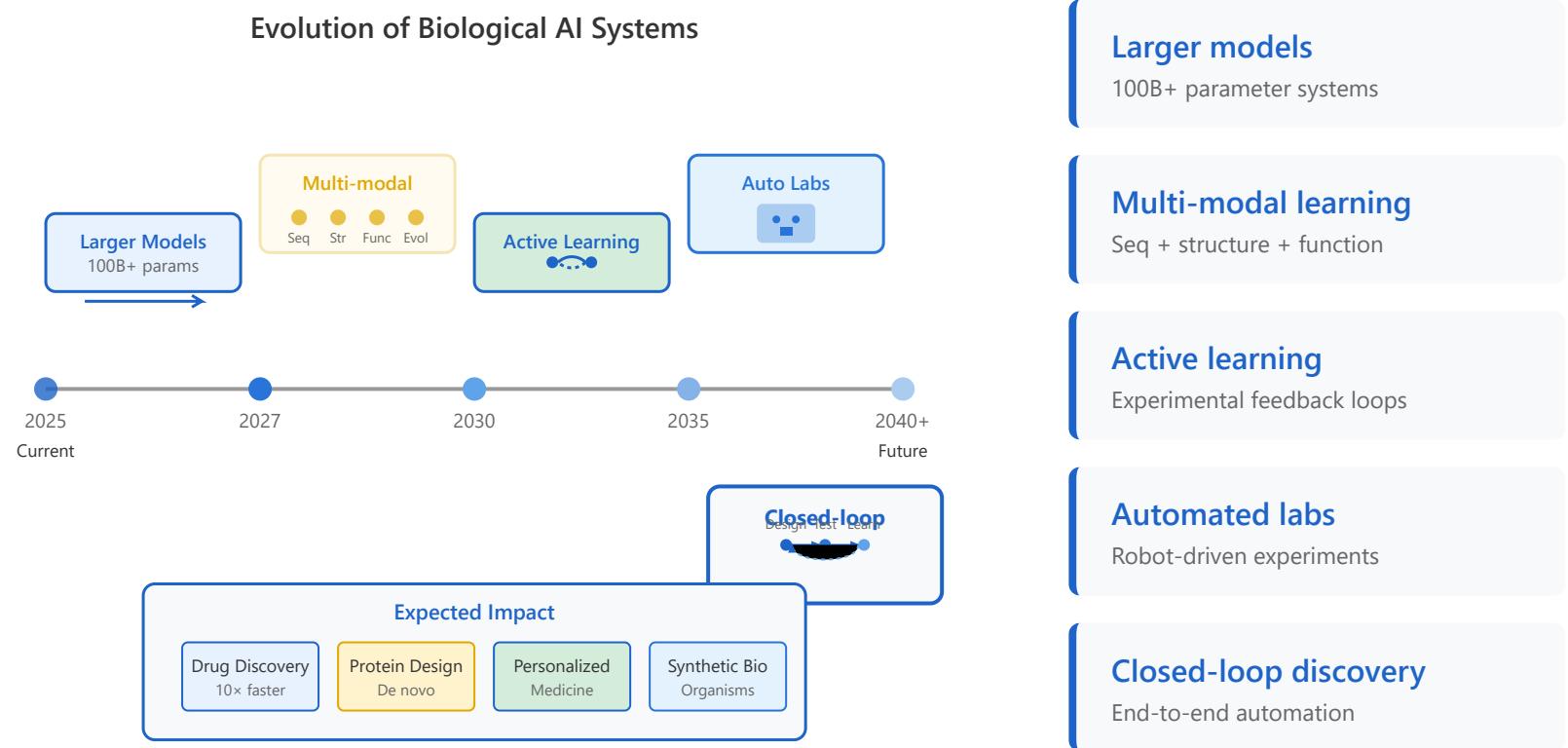


Future Perspectives



1

Larger Models: Scaling to 100B+ Parameters

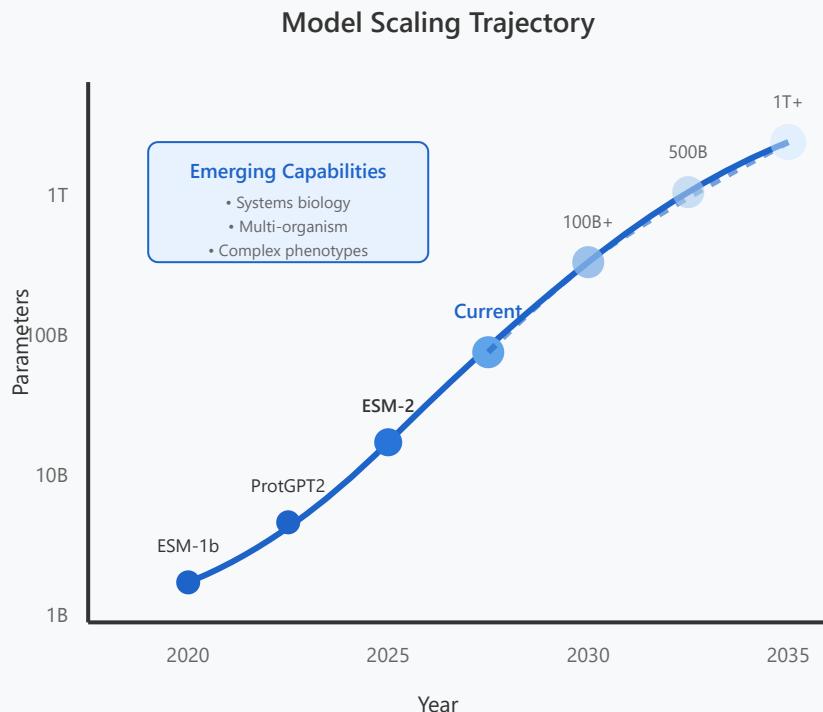
The next generation of biological AI models will scale beyond current architectures, reaching 100 billion parameters or more. This scaling enables models to capture increasingly complex biological patterns, from molecular interactions to systems-level behavior.

Key Capabilities

- ▶ Understanding protein-protein interaction networks at proteome scale
- ▶ Predicting complex phenotypes from genomic sequences
- ▶ Modeling cellular pathways and metabolic networks
- ▶ Cross-species transfer learning for rare organisms
- ▶ Integration of evolutionary information across phylogenies

Expected Impact

Larger models will enable prediction of complex biological phenomena that current models cannot address, such as multi-gene disease mechanisms, organism-level responses to perturbations, and emergent properties in synthetic biological systems.



2 Multi-modal Learning: Integration Across Data Types

Overview

Future AI systems will seamlessly integrate multiple biological data modalities including sequences, 3D structures, functional annotations,

and evolutionary information. This holistic approach mirrors how biologists naturally reason about biological systems.

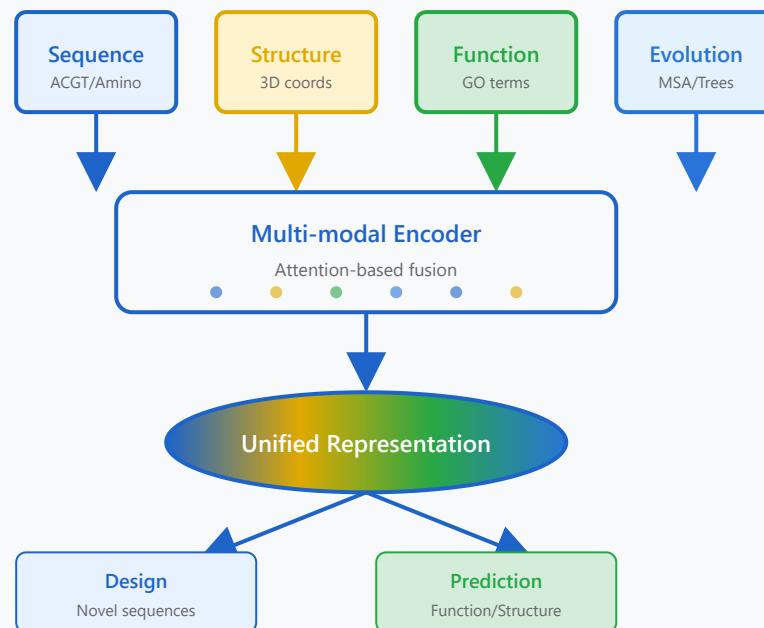
Data Modalities

- ▶ **Sequence:** Genomic and protein sequences with contextual information
- ▶ **Structure:** 3D conformations, dynamics, and structural ensembles
- ▶ **Function:** Biochemical activities, cellular localization, interactions
- ▶ **Evolution:** Phylogenetic relationships and conservation patterns
- ▶ **Expression:** Temporal and spatial gene expression profiles

Expected Impact

Multi-modal models will provide comprehensive understanding of biological entities, enabling accurate prediction of functional effects from sequence alone and facilitating the design of proteins with specified structures and functions.

Multi-modal Integration Architecture



3 Active Learning: Experimental Feedback Loops

Overview

Active learning strategies enable AI systems to identify the most informative experiments to conduct next, dramatically improving data efficiency. The model learns from experimental results and iteratively refines its predictions through strategic sampling.

Core Components

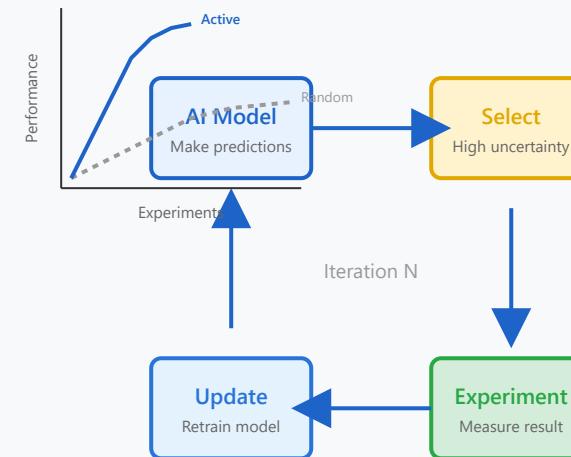
- ▶ **Uncertainty estimation:** Quantifying confidence in model predictions
- ▶ **Acquisition functions:** Selecting maximally informative experiments
- ▶ **Batch optimization:** Planning multiple parallel experiments
- ▶ **Transfer learning:** Leveraging knowledge across related tasks
- ▶ **Cost-aware selection:** Balancing information gain with experimental cost

Expected Impact

Active learning can reduce the number of required experiments by 10-100x, accelerating discovery cycles and enabling exploration of vast sequence spaces that would be prohibitively expensive with traditional approaches.

Active Learning Cycle

Efficiency Comparison



4

Automated Laboratories: Robot-Driven Experiments

Overview

Automated laboratories combine robotic systems, microfluidics, and AI control to execute thousands of experiments in parallel with minimal human intervention. These systems can operate continuously, generating high-quality data at unprecedented scales.

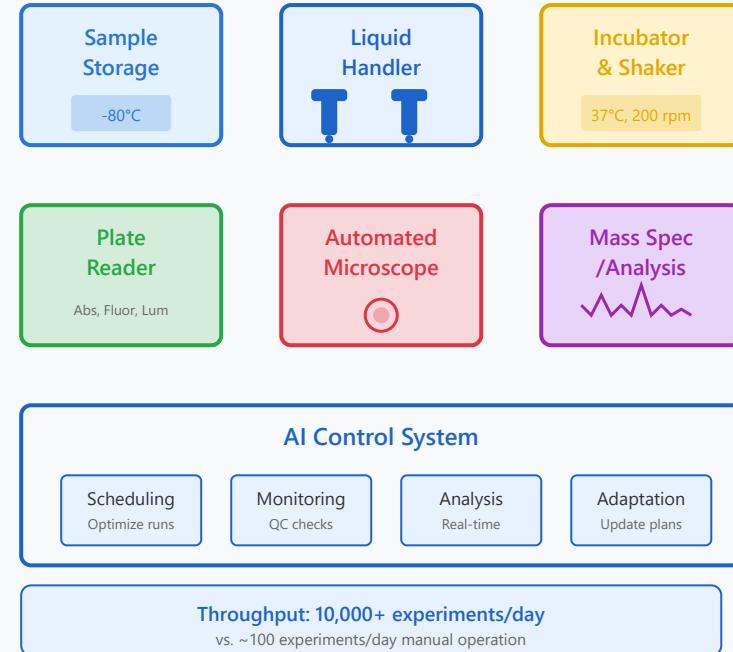
Key Technologies

- ▶ **Liquid handling robots:** Precise pipetting and sample preparation
- ▶ **High-throughput screening:** Parallel assays in microplate formats
- ▶ **Microfluidics:** Miniaturized reactions and cell manipulation
- ▶ **Real-time monitoring:** Automated imaging and spectroscopy
- ▶ **Adaptive protocols:** AI-driven experimental adjustments

Expected Impact

Automated labs will enable 24/7 experimentation with throughput 1000× higher than manual approaches, while maintaining reproducibility and reducing costs. This will democratize access to advanced experimental capabilities.

Automated Laboratory Architecture



5

Closed-loop Discovery: End-to-End Automation

Overview

Closed-loop discovery represents the ultimate integration of AI and automated experimentation, where the entire scientific discovery process operates autonomously. The system designs experiments, executes them, analyzes results, updates models, and iterates without human intervention.

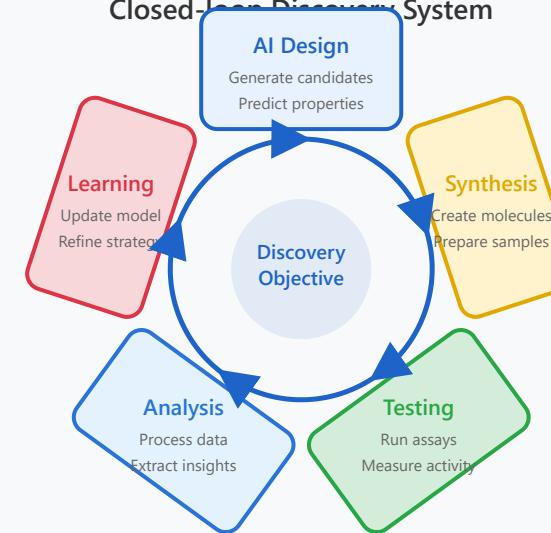
System Components

- ▶ **Autonomous design:** AI generates novel hypotheses and experiments
- ▶ **Robotic execution:** Automated labs perform experiments
- ▶ **Real-time analysis:** Instant processing of experimental data
- ▶ **Model updating:** Continuous learning from new results
- ▶ **Goal optimization:** Multi-objective optimization toward targets

Expected Impact

Closed-loop systems will compress discovery timelines from years to weeks, enable exploration of combinatorially vast design spaces, and accelerate the pace of innovation in drug discovery, materials science, and synthetic biology by 100× or more.

Closed-Loop Discovery System



Key Performance Indicators

Cycle Time: 1-7 days Throughput: 100s/cycle Improvement: 10-100×