David Soden

MARKET ASSESSMENT | AUGUST 2026

The Bottleneck Is Below the Model

AI for scientific discovery: what the record supports, where the capital went, and why the constraint sits in the inputs

David Soden

Independent analyst and practitioner, enterprise automation and applied AI

davidsoden.com/market-assessment

REPORT DS-2026-D07D · 2026 · VERSION 1.0 · PUBLIC

About this report

Every substantive claim in this report carries an evidence classification and, where applicable, a numbered source. Facts are separated from vendor claims, third-party estimates, the author's assessments, and untested hypotheses. Forward-looking sections use scenarios with observable tripwires rather than point forecasts. Stated assumptions are listed before the analysis begins.

Market Assessment reports are published and commissioned at **davidsoden.com/market-assessment**. This document is analysis, not investment, legal, or procurement advice.

Scope and evidentiary standard

This assessment covers the use of AI systems to generate, test, and validate scientific claims, across four segments that are usually reported as one market: property and structure prediction models, literature and data agents, autonomous laboratories, and the automation of research engineering itself. It covers the capital committed to that market through early August 2026, the published record of what those systems have produced, and the condition of the inputs they consume.

Published numbers in this category need heavier discounting than usual, for three reasons. Vendors are the primary source for almost every capability figure, and most of those figures are self-graded: an accuracy rate assessed by the vendor's own evaluators, or a speed-up estimated by beta users comparing an agent run to work they did not actually do. Second, the category has already produced one fabricated headline result that was cited by a central bank and a congressional research service before it was withdrawn, so provenance is not a formality here. Third, the definition of the market itself is unstable, and two credible trackers counting AI-derived drug programs at the same moment differ by roughly a third because they define the term differently.

Every substantive claim below carries one of six labels. Where a widely repeated figure traces only to a weak chain, it is flagged and no conclusion rests on it.

Label	Definition	How to treat it
FACT	Verifiable in the cited primary source: a filing, press release, official documentation, or standards body announcement.	Reliable as stated. Check the source if the exact wording matters.
VENDOR CLAIM	Reported by a vendor about its own business or product. Self-measured	The vendor said it. Directionally useful, not a measurement. Definitions of the metric are

Label	Definition	How to treat it
	and not audited by any third party.	usually undisclosed.
THIRD-PARTY ESTIMATE	A research firm forecast or survey result. Methodology is proprietary and generally not reproducible.	Use for direction and order of magnitude only. Do not build a model on it.
ASSESSMENT	The author's reasoned judgment from the cited evidence. Falsifiable, and the reasoning is shown so it can be argued with.	This is analysis, not reporting. Disagree with it where your evidence differs.
HYPOTHESIS	A proposition offered for testing. Not currently supported by sufficient evidence to assert.	Treat as a research question, not a conclusion. Each one names what would confirm or refute it.
SCENARIO	A structured future state with a stated subjective probability. The probability is the author's judgment, not a calculated figure.	Use the leading indicators attached to each, not the probability. The indicators are observable; the probability is not.

Assumptions this analysis rests on

The first assumption is that the published clinical and materials record is a fair sample of what these systems produce. It probably is not. Failures are disclosed late and selectively, and one large company quietly discontinued an AI-associated candidate in 2025 after a readout showing no responses. If the unpublished record is materially worse than the published one, the segment-level assessments here are too generous, though the argument about inputs gets stronger rather than weaker.

The second is that US federal science policy continues on the trajectory set in late 2025 and July 2026, meaning a large centrally coordinated AI-for-science program running alongside constrained and slow-moving grant funding to universities. If appropriations or the courts reverse that, the two-tier scenario below loses most of its force and the timelines lengthen.

The third is that the value of a scientific claim depends on someone being able to check it. This sounds trivial and is doing a great deal of work. It is why verification infrastructure is treated here as part of the market rather than as overhead, and why a segment that produces unfalsifiable output at high speed is scored as having produced very little. A reader who believes generation alone is the product will disagree with most of the conclusions in this document.

The fourth is that model capability keeps improving at roughly the rate of the past three years. If it improves much faster, the fourth scenario becomes live and the input constraint stops binding. If it stalls, the constraint binds harder and nothing else here changes.

Executive summary

Something real has happened in the last twelve months. The most decorated research team of the current cycle left the largest AI lab in the world to automate the experimental loop, the US government committed more than five billion dollars to AI for science under a Manhattan Project framing, and a generative-AI-designed molecule showed a clinical efficacy signal in a peer-reviewed randomised trial. None of that is hype. All of it is upstream of the question that decides the outcome, which is whether the systems being built can get at trustworthy inputs and whether anyone can check what comes out.

FINDING 1 Four markets are being reported as one.

> **ASSESSMENT** Structure prediction, literature agents, autonomous labs, and research-engineering automation have different proof standards, different capital intensity, and time-to-falsifiable-proof ranging from months to a decade. Aggregate sizing of 'AI for science' is close to meaningless because it sums segments whose evidence is not comparable. [7] [33]

FINDING 2 The talent movement is real and its stated target is the loop, not the model.

> **FACT** Jeff Dean left Google after 27 years, with Sanjay Ghemawat, Oriol Vinyals and Quoc Le, to found Discovery Loop as a Delaware public benefit corporation on 5 August 2026. Google is a founding investor and cloud partner. The stated approach is automating the full experimental cycle, beginning with machine learning research itself. [1] [3] [4]

FINDING 3 The clinical record shows a real gain at the easy end and none yet at the hard end.

> **THIRD-PARTY ESTIMATE** Roughly 173 AI-originated programmes were in clinical development in early 2026. Phase I success rates for AI-native biotechs are reported at 80 to 90 percent against an industry baseline nearer half that. Phase II success is statistically indistinguishable from conventional development, and no AI-discovered drug has been approved. [33] [34] [46]

FINDING 4 The most cited materials result did not survive review, and the field's own correction is the useful signal.

> **FACT** A 2023 model proposed 2.2 million candidate structures and 380,000 predicted-stable ones, and an autonomous lab reported synthesising 41 novel materials in 17 days. Peer-reviewed critiques found scant evidence of compounds that were simultaneously new, credible and useful, and a separate re-examination concluded none of the autonomous syntheses was convincingly demonstrated. [14] [15] [16] [43]

FINDING 5 The binding constraint is the inputs, and the numbers on input quality are worse than the numbers on model capability.

> **FACT** Of 1,792 manuscripts whose authors stated data were available on request, 6.7 percent produced usable data. Over 80 percent of published papers report positive findings. Data leakage has been documented across 17 fields and at least 294 papers. Retraction Watch's managing editor told Congress the observed retraction rate is roughly a tenth of what it should be. [19] [22] [23] [25]

FINDING 6 Washington has already conceded the point in its own implementation plan.

> **FACT** The Genesis Mission executive order gave the Department of Energy 120 days to identify initial data and model assets, specifically including digitisation, standardisation, metadata and provenance tracking. The July 2026 OSTP report calls for high-value datasets and AI-enabled verification infrastructure alongside the models. The flagship national AI-for-science programme's first deliverable is a cataloguing problem. [9] [11] [12]

FINDING 7 Access to the corpus is being priced, per buyer, and that is a structural fact rather than a controversy.

> **FACT** Wiley reported 49 million dollars in AI licensing revenue for the fiscal year ended April 2026 and more than 110 million lifetime, on flat underlying revenue. Taylor and Francis licensed content to Microsoft for about 10 million in the first year. Authors were not asked. The literature is becoming an asset with differential access rather than a commons. [27] [28] [29]

FINDING 8 The most useful thing an outside reader can do is watch the tripwires, not the forecasts.

> **ASSESSMENT** This category's rate of change makes point forecasts worthless. The scenario probabilities below are the least reliable content in this document. The leading indicators are the most useful, because each is observable without private information.

Four markets, one name

ASSESSMENT The organising fact about AI for science is that the phrase covers four businesses that share almost nothing except a customer base. They differ in what they produce, in who is allowed to say whether it worked, and in how long it takes before anyone can tell. Treating them as one market is how a segment with a decade-long verification cycle ends up valued on the evidence of a segment with a three-month one.

FACT Structure and property prediction is the most established. AlphaFold2 arrived in 2021, the associated database now covers more than 200 million predicted protein structures, and its authors shared the 2024 Nobel Prize in Chemistry. Its impact is concentrated at specific points rather than spread across a pipeline: opening targets that had no experimental structure, speeding molecular replacement, and enabling proteome-scale target analysis. It does not predict the ligands, cofactors and metals that drug chemistry turns on. [39] [44]

ASSESSMENT Literature and data agents are the fastest-moving segment and the hardest to score. Output is a hypothesis or a synthesis, and the standard of proof is whatever the buyer applies. This is the segment where vendor self-grading is most prevalent and where the gap between an impressive demonstration and a checked result is widest.

FACT Autonomous laboratories are the most capital-intensive. Lila Sciences raised a 115 million dollar extension in October 2025 at a valuation above 1.3 billion, taking total capital to 550 million, and signed a 235,500 square foot lease in Cambridge, Massachusetts. Periodic Labs closed a 300 million dollar seed in September 2025. As of December 2025 a reporter visiting both found Lila's labs largely empty and Periodic beginning with manual synthesis guided by AI predictions. [5] [6] [7]

ASSESSMENT The fourth segment is the newest and the least discussed: automating research engineering itself, starting with machine learning research, on the theory that the loop

generalises. It is the only segment where the people building the tool are also the domain experts, which removes the translation problem that slows the other three. It is also the segment where a successful result compounds fastest, which is the case for taking it seriously and the case for worrying about it. [1] [2]

Segment	What it produces	Who can say it worked	Time to falsifiable proof
Structure and property prediction	Predicted structures, stability rankings, candidate compositions	Experimentalists, by synthesis and characterisation	Months to years; often never attempted at scale
Literature and data agents	Hypotheses, syntheses, cited analytical reports	Domain experts, then wet-lab follow-up	Weeks for plausibility, years for the claim itself
Autonomous laboratories	Synthesised samples and measured properties	Independent replication and characterisation	Weeks per run, years for a defensible discovery
Research-engineering automation	Working code, reproduced results, improved models	Benchmark scores and independent re-execution	Days to weeks; the shortest loop in the category

Where the capital went

FACT The macro picture is that AI absorbed most of venture capital and then a slice of it turned toward science. Crunchbase recorded roughly 300 billion dollars of global venture investment in the first quarter of 2026, about 80 percent of it into AI companies, a share that stood at 55 percent a year earlier. The OECD put AI at 61 percent of all global venture investment across 2025. [40] [41]

FACT Public money moved on a similar scale and faster. An executive order in November 2025 launched the Genesis Mission, a Department of Energy programme spanning 17 national laboratories with the stated goal of doubling the productivity and impact of American science and engineering within a decade. By July 2026 the White House put the commitment above five billion dollars across more than 15 federal agencies. [9] [10]

ASSESSMENT The composition of that spending matters more than the total. Private capital is concentrated in segments where proof is slow and expensive, which is where it can stay unfalsified longest. Federal capital is concentrated in platform and dataset work, which is where proof is boring and cumulative. Those are complements, and the second is currently the scarcer of the two. [8]

Venture or programme	Announced	Capital	What is not disclosed
Discovery Loop	Aug 2026	Seed co-led by Radical and Khosla; Alphabet a	Round size and valuation; round reported as not closed

Venture or programme	Announced	Capital	What is not disclosed
		founding investor	
Lila Sciences	Oct 2025	$550M total, $1.3B+ valuation	Customer names; utilisation of the new facility
Periodic Labs	Sep 2025	$300M seed led by a16z	Any experimental result to date
Reported OpenAI spin-out	Jul 2026	~$200M sought at ~$2B, per reporting	Company details; the founder disputed the reported figures
Genesis Mission	Nov 2025	$5B+ across 15+ agencies	Allocation between compute, datasets and verification

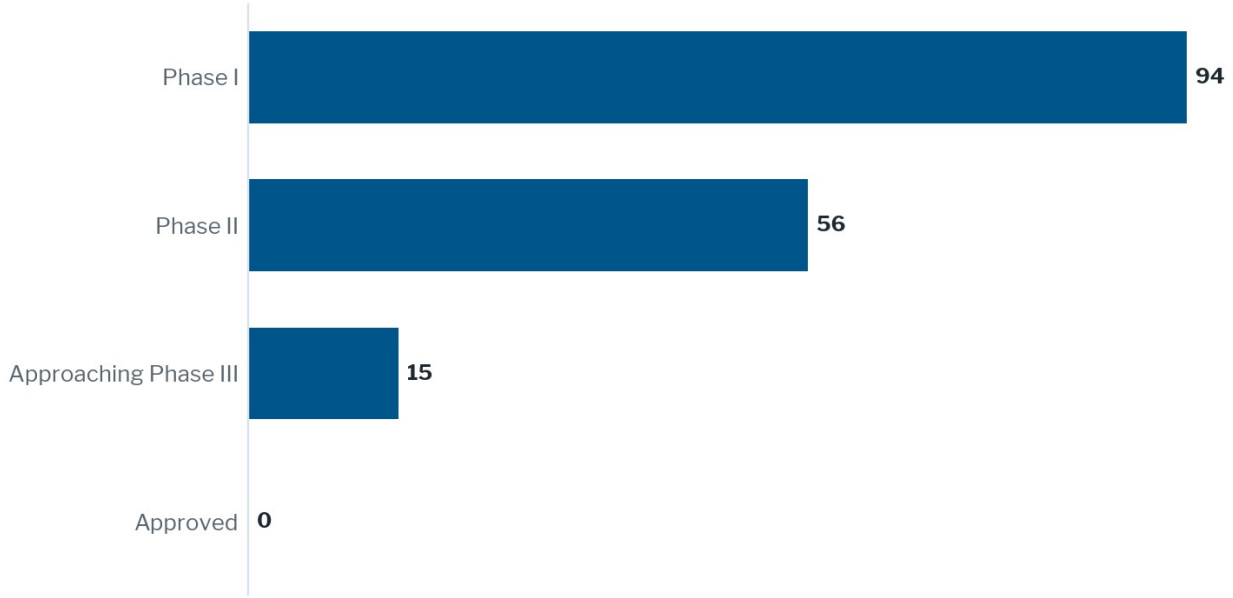
What the record actually supports

ASSESSMENT Read segment by segment, the published record is neither the breakthrough narrative nor the bubble narrative. It is a consistent pattern in which AI systems produce large gains on the tractable part of a problem and no measurable gain yet on the part that decides whether the work was worth doing.

Drug discovery: a real gain, at the wrong end of the pipeline

THIRD-PARTY ESTIMATE Industry trackers counted more than 173 AI-originated drug programmes in clinical development in early 2026, up from roughly two dozen in late 2023. A separate conference abstract counted 117 AI-enabled assets across 63 companies as of December 2025. The gap is definitional, and it is a fair warning about every aggregate number in this section. [33] [45]

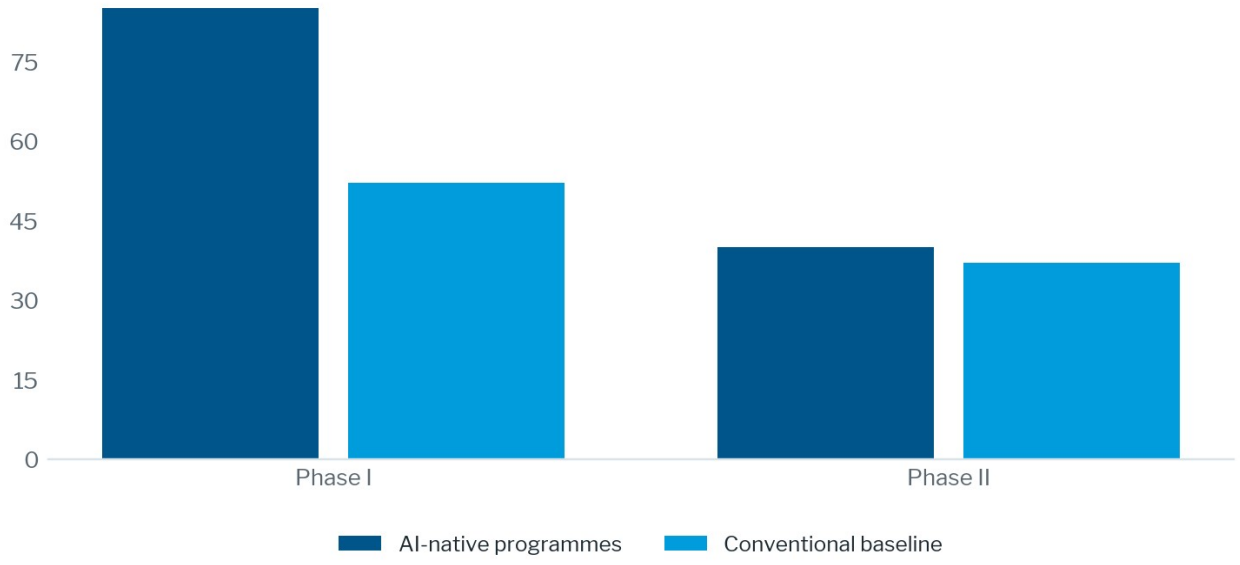
FIGURE 1 **AI-originated drug programmes by stage, early 2026**



THIRD-PARTY ESTIMATE Counts from commercial pipeline trackers, not a registry. 'AI-originated' has no agreed definition and competing trackers differ by roughly a third. The zero is the only figure here that is not disputed. [33] [46]

THIRD-PARTY ESTIMATE The success-rate pattern is the substantive finding. AI-native biotechs are reported at 80 to 90 percent Phase I success against an industry baseline of roughly 40 to 65 percent. At Phase II the reported figure is near 40 percent, against a conventional baseline of about 37 percent. Phase I asks whether a molecule is tolerated. Phase II asks whether the biology was right. [33] [46]

FIGURE 2 **Reported success rates, AI-native programmes against conventional baseline**



THIRD-PARTY ESTIMATE Midpoints of ranges reported by commercial analysts and trade press; methodologies are not public and the AI-native sample is small, young and survivorship-prone. The Phase I gap is the most likely of these figures to shrink as the cohort ages. [33] [46]

FACT One result deserves separating from the aggregates. A generative-AI-discovered TNIK inhibitor for idiopathic pulmonary fibrosis reported randomised Phase IIa results in Nature Medicine in 2025, with a forced vital capacity improvement at the higher dose against a decline on placebo over 12 weeks in 71 patients. That is the first peer-reviewed efficacy signal for a molecule designed this way, and it is a genuine milestone. It is also 71 patients and no approval. [34]

ASSESSMENT The honest reading is that AI has compressed the part of drug discovery that was already the cheapest and fastest, and has not yet touched the attrition that makes drugs expensive. Sixty years of declining research productivity in pharmaceuticals were not caused by a shortage of candidate molecules, and a technology that produces better candidate molecules should not be expected to reverse it on its own. [35]

Materials: the correction is the story

VENDOR CLAIM In late 2023 Google DeepMind announced that its graph-network system had predicted 2.2 million new crystal structures, of which 380,000 were described as the most stable and therefore promising candidates for synthesis. The company characterised this as multiplying the number of technologically viable known materials and as equivalent to about 800 years of knowledge. [14]

FACT A peer-reviewed perspective in Chemistry of Materials examined those claims and found scant evidence of compounds meeting what its authors called the trifecta of novelty, credibility and utility, describing many proposals as dopant variants, symmetry-broken realisations or chemically implausible compositions. A separate published re-examination of the paired autonomous-synthesis paper concluded that none of the claimed syntheses was convincingly demonstrated. [15] [16] [43]

ASSESSMENT The two lead authors of the original work disputed the critiques, and the dispute has not been fully resolved in the literature. That is not a reason to discount either side. It is the reason the segment is scored as unproven rather than as failed: the checking apparatus worked, slowly and in public, and it took eighteen months. The interesting number is not 380,000. It is the throughput of the mechanism that decides which of the 380,000 are real, which is measured in dozens per year. [15]

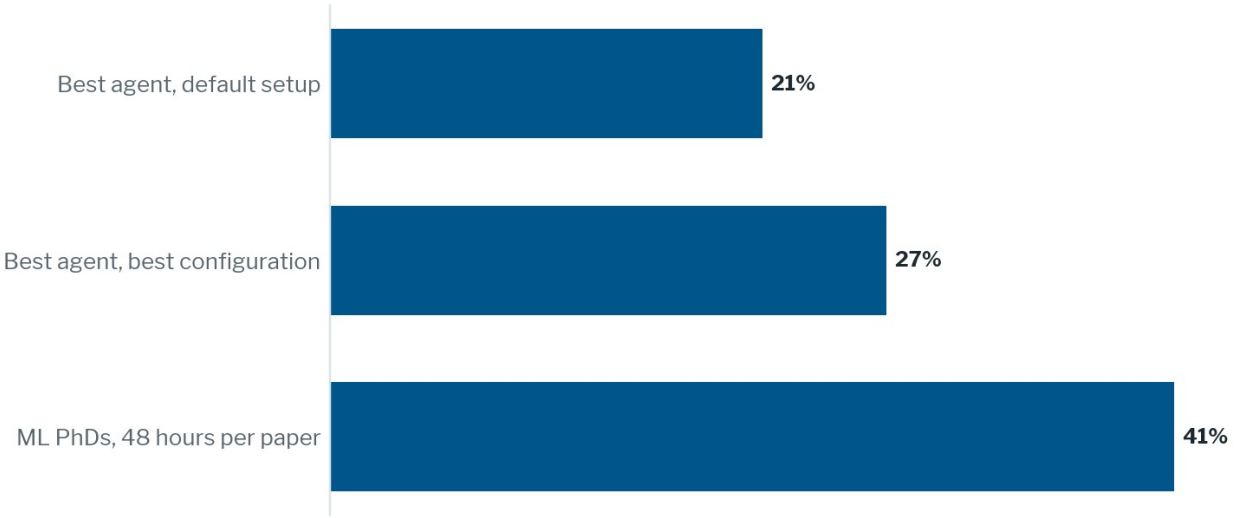
Research agents: fast, plausible, and hard to check

VENDOR CLAIM Edison Scientific, the commercial spin-out of the nonprofit FutureHouse, reports that a single run of its Kosmos system reads about 1,500 papers and writes 42,000 lines of analysis code while maintaining coherence across hundreds of agent trajectories. The company reports that 79.4 percent of the system's conclusions are accurate, and that beta users estimate one day of Kosmos does what would take them six months. [31]

ASSESSMENT Both headline figures are self-graded, and they are graded on different things. The accuracy rate comes from the company's own evaluators assessing whether claims are supported. The six-month comparison is beta users estimating the duration of work they did not perform. Neither is dishonest and neither is evidence in the sense this report uses the word. Critics have disputed the novelty of at least one of the seven announced discoveries, and the company's own leader has said some findings may be flawed and that the system should work alongside scientists. [31] [32]

FACT Where the same class of system has been measured against a rubric built with the original authors, the results are more sober. On a benchmark requiring agents to replicate 20 machine learning papers from scratch across 8,316 individually gradable tasks, the best agent evaluated in the 2025 study scored 21.0 percent, the best configuration reached 27 percent, and machine learning PhDs given 48 hours per paper scored about 41 percent. [30]

FIGURE 3 **Replication of machine learning papers, scored against author-built rubrics**



FACT Replication score on 20 ICML papers, 2025 evaluation. This is the easiest possible case for an agent: the source material is machine learning, the artefact is code, and nothing physical must be synthesised. Later scaffolds report much higher figures on different metrics that are not comparable to these. [30]

ASSESSMENT That benchmark is the most favourable ground these systems will ever be measured on. The papers are in the agents' own field, the deliverable is code, and no reagent has to be ordered. A 21 percent replication score on the easy case is the right anchor for expectations about the hard cases, and it is why the six-month claim above should be treated as a description of speed rather than of work completed.

The one result the field cannot cite

FACT In November 2024 a preprint reported that access to an AI materials-discovery tool at a large industrial laboratory produced 44 percent more materials, 39 percent more patents and 17 percent more product innovations, alongside a fall in job satisfaction. It was covered widely, praised by prominent economists, and cited by the European Central Bank and the Congressional Research Service. In May 2025 the author's university stated it had no confidence in the provenance, reliability or validity of the data, and asked for the preprint to be withdrawn. [17] [18]

ASSESSMENT Months after the withdrawal, a senior European research official appeared to cite the same result in a speech, and reporters could find no other plausible source for the figure. That is the failure mode this document is built to resist. A fabricated number about AI accelerating science travelled through preprint servers, prestige endorsement, press coverage and policy documents without once being checked, in a field whose entire premise is that machines will soon be doing the checking. [18]

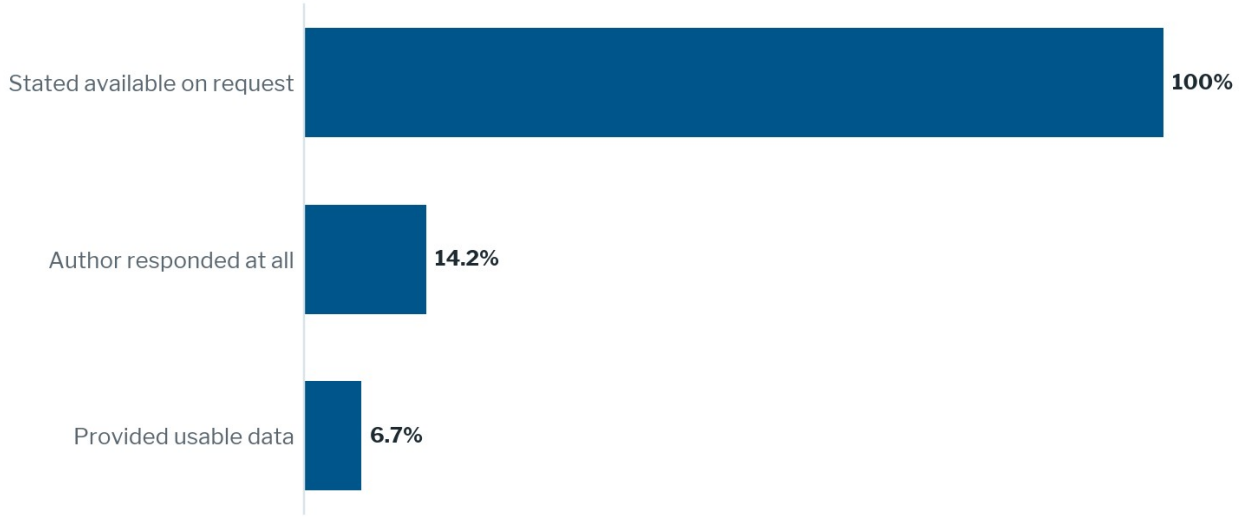
The constraint is in the inputs

ASSESSMENT The prevailing account of what is holding AI back in science is capability: better reasoning, more compute, more researchers. The evidence does not support it. What these systems consume is the recorded output of science, and by every measure that has been taken, that record is partial, biased toward positive findings, substantially unavailable in practice, and contaminated at a rate nobody can currently quantify. Point a better model at an unreliable corpus and the improvement is in the confidence of the answer, not its truth.

The data that does not arrive

FACT A 2022 study in the Journal of Clinical Epidemiology contacted the corresponding authors of 1,792 manuscripts whose data availability statements said data were available on request. Fourteen percent responded. Six point seven percent, 120 of 1,792, provided usable data. [19]

FIGURE 4 **What 'data available on request' produces in practice**



FACT Base of 1,792 biomedical and health manuscripts, 2022. A single study in one field, and the response rate may be higher where a requester is known to the author. An independent institutional review of a smaller 2020 sample found data obtainable from about a fifth of such publications. [19] [20]

FACT An independent review of one research institution's 2020 output examined 84 publications carrying the same phrase and obtained data from about 20 percent, with roughly 8 percent of corresponding authors unreachable because the email address had stopped working. Different field, different method, same direction. [20]

ASSESSMENT This is the single most consequential number in the assessment. Almost every proposal for accelerating science with AI assumes the recorded evidence base is retrievable. For the portion of it governed by availability-on-request, it is not, and the failure is silent: an agent querying the literature sees a data availability statement, not the ninety-three percent of the time that statement resolves to nothing.

The results that were never written down

FACT More than 80 percent of papers published in 2007 reported positive findings, exceeding 90 percent in some disciplines, and the share of positive conclusions rose 22 percent between 1990 and 2007. In a study with full accounting of a known population of 221 social science experiments, 10 of 48 null results were published against 56 of 91 strongly significant ones. [21] [22]

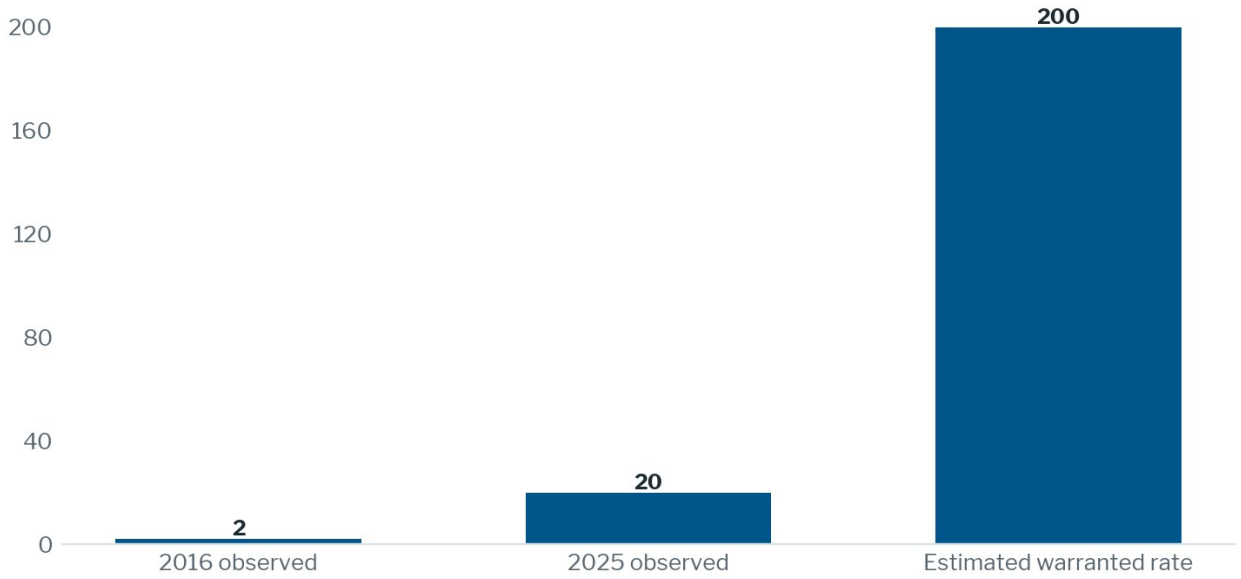
ASSESSMENT For a human reader, publication bias distorts effect sizes. For a system trained or retrieved on the corpus, it removes the negative half of the search space. Every experiment that failed, every synthesis route that did not work, every target that turned out to be undruggable is knowledge that exists and was never recorded in a form anything can read. One venture in this market has made generating that missing negative data an explicit part of its product, which is the correct read of where the scarcity is. [22]

The corpus is contaminated

FACT Researchers have assembled a database of more than 32,700 suspected paper-mill articles and warn it is an undercount, with the volume of fake papers doubling roughly every 18 months between 2016 and 2020. At one large open-access journal, 45 editors comprising a quarter of one percent of the editorial board handled 1.3 percent of 2024 articles but oversaw 30.2 percent of the journal's retractions. [24]

FACT In April 2026 testimony to a congressional hearing on scientific publishing, Retraction Watch's managing editor reported the retraction rate at roughly 0.2 percent for 2025, up from 0.02 percent in 2016, and stated the organisation's confidence that the rate should be about 2 percent, ten times what it is. [25]

FIGURE 5 **Retractions per 10,000 papers, observed against the rate practitioners believe is warranted**



THIRD-PARTY ESTIMATE Restated per 10,000 papers from rates of 0.02 percent, 0.2 percent and 2 percent, because the first figure rounds to zero on a percentage axis. The first two are observed rates reported in congressional testimony. The third is the testifying organisation's own judgment, not a measurement. [25]

FACT Contamination propagates. A cross-sectional study of 200,000 systematic reviews published between 2013 and 2024 found that 0.15 percent incorporated retracted paper-mill articles into their evidence synthesis, including 124 citations occurring after retraction and 13 more than 500 days after. Systematic reviews are what clinical guidance is built from. [26]

FACT The methodological problem is separate from the fraud problem and at least as large. A 2023 survey in Patterns documented data leakage across 17 fields, affecting at least 294 papers, and set out a taxonomy of eight distinct leakage types ranging from textbook errors to open research problems. Each of those fields discovered the issue independently, which suggests the true figure is a floor. [23]

Access is being priced

FACT Wiley reported 49 million dollars of AI licensing revenue for the fiscal year ended 30 April 2026, with lifetime AI revenue above 110 million, while total revenue stayed essentially flat at 1.67 billion and net income rose 163 percent. Taylor and Francis licensed content to Microsoft for about 10 million dollars in the first year with recurring payments structured through 2027. [27] [28]

FACT Authors were not consulted, and the publisher confirmed there was no opt-out. Large publishers with cleared back catalogues are capturing recurring revenue while society and small publishers report little comparable. [27] [29]

ASSESSMENT Read structurally rather than ethically, this establishes that the scientific corpus is a priced input with differential access. That is the condition under which a capability advantage becomes a durable position: not because one lab has a better model, but because one lab has a licence, a private instrument archive and a negative-results dataset that its competitors cannot buy. This is the mechanism behind the two-tier scenario below, and it is already operating.

The government's own first deliverable

FACT The Genesis Mission executive order directed the Department of Energy to identify, within 120 days, an initial set of data and model assets including digitisation, standardisation, metadata and provenance tracking. Within 60 days it was to identify at least 20 national science and technology challenges. [9] [11]

FACT The July 2026 OSTP report to the President, a 123-page document framed as the first rewrite of US science policy since 1945, calls for developing domain-specific scientific foundation models and high-value datasets, and for investment in AI-enabled verification infrastructure and autonomous laboratories. Agencies with three billion dollars or more in FY2026 research budget authority must submit implementation plans within 90 days. [12] [13] [42]

ASSESSMENT Strip the framing and the flagship federal AI-for-science programme has told the country that before it can do AI for science it needs to find, digitise, standardise and provenance-track its own data. That is not a criticism of the programme. It is the strongest available corroboration that the constraint is where this report says it is, published by the party with the most to gain from claiming otherwise.

FACT The same period saw the grant system that produces most of the corpus operating far below capacity. The National Science Foundation had awarded 613 grants by April 2026, about a fifth of the level at the same point in fiscal years 2021 through 2024, and the National Institutes of

Health about 10,000 awards against roughly 18,000 at comparable points. Congress had rejected the deep proposed cuts, and the shortfall is in disbursement rather than appropriation. [36] [37] [38]

Where value accrues

HYPOTHESIS The defensible position in this market is the discovery layer: the infrastructure that makes scientific evidence findable, resolvable, licensed, provenanced and checkable by machines. This proposition is testable. It is confirmed if, over the next 24 months, the ventures that gain durable commercial advantage are those with proprietary or licensed access to structured evidence rather than those with the strongest models. It is refuted if a lab with no privileged data access produces a validated discovery its competitors cannot match, or if a general model reaches high accuracy on the raw corpus without curation.

ASSESSMENT The economics favour this layer for an unglamorous reason: it is the only part of the stack where the work compounds and does not depreciate. A model is superseded in eighteen months. A resolved identifier, a verified provenance chain, a negative-result record and a licence are still worth what they were worth. Every model generation that arrives has to be pointed at something, and it will be pointed at whatever has been made addressable by then.

ASSESSMENT The layer is also, at present, unowned in a way the other three segments are not. Publishers hold the text and are monetising it, but hold neither the instrument data nor the negative results. National laboratories hold enormous federal datasets and are being told to catalogue them, on government timelines. Autonomous-lab ventures generate proprietary experimental data including negative results, but only about their own narrow programmes. Nobody currently holds the join. [27]

Why point forecasts fail here

ASSESSMENT Consider the twelve months to August 2026. A generative molecule posted a peer-reviewed efficacy signal. A national AI-for-science programme went from executive order to more than five billion dollars across fifteen agencies. The chief scientist of the world's largest AI research organisation left to start a company, its CEO stepped back from day-to-day leadership, and a Nobel laureate from the same lab moved to a competitor. A single quarter set an all-time venture record. Any forecast published twelve months ago that specified a number for August 2026 would have been wrong in both directions at once. [1] [10] [34] [40]

ASSESSMENT What follows is therefore four scenarios with subjective probabilities and, for each, the earliest sign that it is happening. The probabilities are the author's judgment and are the least reliable content in this document. The signs are the point.

Scenarios to 2029

SCENARIO A Instrumented science 30 percent

Verification and data infrastructure is built alongside the models. Federal dataset work lands, negative results become publishable and machine-readable, and provenance tracking becomes a condition of funding.

Consequence. Returns compound. AI-derived claims become checkable at close to the rate they are generated, and the Phase II and materials-validation bottlenecks start to move for the first time.

Earliest visible sign. Genesis Mission agency implementation plans, due around mid-October 2026, name dataset and provenance deliverables with owners and dates.

SCENARIO B Two-tier science 42 percent

Access concentrates. A small number of well-capitalised organisations combine licensed corpora, private instrument archives and proprietary negative-result datasets. Everyone else works from a public record that is free, partial and contaminated.

Consequence. The capability gap between funded and unfunded science widens faster than the gap between models. Reproducibility worsens because the best results are produced from inputs nobody else can obtain.

Earliest visible sign. A second major publisher reports AI licensing as a material earnings line, and at least one exclusive rather than non-exclusive corpus agreement is signed.

SCENARIO C Credibility shock 20 percent

A prominent AI-derived result fails publicly: a Phase III readout, a retracted high-profile paper, or a materials claim that does not survive independent synthesis. The failure is attributed to the method rather than to the specific team.

Consequence. Capital retrenches for four to eight quarters and regulatory scrutiny of AI-derived evidence tightens. Verification infrastructure gets funded defensively, a slower route to scenario A.

Earliest visible sign. Two or more AI-associated clinical programmes are discontinued in the same quarter with efficacy rather than strategy given as the reason.

SCENARIO D Capability jumps the gap 8 percent

Model capability improves fast enough that systems reliably distinguish reliable from unreliable evidence in the raw corpus, and generate their own verification, without curated inputs.

Consequence. The discovery layer stops being a constraint and becomes a commodity. Most of the analysis in this document expires.

Earliest visible sign. Replication scores on author-built rubrics for unmodified published papers exceed expert human performance, sustained across more than one independent benchmark.

Leading indicators

ASSESSMENT Each of the following is observable from public sources. None requires access to a private data room, and each moves the probability mass in a specified direction.

If this happens	It means	Moves toward	Where to watch
An AI-discovered molecule reports a positive Phase III	The efficacy gain is real, not a Phase I artefact	A, and away from C	Company releases; peer-reviewed clinical journals
Agency implementation plans name dataset owners and dates	Federal data work is being executed, not announced	A	Agency responses to the FY2028 R&D priorities memorandum
An exclusive corpus licence is signed	Access is becoming a moat rather than a revenue line	B	Publisher investor releases and earnings calls
A funder mandates machine-readable negative results	The missing half of the record starts being captured	A	Funder policy announcements; journal submission rules
A materials claim from an autonomous lab is independently synthesised	The validation bottleneck is loosening	A	Peer-reviewed materials chemistry literature
Two AI-associated clinical programmes stop for efficacy in one quarter	The method, not the team, is being questioned	C	Trial registries and discontinuation notices
An agent exceeds expert humans on replication rubrics	Curation may be becoming optional	D	Independent benchmark results, not vendor evaluations

Implications

For buyers of these systems

ASSESSMENT Ask what a vendor's system does when the evidence is missing rather than when it is present. Every demonstration is run on a well-covered question. The operational question is what the agent produces when the underlying data is behind a licence, sitting in a dead author's inbox, or never recorded because the experiment failed. If the answer is a confident synthesis, that is the product's most important limitation and it will not appear in the benchmark.

ASSESSMENT Treat any vendor accuracy figure as unusable unless you know who graded it and against what. The distinction that matters is between claims assessed by the vendor's own evaluators and claims checked against an external standard the vendor did not design. [31]

For investors

ASSESSMENT The four segments should not carry the same multiple. Time-to-falsifiable-proof is the underwriting variable, because it determines how long a position can remain unfalsified, and therefore how long capital can be raised on a narrative. The segment with the longest verification cycle is the one where a wrong thesis stays invisible longest, which is a risk rather than a feature.

ASSESSMENT The under-priced position is infrastructure that makes evidence addressable. It is unglamorous, it does not demonstrate well, and it is the only asset in the category that a model generation change does not depreciate.

For builders

ASSESSMENT The competitive question over the next two years is not whose model reasons better. It is who can reach evidence that others cannot, and prove where it came from. Access, resolution and provenance are product features, not compliance overhead, and the organisations treating them as overhead are the ones that will find themselves renting their moat from a publisher.

ASSESSMENT Publishing negative results is currently a cost with no return for the party that generates them. Whoever makes it cheap, standard and machine-readable captures a dataset that cannot be reconstructed later, because the experiments will not be re-run.

What this document cannot answer

How much AI-derived scientific work is failing quietly. The published record is a filtered sample, and no registry exists for AI-originated programmes comparable to clinical trial registration. Establishing the true denominator would need primary work with the organisations running these programmes.

What the actual rate of contamination in the scientific corpus is. The retraction rate is observed, the warranted rate is an informed guess, and the gap between them is where the answer lives. Nothing in the public record narrows it.

Whether the Phase I advantage in AI-designed molecules survives cohort maturation. The sample is small, young, and selected. Resolving this requires waiting, not analysis.

What proportion of instrument-generated laboratory data is recoverable at all. Estimates of unusable or unreachable research data circulate widely and trace to weak sources. A defensible figure would require an audit of a defined institutional population, which the author has not seen performed.

Whether the four segments here are the right cut. This taxonomy is the author's, built to separate claims by proof standard. A reader organising by customer or by capital intensity would draw different lines and reach different conclusions about sizing.

Half-life of this assessment

Decaying within six months: all funding figures, valuations and pipeline counts; the status of Discovery Loop's seed round; the disbursement position at NSF and NIH; agency implementation plans, which are due around mid-October 2026 and will either specify deliverables or will not.

Decaying within twelve months: the Phase I and Phase II success rates, as the AI-native cohort ages and survivorship effects wash out; the replication benchmark figures, which move with every scaffold release; the specific tally of AI-originated clinical programmes; the resolution, or continued non-resolution, of the materials synthesis dispute.

Decaying within twenty-four months: the scenario probabilities, all of which should be re-derived once the leading indicators have had time to fire; the claim that the discovery layer is unowned, which is the assessment most likely to be overtaken by a single acquisition.

Architectural, and unlikely to decay: that verification throughput rather than generation throughput determines the rate of real discovery; that publication bias removes the negative half of the search space from any system reading the literature; that priced, differential access to the corpus tends toward concentration. These follow from how science is organised rather than from anything about current models, and they will still hold if every capability figure in this document is superseded.

Limitations

The author is an independent analyst and practitioner. No organisation named in this document commissioned, funded, reviewed or was given advance sight of it, and the author holds no financial position in any of them. The author works on discovery and access infrastructure for agentic systems, which is the subject of the central assessment in this report. A reader should weight the finding that the discovery layer is the constraint accordingly, and should note that the finding is supported here primarily by evidence from parties with no such interest, including a federal implementation order and peer-reviewed meta-research.

The evidence base has specific weaknesses. Capability figures come overwhelmingly from vendors and are self-graded. Pipeline and success-rate figures come from commercial trackers whose methodologies are not published and whose definitions of AI-originated differ materially from one another. Coverage is skewed toward the United States, toward biomedicine and materials, and toward English-language sources, and the input-quality studies cited are concentrated in biomedical and health research, where the meta-research tradition is strongest. It does not follow that other fields are better.

One widely repeated figure was excluded. An estimate of the annual economic cost of research data not meeting findability and reusability standards circulates at over ten billion euros a year and traces, in the chain the author could follow, to secondary summaries of a 2018 study rather than to a verifiable primary source. It is noted here because its absence from the argument is deliberate, and because it is the kind of number that would have made the case for the discovery layer look stronger than the evidence warrants.

Sources

[1] Wiggers, K. "Jeff Dean and other top AI researchers are leaving Google to launch their own startup." TechCrunch, 5 August 2026. *Independent press* techcrunch.com

[2] Dean, J. Founding announcement post. X, 5 August 2026. *Vendor primary* x.com

[3] "Jeff Dean leaving Google after 27 years to co-found Discovery Loop." Quartz, 5 August 2026. *Independent press* qz.com

[4] "The startup idea that convinced a UW computer science legend to leave Google after 27 years." GeekWire, August 2026. *Independent press* geekwire.com

[5] Hu, K. "AI lab Lila Sciences tops $1.3 billion valuation with new Nvidia backing." Reuters, 14 October 2025. *Independent press* reuters.com

[6] "Lila Sciences: business breakdown and founding story." Contrary Research. *Third-party comparison (commercial interest possible)* contrary.com

[7] "AI materials discovery now needs to move into the real world." MIT Technology Review, 15 December 2025. *Independent press* technologyreview.com

[8] "OpenAI researcher in talks to launch AI drug discovery startup valued at $2B." TechCrunch, 14 July 2026. Figures disputed by the subject. *Independent press* techcrunch.com

[9] Executive Order, "Launching the Genesis Mission." The White House, 24 November 2025. *Government primary* whitehouse.gov

[10] "Trump Administration Announces More Than $5 Billion for the Genesis Mission." The White House, July 2026. *Government primary* whitehouse.gov

[11] Mosley, B. "White House announces Genesis Mission with the aim to accelerate scientific discovery through AI." Computing Research Association, 1 December 2025. *Practitioner aggregate* cra.org

[12] "OSTP Director Releases Landmark Report and Recommendations for Renewing American Scientific Discovery." The White House, 21 July 2026. *Government primary* whitehouse.gov

[13] Cobb, G. "OSTP releases Science: A New Golden Age, charting overhaul of federal R&D." FASEB, 30 July 2026. *Practitioner aggregate* faseb.org

[14] "Millions of new materials discovered with deep learning." Google DeepMind, November 2023. *Vendor primary* deepmind.google

[15] Quach, K. "Google DeepMind's AI-discovered materials claims disputed." The Register, 31 January 2024. *Independent press* theregister.com

[16] "Computed materials proposals depart from the structural memory of experimental discovery." arXiv preprint 2606.30967, 2026. Summarises the peer-reviewed critiques. *Second-hand (unverified preprint)* arxiv.org

[17] "MIT disavows doctoral student's paper on AI's productivity benefits." TechCrunch, 17 May 2025. *Independent press* techcrunch.com

[18] "Research commissioner appears to cite discredited study in AI speech." Science|Business, 9 October 2025. *Independent press* sciencebusiness.net

[19] Gabelica, M. et al. Journal of Clinical Epidemiology, 2022, reported in "Many researchers say they'll share data, but don't." Nature, June 2022. *Peer-reviewed primary via press* nature.com

[20] "Handling open research data within the Max Planck Society: looking closer at the year 2020." arXiv 2402.18182. *Practitioner aggregate* arxiv.org

[21] Franco, A., Malhotra, N. & Simonovits, G. "Publication bias in the social sciences: unlocking the file drawer." Science, 2014. *Peer-reviewed primary* science.org

[22] Nissen, S. et al. "Publication bias and the canonization of false facts." eLife, 2016. Reports the Fanelli 2012 positive-results figures. *Peer-reviewed primary* elifesciences.org

[23] Kapoor, S. & Narayanan, A. "Leakage and the reproducibility crisis in machine-learning-based science." Patterns, 2023. *Peer-reviewed primary* cell.com

[24] Richardson, R. et al. PNAS, 2025, reported in "Uncovering the fraudsters and their schemes responsible for polluting the scientific literature." Chemistry World. *Peer-reviewed primary via press* chemistryworld.com

[25] "Paper mills and the fight against scientific fraud." Inside Precision Medicine, 2 July 2026. Reports April 2026 congressional testimony. *Independent press* insideprecisionmedicine.com

[26] Tang, G. & Cai, H. "Citation contamination by paper mill articles in systematic reviews of the life sciences." JAMA Network Open, 2025. *Peer-reviewed primary* jamanetwork.com

[27] "Scholarly publisher AI licensing deals: inside the fiscal 2026 numbers." CASRAI, 2026. *Third-party analysis* casrai.org

[28] "Taylor & Francis AI deal sets worrying precedent." Inside Higher Ed, 29 July 2024. *Independent press* insidehighered.com

[29] "Authors Guild demands prior consent for AI use of academic and news content." The Authors Guild. *Advocacy primary* authorsguild.org

[30] Starace, G. et al. "PaperBench: evaluating AI's ability to replicate AI research." arXiv 2504.01848, OpenAI, 2025. *Vendor primary* arxiv.org

[31] "Kosmos: an AI Scientist for autonomous discovery." Edison Labs research announcement. *Vendor primary* edisonscientific.com

[32] Coverage roundup of the Kosmos launch and the novelty dispute. Ground News, November 2025. *Second-hand (aggregator)* ground.news

[33] "AI-discovered drugs in clinical trials 2026: full pipeline." IntuitionLabs, July 2026. *Third-party comparison (commercial interest possible)* intuitionlabs.ai

[34] Xu, Z., Ren, F., Wang, P. et al. "A generative AI-discovered TNIK inhibitor for idiopathic pulmonary fibrosis: a randomized phase 2a trial." Nature Medicine, 2025. *Peer-reviewed primary* nature.com

[35] Scannell, J., Blanckley, A., Boldon, H. & Warrington, B. "Diagnosing the decline in pharmaceutical R&D efficiency." Nature Reviews Drug Discovery, 2012. *Peer-reviewed primary* nature.com

[36] "A huge rupture in everything: US science faced major upheaval in 2025." Chemical & Engineering News, December 2025. *Independent press* cen.acs.org

[37] "NSF lags in grant awards and Trump again proposes deep cuts." APS News, 14 April 2026. *Practitioner aggregate* aps.org

[38] "NSF slashes research programs to support new tech initiative, insiders say." Science, 23 June 2026. *Independent press* science.org

[39] "AlphaFold in drug discovery: what protein structure prediction has and hasn't changed." Drug Discovery News, July 2026. *Independent press* drugdiscoverynews.com

[40] "Q1 2026 shatters venture funding records as AI boom pushes startup investment to $300B." Crunchbase News, 1 April 2026. *Third-party estimate* crunchbase.com

[41] OECD. "Venture capital investments in artificial intelligence through 2025." OECD Policy Briefs No. 50, 2026. *Analyst firm* oecd.org

[42] "OSTP's Golden Age report reshapes federal R&D strategy." Environment+Energy Leader, July 2026. *Independent press* environmentenergyleader.com

[43] Cheetham, A. K. & Seshadri, R. "Artificial intelligence driving materials discovery? Perspective on the article: scaling deep learning for materials discovery." Chemistry of Materials 36(8), 3490-3495, 2024. *Peer-reviewed primary* pubs.acs.org

[44] Varadi, M. et al. "The impact of AlphaFold Protein Structure Database on the fields of life sciences." PROTEOMICS, 2023. *Peer-reviewed primary* wiley.com

[45] "AI applications in the drug development pipeline." IntuitionLabs, February 2026. *Third-party comparison (commercial interest possible)* intuitionlabs.ai

[46] "173 AI drugs in trials, zero approved: what 2026 is missing." Science Reader, March 2026. *Second-hand (secondary analysis)* sciencereader.com

Rights and permitted use

This document is analysis, not investment, legal, or procurement advice. It was not commissioned, reviewed, or funded by any company named in it, and the author holds no position in any of them.

Market Assessment reports, commissioned research, and briefings: **davidsoden.com/market-assessment**