



Artificial Intelligence in R&D: Trends in AI Investment, Business Models & Partnerships

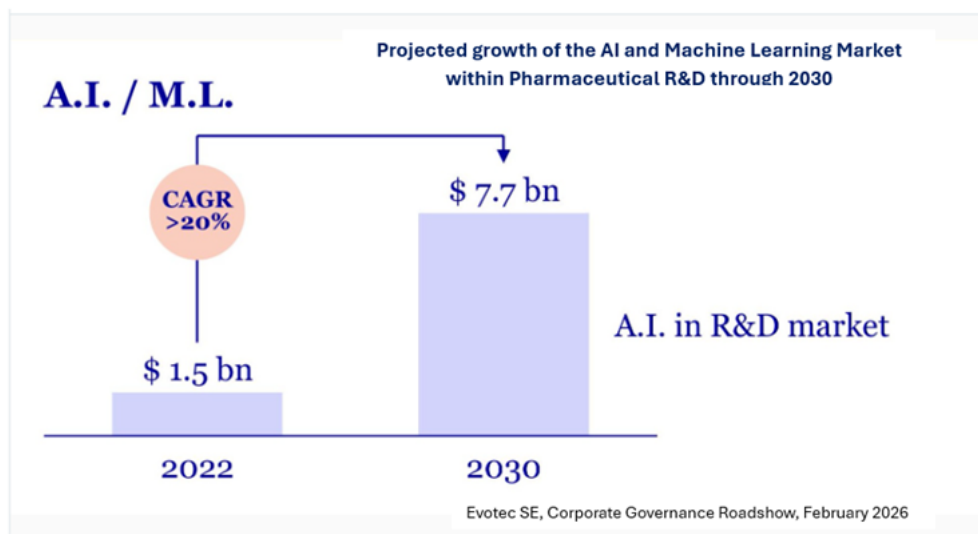
In Pullan's Pieces #223 March 2026

Kristine Dorward

kristine@pullanconsulting.com

44% of biopharma managers representing \$10B in aggregate R&D spending, who were queried in a recent survey, expect to allocate more than **10%** of their total R&D budget to AI in 2026, up from just 30% in 2025

By 2030, **60%** of these managers expect AI to command over **10%** of R&D budgets, with many anticipating that figure to exceed **20%** of total R&D expenditure in 2030



Questions & topics in this article:

- Inference AI Across R&D
- What are the main AI categories of investment by pharma/biotech and what do they involve?
- What is the MOAT for AI in R&D now?
- Which business models are predominating and why?
- What types of deals have been executed in AI R&D?
- What benefits are these deals likely to provide partners?

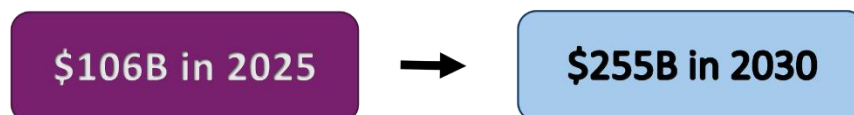
What Is Inference AI?

Inference is the act of running a trained AI model to generate predictions, answers and outputs; training AI models on the other hand involves building the model in the first place and requires occasional investment when models are updated

A model is trained once but queried continuously. As a result, inference can account for **80–90% of the lifetime AI production system cost because it runs continuously** and every prediction consumes compute and power

The AI compute market is currently undergoing a significant shift: in 2024, roughly 80% of AI spending went to training and 20% to inference - **but in 2026, that ratio is expected to reverse** and Gartner predicts that AI inference will represent **65% of all AI compute by 2029**

The AI Inference Market is predicted to reach \$255B by 2030, representing a 19.2% CAGR



Hyperscale Compute Infrastructure

The single largest category of absolute capital deployment now is compute infrastructure

Inference AI is highly embedded in the hyperscale compute infrastructure segment; Blackwell chips are architected to excel at inference workloads, not just training workloads

Roche's recent deployment of 3,500 NVIDIA Blackwell GPUs and [Eli Lilly's \\$1B NVIDIA supercomputer investment](#) both provide massive inference-capable infrastructure

AI-biopharma leaders who can establish access to secure, domain-specific inference at industrial scale will be winners and dominant players in the AI R&D race

Proprietary Data Generation – An Enabler of Inference AI

Proprietary data generation for AI training involves wet labs, multi-omics assays, biomarkers, organoid experiments, clinical and real-world evidence, longitudinal evidence etc.

Inference AI actively consumes proprietary data downstream, however capital allocation in this category is directed at data creation, not towards running models, per se

The ability to continuously generate, aggregate and integrate high-quality, scalable proprietary data is widely viewed as a critical differentiator for leading drug discovery platforms – why? – because foundational AI architectures are increasingly open-source and/or easily replicable with sufficient compute

AI-driven drug discovery companies are investing extensively in proprietary data generation in order to thrive in this market, such as fortifying capital in automated wet-labs capable of running millions of physical experiments weekly to feed predictive models, generating clinical and advanced multi-omics data

Agentic AI Platforms – Inference AI is deeply embedded

Agentic platform investment is, by its very nature, inference AI investment; end-to-end agentic pipeline requires continuous, high-throughput inference

An estimated ~ 73% of pharma organizations are deploying agentic AI, which in operational terms is akin to deploying inference at scale

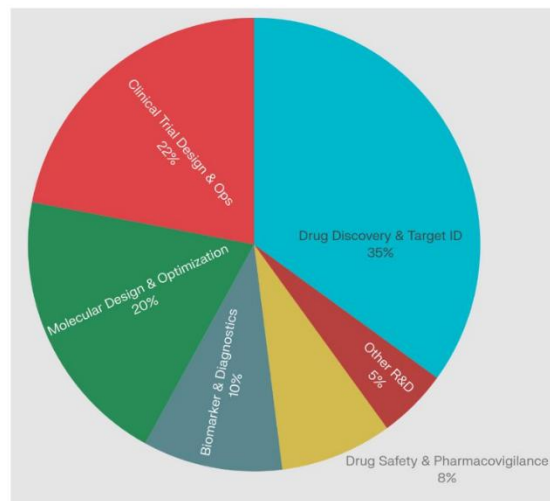
Every action taken by an autonomous agentic AI agent, such as generating a molecule, predicting a clinical outcome or routing a workflow, involves inference decision

AI Investment By R&D Application Area

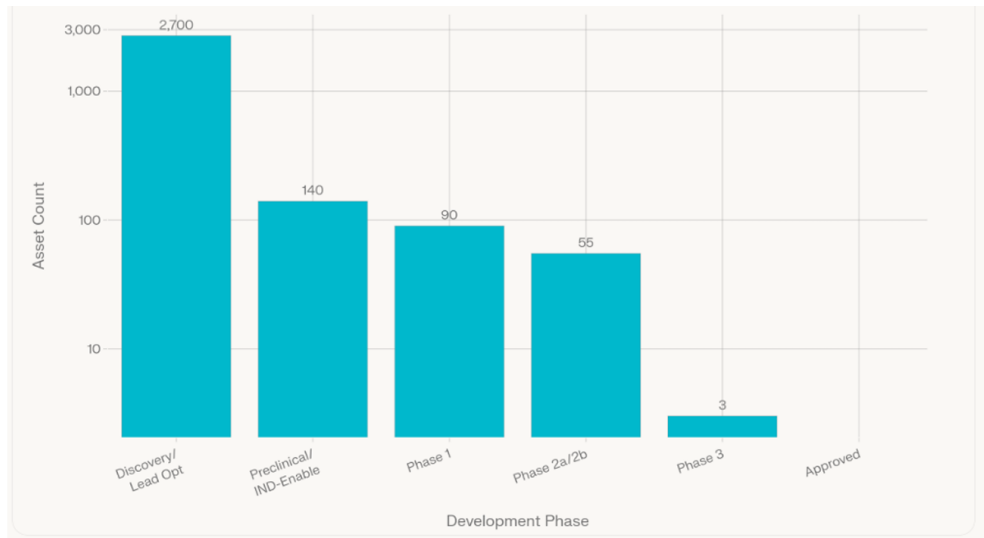
% of Total Pharma AI Spend, 2025



Source: CB Insights, ZS Associates



AI-Discovered Industry Pipeline By Phase



Source: [Intuition Labs](#)

March 2026

AI is delivering measurable, stage-specific efficiency improvements throughout the R&D value chain

- Target identification speed improvements of ~70% and early failure detection rates of ~70% are among the most impactful, as they eliminate weak candidates before costly clinical stage work is initiated
- Clinical trial cost savings of ~60% and preclinical cost reductions of ~40% will directly ameliorate an increasingly unmanageable, ~ \$2.6B average cost-per-new-drug problem
- The notoriously recognized “valley of death” transition from Phase 1 to Phase 2 that claims nearly half of all drug candidates is significantly improved with AI efficiency gains

- BCG data shows that companies investing in end-to-end AI infrastructure are seeing 25%+ reductions in early discovery timelines and 5-10% revenue uplift from AI-powered commercial personalization

Overall, estimated time-to-market is anticipated to be compressed by ~ 50–60% in the near future, a figure that has begun influence how licensing deals and milestone structures are priced in BD transactions

Shift Towards AI Platform Deals, Not Just Single Asset Deals

The average deal value for end-to-end AI platforms is expected to double from ~ \$85M in 2024 to a projected \$175M in 2026, reflecting pharma's trend from single-asset bets to broader infrastructure investments

Companies are [prioritizing platforms that deliver reusable capabilities](#) across multiple programs

The January 2026 deals between **Eli Lilly-Chai Discovery**, **GSK-Noetik**, and **Pfizer-Boltz** exemplify this shift: all three are multi-year, multi-target platform collaborations rather than single-asset licenses

Moat Is Derived from Owning the Dynamic, Integrated Data & Outcomes AI Generates

A recent article from ZS Insights highlights that strategically significant, independent decisions in large pharma have transitioned to become [learning systems and decision flywheels](#)

With agentic AI, signals are sensed and models continuously updated as conditions change, scenarios are evaluated dynamically, evidence-based decisions drive outcomes which are measured. Quality improves over time.

Agentic systems, whether they are screening molecules or routing enterprise workflows, execute decisions through inference calls


The competitive moat being built with agentic AI is both the agent orchestration layer as well as the ability to run secure, domain-specific, rapid inference at industrial scale


The inference engine is the strategic product, along with the compute underneath it


Sovereign AI inference over a closed feedback loop is a biopharma's invaluable moat – articulated by the question “who owns the learning loop?”


A feedback loop functions as a moat when it runs on sovereign, on-premises inference infrastructure

Four Strategic AI Archetypes Emerging

 **Infrastructure Kings (The Foundries)** -- Eli Lilly, Roche; betting on hyperscale sovereign compute as the platform others will eventually access

 **Defensive Incumbents (The Walled Gardens)** -- large pharma protecting existing pipeline value with agentic AI layered over decades of proprietary data

 **Embedded Operating Systems (The Workflow Players)** -- Ex: KALA BIO, Tamarind Bio: these companies are building sovereign inference and agentic middleware that sits inside other companies' R&D workflows

 **Native Disruptors (High Risk/High Reward)** -- Ex: Insilico Medicine, Recursion, Exscientia: for these companies their AI model fully underpins their drug pipeline, with external partnership stakes on AI performance

The winners across all four archetypes share one structural feature: they own the inference layer running on their own data, inside their own walls

Business Models of AI Firms in the Pharma/Biotech Sector

- SaaS / Subscription

Companies like Dotmatics and Benchling charge recurring access fees for cloud-based tools (molecular modeling, ELN, data management) with no IP transfer

- SaaS / Hybrid – Internal Pipeline

Recursion, Insilico Medicine and Exscientia run a dual model: advancing their own proprietary drug candidates, while also monetizing their AI platforms through co-development partnerships with pharma

- FFS / Compound Ownership Transfer

A pure service engagement where the client owns all IP outputs -- typically used by CROs/CDMOs with AI capabilities (e.g., WuXi AppTec AI services). Compensation for service is milestone- or deliverable-based.

- Platform Licensing / Perpetual

Biopharma clients receive a one-time or multi-year license for a platform or model (e.g., Schrödinger's perpetual computational licenses to large pharma)

Distinct from SaaS in that access is not per user or seat-based and often includes on-premises deployment.

- Co-Development / Risk-Sharing – AI Powered Research

AI company and pharma partner share R&D costs and downstream economics (milestones, royalties) Examples include Isomorphic Labs' arrangements with Eli Lilly and Novartis

Isomorphic does not sell platform access or advance its own proprietary branded pipeline in parallel

Isomorphic is purely an AI-powered research partner delivering molecules against partner targets for milestone + royalty compensation.

Asset-light by design, with no physical lab overhead, relying entirely on IP and data as its core assets.

- Data Partnership / Access License

Revenue from licensing proprietary training datasets, biomarker databases, or patient cohort data (e.g., Flatiron Health, HealthVerity)

Increasingly distinct from software licensing as data becomes the core asset

- Custom Exclusive / Bespoke AI Build

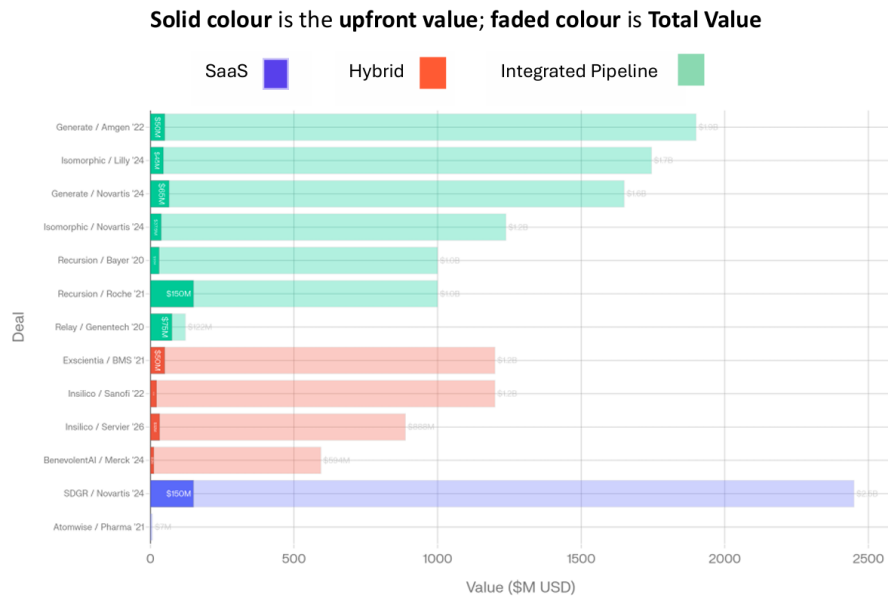
Fully bespoke AI system built exclusively for one pharma client, often with exclusivity provisions

Overlaps with Custom Research Agreements common in large pharma-AI vendor deals

Agentic AI & Inference AI Companies

Company	Agentic AI Platform	Inference Focus	Notable Activity
Variant Bio	<i>Inference</i> -- world's first agentic AI genomic drug discovery platform	Continuous inference over proprietary transcriptomic, proteomic & metabolomic data	Launched Jan 2026; partnership collaborations totaling \$200M+; VB-Predict and VB-Atlas inference agents
KALA BIO (NASDAQ: KALA)	<i>Reseagency.ai</i> -- agentic enterprise platform for biotech/pharma; described as "Palantir for biotech"	On-premises, data-sovereign inference; agents never upload proprietary molecular data to third-party cloud	First commercial AI agent shipped March 2026; exclusive worldwide license from Younet.ai; first client: Red Light Holland (psilocybin clinical dev)
Tamarind Bio	Molecular AI model coordination platform -- enables scientists to run agentic workflows with no ML expertise	Pure inference infrastructure play -- the core product is molecular AI inference-as-a-service	Raised \$13.6M Series A (Dimension Capital + YC) Feb 2026; 700% growth in one year; serves 8 of top 20 pharma including Bayer & Boehringer Ingelheim
Insilico Medicine	<i>Pharma.AI</i> --end-to-end autonomous pipeline from target ID to clinical prediction	Continuous generative inference across biology, chemistry & clinical data	Flagship full-pipeline agentic AI platform; strong presence in global pharma, including China market
Recursion Pharmaceuticals	Multi-agent biological discovery platform	High-throughput inference on massive phenomics image datasets	NVIDIA partnership; one of the most data-rich biotech AI platforms
Atomwise	<i>AtomNet</i> -- autonomous compound screening agents	Structure-based inference against up to 8 billion molecular candidates	Partners with 250+ research institutions
Exscientia (acquired by Recursion)	End-to-end autonomous drug design	Inference-driven molecular optimization and ADMET prediction	Pioneer in AI-designed drugs entering clinical trials
Model Medicines	Agentic AI for ultra-large virtual screening	Inference over 2M–8B compound libraries	Presented agentic + screening breakthroughs at conference March 2026

AI Drug Discovery Deals By Model (2020 – 2026)



SaaS / Subscription Deals

Company	Partner	Year	Upfront	Total Potential	Notes
Atomwise	Multiple Pharma Partners (per-target)	2021	~\$1M/target	~\$7M/target	Per-target SaaS model; pure software economics
Certara	Undisclosed Pharma	Ongoing	N/D	N/D	Pure subscription; no disclosed deal economics

Hybrid Co-Development Deals with Platform Licensing

Company	Partner	Year	Upfront	Total Potential	Notes
Schrödinger	Novartis	2024	\$150M	~\$2.45B + royalties	Bundled software subscription + multi-target collaboration; milestone/royalty structure is pipeline-tied, not platform-usage-tied
Schrödinger	Eli Lilly	2024	N/D	~\$125M near-term	Expanded deal includes co-development obligations beyond subscription renewal
Insilico Medicine	Sanofi	2022	\$21.5M	>\$1.2B + royalties	AI platform + pipeline co-development
Insilico Medicine	Servier	2026	\$32M	~\$888M	AI platform + pipeline co-development
BenevolentAI	Merck KGaA	2024	~\$12M	~\$594M	Knowledge graph platform + co-development
Exscientia (now owned by Recursion)	Bristol Myers Squibb	2021	\$50M	>\$1.2B + royalties	End-to-end AI design + co-development

Integrated “TechBio” Companies with Platform Deals

Company	Partner	Year	Upfront	Total Potential	Notes
Isomorphic Labs	Eli Lilly	2024	\$45M	~\$1.75B + royalties	AlphaFold-based structure prediction → drug design
Isomorphic Labs	Novartis	2024	\$37.5M	~\$1.24B + royalties	AlphaFold-based structure prediction → drug design
Recursion	Roche/Genentech	2021	\$150M	Several Billion	Phenomix platform; largest single upfront in category
Recursion	Bayer	2020	\$30M	>\$100M/prgm	Per-program deal structure
Generate Biomedicines	Amgen	2022	\$50M	~\$1.9B	Generative protein design; up to 10 programs
Generate Biomedicines	Novartis	2024	\$65M	~\$1.65B	Generative protein design
Relay Therapeutics	Genentech	2020	\$75M	\$121.8M <i>(actual amount paid prior to termination)</i>	Terminated in 2024 Target was an oral SM SHP2; Roche intended to pair with its KRAS G12C to overcome KRAS resistance

A Few Key Takeaways.....

- Inference AI has become the dominant force in pharma R&D computing
 - o The AI compute balance is fundamentally shifting, from ~ 80% training and ~ 20% inference in 2024, to an expected reversal in 2026

o Inference AI is now the operational engine of drug discovery, projected to represent ~ 65% of all AI compute by 2029

▪ Proprietary data provides true competitive differentiation, not the model itself

o Foundational AI architectures are increasingly open source and replicable, as a consequence, the real moat lies in the ability to continuously generate, aggregate, and integrate high-quality proprietary data

▪ Owning the "learning loop" creates the winners

o Sovereign, on-premises inference running over a closed feedback loop, where outputs continuously improve the model, is the most durable competitive advantage for biopharma

▪ Deal structures in AI are evolving from single-asset to platform-scale

▪ AI is delivering measurable R&D efficiencies that are beginning to reshape deal economics

▪ Four strategic archetypes are emerging, yet the common thread is sovereign inference

o The structural feature shared by dominant leaders in AI biopharma is the same: these companies own the inference layer running on their own data, inside their own walls.

o Companies that outsource this layer to third-party

clouds risk commoditizing their most strategically
valuable asset