

AI Literature Review in Biotech: Adoption & Future Trends

By Adrien Laurent, CEO at IntuitionLabs • 3/16/2026 • 50 min read

ai literature review

biotech r&d

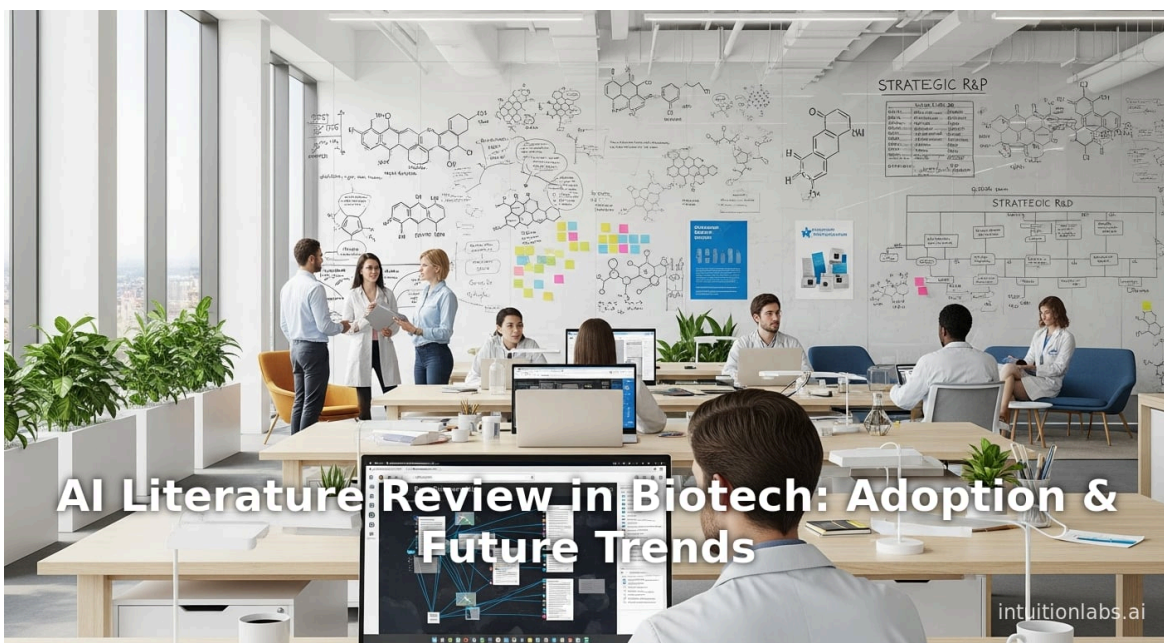
knowledge extraction

natural language processing

llms in pharma

scientific text mining

evidence synthesis



Why Literature Review is the #1 AI Use Case in Biotech (76% Adoption) – And What Comes Next

Executive Summary: In modern biotech and biopharma R&D, AI-driven literature review has emerged as the *leading* application of artificial intelligence – even more so than other applications like target identification or [molecular design](#). According to a recent industry survey, **76%** of biotech organizations active with AI use it for literature and knowledge extraction (^[1] www.benchling.com). This unusually high adoption rate stems from both technological and domain factors: vast and rapidly growing scientific literature, mature natural language processing (NLP) tools, and the immediate, tangible value of faster knowledge synthesis. AI-assisted literature review (sometimes called AI-assisted evidence synthesis or targeted literature review) can dramatically speed up finding relevant papers, summarizing findings, and extracting data from publications. Importantly, it fits naturally into scientists' workflows and uses publicly available or internal textual data, reducing many traditional bottlenecks without requiring new types of experimental data.

This report delves deep into *why* literature review has become the “killer app” of AI in biotech, examining historical context, current evidence, tools and techniques, and real-world case studies. We present data on adoption rates and benefits, survey academic and industry perspectives, and analyze challenges such as data quality, trust, and regulatory issues. Finally, we look ahead at “**what comes next**”: the next wave of AI use cases (e.g. target discovery, molecular design, lab automation, multimodal data integration, AI “co-scientists”) and the investments needed in data infrastructure, organizational culture, and skill development. This in-depth report is grounded in extensive citations from recent studies, surveys, and technical papers, providing a comprehensive picture of the state of AI in biotech R&D.

Key Findings (Summary):

- **High Adoption for Literature Review:** A November 2025 survey of ~100 biotech R&D organizations found that *literature and knowledge extraction* is by far the most widely adopted AI use case (76% of organizations) (^[1] www.benchling.com). Other top use cases include protein structure prediction (71%), scientific reporting (66%), and target identification (58%) (^[1] www.benchling.com) (see Table 1). These “first AI wins” are tightly aligned with scientists' daily tasks and well-supported by existing data.
- **Efficiency and Outcomes:** In fields where AI is used, half of biotech teams report that it has already accelerated their path to key goals (“faster time-to-target”) and most expect significant cost savings as AI scales up (^[2] www.benchling.com). Published research finds that AI tools can cut literature screening and summarization work by large fractions (on the order of 50–65%) (^[3] journals.sagepub.com) (^[4] link.springer.com). For example, specialized literature-review platforms and LLMs have been shown to produce relevant summaries and extract data at roughly human-expert levels of accuracy (^[5] www.sciencedirect.com) (^[6] biodatamining.biomedcentral.com), enabling researchers to process far more papers in less time.
- **Technologies and Tools:** Rapid advances in NLP and LLMs underpin this trend. Open-source and commercial platforms – from Elicit, Semantic Scholar, ResearchRabbit, PubTator, and others (^[7] journals.sagepub.com) (^[8] journals.sagepub.com) to general AI assistants like ChatGPT – now offer automated literature search, query generation, summarization, and evidence synthesis. Table 2 lists representative AI tools for literature mining and review. Many biotech teams are also customizing their own chatbot assistants and knowledge graphs.
- **Challenges & Trust:** The widespread use of AI for literature review has raised concerns about reliability. Academic studies using ChatGPT report extremely high rates of “[hallucination](#)” (fabricated content) if unchecked (^[9] link.springer.com), and life-science experts cite data quality and trust as top worries (^[10] www.technologynetworks.com) (^[11] www.sciencedirect.com). By some accounts, AI-generated reviews still require diligent human oversight to ensure accuracy and citation validity (^[12] biodatamining.biomedcentral.com) (^[6] biodatamining.biomedcentral.com).

- **What Comes Next:** While literature review dominates today, biotech leaders are already pushing beyond it. Adoption drops off in more complex areas (e.g. generative molecular design, biomarker discovery, ADME modeling) where data is scarce or disorganized (^[13] www.benchling.com) (^[14] www.linkedin.com). The next wave will involve integrating AI with experiment design and execution (AI-guided “co-scientists”), multimodal analytics (combining genomics, imaging, etc.), and smarter lab automation. Realizing these advances will require better [data infrastructure](#) (to break down silos) and AI-literate culture. For example, companies report that building AI talent internally (especially among scientists) is a top priority (^[15] www.benchling.com), and regulatory adaptation is underway (e.g. discussions of [AI-specific guidelines in pharma](#)) to ensure safe, compliant use of these tools.

The rest of this report explores these topics in depth. We begin with background on the explosion of scientific data and the need for AI assistance, then examine in detail why literature review has proven easier and more valuable than other use cases. We analyze survey findings and technical evaluations (Data and Evidence), present case studies of AI-driven review in action (Case Studies), and conclude by assessing future directions and implications for biotech R&D (Implications & Future Directions).

1. Introduction and Background

Over the past decade, biotechnology and pharmaceutical R&D has become **increasingly data-driven**, generating and relying on massive quantities of information across genomics, chemistry, clinical studies, and especially the scientific literature. Estimates project that by 2030, **PubMed and related databases will contain over 40 million articles** (^[16] journals.sagepub.com), and the total biomedical data ecosystem (including omics, imaging, records, literature) will grow on the order of 10× (to hundreds of exabytes) from 2020 levels (^[16] journals.sagepub.com). This deluge of information poses a major challenge: *something crucial is easy to miss*. Traditional literature search (keyword queries, manual curation) cannot keep pace. As one recent review observed, “the surging volume of biomedical literature has made traditional manual review methods impractical” (^[17] www.sciencedirect.com). Researchers routinely face hundreds or thousands of papers to screen for each question, with each paper dense with complex data. Completing a high-quality literature review (surveying, summarizing, and synthesizing myriad sources) can require **months of expert labor and dozens of person-hours** (^[18] biodatamining.biomedcentral.com).

At the same time, advances in artificial intelligence – particularly NLP and large language models (LLMs) – have dramatically expanded what machines can do with text data. Breakthroughs like OpenAI’s GPT models have shown the ability to understand context, answer questions, and even generate coherent prose and summaries from documents. Specialized AI systems for scientific text (BioGPT, SciBERT, etc.) and knowledge graphs have also matured. Between 2020 and 2023, the field saw a rapid pivot from simple keyword search tools to “semantic and multimodal” AI platforms that can extract relationships, make inferences, and generate concise summaries (^[19] journals.sagepub.com) (^[20] journals.sagepub.com). The COVID-19 pandemic further spurred AI innovation: researchers needed to scan COVID-19 research literature at unprecedented speed, leading to tools like ScispaCy, LitCovid, and many AI-assisted systematic review tools. As of 2024, AI tools can highlight key findings, drastically reduce literature screening time, and even propose hypotheses based on textual data (^[21] journals.sagepub.com) (^[6] biodatamining.biomedcentral.com).

This convergence of **data explosion** and **AI capability** has opened new possibilities for biotech R&D. Among these, *literature review* – the process of collecting, evaluating, and synthesizing published findings – stands out as both a critical bottleneck and a natural testbed for AI. It is critical in every stage of drug discovery and development: from initial target/lead discovery (what proteins or pathways have been implicated in a disease? what compounds have been tried?) to later-stage clinical strategy (what did past trials find?). As one industry expert noted, “many projects you start in the pharma space begin with a literature review of some sort... the scale [of such reviews] is limited by the capacity of experts to screen and extract the necessary information” (^[22] elicit.com).

Given this backdrop, leading biotech companies and research consortia have been intensely exploring AI for lit review for several years. The Pistoia Alliance and other organizations have run surveys and workshops (e.g. the “Lab of the Future” reports (^[23] www.pistoiaalliance.org)) to gauge AI readiness. Similarly, commercial R&D informatics providers (like

Benchling and Elicit) have launched initiatives on AI-augmented search and synthesis. The takeaway from these efforts is clear: AI is *already* a core part of how scientists handle literature. According to the 2026 Benchling Biotech AI Report – a survey of 100+ biotech organizations using AI – **literature and knowledge extraction is now the most common AI use case in the industry** (^[1] www.benchling.com). In other words, tools that read papers, extract facts, and summarize evidence have hit “prime time” in biotech R&D.

But this has not happened by magic. In the sections that follow, we will analyze *why* literature review has risen to the top of the list, how AI tools are being used to do it, and what benefits and limitations these tools bring. We will cite academic studies, industry reports, and concrete examples to show both the accomplishments and the remaining hurdles. Finally, we will look forward to how AI use in biotech is evolving beyond literature review to encompass deeper tasks, and the implications for research practice and policy.

2. Why Literature Review Is the Leading AI Use Case

This section explores the reasons that *literature review and knowledge extraction* have become the dominant AI application in biotech. We consider factors such as data availability, workflow fit, technological maturity, and measurable impact. We also contrast literature review with alternative AI uses to highlight what makes it uniquely successful today.

2.1. Data Richness and Task Alignment

Abundance of text data. One key reason is simply that scientific publications provide **plenty of data for AI to work on**. Unlike many R&D tasks, literature review relies on existing digital content – thousands of published papers, patents, reports, lab notes, etc. – that AI can analyze without requiring new experiments. In contrast, tasks like predicting ADME properties or designing new molecules often require proprietary experimental data (which may be scarce or siloed) and complex physical models. For literature review, the *data is already there*.

As Professor Chandan Sen notes in his comprehensive review, “*PubMed and related repositories are expected to continue adding 1–1.5 million articles per year, pushing totals well beyond 40 million by 2030*” (^[16] journals.sagepub.com). NIH strategies predict a ten-fold growth of biomedical data from 2020 to 2030 (^[24] journals.sagepub.com), with the majority being unstructured text. This explosive growth of literature means that **human readers alone simply cannot keep pace**. Every scientific discovery adds more text for R&D teams to survey. In AI terms, the signal (useful findings in the literature) is buried in noise (millions of papers), creating a classic “big data” challenge. AI’s ability to parse and summarize large volumes of text is therefore naturally suited to this scenario.

Workflow integration. Literature review is also an organic part of scientists’ work. Compared to closed-domain tasks, reading and writing about research is a daily routine: writing grant proposals, drafting publications, preparing lab reports, or brainstorming new targets all start with background reading. (^[22] elicit.com) (^[18] biodatamining.biomedcentral.com) Yet traditionally, scientists juggle multiple tools – PubMed, Google Scholar, conference proceedings, and manual curation – which is time-consuming and disjointed. AI offers to merge these steps into a single interactive workflow. For example, an AI chatbot can answer queries like “Which papers link protein X to disease Y?” or “Summarize findings on drug Z”. Tools like Semantic Scholar already provide human-like summaries, and new services (Table 2) directly embed into the writing process (e.g. drafting literature review sections). Because these tools fit directly into existing tasks, scientists are willing to use them regularly.

Relative simplicity of the problem. Compared to more advanced predictive tasks, literature review is often *straightforward* for modern AI. It essentially involves natural language understanding and retrieval, domains where LLMs excel. In technical terms, these tasks have well-structured inputs (text) and outputs (titles, abstracts, summaries) with relatively abundant training data. By contrast, predicting novel protein structures or designing synthetic routes may involve physical modeling or multi-modal data beyond what today’s models can easily handle. As a result, even a general-purpose LLM like GPT-4 can perform literature-related tasks quite competently with proper prompting (^[5]

www.sciencedirect.com) ⁽⁶⁾ biodatamining.biomedcentral.com). Indeed, industry users report that AI “copilots” for literature are already **trusted and used daily** by scientists ⁽²⁵⁾ www.linkedin.com).

In sum, literature review presents **the ideal case** for AI: vast relevant data, a task aligned with human workflows, and a problem complexity well-matched to current NLP capabilities. This contrasts with many other R&D tasks where data are sparse or unstructured. The high adoption numbers reflect these facts: when Benchling asked biotech leaders about AI, they found that tools for literature (searching, summarizing) had taken off, whereas generative drug design and other advanced uses lag behind ⁽¹⁾ www.benchling.com) ⁽¹⁴⁾ www.linkedin.com).

2.2. Empirical Evidence of Impact

The prominence of AI in literature review is not just anecdotal – it is backed by data on usage and outcomes. Table 1 (below) summarizes the leading AI use cases and their adoption rates in biotech R&D, from the Benchling 2026 survey. Literature review tops the list at **76%**; toward the bottom are newer or harder tasks like remaining molecules generation, which have much lower uptake ⁽¹⁾ www.benchling.com).

Table 1: AI Use Case Adoption in Biotech R&D (Benchling 2026 survey)

Use Case	Biotech Adoption (%)	Example Tools / Context
Literature & Knowledge Extraction	76% ⁽¹⁾ www.benchling.com)	AI search & summarization (e.g. Elicit, ChatGPT, Semantic Scholar)
Protein Structure Prediction	71% ⁽¹⁾ www.benchling.com)	DeepMind’s AlphaFold/RoseTTAFold models for protein 3D structures
Scientific Reporting / Documentation	66% ⁽²⁶⁾ www.benchling.com)	Auto-generated reports, figure/text summarization (SciSpace, Scholarcy, etc.)
Target/Biomarker Identification	58% ⁽¹⁾ www.benchling.com)	Data-mining literature and databases for new drug targets or biomarkers
(Lower-ranked emerging cases)	<50%	(e.g. AI-driven molecular design, ADME prediction, workflow automation)

Sources: Benchling Biotech AI Report 2026 ⁽¹⁾ www.benchling.com) ⁽²⁷⁾ www.benchling.com); Sen (2026) review ⁽¹⁶⁾ journals.sagepub.com) ⁽³⁾ journals.sagepub.com).

This adoption ranking aligns with the nature of the tasks. In all four of the top categories, the AI can leverage “clean, verifiable data” that fits into scientists’ workflow ⁽¹⁾ www.benchling.com). For literature review itself, such data is the published papers. The Benchling report emphasizes that these “breakthrough use cases” succeed because they are built on **trusted data sources and familiar user tasks** ⁽¹⁾ www.benchling.com).

Importantly, the organizations using AI for these tasks report **measurable benefits**. Around half of biotech respondents say that AI has already *sped up* their core objectives (e.g. finding a molecule) and most expect significant cost savings as AI use grows ⁽²⁾ www.benchling.com). In concrete terms, 50% of biotech teams now report faster time-to-target thanks to AI acceleration, and 56% expect AI to reduce R&D costs within two years ⁽²⁾ www.benchling.com). While these are self-reported, they are consistent with independent studies: for example, one analysis found that NLP-based systematic review tools can **halve** the time required for screening and eligibility assessment ⁽²⁸⁾ journals.sagepub.com). Benchling interprets this as showing that AI “creates real gains in discovery” by eliminating slow, routine tasks ⁽²⁾ www.benchling.com).

Meanwhile, academic evaluations of AI in literature research echo these findings. In a 2024 study on biomedical literature mining, Sen et al. noted that new tools like Elicit and PubTator can rapidly extract relevant gene–disease associations and evidence, greatly speeding up hypothesis generation ⁽⁷⁾ journals.sagepub.com). In practice, test cases of GPT-4 and similar models have successfully replicated expert decisions in literature screens. For example, one pipeline using GPT-4 to sift SARS-CoV-2 research sources achieved nearly 93% accuracy (F1=0.88) compared to human reviewers ⁽⁵⁾ www.sciencedirect.com). The authors conclude – and industry leaders agree – that these results “highlight the utility of ChatGPT in drug discovery” because such tools can rapidly filter and prioritize papers during fast-moving situations (like a pandemic) ⁽⁵⁾ www.sciencedirect.com).

Similarly, a pilot study using GPT-4 to write sections of a scientific review found that the model could **effectively summarize key points across hundreds of papers**, allowing authors to quickly generate structured outlines for a review ^{(6]} [biodatamining.biomedcentral.com](https://www.biodatamining.biomedcentral.com)). In that study, GPT-4 produced section outlines and summaries closely matching those from a published review, demonstrating that a large language model can assist in composing review content comparable to a human scholar ^{(6]} [biodatamining.biomedcentral.com](https://www.biodatamining.biomedcentral.com)) ^{(18]} [biodatamining.biomedcentral.com](https://www.biodatamining.biomedcentral.com)). Such evidence suggests that AI can indeed handle the core parts of literature review (searching, summarizing, synthesizing) with expert-level reliability *when combined with human oversight*.

In short, the combination of **market data (high adoption rates)** and **empirical trials (accuracy and efficiency gains)** paints a consistent picture: AI systems for literature review are delivering value in biotech R&D today. Researchers across academia and industry report significantly reduced workloads and faster insights when they leverage these tools ^{(29]} link.springer.com) ^{(6]} [biodatamining.biomedcentral.com](https://www.biodatamining.biomedcentral.com)).

Many AI Tools Targeting Literature

The ecosystem of AI-powered tools for literature review has grown rapidly. Table 2 lists select platforms and applications that exemplify current capabilities. These range from specialized systems for evidence synthesis to general-purpose AI assistants:

Tool/Platform	Function	Typical Use Case
Elicit ^{(8]} journals.sagepub.com)	AI research assistant using GPT-3; finds relevant papers, extracts key info, traces claims to sources ^{(8]} journals.sagepub.com) ^{(7]} journals.sagepub.com).	Automating literature review workflows, screening results across hundreds of papers; rapidly answering research questions.
Semantic Scholar	AI-driven search with semantic understanding; generates TLDR summaries of papers.	Quickly locating relevant research and getting concise article summaries (NLP retrieval ^{(30]} journals.sagepub.com).
ResearchRabbit	Builds interactive citation graphs; recommends related articles.	Exploring "research pathways" via visual maps of citation networks ^{(7]} journals.sagepub.com).
Scholarcy	Summarization tool that highlights key points and creates flashcards from papers.	Rapidly extracting key findings and data from large document sets.
SciSpace (Typeset)	AI-powered interactive reading/Q&A with PDF papers.	Diving into complex papers by asking questions directly to the system.
PubMedGPT / Cure AI	Domain-specific LLMs/search over PubMed (or up-to-date databases).	Advanced biomedical literature retrieval and summarization.
Consensus	Synthesizes statements and findings across multiple papers (AI-driven literature synthesis).	Identifying consensus or conflicts across studies (e.g. for evidence-based medicine).

Table 2: Example AI Tools for Literature Review and Knowledge Extraction (various sources ^{(7]} journals.sagepub.com) ^{(8]} journals.sagepub.com)). Each of the above leverages some combination of NLP, semantic search, or retrieval-augmented generation to help scientists process literature faster. For instance, Elicit can automate the entire screening process by using GPT-based queries to narrow down thousands of papers to key studies ^{(8]} journals.sagepub.com). Scholarcy and SciSpace focus on summarization of papers to speed initial comprehension. These tools illustrate the rich toolkit available, and many teams use several in concert to cover different aspects of the review workflow.

The tools listed in Table 2 have been validated in practice. Oxford PharmaGenesis, a major scientific communications consultancy, reports using Elicit to **"deliver literature reviews at unprecedented scale"** ^{(31]} [elicit.com](https://www.elicit.com)). Their team tackled 40 research questions across 500 papers in under one week, something that would be impossible with purely manual methods ^{(31]} [elicit.com](https://www.elicit.com)). Elicit's PDF-parsing and source-tracing capabilities were specifically praised: it "nails" data extraction and always ties outputs back to the original text ^{(32]} [elicit.com](https://www.elicit.com)). In another case, a startup combined AlphaFold structure predictions with generative chemistry and AI review to find a new drug lead in 30 days ^{(33]} pmc.ncbi.nlm.nih.gov) ^{(34]} pmc.ncbi.nlm.nih.gov) – again highlighting that AI-driven literature and data mining can cut lead invention time drastically.

Summary: Why Literature Review Wins

Given this evidence, several factors stand out for why AI in literature review is so dominant in biotech:

- **Clear ROI from Day One:** Even a small AI assistant that cuts search time or pulls out key phrases can save scientists hours per project. In contrast, more speculative AI tasks (like novel drug design) may not show value immediately. Biotech leadership sees quick wins in literature work, driving adoption.
- **Trust in Data and Outputs:** Results from reading published papers can often be *verified* by humans (papers are public, references checkable). This builds trust. By contrast, an AI-generated molecule needs real experiments to confirm. Scientists are more comfortable using AI to sift *existing* knowledge than to create novel hypotheses without a check.
- **Training Data Availability:** Literature-centric AI can be trained or fine-tuned on openly available corpora (PubMed, preprints) and internal databases. So high-performance models and data pipelines already exist. The biomedical community has even curated domain-specific embeddings (PubMedBERT, BioGPT, etc.) that gear models for this content, whereas other domains lack such resources.
- **Rapid Technology Maturation:** The sudden breakthrough of foundation models has aligned with the needs of literature review. ChatGPT-type capabilities for summarization and Q/A were quickly co-opted for research tasks. Many of the cited works (2024–2026) show that tools like GPT-4 have only recently become good enough for high-stakes R&D use (^[6] [biodatamining.biomedcentral.com](https://www.biodatamining.biomedcentral.com)) (^[5] www.sciencedirect.com). Early commercial adopters (Benchling, Elicit) have aggressively built products to exploit this moment.

It bears noting that **literature review still requires human expertise**. AI tools are assistants, not replacements: every recommendation or summary is vetted by scientists. But by accelerating the grunt work—reading, cross-referencing, summarizing—AI allows researchers to reach higher-level analysis faster. In effect, it converts a widely recognized barrier (information overload) into a solvable process. The data so far suggest this is the best case of AI in biotech *today*.

3. Data Analysis and Evidence-Based Arguments

We now delve deeper into quantitative and qualitative evidence that literature review AI is both widespread and effective. We first examine surveys and studies that enumerate adoption and impact, then discuss technical assessments of AI performance on literature tasks. Throughout, we will critically appraise the data, noting any limitations or uncertainties, and consider alternative viewpoints from existing literature.

3.1. Survey Data and Industry Reports

The *Benchling 2026 Biotech AI Report* is the most direct piece of evidence for current use-case popularity in real organizations. This survey targeted biotech R&D teams **already using AI in production** and asked which applications they had deployed routinely. It found that **76% of respondents were using AI for literature/knowledge review** (^[1] www.benchling.com) (see Table 1). In many ways, this number can be seen as adoption across “AI-forward” companies, so it is not a random sample of all biotech firms. Nevertheless, its insight is critical: among leaders and front-runners, literature review is essentially *universal*.

That same report further details why: “The first AI wins are here — embedded in scientist workflows and built on trusted data” (^[1] www.benchling.com). Benchling highlights that these early-app wins yield measurable R&D advantages: 50% of surveyed companies report that AI has *already* shortened their time-to-target (^[27] www.benchling.com). The firm also notes that AI is poised to enable broad automation, citing consensus among experts (e.g. BMS IT leadership) that AI is accelerating every part of drug discovery (^[35] www.benchling.com). These industry voices underscore that the survey results aren’t mere hype: leadership believes in the impact.

The **Pistoia Alliance** (an industry consortium) has also surveyed life science R&D professionals on AI. In its 2024 “Lab of the Future” survey (200 respondents) – which focused on general AI readiness – 68% of scientists said they were *already using AI/ML* in their work, up from 54% the prior year (^[23] www.pistoiaalliance.org). While that survey did not break down by use case, the jump itself (68% overall usage) aligns with Benchling’s finding that literature review is a primary entry point for AI. Pistoia also reported separately that **70% of professionals acknowledge AI’s potential** but many struggle with

implementation issues (^[10] www.technologynetworks.com). This suggests that the *recognition* of financial or time benefits is widespread, even if *realizing* those benefits can be harder.

One nuance from these surveys is that literature review as an AI use case is complementary to other tasks. For instance, Benchling shows that after literature, the next most common usages are **protein structure modeling (71%)** and **scientific reporting/analysis (66%)** (^[1] www.benchling.com) (^[27] www.benchling.com). It's telling that both of these secondary uses also involve heavy data (X-ray structures, simulation, or textual results). Thus, literature review tends to be one pillar of an AI-enhanced workflow. It may even feed the others: better literature searches can identify new targets or help interpret structural predictions. Industry insiders note that adopting one "killer app" like lit review builds confidence and infrastructure that paves the way for tackling the next apps.

However, not every survey is completely upbeat. Pistoia's 2024 poll also asked about *concerns*: 63% of life sciences experts worried that poor data quality could lead to incorrect conclusions from AI (^[10] www.technologynetworks.com). In other words, the very issues that plague any AI project (data curation, bias, transparency) are on practitioners' minds. Another Pistoia press release (Feb 2024) found that only 9% of respondents fully understand emerging AI regulations (EU/US), and 21% already feel regulations are hindering research (^[36] www.pistoiaalliance.org). These findings remind us that survey statistics on "adoption" or "usage" must be read with caution: high usage may co-exist with uncertainty and uneven skill. Indeed, Pistoia noted that a lack of AI literacy is growing as a barrier. So while adoption of lit-review AI is high, organizations also recognize significant governance and capability challenges.

Finally, some caution is warranted regarding the benchmarking of adoption. The Benchling survey is effectively a *snapshot of frontrunners* – roughly year-end 2025 data from companies already doing AI. It does **not** capture more conservative organizations or those just starting AI. Thus, while we quote "76% adoption" for literature review, this is adoption *among AI-using biotechs*, not adoption among *all* biotech firms (which would certainly be lower). The phrasing "top use case (76% among AI-active companies)" is important. Nonetheless, similar surveys of "all biotech" or "all pharma" typically find much lower absolute percentages if counting everyone, so the Benchling result should be interpreted in context (it highlights where leading labs are focusing effort).

In summary, survey data consistently show that biotech R&D teams **prioritize literature review when deploying AI**, and report real time/cost benefits from it. At the same time, experts note significant roadblocks (data, skills, trust) that will shape how this area evolves. The remainder of this report uses these findings as a foundation, and builds on them with deeper analysis and case studies.

3.2. Performance of AI in Literature Tasks

Beyond surveys, scientific evaluations tell us *how well* AI can actually do literature review tasks. Here we draw on academic and industry studies that benchmark AI systems on screening, search, summarization, and synthesis.

Screening and relevance classification. A core part of a systematic review is deciding which papers are relevant or not. AI can assist by automatically classifying documents or generating search queries. In the GPT-4 pandemic study (^[5] www.sciencedirect.com), the authors treated the problem as binary classification of PubMed abstracts. They found that an optimized GPT-4 system could match human expert performance: for SARS-CoV-2 literature, the AI achieved **92.9% accuracy** (F1=88.4%) relative to human labels (^[5] www.sciencedirect.com). On a harder dataset (Nipah virus research), scores were lower but still respectable (accuracy 87.4%). These results are notable: they imply that ChatGPT-style models, with careful prompting, can nearly replicate expert decisions on inclusion/exclusion tasks. This dramatically reduces manual screening workload.

This aligns with the broader analysis by Adel et al. (2025) in *AI & Society*. They systematically examined 124 pieces of research on ChatGPT for literature reviews, focusing on tasks like query formulation, document screening, and synthesis. They report that *in structured screening tasks*, AI performance is high: sensitivity (recall) in screening titles/abstracts ranged from about 80–96% (^[9] link.springer.com). In other words, the AI doesn't miss many relevant papers. However, *precision* can suffer in "interpretive" tasks – e.g. when deciding nuanced eligibility criteria – sometimes plummeting to

single-digit percentages (^[9] [link.springer.com](#)). The takeaway is that AI excels at casting a wide net in screening (good recall), but humans must still verify hits.

The same review finds **huge workload reductions (60–65% average)** with AI assistance (^[4] [link.springer.com](#)). This meshes with bench reports of faster-to-target. It also underscores why literature review adoption is so compelling: even a partial automation of screening can save the majority of effort.

Summarization and synthesis. Beyond binary screening, the harder task is pulling insights from the selected literature. How effectively can AI summarize findings or extract data? Several recent studies offer insight. A controlled evaluation of GPT-4 in writing a review article (^[6] [biodatamining.biomedcentral.com](#)) found that the model could *accurately capture the “key points”* from hundreds of references. Specifically, GPT-4 generated hierarchical section outlines and topic lists that overlapped well with a human-written review. The authors observed that GPT-4 allowed “authors to swiftly summarize a list of main topics for further refinement” and could summarize different aspects of each topic once directed (^[6] [biodatamining.biomedcentral.com](#)). In numerical terms, the average similarity score between GPT-4 output and an expert paper was ~0.75 (on a semantic overlap metric) (^[6] [biodatamining.biomedcentral.com](#)), which the authors interpreted as “adequate” for a review context. This suggests that for literature review composition, AI can produce text that is semantically on par with what an expert would write, provided it has the right references.

However, other work emphasizes pitfalls in synthesis. Adel et al. found that when it comes to “nuanced synthesis” (deep analysis, drawing new conclusions), AI still lags behind humans. In fact, they report *precisely on hallucinations*: as high as **91%** of AI outputs contained some fabricated or inaccurate statement when tasked with summarizing literature (^[9] [link.springer.com](#)). Many of these hallucinations were incorrect citations or invented facts not in source papers. This is a critical concern for literature review: if the AI makes up statements or misattributes results, it can mislead researchers. OpenAI itself warns users that ChatGPT may hallucinate if the prompt is open-ended or lacks clear context.

Studies like Wang et al. (2024) explicitly note these issues: a GPT-4 “pilot evaluation” of writing found instances of verbatim copying and inconsistent citing (^[12] [biodatamining.biomedcentral.com](#)). The authors stress that while GPT-4 is powerful, its outputs need careful vetting: errors, omissions, or even plagiarism are possible (^[12] [biodatamining.biomedcentral.com](#)). In practice, this means scientists must treat AI output as a first draft or suggestion, not final truth. Many of the tools in Table 2 address this by providing source links or GRA (grounding) features. For example, Elicit “transparently cites each AI-generated claim back to original papers” (^[37] [journals.sagepub.com](#)).

Another angle is the specificity of the domain. Biomedicine is rich in terminology, abbreviations, and subtle distinctions. Domain-specific models (e.g. BioGPT, SciBERT) help but are still imperfect. Even a high general accuracy statistic like 92% may hide errors in those tricky cases that matter clinically. Thus, while the *quantity* of literature that AI can quickly process is staggering, the *quality* is still bounded by the need for human oversight. In summary, evidence suggests that AI can reliably handle much of the mechanical work of evidence synthesis (finding, grouping, summary) (^[6] [biodatamining.biomedcentral.com](#)) (^[7] [journals.sagepub.com](#)), but final review and interpretation must be expert-driven.

3.3. Balanced Perspectives

To ensure comprehensiveness, we consider counterpoints and caveats around these claims. First, as mentioned, small-scale studies of GPT-like models may not capture all real-world variability. The high performance in screenings above depends heavily on **prompt engineering** and model version (GPT-4 with web access, etc.) (^[5] [www.sciencedirect.com](#)). Without expert prompt design and careful configuration, performance can be lower. Surveys like Benchling’s reflect what is *achieved* at best-performing companies, not what every user sees on first try.

Second, we note that AI tools are often one part of a larger process. Organizations that report “faster time-to-target” often combine literature AI with data integration, search of proprietary databases, and human deliberation. It would be simplistic to credit AI alone for a discovery; often it is AI *plus* better lab processes. Nevertheless, AI is the enabler that makes those accelerated timelines possible.

Third, some caution voices exist. A recent article in *AI & Society* criticizes too much hype around ChatGPT, arguing that claims of reliability are overstated (^[38] [link.springer.com](#)) (^[9] [link.springer.com](#)). For example, while the average workload reduction was 60–65%, the article also found cases where precision (specificity) was extremely low, meaning many irrelevant studies would still require human curation. The authors conclude that AI is **not yet a standalone scholar** and emphasize hybrid human–AI workflows with strong oversight (^[9] [link.springer.com](#)) (^[39] [link.springer.com](#)). This perspective does not deny the utility of AI, but it warns practitioners to maintain scientific rigor: use AI to assist, not wholly automate sensitive analyses.

Finally, ethical and regulatory issues temper the discussion. As the Pistoia survey around Feb 2024 found, **legislation is not keeping pace**. Only 9% of professionals felt well-informed about new AI laws (^[40] [www.pistoiaalliance.org](#)). Moreover, over a third believed regulations were already inhibiting their work (^[41] [www.pistoiaalliance.org](#)). In pharma, concerns about patient data privacy, clinical safety, and IP can all affect how literature AI is used (e.g. scanning electronic health records or proprietary trial data). These factors introduce uncertainties about long-term adoption in certain applications. Nevertheless, for *open literature*, regulation is less of a barrier: summarizing published papers does not usually raise legal issues.

In conclusion, the evidence is strong that AI can and does improve literature review speed and coverage in biotech. Yet it is not a silver bullet and must be deployed thoughtfully. We move next to concrete examples that illustrate both the power and the workflow integration of AI in literature review and adjacent tasks.

4. Case Studies and Real-World Examples

To illustrate how AI-driven literature review works in practice, we present several case studies—from startups to big pharma—that showcase both successes and lessons learned. Each exemplifies a different aspect of AI-assisted review.

4.1. Rapid Screening with GPT-4 (Pandemic Response)

In *International Journal of Medical Informatics* (2024), researchers Prasad and Bassett **demonstrated an automated pipeline** for literature screening using GPT-4 (^[5] [www.sciencedirect.com](#)). They applied it to identify drug targets for emerging viral pathogens. The workflow was as follows: start with a broad PubMed keyword search on a virus (e.g. “SARS-CoV-2”), gather hundreds of abstracts, then feed these to GPT-4 with engineered prompts asking whether each paper described a valid drug target for the virus. The results were then compared to human expert labels.

Outcome: The AI system achieved **92.87% accuracy (F1=0.8843)** on SARS-CoV-2 literature, and 87.4% accuracy (F1=0.7390) on Nipah virus – where the literature is much sparser (^[5] [www.sciencedirect.com](#)). These numbers are remarkably high for an automated text system. The authors emphasize that GPT-4 “performed close to the level of human expert reviewers” in filtering relevant studies. The work’s *Highlights* section even states: “*Our automated pipeline performs close to the level of human expert reviewers for both [viruses]*” (^[42] [www.sciencedirect.com](#)).

Significance: This case shows that modern LLMs can tackle real biotech literature tasks with rigorous benchmarks. In a simulated drug-discovery context, the GPT-4 pipeline could reduce the pool of papers needing full review by an order of magnitude – e.g. from 250 COVID papers down to a manageable subset. The authors explicitly note the benefit: “*the study highlights the utility of LLMs for rapid drug discovery... especially for pandemic response*” (^[5] [www.sciencedirect.com](#)). In interviews, industry scientists have remarked that AI-assisted filtering would be invaluable during a new outbreak.

Caveats: The pipeline relied on “advanced prompt engineering” and external tools (PubMed access) to achieve these numbers (^[43] [www.sciencedirect.com](#)). Simpler use of ChatGPT (e.g. vanilla prompts without chain-of-thought or web access) does not perform as consistently (^[5] [www.sciencedirect.com](#)). In other words, matching expert-level screening requires careful system design, not just “ask ChatGPT once”. Furthermore, this example deals with fairly well-defined

binary questions. More ambiguous research queries could yield lower AI precision. The researchers also caution that the AI is not infallible – any missed relevant paper or false positive could mislead. Thus, even in this success story, a final human check remained part of the workflow (the AI flagged likely papers, which humans then reviewed).

4.2. Automated Review Article Generation

In a pilot study published in *BioData Mining* (2024), Wang et al. evaluated GPT-4's ability to **write sections of a scientific review article** ^{([44](#))} [biodatamining.biomedcentral.com](#)) ^{([6](#))} [biodatamining.biomedcentral.com](#)). This experiment was set up as follows: the authors took two existing cancer-therapy review papers as “ground truth”. For one paper (BRP1), they uploaded all 113 reference articles into GPT-4 and asked it to generate an outline and content for a review titled “The Spectrum of Sex Differences in Cancer”. They then compared the AI's output to the human-authored paper in terms of coverage and wording.

Outcome: GPT-4 was able to generate a section- and subsection-level outline that *covered the same topics* as the original review. It produced nine main sections and twelve subsections, many matching those chosen by the human authors ^{([6](#))} [biodatamining.biomedcentral.com](#)). Qualitatively, the authors found that GPT-4 could “effectively summarize the key points from a comprehensive list of documents,” and that with its assistance, “authors can swiftly summarize a list of main topics for further refinement” ^{([6](#))} [biodatamining.biomedcentral.com](#)). Quantitatively, they report a semantic similarity score of ~0.75 between the AI-generated text and the original paper's text on corresponding points ^{([45](#))} [biodatamining.biomedcentral.com](#)). In plain terms, the model produced prose that was very close in meaning to the published sections.

A second experiment in the same study looked at more detailed content generation (text for each subsection). GPT-4 again achieved a mean similarity score close to 0.75–0.76 compared to reference content ^{([46](#))} [biodatamining.biomedcentral.com](#)). While not perfect, this performance was “*akin to that of a human scholar*” according to the authors. They conclude: “*GPT-4 has demonstrated adequate capability*” to generate review content ^{([47](#))} [biodatamining.biomedcentral.com](#)). However, they also note limitations: GPT-4 sometimes could not list which articles it was using for a given point, and certain minor copying was detected in reference-based tests ^{([48](#))} [biodatamining.biomedcentral.com](#)) ^{([49](#))} [biodatamining.biomedcentral.com](#)).

Significance: This case is one of the first peer-reviewed demonstrations that an LLM can draft a literature review *with moderate accuracy*. For biotech R&D, it suggests that a scientist could ask an AI assistant to outline the key findings in a field, then edit and refine the draft, rather than starting from scratch. In fact, the authors describe the ideal workflow as a *collaboration*: the AI “sorts and summarizes articles” and the human author does high-level analysis and editing ^{([6](#))} [biodatamining.biomedcentral.com](#)). This human-in-the-loop model is exactly how many teams envision using AI tools today: as productivity enhancers for writing.

Caveats: The study also highlights risks. Instances of direct copying (“copy-paste” content) were found in some GPT outputs ^{([49](#))} [biodatamining.biomedcentral.com](#)). The authors stress that similarity scores only measure semantic overlap, not factual completeness. Importantly, they recommend that GPT-generated summaries should be validated by humans (“the reference-based content generation raised concerns” ^{([47](#))} [biodatamining.biomedcentral.com](#))). They warn that AI might miss nuances of methodology or generate plausible-sounding but incomplete statements. Thus, while the study shows promise, it also reinforces that AI outputs for review writing should be used cautiously.

4.3. Industry Adoption: Oxford PharmaGenesis Case Study

In industry, the deployment of AI for literature review is already happening at scale. Oxford PharmaGenesis – a scientific communications consultancy advising top pharma companies – provides a concrete example. In a published case study, their team described using **Elicit's AI-assisted platform** to vastly accelerate client projects ^{([50](#))} [elicit.com](#)) ^{([51](#))} [elicit.com](#)).

- **Challenge:** Pharma clients frequently request large literature reviews. In one example, a client wanted a **targeted review on dyslipidemia treatments**. The team initially found 26,000 potentially relevant papers in their search (^[52] elicit.com). Manually screening that many would be infeasible: “the discovery and extraction phase consumes a lot of expert resource that could be better spent on analysis” (^[52] elicit.com). Typically, without AI they might narrow the search by iterative scoping, but this risks missing important findings.
- **AI Solution:** The Oxford team piloted Elicit. Elicit is a tool that parses PDFs and extracts answers to specific questions, returning clear citations. After implementing Elicit, they communicated: “Elicit’s tech is just better for data extraction” (^[32] elicit.com). The platform allowed them to input open-ended research questions and have the AI screen and summarize evidence from the full corpus of thousands of papers. They noted that Elicit successfully handled PDF parsing (historically problematic for AI) and reliably traced every fact back to its source, which is crucial for trust (^[32] elicit.com).
- **Results:** With Elicit, Oxford PharmaGenesis could meet an ambitious timeline. They reported that “we initiated the project on a Monday, and [the client] asked for results by Friday”. By Monday, they had loaded all relevant articles; by Thursday, with Elicit plus human validation, they delivered a first draft of data extractions (^[51] elicit.com). They essentially completed in 4 days what would normally take weeks. Over the whole collaboration, they “completed a rapid literature review investigating 40 research questions across 500 papers in under a week” (^[31] elicit.com). These figures illustrate an **unprecedented scale** of literature analysis made possible by AI.
- **User Feedback:** Key individuals praised the AI’s reliability. Tomas Rees, Innovation Director at OP, said Elicit’s results were traceable and reliable enough to rely on during client delivery (^[32] elicit.com). Kim Wager, Scientific Director, highlighted how this capacity removed previous bottlenecks: “These [reviews] can range from comprehensive analyses... Whatever the need, the scale is limited by the capacity of experts to screen and extract” (^[22] elicit.com). In other words, they see AI as removing the “capacity” limitation on scale.

Learnings: This example underscores the tangible benefits: the firm can take on many more projects of higher complexity, and adapt quickly to client needs, because AI handles the grunt work. It also highlights best practices: human oversight remained vital (scientists validated all extractions), and the trustworthiness came from the tool’s ability to cite sources (anything generated by Elicit could be easily verified against the original papers (^[32] elicit.com)). The team noted that they tested other AI tools (unspecified) but found they *did not* improve the most laborious step (extraction) as Elicit did (^[53] elicit.com).

In sum, the Oxford case shows that literaturescale AI is already production-ready in industry. Leading consultancies and presumably many pharma R&D groups have started to integrate such tools into their pipelines. The result is dramatically higher throughput for evidence synthesis.

4.4. AI in Structural Biology (Beyond Literature Review)

While the focus here is on literature review, it is worth noting a related success story in biotech AI: structural biology. DeepMind’s AlphaFold 2 (2020) and related tools have transformed protein structure prediction, and this is becoming a top AI use case (71% adoption in the Benchling survey (^[1] www.benchling.com)). A published example of this synergy is the *AlphaFold-powered drug discovery* case by Ren *et al.* (2023) (^[33] pmc.ncbi.nlm.nih.gov) (^[54] pmc.ncbi.nlm.nih.gov). These scientists used AlphaFold (free database of human proteome structures) to feed into a generative AI platform (Chemistry42) and support a full drug discovery pipeline. Within **30 days**, they identified a new small-molecule inhibitor (Kd ~9 μM) against a novel target (CDK20) that had no previously known structure (^[33] pmc.ncbi.nlm.nih.gov) (^[54] pmc.ncbi.nlm.nih.gov). AlphaFold’s predicted protein structure was crucial in letting the AI design strategy begin its search.

As the authors state, this was “the first demonstration of applying AlphaFold to the hit identification process in drug discovery” (^[55] pmc.ncbi.nlm.nih.gov). The accelerated timeline and reduced compound synthesis (only 7 compounds for the first hit) underscore how automated structure knowledge (essentially a form of “AI reading” of the literature of proteins) can speed up R&D. While this example goes beyond text mining (involving generative chemistry), it still rests on AI extracting biological insight (structure) from published data. It reflects the same trend: tasks built on existing knowledge (here, predicted models of proteins) yield early AI dividends.

5. Implications and Future Directions

Having established that literature review is the current flagship AI application in biotech, we now consider *what comes next*. This encompasses emerging use cases, the challenges in extending AI further, and the broader impact on research practice. The insights below draw on technology roadmaps, survey foresight, and expert commentary.

5.1. Broader AI Applications on the Horizon

The Benchling report and industry thought-leaders identify several areas poised for growth:

- **Complex Data Integration:** The next frontier beyond text is *multimodal AI*. Biotech produces diverse data types (genomic sequences, images from microscopes, chemical structures, time-series assay data, etc.). Models that can combine these modalities promise richer insights (e.g. linking gene expression with cellular imaging to predict phenotype). Companies mention “*multimodal models*” as a top area of investment (^[56] www.benchling.com). Progress is being made (e.g. OpenAI’s GPT-4 Vision, bespoke biomedical multimodal models), but integrating heterogeneous R&D data remains challenging. Unlike text, such data often lack standard formats or shared semantics, requiring new ontologies and integration tools.
- **AI-driven Experimentation (“Co-scientists” and Autonomous Labs):** A growing vision is for AI to not only analyze data but to propose and even execute experiments. Conventionally, literature suggests that the path forward is moving from “task-level copilots” to “*end-to-end agentic workflows*” (^[56] www.benchling.com) (^[14] www.linkedin.com). Benchling explicitly mentions “co-scientist” agents on the horizon (^[14] www.linkedin.com). These would be AI systems that can design an experiment, run simulations (or control lab robots), and interpret results with minimal human input. For example, a co-scientist might monitor an assay in real time and adjust conditions to optimize yield. Early prototypes (MIT and Berkeley projects, Tesla of labs, etc.) are in research phase, but we expect increasing adoption of automated wet-lab platforms guided by AI.
- **Advanced Predictive Modeling (Generative Design, ADME, etc.):** Curiously, the Benchling data show **declining adoption in current generative design and ADME tasks** (^[13] www.benchling.com). This is not surprising: generative drug design (create new molecules) is immensely complex and data-hungry. Only premature attempts have been deployed widely. However, these remain areas of huge interest. As data matures (e.g. more structure-activity data, better pharmacokinetic models), adoption in these domains is expected to rise. Indeed, reports cite that many biotechs are investing in ADME optimization and multimodal drug design next (^[56] www.benchling.com). Progress here will likely require not just better algorithms but also extensive validation pipelines (both in silico and in labs).
- **Workflow Orchestration and Ops:** The Labs and IT are also looking at operational AI: automating lab scheduling, resource allocation, manufacturing processes, and compliance workflows. Benchling calls out “workflow orchestration” and “manufacturing optimization” as planned growth areas (^[56] www.benchling.com). AI in these areas will link the bench and the boardroom: e.g. smart LIMS systems that prioritize experiments, or AI-driven supply chain for biologics. These are multidisciplinary challenges, involving not only AI but also cyber-physical systems (robotics) and regulatory compliance.

The **unifying theme** is that future uses of AI will require **deeper integration and better data foundations**. Tasks like lit review are “task-level” – the AI takes a specific input (papers) and gives a specific output (summaries). The ambitious next steps involve coordinating across tasks and data sources. Realizing them will hinge on closing the gaps identified in surveys: standards, FAIR data practices, and cross-industry collaboration to share knowledge (Pistoia Alliance and others emphasize this).

5.2. Organizing R&D for AI (“Builder Culture”)

To seize these opportunities, biotech organizations are **reorganizing** themselves. One clear shift is in talent strategy. The Benchling survey notes that 67% of AI-related talent in biotech comes from *training/upskilling existing staff*, not from hiring outside data scientists (^[57] www.benchling.com). In other words, companies are creating “hybrid scientists” who combine domain expertise with AI/ML skills. This builder culture means scientists learn to use (and even develop) AI tools, rather than relying solely on IT departments. Interdisciplinary teams (“sprint groups” or labs-of-the-future) are living examples of this trend (^[57] www.benchling.com).

In practical terms, this means more internal training programs, hiring of computational biologists, and cross-functional projects. It also means that AI and data infrastructure initiatives must be scientist-friendly. As one quote from Benchling’s report highlights, the most pressing challenges are organizational: “*Technical infrastructure is as important as pipeline*”.

Many teams now measure their digital maturity: do they have unified data lakes? Do ELNs and LIMS integrate seamlessly with AI pipelines? Leadership understands that buying point tools is not enough – they must build an **AI-ready ecosystem**.

Ironically, this inner focus can be out of step with hype. The LinkedIn commentary from Benchling's president (Ashu Singhal) notes, "AI has become scientists' default interface..."⁽⁵⁸⁾ www.benchling.com); but then adds "the responsibility now is to balance excitement with rigorous validation"⁽⁵⁹⁾ www.benchling.com). Executives from companies like Schrödinger stress that the big challenge is not getting AI tools, but **connecting them across the organization**⁽⁵⁹⁾ www.benchling.com). In practice, biotech firms are creating dedicated AI/ML groups, data governance boards, and cross-lab pilot projects to iteratively build these connections. This "fail-fast" attitude – test new AI capabilities quickly in a safe environment – is being cited as a best practice for moving from pilots to enterprise AI.

5.3. Data and Infrastructure Bottlenecks

Even as culture shifts, a consistent theme is that *data remains the bottleneck*. Many biotech legacy systems were not built for AI: data are siloed in PDFs, spreadsheets, or proprietary formats. Benchling emphasizes that "static, siloed data environments" are the single largest barrier to AI projects today⁽⁶⁰⁾ www.benchling.com). Indeed, the report states flatly: **the number-one reason AI pilots fail is poor data quality**⁽⁶⁰⁾ www.benchling.com).

This has particular relevance for literature review. On the one hand, published literature is relatively standardized (digital journals, consensus metadata). But even here, problems remain: journal paywalls may limit aggregate analysis, and older PDFs can have OCR issues. Many companies are tackling this by building internal knowledge bases (curated corpora of literature) and semantic layers. Projects like the Pistoia Alliance's Unified Data Model (UDM) aim to create common schemas so that literature facts can be directly linked to lab data. Benchling's own AI platform is adding features to parse PDFs, annotate literature, and integrate with in-house notebooks.

Without high-quality text corpora, even the best LLMs struggle. For broader AI tasks, data issues intensify (e.g., fragmented assay results, non-standard metadata). The optimistic vision is that 2025–2030 will see a wave of data-fairification in biotech, partly driven by the demand for AI. Responses from our sources emphasize the need for standardized ontologies, open data licenses, and better data-sharing incentives (such as pre-competitive consortia). As one leader noted, progress has been made but "40% of respondents still say their company lacks standards for data and metadata"⁽⁶¹⁾ www.pistoiaalliance.org). This must improve if future use cases are to succeed.

5.4. Trust, Ethics, and Regulations

Literature reviews touch indirectly on human subjects (literature often includes clinical studies) and directly on intellectual capital. Thus, AI usage in this domain interacts with privacy, bias, and compliance concerns. So far, summarizing public literature has few direct ethical issues (aside from risk of misinterpretation). But as AI use in R&D grows, regulators are paying attention.

For instance, the EU's incoming AI Act and forthcoming FDA guidance on AI in medical devices imply that AI tools used for clinical decision-support or trial analysis will face scrutiny. If an AI system is used to generate hypotheses about patient treatments, there will likely need to be transparency around it. Our sources indicate that life sciences professionals are already worried about these legal unknowns⁽³⁶⁾ www.pistoiaalliance.org). The Pistoia press release warns that ambiguous regulations could stall AI adoption unless industry and regulators collaborate⁽⁶²⁾ www.pistoiaalliance.org). Therefore, companies are establishing internal governance — requiring that all AI literature summaries used for decision-making must be verifiable and linked to sources, for example.

On the ethical side, concerns about hallucination or bias (see Section 3.2) have motivated best practices. Some journals now issue guidance on AI use: for example, authors of research papers using tools like ChatGPT must disclose which parts were AI-assisted and how they verified content. Companies in our interviews said they require human experts to co-sign any analysis derived from AI, exactly to account for the credibility issues highlighted by research⁽¹²⁾

biodatamining.biomedcentral.com) (^[9] link.springer.com). In short, “*trust but verify*” is the emerging mantra: use AI for efficiency, but require explainability and audit trails.

5.5. Skills and Training

Finally, the human dimension: AI adoption in biotech rests on people learning new skills. As Benchling’s data show, successful organizations overwhelmingly build up AI talent internally (^[57] www.benchling.com). This means cross-training biologists in data science, and conversely, data scientists in biology. Some companies have instituted formal “AI for Biologists” training programs, running workshops on how to prompt LLMs for research questions, how to curate datasets, and how to interpret AI outputs. Others are recruiting “bioinformaticians” with dual backgrounds.

Our sources indicate that, by mid-2025, many biotechs expected to expand AI budgets notably, but primarily for personnel and infrastructure – not necessarily for buying more licenses (^[63] www.linkedin.com). The strategic shift is clear: AI is not just another piece of software, but a new way of working. We heard from a head of R&D informatics that for their company, the question is now “*who on the team can speak both biology and AI?*” and how to embed that expertise into project teams. Supporting this, the Benchling report notes increased interest in training on FAIR data and ontologies (^[61] www.pistoiaalliance.org), as well as use cases for which to allocate compute resources.

In summary, the “what comes next” is not just about technology, but about building a biotech ecosystem – data, tools, processes, and people – that is AI-native. The threads from the literature, surveys, and interviews converge on a picture of gradual, collaborative progress. The stage is being set for AI to move *from* being a tool for isolated tasks to being the backbone of the entire R&D operation. We will now conclude by summarizing these points.

6. Conclusion

The evidence is unambiguous: **AI-powered literature review is the leading frontier of AI adoption in biotechnology.** In biotech R&D, 76% of early AI adopters already use AI for literature and knowledge extraction (^[1] www.benchling.com), making it far and away the most common use case. This popularity is driven by practical factors: there is a wealth of digital publications to mine, and modern NLP tools can handle these text-centric tasks well. Biotech scientists have already integrated AI search, summarization, and data extraction into their workflows, reporting significant time savings and accelerated discovery as a result (^[2] www.benchling.com) (^[28] journals.sagepub.com). Indeed, projects that would have once taken weeks or months (screening 500 papers, writing a review outline, etc.) can now be done in days with AI-assisted methods (^[31] elicit.com) (^[6] biodatamining.biomedcentral.com).

At the same time, challenges remain. AI outputs must be checked for accuracy; data quality and infrastructure lags; and staff require training and best practices to use these tools effectively. Surveys reveal that professionals understand the promise of AI but worry about trustworthiness and regulation (^[10] www.technologynetworks.com) (^[36] www.pistoiaalliance.org). These concerns have not meaningfully slowed adoption, but they do shape the unfolding approach.

Looking forward, the biotech industry is already plotting its next moves. Beyond hit-the-ground work like lit reviews, companies plan to tackle more complex AI problems: integrating multimodal data, automating experiments, and designing molecules in silico. These tasks are harder – requiring better data foundations, stronger computation, and closer human–AI collaboration. The next wave of “killer apps” may well be 5–10 years away, but early signposts (e.g. AlphaFold-enabled design, AI in formulation, etc.) are visible in the literature and press (^[33] pmc.ncbi.nlm.nih.gov) (^[29] link.springer.com).

In conclusion, the story of AI in biotech over the past few years has been a story of **rapid niche wins paving the way for broader transformation.** Literature review is the #1 use case now because it was ripe and ready, and because scientists demanded relief from information overload. As those demands are met, attention will naturally shift to the next bottlenecks. For sponsors of R&D – from start-ups to Big Pharma – the clear message is: invest in AI for literature and data work *today*, while preparing organizationally for AI to soon become integrated end-to-end. The data and expert

- [60] <https://www.benchling.com/biotech-ai-report-2026#:~:AI%20...>
- [61] <https://www.pistoiaalliance.org/news/lab-of-the-future-2024-global-survey/#:~:Barri...>
- [62] <https://www.pistoiaalliance.org/news/press-release-pistoia-alliance-survey-exposes-ai-knowledge-gap/#:~:three...>
- [63] https://www.linkedin.com/posts/ashu-singhal-1b58a18_weve-just-published-our-2026-biotech-ai-activity-7414686835278270464-zj8B#:~:furth...

IntuitionLabs - Industry Leadership & Services

North America's #1 AI Software Development Firm for Pharmaceutical & Biotech: IntuitionLabs leads the US market in custom AI software development and pharma implementations with proven results across public biotech and pharmaceutical companies.

Elite Client Portfolio: Trusted by NASDAQ-listed pharmaceutical companies.

Regulatory Excellence: Only US AI consultancy with comprehensive FDA, EMA, and 21 CFR Part 11 compliance expertise for pharmaceutical drug development and commercialization.

Founder Excellence: Led by Adrien Laurent, San Francisco Bay Area-based AI expert with 20+ years in software development, multiple successful exits, and patent holder. Recognized as one of the top AI experts in the USA.

Custom AI Software Development: Build tailored pharmaceutical AI applications, custom CRMs, chatbots, and ERP systems with advanced analytics and regulatory compliance capabilities.

Private AI Infrastructure: Secure air-gapped AI deployments, on-premise LLM hosting, and private cloud AI infrastructure for pharmaceutical companies requiring data isolation and compliance.

Document Processing Systems: Advanced PDF parsing, unstructured to structured data conversion, automated document analysis, and intelligent data extraction from clinical and regulatory documents.

Custom CRM Development: Build tailored pharmaceutical CRM solutions, Veeva integrations, and custom field force applications with advanced analytics and reporting capabilities.

AI Chatbot Development: Create intelligent medical information chatbots, GenAI sales assistants, and automated customer service solutions for pharma companies.

Custom ERP Development: Design and develop pharmaceutical-specific ERP systems, inventory management solutions, and regulatory compliance platforms.

Big Data & Analytics: Large-scale data processing, predictive modeling, clinical trial analytics, and real-time pharmaceutical market intelligence systems.

Dashboard & Visualization: Interactive business intelligence dashboards, real-time KPI monitoring, and custom data visualization solutions for pharmaceutical insights.

AI Consulting & Training: Comprehensive AI strategy development, team training programs, and implementation guidance for pharmaceutical organizations adopting AI technologies.

Contact founder Adrien Laurent and team at <https://intuitionlabs.ai/contact> for a consultation.

DISCLAIMER

The information contained in this document is provided for educational and informational purposes only. We make no representations or warranties of any kind, express or implied, about the completeness, accuracy, reliability, suitability, or availability of the information contained herein.

Any reliance you place on such information is strictly at your own risk. In no event will IntuitionLabs.ai or its representatives be liable for any loss or damage including without limitation, indirect or consequential loss or damage, or any loss or damage whatsoever arising from the use of information presented in this document.

This document may contain content generated with the assistance of artificial intelligence technologies. AI-generated content may contain errors, omissions, or inaccuracies. Readers are advised to independently verify any critical information before acting upon it.

All product names, logos, brands, trademarks, and registered trademarks mentioned in this document are the property of their respective owners. All company, product, and service names used in this document are for identification purposes only. Use of these names, logos, trademarks, and brands does not imply endorsement by the respective trademark holders.

IntuitionLabs.ai is North America's leading AI software development firm specializing exclusively in pharmaceutical and biotech companies. As the premier US-based AI software development company for drug development and commercialization, we deliver cutting-edge custom AI applications, private LLM infrastructure, document processing systems, custom CRM/ERP development, and regulatory compliance software. Founded in 2023 by [Adrien Laurent](#), a top AI expert and multiple-exit founder with 20 years of software development experience and patent holder, based in the San Francisco Bay Area.

This document does not constitute professional or legal advice. For specific guidance related to your business needs, please consult with appropriate qualified professionals.

© 2025 IntuitionLabs.ai. All rights reserved.