AIxBio Horizon Scan: Spring 2026

Introduction

Since the AIxBio Horizon Scan Winter 2025-2026 published in March 2026, there has been steady, incremental progress across AI-enabled biological tools.

Protein design tools have continued to improve, agentic coding tools have matured in ways that lower barriers to computational biology, and commercial AI companies are making significant investments in the life sciences. The more consequential movements have been in policy, including the first serious bipartisan legislation mandating DNA synthesis screening and continued disruption from U.S. federal funding changes and geopolitical competition.

Key Trends

Continued Incremental Improvement of Protein and Genomic Language Models

The release of RFdiffusion3 by the University of Washington's Institute for Protein Design in December 2025 consolidated capabilities that previously required separate specialized models into a single framework operating at full atomic resolution. The tool is approximately ten times faster than its predecessor and has been released as open-source software. Separately, new work on protein binder design using RFdiffusion was published in *Nature Communications* in early 2026, achieving binding affinities comparable to therapeutic monoclonal antibodies. While these developments represent meaningful improvements, experts characterized the period as one of continued refinement rather than fundamental shifts.

New evaluations of biological foundation models continue to struggle with generalization, particularly for the prediction of viral mutation effects and cases that differ significantly from the training data. Purpose-built tools continue to outperform large foundation models on specific tasks. For example, a November 2025 benchmarking study in *Nature Communications* found that general-purpose DNA foundation models were less effective

than specialized models at predicting gene expression and identifying causal mutations that change gene expression, although fine-tuning may close this gap.

Frontier LLM Labs are Betting on Biology with Cautious Public Access

Some companies' policies are limiting how quickly the public can access LLM agents using biological AI tools. Agentic coding products, such as Claude Code and OpenAI's Codex, have increased the accessibility of computational biology workflows. These tools allow researchers to give high-level direction to an AI agent that can write, execute, and debug code with minimal intervention, translating those directions directly to bioinformatics and biological design tool usage, but interviews suggest that model safeguards often restrict benevolent use. Code generation has improved enough that biological tool codebases are now more accessible than they were with the dedicated scientific agents as of mid-2025, including novel tool generation. However, some experts noted a pause in aggressive integration of LLMs with external biological tools, possibly reflecting a more cautious policy by Anthropic and OpenAI.

Commercial AI companies are making significant bets on biology. Anthropic acquired Coefficient Bio, a biotech AI startup focused on drug discovery, and partnered with the Allen Institute and HHMI for frontier scientific research. OpenAI has established partnerships with Ginkgo Bioworks and Retro Biosciences. The Chan-Zuckerberg Biohub's acquisition of EvolutionaryScale and its pivot toward AI-first life science research represent a significant new center of gravity. These moves suggest that large AI companies and labs see biology as a major commercial and innovation opportunity, which could accelerate progress and also shift innovation from academia toward industry.

Laboratory Automation and Cloud Labs Are Expected to Advance

Progress in laboratory automation continues but has not yet produced the transformative changes some anticipated. Ginkgo Bioworks has reinvented itself as a cloud lab platform, and its collaboration with OpenAI demonstrated GPT-5 autonomously designing and executing 36,000 experiments in a closed-loop workflow. However, experts noted that cloud labs remain largely "monomodular" and non-agentic: able to automate specific workflows but unable to flexibly handle the range of tasks a human researcher would perform. The gap between AI's computational capabilities and physical laboratory execution remains a significant constraint.

Data Gaps Will Widen the Open vs. Proprietary Divide

Data remains a critical bottleneck. Large-scale data collection remains driven by structures of conventional scientific inquiry rather than novel sampling systems optimized to improve model performance. Currently, optimization is targeted at reducing design-build-test-learn-loop time.

Proprietary datasets are emerging as a key way to overcome this bottleneck. Isomorphic Labs announced its Drug Design Engine (IsoDDE) with substantial performance gains attributed to proprietary data and methods not made publicly available. Boltz has signed agreements with Pfizer to train models on proprietary pharmaceutical data. Although the OpenFold consortium released all of its training data, this does not seem to be representative of the larger trend. If key capability drivers increasingly reside in private datasets, this could concentrate advanced capabilities among well-resourced commercial actors rather than distributing them broadly. A November 2025 Executive Order from the White House meant to advance U.S. AIxBio capabilities requires "uniform and stringent data access and management processes… for non-Federal collaborators" suggesting that even government sponsored work in this space may not be open.

These trends are not restricted to the United States. Shanghai's Zhangjiang science hub is driving AI-driven drug discovery by integrating molecular design, synthesis, cell, and animal studies all under one roof. This integrated model magnifies the opportunities AI presents to speed iteration and for short-term health advances, especially in conjunction with China's streamlined clinical trial system. Further, Alibaba may be reversing course—shifting from open-source models to closed releases to focus on profit.

The open-source biological AI ecosystem remains active: Boltz-2 is widely used for protein structure prediction, OpenFold has released preview models, and RFdiffusion3 was released open-source. Meanwhile, proprietary tools are reserved for private use (companies like Nabla and Isomorphic Labs do not provide external access to their latest models). While this limited access may protect against some potential misuse scenarios, it also may decrease visibility into the capabilities of those models, which could complicate ongoing governance discussions.

An interviewee noted a growing number of companies offer biological design services through web interfaces or APIs, and the proliferation of dedicated "AI scientist" chatbots further illustrates commercialization. Such services provide new examples of managed access approaches, straddling the gap between open and closed, providing relatively open access to capabilities while keeping the underlying weights and data obscured.

Policy and Geopolitical Developments

The most significant policy development is the introduction on February 4, 2026 of the Biosecurity Modernization and Innovation Act (S.3741) by U.S. Senators Tom Cotton and Amy Klobuchar. This bipartisan legislation would establish mandatory federal requirements for DNA synthesis providers to screen orders and verify customer identities, replacing the current voluntary framework. However, analysts have noted that the bill's reliance on homology-based screening may be insufficient in an era when AI design tools can generate functional but sequence-novel agents that would not match existing

watchlists. Current proposals recommend exploring function-based screening approaches.

Geopolitically, U.S. academic research funding cuts are beginning to take effect, with some labs sustaining existing work but unable to launch new efforts. The gap between leading AI programs globally has narrowed, with Chinese AI models demonstrating increasingly competitive capabilities. Technology diffusion across borders has proven difficult to constrain, and U.S. export controls on advanced hardware have had limited effect on the pace of international AI development. ByteDance has expanded its AIxBio work and is gaining visibility. The geopolitical environment often positions the U.S. and China as competitors, which makes official cooperation on biosecurity difficult. However, track II dialogues on biosecurity and AI safety continue to serve as an effective channel for communication between American and Chinese experts. Opportunities to raise U.S.-China cooperation on biosecurity as a political priority will likely depend on the outcome of the May 14-15 Trump-Xi summit.

Emerging Issues

Understanding Risk Tolerance to Improve Evaluation Frameworks

Multiple experts have flagged an urgent need to move beyond evaluations as the primary governance tool. The field has focused heavily on producing evaluation metrics without first establishing risk tolerances or red lines. Without defined risk tolerances or clear red lines, evaluation results cannot be translated into actionable policy.

Uplift Studies as Early Capability Indicators

Multiple recent studies (1, 2, 3) have begun to assess whether LLMs provide meaningful uplift to non-experts attempting wet lab biological work. Early results present a mixed picture:

- No significant uplift for novice lab work: A randomized control trial found no statistically significant LLM uplift for novices performing hands-on laboratory tasks, suggesting that, as of mid-2025, tacit knowledge remains a significant barrier to wet lab work.
- Higher completion rates, but underpowered: This study was lacked statistical power for significance, but did show higher completion rates on wet-lab objectives in the LLM group.
- Substantial uplift for *in silico* tasks: Found substantial uplift for *in silico* biology tasks.

These empirical uplift studies are substantially more informative than typical benchmarking for understanding actual biosecurity risk, because they measure whether

strong model performance on written biology tests actually translates into real-world capability gains for the users of concern.

LLMs as Scientific Thought Partners

Experts noted a relatively underexplored dimension of how scientists actually use LLMs today: not primarily for autonomous research, but as thought-partners that synthesize literature, suggest better experimental questions, and present information in the form most useful to the researcher. This augmentation may have significant near-term impact on research productivity even without breakthroughs in autonomous agents.

Future Predictions

In the next 6-18 months, experts anticipate continued steady improvement across LLMs, biological design tools, and agent capabilities. Key likely developments include further integration of AI agents with wet lab systems, additional generations of protein and genomic language models, and LLMs with more advanced biological capabilities. Whether S.3741 advances through Congress in this time frame is unclear.

On a 2-5 year horizon, the most transformative capability is likely to be the emergence of lab-in-the-loop biological models, where molecular design, experimental testing, and model retraining occur in continuous, automated cycles. The self-reinforcing nature of these systems could produce compounding capability gains that are rapid, difficult to anticipate in advance, and hard to reverse once underway. They may also eliminate the wet-lab-expertise bottleneck that currently limits who can translate computational designs into functional biological systems.

The trajectory identified in the Winter 2025-26 scan continues largely as expected, with incremental advances across multiple fronts. Capability development continues to outpace the governance frameworks needed to manage it, and the window for establishing effective oversight mechanisms continues to narrow.

This Horizon Scan captures developments and interviews through the beginning April 1, 2026. Several notable frontier model releases after this date warranted inclusion given their relevance to the AIxBio landscape.

Step-Changes in Frontier LLM Capabilities

Two major frontier models were released in mid-April. Anthropic's Claude Mythos Preview (April 8) represents a new, higher-capability model tier, while OpenAI's GPT-5.5 (April 23) continues a roughly six-week release cadence for its GPT-5 family. Both models showed meaningful gains on biological benchmarks:

- GPT-5.5 improved over its predecessor on BixBench (bioinformatics) and GeneBench (multi-stage genetics data analysis)
- Anthropic published BioMysteryBench, a new bioinformatics benchmark on which Mythos solved 30% of problems that a panel of domain experts could not.

Both companies conducted biosecurity-specific evaluations and classified their models at equivalent risk tiers: meaningful uplift for users with some technical background, but not equivalent to expert-level assistance for novel catastrophic weapons development. As stated in their system cards, SecureBio noted uncertainty about the robustness of GPT-5.5's biological safeguards to circumvention while Anthropic characterized its own mitigations as making catastrophic biological risk “very low but not negligible”.

Mythos's most striking capabilities are in cybersecurity, where it saturated existing vulnerability benchmarks and demonstrated the ability to discover zero-day vulnerabilities in real codebases. Anthropic withheld the model from general release, representing the first such action by a major lab since OpenAI briefly withheld GPT-2 in 2019. Instead, they provided it only to vetted cybersecurity partners through an initiative called Project Glasswing. This managed-access approach represents a novel governance model. However, reports of unauthorized access within days of the announcement illustrate the difficulty of maintaining such controls. The White House has called for Anthropic to refrain from a wider release of this model. Cybersecurity capabilities could present concerns for high-containment labs, medical countermeasure infrastructure, or corruption of biological databases. While the cybersecurity step-change is distinct from biological capabilities, it suggests that similar leaps in biological reasoning could arrive in future model generations. This could present significant risks since biological vulnerabilities are drastically harder to patch than their cyber counterparts.

Both OpenAI and Anthropic are also building life-science-specific products. On April 16, OpenAI released GPT-Rosalind, its first domain-specific frontier model, purpose-built for biology, drug discovery, and translational medicine. The model is available as a research preview through OpenAI's trusted access program, with early partners including Amgen, Moderna, the Allen Institute, and Thermo Fisher Scientific. Unlike general-purpose models applied to biology tasks, GPT-Rosalind is optimized for biological reasoning and comes with a Life Sciences plugin for Codex that connects to more than 50 scientific tools and data sources, serving as an orchestration layer for workflows in areas like human genetics, functional genomics, and protein structure. On LABBench2, the model outperformed GPT-5.4 on six of eleven tasks, with the largest gains in molecular cloning reagent design.

Anthropic made a similar push in late 2025 with a suite of life-science connectors and skills for scientific collaboration.It followed in January 2026 with Claude for Healthcare , featuring models trained specifically for healthcare and life sciences tasks and native integrations to medical and scientific databases including PubMed and ICD-10 codes. Anthropic's approach emphasizes connector-based access to platforms like Benchling for lab notebooks, positioning Claude as an interface layer across existing research infrastructure. The emergence of competing, purpose-built life science model products

from both leading frontier labs marks a shift from the general-purpose LLM applications described earlier in this scan toward dedicated scientific tooling. This shift accelerates the integration trends noted above while also concentrating advanced biological AI capabilities behind managed-access enterprise programs.

###