IntuitionLabs

INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

Novo Nordisk Anthropic Collaboration Pilot Blueprint

> Build practical AI for pharma and biotech with IntuitionLabs. We help life-science teams turn complex information and workflows into useful software, governed knowledge systems and AI tools.

Published by IntuitionLabs · 2026-09-19 · novo nordisk anthropic collaboration · novo nordisk claude · claude science · pharmaceutical r&d · generative ai pilot · enterprise ai governance · drug discovery ai · agentic software engineering

Original article: https://intuitionlabs.ai/articles/novo-nordisk-anthropic-rd-pilot

In this article

Executive Summary

The **Novo Nordisk Anthropic collaboration**, announced on **September 16, 2026**, is best read as a jointly developed research and development pilot, not as evidence that Claude has shortened a discovery program. Novo said it will test Claude Science on selected workflows, jointly identify high-impact scientific problems, develop targeted workflows for biological reasoning, and strengthen agentic software engineering (novonordisk.com) (novonordisk.com). The release also promises robust data governance and human oversight (novonordisk.com). It does not disclose contract value, term, exclusivity, workloads, model versions, deployment route, validation datasets, intellectual-property allocation, success thresholds, or results.

Claude Science is a beta scientific workbench rather than a single drug-discovery model. Anthropic describes a research environment with more than **60 curated skills and connectors**, auditable artifacts, local or high-performance computing access, and administrator enablement for Team and Enterprise users (anthropic.com) (anthropic.com). Those capabilities make workflow instrumentation possible, but they do not determine Novo's implementation. The earlier **NovoScribe** documentation system is a separate use case. Its vendor case studies report clinical-study-report creation falling from **12 weeks to 10 minutes**, with source lineage for author verification (mongodb.com) (mongodb.com). Those company-reported documentation metrics cannot be transferred to discovery productivity.

A defensible **90-day pilot** should therefore establish its own baseline, freeze test sets, log resource use, and require scientific review before any output influences an experiment or regulated record. For workloads that will support FDA regulatory decision-making, FDA's January 2025 draft, nonbinding guidance starts with the specific question and context of use, and calls for independent test data (fda.gov) (fda.gov). EMA similarly recommends lifecycle risk management and prospective testing for higher-impact uses (ema.europa.eu).

The decision rule is practical: scale only when the pilot beats an organization-supplied baseline on reproducibility, scientific utility, total resource consumption, and integration burden without exceeding agreed risk tolerances. Published benchmarks range from **17%** accuracy on independent bioinformatics tasks to **69.67%** on an independent clinical-agent benchmark, showing that performance depends heavily on task and environment (arxiv.org) (arxiv.org). No universal return-on-investment benchmark substitutes for a matched local comparison.

Introduction and Background

The announcement arrives at the intersection of two different enterprise-AI agendas. One is **scientific reasoning**, where researchers ask an agent to coordinate literature, code, data tools, and compute. The other is **agentic software engineering**, where an agent plans and changes code under human control. Novo explicitly named both tracks, which means the collaboration should not be reduced to a generic "AI drug discovery" headline (novonordisk.com).

Claude Science provides the product context. Anthropic says it brings previously fragmented tools into one research environment and names PubMed, Jupyter, R, and cluster terminals among the researcher interfaces it connects (anthropic.com) (anthropic.com). Its value proposition is orchestration and traceability across a workflow, not replacement of the specialist models, datasets, scientists, or laboratory systems that generate evidence.

That distinction matters to R&D digital leaders. A language model can draft a hypothesis, write analysis code, call a structure tool, or prepare a figure. It still needs an explicit **context of use**, meaning the role, scope, and intended use of its output. For an AI workload that will support FDA regulatory decision-making, FDA's January 2025 draft, nonbinding guidance defines that concept directly and ties it to how model output will inform a decision (fda.gov).

For an adjacent advisor such as IntuitionLabs, the relevant perspective is operating-model discipline, not a place in a vendor comparison. Its stated approach is to select one department, implement a small portfolio of governed workflows, and decide what to scale from observed evidence (intuitionlabs.ai). That posture fits this report's central recommendation: use the announcement as a design prompt, not a performance claim.

Key Changes

From general model access to jointly selected scientific problems

The announced operating model begins with selection. Novo scientists identify drug-discovery problems, while Novo and Anthropic jointly choose areas where their combined capabilities are expected to matter. The release then points to targeted solutions and workflows supporting biological reasoning (novonordisk.com). This is narrower and more useful than procuring seats first and looking for use cases later.

Table 1 separates the public commitments from decisions that a pilot charter still has to make.

Area	Announced on September 16, 2026	Not publicly specified, so the pilot must decide
Scientific scope	Joint selection of Novo-identified problems and targeted biological-reasoning workflows.	Disease area, target class, modality, datasets, specialist tools, and number of workflows.
Engineering scope	Use of frontier models for agentic software engineering.	Repositories, allowed actions, test coverage, merge authority, and production access.
Controls	Robust data governance, ethical standards, compliance standards, and human oversight.	Data classifications, retention, training use, audit storage, reviewer qualifications, and escalation path.

Area	Announced on September 16, 2026	Not publicly specified, so the pilot must decide
Technology	Anthropic models and Claude Science will be tested.	Model name and version, direct or cloud access, region, network pattern, connectors, and compute provider.
Economics and results	No figure or outcome appears in the release.	Contract value, seats, token budget, compute budget, baseline, success thresholds, term, and exit rights.

The table shows why procurement, architecture, and evidence teams cannot infer implementation details from the announcement. Even the platform's normal availability does not answer the Novo-specific questions. The release does not identify Novo's route or model.

A scientific-reasoning workstream and a software-engineering workstream

The two workstreams need different controls and measures.

- **Scientific reasoning:** evaluate literature synthesis, analysis planning, tool selection, hypothesis quality, traceability, and reproducibility.
- **Agentic software engineering:** evaluate issue resolution, test pass rates, code review findings, security checks, rollback success, and maintainability.
- **Shared foundation:** apply identity, least privilege, approved connectors, immutable logs, model/version capture, cost telemetry, and human approval.
- **Separate release authority:** let scientific reviewers accept evidence artifacts and engineering owners accept code changes. Neither agent should approve its own work.

This division avoids treating a coding benchmark as evidence of biological validity. Anthropic reported **49%** resolution on the 500-task SWE-bench Verified set using a minimal agent scaffold, but that result measures software issue resolution under its specified harness (anthropic.com). An independent machine-learning-agent study reported **37.5%** average success across its own task set, further illustrating harness dependence (arxiv.org). Neither establishes that an R&D hypothesis is novel, experimentally tractable, or correct.

Coexistence with earlier Novo AI programs

The Anthropic work should not be framed as exclusive or as replacing other relationships. Novo's April 2026 OpenAI announcement covered activities from drug discovery to commercial operations and described work with varied technology partners (novonordisk.com). An OpenAI update dated September 11 still described its Novo partnership and GPT-Rosalind (openai.com). A sensible portfolio can evaluate multiple models or routes against the same frozen tasks.

NovoScribe Versus the Claude Science R&D Scope

NovoScribe is useful evidence about organizational capability, but it answers a different question. The documentation project began in mid-2023 and used an architecture including Claude 3 and Titan on Amazon Bedrock, a private ChatGPT instance, LangChain, and MongoDB Atlas Vector Search (mongodb.com) (mongodb.com).

Table 2 shows why the earlier system's reported gains are not a discovery benchmark.

Dimension	NovoScribe documentation workflow	New Claude Science R&D collaboration
Primary output	Draft clinical-study-report content with source lineage for author checking.	Selected scientific workflows for biological reasoning, plus agentic software engineering.
Evidence base	Internal documents retrieved into a drafting workflow.	Undisclosed; it may involve literature, data, code, specialist tools, and compute.
Human role	Authors verify text against presented sources.	Human oversight is promised, but roles and approval points are unpublished.
Public metric	Vendor story reports **12 weeks to 10 minutes** for one report-creation process.	No public time, quality, cost, discovery, or reproducibility result as of September 19, 2026.
Transferability	Demonstrates document-orchestration experience.	Must establish scientific validity and workflow utility independently.

The AWS account adds a second company-reported description: work historically involving **40 to 50 people for up to 15 weeks** could be completed in minutes by a team of three (aws.amazon.com). The proper interpretation is narrow. It is evidence that a structured documentation process can be redesigned around retrieval, generation, and review. It is not evidence that an agent can select a valid target, improve an assay, or reduce the duration of a medicine-development program.

The same caution applies to other customer stories. An AWS and Pfizer case study estimated up to **16,000 scientist hours** saved annually in a search-and-extraction workflow (aws.amazon.com). An AWS and Genentech story projected automation of more than **43,000 hours** in biomarker validation (aws.amazon.com). A Benchling customer story separately claims up to **two weeks** saved on complex-data transformation (claude.com). All three are useful workflow hypotheses, but none is a matched, independent outcome study for Novo's new pilot.

Implementation Considerations and Process Changes

Scientific-problem selection funnel

Candidate selection should happen before architecture selection. For an AI workload that will support FDA regulatory decision-making, FDA's January 2025 draft, nonbinding credibility framework begins by describing the specific question, decision, or concern addressed by a model (fda.gov). A useful funnel scores each candidate on six dimensions:

- **Decision clarity:** name the downstream decision and the accountable human.
- **Baseline availability:** require current elapsed time, scientist time, compute, quality, and repeat rate.
- **Ground-truth access:** identify a frozen test set or a repeatable expert-review protocol.
- **Tool readiness:** confirm that required databases, code, compute, and licenses can be accessed safely.

- **Risk and reversibility:** prefer advisory outputs whose errors can be caught before an experiment or record changes.
- **Learning value:** select tasks that reveal platform limits, not only tasks likely to produce a favorable demo.

Each candidate can receive a **1 to 5** score per dimension. Weight decision clarity, ground truth, and reversibility twice; use the sum for prioritization, not as a universal pass mark. Reject candidates without an accountable reviewer or testable output even if their potential value appears high. NIST's mapping stage likewise places a go or no-go decision before design, development, or deployment (airc.nist.gov).

Data boundary, access, human review, and IP

The phrase "robust data governance" needs to become a data-flow diagram and control register. The minimum architecture decisions are:

- **Data classes:** public literature, licensed content, internal nonclinical data, clinical data, source code, personal data, and regulated records.
- **Permitted route:** direct commercial service, Amazon Bedrock, Google Cloud, or another approved environment, documented per workload.
- **Retention:** policy, contractual override, exception handling, log retention, and deletion verification.
- **Training use:** whether prompts, outputs, feedback, or artifacts may be used for model improvement.
- **Identity:** named users, service principals, privileged roles, connector scopes, and emergency revocation.
- **Review:** required scientific specialty, independence, evidence checklist, and final decision owner.
- **Intellectual property:** ownership of inputs, outputs, code changes, inventions, and third-party licensed material.

These cannot be inferred from the public collaboration. Anthropic's standard API policy says it automatically deletes inputs and outputs within **30 days**, subject to stated exceptions (privacy.claude.com). Its commercial documentation describes Anthropic as a processor acting for the customer (support.claude.com). Retention must be treated as a configured, contractual property, not as a universal constant (privacy.claude.com). Those general terms still require confirmation against the exact Novo contract and product surface.

Cloud routes can change the control plane. AWS states that Bedrock inputs and outputs are not shared with model providers or used to train base models (aws.amazon.com). Google's managed-service documentation states that customer data is not used to train or fine-tune models without permission or instruction (docs.cloud.google.com). Neither statement proves which route Novo selected.

For ownership, the governing agreement matters more than a generic assumption. Anthropic's published Amazon Bedrock terms say the customer owns outputs and that Anthropic disclaims rights received in customer content (www-cdn.anthropic.com). A pilot still needs counsel and procurement to map the applicable terms, employee invention rules, third-party dataset licenses, and jointly developed workflow assets.

A practical responsibility map

The accountable roles should be explicit before the first live dataset is connected. ICH E6(R3) states that agreements should clearly define roles, activities, and responsibilities (database.ich.org). A workable responsibility assignment is:

- **R&D problem owner, accountable:** defines the decision, scientific acceptance criteria, and stop conditions.
- **Scientific evaluator, responsible:** builds the gold set, performs blinded review, and adjudicates disagreements.
- **Platform owner, responsible:** implements identity, connectors, telemetry, version pinning, and rollback.
- **Data owner, accountable:** approves each source, transformation, retention rule, and downstream use.
- **Quality and governance, consulted:** sets validation depth, record requirements, risk tier, and change control.
- **Security and privacy, consulted:** approves threat model, regions, keys, logging, and personal-data controls.
- **Anthropic and integrators, responsible by contract:** document service behavior, support boundaries, and agreed deliverables.
- **Procurement and legal, accountable for terms:** confirm costs, IP, confidentiality, audit rights, and exit assistance.
- **End scientists, informed and consulted:** report usability, corrections, failure modes, and support burden.

This is a template, not a claim about the parties' unpublished allocation. The ICH expectation that systems be appropriately validated before use reinforces why the quality owner cannot be added only after a build is complete (database.ich.org).

The 90-Day Pilot Blueprint and Stage Gates

Days 0 to 30: charter, baseline, and sandbox

The first month should produce evidence before it produces scale.

- **Freeze two or three tasks:** one bounded reasoning task, one reproducible analysis task, and optionally one low-risk code task.
- **Record the baseline:** measure elapsed time, active scientist hours, compute, retries, accepted outputs, and review defects on the current process.
- **Create the test pack:** separate development examples from held-out evaluation data. For a workload that will support FDA regulatory decision-making, FDA's January 2025 draft, nonbinding guidance states that test data should be independent of development data (fda.gov).

- **Build the sandbox:** restrict connectors and write access, log every run, capture model and tool versions, and prohibit direct submission to regulated repositories.
- **Set stop conditions:** define unacceptable data exposure, unexplained result divergence, runaway cost, and inability to reconstruct an artifact. NIST frames safety as demonstrated performance with residual risk within organizational tolerance (airc.nist.gov).

Days 31 to 60: repeated evaluation and controlled integration

Run each task enough times to expose nondeterminism, tool failures, and review burden. Claude Science's artifact design can help because generated figures can include the exact code and environment that produced them (anthropic.com). The pilot team should still export and retain the artifacts under its own records policy.

- **Blind review:** hide whether an output came from the baseline or agent where practical.
- **Repeat runs:** rerun frozen inputs across sessions, model versions, and infrastructure conditions. EMA says higher-risk or higher-impact models should be prospectively tested with newly acquired data (ema.europa.eu).
- **Adversarial cases:** include missing data, conflicting sources, malformed files, tool timeouts, and ambiguous instructions.
- **Integration trial:** connect one approved read-only source before any write-enabled system.
- **Change log:** treat model, prompt, skill, connector, and dataset changes as versioned configuration.

NIST recommends a go or no-go decision before design or deployment and a pre-deployment demonstration that the system is valid and reliable (airc.nist.gov) (airc.nist.gov).

Days 61 to 90: decision package

The final month should compare results with the frozen baseline, document residual risk, and test operational ownership. ICH guidance supports validating relevant computerized systems before use and maintaining adequate backups (database.ich.org) (database.ich.org). The decision package should include:

- **Context-of-use statement** and excluded uses.
- **Data-flow diagram** with retention, regions, training-use policy, and connector permissions.
- **Evaluation report** with frozen cases, confidence intervals where applicable, reviewer agreement, and failures.
- **Reproducibility bundle** containing inputs, code, environment, tool versions, outputs, and reviewer decisions.
- **Economic worksheet** covering labor, tokens, compute, integration, validation, support, and repeated experiments.
- **Scale recommendation** for proceed, narrow and retest, hold, or stop.

EMA recommends lifecycle risk management across development, deployment, and performance monitoring (ema.europa.eu). A passed 90-day pilot is therefore an authorization for a defined next phase, not permanent approval of a changing agent.

Data Analysis and Evidence

Public evidence supports testing, but it does not support a universal productivity forecast. Benchmark results vary by task, scaffold, model, and compute. Anthropic reported **0.83** for Sonnet 4.5 on its Protocol QA evaluation versus a **0.79** human baseline, using its stated multiple-choice, 10-shot setup (anthropic.com). In a separate vendor run, Claude 3.7 Sonnet reached **84.8%** on GPQA with **256** samples and a maximum **64k-token** thinking budget (anthropic.com). These are informative capability signals, not a Novo business case. The vendor-reported Protocol QA figure should therefore be preserved with its task conditions whenever it is cited (anthropic.com).

Independent agent evaluations are more sobering and more relevant to pilot design. BixBench tested more than **50** bioinformatics scenarios and nearly **300** open-answer questions; its best reported frontier-model accuracy was **17%** (arxiv.org). MedAgentBench used **300** physician-written tasks, **100** patient profiles, and more than **700,000** data elements; Claude 3.5 Sonnet v2 reached **69.67%** overall (arxiv.org). MLAgentBench found Claude 3 Opus averaged **37.5%** across its machine-learning experimentation tasks (arxiv.org).

Table 3 turns that variability into a local scorecard. Every threshold should be set before results are revealed.

Gate	Measure and calculation	Evidence required for a pass
Feasibility	Completed frozen cases divided by attempted cases.	Tools, data, and compute function across the agreed task set; failures are classified.
Reproducibility	Independently reproduced frozen analyses divided by analyses attempted.	A different qualified expert obtains the same result from the same data, the NIST definition (csrc.nist.gov).
Scientific utility	Qualified hypotheses or accepted analyses per scientist hour.	Blinded experts score usefulness, correctness, novelty, and actionability against the baseline.
Safety and control	Critical-control breaches, unsupported claims, and unapproved actions per run.	Zero breaches of predefined stop conditions; residual risk is within approved tolerance.
Integration	Successful connector and compute jobs divided by attempts; median recovery time.	Read-only data flow, identity, logging, version capture, and rollback operate as designed.
Economics	Labor cost + model/API spend + allocated compute + integration + validation + repeat costs.	Total cost per accepted output improves against the organization-supplied baseline.
Scale readiness	Passed gates divided by required gates, with no critical gate averaged away.	Named owner, support model, budget, change control, and next-phase scope are approved.

The worksheet should record **scientist hours**, **GPU hours**, **model/API spend**, **experiment repeats**, **qualified hypotheses**, and **reproducibility rate** before and after. Anthropic's analytics interface exposes user and organization token usage and cost over time (platform.claude.com). That makes model cost an observable input rather than an estimate (platform.claude.com). Slurm exposes elapsed seconds and allocated resources,

permitting allocated GPU-hours to be calculated as elapsed hours multiplied by allocated GPUs (slurm.schedmd.com). The scheduler field should be captured for every instrumented compute job (slurm.schedmd.com). NVIDIA's Data Center GPU Manager adds utilization by measuring the fraction of time graphics or compute engines were active (docs.nvidia.com). Allocated GPU-hours and active-engine time should be reported separately (docs.nvidia.com).

Experiment repeats also need a definition. Technical replicates do not equal independent occasions. NIH examples explicitly distinguish points run in triplicate from compounds tested on three separate occasions (grants.nih.gov). A qualified hypothesis should be clear, testable, supported by citations or preliminary data, and paired with an evaluation plan, consistent with NIH grant-design guidance (niaaa.nih.gov).

The interpretation after the table is deliberately conservative. A pilot that saves scientist time but increases experiment repeats, compute consumption, or review defects may not create value. Conversely, a workflow that takes slightly longer but improves reproducibility and produces auditable artifacts may merit scale in a high-value context. The scorecard prevents one attractive number from masking the rest of the system.

Implications and Future Directions

The collaboration can be strategically meaningful without yet being outcome evidence. Its structure signals that the unit of adoption is a **workflow**, not merely a model endpoint. It also recognizes that biological reasoning and software engineering share infrastructure but require separate evaluation. As models, skills, connectors, and scientific tools change, the durable asset will be the evaluation system and governed information layer.

Three implications follow for pharma R&D leaders:

- **Portfolio governance should compare routes, not declare a winner early.** Direct services, cloud platforms, and other model providers can be evaluated on the same frozen cases. General cloud policies and public terms must be mapped to the actual contract.
- **Evidence should travel with the artifact.** Code, environment, tool calls, source references, model version, and reviewer decisions should be exportable and retained. NIST's generative-AI profile calls for model details including proposed use, value, assumptions, provenance, data quality, architecture, and evaluation data (nvlpubs.nist.gov). That documentation should remain linked to every evaluation release (nvlpubs.nist.gov).
- **Scale should be conditional.** NIST says safety should be demonstrated and residual risk kept within organizational tolerance (airc.nist.gov). EMA adds that human-centricity should guide development and deployment (ema.europa.eu).

IntuitionLabs' public measurement model emphasizes workflow penetration, time recovered, quality, risk signals, reliability, and support burden before expansion (intuitionlabs.ai). As an adjacent consultancy rather than a model vendor, that perspective belongs in operating guidance, not in a product-comparison row.

Frequently Asked Questions (FAQs)

What did Novo Nordisk and Anthropic actually announce?

They announced a collaboration to test Anthropic models and Claude Science on selected Novo R&D workflows, jointly address scientific problems, build targeted biological-reasoning workflows, and advance agentic software engineering. Novo also stated that human oversight and data governance will apply. No public result was announced.

Is Claude being used for drug discovery at Novo Nordisk?

The disclosed scope is an initial test for specific R&D workflows and biological reasoning. "Drug discovery" describes the program area, not a published outcome. The companies have not named a disease area, model version, workload, dataset, architecture, or validated improvement.

How can a pharmaceutical company use Claude in R&D?

A company can start with bounded, reviewable tasks such as literature triage, analysis planning, code generation, data transformation, or figure production. It should define context of use, use held-out evaluation cases, restrict tools and data, retain artifacts, and require qualified human approval. FDA's 2026 principles emphasize a risk-based approach with proportionate validation and oversight (fda.gov).

Does the NovoScribe result prove Claude can accelerate discovery?

No. NovoScribe is a documentation workflow. The reported reduction from 12 weeks to 10 minutes is a vendor case-study claim about clinical-study-report creation, not a controlled measure of hypothesis quality, experimental success, or medicine-development time (mongodb.com).

What should a Claude Science pilot measure?

At minimum: completion rate, accepted outputs, reviewer defects, scientist hours, token and API spend, allocated and utilized GPU hours, repeat experiments, qualified hypotheses, reproducibility, integration failures, recovery time, and support burden. NIST says AI systems should be tested before deployment and regularly during operation (airc.nist.gov).

Is there a standard return on investment for generative AI in pharma R&D?

No defensible universal benchmark was found. Public figures use different tasks, baselines, models, scaffolds, and accounting boundaries. A pilot should compare total cost per accepted output and scientific-quality measures against its own pre-registered baseline, with no critical safety or reproducibility gate averaged away.

Conclusion

The **Novo Nordisk Anthropic collaboration** is a useful operating-model case precisely because its public claims are limited. It commits the parties to jointly selected scientific problems, Claude Science workflow development, biological reasoning, agentic software engineering, data governance, and human oversight. It does not disclose the technical, economic, contractual, or validation details needed to judge performance.

Decision-makers should preserve that boundary. NovoScribe shows that Novo has experience redesigning a documentation workflow around multiple models, retrieval, lineage, and human checking. Other vendor case studies show that search and biomarker workflows may offer substantial time-saving hypotheses. Independent agent benchmarks, however, range widely across task designs. None establishes that the new collaboration has improved discovery timelines.

The credible next step is a controlled **90-day pilot** with a defined context of use, two or three frozen workflows, independent test data, restrictive access, artifact-level traceability, qualified review, and a before-and-after cost and quality worksheet. Scale should depend on reproducibility, scientific utility, integration reliability, total resource use, and residual risk, not on a press release or a transferred case-study metric. That approach turns a partnership announcement into a testable R&D capability decision while keeping claims aligned with the evidence available as of September 19, 2026.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://www.novonordisk.com/news-and-media/news-and-ir-materials/news-details.html?id=916768
2. https://intuitionlabs.ai/articles/pharma-ai-governance-gxp-compliance
3. https://www.anthropic.com/news/claude-science-ai-workbench
4. https://www.mongodb.com/solutions/customer-case-studies/novo-nordisk
5. https://intuitionlabs.ai/articles/genai-proof-of-concept-in-pharma
6. https://www.fda.gov/media/184830/download?_bhlid=4d54a7e74a14ce20df123608285aebfd01738f67
7. https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf
8. https://arxiv.org/abs/2503.00096
9. https://arxiv.org/abs/2501.14654
10. https://intuitionlabs.ai/
11. https://www.anthropic.com/engineering/swe-bench-sonnet
12. https://arxiv.org/abs/2310.03302
13. https://intuitionlabs.ai/articles/novo-nordisk-openai-partnership-pharma-ai-strategy
14. https://www.novonordisk.com/news-and-media/news-and-ir-materials/news-details.html?id=916532
15. https://openai.com/index/introducing-new-capabilities-to-gpt-rosalind/
16. https://intuitionlabs.ai/articles/claude-science-vs-gpt-rosalind-vs-isomorphic-labs
17. https://aws.amazon.com/th/video/watch/a9a9399c66c/
18. https://aws.amazon.com/solutions/case-studies/pfizer-PACT-case-study/
19. https://aws.amazon.com/solutions/case-studies/genentech-generativeai-case-study/
20. https://claude.com/customers/benchling?target=_blank
21. https://airc.nist.gov/airmf-resources/airmf/5-sec-core/
22. https://privacy.claude.com/en/articles/7996866-how-long-do-you-store-my-organization-s-data

23. https://support.claude.com/en/articles/9267385-does-anthropic-act-as-a-data-processor-or-controller
24. https://aws.amazon.com/bedrock/security-privacy-responsible-ai/
25. https://docs.cloud.google.com/gemini-enterprise-agent-platform/resources/zero-data-retention
26. https://www-cdn.anthropic.com/files/4zrzovbb/website/6b68a6508f0210c5fe08f0199caa05c4ee6fb4dc.pdf
27. https://database.ich.org/sites/default/files/ICH_E6%28R3%29_Step4_FinalGuideline_2025_0106.pdf
28. https://www.anthropic.com/news/claude-for-life-sciences
29. https://www.anthropic.com/news/visible-extended-thinking?cid=2554
30. https://csrc.nist.gov/glossary/term/reproducibility
31. https://platform.claude.com/docs/en/manage-claude/analytics-api
32. https://slurm.schedmd.com/sacct.html
33. https://docs.nvidia.com/datacenter/dcgm/latest/learn/modules/profiling.html
34. https://www.grants.nih.gov/policy-and-compliance/policy-topics/reproducibility/resources
35. https://www.niaaa.nih.gov/sites/default/files/publications/Training/Training_Quick_Guide_for_Grant_Applications-rev-2010.pdf
36. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf
37. https://www.fda.gov/about-fda/artificial-intelligence-drug-development/guiding-principles-good-ai-practice-drug-development

THE PUBLISHER

About IntuitionLabs

IntuitionLabs is an AI consulting, custom software development and data engineering firm serving pharmaceutical, biotechnology, medical-device and other life-science organizations. We work with clinical, regulatory, medical-affairs, commercial, quality and IT teams to connect technology decisions with the work people need to accomplish.

AI consulting and adoption

Our AI enablement services cover readiness assessments, use-case selection, governance and policies, team workshops, adoption measurement and ongoing advisory support. We help organizations structure the information layer behind AI: source material, context, permissions and maintained knowledge that make generated answers useful and reviewable. Private LLM inference and hosted AI options support teams evaluating how to operate AI with appropriate control over their data and infrastructure.

Software, data and life-science workflows

IntuitionLabs develops custom software for pharma and biotech, integrates enterprise systems, and builds data engineering and business intelligence solutions. Areas of focus include AI agents, regulatory research, medical writing, medical affairs, CMC information, competitive intelligence and clinical-document workflows. Our eTMF intelligence work includes cross-system reconciliation and inspection-readiness support.

Enterprise platforms and regulated delivery

We provide Veeva services, application support, managed services, integrations and custom applications, alongside enterprise content work involving platforms such as Egnyte. For regulated workflows, our services include GxP enablement, computer-system validation and software development addressing 21 CFR Part 11 requirements. The applicable controls, validation responsibilities and acceptance criteria are defined for each engagement.

Work with IntuitionLabs

Explore AI enablement, pharma and biotech software development, data engineering and BI, and Veeva services. Contact IntuitionLabs to discuss your workflow, information sources and implementation needs.

IntuitionLabs publishes educational research to help life-science teams make informed technology decisions. Coverage of a product or organization does not imply a client relationship, endorsement or partnership.

Website: https://intuitionlabs.ai