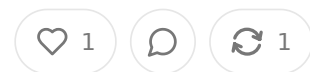


The AI Drug Discovery Capital Stack in 2026: Who Has Raised the Most, Why Their Technical Approaches Actually Differ, and Which Recent Industry and Academic Papers Are Worth a Real Read

APR 21, 2026



Share

Abstract

This essay maps the best capitalized AI drug discovery companies as of April 2026 and separates their platforms by what they actually do under the hood. Key points covered:

- Top of the funding stack: Xaira (\$1.3B disclosed), Eikon (~\$1.5B incl. 2026 IPO), Isomorphic Labs (\$600M external + ~\$3B in Lilly/Novartis deal value), Recursion (post-Exscientia), insitro (\$643M+), Iambic (\$300M+), Genesis Therapeutics (\$200M Series B), Chai (\$225M+), Insilico (\$500M+ private and a \$293M HKEX IPO Dec 2025)
- Four real technical lanes, not one: structure foundation models, generative chemistry, phenomics and perturbational biology, translational prediction
- Industry papers worth reading: AlphaFold 3, Chai-1, Boltz-1/2, insitro POSH, Iambic Enchant, RFdiffusion family
- Academic papers worth reading: PoseBench, AI-guided competitive docking, Target ID review in Nat Rev Drug Disc, Cell gene-expression de novo design, Science active learning transcriptomics
- Clinical reality check: Insilico is the only one on the list with a Phase 2 readout in humans for a fully AI-discovered, AI-designed asset (rentosertib in IPF)
- Where the moat is going: not any single layer, more the integration of proprietary perturbational data, generative models, automated wet labs, and clinical translation infrastructure

Table of contents

Why the funding question has two answers

The capital stack as of April 2026

The four technical lanes and why blurring them is lazy

Isomorphic vs Chai vs Boltz, the structure for

insitro and Recursion, the phenomics lane

Iambic and Genesis, the translational and ge

Insilico, the only one with a Phase 2 human r

The papers that actually matter

What the moat is becoming

Why the funding ques answers

There are two clean ways to answer who has they give different rankings, so anyone who sell something. Method one is largest single largest disclosed total capital raised over the splashy launches and IPOs. Method two rewards persistence, quiet follow-ons, and being old enough to have stacked rounds. The most useful version of the answer is to keep them separate, then layer on a third lens, which is the value of the pharma deal book, since for some of these companies that money is functionally part of the runway even if it is technically contingent on milestones.

The capital stack as of April 2026

By largest single financing event, the top of the heap is still Xaira launching in April 2024 with more than \$1B of committed capital, Isomorphic raising \$600M in its first external round in March 2025, Exscientia closing \$510.4M of aggregate IPO financing back in 2021 (now folded into Recursion), Recursion at \$436.4M in its 2021 IPO with substantial follow-ons since, and Eikon at \$350.7M in a Series D in February 2025 followed by a \$381.2M IPO in February 2026.

By largest disclosed total, the picture shifts. Eikon now sits around \$1.5B if you add the post-IPO capital to the \$1.1B+ it had said it raised privately by 2025. Xaira has quietly grown to roughly \$1.3B in total disclosed funding, not the \$1B headline number that still gets cited everywhere. Isomorphic is \$600M of external financing



Discover more from Thoughts on Healthcare Markets and Technology

We cover broad topics at the intersection of health tech, health policy, health entrepreneurship and health investing.

Over 3,000 subscribers

Subscribe

By subscribing, you agree Substack's [Terms of Use](#), and acknowledge its [Information Collection Notice](#) and [Privacy Policy](#).

Already have an account? [Sign in](#)

plus whatever Alphabet has been pouring in internally for years before the external round, plus a deal book with Lilly and Novartis worth nearly \$3B in upfront and milestone value, with Novartis having expanded the partnership in February 2025 to add up to three more programs. insitro is at least \$643M from its \$100M+ Series A, \$143M Series B, and \$400M Series C. Iambic is at \$300M+ across a \$53M Series A, a \$150M+ Series B, and an oversubscribed \$100M+ raise in late 2025. Genesis Therapeutics, which often gets left off these lists for some reason, is at roughly \$280M total after its \$200M Series B co-led by Andreessen Horowitz. Chai is at \$225M+ after its December 2025 Series B. And Insilico, the only one in this group that has actually tapped public equity markets, raised about \$293M (\$2.277B HKD) in its December 30, 2025 Hong Kong IPO on top of more than \$500M raised privately, which puts it somewhere around \$800M total disclosed.

So the right shortlist of best-capitalized AI-native or AI-centric drug discovery players to watch in 2026, in roughly the right order, is Eikon, Xaira, Isomorphic, Recursion (with the absorbed Exscientia capital), insitro, Insilico, Iambic, Genesis, and Chai. That ordering changes a bit depending on whether you count the Isomorphic deal book as capital, whether you count public-market dollars at par with private, and whether you treat Recursion plus Exscientia as one entity or two. None of those framing choices is wrong. They are just different.

The four technical lanes and why blurring them is lazy

The single most underrated point about AI drug discovery in 2026 is that it is not one category anymore. It is at least four real technical lanes, and the model classes, the data moats, the validation strategies, and the failure modes are pretty different across them.

Lane one is structure prediction and biomolecular foundation models. This is AlphaFold 3, Chai-1, Boltz-1, Boltz-2. The bet is that if you can model proteins, nucleic acids, ligands, ions, and modified residues all at once, much more of medicinal chemistry can move in silico. Lane two is generative chemistry, which is about proposing actual molecules with desired properties, often through diffusion models, language models for molecules, or graph neural nets. Lane three is phenomics and perturbational biology, which is about generating massive amounts of cellular data and learning representations over biological state, rather than over atomic geometry. Lane four is translational prediction, which is the layer trying to predict whether a preclinical candidate will actually survive ADME, tox, PK, and human trials. Most

slide decks blur these. They should not be blurred. A company optimized for lane one will not necessarily fix the problems in lane four, and vice versa.

Isomorphic vs Chai vs Boltz, the structure foundation lane

Isomorphic Labs is the most structure-centric of the top tier. Its bet is essentially that if you can model biomolecular complexes well enough, structure-based drug design becomes radically more productive. AlphaFold 3 is the technical anchor, and its core contribution is a diffusion-based architecture for joint complex prediction, which is a very different philosophy from the older AlphaFold 2 design and a totally different philosophy from classic QSAR or phenotypic screening. The commercial proof is the Lilly and Novartis deals signed in early 2024, which together had roughly \$3B in upfront and milestone value, plus the Novartis expansion in February 2025 adding more programs. Then there is the not-so-small fact that Demis Hassabis and John Jumper picked up the 2024 Nobel in Chemistry, which is the kind of institutional validation no other company on this list has.

Chai is closest to Isomorphic in spirit but very different in posture. Chai-1 is also a multimodal foundation model for biomolecular structure prediction, but it is openly accessible for non-commercial use, can run in single-sequence mode without multiple sequence alignments while preserving most of its performance, and can optionally be prompted with experimental restraints. The most underdiscussed differentiator between these two is not raw model quality but licensing and posture. Isomorphic gates AlphaFold 3 for commercial use pretty tightly, which has pushed a meaningful chunk of industry computational chemistry and biotech R&D toward Chai, Boltz, and the open lane. That is a moat question, not a quality question, and it is the kind of thing that will matter more than benchmark scores over the next few years.

Boltz, out of MIT, deserves more attention than it usually gets. Boltz-1 and the more recent Boltz-2 are fully open weights and training code, which neither AlphaFold 3 nor Chai-1 are. For academic groups, smaller biotechs, and any team that needs to fine-tune on its own proprietary data without sending that data into someone else's API, Boltz is increasingly the default. Boltz-2 in particular has made meaningful gains on affinity prediction, which has historically been the place where structure foundation models embarrass themselves. A useful frame is that Isomorphic owns the lab, Chai owns the playground, and Boltz owns the open commons. All three matter.

The bigger meta point about the structure lane is that being able to predict structures is now table stakes. The field has slowly figured out that protein-ligand geometry alone is not the actual bottleneck to successful programs. Translation, ADME, tox,

PK, manufacturability, patient selection, those are the bottlenecks. Structure prediction is necessary, not sufficient.

insitro and Recursion, the phenomics lane

insitro is much less about structure prediction and much more about building a data engine around human biology. The official positioning is integration of in vitro cellular data from its own labs with human clinical data, genetics, and machine learning. The CellPaint-POSH paper published in Nature Communications in 2025 makes the technical nuance much clearer. POSH combines pooled CRISPR perturbation, Cell Painting, and self-supervised representation learning to infer gene function and disease biology at scale. So insitro's comparative advantage is upstream and translational. It is trying to learn disease state and intervention biology from richer human-relevant data, rather than guessing whether a small molecule will fit a binding pocket. Whether that bet pays off depends on whether the resulting models generalize beyond the cell types and perturbations in the training set, which is honestly still an open question for the entire phenomics field.

Recursion is the clearest phenomics-first player and has been since well before the rest of the field caught on. Its platform language is about Maps of Biology and Chemistry, high-content perturbational data, and large proprietary biological and chemical datasets. The bet is similar in spirit to insitro but the scale and the wet lab automation are different. Recursion has been generating petabyte-scale image data for a long time and the data moat is real. The harder question is what to do with all of it. Recursion absorbed Exscientia in November 2024, which gave it a generative chemistry leg the original platform did not really have. The industrial logic of that deal is sound. The integration story has been bumpy in practice, with program shedding and headcount changes through 2025, and the combined entity has not yet shown the world the integrated end-to-end story it promised at deal announcement. The capital base is still impressive, the platform is still differentiated, but there is some operational risk that gets glossed over in the bull case.

The fair summary on the phenomics lane is that the data moat is durable, the model story is improving fast, but the translation from cellular phenotype to actual clinical benefit remains the hardest leap, and nobody has fully cracked it yet.

Iambic and Genesis, the translational and generative chem lane

Iambic is aiming at a different bottleneck than the structure folks or the phenomics folks. Enchant is positioned as a multimodal transformer trained across many data sources to predict key clinical properties from mostly preclinical information, with the explicit claim that it helps bridge the data wall between discovery-stage and human-stage R&D. So Iambic is less about target ID, less about protein-ligand pose, and more about translational risk reduction layered on top of medicinal chemistry and candidate selection. In plain terms, Isomorphic and Chai are asking what binds and how, insitro and Recursion are asking what biology matters and in whom, and Iambic is asking which candidates are most likely to survive the trip from preclinical to clinic. That is a real and underserved bottleneck. The honest caveat is that Enchant is still presented primarily through company materials and press coverage rather than through a peer-reviewed flagship methods paper, so the external validation is thinner than the structure prediction work.

Genesis Therapeutics often gets left off these lists, which is strange because its \$200M Series B co-led by Andreessen Horowitz puts it in the same neighborhood as Iambic, and its GEMS platform is a meaningfully different technical bet. Genesis leans on graph neural networks for molecular property prediction, with a focus on potency, selectivity, and ADME prediction in the design stage, rather than on structure foundation models or pure phenomics. The closest analog is probably Iambic in terms of where in the pipeline it is trying to add value, but the model architecture is different, and the company is older, with a longer track record of internal asset development. For investors who want exposure to the design and optimization layer specifically, Genesis is closer to the front of the field than its press footprint suggests.

Insilico, the only one with a Phase 2 human readout

Insilico Medicine is the most product-shaped of the AI-native discovery companies, and it is the only one on this list with a clinically validated AI-discovered asset. The Pharma.AI platform is split across PandaOmics for target discovery and Chemistry42 for molecule generation and optimization, and Chemistry42 in particular combines generative AI with physics-based methods, which is a more nuanced story than the usual “language model for molecules” pitch. The asset that matters here is rentosertib, also known as ISM001-055, a TNIK inhibitor for idiopathic pulmonary fibrosis. The Phase 2a results were published in Nature Medicine in June 2025, with further studies in kidney fibrosis and an inhaled IPF formulation planned for 2026. There is also ISM5411, a gut-restricted PHD1/2 inhibitor for inflammatory bowel disease that has completed Phase 1.

The other very real thing about Insilico is that it is the only one on this list that has actually tapped public equity markets. Insilico raised about \$293M in its December 30, 2025 Hong Kong Stock Exchange IPO, becoming the first AI-driven biotech to list on the HKEX Main Board under Chapter 8.05 listing rules. That offering was the largest biotech IPO in Hong Kong in 2025 by funds raised, and the cornerstone book included Lilly, Tencent, Temasek, Schroders, UBS AM, Oaktree, E Fund, and Taikang Life Insurance. Lilly and Tencent each subscribed for the first time as cornerstone investors in a biotechnology company, which is a small but meaningful signal about cross-industry conviction in AI-native R&D. Combined with more than \$500M raised privately across rounds backed by Warburg Pincus, Qiming, WuXi AppTec, B Capital, Prosperity7, OrbiMed, Deerfield, and others, Insilico is now sitting on a roughly \$800M total disclosed capital base, with revenue (yes, real revenue) of \$85.8M for 2024 and a net loss of \$17.4M, per the prospectus.

Whatever someone thinks of any individual platform claim, the asymmetry is real. Insilico is the only company in this group that can point to a Phase 2 readout in humans for a fully AI-discovered, AI-designed asset. Everyone else is still arguing about model architectures and benchmark scores. Clinical data is the only real moat in this industry over the long run, and Insilico is the first to get there at meaningful scale.

The papers that actually matter

For the industry-led reading list, four papers are unavoidable. AlphaFold 3, “Accurate structure prediction of biomolecular interactions with AlphaFold 3,” published in Nature in 2024, is the core Isomorphic and Google DeepMind paper extending structure prediction to joint complexes across proteins, nucleic acids, small molecules, ions, and modified residues. [Chai-1: Decoding the molecular interactions of life](#), published as a 2024 technical report and bioRxiv preprint, is Chai Discovery’s main structure prediction paper and the cleanest comparison point to AlphaFold 3 for anyone who wants to actually use a model commercially without negotiating with Alphabet. [Boltz-1 and Boltz-2 from MIT](#) belong on the same shelf for anyone who wants the open weights and training code path. The third is “[A pooled Cell Painting CRISPR screening platform enables de novo inference of gene function](#)” from insitro, published in Nature Communications in 2025, which is the strongest recent insitro methods paper and the cleanest articulation of the phenomics-plus-CRISPR-plus-self-supervised-learning thesis. The fourth is the [Iambic Enchant white paper](#), which is not a peer-reviewed journal article and should be read with that caveat, but is still the clearest articulation of the translational prediction lane right now.

The RFDiffusion and RFantibody work coming out of David Baker's lab at the University of Washington is also unavoidable for anyone trying to understand where Xaira comes from intellectually. Baker is a Xaira co-founder and the researchers who built RFDiffusion and RFantibody in his lab are now part of Xaira. Anyone serious about generative biologics in the next two years should be reading Baker lab output continuously.

For the academia-led or academia-heavy reading, the most useful set is more about evaluation, target ID, and closed-loop discovery than about splashy company launches. The 2026 Nature Machine Intelligence paper "[Assessing the potential of deep learning for protein-ligand docking](#)" is one of the most useful reality-check papers in the field. It introduces PoseBench and shows that co-folding methods can beat older docking baselines, but also that models still struggle with novel binding poses, multiligand settings, and chemical specificity.

The 2026 npj Drug Discovery paper "[AI-guided competitive docking for virtual screening and compound efficacy prediction](#)" is notable because it pushes beyond pose prediction toward rank-ordering active vs inactive compounds and using pairwise competitive docking for prioritization.

The 2026 Nature Reviews Drug Discovery review "[Target identification and assessment in the era of AI](#)" is probably the cleanest recent synthesis if the interest is upstream target discovery rather than only structure prediction.

The 2026 Cell paper "[Deep-learning-based de novo discovery and design of therapeutic molecules guided by gene-expression signatures](#)" points to a transcriptomics-driven route for molecule generation rather than pure structure-first design.

And the 2025 Science paper "[Active learning framework leveraging transcriptomics identifies modulators of disease phenotypes](#)" matters because it moves the conversation toward closed-loop wet-lab learning systems instead of static benchmark chasing.

If a reader only reads one company-led paper and one academic paper, the highest signal pair is probably AlphaFold 3 from Nature 2024 plus the PoseBench paper from Nature Machine Intelligence 2026. AlphaFold 3 is the unlock. PoseBench is the cold shower.

What the moat is becoming

The capital is still flowing most aggressively into firms trying to own the full stack. Proprietary data generation, multimodal foundation models, generative chemistry, automated wet labs, translational prediction, and at least some path to internal asset creation. The paper frontier, meanwhile, is shifting from “can we predict structures” toward “can we rank actives, generalize to new chemistry, incorporate phenotypes, reduce downstream attrition, and run closed-loop experiments.” That is the right shift. Structure prediction was the unlock around 2020 to 2024. It is not the full moat in 2026.

The blunt version is this. Isomorphic and Chai are leading the structure-foundation-model lane. Boltz is leading the open structure-foundation lane. insitro and Recursion are leading the biology-data and phenomics lane. Iambic and Genesis are leading the translational and generative chem lane. Insilico is the most modular, most productized, and the only one with a Phase 2 human readout. Xaira is the wildcard with the deepest capital and the strongest generative biologics talent density. Eikon is the new public-market entrant with one of the largest total capital bases in the field. Recursion plus Exscientia is the most ambitious integration story but with real operational risk in the near term.

The harder truth underneath all of this is that no single technical layer is the moat anymore. The moat is becoming the integration of proprietary perturbational data, generative models, automated wet labs, and clinical translation infrastructure, with patient-relevant data as the actual scarce input. That is exactly why the well-capitalized players are all trying to own the full stack, and exactly why the question of who has the best paper is becoming less predictive of who will have the best platform than it was three years ago. Clinical assets in humans are now the differentiating column on any honest market map. Right now, only one company in this group has a real one. Everyone else is trying to catch up to that fact.

Subscribe to Thoughts on Healthcare Markets and Technology

By Special Interest Media · Hundreds of paid subscribers

We cover broad topics at the intersection of health tech, health policy, health entrepreneurship and health investing.

By subscribing, you agree Substack's [Terms of Use](#), and acknowledge its [Information Collection Notice](#) and [Privacy Policy](#).



1 Like • 1 Restack

← Previous

Next →

Discussion about this post

Comments

Restacks



Write a comment...