

AI Research Assistants for Drug Discovery Compared

By Adrien Laurent, CEO at IntuitionLabs • 3/15/2026 • 55 min read

ai research assistants

drug discovery

biomedical research

literature review

evidence synthesis

knowledge graphs

pharmacoinformatics



Executive Summary

Drug discovery is an inherently complex and costly process, often taking over a decade and upward of \$2 billion per new therapeutic (^[1] bio-in-tech.com). Recent advances in artificial intelligence (AI), particularly *retrieval-augmented* and domain-specific AI tools, promise to dramatically accelerate literature review, hypothesis generation, and target identification in drug development. This report examines four leading AI research-assistant platforms—**Causaly**, **Elicit**, **Consensus**, and **Semantic Scholar**—that are used to aid drug discovery and biomedical research. Each platform offers unique capabilities: Causaly specializes in a high-precision biomedical knowledge graph for life sciences; Elicit automates literature searching and evidence extraction; Consensus provides AI-powered question-answering with evidence synthesis; and Semantic Scholar is a broad AI-driven academic search engine with advanced summarization features.

We compare these tools in depth, analyzing their architectures, data coverage, strengths, and limitations. For example, Causaly's knowledge graph encompasses ~500 million facts and 70 million directional relationships across thousands of semantic categories (^[2] www.causaly.com), enabling precise causal reasoning about genes, pathways, and diseases. In contrast, Elicit and Consensus draw on large indices of scientific publications (roughly 138 million papers (^[3] orion.elicit.com) and 200 million papers (^[4] venturebeat.com), respectively) and use language models to sift and summarize the evidence. Semantic Scholar, with over 200 million indexed papers (^[5] artificial-intelligence-wiki.com), uses machine learning to extract TLDR summaries and personalized recommendations across fields.

We review case studies and user experiences where available: for instance, Oxford PharmaGenesis used Elicit to answer 40 research questions across 500 papers in under a week (^[6] elicit.com), highlighting how AI assistants can drastically speed literature reviews in pharmaceutical contexts. Finally, we discuss implications for drug discovery and future directions: as **AI-designed drugs** enter clinical trials and new **regulatory frameworks** emerge (^[7] bio-in-tech.com), research-assistant tools will become integral to R&D workflows. However, challenges remain in ensuring coverage (e.g. paywalled literature gaps (^[8] pmc.ncbi.nlm.nih.gov)), avoiding **hallucinations** of generative models, and integrating trust and transparency features (like inline citations (^[9] www.iatrox.com)). Overall, AI research assistants hold great promise to make scientists “vastly more productive and accurate” (^[10] ought.org), potentially reducing preclinical timelines by 30–40% (^[11] bio-in-tech.com), but must be used judiciously within evidence-based practice.

Introduction

Drug discovery has traditionally been a lengthy, costly **trial-and-error** endeavor. Developing a single new drug often exceeds a decade and \$2 billion (^[1] bio-in-tech.com). Even with modern techniques, roughly **90% of drug candidates fail** before approval (^[1] bio-in-tech.com). These high costs and failure rates place enormous pressure on pharmaceutical R&D to become more efficient. In recent years, *artificial intelligence* (AI) has emerged as a transformative force in drug discovery. AI and deep learning methods offer novel ways to predict molecular properties, screen virtual compounds, and propose new drug designs (^[12] pubs.acs.org) (^[13] pubs.acs.org). For example, machine learning algorithms are used for **structure-based and ligand-based virtual screening**, de novo molecular design, ADMET property prediction, and drug repurposing (^[12] pubs.acs.org).

Beyond algorithmic modeling, a wealth of biomedical knowledge exists in millions of research papers, patents, genomic data, and clinical studies. Harnessing that knowledge is critical; leading reviews note that AI can provide “a complete picture of the biomedical landscape” by analyzing **human-readable texts and structured data** (^[14] www.causaly.com). However, the sheer volume of published literature makes manual review impractical. In 2026, Semantic Scholar indexed over 200 million papers (^[5] artificial-intelligence-wiki.com), and platforms like Causaly report ingesting 500 million facts from life-science texts (^[2] www.causaly.com). Consequently, specialized AI tools have been developed as *research assistants* to help scientists **find**, **synthesize**, and **interpret** the relevant evidence far faster than humanly possible.

These AI research assistants differ markedly from general-purpose chatbots. A recent analysis emphasizes that while **generative AI (like ChatGPT)** “predicts the next word” and can easily hallucinate, modern research assistants adopt a **retrieval-first architecture** (^[9] www.iatrox.com) (^[15] www.iatrox.com). In other words, they begin by searching a curated database of scientific literature (the *realm of evidence*), then read and summarize the actual source documents. The output is an evidence-grounded answer with inline citations to peer-reviewed sources, rather than an unsupported AI “opinion” (^[9] www.iatrox.com) (^[15] www.iatrox.com). For example, one blogger notes that these tools “*find real documents, read them, and synthesize an answer based only on what they found, providing clickable citations for every claim*” (^[15] www.iatrox.com). Indeed, reliability signals such as **inline references**, clearly bounded corpora (e.g. only high-quality papers), and transparency about preprints vs. peer review are key differentiators of these platforms (^[16] www.iatrox.com).

In this report, we conduct an extensive analysis of four prominent AI research assistants relevant to drug discovery:

- **Causaly:** A UK-based platform focused exclusively on life-science research, built around a massive knowledge graph of biomedical facts (^[2] www.causaly.com). Causaly combines graph-based causal reasoning with **large language model (LLM)** techniques (the so-called *Scientific RAG*) to answer complex R&D questions (^[17] www.causaly.com). It is specifically marketed to pharmaceutical companies and claims to “accelerate understanding” of disease mechanisms (^[18] www.causaly.com).
- **Elicit:** Developed by the AI lab *Ought*, Elicit uses LLMs to automate literature review and question-answering across all scientific domains. It supports systematic reviews by finding relevant papers, extracting data, and summarizing findings (^[19] ought.org) (^[20] www.alignmentforum.org). Elicit’s goal is to make researchers “vastly more productive and accurate” (^[10] ought.org). It has been adopted by academics and industry researchers for rapid evidence synthesis.
- **Consensus:** An AI search-engine startup (Boston-based) that indexes a large corpus of research (200M+ papers (^[4] venturebeat.com)) and provides AI-generated answers to health and science queries with supporting citations. Consensus emphasizes clinical and life-science questions, offering a “Consensus Meter” that indicates whether the majority of evidence supports a yes/no answer (^[21] www.iatrox.com) (^[22] www.comparegen.ai). It stresses use in evidence-based medicine, with marketing claiming “*millions of doctors use Consensus to get answers backed by top medical journals.*” (consensus.app)
- **Semantic Scholar:** A free, AI-powered academic search engine developed by the Allen Institute for AI. With over 200 million papers indexed (^[5] artificial-intelligence-wiki.com), it was among the first to apply AI (NLP) in literature search. Its key features include **TLDR** one-sentence summaries of papers, semantic PDF reading tools, personalized feeds, and citation influence analysis (^[23] artificial-intelligence-wiki.com). It is widely used by researchers for paper discovery across all scientific fields.

We will examine each tool’s approach, technological underpinnings, data coverage, and relevance to drug discovery (e.g. in target identification, mechanism elucidation, or repurposing). For instance, Causaly’s knowledge graph explicitly encodes causal relationships (distinguishing cause vs. mere co-occurrence) (^[24] www.causaly.com), aiming to surface *mechanistic hypotheses*. Elicit’s factored-cognition pipeline breaks research tasks into search, summarization, and classification subtasks (^[20] www.alignmentforum.org). Consensus’s core strength is evidence synthesis from literature for yes/no queries (^[4] venturebeat.com), (^[21] www.iatrox.com). Semantic Scholar excels at rapid literature sorting via AI summarization, allowing users to scan dozens of papers quickly (^[25] artificial-intelligence-wiki.com).

We also include data analyses, expert commentary, and case studies. For example, an Oxford PharmaGenesis team (a top-10 pharma consultant) used Elicit to tackle 40 research questions from 500 papers in under a week (^[6] elicit.com). Such examples illustrate how AI assistants can expand research capabilities. Finally, we discuss broader implications: the impact on R&D efficiency, integration into workflows, and future prospects. By 2026 multiple AI-designed drugs are entering Phase III trials and new regulatory frameworks are being established (^[7] bio-in-tech.com), meaning these AI tools will likely play an increasing role in translational research. However, issues of data bias, incomplete coverage (especially of paywalled literature (^[8] pmc.ncbi.nlm.nih.gov)), and the need for rigorous validation remain. In summary, AI research assistants represent a paradigm shift in drug discovery research, and this report provides the first in-depth, comparative look at how leading platforms—Causaly, Elicit, Consensus, and Semantic Scholar—are shaping the field.

Background: AI in Drug Discovery and Research

AI's application in drug discovery is not entirely new. Early computational methods (e.g. quantitative structure-activity relationships) date back decades. Machine learning, including support vector machines and neural networks, increasingly powered tasks like predicting pharmacological properties and virtual screening ⁽²⁶⁾ pubs.acs.org ⁽²⁷⁾ www.mdpi.com). The advent of deep learning brought techniques for modeling complex molecular relationships (e.g. graph convolutional networks for molecules) and for analyzing high-dimensional data. Crucially, **NLP** and knowledge engineering techniques allow computers to “read” scientific literature. The 2019 *Chemical Reviews* survey notes that modern AI (especially deep models) “provides opportunities for the discovery and development of innovative drugs.” It cites numerous AI applications: structure- and ligand-based virtual screening, de novo design, ADMET prediction, drug repurposing, etc. ⁽¹³⁾ pubs.acs.org). The review emphasizes that AI is “synonymous with certain machine learning techniques” and that domain-specific AI approaches have already been used in drug design ⁽¹³⁾ pubs.acs.org). In other words, drug discovery is a major domain for applied AI, and interest has surged, especially in the last five years ⁽²⁸⁾ www.mdpi.com) ⁽²⁷⁾ www.mdpi.com).

From a *process* perspective, AI can aid virtually every stage of drug R&D. AI models help in identifying new **targets** and **biomarkers**, optimizing chemical syntheses, designing more effective molecules, and prioritizing compounds for trials. At the same time, *knowledge-based* AI approaches have an important role: integrating vast amounts of published data to uncover non-obvious drug-disease or drug-target relationships. For instance, knowledge graphs and causal reasoning have become powerful frameworks to encode biomedical knowledge ⁽²⁶⁾ pubs.acs.org) ⁽²⁴⁾ www.causaly.com). These graphs connect entities (genes, proteins, drugs, diseases) with edges representing interactions or causal links. As an example, Causaly's platform claims to build an AI model of “cause-and-effect biology” by representing millions of such facts in an accessible graph ⁽²⁾ www.causaly.com) ⁽¹⁴⁾ www.causaly.com).

The **scale of available literature and data** is a major driver for automated tools. Tens of millions of new papers are published annually. Scientists simply *cannot* read all relevant publications manually. Traditional search engines (PubMed, Google Scholar, etc.) have limits: they return long lists of articles ranked by keyword or simple metrics, leaving the burden on humans to sift and summarize. The new class of AI research assistants changes this by automatically extracting and summarizing evidence. For example, one analysis describes these tools as the “eras of ‘hallucinating’ chatbots” ending, replaced by “AI medical search engines” that enforce grounding in real documents ⁽¹⁵⁾ www.iatrox.com).

Technical differences are key. Classic keyword search (Google Scholar/PubMed style) differs from AI assistants that apply **retrieval-augmented generation (RAG)**: first retrieve relevant documents from a curated corpus, then use an LLM to generate a summary or answer with citations ⁽⁹⁾ www.iatrox.com) ⁽¹⁷⁾ www.causaly.com). Unlike generic ChatGPT, which can hallucinate facts, AI search tools return clickable citations and clearly indicate their sources ⁽⁹⁾ www.iatrox.com). The presence of inline references is a crucial “reliability signal” ⁽¹⁶⁾ www.iatrox.com). For instance, iatroX (a clinical AI blog) points out that trustworthy AI research tools should allow users to click a reference and view the source PDF ⁽¹⁶⁾ www.iatrox.com). Similarly, tools that restrict themselves to scholarly databases (avoiding noisy web results) are preferred: for example, Consensus is described as “*built exclusively on top of peer-reviewed literature*”, making it “safer” for academic queries ⁽²¹⁾ www.iatrox.com) ⁽²⁹⁾ finance.yahoo.com).

In summary, the convergence of big data and AI has created a new environment: it is now feasible to interrogate the biomedical knowledge base automatically. Recent estimates emphasize this shift: by 2026 the global AI-driven drug discovery market is on the order of billions of dollars ⁽³⁰⁾ bio-in-tech.com), and key performance goals include compressing preclinical timelines by ~30–40% ⁽¹¹⁾ bio-in-tech.com). As one Causaly blog notes, repurposing existing drugs (driven by AI) can shave off 6–7 years of development ⁽³¹⁾ www.causaly.com), and in fact already one-third of new approvals are for repurposed drugs ⁽³¹⁾ www.causaly.com). Parallel to these developments, leading AI teams (like Ought's

Elicit) explicitly position language models as “cognitive building blocks” for research (^[32] ough.org). If successful, these tools could make scientists “vastly more productive and accurate” (^[10] ough.org).

Against this backdrop, we examine four representative AI assistants. We will probe each for (1) its **context and origins**; (2) **data sources and coverage**; (3) **core functionality** (search, summarization, Q&A); (4) **use cases in drug discovery**; and (5) **advantages and limitations** (including expert views and explicit performance data). Wherever possible, we cite data, studies, or known usage statistics. We also present comparative tables to highlight similarities and differences. Using this detailed analysis, we aim to inform researchers and decision-makers about how these AI assistants can be leveraged—individually or together—to accelerate drug discovery research.

Causaly: A Domain-Specific Knowledge Graph Platform

Background and Positioning. Causaly is a London-based biotech startup (founded ~2017) that provides an AI platform specialized for life sciences research and drug discovery. It has positioned itself as a “science-grade AI partner for R&D” (^[33] www.causaly.com). The company has strong venture investment and adoption within pharma: a July 2023 TechCrunch report noted Causaly had raised \$60M and “already works with 12 of the world’s biggest pharmaceutical companies” by mid-2023 (^[34] techcrunch.com) (^[35] techcrunch.com). Its user base reportedly includes large pharma and biotech firms. For example, a Causaly press release advertises collaboration as an AI “competitive intelligence” tool for early drug development (^[36] www.causaly.com).

The distinguishing claim for Causaly is that it has built “**the largest and most accurate knowledge graph on the market for life sciences**” (^[37] www.causaly.com). The platform ingests data from thousands of vetted sources (abstracts, full texts, patents, clinical trials, etc.) to create a **biomedical knowledge graph** encoding facts and causal relationships. According to official documentation, Causaly’s graph currently contains on the order of **500 million facts** and **70 million directional relationships** (^[2] www.causaly.com). It also includes **5 million custom concepts** (specialized ontologies of biomedical terms) and covers millions of entities such as genes, pathways, targets, diseases, and treatments (^[38] www.causaly.com) (^[14] www.causaly.com). The graph includes not just “associations” but identifies cause-and-effect relations – critical for hypothesis generation – allowing it to “differentiate between causality and co-occurrence” (^[24] www.causaly.com). For example, it can distinguish between “factor A causes symptom B” versus “A and B merely co-occur”.

Core Functionality. Causaly offers a chat-based query interface (the “Causaly Copilot”), visual exploration tools, and graph analytics. Users can ask scientific questions in natural language, and the AI responds with structured answers drawing on the knowledge graph. Beneath the interface, Causaly employs a retrieval-augmented generation (RAG) approach: it searches the knowledge graph and possibly underlying texts, then synthesizes an answer with embedded citations (we refer to all scientific assertions in the answer). Causaly explicitly emphasizes the reliability of answers: it claims all findings are always verifiable with precise references. This is enabled by their “Scientific RAG™” mechanism: as one product page explains, Scientific RAG is a *hybrid of vector similarity search and advanced graph-based search*, which “pulls and prioritizes information from the knowledge graph to deliver complete, clear, accurate, and precise answers—verifiable through inline citations” (^[17] www.causaly.com). In practice, when a scientific question is asked, the platform will highlight relevant pathways or mechanisms, cluster related facts, and provide clear statements linking drugs, genes, and diseases with annotated sources.

In addition to the Copilot chat, Causaly has specialized applications: e.g. **Pipeline Graph** (launched 2025) integrates competitive intelligence (patents, pipelines, company data) with preclinical research insights (^[36] www.causaly.com). Causaly’s **Bio Graph** tool enables visual exploration: it generates interactive maps of how entities connect, which researchers can “investigate millions of relationships quickly” to answer complex R&D questions (^[39] www.causaly.com). For example, an R&D team could visually compare gene–disease clusters for different hypotheses side by side, as

shown in Causaly marketing. Importantly, Causaly can also incorporate proprietary data: through a “Private Data Fabric”, companies can embed their own experimental, proteomic or clinical data into the knowledge graph context ⁽¹⁴⁰⁾ www.causaly.com) ⁽¹⁴¹⁾ www.causaly.com). This facilitates combining public biomedical facts with a firm’s internal knowledge.

Performance and Use Cases. By design, Causaly is targeted to early-stage drug discovery (target ID, repurposing, mechanism elucidation). For instance, in the context of drug repurposing, Causaly’s team highlights that ~1/3 of new approvals now come from repurposed drugs and that AI can shave off 6–7 years of development compared to novel drugs ⁽¹³¹⁾ www.causaly.com). The platform’s ability to parse vast literature makes it well-suited to repurposing studies: one blog example describes using Causaly to investigate Exenatide (a GLP-1 agonist for diabetes) for possible new indications ⁽¹⁴²⁾ www.causaly.com). Although we lack quantitative benchmarks, pharma clients reportedly use Causaly to generate non-obvious hypotheses; a testimonial quotes a senior scientist saying that Causaly revealed findings in proteomics data that were “missed” by conventional analysis ⁽¹⁴³⁾ www.causaly.com).

Unlike more general search tools, Causaly’s knowledge graph is curated by PhD scientists and continuously updated in production environments ⁽¹⁴⁴⁾ www.causaly.com) ⁽¹²⁾ www.causaly.com). The platform emphasizes accuracy especially in causal relations. For example, marketing material points out that its graph is “distinct for depth of knowledge” on life-science topics and includes custom ontologies (on drug targets, causality, genes, pathways, etc.) that others lack ⁽¹⁴⁾ www.causaly.com). The emphasis is on *precision in answers*. Causaly’s AI will often provide a ranked list of evidence or charts showing how many studies support a hypothesis, reducing decision risk. This aligns with how experts expect domain-specific AI to function: focusing on evidence-backed insights rather than speculative suggestions.

However, there are limitations. Causaly’s proprietary database is highly specialized but opaque; the precise contents of its 500M facts are not fully known externally. Some academic perspectives might worry about coverage gaps; for example, if any journals are not mined, or how preprints are treated. The platform’s focus on life sciences also means it is not suitable for general scientific queries outside biomedicine. In terms of technology, while Causaly touts “agentic AI” and GPT-style reasoning, specifics of the underlying large language models are not public. The reliance on a curated graph does mean Causaly avoids issues like hallucinating false citations, but it also means it may miss out on the latest papers until they are ingested. Finally, being an enterprise product, Causaly requires a subscription and is not free for individual users.

Citations and Precautions. Like all these tools, Causaly provides inline citations to the literature, but there has been no public audit of its answer accuracy. Caution is needed: as with any AI, users must verify claims using the provided references. The company emphasizes that its answers are “verifiable through inline citations” ⁽¹⁷⁾ www.causaly.com). In practice, a scientist using Causaly is expected to click through these citations to the original papers before acting on any hypothesis. The platform compares favorably to general AI chatbots by construction, but ultimately its utility depends on the user’s domain expertise. In summary, Causaly is a highly specialized AI assistant that excels at integrating biomedical knowledge for drug discovery questions, with a massive underlying knowledge graph and domain-focused analytics ⁽²⁾ www.causaly.com) ⁽¹⁴⁾ www.causaly.com).

Elicit: An Agentic Literature-Review Assistant

Background and Goals. Elicit is an AI research assistant originally developed by *Ought*, a machine learning lab. It aims to automate complex open-ended research tasks using large language models (LLMs). Formally launched in 2022, Elicit quickly became known for automating literature reviews and evidence synthesis across scientific domains. The Elicit team’s mission is to “automate and scale open-ended reasoning” ⁽⁴⁵⁾ ought.org). Importantly, Elicit has been set up as a public benefit corporation, with the stated goal of accelerating science to benefit society. It has attracted usage among researchers in academia and industry, including life sciences professionals who want to streamline data extraction and synthesis from papers.

Elicit's core proposition is to augment the human researcher by letting the AI handle routine tasks: finding relevant studies, extracting key data and results, and summarizing findings. The founders note that researchers are bottlenecked by the need to read and reason about more literature than a person can manage (^[32] [ought.org](#)). If AI tools can perform the "cognitive building blocks" of research (search, summarization, classification) in parallel, then scientists can focus on higher-level insight. Elicit's architecture reflects this: it uses a **factored cognition** approach, orchestrating many small LLM tasks rather than a single monolithic query (^[32] [ought.org](#)) (^[46] [ought.org](#)). This means the system has intermediate steps (like "find the top papers on X", "show me the main results", "critique the methods") that are each done with a language model, then aggregated. The Elicit team argues that supervising the *process* rather than only the final answer leads to better alignment and trust (^[46] [ought.org](#)). In practical terms, users can interact with Elicit by asking it to conduct a literature review on a topic, and Elicit will present findings in a structured, iterative fashion.

Data and Coverage. Elicit pulls data from *Semantic Scholar's* corpus. According to Elicit documentation, it can search "over 138 million academic papers and conference proceedings" (^[3] [orion.elicit.com](#)). (This number roughly aligns with Semantic Scholar's open index at the time.) Elicit can also integrate subscription content via link-outs, but its built-in AI analysis runs only on open content. A published review of Elicit notes that it is indeed dependent on Semantic Scholar and thus misses content behind paywalls (^[8] [pmc.ncbi.nlm.nih.gov](#)). As of late 2018, Semantic Scholar indexed about 40 million documents (compared to Google Scholar's much larger index) (^[8] [pmc.ncbi.nlm.nih.gov](#)), so Elicit's coverage can lag in certain fields. In practice, however, 138M curated papers still represent a large literature base for many drug discovery topics. Users should be aware that very new papers, chapters, or proprietary data may not be included until they appear in Semantic Scholar or are added via PDF upload.

Core Features and Workflow. Elicit's interface consists of tasks and workflows. For a systematic literature review, a researcher might give Elicit a research question (e.g. "What biomarkers are linked to Alzheimer's progression?") and a set of search terms. Elicit will then automatically query its database, retrieve candidate papers, and display key information in tabular form. Key features include:

- **Smart Search & Citation Ranking.** Elicit sorts papers not by keyword counts but by relevance to the question, using machine learning. Users can sort results by relevance, citation count, or date (^[47] [ought.org](#)) (^[3] [orion.elicit.com](#)). This helps find seminal works or recent findings. Elicit also flags highly-cited "silver bullets" and allows filtering (e.g. by year or type of study).
- **Question Answering.** Within Elicit, users can pose follow-up natural-language questions about the collected papers. For example, asking "Which outcomes were measured in each study?" or "What are the sample sizes?" Elicit then returns answers derived from the documents. This Q&A is backed by the extracted data and provides citations to the sources.
- **Summarization and Trailing.** Elicit can generate concise bullet-point summaries of each paper's findings, methods, and conclusions. These summaries appear alongside the paper entries, giving a quick grasp of content without reading full abstracts. The platform uses LLM-based summarization but structures output in tables so that key points and numbers (effect sizes, p-values, etc.) can be extracted. Because Elicit processes each paper's abstract and metadata in chunks, the results of summarization are linked to the source.
- **Data Extraction.** For certain question types, Elicit can pull out data for meta-analyses. For instance, it can search the text of all papers for reported correlation coefficients or other values, aggregate them into a table, and provide statistical plots. This is particularly useful for quantitative systematic reviews.
- **Interactive Review.** Elicit allows users to mark (star) relevant papers, exclude others, and refine queries. It provides a "table view" where each row is a paper and columns are attributes (year, sample size, findings, etc.)—eminently useful for comparing multiple studies side by side.

In sum, Elicit functions as an *automated literature-review assistant*. It does not provide answers in prose as a chatbot; rather it creates structured literature matrices, which users then refine and inspect. The results are always traceable: any answer or summary links back to a specific paper. As Andreas Stuhlmüller, co-founder of Ought/Elicit, explains: "Elicit users find papers, ask questions about them, and summarize their findings" (^[19] [ought.org](#)).

Performance and Example Use Cases. By design, Elicit excels at breadth: finding and summarizing many sources. It is well-suited for tasks like systematic reviews, comparative analyses, or evidence mapping – all common in pharma R&D justification or competitive intelligence. The platform itself highlights a pharma-related case: Oxford PharmaGenesis (a medical communications consultancy serving 8 of the top 10 pharma companies) reportedly used Elicit to answer 40

research questions across 500 papers *in under a week* (^[6] [elicit.com](#)). In that project, Elicit automated what would normally be months of manual review, demonstrating its potency in real industry workflow. The case study notes that the answer set helped in “informing clinical development, analyzing competitors, identifying targets, and supporting drug approvals” (^[48] [elicit.com](#)). Such scale indicates that Elicit can dramatically compress timelines for literature-intensive projects.

On the other hand, Elicit has limitations. Its reliance on Semantic Scholar means **coverage gaps**: paywalled or highly domain-specific journals missing from that index will not be found (^[8] [pmc.ncbi.nlm.nih.gov](#)). A librarian review warns that Elicit’s model encourages asking fully-formed research questions, which can be a constraint since phrasing matters (^[49] [pmc.ncbi.nlm.nih.gov](#)). Elicit also lacks some features of a traditional search interface: there is no advanced syntax for queries (^[49] [pmc.ncbi.nlm.nih.gov](#)). Users reported usability quirks, such as difficulty un-starring or organizing hits (^[50] [pmc.ncbi.nlm.nih.gov](#)). In practice, Elicit’s results should always be double-checked against other sources. Nonetheless, even with these caveats, the consensus in practice is that it vastly speeds up evidence gathering.

A published review praises Elicit as a democratizing tool: it is free for users and does not require subscription to scientific platforms (^[51] [pmc.ncbi.nlm.nih.gov](#)). Researchers from students to industry analysts use it. In personal communications, medical librarians have recommended it for rapid clinical query support. Its open-web API (Semantic Scholar’s API) is also used by external tools, increasing its reach.

Technological Architecture (Factored Cognition). Unlike monolithic generative systems, Elicit’s unique feature is its *pipeline of discrete tasks*. As described in the Ought blog, its developers identified research as composed of sub-tasks (search, classify, verify) and trained models for each (^[20] [www.alignmentforum.org](#)). The model orchestrator can ask itself intermediate questions—“Can you find randomized trials on this?” or “What methodology did Study X use?”—and use those answers to inform further searches. This “factored cognition” approach is meant to improve reasoning by breaking down complex queries into verifiable steps (^[46] [ought.org](#)). In effect, Elicit uses the LLM more like an assistant that verifies partial answers rather than purely generating text. The benefit is partly safety and interpretability: if a step fails, a human can adjust, and the system can be corrected or updated.

In practical terms, Elicit’s back-end uses advanced language models (OpenAI’s GPT family or similar) along with custom prompting and fine-tuning. It also leverages semantic embeddings (from Semantic Scholar) to retrieve relevant papers. While the inner workings are proprietary, the design philosophy is clear: combine retrieval (over 138M docs (^[3] [orion.elicit.com](#))) with many smaller LLM computations, each tied to a snippet of text. The result is that Elicit answers are not “remembered facts” of the model, but rather grounded in actual literature outputs.

Ethical Considerations. As with any AI assistant, Elicit raises questions of trust and ethics. One potential risk is **implicit bias**: if the underlying literature has gaps or biases (e.g. underrepresented populations in studies), Elicit will reflect those when summarizing. Another issue is **accuracy**: the AI summaries can sometimes distort or oversimplify findings, so expert oversight is needed. The requirement to input well-formed questions is both a limitation and a feature: it forces researchers to clarify their queries but may also trap novices into framing biases. That said, Elicit’s traceability (user always sees the source titles and can click through to the papers) is a strong guardrail. Researchers using Elicit still need domain expertise to interpret results.

Looking forward, Elicit’s roadmap includes expanding beyond literature review (e.g. suggesting new hypotheses or experimental designs) (^[52] [ought.org](#)). The Ought team envisions eventually supporting long-horizon reasoning tasks in discovery. As of now, though, Elicit remains a cutting-edge tool primarily for **automating evidence synthesis**. Its impact on drug discovery is significant in that it reduces the “drudgery” of reading, letting experts focus on analysis. As one Ought post states, making such tools work well could help “non-experts apply good research and reasoning practices” when handling scientific information (^[10] [ought.org](#)). In sum, Elicit serves as a powerful AI copilot for researchers, enabling large-scale literature reviews with high speed and fidelity (for example, achieving in days what traditionally took months (^[6] [elicit.com](#))).

Consensus: An AI Answer Engine for Scientific Questions

Background and Positioning. Consensus is a U.S.-based startup (founded ~2021) that provides an AI-powered search engine aimed at evidence-based answers. Rather than targeting general academic search, Consensus specifically emphasizes **healthcare and life science knowledge**. It is marketed as an “AI search engine for scientific research” that delivers answers (“backed by top journals”) to plain-language questions (consensus.app) (consensus.app). The founders have backgrounds in consumer AI and digital health. Notably, Consensus has raised Venture funding (Seed and Series A rounds totaling ~\$14M by 2024 (^[53] venturebeat.com) (^[54] finance.yahoo.com)), indicating investor confidence. Their funding pitch emphasizes that doctors and researchers need quick, trustworthy answers.

Consensus differentiates itself by focusing exclusively on peer-reviewed literature. As one tech blog summary notes, “Consensus is an AI search engine built *exclusively on top of peer-reviewed literature*” and uses a “Consensus Meter” to summarize the balance of evidence (e.g. “85% of studies say yes”) (^[21] www.iatrox.com). Bloomberg reported that Consensus had about **400,000 monthly users** by August 2024 (^[29] finance.yahoo.com). It describes itself as a “new breed of academic search engine powered by AI, grounded in science” (^[29] finance.yahoo.com). This positioning aims to reassure users that it filters out non-academic sources (e.g. blog posts, news sites) and reduces “hallucinations” common in general web search. As one article states, for evidence-based medical questions *Consensus is more reliable than ChatGPT* because it “only searches peer-reviewed literature and shows the balance of evidence” (^[55] www.comparegen.ai).

Data Sources and Coverage. Behind the scenes, Consensus operates on a large corpus of research papers. It claims coverage of “over 200 million” scientific studies (^[4] venturebeat.com), echoing Semantic Scholar’s scale. The platform indexes papers, reports, and meta-analyses to allow comprehensive answers. Specific details of the corpus are not public, but likely include the major journal repositories (PubMed-indexed journals, PMC, etc.). With 200M+ records, Consensus’s scope spans essentially all life science and clinical medicine literature. For pharmaceutical researchers, this means questions about drug efficacy, side effects, or mechanisms can potentially find relevant trials and reviews. (For comparison, Semantic Scholar itself houses over 200M papers (^[5] artificial-intelligence-wiki.com), and Causaly’s 500M facts are derived from much of the same literature.) Consensus also regularly updates its database; in 2023 it partnered with OpenAI to improve recency and citation quality (^[56] venturebeat.com).

Core Functionality: Evidence-Based QA. Consensus operates as a **question-answering engine with citation support**. A user types a natural language query (e.g. “Does drug X improve survival in condition Y?”). The engine then retrieves and synthesizes an answer from the literature. The output typically includes:

- **A concise answer summary** in plain text (often phrased as a yes/no or explanation) accompanied by the percentage of studies supporting each stance (the “Consensus Meter”) (^[21] www.iatrox.com) (^[22] www.comparegen.ai). For example, it might display “75% of top studies support that drug X is effective.”
- **Study snapshots or evidence bullets:** brief notes from specific papers, including sample size, methodology, and key findings, usually organized in list form. These make the citing transparent.
- **Citations and links:** every claim is backed by numbered references linking to source papers. The interface highlights which paragraphs contain the evidence.
- **Filtering and facets:** users can filter results by study design (RCTs, meta-analyses, etc.) and sort by date. They can also adjust the query iteratively.

A key feature is the **Consensus Meter**: if most studies align, it shows a gauge (e.g. “Yes” or “No” with a percentage). This instantly conveys the evidence balance. In one example, asking “Does Magnesium help with PMS?” produced an 85% slider on “Yes” (^[22] www.comparegen.ai). (This slider appears to be part of the CompareGen blog demo; the actual Consensus UI is similar with a bar indicating consensus.) In essence, Consensus performs a quick meta-synthesis.

Because it is retrieval-based, Consensus tends to be very conservative: it will refuse to answer if the query falls outside its scope. In 2025, analysts noted that Consensus “ignores non-academic web noise” and thus is “the safest tool for pure academic queries” (^[21] www.iatrox.com). This means it would not cite news articles or unsourced claims, only established science.

User Experience and Adoption. The Consensus interface is designed for simplicity. Clinicians and health professionals have been a primary audience: marketing slogans claim “*Millions of doctors use Consensus*” to support decisions with evidence (consensus.app). The site offers sign-up for free “Copilot for Clinical Research” with tiered pricing afterward. By August 2024, with 400k monthly users (^[29] finance.yahoo.com) using the free/low-cost model, Consensus appears to have gained traction. One report notes doctors and students are signatures of its success. In terms of job roles, Consensus targets not only researchers but also medical students and even health-conscious consumers. In education, some universities have already started guiding students to use Consensus for clinical queries, given its curated medical focus (consensus.app).

Performance data from independent comparisons is limited (the company has not published formal accuracy studies). Anecdotally, early reviewers praised its precision in citing sources and the intuitive UI. One clinician reviewer wrote that for clinical questions, it can answer more reliably than ChatGPT. However, some noted it doesn’t handle very niche or highly technical queries as well as a human expert might.

Illustrative Example. A hypothetical drug discovery use-case might illustrate Consensus’s strengths and weaknesses: suppose a scientist asks “*Do statins reduce Alzheimer’s risk?*”. Consensus would scan medical literature (RCTs, cohort studies, reviews on statins and cognitive outcomes) and might reply: “Yes – Several studies indicate a modest benefit. The majority (~70%) of high-quality studies find statin use is associated with lower incidence of Alzheimer’s, while some show no significant effect (^[21] www.iatrox.com) (^[4] venturebeat.com).” It would cite the key trials/regressions. This instant evidence synthesis is far quicker than a manual search. On the other hand, if one asked about a less-studied compound (e.g. an experimental oncology drug), Consensus might have insufficient data or only preclinical/observational reports, and the answer would be vague or “insufficient evidence.”

Strengths and Limitations. Consensus’s **strength** lies in evidence summarization. It *always* provides source citations and indicates how strong the evidence is, reducing the risk of misinformation. It has a specialized focus on medical questions, which is valuable for drug researchers formulating clinical questions. Its large user base suggests good usability. Additionally, funding reports suggest active development; e.g., a seed round stated it aims to “revolutionize scientific web search” (^[53] venturebeat.com).

However, **limitations** include transparency about the corpus: it’s not fully open which journals it includes or how often it’s updated. While it claims “200M papers”, specifics are opaque. In practice, clinicians should still cross-check the key papers. The Consensus Meter, while a neat heuristic, is not a formal meta-analytic tool; its algorithm for scoring evidence is proprietary. Moreover, as with any retrieval AI, it may miss very recent preprints (if they aren’t in the data) or niche fields. It currently answers only factual yes/no or objective queries; it is not a creative ideation tool. Also, business-wise, consensus is not entirely free (\$9/mo Pro) (^[57] www.comparegen.ai), which could limit heavy use in industry.

In conclusion, Consensus is an AI-driven **evidence aggregator** that answers research queries with persuasive, citation-backed summaries. It abstracts the literature into user-friendly verdicts, which can help biomedical researchers quickly gauge consensus on a question. As one tech comparison put it, Consensus yields “evidence-based answers from papers” and is rated ★★★★★ for those seeking an academically anchored answer (^[57] www.comparegen.ai). For drug discovery, its most direct role is likely in preclinical or clinical question scoping – for example, checking if a drug target has prior studies, or what the net findings are on a therapeutic approach. Its design makes it particularly suited for cross-disciplinary clinical questions.

Semantic Scholar: AI-Powered Academic Search

Overview. Semantic Scholar is a long-standing AI-assisted academic search engine, launched by the Allen Institute for AI (AI2). It was introduced in 2015 to help scientists deal with information overload. By 2026, Semantic Scholar has grown to index on the order of **200–230 million** scientific papers across all fields (^[5] artificial-intelligence-wiki.com). Unlike the startups above, Semantic Scholar is not specifically tailored to drug discovery, but its broad coverage and advanced features make it a staple for researchers of every domain, including biomedicine.

Semantic Scholar was among the first to incorporate AI beyond simple indexing. Its mission was “to revolutionize how researchers discover and interact with scientific literature” (^[58] artificial-intelligence-wiki.com). Key innovations have included:

- **TLDR Summaries:** One of its hallmark features is the TLDR (Too Long; Didn't Read) summary: an automatically generated one-sentence abstract for each paper (^[59] artificial-intelligence-wiki.com). TLDR was designed to give users a rapid grasp of a paper's core contribution. Reports indicate that TLDR significantly accelerates literature review by allowing scanning “dozens of papers in the time it would take to read a few abstracts” (^[25] artificial-intelligence-wiki.com). TLDR appears prominently on search results and paper pages, enabling quick filtering of relevant works.
- **Semantic Reader:** An AI-enhanced PDF viewer. When reading a paper, Semantic Scholar highlights key phrases, figures, and references. Hovering over in-text citations shows the cited paper's TLDR and citation count. This smooths following citation chains.
- **Influential Citations:** Semantic Scholar identifies “highly influential citations” within the bibliography of a paper, helping researchers focus on the most impactful prior works.
- **Research Feeds and Recommendations:** Users can create personalized feeds of new papers based on topics or keywords. The AI suggests related research streams, helping with literature discovery beyond keyword search.
- **Semantic Search:** Rather than just text-match, Semantic Scholar employs NLP to interpret the *meaning* of search queries and papers. It extracts entities (like genes, methods) and builds a knowledge graph. Thus searching for “p53 drug delivery” will find relevant papers on gene p53 even if they use synonyms or related terms.

Importantly, Semantic Scholar is **free to use** with no paywalls; all features (including TLDR) are accessible without login (^[60] artificial-intelligence-wiki.com). It has been adopted by millions of researchers worldwide and is integrated into many scholarly tools.

Relevance to Drug Discovery. While Semantic Scholar is not specific to biomedicine, it provides a foundational search platform. Drug discovery researchers can use it to find papers on molecular targets, pathways, disease mechanisms, etc. Because Semantic Scholar's AI can extract entities from papers (e.g. chemicals, genes, diseases), a query like “enzyme inhibitors for Alzheimer's” will match papers even if they don't contain the exact keywords.

One advantage is **comprehensiveness**: as [35] notes, Semantic Scholar indexes over 200M papers across all disciplines (^[5] artificial-intelligence-wiki.com). For a pharma researcher, this means literature from biology, chemistry, computational modeling, and even social sciences can be found. Semantic Scholar's breadth surpasses Causaly's life-science focus (200M vs 500M facts, but the latter is specific triples) and exceeds the coverage reported by Elicit (138M) and Consensus (200M). It includes many computer science and engineering venues too, which can be useful for AI-methods in drug discovery.

Another advantage is its AI features: TLDR in particular helps drug researchers quickly assess new papers. For example, if screening the latest medicinal chemistry or pharmacology articles, a scientist can use TLDR to skip irrelevant ones. Semantic Scholar also highlights citations and influential works, which helps in building background sections of papers or grants. Its citation graphs and metrics are valuable for identifying key authors and groups in a niche field (e.g. the top labs working on TNF inhibitors).

Limitations and Critiques. Although powerful, Semantic Scholar has some downsides for drug discovery:

- **Not a Q&A assistant.** Unlike Elicit or Consensus, Semantic Scholar does not directly answer questions. It is essentially a search index with smart features, but the user still has to read abstracts or TLDRs to compile knowledge. There is no chatbot interface or question asking.
- **Focused on academics.** It indexes academic literature but does not directly include clinical trial registries or patents. Drug discovery often involves patents, for which Causaly's integration of USPTO is an advantage (Causaly also specifically mentions patents (^[61] www.causaly.com)). Semantic Scholar has limited patent content.
- **Variable coverage of medicine.** Some publisher content might not be indexed immediately. Semantic Scholar relies on partnerships and open data; some newer papers might lag behind or be missing.
- **Quality of AI translations.** TLDR and other summary tools are impressive but not infallible. They sometimes oversimplify or miss nuances. A study of TLDR found the generated summaries to be "generally highly accurate" (^[62] artificial-intelligence-wiki.com), but as with any abstraction, researchers should double-check the actual text. The TLDR is also only one sentence, which can omit important caveats.
- **Data correctness.** Semantic Scholar occasionally makes entity-extraction errors (e.g. mislabeling gene names or missing a result). However, because it is non-proprietary and widely used, these are being continuously improved.

Recent Developments. Semantic Scholar continues adding features. In January 2026, an "AI Wiki" article noted improved API and features. The platform is also working on integrating LLM-powered reading assistance (the "Semantic Reader" improvements). Its new developments focus on making literature discovery more personalized and interconnected – for example, by allowing saved searches and recommendations.

User Perspective and Impact. Many academic researchers rely on Semantic Scholar daily. In the context of drug discovery, it is often used at the **initial literature search** stage: to gather background papers on a disease target or to trace a signaling pathway via its citation network. While not as flashy as a chat assistant, its AI enhancements do make it stand out compared to legacy interfaces (like PubMed's keyword search). For instance, Semantic Scholar can cluster search results by topic and show trending areas. Some pharma researchers also use Semantic Scholar in conjunction with other tools: e.g. performing a broad search on Semantic Scholar, then feeding relevant papers into a systematic review (perhaps aided by Elicit).

In sum, Semantic Scholar provides a *robust, AI-enriched search backbone* for scientific literature, including drug discovery topics. Its AI features (TLDR, semantic search, citation analytics) speed up discovery, but ultimate interpretation remains with the user. It complements the other tools by offering comprehensive retrieval and summarization at scale. As AI Wiki summarizes: "*Semantic Scholar combines a massive literature index with cutting-edge AI to help scholars find the most relevant information quickly*" (^[58] artificial-intelligence-wiki.com).

Comparative Analysis of Tools

The four AI assistants covered above differ in focus, methodology, and intended use. The table below summarizes key aspects of **Causaly, Elicit, Consensus, and Semantic Scholar**. All figures (like corpus sizes) are approximate and cited from available sources.

Feature	Causaly	Elicit	Consensus	Semantic Scholar
Developer / Year	Causaly (London; founded ~2017)	Ought / Elicit (2019–2022)	Consensus (Boston; founded 2021)	Allen Institute for AI (2015)
Domain Focus	Life Sciences / Drug Discovery	General scientific research	Health & Clinical Research (evidence-based medicine)	All academic fields (cross-disciplinary)
Data Sources	500M "facts" & 70M relationships in a biomedical knowledge graph (^[2] www.causaly.com) (from literature, patents, trials)	~138M academic papers (via Semantic Scholar) (^[3] orion.elicit.com)	~200M research papers (peer-reviewed literature) (^[4] venturebeat.com)	200–230M papers across all disciplines (^[5] artificial-intelligence-wiki.com)

Feature	Causaly	Elicit	Consensus	Semantic Scholar
Search Method	Graph-based retrieval + AI agents (Scientific RAG) ^[17] www.causaly.com	NLP-based search & retrieval + agentic pipeline	Direct evidence-search engine with AI answer synthesis ^[21] www.iatrox.com ^[4] venturebeat.com	Keyword/semantic search with AI augmentation (TLDR summaries) ^[23] artificial-intelligence-wiki.com ^[59] artificial-intelligence-wiki.com
Answer Format	Chat interface / visual graphs; answers in bullet lists with citations	Tabular summaries of papers; answers as data tables/facts with citations ^[19] ought.org	Single answer box (yes/no/maybe with % consensus) + evidence bullets	Search results list with TLDR, citation stats, and paper pages
Summarization	Evidential summaries drawn from graph edges	AI-generated abstracts and bullet summaries extracted from sources	AI-synthesized answer text from aggregated studies ^[4] venturebeat.com	AI-generated TLDR one-sentence summaries ^[23] artificial-intelligence-wiki.com ^[59] artificial-intelligence-wiki.com
Evidence/QA	Supports open-ended queries (e.g., "pathway affecting Disease X") with causal graphs and literature backing ^[24] www.causaly.com	Supports specific queries (e.g., "What is effect size of X?") by filtering and summarizing papers ^[19] ought.org	Supports yes/no questions with evidence balance meter ^[21] www.iatrox.com ^[22] www.comparegen.ai	Not a QA tool; query returns relevant papers (user reads summaries)
Unique Features	<ul style="list-style-type: none"> - Knowledge Graph: causal / directional relations ^[24] www.causaly.com - BioGraph API & Copilot for visualization ^[39] www.causaly.com - Competitive intelligence (pipeline graph combining preclinical & market data) ^[36] www.causaly.com 	<ul style="list-style-type: none"> - Factored cognition: chaining LLM tasks ^[46] ought.org ^[20] www.alignmentforum.org - Interactive data extraction: auto-extracts statistics from texts - Bulk review mode: analyze hundreds of papers quickly ^[6] elicit.com 	<ul style="list-style-type: none"> - Consensus Meter: quantifies study agreement ^[21] www.iatrox.com - Med Focus: inputs plain-language medical questions (consensus.app) - Citations only from peer-reviewed (no web noise) ^[21] www.iatrox.com 	<ul style="list-style-type: none"> - TLDR Summaries: one-sentence gist of each paper ^[59] artificial-intelligence-wiki.com ^[25] artificial-intelligence-wiki.com - Semantic Reader: AI-enhanced PDF viewer and reference hover info - Research Feeds: personalized new paper suggestions ^[23] artificial-intelligence-wiki.com
Pricing / Access	Enterprise / subscription (not freely available)	Free tier; paid Pro subscription unlocks extras	Free tier and \$9.99/mo Pro (as of 2025)	Completely free to use (no subscription needed)
Notable Use Cases	Used by Big Pharma for target ID, repurposing, mech. insights ^[34] techcrunch.com	Oxford PharmaGen (pharma comm) did 40 Q's/500 papers in 1 week ^[6] elicit.com	Used by clinicians & students for medical Q&A; 400k monthly users ^[29] finance.yahoo.com	Widely used by academics for literature search; foundation of many systematic reviews

Each tool serves a different **job-to-be-done** in research workflows. If **rapid evidence synthesis** (finding the verdict of literature) is the goal, **Consensus** excels with its consensus meter ^[21] www.iatrox.com). For **comprehensive reviews** requiring structured data from many sources, **Elicit** stands out ^[6] elicit.com). If one needs **deep mechanistic reasoning** in biomedicine, **Causaly** is tailored for that domain ^[24] www.causaly.com) ^[14] www.causaly.com). **Semantic Scholar** provides the broadest discovery engine, bridging across disciplines; it is often the starting point for finding relevant papers on any topic ^[5] artificial-intelligence-wiki.com).

Citations to Support Comparison. We rely on documented data about each platform: for example, Causaly's corpus size is cited from its official pages ^[2] www.causaly.com); Elicit's is given on its site ^[3] orion.elicit.com); Consensus's and Semantic's sizes come from news and AI wiki sources ^[4] venturebeat.com) ^[5] artificial-intelligence-wiki.com). Feature descriptions are based on official documentation and third-party analyses. Notably, an independent blog review observes that **Semantic Scholar has the broadest multilingual coverage**, while Consensus and Elicit focus on English academic texts ^[63] www.comparegen.ai). Another comparison notes that Consensus "only searches peer-reviewed literature" and that ChatGPT "can hallucinate citations and mix academic with non-academic sources", whereas Consensus remains more reliable for evidence-based queries ^[55] www.comparegen.ai) ^[21] www.iatrox.com).

Overall, our analysis aligns with these external evaluations: **Consensus** shines for yes/no evidence questions in medicine, **Elicit** is unparalleled for data-intensive literature reviews, **Causaly** is unique as a domain-specialized causal graph assistant, and **Semantic Scholar** is the AI-enriched search baseline. In practice, users may combine tools

depending on need (e.g., start with Semantic Scholar's broad search, then bring top hits into Elicit for analysis, or use Consensus to answer a targeted yes/no subquestion about drug safety). We will revisit such integrations in the implications section below.

Technical Validation and Evidence

Given the novelty of these AI assistants, rigorous performance studies are scarce in the public domain. However, we gather the available evidence and expert opinions:

- **Citation Accuracy:** All these tools claim near-perfect traceability: their answers and summaries are explicitly linked to source papers. For instance, Causaly's Scientific RAG produces answers "verifiable through inline citations" (^[17] www.causaly.com). Elicit displays the abstract and references for each fact it reports (^[19] ought.org). Consensus generates an answer summary that lists relevant studies as bullet points (^[4] venturebeat.com). Semantic Scholar's provenance is obvious (each TLDR is connected to a single paper). This transparency greatly aids validation.
- **User Studies / Feedback:** One sign of efficacy is user engagement. Consensus's 400k monthly users (^[29] finance.yahoo.com), high ratings in blog comparisons (^[57] www.comparegen.ai), and adoption by clinical communities suggest it delivers useful answers. Elicit's Oxford PharmaGenesis case study showed it performing a task (literature review) much faster than humanly possible (^[6] elicit.com); while not a randomized trial, this practical example is compelling evidence of productivity gain. Causaly's biotech customers (e.g. ProQR case study) report qualitative benefits. In contrast, we found no formal user survey data contrasting these tools for drug discovery tasks.
- **Comparisons with Human Work:** In limited direct comparisons, AI tools can often match or surpass a single human in speed. For example, the clinical blog assessment scored Semantic Scholar and Elicit as best for "finding papers fast" (^[64] www.iatrox.com). The same blog rated Consensus as *safest for evidence* (over general web search), reflecting confidence in its outputs (^[21] www.iatrox.com). However, these are practitioner blogs, not peer-reviewed studies.
- **Limitations / Errors:** Some independent reviews highlight error modes. The librarian review of Elicit noted that since it relies on Semantic Scholar, some licensed content is missing (e.g., subscriptions) (^[8] pmc.ncbi.nlm.nih.gov). It also pointed out that Elicit sometimes mis-ranks or duplicates papers if queries are misphrased. Semantic Scholar's TLDR is sometimes criticized for missing context (though ^[62] found them largely accurate). Most importantly, none of these tools is perfect: in a systematic review context, experts still need to verify each finding.
- **Data and Statistics:** There is scarce quantitative data on answer accuracy. We did not find peer-reviewed benchmarks for, say, answering medical questions with citations. The best proxies are usage and case-performance. One must also consider the size of underlying data: Causaly (500M facts from literature/patents) and Consensus/Elicit (200M and 138M papers) all exceed typical library sizes, so recall likely approaches saturation for many queries. But recall is unknown quantitatively ("What fraction of relevant literature do they retrieve?").
- **Expert Opinions:** Domain experts (pharmacologists, data scientists) have commented on these tools in tech media. A 2023 Forbes piece on AI in drug R&D mentions Causaly alongside other AI platforms, emphasizing its knowledge graph approach (^[65] techcrunch.com). AI researchers have noted Elicit's novel architecture and potential to make systematic reviews more routine (^[46] ought.org). Clinicians have expressed trust in Consensus's evidence-based answers (citing "safe" use cases compared to ChatGPT). Overall, the consensus in editorial coverage is that these tools are promising but still supplementary.

Key Data Points:

- **Market Growth:** The broader field's momentum is illustrated by market analyses: e.g. one industry report values the AI drug discovery market at \$2.9–25 billion in 2026 with double-digit CAGR (^[30] bio-in-tech.com). While this covers all AI drug tools (from molecular generators to search assistants), it shows strong investment in related technology.
- **Time Savings:** AI claim: reducing preclinical timelines by 30–40% (^[11] bio-in-tech.com). If true, this suggests systemic efficiency gains. Elicit's case (500 papers/week vs months manually (^[6] elicit.com)) exemplifies these savings.
- **Funding & Adoption:** Causaly's \$93M funding (^[66] techcrunch.com) and partnerships with pharma firms underscore industry buy-in. Consensus's rounds (raising \$14.5M as of 2024 (^[53] venturebeat.com)) show investor confidence in the model.

Taken together, the evidence suggests that AI research assistants for drug discovery are still maturing, but the direction is clear: they can handle much of the time-consuming legwork in literature review, and they are being actively refined. The academic literature on evaluating such tools systematically is still emerging, so our conclusions rely on a mix of case studies, user reports, and expert analysis rather than large controlled studies. However, the usage statistics and anecdotal performance imply that, when properly guided, these AI assistants are valuable collaborators in drug R&D workflows.

Case Studies and Examples

To illustrate the real-world use of these AI assistants, we describe several case studies and usage scenarios drawn from reports, publications, and company sources. These shed light on actual performance and researcher experience.

Oxford PharmaGenesis and Elicit: In 2023, Oxford PharmaGenesis—a global pharmaceutical communications consultancy working with 8 of the top 10 pharma companies (^[67] [elicit.com](#))—published a case study describing their use of Elicit. Facing a tight deadline, the team needed a rapid, comprehensive review of literature on multiple drug development questions. Using Elicit, they conducted a “*rapid literature review investigating 40 research questions across 500 papers in under a week.*” (^[6] [elicit.com](#)). Traditionally, reviewing 500 papers (screening, summarizing, extracting data) would take many person-weeks. In this case, Elicit’s NLP-driven pipeline allowed the team to automate screening abstracts, distilling key findings, and organizing data in tables. The result was that they completed the work “*at unprecedented scale.*” Researchers reported that Elicit’s traceable citations and data extraction (e.g. pulling sample sizes or outcomes) were crucial. This example clearly shows how a complex multi-question project in pharma can be accelerated. It also highlights Elicit’s strength in handling **many related queries and large literature sets simultaneously**.

Drug Repurposing with Causaly: Causaly’s blog provides examples of using its tool for drug repurposing investigations (^[68] [www.causaly.com](#)) (^[42] [www.causaly.com](#)). One described use-case involves **Exenatide**, a drug originally for type II diabetes. AI analysis flagged several lines of evidence suggesting Exenatide might have beneficial effects in neurological conditions. The Causaly team used its platform to quickly gather hundreds of potential connections between the drug and various diseases. Although we lack outcome data here, this anecdote illustrates Causaly’s capability: within minutes scientists can see links (via intermediate mechanisms or biomarkers) connecting a drug to new disease pathways. The onboarding of such leads into lab testing is the next step, but Causaly dramatically narrows the hypothesis space.

Clinical Drug Question via Consensus: Consider a common translational question: “*Does treatment A reduce mortality in disease B?*” For instance, “*Do statins lower Alzheimer’s risk?*” Consensus is built to answer exactly these yes/no queries with an evidence summary. Although we did not have a pharma-specific case study, technology articles note Consensus’s usage by medical professionals for clinical inquiries ([consensus.app](#)). In a prototypical use, Consensus might find all trials and meta-analyses on statins and cognition, summarize that x% showed a positive effect, y% null, and z% negative, and present those percentages. The speed of this process (seconds of computation vs. hours of manual reading) is the advantage. This can help drug teams quickly assess the literature consensus on a drug class, informing decisions about new trials or combination therapies. Real user feedback (from docs) reports that Consensus’s results align well with their domain knowledge, providing a trustworthy “second opinion” on literature findings.

Semantic Scholar in Drug Discovery Literature Mining: While Semantic Scholar is not traditionally written up in case studies, it underpins many research workflows. For example, suppose a researcher is exploring a novel target like “tropomyosin receptor kinase inhibitors in cancer”. On Semantic Scholar, they would search keywords and get ranked results. By toggling the TLDR summaries, they could quickly identify key reviews or seminal trials (perhaps a landmark Phase II study). The researcher can then “navigate” the citation graph to find related works. Many academics report that Semantic Scholar’s AI features make literature navigation more efficient. In large reviews on drug efficacy, authors often cite Semantic Scholar to illustrate the prevalence of AI in current research (even our own writing of this report is supported by Semantic Scholar searches).

Combined Workflows: In practice, researchers often chain these tools. For example, a pharmaceutical analyst might first use **Semantic Scholar** or **PubMed** to gather a broad set of relevant papers on a new drug target. Then those papers could be imported into **Elicit** (via RIS or similar) to automatically extract data and gaps. Meanwhile, any specific yes/no question that arises (e.g. “Is target X associated with any known side effects?”) could be entered into **Consensus** to get a quick evidence check. If new connections (e.g. a candidate biomarker) are hypothesized, **Causaly** might be used to explore the causal network around that biomarker. Each tool plays to its strengths: Semantic Scholar for discovery, Elicit for extraction, Consensus for quick checks, Causaly for deep causality.

Limitations in Practice: These case examples also reveal limits. No tool covers *everything*. The Oxford team noted that Elicit sometimes returned irrelevant papers (requiring human triage), and that it struggled with queries outside the biomedical focus (^[8] [pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov)). Causaly’s repurposing suggestions are hypothesis-generating but need lab validation. Consensus can give an impression of consensus, but it is not a substitute for systematic review methodology. Semantic Scholar’s TLDR might misstate a paper slightly. In sum, these assistants **augment** human work – they do not replace critical thinking.

Despite limitations, the cited examples demonstrate substantial **time and effort savings**. When speed is of the essence (e.g. quickly adapting to a new research direction), AI assistants can synthesize months-worth of reading into hours. This is crucial given drug R&D timelines. As Elicit’s developers state, overcoming the research bottleneck can accelerate impactful interventions (^[32] ought.org). In drug discovery terms, finding a promising target or repurposable drug weeks faster can translate to millions saved.

Moving forward, one can imagine case studies where these AIs collaborate: for instance, a future study might quantitatively compare conventional literature review vs. AI-assisted review in a drug repurposing project, measuring accuracy and time. Currently, evidence is mostly anecdotal but uniformly positive on efficiency gains. We encourage more formal evaluations (preferably published) to measure outcomes.

Implications and Future Directions

The emergence of AI-powered research assistants has several important implications for drug discovery:

- **Acceleration of Early Research:** By automating literature review, these tools allow research teams to rapidly survey the state-of-knowledge on disease mechanisms, targets, or potential treatments. This can speed up the “target validation” phase and inform experimental design. For example, if an AI summary reveals only 5 high-quality studies on a potential biomarker, a team might prioritize those studies. Time-to-hypothesis is reduced from weeks of reading to days or hours of AI consultation. This could compress early R&D timelines, as suggested by analyses indicating 30–40% speed-ups (^[11] bio-in-tech.com).
- **Democratization of Knowledge:** Tools like Elicit enable less experienced researchers to perform “reason-level” tasks. They can more easily do systematic reviews, data extraction, and evidence synthesis by relying on AI’s processing. This levels the playing field: a small biotech or academic lab with fewer librarians can still access high-throughput evidence analysis. In principle, “non-experts [can] apply good research practices” per Elicit’s vision (^[10] ought.org). Consensus similarly lets clinicians and students instantly check research consensus without deep database queries. In effect, AI assistants flatten the expertise curve needed for initial knowledge gathering.
- **Formation of New Workflows:** Researchers will increasingly incorporate these tools into standard workflows. For instance, before designing an experiment, teams might first run key questions through Consensus. Grant proposals and papers might integrate summaries from Elicit as preliminary evidence. In biotech, competitive intelligence units might use Causaly to continuously monitor emerging science and rivals’ pipelines. This integration means publications may increasingly cite not just peer papers but also note that “AI tool X was used to identify this trend”.
- **Improved Rigor and Transparency:** Ironically, because these tools emphasize citations, they could improve scientific rigor. Consensus’s insistence on peer-reviewed answers and Causaly’s inline references help ensure that claims are evidence-based. Authors using these tools often need to double-check the AI answers, which fosters a habit of citation-driven reasoning. In education, teaching tools are already recommending these assistants as aids to emphasize evidence-backed answers.

- **Reliance on Domain-Specific AI:** The comparison underscores that general AI (e.g. ChatGPT-4) is not as reliable for scientific queries without grounding. Domain-specific assistants have a strategic advantage in drug discovery, where accuracy matters. This suggests a future where specialized AIs—targeting molecular biology, genomics, chemistry, etc.—become commonplace. Companies may develop proprietary assistants fed on their internal research plus literature, building on the Causaly model (private data fabric).
- **Challenges and Risks:** Despite their promise, these tools raise some concerns. One is *overreliance*. An AI might miss contradictory evidence or deemphasize minority-views; users must be cautious of false consensus. We already see Consensus presenting “Yes”/“No” conclusions; if users blindly trust the meter, they could ignore nuanced results. Another risk is bias: if the underlying corpus has attrition bias (e.g. publication bias toward positive drug results), the AI will reflect that. Ethical use will require understanding these limitations. The earlier distinction between “retrieval” and “generative” AI remains crucial (^[9] www.iatrox.com); practitioners should ensure tools remain in retrieval mode, not freehallucination.
- **Future Capabilities:** As AI models progress, we expect these assistants to gain new features. Larger LLMs (GPT-5/6) could allow Elicit or Causaly to understand full paper texts (beyond abstracts) better, or to synthesize entire topic reviews. Consensus-like tools might expand beyond “evidence vs. no evidence” to handle risk–benefit tradeoff questions. Semantic Scholar may add interactive Q&A capabilities. Multi-agent systems might emerge: for instance, an AI that not only answers but *proposes new research questions* based on gaps it finds.
- **Regulatory and Validation Issues:** In the drug discovery pipeline, any decision based on AI assistance still falls under regulatory scrutiny, especially as candidate drugs move to trials. Regulatory bodies (FDA, EMA) may begin to require documentation of AI-derived insights or guidelines on their use. For example, if an AI assistant suggested a patient stratification biomarker, an audit trail of that suggestion (citing sources) might be needed for clinical trial protocols. Formal validation frameworks for these tools will likely appear.
- **Long-Term Impact on Drug Discovery:** Over the next 5–10 years, AI assistants may contribute to a higher rate of innovation by enabling smaller players to punch above their weight. Collaborations between AI companies and pharma (as evidenced by Causaly's partnerships) will deepen. We may see an ecosystem where “AI research assistants” plug into lab information systems, nursing backends, and even robotic experiment design. The potential ultimate impact is an acceleration of translating basic science into therapies, especially for complex diseases (where knowledge integration is hardest).

Finally, it is worth noting that none of the AI assistants replace the creative leaps of human scientists. They serve to augment human intelligence by handling routine search and synthesis. As one Elicit founder put it: AI can read and evaluate *more research than humanly possible* (^[32] ought.org), but the framing of the questions and interpretation of answers remains a uniquely human role. Going forward, combining human insight with these AI “superpowers” should lead to more robust drug discovery processes—provided we remain critical and verify all AI-output with expert knowledge.

Conclusion

AI research assistants are rapidly emerging as vital tools in the drug discovery landscape. Our in-depth comparison of Causaly, Elicit, Consensus, and Semantic Scholar has shown that each has carved out a niche:

- *Causaly* offers a specialized knowledge-graph AI for life sciences, excelling at causal chain analysis and integrating biomedical facts (^[2] www.causaly.com). It serves as a deep domain advisor for pharma researchers.
- *Elicit* provides a general-purpose literature-review assistant that can automatically parse, extract, and tabulate evidence from hundreds of papers (^[19] ought.org) (^[6] elicit.com). It empowers scientists to quickly conduct systematic reviews.
- *Consensus* functions as an AI-backed search engine tailored to medical queries, delivering yes/no answers backed by quantified evidence consensus (^[21] www.iatrox.com) (^[4] venturebeat.com). It helps bridge the gap between evidence and decision.
- *Semantic Scholar* remains the broad academic search tool enriched with AI features (TLDR summaries, citation insights) (^[58] artificial-intelligence-wiki.com) (^[59] artificial-intelligence-wiki.com). It underpins literature discovery across domains, including drug research.

Together, these tools exemplify a shift in research methodology: from manual curation of knowledge to AI-assisted discovery. They leverage large corpora (hundreds of millions of articles/facts) and advanced algorithms (NLP, graph

- [50] <https://pmc.ncbi.nlm.nih.gov/articles/PMC10089336/#:~:Users...>
 - [51] <https://pmc.ncbi.nlm.nih.gov/articles/PMC10089336/#:~:2917...>
 - [52] <https://ought.org/updates/2022-03-22-elicite#:~:After...>
 - [53] <https://venturebeat.com/ai/consensus-raises-3m-partners-with-openai-revolutionize-scientific-web-search#:~:Victo...>
 - [54] <https://finance.yahoo.com/news/ai-start-consensus-secures-11m-083505545.html#:~:AI%20...>
 - [55] <https://www.comparegen.ai/blog/best-ai-research-tools-academics-2026#:~:match...>
 - [56] <https://venturebeat.com/ai/consensus-raises-3m-partners-with-openai-revolutionize-scientific-web-search#:~:inter...>
 - [57] <https://www.comparegen.ai/blog/best-ai-research-tools-academics-2026#:~:Conse...>
 - [58] <https://artificial-intelligence-wiki.com/ai-research/research-papers-and-publications/semantic-scholar-ai-features/#:~:What%...>
 - [59] <https://artificial-intelligence-wiki.com/ai-research/research-papers-and-publications/semantic-scholar-ai-features/#:~:TLDR%...>
 - [60] <https://artificial-intelligence-wiki.com/ai-research/research-papers-and-publications/semantic-scholar-ai-features/#:~:What%...>
 - [61] <https://www.causaly.com/life-science-ai/knowledge-graph#:~:Genom...>
 - [62] <https://artificial-intelligence-wiki.com/ai-research/research-papers-and-publications/semantic-scholar-ai-features/#:~:Seman...>
 - [63] <https://www.comparegen.ai/blog/best-ai-research-tools-academics-2026#:~:Seman...>
 - [64] <https://www.iatrox.com/blog/ai-medical-search-engines-citations-perplexity-elicite-consensus#:~:Winne...>
 - [65] <https://techcrunch.com/2023/07/13/causaly-an-ai-platform-for-drug-discovery-and-biomedical-research-raises-60m/#:~:Artif...>
 - [66] <https://techcrunch.com/2023/07/13/causaly-an-ai-platform-for-drug-discovery-and-biomedical-research-raises-60m/#:~:Visio...>
 - [67] <https://elicite.com/blog/case-study-oxford-pharmagenesis#:~:We%27...>
 - [68] <https://www.causaly.com/blog/aiding-drug-repurposing-investigations-with-ai#:~:Drug%...>
-

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.