Capgemini invent



Closing the Loop: AI, Data and Discovery United in Preclinical Research

Executive Summary

Finding and testing new medicines has always been slow, expensive and full of trial and error. Today, three technologies are changing that, bioinformatics (making sense of biological data), computational biology (simulating how biology behaves), and AI, including AI “agents” that can carry out multi-step scientific work on their own. Used together, on a connected digital platform, they let scientists ask better questions, test more ideas safely in a computer before ever touching a lab bench, and catch problems years earlier than before.

This is not about replacing scientists. It is about giving them a well-organized, trustworthy digital “co-pilot” that removes the manual, repetitive work of finding and preparing data, so scientists can spend their time on judgment calls that matter. Capgemini brings together life sciences expertise, data engineering, AI, cloud and high-performance computing, and change management to help organizations move from small pilots to a dependable, everyday way of working.

What the Data Is Telling Us

This isn't just a technology trend - biopharma leaders themselves say the shift is under way. In a 2026 Capgemini Research Institute survey of 500 senior biopharma executives, the appetite for AI in R&D was striking:

82%
believe AI will fundamentally transform biopharma R&D

63%
expect most new medicines to originate from AI-driven platforms within a decade

74%
see significant potential in generative AI for drug discovery

43%
have already deployed AI for target identification, saving 28% of the time on average

Source: Capgemini Research Institute, "Smart bet, only option, or both? Biopharma R&D turns to AI," 2026

1. Why Preclinical Research Needs to Change

Bringing a medicine from an idea to a clinical trial takes years and involves six major steps. As a program moves through these steps, cost and risk both climb steeply. The earlier a bad idea is ruled out, and the earlier a good idea is strengthened with evidence, the more time and money is saved later.

Step	What Happens	The Big Question It Answers
Target identification	Find the gene or protein linked to the disease	What should we aim at?
Target validation	Confirm that acting on it actually helps, safely	Will hitting this target work?
Hit identification	Find molecules that interact with the target	Which molecules show promise?
Lead optimization	Improve the best molecules' strength and safety	How do we make it better?
Preclinical testing	Test efficacy and safety before humans	Is it safe and effective enough?
Clinical development	Test in people, in three phases	Does it work in humans?

Figure 1: *The preclinical-to-clinical journey - cost and uncertainty both rise at every step, so the goal is to build confidence as early as possible*

The most important shift is not just going faster. It is making better "go/no-go" decisions earlier - before a program reaches the expensive, high-stakes clinical stage - by combining genetic, biological, chemical and safety evidence into one clear picture.

2. The Capgemini Integrated Scientific Ecosystem

Too often, data platforms, biology software and AI tools are treated as separate projects with separate owners. Capgemini's view is that they work best as one connected system, where each layer feeds the next, so that every experiment makes the whole system a little smarter. This layered model is the central framework of this paper.

Layer	What It Does	Everyday Examples
Scientific data	Collects and organizes all research data so it can be trusted and reused	Genetic data, lab results, images, compound records
Bioinformatics	Turns raw biological data into clean, usable evidence	Quality checks, gene analysis, biomarker discovery
Computational biology	Simulates how molecules and biology behave, before testing in real life	Molecule docking, disease simulations, dose prediction
AI & agentic intelligence	Spots patterns, predicts outcomes, and runs multi-step scientific work with oversight	Predicting toxicity, designing molecules, orchestrating workflows
Connected lab	Links robots, instruments and lab software so results flow back automatically	Automated lab equipment, electronic lab notebooks
Trust & governance	Keeps everything secure, auditable, and scientifically sound	Data security, model checks, compliance records

Figure 2: *The integrated scientific ecosystem - six connected layers, working as one continuously learning system.*

Seen together, the six layers form the connected system described above: scientific data provides a trusted foundation; bioinformatics and computational biology turn that data into usable evidence and simulations; AI supports prediction and multi-step scientific work; the connected lab brings results back into the system; and trust and governance keep the environment secure, auditable and scientifically sound. The value comes from these layers working together rather than as separate projects.



3. Autonomous Scientific Workflows: AI Agents at Work

Most conversations about AI in R&D focus on individual tools; one model that predicts toxicity, another that reads scientific papers. The bigger opportunity is connecting these tools together with AI "agents" that can plan and carry out a multi-step piece of scientific work from start to finish, with a scientist checking and approving the important decisions.

In practice, this means an agent can gather, clean and organize data that a scientist would otherwise pull manually from multiple systems. Teams of agents can also connect the steps of an experiment by helping design it, handing it to lab robots, reading back the results and suggesting the next experiment. Throughout this process, every agent action remains logged and reviewable, so scientists stay in control and can understand what the system did and why.

A closely related idea is data that organizes itself. Rather than scientists manually tagging and structuring data for later use, the future goal is data that automatically arranges itself into a clean, well-labelled format, known as FAIR (Findable, Accessible, Interoperable, Reusable), the moment it is created, so both people and AI agents can use it right away, without weeks of manual clean-up.

This "agentic core" is what gives scientists a dependable digital partner: a system that can act with real autonomy on routine, well-defined tasks, while remaining fully transparent and accountable to the humans who make the final call.

4. From Molecules to Patients: A Multi-Scale View of Discovery in Translation Science

A common misconception is that computational simulations and validations happen in one straight line: first at the molecular level, then cellular, and then at the clinical level. In reality, these levels are studied side-by-side and constantly checked against each other, with real lab evidence at the centre of it all.

These different scales provide complementary views of the same biological challenge. Molecular modelling examines how a drug and its target interact, while cellular and pathway models explore the biological response that follows. Tissue and organism-level modelling considers broader biological effects, and translational modelling helps connect laboratory and animal findings to their likely relevance in humans.

Scale	What It Looks At	Why It Matters
Molecular	How a drug molecule and its target fit and interact	Confirms the molecule can physically do its job
Cellular / pathway	How a cell responds once the target is affected	Shows the biological chain reaction inside the cell
Tissue / organism	How an organ or whole body responds to exposure	Predicts real biological effects and side effects
Translational	How results are likely to carry over to humans	Bridges lab and animal findings to people

Figure 3: *A multi-scale modelling ecosystem - four views of research and discovery, cross-checked continuously, with lab evidence and uncertainty assessment always at the centre*

Keeping real experimental evidence and an honest account of uncertainty at the centre of this picture is what keeps computer predictions scientifically trustworthy, rather than treating them as guaranteed answers.

5. Where AI Adds the Most Value

AI is not one technology; it is a family of tools, each suited to a different scientific question. The table below connects each stage of research to the AI capability that helps most, and the real business value it delivers.

> *One of the biggest misconceptions with AI in drug discovery is that it is separate from laboratory science*
>
> - Thorsten Rall, Global Industry Lead for Life Sciences, Capgemini (AI Magazine, 2026)

Research Stage	AI Capability	Value Delivered
Target discovery	Pattern-finding across genetic and biological data	Finds promising, well-evidenced targets faster
Hit discovery	Virtual screening & active learning	Fewer wasted lab experiments
Lead optimization	AI-guided molecule design	Better-quality candidates, explored faster
Safety & translation	Toxicity and outcome prediction	Catches safety risks earlier and more cheaply
Decision-making	Evidence synthesis & agentic orchestration	Faster, better-documented go/no-go decisions

Figure 4: *AI use-case and value matrix across the discovery journey*

Across these stages, the role of AI changes with the scientific question. It can help identify stronger targets, reduce unnecessary laboratory experiments, guide molecule design, flag safety risks earlier, and bring evidence together for better-documented go/no-go decisions. The common thread is using the right AI capability at the point where it can add the most value to the research process.

6. From Point Solutions to an Integrated Preclinical Research Engine

Many organizations already use good individual tools - a genomics platform here, a modelling tool there, an AI model somewhere else. The real transformation happens when these are connected into one engine that carries a scientific question, with all its supporting evidence, cleanly from one stage to the next, instead of scientists re-explaining context every time work moves between teams or systems.

A connected research engine brings together target intelligence by combining genetics, biology, published research and safety evidence into one view. It links computer screening with laboratory testing for adaptive hit discovery, while lead optimization considers strength, safety and manufacturability together. Safety, dosing and biomarker evidence then support translational confidence, while clear evidence packages make go/no-go decisions traceable. Crucially, results, whether good or bad, feed back into the system so that the research engine continues to learn.

This is the shift from a collection of clever point solutions to a single, dependable research engine: less time spent hunting for information, more time spent making confident scientific calls.

7. A Modular Platform Blueprint

To make this real, organizations need a well-structured technology platform. Importantly, this should be built from modular, interoperable pieces that connect with existing tools: not one giant system that replaces everything at once, which is slower, riskier and harder to change later.

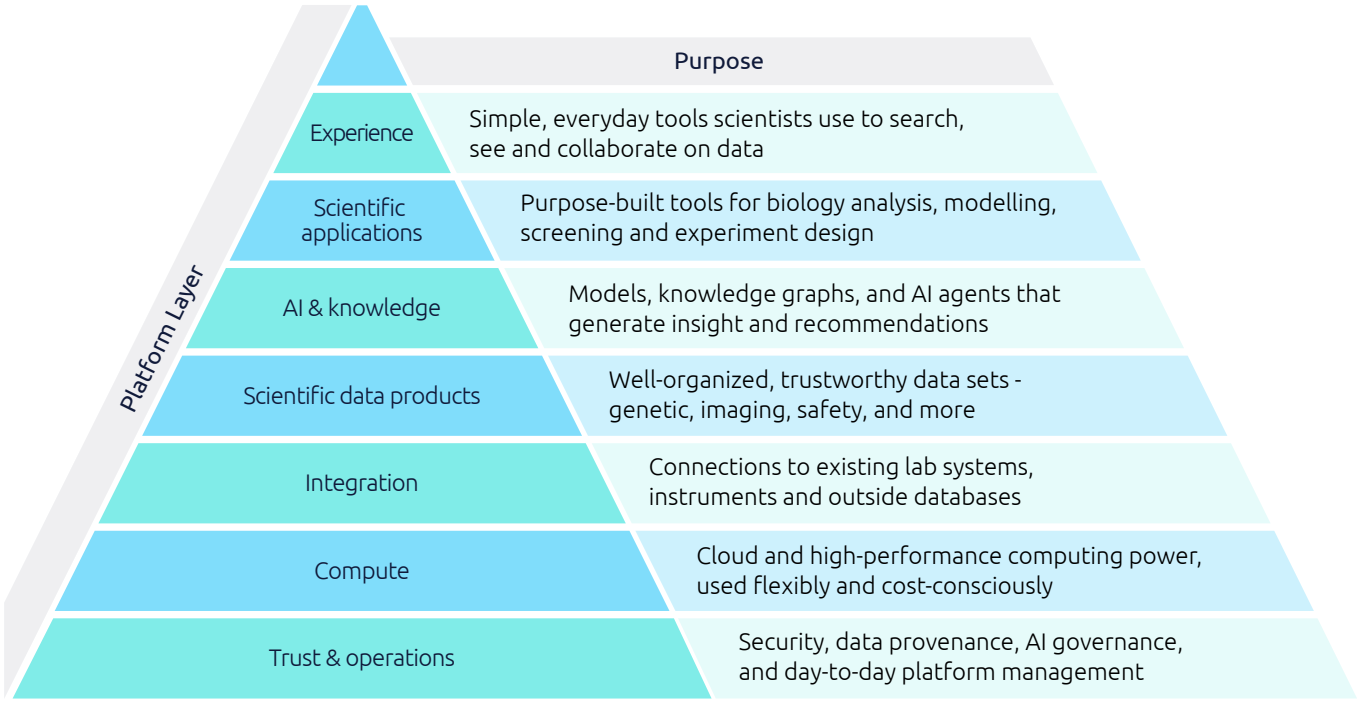


Figure 5: *A modular, interoperable platform blueprint - built to connect with what already exists*

Together, these layers create a platform that connects the scientists' everyday experience with the applications, AI capabilities and trusted scientific data underneath. Integration links this environment to existing laboratory systems, instruments and external databases, while cloud and high-performance computing provide the required compute capacity. Trust and operations underpin the platform through security, data provenance, AI governance and day-to-day management.

8. Operating Model: A Federated Center of Excellence

Technology alone will not create this future; people and ways of working do the heavy lifting. Instead of putting one central technology team in charge of every decision, we recommend a federated Scientific, Digital and AI Center of Excellence (CoE), so that standards are shared but decisions stay close to the science.

> *we recommend a federated Scientific, Digital and AI Center of Excellence (CoE)*

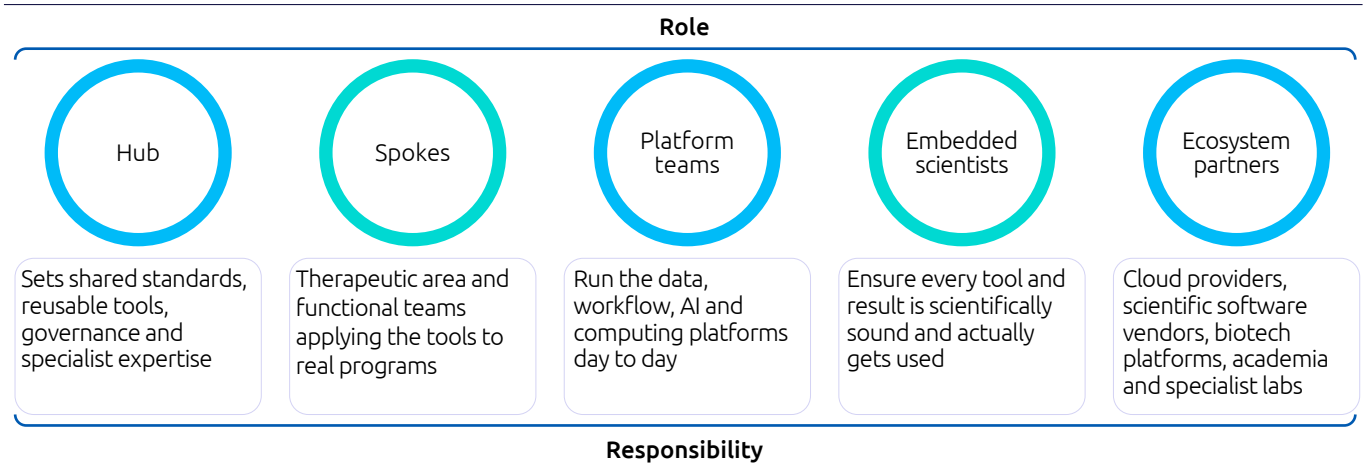


Figure 6: *A federated operating model - shared standards, locally accountable teams. Capgemini's role is to orchestrate this ecosystem and help implement it, not simply supply technology*

In this federated model, the hub provides shared standards, governance, reusable tools and specialist expertise, while the spokes apply them within therapeutic-area and functional teams. Platform teams run the underlying data, workflow, AI and computing environment, embedded scientists keep the work scientifically grounded, and ecosystem partners contribute the external technology and specialist capabilities required.



9. A Practical Path Forward

The path forward is deliberately progressive. It starts by understanding pain points and readiness, then prioritizing a short list of evidence-based use cases. One or two high-value workflows can then be tested with real scientists before reusable platform services are industrialized. Once those workflows are reliable, governed and repeatable, they can be extended across therapeutic areas and embedded into everyday use.

> *Technology alone will not create this future; people and ways of working do the heavy lifting.*

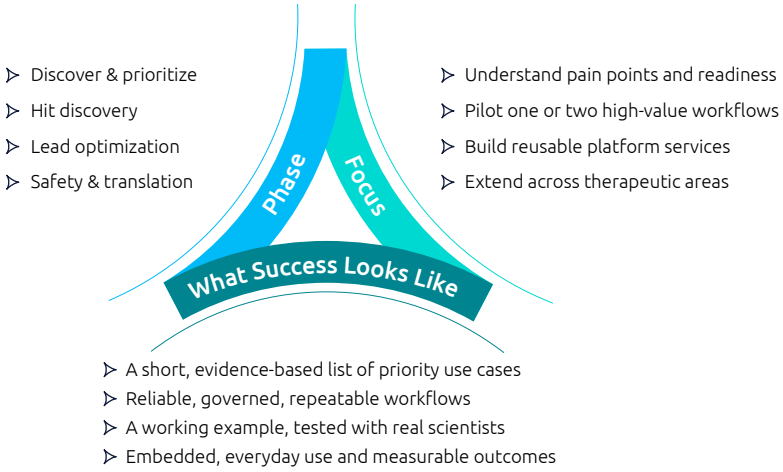


Figure 7: *A practical four-phase roadmap - start small, prove value, then scale with confidence*

10. Measuring What Matters

Success should not be judged only by technical measures like model accuracy or processing speed. A rounded view of value includes ***four categories***.

These measures provide a broader picture of whether the approach is delivering value. Scientific outcomes consider the quality and speed of research decisions, operational outcomes capture improvements in how work gets done, and adoption measures show whether scientists actively use and trust the tools. Economic outcomes complete the picture by considering avoided low-value experiments, computing spend and the value delivered by individual use cases.

Category	Example Measures
Scientific outcomes	Faster time from data to decision; results that hold up when repeated; better-quality candidates moving forward
Operational outcomes	Shorter analysis cycles; less manual data preparation; reusable workflows
Adoption outcomes	How many scientists actively use the tools; trust and satisfaction; less "shadow" spreadsheet work
Economic outcomes	Low-value experiments avoided; better use of computing spend; value delivered per use case

Figure 8: *A balanced KPI framework - scientific, operational, adoption and economic value together*

We deliberately avoid promising a reduction in clinical failure rates unless that link can be clearly measured and demonstrated for a specific program - over-claiming on outcomes this complex undermines trust.

Conclusion: From Digital Experiments to Translational Confidence

The future of preclinical research will not be decided by algorithms alone. It will be decided by whether organizations can connect trustworthy data, real biological understanding, computational intelligence and hands-on lab evidence into one dependable, learning system - with scientists firmly in charge of the decisions that matter most.

Capgemini is positioned to help Life Sciences organizations make this shift: defining the strategy, building the platforms, connecting the lab and computing environment, putting responsible AI controls in place, and scaling adoption from a single use case to genuine, enterprise-wide scientific value.

This point of view draws on Capgemini's life sciences delivery experience together with the Capgemini Research Institute's 2026 global survey of 500 senior biopharma executives.

How Capgemini Can Help

Capgemini can support this shift through Data-Driven R&D for Life Sciences, helping build FAIR and trustworthy scientific data foundations; Agentic AI for Enterprise through Capgemini RAISE, supporting governed agents for multi-step workflows; and Next-Gen Laboratories and Lab of the Future, connecting instruments, robotics and data. Digital Core and Industry Platform Software for Life Sciences provides the modular, cloud-native platform foundation, while Cambridge Consultants (Capgemini's deep-tech lab) brings capabilities in closed-loop AI and wet-lab experimentation.

The opportunity, ultimately, is to create a preclinical research environment in which data, biology, computational intelligence and laboratory evidence work together rather than in isolation. Connecting these capabilities within a trusted, continuously learning ecosystem can give scientists a stronger foundation for the decisions that matter most, while helping organizations move from individual digital initiatives towards a more integrated approach to scientific discovery.

Public References

Capgemini Research Institute: "Smart bet, only option, or both? Biopharma R&D turns to AI" (2026)
Capgemini Research Institute: Life Sciences Engineering and R&D Pulse 2026
Capgemini and Insilico Medicine: Partnering to Accelerate AI-Powered Drug Discovery
Data-driven R&D for life sciences - Capgemini
How AI is Transforming Pharma Drug Discovery (AI Magazine, featuring Capgemini)
Exploring the Impact of AI on Biopharma (AI Magazine, featuring Capgemini)
Agentic AI for Enterprise by Capgemini
Capgemini accelerates enterprise adoption of agentic AI for industries with NVIDIA
Next-Gen Laboratories Solutions - Capgemini
Industry Platform Software - Life Sciences - Capgemini
Integrating AI into the drug discovery phase of pharmaceutical R&D - Capgemini client story

Authors



Dr Sarika Vanarse
Senior Director
sarika.vanarse@capgemini.com



Chitra Thakur
Product Owner
chitra.thakur@capgemini.com



Sandhya Srinivasan
Senior Consultant
sandhya.a.srinivasan@capgemini.com

About Capgemini Invent

As the digital innovation, design and transformation brand of the Capgemini Group, at Capgemini Invent we enable CxOs to envision and shape what's next for businesses. Located in over 40 studios and more than 70 offices around the world, we are a 13,500 strong team of strategists, data scientists, product and experience designers, brand experts and technologists who develop new digital services, products, experiences and business models for sustainable growth.

Capgemini Invent is an integral part of Capgemini, an AI-powered global business and technology transformation partner, delivering tangible business value. We imagine the future of organizations and make it real with AI, technology and people. With our strong heritage of nearly 60 years, we are a responsible and diverse group of over 420,000 team members in more than 50 countries. We deliver end-to-end services and solutions with our deep industry expertise and strong partner ecosystem, leveraging our capabilities across strategy, technology, design, engineering and business operations. The Group reported 2025 global revenues of €22.5 billion.

www.capgemini.com

Make it real.

©2026 Capgemini. All rights reserved. This presentation contains information that may be privileged or confidential and is the property of the Capgemini Group.