of scientific integrity throughout the AI development process.

Responsible development of AI tools for biomedical research depends on maintaining high scientific standards [108]. Important principles include careful validation to understand performance and limitations, transparent reporting of how models are developed and assessed, and clear definitions of the roles and boundaries of human oversight. In particular, expert oversight is critical for curating high-quality input data and literature, filtering out low-quality or misleading information, and validating AI-generated outputs before they inform scientific conclusions. This close involvement of domain experts will help develop, judge, and rate AI tools, to maintain scientific credibility and relevance. The c-brAIn brings together a broad network of biomedical researchers to enable collective expert review of both

inputs and outputs.

By thoughtfully addressing these challenges, AI has the potential to become a valuable tool for biomedical research, supporting scientists in navigating complex data and generating new insights, particularly in fields like AD. Maintaining scientific rigor, transparency, and human oversight will be essential to ensuring its responsible and effective use.

9. Anticipated impact on Alzheimer's research

Integrating AI into biomedical research has the potential to offer measurable benefits for the study of AD. The envisioned AI Biomedical Scientist, with its capacity to synthesize complex data and literature, is expected to support researchers in accelerating key scientific processes. Early estimates suggest that AI-assisted literature review could reduce the time required by ~25 % compared to traditional manual approaches [97]. Such improvements are particularly relevant in AD research, where timely insights can contribute to advancing therapeutic development. Looking ahead, broader integration of AI systems in Phase 2 may enable even greater efficiencies, potentially achieving gains of two- to ten-fold in certain research workflows [51].

Beyond improving efficiency, the assistant is intended to support the rigor and reproducibility of scientific research. By systematically integrating information from diverse sources, the AI system may help researchers identify consistent patterns and avoid pursuing directions less likely to yield meaningful results. This capability could contribute to more effective target identification and validation, potentially supporting higher success rates in experimental studies and subsequent clinical trials, although precise estimates of such impacts remain uncertain.

The potential benefits of AI tools in AD research extend beyond individual laboratories. Advanced analytics and knowledge synthesis capabilities, which have typically been available to large research institutions with substantial resources, could become more accessible to smaller labs and early-career investigators. Increasing the availability of these tools may help reduce disparities in research capacity and promote contributions from a more diverse scientific community.

In addition, AI systems could play a supportive role in training the next generation of biomedical researchers. By providing examples of literature analysis, hypothesis development, and experimental critique, the AI Biomedical Scientist may help shorten learning curves for trainees and early-career scientists. Access to such resources could enhance scientific literacy and build confidence in working with complex, data-driven questions, potentially contributing to a more skilled and adaptable research community.

While precise metrics will continue to emerge as these systems are further developed and tested, integrating AI into AD research offers meaningful opportunities to improve scientific workflows, enhance reproducibility, and expand access to advanced analytical capabilities.

10. Future vision

The longer-term vision for AI in biomedical research extends beyond developing individual tools or even unified platforms. As foundational models and integrated systems mature, one key aspiration for future development lies in enabling semi-autonomous experimental design. In this scenario, AI systems could propose experimental protocols, suggest statistical methodologies, and forecast potential outcomes by synthesizing prior literature and integrated data models. While ultimate decision-making and oversight would remain in scientists' hands, the ability of AI to rapidly generate and evaluate experimental scenarios has the potential to increase research efficiency and facilitate exploration of more diverse scientific hypotheses. This capability would effectively yield a multiplier effect on capacity for research studies in the lab.

Expanding the application of AI tools beyond AD represents another important frontier. Technical architectures and methodological insights developed through neurodegenerative research are anticipated to be adaptable to other biomedical domains, including oncology, rare diseases, immunology, and complex chronic conditions. For example, AI systems capable of integrating diverse datasets, such as genomic profiles, imaging, and real-world clinical data, could help uncover shared pathways across diseases or identify patient subgroups more likely to benefit from specific therapies. Such cross-disciplinary insights might accelerate progress in fields where traditional research approaches have faced persistent challenges.

Beyond individual research programs, the broader vision for AI in biomedical science envisions a fundamental evolution in how discoveries are made. The traditional paradigm, characterized by linear hypothesis testing and siloed data sources, could increasingly give way to iterative, data-driven exploration powered by AI systems able to synthesize information across domains and scales. While realizing such capabilities will require significant advances in AI technologies, robust validation, and sustained collaboration between computational scientists and biomedical experts, this evolution holds the promise of accelerating the translation of basic research into clinical advances, ultimately contributing to improved diagnostics, therapies, and preventive strategies across a wide range of diseases.

11. Call to action

The development of an AI Biomedical Scientist marks a step toward transforming scientific discovery in complex fields like AD. The urgency of this challenge and the scale of resources and expertise it demands has led to the formation of a dedicated consortium committed to designing, building, and refining this new class of scientific tools.

A white paper outlining the scientific and technical vision for this initiative has been published and is available to the research community. But moving from vision to practical reality requires deep collaboration across disciplines, institutions, and sectors. The AI Biomedical Scientist is being developed as a tool built by scientists, for scientists. Its success will depend on diverse input and engagement from those who understand the complexities of biomedical research and the pressing need for new solutions.

For those interested in contributing or learning more, additional information is available at: https://c-brain.org. We invite researchers, clinicians, data scientists, and technology developers to join this effort. There are many ways to participate, from contributing domain expertise and engaging in pilot projects, to building and testing emerging tools, offering feedback on usability and performance, and sharing perspectives on critical scientific questions where AI could make a difference. Funders and philanthropic organizations interested in accelerating scientific progress are also encouraged to explore ways to support the development and broad dissemination of these tools. All interested parties can begin by filling out the survey on the c-brAIn website.

AD poses enormous challenges, but it also presents a profound opportunity: to harness innovative technologies and collaborative spirit to unlock new understanding and improve outcomes for patients and families affected by these devastating conditions. We invite the scientific community to help shape and build the next generation of tools that could redefine how discovery happens.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to improve language and readability. After using this tool/service, the authors reviewed and heavily edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Randall J. Bateman reports financial support was provided by Avid

Radiopharmaceuticals Inc. Randall J. Bateman reports financial support was provided by Janssen Pharmaceuticals Inc. Randall J. Bateman reports financial support was provided by Roche Diagnostics GmbH. Randall J. Bateman reports financial support was provided by Genentech Inc. Randall J. Bateman reports financial support was provided by Eli Lilly and Company. Randall J. Bateman reports financial support was provided by Eisai Inc. Randall J. Bateman reports financial support was provided by Biogen Inc. Randall J. Bateman reports financial support was provided by AbbVie Inc. Randall J. Bateman reports financial support was provided by Bristol-Myers Squibb Company. Randall J. Bateman reports financial support was provided by Novartis Pharmaceuticals Corporation. Randall J. Bateman reports a relationship with C2N Diagnostics, LLC that includes: board membership, consulting or advisory, and equity or stocks. Randall J. Bateman reports a relationship with Roche that includes: board membership. Randall J. Bateman reports a relationship with Biogen Inc that includes: board membership. Robert Lazar reports a relationship with Booz Allen Hamilton Inc that includes: employment. Sergio Pablo Sardi reports a relationship with Sanofi that includes: employment. Marc D. Voss reports a relationship with Sanofi that includes: employment. Joel W. Schwartz reports a relationship with Bristol Myers Squibb Co that includes: employment. Thomas Sandmann reports a relationship with Denali Therapeutics Inc that includes: employment. Gayle Wittenberg reports a relationship with Johnson & Johnson Inc that includes: employment. Niranjan Bose reports a relationship with Gates Ventures that includes: employment. Gregory J. Moore reports a relationship with Gates Ventures that includes: independent contract. Ludovico Mitchener reports a relationship with FutureHouse that includes: employment. Gabrielle Strobel reports a relationship with Alzforum that includes: employment. Elizabeth Wu reports a relationship with Alzforum Foundation Inc that includes: employment. Luca Foschini reports a relationship with Sage Bionetworks that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This initiative is supported by major consortium partners including Washington University, Gates Ventures, Alzforum, AD Data Initiative, Alzheimer's Association, Anonymous Foundation, Bristol Myers Squibb, Sanofi, Tau Consortium, The Dolby Family, The 10,000 Brains Project, The Robertson Foundation, Eisai, Johnson & Johnson, and over 100 engaged biomedical researchers. Figures were created in BioRender. https://BioRender.com/gq38ex5 and https://BioRender.com/tw98m0y

References

[1] Barthélemy NR, Salvadó G, Schindler SE, He Y, Janelidze S, Collij LE, et al. Highly accurate blood test for Alzheimer's disease is similar or superior to clinical cerebrospinal fluid tests. Nat Med 2024;30(4):1085–95. Apr.
[2] Bateman RJ, Li Y, McDade EM, Llibre-Guerra JJ, Clifford DB, Atri A, et al. Safety and efficacy of long-term gantenerumab treatment in dominantly inherited Alzheimer's disease: an open-label extension of the phase 2/3 multicentre, randomised, double-blind, placebo-controlled platform DIAN-TU trial. Lancet Neurol 2025;24(4):316–30. Apr 1.
[3] Sims JR, Zimmer JA, Evans CD, Lu M, Ardayfio P, Sparks J, et al. Donanemab in early symptomatic Alzheimer disease: the TRAILBLAZER-ALZ 2 randomized clinical trial. JAMA 2023;330(6):512–27. Aug 8.
[4] Vigneswaran S, Vijverberg EGB, Barkhof F, van de Giessen E, Lemstra AW, Pijnenburg Y, et al. Real-world" eligibility for anti-amyloid treatment in a tertiary memory clinic setting. Alzheimers Dement 2025;21(6):e70375. June 12.
[5] Li Y, Yen D, Hendrix RD, Gordon BA, Dlamini S, Barthélemy NR, et al. Timing of biomarker changes in sporadic Alzheimer's disease in estimated years from symptom onset. Ann Neurol 2024;95(5):951–65. May.
[6] Ioannidis JPA. Why most clinical research is not useful. PLoS Med 2016;13(6):e1002049. June 21.
[7] Myszczynska MA, Ojamies PN, Lacoste AMB, Neil D, Saffari A, Mead R, et al. Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. Nat Rev Neurol 2020;16(8):440–56. Aug.
[8] Yiannopoulou KG, Anastasiou AI, Zachariou V, Pelidou SH. Reasons for failed trials of disease-modifying treatments for Alzheimer disease and their contribution in recent research. Biomedicines 2019;7(4):97. Dec 9.
[9] Pfeffer C, Olsen BR. Editorial: journal of negative Results in biomedicine. J Negat Results Biomed 2002;1(1):2. Nov 12.
[10] Bik EM. Publishing negative results is good for science. Access Microbiol 2024;6(4):000792. Apr 2.
[11] Fanelli D. Negative results are disappearing from most disciplines and countries. Scientometrics 2012;90(3):891–904. Mar 1.
[12] Kearns WG, Stamoulis G, Glick J, Baisch L, Benner A, Brough D, et al. The application of knowledge engineering via the use of a biomimetic digital twin ecosystem, phenotype-driven variant analysis, and exome sequencing to understand the molecular mechanisms of disease. J Mol Diagn 2024;26(7):543–51. July.
[13] Brazil R. Illuminating 'the ugly side of science': fresh incentives for reporting negative results. Nat [Internet] 2024. May 8 [cited 2025 Aug 22]Available from, https://www.nature.com/articles/d41586-024-01389-7.
[14] Park M, Leahey E, Funk RJ. Papers and patents are becoming less disruptive over time. Nature 2023;613(7942):138–44. Jan.
[15] Hanson MA, Barreiro PG, Crosetto P, Brockington D. The strain on scientific publishing. Quant Sci Stud 2024;5(4):823–43. Nov 1.
[16] Wang H, Fu T, Du Y, Gao W, Huang K, Liu Z, et al. Scientific discovery in the age of artificial intelligence. Nature 2023;620(7972):47–60. Aug.
[17] Tu T, Fang Z, Cheng Z, Spasic S, Palepu A, Stankovic KM, et al. Genetic discovery enabled by A large language model [Internet]. bioRxiv, https://www.biorxiv.org/content/10.1101/2023.11.09.566468v1; 2023.
[18] Hulsen T. Literature analysis of artificial intelligence in biomedicine. Ann Transl Med 2022;10(23):1284. Dec.
[19] Athanasopoulou K, Daneva GN, Adamopoulos PG, Scorilas A. Artificial intelligence: the milestone in modern biomedical research. BioMedInformatics 2022;2(4):727–44. Dec.
[20] Kwa T, West B, Becker J, Deng A, Garcia K, Hasin M, et al. Measuring AI ability to complete long tasks [Internet]. arXiv, http://arxiv.org/abs/2503.14499; 2025.
[21] Gao S, Fang A, Huang Y, Giunchiglia V, Noori A, Schwarz JR, et al. Empowering biomedical discovery with AI agents. Cell 2024;187(22):6125–51. Oct 31.
[22] Huang K., Zhang S., Wang H., Qu Y., Lu Y., Roohani Y., et al. Biomni: a general-purpose biomedical AI agent. bioRxiv. 2025 June 2;2025.05.30.656746.
[23] Shah C. From prompt engineering to prompt science with Human in the loop [Internet]. arXiv, http://arxiv.org/abs/2401.04122; 2024.
[24] Naveed H, Khan AU, Qiu S, Saqib M, Anwar S, Usman M, et al. A comprehensive overview of large language models [Internet]. arXiv, http://arxiv.org/abs/2307.06435; 2024.
[25] Pantha N, Ramasubramanian M, Gurung I, Maskey M, Ramachandran R. Challenges in guardrailing large language models for science [Internet]. arXiv, http://arxiv.org/abs/2411.08181; 2024.
[26] Massenon R, Gambo I, Khan JA, Agbonkhese C, Alwadain A. My AI is lying to me": user-reported LLM hallucinations in AI mobile apps reviews. Sci Rep 2025;15(1):30397. Aug 19.
[27] Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 2020;36(4):1234–40. Feb 15.
[28] Luo R, Sun L, Xia Y, Qin T, Zhang S, Poon H, et al. BioGPT: generative pre-trained transformer for biomedical text generation and mining. Br Bioinform 2022;23(6):bbac409. Nov 1.
[29] Bolton E, Venigalla A, Yasunaga M, Hall D, Xiong B, Lee T, et al. BioMedLM: a 2.7B parameter language model trained on biomedical text [Internet]. arXiv, http://arxiv.org/abs/2403.18421; 2024.
[30] Saab K, Tu T, Weng WH, Tanno R, Stutz D, Wulczyn E, et al. Capabilities of Gemini models in medicine [Internet]. arXiv, http://arxiv.org/abs/2404.18416; 2024.
[31] Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature 2023;620(7972):172–80. Aug.
[32] Gao Y, Xiong Y, Gao X, Jia K, Pan J, Bi Y, et al. Retrieval-augmented generation for large language models: a survey [Internet]. arXiv, http://arxiv.org/abs/2312.10997; 2024.
[33] Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc.; 2020. p. 9459–74. NIPS '20.
[34] Han H, Wang Y, Shomer H, Guo K, Ding J, Lei Y, et al. Retrieval-augmented generation with graphs (GraphRAG) [Internet]. arXiv, http://arxiv.org/abs/2501.00309; 2025.
[35] Nisa U, Shirazi M, Saip MA, Pozi MSM. Agentic AI: the age of reasoning—A review. J Autom Intell [Internet] 2025. Aug 28 [cited 2025 Oct 6]Available from, https://www.sciencedirect.com/science/article/pii/S2949855425000516.
[36] Acharya DB, Kuppan K, Divya B. Agentic AI: autonomous intelligence for complex goals—A comprehensive survey. IEEE Access 2025;13:18912–36.
[37] Zhou, J., Chen, S., Wu, Y., Li, H., Zhang, B., Zhou, L., et al. PPML-Omics: a privacy-preserving federated machine learning method protects patients' privacy in omic data. Sci Adv. 2024, 10 (5):eadh8601.
[38] Avraam D, Wilson RC, Aguirre Chan N, Banerjee S, Bishop TRP, Butters O, et al. DataSHIELD: mitigating disclosure risk in a multi-site federated analysis platform. Bioinform Adv 2025;5(1). Mar 10vbaf046.
[39] Escriba-Montagut X, Marcon Y, Anguita-Ruiz A, Avraam D, Urquiza J, Morgan AS, et al. Federated privacy-protected meta- and mega-omics data

analysis in multi-center studies with a fully open-source analytic platform. PLoS Comput Biol 2024;20(12):e1012626. Dec 9.
[40] Federated Data Sharing Appliance full guide - v1.3.32 - resources - Federated Data Sharing Appliance (FDSA) - AD Connect [Internet]. https://community.addi.ad-datainitiative.org/fdsa/m/resources/557; 2024.
[41] McMahan B, Moore E, Ramage D, Hampson S, Arcas BA. Communication-efficient learning of deep networks from decentralized data. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics [Internet]. PMLR; 2017. p. 1273–82 [cited 2025 Oct 6]Available from, https://proceedings.mlr.press/v54/mcmahan17a.html.
[42] Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Sci Rep 2020;10(1):12598. July 28.
[43] Sadilek A, Liu L, Nguyen D, Kamruzzaman M, Serghiou S, Rader B, et al. Privacy-first health research with federated learning. NPJ Digit Med 2021;4(1):132. Sept 7.
[44] Zhang F, Kreuter D, Chen Y, Dittmer S, Tull S, Shadbahr T, et al. Recent methodological advances in federated learning for healthcare. Patterns [Internet] 2024;5(6). June 14 [cited 2025 Oct 6]Available from, https://www.cell.com/patterns/abstract/S2666-3899(24)00131-4.
[45] Xu T, Feng J, Melendez J, Roberts K, Cai D, Zhu M, et al. Addressing accuracy and hallucination of LLMs in Alzheimer's disease research through knowledge graphs [Internet]. arXiv, http://arxiv.org/abs/2508.21238; 2025.
[46] Skarlinski MD, Cox S, Laurent JM, Braza JD, Hinks M, Hammerling MJ, et al. Language agents achieve superhuman synthesis of scientific knowledge [Internet]. arXiv, http://arxiv.org/abs/2409.13740; 2024.
[47] Pu K, Feng KJK, Grossman T, Hope T, Mishra BD, Latzke M, et al. IdeaSynth: iterative research idea development through evolving and composing idea facets with literature-grounded feedback. In: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems [Internet]; 2025. p. 1–31 [cited 2025 July 18]Available from, http://arxiv.org/abs/2410.04025.
[48] Lu C, Lu C, Lange RT, Foerster J, Clune J, Ha D. The AI scientist: towards fully automated open-ended scientific discovery [Internet]. arXiv, http://arxiv.org/abs/2408.06292; 2024.
[49] Wei J, Yang Y, Zhang X, Chen Y, Zhuang X, Gao Z, et al. From AI for science to agentic science: a survey on autonomous scientific discovery [Internet]. arXiv, http://arxiv.org/abs/2508.14111; 2025.
[50] Gottweis J, Weng WH, Daryin A, Tu T, Palepu A, Sirkovic P, et al. Towards an AI co-scientist [Internet]. arXiv, http://arxiv.org/abs/2502.18864; 2025.
[51] Ghareeb AE, Chang B, Mitchener L, Yiu A, Szostkiewicz CJ, Laurent JM, et al. Robin: a multi-agent system for automating scientific discovery [Internet]. arXiv, http://arxiv.org/abs/2505.13400; 2025.
[52] Lahiri AK, Hu QV. AlzheimerRAG: multimodal retrieval augmented generation for clinical use cases using PubMed articles [Internet]. arXiv, http://arxiv.org/abs/2412.16701; 2025.
[53] Naumov V, Zagirova D, Lin S, Xie Y, Gou W, Urban A, et al. DORA AI scientist: multi-agent virtual research team for scientific exploration discovery and automated report generation [Internet]. bioRxiv, https://www.biorxiv.org/content/10.1101/2025.03.06.641840v1; 2025.
[54] Sedler AR, Mitchell CS. SemNet: using local features to navigate the biomedical concept graph. Front Bioeng Biotechnol 2019;7:156.
[55] Swanson K, Wu W, Bulaong NL, Pak JE, Zou J. The Virtual Lab of AI agents designs new SARS-CoV-2 nanobodies. Nature 2025:1–3. July 29.
[56] Ifargan T, Hafner L, Kern M, Alcalay O, Kishony R. Autonomous LLM-driven research — From data to Human-verifiable research papers. NEJM AI 2025;2(1):AIoa2400555. Jan.
[57] Istrate AM, Milletari F, Castrotorres F, Tomczak JM, Torkar M, Li D, et al. rbio1-training scientific reasoning LLMs with biological world models as soft verifiers [Internet]. bioRxiv, https://www.biorxiv.org/content/10.1101/2025.08.18.670981v2; 2025.
[58] Chai J, Tang S, Ye R, Du Y, Zhu X, Zhou M, et al. SciMaster: towards general-purpose scientific AI agents, part I. X-Master Found: Can We Lead Humanity 19s Last Exam? [Internet] 2025. arXiv[cited 2025 Aug 29]Available from, http://arxiv.org/abs/2507.05241.
[59] Luo Y, Shi L, Li Y, Zhuang A, Gong Y, Liu L, et al. From intention to implementation: automating biomedical research via LLMs. Sci China Inf Sci 2025;68(7):170105. June 23.
[60] Roohani Y, Lee A, Huang Q, Vora J, Steinhart Z, Huang K, et al. BioDiscoveryAgent: an AI agent for designing genetic perturbation experiments [Internet]. arXiv, http://arxiv.org/abs/2405.17631; 2025.
[61] Ding N, Qu S, Xie L, Li Y, Liu Z, Zhang K, et al. Automating exploratory proteomics research via language models [Internet]. arXiv, http://arxiv.org/abs/2411.03743; 2024.
[62] Jin R, Zhang Z, Wang M, Cong L. STELLA: self-evolving LLM agent for biomedical research [Internet]. arXiv, http://arxiv.org/abs/2507.02004; 2025.
[63] Andrieu S, Bateman RJ, Bereczki E, Bose N, Brookes AJ, Doraiswamy PM, et al. Harnessing artificial intelligence to transform Alzheimer's disease research. Nat Med 2025;31(5):1384–5. May.
[64] Veitch DP, Weiner MW, Miller M, Aisen PS, Ashford MA, Beckett LA, et al. The Alzheimer's Disease Neuroimaging Initiative in the era of Alzheimer's disease treatment: a review of ADNI studies from 2021 to 2022. Alzheimers Dement 2024;20(1):652–94. Jan.
[65] Daniels A.J., McDade E., Llibre-Guerra J.J., Xiong C., Perrin R.J., Ibanez L., et al. 15 Years of longitudinal genetic, clinical, cognitive, imaging, and biochemical measures in DIAN. medRxiv. 2024 Aug 9;2024.08.08.24311689.
[66] Wagemann O, Liu H, Wang G, Shi X, Bittner T, Scelsi MA, et al. Downstream biomarker effects of Gantenerumab or Solanezumab in dominantly inherited Alzheimer disease: the DIAN-TU-001 randomized clinical trial. JAMA Neurol 2024;81(6):582–93. June 1.
[67] Greenwood AK, Montgomery KS, Kauer N, Woo KH, Leanza ZJ, Poehlman WL, et al. The AD Knowledge Portal: a repository for multi-omic data on Alzheimer's disease and Aging. Curr Protoc Hum Genet 2020;108(1):e105. Dec.
[68] McHugh CP, Clement MHS, Phatak M. AD Workbench: transforming Alzheimer's research with secure, global, and collaborative data sharing and analysis. Alzheimers Dement 2025;21(5):e70278. May 19.
[69] Lovestone S, Imam F. The GNPC provides a proteomic resource for biomarker discovery and mechanistic insight in neurodegenerative disease. Nat Aging 2025;5(7):1181–5. July.
[70] Crawford KL, Neu SC, Toga AW. The Image and Data Archive at the Laboratory of Neuro Imaging. Neuroimage 2016;124:1080–3. Jan 1Pt B.
[71] Beekly DL, Ramos EM, Lee WW, Deitrich WD, Jacka ME, Wu J, et al. The National Alzheimer's Coordinating Center (NACC) database: the Uniform Data Set. Alzheimer Dis Assoc Disord 2007;21(3):249–58.
[72] Leung YY, Lee WP, Kuzma AB, Nicaretta H, Valladares O, Gangadharan P, et al. Alzheimer's Disease Sequencing Project release 4 whole genome sequencing dataset. Alzheimers Dement 2025;21(5):e70237. May.
[73] Beecham GW, Bis JC, Martin ER, Choi SH, DeStefano AL, van Duijn CM, et al. The Alzheimer's Disease Sequencing Project: study design and sample selection. Neurol Genet 2017;3(5):e194. Oct.
[74] Ellis KA, Bush AI, Darby D, De Fazio D, Foster J, Hudson P, et al. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. Int Psychogeriatr 2009;21(4):672–87. Aug.
[75] Pichet Binette A, Gaiteri C, Wennström M, Kumar A, Hristovska I, Spotorno N, et al. Proteomic changes in Alzheimer's disease associated with progressive Aβ plaque and tau tangle pathologies. Nat Neurosci 2024;27(10):1880–91. Oct.
[76] Jessen F, Spottke A, Boecker H, Brosseron F, Buerger K, Catak C, et al. Design and first baseline data of the DZNE multicenter observational study on predementia Alzheimer's disease (DELCODE). Alzheimer 19s Res Ther 2018;10(1):15. Feb 7.
[77] Saunders S, Gregory S, Clement MHS, Birck C, der Geyten S van, Ritchie CW. The European Prevention of Alzheimer's Dementia Programme: an Innovative Medicines Initiative-funded partnership to facilitate secondary prevention of Alzheimer's disease dementia. Front Neurol [Internet] 2022:13. Nov 22 [cited 2025 Aug 29]Available from, https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.2022.1051543/full.
[78] Pérez-González AP, García-Kroepfly AL, Pérez-Fuentes KA, García-Reyes RI, Solis-Roldan FF, Alba-González JA, et al. The ROSMAP project: aging and neurodegenerative diseases through omic sciences. Front Neuroinform 2024;18:1443865. Sept 16.
[79] Sperling RA, Donohue MC, Raman R, Rafii MS, Johnson K, Masters CL, et al. Trial of Solanezumab in preclinical Alzheimer's disease. N Engl J Med 2023;389(12):1096–107. Sept 20.
[80] Rafii MS, Sperling RA, Donohue MC, Zhou J, Roberts C, Irizarry MC, et al. The AHEAD 3–45 study: design of a prevention trial for Alzheimer's disease. Alzheimers Dement 2023;19(4):1227–33. Apr.
[81] Holstege H, Beker N, Dijkstra T, Pieterse K, Wemmenhove E, Schouten K, et al. The 100-plus Study of cognitively healthy centenarians: rationale, design and cohort description. Eur J Epidemiol 2018;33(12):1229–49.
[82] Bianchi DW, Brennan PF, Chiang MF, Criswell LA, D'Souza RN, Gibbons GH, et al. The All of Us Research Program is an opportunity to enhance the diversity of US biomedical research. Nat Med 2024;30(2):330–3. Feb.
[83] Cai Y, Zhou J, Scott PW, Tian Q, Wanigatunga AA, Lipsitz L, et al. Physical activity complexity, cognition, and risk of cognitive impairment and dementia in the Baltimore Longitudinal Study of Aging. Alzheimers Dement (N Y) 2025;11(2):e70077.
[84] Shafto MA, Tyler LK, Dixon M, Taylor JR, Rowe JB, Cusack R, et al. The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. BMC Neurol 2014;14:204. Oct 14.
[85] Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature 2023;613(7944):508–18. Jan.
[86] Fernandez MV, Liu M, Beric A, Johnson M, Cetin A, Patel M, et al. Genetic and multi-omic resources for Alzheimer disease and related dementia from the Knight Alzheimer Disease Research Center. Sci Data 2024;11(1):768. July 12.
[87] Bookheimer SY, Salat DH, Terpstra M, Ances BM, Barch DM, Buckner RL, et al. The Lifespan Human Connectome Project in Aging: an overview. Neuroimage 2019;185:335–48. Jan 15.
[88] Schwarz CG, Kremers WK, Prakaashana CM, Przybelski SA, Christenson LR, Wiliams JM, et al. A large public release of clinical and imaging data from the Mayo Clinic study of aging. Alzheimers Dement 2025;20(Suppl 9):e093966. Jan 9.
[89] Nilforooshan R., Barnaghi P. The RESILIENT dataset: multimodal monitoring of ageing-related comorbidities and cognitive decline [Internet]. Zenodo; 2025 [cited 2025 Aug 29]. Available from: https://zenodo.org/records/16755408.
[90] Huang X, Han X, Chang H, Yu T, Dong Y, Mao M, et al. Associations between trajectories of plasma biomarkers for Alzheimer's disease, brain structures, and cognitive function: a prospective cohort study in the UK Biobank. Mol Psychiatry 2025. Aug 28.

[91] Palchuk MB, London JW, Perez-Rey D, Drebert ZJ, Winer-Jones JP, Thompson CN, et al. A global federated real-world data and analytics platform for research. JAMIA Open 2023;6(2). Julyooad035.
[92] Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016;3:160011. Mar 3.
[93] Lentzen M, Vairavan S, Muurling M, Alepopoulos V, Atreya A, Boada M, et al. RADAR-AD: assessment of multiple remote monitoring technologies for early detection of Alzheimer's disease. Alzheimers Res Ther 2025;17(1):29. Jan 27.
[94] Palermo F, Chen Y, Capstick A, Fletcher-Loyd N, Walsh C, Kouchaki S, et al. TIHM: an open dataset for remote healthcare monitoring in dementia. Sci Data 2023;10(1):606. Sept 9.
[95] Tomczyk P, Brüggemann P, Mergner N, Petrescu M. Are AI tools better than traditional tools in literature searching? Evidence from E-commerce research. J Librariansh Inf Sci 2024. Nov 1509610006241295802.
[96] Lau O, Golder S. Comparison of elicit AI and traditional literature searching in evidence syntheses using four case studies. Cochrane Evid Synth Methods 2025;3(6):e70050. Nov.
[97] Wang Z, Cao L, Jin Q, Chan J, Wan N, Afzali B, et al. A foundation model for human-AI collaboration in medical literature mining [Internet]. arXiv, http://arxiv.org/abs/2501.16255; 2025.
[98] Liang W, Zhang Y, Cao H, Wang B, Ding DY, Yang X, et al. Can large language models provide useful feedback on research papers? A large-scale empirical analysis. NEJM AI 2024;1(8):AIoa2400196. July 25.
[99] Qi B, Zhang K, Tian K, Li H, Chen ZR, Zeng S, et al. Large language models as biomedical hypothesis generators: a comprehensive evaluation [Internet]. arXiv, http://arxiv.org/abs/2407.08940; 2024.
[100] Cutillo CM, Sharma KR, Foschini L, Kundu S, Mackintosh M, Mandl KD. Machine intelligence in healthcare—Perspectives on trustworthiness, explainability, usability, and transparency. npj Digit Med 2020;3(1):47. Mar 26.
[101] Gabriel I, Manzini A, Keeling G, Hendricks LA, Rieser V, Iqbal H, et al. The ethics of advanced AI assistants [Internet]. arXiv, http://arxiv.org/abs/2404.16244; 2024.
[102] Tang X, Jin Q, Zhu K, Yuan T, Zhang Y, Zhou W, et al. Risks of AI scientists: prioritizing safeguarding over autonomy [Internet]. arXiv, http://arxiv.org/abs/2402.04247; 2025.
[103] Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell 2019;1(9):389–99. Sept.
[104] Price WN, Cohen IG. Privacy in the age of medical big data. Nat Med 2019;25(1):37–43. Jan.
[105] McMurry JA, Juty N, Blomberg N, Burdett T, Conlin T, Conte N, et al. Identifiers for the 21st century: how to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. PLoS Biol 2017;15(6):e2001414. June 29.
[106] Davis KR, Peabody B, Leach P. Universally unique IDentifiers (UUIDs) [Internet]. Internet Eng Task Force 2024. May [cited 2025 Oct 6]. Report No.: RFC 9562. Available from, https://datatracker.ietf.org/doc/rfc9562.
[107] Hofmann B. Biases in AI: acknowledging and addressing the inevitable ethical issues. Front Digit Health 2025;7:1614105.
[108] Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3(1):160018. Mar 15.