The Life Sciences AI Handbook

AI for Biomedical Discovery, Biotechnology, and Translational Research

Bryan Tegomoh, MD, MPH

May 2026

Table of contents

Welcome

Biomedical discovery, biotechnology, and translational research

The Life Sciences AI Handbook

AI for Biomedical Discovery, Biotechnology, and Translational Research

A professional handbook for researchers, biotechnology teams, computational biologists, physician-scientists, and students evaluating AI systems across molecules, cells, experiments, and therapeutic development.

Start Reading Model Index

Core Reading Path

Life sciences AI is best read as an experimental discipline. Models that predict structures, design proteins, rank targets, generate molecules, or forecast perturbations need evidence tied to assays and decisions.

Demonstrated: protein structure prediction, selected protein design tasks, molecular property benchmarks, single-cell representation learning, perturbation prediction in bounded settings, and closed-loop experiments.

Theoretical: integrated models that connect molecules, cells, tissues, and experiments across laboratories.

Beyond current capabilities: fully autonomous biological discovery without experimental validation or human governance.

Handbook Map

Part VEngineering and Automation

Part VIPractice and Governance

- **Part I: Foundations** covers scope, data infrastructure, foundation models, and evaluation.
- **Part II: Molecular AI** covers protein structure, protein design, antibodies, genome models, and variant effects.
- **Part III: Therapeutics AI** covers target identification, small molecules, RNA and vaccines, trials, and translational failure modes.
- **Part IV: Cellular and Systems Biology** covers single-cell models, spatial omics, image phenotyping, virtual cells, and multi-omics.
- **Part V: Engineering and Automation** covers self-driving labs, robotic labs, synthetic biology design tools, and agentic workflows.
- **Part VI: Practice and Governance** covers benchmarks, reproducibility, information hazards, workforce, and institutional readiness.

Explore the Handbook Series

The Physician AI Handbook

Clinical AI across medical specialties, implementation, safety, workflow, liability, and physician-facing evaluation.

Visit handbook

The Public Health AI Handbook

AI for surveillance, forecasting, public health operations, population health analytics, and deployment in health agencies.

Visit handbook

The Biosecurity Handbook

Biological risk, dual-use research oversight, biosafety, AI-bio convergence, governance, and risk evaluation.

Visit handbook

The Life Sciences AI Handbook

AI for molecular design, cellular systems, biomedical discovery, biotechnology, automation, and translational research.

Visit handbook

License and Citation

Suggested citation: Tegomoh, B. (2026). *The Life Sciences AI Handbook: AI for Biomedical Discovery, Biotechnology, and Translational Research.* DOI pending. URL: https://lifesciencesaihandbook.com

Preface

I wrote this handbook because the constructive side of AI in biology deserves a practical, evidence-centered home. Clinical AI, public health AI, and biosecurity each answer different questions. This book focuses on discovery and engineering: molecules, cells, experiments, therapeutics, and the research systems that connect them.

The goal is not to rank every model or chase every release. The goal is to give researchers a stable way to read claims, design evaluations, and decide when an AI output deserves experimental attention.

Author Information

Bryan Tegomoh, MD, MPH is a physician-scientist and epidemiologist. His work spans biomedical research, public health, biosecurity, and AI evaluation.

How the Book Uses Evidence

The handbook uses three evidence tiers throughout:

- **Demonstrated:** supported by published evidence, official documentation, or reproducible benchmark results.
- **Theoretical:** plausible based on current methods, but not yet established for routine use.
- **Beyond current capabilities:** not supported by credible evidence with current systems.

Claims about AI systems in biology are placed inside those tiers so readers can separate current utility from research aspiration.

Executive Summary

Life sciences AI has crossed from isolated modeling tasks into the operating layer of biomedical discovery. Protein structure prediction, protein design, genome models, single-cell foundation models, molecular generation, self-driving laboratories, and agentic research workflows now affect how teams choose experiments.

The practical risk is not only overclaiming. It is using the wrong evidence for the wrong decision.

Core Takeaways

- Protein structure prediction is now a routine research input for many proteins, but structure is not function, mechanism, safety, or clinical value.
- Protein design and antibody design are experimental disciplines. Designed sequences need expression, binding, specificity, stability, immunogenicity, and manufacturability review.
- Genome and variant models are strongest when tied to measured functional genomic outputs. Clinical or organism-level claims require additional evidence.
- Therapeutics AI adds value when it improves a decision in target selection, chemistry, trial design, or evidence generation. It does not remove attrition.
- Cellular AI is moving from annotation toward perturbation prediction. Virtual cell claims should be read as task-specific unless proven otherwise.
- Self-driving labs and agentic workflows require protocol standards, instrument reliability, provenance, and human authorization gates.
- Information hazards are handled through deliberate disclosure, reproducibility planning, and review of release details.

Reading Rule

Every major capability claim in the handbook should land in one of three tiers: Demonstrated, Theoretical, or Beyond current capabilities.

Practical Use

Use this handbook to evaluate a model, design a validation plan, brief a research team, review a vendor claim, or decide whether a paper changes a program decision.

How to Read This Handbook

This handbook is organized by research object: data, molecules, therapeutics, cells, laboratories, and practice. Each chapter uses the same three-tier structure so a reader can locate the evidence level before acting on a claim.

Fast Path

- For model selection, start with Evaluation Principles and Benchmarks.
- For molecular design, read Protein Structure Prediction, Protein Design, and Antibody and Biologic Design.
- For therapeutics work, read Target Identification, Small Molecule Generation and ADMET, and Translational Evidence.
- For cell modeling, read Single-Cell Foundation Models and Perturbation Prediction.
- For automation, read Self-Driving Laboratories and Agentic Science Workflows.

Evidence Tiers

- **Demonstrated** means the claim is supported by published evidence, official documentation, or reproducible benchmark results.
- **Theoretical** means the claim is plausible but not established for routine practice.
- **Beyond current capabilities** means the claim is not supported by current evidence.

How to Use Citations

Citations are inline links. Peer-reviewed sources use author-year labels. Official program and database sources use organization-year labels. Preprints are labeled when used.

Scope and Limitations

This handbook is for education, research planning, and technical evaluation. It is not medical advice, public health guidance, legal advice, biosafety guidance, or regulatory advice.

In Scope

- AI models for biological sequence, structure, molecular, cellular, imaging, and experimental data.
- Biomedical discovery, biotechnology, and translational research workflows.
- Evaluation, reproducibility, information hazards, and institutional readiness.

Out of Scope

- Patient-specific medical advice.
- Operational biosafety protocols.
- Step-by-step instructions for biological misuse.
- Product endorsement.
- Legal or regulatory determinations.

Claim Calibration

The handbook uses Demonstrated, Theoretical, and Beyond current capabilities to prevent unsupported certainty. When evidence is mixed, the text favors the lower-confidence tier.

Part I: Foundations

AI for the Life Sciences

Life sciences AI is the use of machine learning systems to represent, predict, design, and evaluate biological objects. Its center of gravity is upstream of clinical care: molecules, cells, organisms, experiments, datasets, and translational research decisions.

Learning Objectives

- Distinguish discovery AI from clinical AI and public health AI.
- Separate prediction, generation, prioritization, and experimental control tasks.
- Use the Demonstrated, Theoretical, and Beyond current capabilities tiers when reading claims.

TL;DR

Life sciences AI is not one field. It is a set of modeling practices that share biological data constraints, experimental validation requirements, and high error costs. The first rule is to ask what biological object the model represents and what experiment would falsify the output.

Introduction

Biomedical AI programs depend on data that are structured for computation, not only archived for human reading. NIH Bridge2AI was created around this problem: AI-ready datasets require metadata, ethics, quality control, and workforce development rather than only larger files (NIH Bridge2AI, 2022). The handbook therefore treats data infrastructure, model capability, and validation as a single operating problem.

Demonstrated

Demonstrated capability includes structure prediction, sequence modeling, perturbation modeling, molecular property prediction, and closed-loop experimentation in bounded settings. AlphaFold 2 demonstrated high-accuracy single-chain protein structure prediction in CASP14-related evaluation (Jumper et al., 2021). RFdiffusion demonstrated experimental protein

design across several design tasks (Watson et al., 2023). GEARS demonstrated transcriptional perturbation prediction for selected genetic perturbation settings (Roohani et al., 2024).

Evidence Anchor	What It Supports	Practical Constraint
AlphaFold 2	Protein structure prediction at near experimental accuracy for many single-chain targets	Confidence measures, disorder, complexes, ligands, and conformational state remain separate questions
NIH Bridge2AI	AI-ready biomedical data as a fundable infrastructure problem	Dataset governance and metadata quality shape model value
ARPA-H IGoR	AI-linked biomedical research infrastructure as a current public program	Agentic systems require protocol standardization and laboratory verification

Theoretical

Theoretical capability includes general biological reasoning across scales, unified models of disease mechanism, and experiment selection that transfers across laboratories. These goals are plausible because models already operate over sequences, structures, images, single-cell profiles, and assay outputs. They remain theoretical when the same system has not shown external validity across organisms, laboratories, disease areas, and measurement platforms.

Beyond Current Capabilities

Beyond current capabilities includes automated discovery systems that replace experimental biology, universal virtual cells that reliably forecast organism-level phenotypes, and general-purpose therapeutic design without iterative measurement. Existing systems help choose better experiments. They do not remove the need for biological measurement.

Practice Notes

- Name the biological object first: sequence, structure, ligand, cell state, tissue, experiment, or clinical endpoint.
- Name the validation object second: assay, benchmark, prospective experiment, or independent dataset.
- Do not equate a model score with biological truth.

- Treat every vendor claim as a claim about a specific data distribution until proven otherwise.

Biological Data Infrastructure

Life sciences AI inherits the strengths and weaknesses of biological archives. Protein structures, sequences, small molecules, perturbation screens, and imaging datasets carry different biases, missingness patterns, and validation traditions.

Learning Objectives

- Map the main data classes used by life sciences AI systems.
- Identify when a dataset is machine-readable but not AI-ready.
- Explain why metadata, assay provenance, and negative examples matter.

TL;DR

Better models do not rescue poorly specified biological data. AI-ready data require provenance, assay context, versioning, licensing, and negative controls. The most useful model card is often a dataset card.

Introduction

The Protein Data Bank has served as a global archive for experimentally determined macromolecular structures since 1971 (wwPDB, 2026). RCSB PDB now also presents computed structure models beside experimental structures, which forces users to distinguish measurement from prediction (RCSB PDB, 2026). For biomedical AI, the same distinction applies across sequence archives, chemical databases, single-cell atlases, and image repositories.

Demonstrated

Demonstrated capability includes training and evaluating models on curated public resources. ChEMBL provides curated bioactivity data for drug-like molecules (Zdrazil et al., 2024). PubChem provides chemical substance, compound, and bioassay records through NIH infrastructure (Kim et al., 2023). AlphaFold DB and the ESM Metagenomic Atlas show how predicted structures became research resources at database scale (AlphaFold Protein Structure Database, 2026; ESM Metagenomic Atlas, 2026).

Evidence Anchor	What It Supports	Practical Constraint
PDB	Experimentally determined structure archive	Coverage follows what structural biology could measure
ChEMBL and PubChem	Chemical structure and bioactivity resources	Assay context and curation level differ across entries
Bridge2AI	AI-ready data standards as a program objective	Ethical sourcing, metadata, and fairness are part of data quality

Theoretical

Theoretical capability includes cross-database models that learn from raw sequence, structure, chemical, image, and text records without manual harmonization. This remains theoretical in many workflows because identifiers, assay conditions, version histories, and licensing terms often fail to align cleanly.

Beyond Current Capabilities

Beyond current capabilities includes biological datasets that fully encode the causal context of an experiment. No public archive contains all cell states, reagent histories, operator choices, instrument behavior, and environmental variables needed to remove experimental ambiguity.

Practice Notes

- Record data version, download date, accession source, and filtering logic.
- Separate experimental structures from computed structure models in tables and figures.
- Keep negative, failed, and inconclusive experiments when training prioritization systems.
- Audit license terms before mixing public, consortium, and commercial data.

Foundation Models for Biology

Biological foundation models learn reusable representations from large collections of biological measurements. The practical question is whether those representations transfer to a specific biological decision.

Learning Objectives

- Distinguish sequence, structure, cellular, and multimodal foundation models.
- Read claims about scale without confusing scale with external validity.
- Use task-specific evaluation before adopting a general representation.

TL;DR

A foundation model is useful when pretraining improves a downstream biological task under realistic validation. Model size, modality count, and dataset volume matter less than task transfer, assay fidelity, and external testing.

Introduction

Biology is well suited to representation learning because many data types are symbolic or structured: DNA, RNA, protein sequence, molecular graphs, images, and expression matrices. ESMFold showed that protein language model representations could support atomic-level structure prediction from sequence (Lin et al., 2023). Geneformer and scGPT moved the same broad idea into single-cell transcriptomics (Theodoris et al., 2023; Cui et al., 2024).

Demonstrated

Demonstrated capability includes reusable embeddings and task transfer in published benchmark settings. ESMFold demonstrated fast protein structure prediction from protein language model representations (Lin et al., 2023). Geneformer demonstrated transfer across gene network tasks in limited-data settings (Theodoris et al., 2023). scGPT demonstrated pretraining over large single-cell repositories with downstream tasks in cell type annotation, perturbation, and integration (Cui et al., 2024).

Evidence Anchor	What It Supports	Practical Constraint
ESMFold	Protein language representations linked to structure prediction	Accuracy and confidence differ from MSA-based methods
Geneformer	Single-cell transfer learning for gene network tasks	Training data and cell context shape transfer
scGPT	Single-cell pretraining across large repositories	Benchmark selection determines apparent gains

Theoretical

Theoretical capability includes a single model that supports molecular design, cell-state forecasting, tissue interpretation, and experiment planning. Current models usually specialize by modality or task family. Multimodal systems are expanding, but general biological validity remains an empirical question.

Beyond Current Capabilities

Beyond current capabilities includes foundation models that infer causal biology from observational pretraining alone. Perturbational data, experimental design, and mechanistic testing remain necessary for causal claims.

Practice Notes

- Ask what was masked or predicted during pretraining.
- Compare foundation-model features against simple baselines.
- Evaluate by biological split, not random row split, when testing transfer.
- Track whether the model saw related cell types, homologs, assays, or structures during training.

Evaluation Principles for Biomedical Discovery AI

Evaluation in life sciences AI must respect biology, not only machine learning convention. The wrong split, metric, or benchmark often makes a model look useful before it meets a new assay or a new laboratory.

Learning Objectives

- Choose evaluation splits that match biological use.
- Match metrics to experimental decisions.
- Detect leakage, scaffold bias, homolog leakage, and assay leakage.

TL;DR

The core evaluation question is not whether a model performs well on held-out rows. The core question is whether it improves a real experimental decision under the distribution where it will be used.

Introduction

Benchmarks in structural biology, molecular docking, and single-cell perturbation illustrate the same principle from different angles. CASP evaluates protein structure methods against blinded targets (Kryshtafovych et al., 2024). CAMEO provides continuous automated structure-prediction evaluation (CAMEO, 2026). PoseBusters showed that RMSD alone misses physically implausible docking outputs (Buttenschoen et al., 2024).

Demonstrated

Demonstrated capability includes evaluation regimes that expose failure modes hidden by simple metrics. CASP and CAMEO support community-level structure prediction assessment (Kryshtafovych et al., 2024; CAMEO, 2026). PoseBusters demonstrated that docking outputs

need chemical and physical validity checks in addition to geometric error (Buttenschoen et al., 2024).

Evidence Anchor	What It Supports	Practical Constraint
CASP	Blinded community assessment for structure prediction	Targets and categories change across rounds
CAMEO	Continuous server evaluation	Automated evaluation depends on target release and criteria
PoseBusters	Physical plausibility checks for docking poses	RMSD-only evaluation rewards incomplete success

Theoretical

Theoretical capability includes benchmark suites that forecast real discovery productivity. Such benchmarks are plausible when they contain prospective experiments, cost-aware decisions, and multiple failure categories. They remain incomplete when they only compare model scores.

Beyond Current Capabilities

Beyond current capabilities includes a universal score that ranks models across every biological domain. Structure prediction, compound screening, cellular response, and clinical translation require different ground truth and different error costs.

Practice Notes

- Use scaffold, sequence-family, cell-line, target, and time splits when those match deployment.
- Report calibration, uncertainty, and failure modes beside headline metrics.
- Include physical, chemical, and biological validity checks.
- Prefer prospective validation when model output drives experiment selection.

Part II: Molecular AI

Protein Structure Prediction

Protein structure prediction changed the operating baseline for structural biology. The hard questions moved from whether a predicted fold is useful to which molecular states, partners, ligands, and experimental contexts the prediction represents.

Learning Objectives

- Explain the practical difference between single-chain and interaction prediction.
- Read confidence outputs as decision aids rather than truth labels.
- Separate structure prediction from function, mechanism, and binding validation.

TL;DR

Structure models are now routine inputs to biology, but they are not substitutes for experiments. Confidence, conformational state, ligand geometry, and biological context determine whether a predicted structure supports a downstream decision.

Introduction

AlphaFold 2 established a new baseline for many protein structure prediction tasks (Jumper et al., 2021). ESMFold showed a different route through protein language models (Lin et al., 2023). AlphaFold 3 expanded the prediction target from proteins alone toward biomolecular interactions involving proteins, nucleic acids, small molecules, ions, and modifications (Abramson et al., 2024).

Demonstrated

Demonstrated capability includes high-quality prediction for many protein domains and useful interaction modeling for selected biomolecular systems. AlphaFold DB made predicted structures available at large scale (AlphaFold Protein Structure Database, 2026). AlphaFold 3 demonstrated improved interaction prediction across several molecular classes in the published evaluation (Abramson et al., 2024). Boltz-1 demonstrated the current pressure for open, reproducible interaction modeling systems (Boltz-1, 2024).

Evidence Anchor	What It Supports	Practical Constraint
AlphaFold 2	High-accuracy structure prediction for many proteins	Disorder, conformational ensembles, and ligands need caution
AlphaFold 3	Interaction prediction across biomolecular classes	Access, training overlap, and chemical validity remain review points
ESMFold	Fast sequence-only structure prediction	Speed does not remove accuracy checks

Theoretical

Theoretical capability includes routine modeling of dynamic complexes, alternate conformations, and condition-specific structural states. These are active research directions, but a single predicted structure often represents one plausible state rather than an ensemble.

Beyond Current Capabilities

Beyond current capabilities includes direct inference of biological mechanism from a predicted structure alone. Mechanism needs kinetics, thermodynamics, localization, expression, regulation, and perturbation evidence.

Practice Notes

- Check confidence metrics and domain-level uncertainty before using a model.
- Compare predicted structures against experimental homologs when available.
- Treat low-confidence regions as hypotheses about disorder or missing context, not as solved geometry.
- Use docking, mutagenesis, binding assays, and orthogonal structure methods for consequential decisions.

Protein Design and Engineering

Protein design turns the structure problem around. Instead of asking what shape a sequence adopts, the design problem asks which sequence, scaffold, or assembly satisfies a target constraint.

Learning Objectives

- Distinguish backbone generation, inverse folding, and function design.
- Use experimental validation as the boundary between design and speculation.
- Recognize why design tasks differ from natural structure prediction.

TL;DR

Protein design is strongest when the target is structurally specified and the success assay is direct. Claims become weaker as design moves from fold, to binding, to catalysis, to cellular phenotype.

Introduction

RFdiffusion and ProteinMPNN are central tools in the current design stack. RFdiffusion generates protein backbones under structural constraints (Watson et al., 2023). ProteinMPNN designs amino acid sequences for target backbones (Dauparas et al., 2022). ESM3 added a multimodal protein language model route to sequence, structure, and function-conditioned generation (Hayes et al., 2025).

Demonstrated

Demonstrated capability includes de novo protein backbone design, sequence design for specified structures, and experimental success in selected binder and fold design tasks. RFdiffusion demonstrated experimentally tested designs across several categories (Watson et al., 2023). ProteinMPNN demonstrated strong sequence recovery and design utility for fixed backbones (Dauparas et al., 2022). ESM3 demonstrated generation and protein synthesis of esmGFP in the Science report (Hayes et al., 2025).

Evidence Anchor	What It Supports	Practical Constraint
RFdiffusion	Backbone generation for design tasks	Function remains assay-dependent
ProteinMPNN	Sequence design for specified structures	Designed sequence quality depends on backbone realism
ESM3	Multimodal protein generation with experimental protein synthesis example	General design reliability varies by target function

Theoretical

Theoretical capability includes routine design of enzymes, switches, and therapeutic proteins from natural-language specifications. Existing workflows still require explicit constraints, expert review, and wet-lab iteration.

Beyond Current Capabilities

Beyond current capabilities includes general function design where a text prompt produces a safe, manufacturable, active biologic without assay cycles. Protein function is context-dependent and rarely reducible to static structure.

Practice Notes

- Define success as an assay result, not a confidence score.
- Check novelty, developability, immunogenicity signals, aggregation risk, and manufacturability.
- Use negative design constraints when off-target binding or aggregation matters.
- Document all sequence filters before synthesis orders or expression work.

Antibody and Biologic Design

Antibody and biologic design sits where structure, sequence, immunology, and manufacturing meet. A designed binder is useful only if affinity, specificity, expression, stability, safety, and developability survive the same program.

Learning Objectives

- Separate binder design from therapeutic biologic development.
- Identify developability constraints that structure models do not settle.
- Use assay cascades for affinity, specificity, and manufacturability.

TL;DR

AI design methods help generate and prioritize binders. Therapeutic biologics still require assay cascades, liability screening, cell-based testing, and manufacturing review.

Introduction

Protein design systems now support binder and scaffold design, but antibodies add constraints from immune repertoires, CDR loop geometry, epitope context, glycosylation, Fc behavior, and developability. RFdiffusion and AlphaFold 3 are relevant to this space because they address structure and interaction modeling (Watson et al., 2023; Abramson et al., 2024).

Demonstrated

Demonstrated capability includes structure-guided binder design, antibody structure modeling, and sequence optimization workflows for selected targets. RFdiffusion demonstrated protein binder design tasks with experimental follow-up (Watson et al., 2023). AlphaFold 3 demonstrated biomolecular interaction prediction that includes protein complexes relevant to binder assessment (Abramson et al., 2024).

Evidence Anchor	What It Supports	Practical Constraint
RFdiffusion	Designed binders and constrained protein generation	Therapeutic behavior requires downstream assays
AlphaFold 3	Complex prediction for biomolecular interactions	Binding affinity and function are not guaranteed by a pose
ProteinMPNN	Sequence design around backbones	Developability filters remain external

Theoretical

Theoretical capability includes antibody libraries designed around target epitopes with predictable affinity maturation and low developability risk. That goal is plausible for constrained domains, but target biology and manufacturability still determine program success.

Beyond Current Capabilities

Beyond current capabilities includes end-to-end biologic development from antigen sequence to clinical candidate without immunological, pharmacological, and manufacturing testing. No model removes the need for those gates.

Practice Notes

- Keep epitope, paratope, isotype, format, and intended mechanism separate.
- Screen for aggregation, viscosity, expression, liabilities, immunogenicity, and polyspecificity.
- Use orthogonal assays when structural models disagree with binding data.
- Treat high-affinity designs without specificity data as incomplete.

Nucleic Acid and Genome Models

Genome models move life sciences AI from protein sequence toward regulatory sequence, RNA, genome organization, and cellular context. The unit of modeling is no longer only a protein product.

Learning Objectives

- Distinguish coding-sequence models from regulatory-sequence models.
- Explain the importance of context length and assay tracks.
- Read genome-model claims through organism and cell-type coverage.

TL;DR

DNA and RNA models are strongest when the output is tied to measured functional genomic assays. Variant interpretation remains difficult when disease mechanism, cell context, and long-range regulation are uncertain.

Introduction

Evo was presented as a biological foundation model operating from molecular to genome scale (Nguyen et al., 2024). AlphaGenome focuses on regulatory variant-effect prediction from long DNA sequence context (Avsec et al., 2026). These systems differ from protein language models because regulatory function depends on cell type, chromatin context, and measurement modality.

Demonstrated

Demonstrated capability includes sequence-to-function prediction for specific genomic tracks and zero-shot or few-shot transfer for selected molecular tasks. Evo demonstrated modeling across DNA, RNA, and proteins in the Science report indexed by PubMed (Nguyen et al., 2024). AlphaGenome demonstrated regulatory variant-effect prediction using megabase-scale DNA sequence inputs in Nature (Avsec et al., 2026).

Evidence Anchor	What It Supports	Practical Constraint
Evo	Genome-scale sequence modeling across biological modalities	Training domain and organism coverage define use
AlphaGenome	Regulatory variant-effect prediction from long sequence	Human and mouse training context does not equal all biology
Bridge2AI	AI-ready genomic data as an infrastructure need	Ethics and metadata remain part of model validity

Theoretical

Theoretical capability includes genome editing design that forecasts regulatory, transcriptomic, proteomic, and phenotypic outcomes before experiments. The causal path from sequence to phenotype remains too context-rich for routine certainty.

Beyond Current Capabilities

Beyond current capabilities includes whole-organism phenotype prediction from raw genome sequence alone. Development, environment, epigenetics, microbiome, and measurement context prevent that claim.

Practice Notes

- Track organism, genome build, cell type, assay, and context window.
- Separate coding variant interpretation from noncoding regulatory interpretation.
- Use perturbational validation for proposed regulatory edits.
- Keep RNA structure, RNA expression, and RNA therapeutic design as related but distinct tasks.

Variant Effect Prediction

Variant effect prediction asks whether a sequence change alters molecular function, cellular state, or disease risk. The model output is not a diagnosis unless the clinical and biological context also supports that interpretation.

Learning Objectives

- Separate molecular effect, cellular effect, and clinical significance.
- Compare protein and regulatory variant prediction tasks.
- Identify when experimental evidence is required before interpretation.

TL;DR

Variant models help prioritize variants and hypotheses. They do not replace segregation evidence, functional assays, population frequency, disease mechanism, and clinical interpretation.

Introduction

Protein language models, genome models, and functional genomics models now support different variant-effect tasks. AlphaGenome targets regulatory variant effects across genomic signals (Avsec et al., 2026). ESM-family protein models support protein sequence representations relevant to missense variant analysis (Lin et al., 2023).

Demonstrated

Demonstrated capability includes ranking missense and regulatory variants for selected molecular readouts. AlphaGenome demonstrated improved regulatory variant-effect prediction across evaluated tasks in the Nature report (Avsec et al., 2026). Geneformer demonstrated that pretrained gene representations could support gene network predictions in selected settings (Theodoris et al., 2023).

Evidence Anchor	What It Supports	Practical Constraint
AlphaGenome	Regulatory sequence-to-function and variant scoring	Cell type and assay coverage constrain interpretation
Protein language models	Protein sequence representation for missense effects	Clinical classification needs additional evidence
Functional genomics resources	Measured assay tracks for variant interpretation	Assay signal is not the same as disease causality

Theoretical

Theoretical capability includes joint models that connect variant, molecular effect, cellular response, tissue pathology, and patient phenotype. These are plausible research targets when linked to high-quality perturbational and clinical data.

Beyond Current Capabilities

Beyond current capabilities includes clinical-grade interpretation for arbitrary variants without population, family, functional, and phenotype data. Variant interpretation remains evidence integration, not model output transcription.

Practice Notes

- Label the endpoint: molecular activity, expression, splicing, binding, cell state, or clinical classification.
- Avoid treating one model score as decisive evidence.
- Check ancestry, ascertainment, and population frequency limitations.
- Use calibrated thresholds only when validated for the intended disease and assay context.

Part III: Therapeutics AI

Target Identification and Prioritization

Target identification is a decision under uncertainty. AI systems help organize evidence, but the scientific risk is confusing association with therapeutic tractability.

Learning Objectives

- Separate target association, causal support, tractability, and safety.
- Use genetics and functional genomics as evidence layers.
- Evaluate knowledge graph outputs as prioritization, not proof.

TL;DR

AI-assisted target selection is useful when it integrates evidence transparently. The winning target is not the top-ranked node. It is the target with a testable mechanism, feasible modality, safety rationale, and disease-relevant assay path.

Introduction

The Open Targets Platform integrates genetics, genomics, chemistry, literature, and drug evidence to support systematic target-disease prioritization (Ochoa et al., 2021). AI adds ranking and representation learning, but the underlying problem remains biological evidence integration.

Demonstrated

Demonstrated capability includes evidence aggregation, target-disease scoring, literature mining, and genetics-informed prioritization. Open Targets provides a transparent public example of systematic target prioritization infrastructure (Ochoa et al., 2021). FDA also recognizes AI and machine learning use across drug development submissions and discussion papers (FDA, 2026).

Evidence Anchor	What It Supports	Practical Constraint
Open Targets	Integrated target-disease evidence	Scores require biological interpretation
FDA drug AI materials	Regulatory attention to AI in drug development	Regulatory acceptance depends on context of use
Geneformer	Gene-network modeling from single-cell data	Network prediction is not target validation

Theoretical

Theoretical capability includes target selection models that forecast efficacy, toxicity, patient subgroup, and modality fit before a program begins. The causal and translational links remain difficult, especially when disease biology is heterogeneous.

Beyond Current Capabilities

Beyond current capabilities includes reliable target discovery without wet-lab or human genetics validation. Computational ranking alone does not establish disease causality or therapeutic window.

Practice Notes

- Separate evidence for disease association from evidence for intervention.
- Require a modality hypothesis before program launch.
- Use orthogonal evidence layers rather than one graph score.
- Record why lower-ranked targets were rejected.

Small Molecule Generation and ADMET

Small molecule AI sits between chemical imagination and experimental attrition. Generating structures is easy compared with generating useful, selective, soluble, safe, and synthesizable compounds.

Learning Objectives

- Read molecular generation claims through medicinal chemistry constraints.
- Use ADMET and physical plausibility checks early.
- Distinguish benchmark gains from lead optimization value.

TL;DR

The useful output is not a molecule that looks novel. The useful output is a prioritized set of compounds with rationale, feasibility, assay plan, and acceptable risk across potency, selectivity, ADMET, and chemistry.

Introduction

MoleculeNet remains a reference point for molecular machine learning benchmarks (Wu et al., 2018). ChEMBL and PubChem provide major public chemical and bioactivity resources (Zdrazil et al., 2024; Kim et al., 2023). PoseBusters shows why geometric or score-based docking success requires physical plausibility checks (Buttenschoen et al., 2024).

Demonstrated

Demonstrated capability includes property prediction, virtual screening support, molecular representation learning, and generative chemistry under constraints. MoleculeNet demonstrated standardized benchmark tasks for molecular machine learning (Wu et al., 2018). PoseBusters demonstrated that AI docking methods need validity checks beyond RMSD (Buttenschoen et al., 2024).

Evidence Anchor	What It Supports	Practical Constraint
MoleculeNet	Benchmarking molecular property models	Benchmark datasets are not medicinal chemistry programs
ChEMBL and PubChem	Chemical and bioactivity data sources	Assay context and duplicates require curation
PoseBusters	Docking plausibility checks	Physical validity matters beside RMSD

Theoretical

Theoretical capability includes multi-objective compound design that jointly optimizes potency, selectivity, solubility, permeability, metabolism, toxicity, and synthetic route. Current workflows approximate this with staged filters and expert review.

Beyond Current Capabilities

Beyond current capabilities includes one-shot generation of clinical candidates from target name alone. Biology, chemistry, formulation, toxicology, and clinical pharmacology remain program-level work.

Practice Notes

- Use scaffold splits and time splits for virtual screening evaluation.
- Review synthetic feasibility before celebrating novelty.
- Track assay provenance and units when merging activity data.
- Keep medicinal chemistry review inside the loop for design decisions.

mRNA, RNA, and Vaccine Design

RNA and vaccine design combine sequence, structure, immunology, delivery, manufacturing, and population biology. AI helps with parts of that stack, not the whole stack at once.

Learning Objectives

- Separate antigen selection, RNA sequence design, delivery, and immune evaluation.
- Identify the experimental gates for RNA and vaccine programs.
- Apply the three-tier framework to claims about design speed.

TL;DR

RNA and vaccine AI is strongest when the model output is tied to a measurable endpoint: expression, stability, antigenicity, manufacturability, or immune response. Program success still depends on delivery, dosing, safety, and clinical evidence.

Introduction

Genome and protein models matter for vaccine and RNA design because they support antigen analysis, protein design, variant interpretation, and sequence optimization. AlphaFold 3 addresses biomolecular interaction prediction (Abramson et al., 2024). Evo and AlphaGenome illustrate the move toward long-context sequence models (Nguyen et al., 2024; Avsec et al., 2026).

Demonstrated

Demonstrated capability includes protein antigen modeling, epitope-informed design support, and sequence-to-function modeling for selected genomic and molecular outputs. AlphaFold 3 demonstrated interaction modeling relevant to antigen-antibody and protein-nucleic acid questions (Abramson et al., 2024). Evo demonstrated sequence modeling across DNA, RNA, and proteins in the indexed Science report (Nguyen et al., 2024).

Evidence Anchor	What It Supports	Practical Constraint
AlphaFold 3	Biomolecular interaction structure prediction	Immune response is not settled by structure alone
Evo	Long-context biological sequence modeling	RNA therapeutic behavior requires delivery and assay data
FDA drug AI materials	Regulatory attention to AI in drug and biologic development	Context of use drives evidentiary needs

Theoretical

Theoretical capability includes integrated vaccine design systems that jointly model antigen structure, immune escape, expression, delivery, and population-level strain coverage. Such systems require data across immunology, manufacturing, and clinical outcomes.

Beyond Current Capabilities

Beyond current capabilities includes reliable vaccine design from sequence surveillance alone. Immunogenicity, durability, safety, delivery, and real-world effectiveness require experiments and trials.

Practice Notes

- Keep antigen modeling separate from immune-response prediction.
- Validate expression, stability, and formulation before immunogenicity claims.
- Use neutralization, cellular immunity, and safety assays as separate evidence layers.
- Avoid fixed timeline claims unless a specific program supplies verified dates.

Clinical Trial AI for Translational Research

Clinical trial AI belongs in a life sciences handbook because discovery programs fail at the translation step as often as they fail at the molecular step. Model-guided trials still need prespecified protocols, audit trails, and regulatory clarity.

Learning Objectives

- Distinguish trial operations AI from inferential AI.
- Identify where AI affects validity, not only efficiency.
- Use regulatory context when AI affects drug evidence.

TL;DR

AI in trials is safest when the context of use is explicit. Recruitment support, site selection, endpoint extraction, enrichment, and synthetic controls carry different evidentiary and regulatory burdens.

Introduction

FDA and EMA have both published materials addressing AI and machine learning across medicinal product development (FDA, 2026; EMA, 2024). The translation lesson is simple: an AI tool used to generate evidence for a product must be governed as part of the evidence-generation system.

Demonstrated

Demonstrated capability includes operational analytics, eligibility matching, real-world data curation, endpoint extraction, and risk-based monitoring support. FDA reports experience with submissions containing AI components from 2016 to 2023 in drug-development contexts (FDA, 2026). EMA's adopted reflection paper covers AI across the medicinal product lifecycle (EMA, 2024).

Evidence Anchor	What It Supports	Practical Constraint
FDA drug AI page	Agency experience and guidance activity	Specific context of use determines evidence expectations
EMA reflection paper	Lifecycle framing for AI in medicines	Applicants remain responsible for validity and compliance
Bridge2AI	AI-ready biomedical data requirements	Trial data quality and provenance remain central

Theoretical

Theoretical capability includes adaptive trial systems that continuously update enrichment and operational strategy while preserving valid inference. That design requires prespecification, simulation, monitoring, and regulator engagement.

Beyond Current Capabilities

Beyond current capabilities includes replacing randomized evidence with model outputs for most therapeutic claims. Models support evidence generation. They do not remove the need for credible causal inference.

Practice Notes

- Write the AI context of use before selecting metrics.
- Separate operational endpoints from efficacy endpoints.
- Prespecify model updates, monitoring, and audit logs.
- Use regulator-facing documentation when model output affects trial evidence.

Translational Evidence and Failure Modes

The failure pattern in therapeutic AI is not only model error. It is often the mismatch between model endpoint, biological mechanism, assay system, and program decision.

Learning Objectives

- Identify failure modes across the discovery-to-development path.
- Separate model validation from program validation.
- Use negative evidence as a design input.

TL;DR

A model that improves a proxy endpoint may still harm the program if the proxy is poorly linked to disease biology or developability. Failure analysis belongs near the start of the workflow, not after candidate nomination.

Introduction

PoseBusters is a useful example outside its narrow docking domain because it shows how a familiar metric can hide invalid outputs (Buttenschoen et al., 2024). The same pattern appears in target selection, small molecules, biologics, and trials: model success depends on the decision that follows the score.

Demonstrated

Demonstrated capability includes identifying model failures through stricter benchmark design and physical validity checks. PoseBusters demonstrated that docking evaluations need chemical plausibility in addition to RMSD (Buttenschoen et al., 2024). MoleculeNet demonstrated the value and limits of shared molecular benchmarks (Wu et al., 2018).

Evidence Anchor	What It Supports	Practical Constraint
PoseBusters	Docking failure detection	Metric choice changes conclusions
MoleculeNet	Shared molecular benchmark tasks	Program-level value needs external evidence
FDA and EMA materials	Regulatory attention to AI lifecycle risks	Documentation and accountability are expected

Theoretical

Theoretical capability includes AI systems that forecast full program attrition risk across target, chemistry, biology, toxicology, trial execution, and market access. Existing data fragmentation makes this an evidence integration problem rather than a model-size problem.

Beyond Current Capabilities

Beyond current capabilities includes reliable prediction of clinical success for early discovery assets without prospective evidence. Program success depends on human biology, trial design, safety, adherence, and effect size.

Practice Notes

- Map each model endpoint to the next experimental decision.
- Keep failed compounds, failed assays, and failed targets in the learning set.
- Use decision curves and cost-aware evaluation when experiments are expensive.
- Require an explicit stop rule for model-guided programs.

Part IV: Cellular and Systems Biology

Single-Cell Foundation Models

Single-cell foundation models learn representations of cells and genes from large expression atlases. Their value depends on transfer to the cell types, perturbations, and assays that matter for a biological decision.

Learning Objectives

- Compare single-cell foundation model tasks.
- Identify leakage and batch effects in single-cell evaluation.
- Use cell context when reading perturbation claims.

TL;DR

Single-cell foundation models are useful representation systems, not general virtual cells. Evaluation must account for cell type, donor, batch, disease state, and perturbation split.

Introduction

Geneformer pretrained on large collections of single-cell gene-expression profiles to support gene-network tasks (Theodoris et al., 2023). scGPT pretrained across tens of millions of cells and evaluated on cell annotation, perturbation, integration, and related tasks (Cui et al., 2024).

Demonstrated

Demonstrated capability includes representation learning for cell type annotation, batch integration, perturbation-related tasks, and gene network analysis in benchmark settings. Geneformer demonstrated transfer under limited-data scenarios (Theodoris et al., 2023). scGPT demonstrated generative pretraining for single-cell multi-omics tasks (Cui et al., 2024).

Evidence Anchor	What It Supports	Practical Constraint
Geneformer	Gene and cell representation learning	Transfer depends on biological and dataset proximity
scGPT	Single-cell multi-omics pretraining	Evaluation split determines credibility
GEARS	Perturbation prediction	Graph priors and context affect generalization

Theoretical

Theoretical capability includes cell-state models that forecast response to novel perturbations across donors, tissues, and disease states. This needs perturbational data and validation beyond atlas pretraining.

Beyond Current Capabilities

Beyond current capabilities includes a general virtual cell that reliably predicts full cellular behavior across all contexts. Current systems usually model measured expression outputs, not all cellular mechanisms.

Practice Notes

- Use donor, batch, tissue, and time splits where relevant.
- Benchmark against simple gene-level and cell-type baselines.
- Audit whether target genes or similar perturbations were present in training.
- Report uncertainty and failure cases for rare cell populations.

Spatial Omics and Tissue Models

Spatial omics adds tissue position to molecular measurement. The modeling problem changes because cell state, neighborhood, morphology, and tissue architecture now influence interpretation.

Learning Objectives

- Explain why spatial context changes single-cell interpretation.
- Separate image alignment, cell typing, neighborhood analysis, and tissue modeling.
- Identify where spatial resolution and assay sensitivity limit claims.

TL;DR

Spatial AI is most useful when it links molecular signals to tissue structure with clear resolution limits. It is not enough to assign labels to spots or cells. The biological question is whether spatial organization changes mechanism or decision.

Introduction

Spatial transcriptomics and single-cell technologies have produced datasets that connect cell identity and tissue architecture. A Nature Reviews Molecular Cell Biology review describes advances and challenges in characterizing cell states and multicellular neighborhoods, including deep learning methods (Rao et al., 2024).

Demonstrated

Demonstrated capability includes cell-state mapping, neighborhood analysis, tissue segmentation support, and multimodal alignment between histology and gene expression in bounded settings. The spatial transcriptomics review documents the increasing role of deep learning in single-cell and spatial data analysis (Rao et al., 2024).

Evidence Anchor	What It Supports	Practical Constraint
Spatial transcriptomics review	Cell identity and tissue architecture as linked data	Resolution, sensitivity, and platform differences constrain inference
Cell Painting	Image-based phenotyping as cellular morphology data	Morphology is a proxy requiring biological validation
Single-cell models	Cell-state representation learning	Dissociated cells lose native spatial context

Theoretical

Theoretical capability includes tissue foundation models that represent histology, spatial transcriptomics, proteomics, and perturbations together. This is plausible, but dataset harmonization, resolution mismatch, and tissue processing artifacts remain hard.

Beyond Current Capabilities

Beyond current capabilities includes reliable simulation of tissue response to arbitrary perturbations from static spatial data alone. Dynamic experiments and perturbational measurements are still required.

Practice Notes

- State the spatial resolution and whether measurements are spot-level, cell-level, or subcellular.
- Keep histology-derived labels separate from molecular labels.
- Validate inferred neighborhoods with biological markers.
- Avoid comparing platforms without platform-aware normalization.

Cell Painting and Image-Based Phenotyping

Image-based phenotyping turns cells into high-dimensional morphology profiles. The central promise is that perturbations with similar biological effects may produce related image signatures.

Learning Objectives

- Explain Cell Painting as a morphology profiling assay.
- Separate image features from mechanism claims.
- Use controls and batch correction in image-based screens.

TL;DR

Cell images are rich biological measurements, but morphology is not mechanism by itself. High-content imaging requires careful controls, segmentation quality checks, and orthogonal validation.

Introduction

The Cell Painting assay uses multiplexed fluorescent dyes for high-content morphological profiling (Bray et al., 2016). AI methods extend this work through segmentation, representation learning, perturbation matching, and phenotype clustering.

Demonstrated

Demonstrated capability includes automated image analysis, feature extraction, perturbation profiling, and compound or gene clustering by morphology. Cell Painting provides a standardized assay protocol for morphological profiling (Bray et al., 2016).

Evidence Anchor	What It Supports	Practical Constraint
Cell Painting	High-content morphology profiling	Assay design and segmentation quality determine signal
ChEMBL and PubChem	Chemical metadata for compound screens	Chemical identity and batch history require curation
MoleculeNet	Molecular benchmark context	Images and structures need different validation

Theoretical

Theoretical capability includes image foundation models that infer mechanism of action and toxicity across cell types. This goal requires perturbation labels, pathway data, dose-response structure, and external assays.

Beyond Current Capabilities

Beyond current capabilities includes complete mechanism inference from a single microscopy image. Morphology supplies evidence, not a full causal map.

Practice Notes

- Track plate, batch, dose, time, stain, microscope, and segmentation model.
- Use positive and negative controls on every batch.
- Evaluate replicate consistency before biological interpretation.
- Confirm mechanism hypotheses with orthogonal assays.

Perturbation Prediction and Virtual Cells

Perturbation prediction is the most direct route from representation learning to biological action. The model must answer what changes after an intervention, not only what a cell resembles.

Learning Objectives

- Distinguish observational cell embeddings from perturbational models.
- Identify realistic endpoints for virtual cell claims.
- Use causal language only when the design supports it.

TL;DR

Virtual cell work is promising when framed as perturbation prediction for defined outputs. It becomes misleading when a transcriptomic forecast is treated as a full model of the cell.

Introduction

GEARS uses graph-enhanced modeling to predict transcriptional outcomes of novel multigene perturbations in selected settings (Roohani et al., 2024). The Arc Institute Virtual Cell Challenge illustrates the field's move toward public benchmarks for perturbation response prediction (Arc Institute, 2025).

Demonstrated

Demonstrated capability includes transcriptional perturbation prediction for selected genes, cell contexts, and benchmark designs. GEARS demonstrated prediction of some novel multigene perturbation outcomes (Roohani et al., 2024). Public challenge efforts demonstrate growing demand for independent evaluation of virtual cell models (Arc Institute, 2025).

Evidence Anchor	What It Supports	Practical Constraint
GEARS	Multigene perturbation outcome prediction	Generalization depends on graph priors and cell context
Virtual Cell Challenge	Public benchmark framing for cell response prediction	A challenge metric is not a full cell model
scGPT	Single-cell foundation model baseline	Pretraining alone does not establish causal validity

Theoretical

Theoretical capability includes models that rank interventions before CRISPR, drug, or combination screens. This is plausible for defined cell systems and measured outputs, especially when active learning adds new experiments.

Beyond Current Capabilities

Beyond current capabilities includes a complete executable cell model that forecasts all molecular, phenotypic, and temporal consequences of arbitrary interventions. Current data cover slices of cellular behavior.

Practice Notes

- Name the perturbation type: knockout, knockdown, activation, compound, dose, timing, or combination.
- Name the output: expression, morphology, viability, secretion, or functional assay.
- Use held-out perturbations and held-out cell contexts separately.
- Avoid causal claims from observational pretraining alone.

Microbiome and Multi-Omics AI

Multi-omics AI tries to combine biological layers that were often measured separately. The challenge is not only more data. It is alignment across molecules, cells, organisms, time, and environment.

Learning Objectives

- Separate multi-omics integration from causal mechanism.
- Identify microbiome-specific data constraints.
- Use modality-aware validation for integrated models.

TL;DR

Multi-omics models are useful when each modality has clear provenance and the validation endpoint is explicit. Integration can hide weak measurements if the workflow does not track missingness and batch effects.

Introduction

Bridge2AI frames AI-ready biomedical data as a cross-disciplinary infrastructure problem rather than a single dataset problem (NIH Common Fund Bridge2AI, 2026). That framing fits multi-omics work because each modality brings its own measurement noise, preprocessing choices, and biological interpretation.

Demonstrated

Demonstrated capability includes modality-specific representation learning, cross-modal alignment in selected datasets, metagenomic protein structure prediction resources, and integrated analysis pipelines. The ESM Metagenomic Atlas demonstrates structure prediction resources for metagenomic protein space (ESM Metagenomic Atlas, 2026).

Evidence Anchor	What It Supports	Practical Constraint
Bridge2AI	AI-ready data and workforce infrastructure	Ethics, metadata, and standards affect integration
ESM Metagenomic Atlas	Predicted structures for metagenomic proteins	Predicted structure does not imply organismal function
PubChem and ChEMBL	Chemical and bioactivity layers	Molecule data need assay context

Theoretical

Theoretical capability includes disease models that combine microbiome, metabolome, proteome, transcriptome, imaging, and clinical phenotypes. This requires alignment across time, sampling conditions, measurement platforms, and causal hypotheses.

Beyond Current Capabilities

Beyond current capabilities includes general health prediction from a single multi-omics snapshot. Biological state changes over time, and many omics associations are not causal.

Practice Notes

- Track sample handling, extraction, sequencing, and batch effects.
- Model missingness instead of silently dropping incomplete subjects.
- Validate integrated signals against modality-specific controls.
- Use longitudinal designs when claims involve dynamics.

Part V: Engineering and Automation

Self-Driving Laboratories

Self-driving laboratories combine experimental hardware, data systems, and model-guided experiment selection. Their value lies in faster learning cycles, not in removing scientific judgment.

Learning Objectives

- Define closed-loop experimentation in practical terms.
- Separate automation, active learning, and scientific inference.
- Identify the reproducibility requirements for self-driving labs.

TL;DR

A self-driving lab is an experimental system with a model in the loop. It needs reliable instruments, machine-readable protocols, calibration, error handling, and human review of objectives and stopping rules.

Introduction

Self-driving labs are machine-learning-assisted modular experimental platforms that iteratively select experiments to reach a user-defined objective, as described in Nature Synthesis (Abolhasani and Kumacheva, 2023). The mobile robotic chemist demonstrated autonomous experimental execution in a physical laboratory setting (Burger et al., 2020).

Demonstrated

Demonstrated capability includes closed-loop optimization in chemistry and materials settings, robotic execution of defined workflows, and distributed self-driving laboratory coordination. Burger and colleagues demonstrated a mobile robotic chemist (Burger et al., 2020). A dynamic knowledge graph approach demonstrated distributed self-driving laboratory coordination (Mehr et al., 2024).

Evidence Anchor	What It Supports	Practical Constraint
Mobile robotic chemist	Physical robotic execution and closed-loop search	Workflow scope is bounded by equipment and protocols
Nature Synthesis review	SDL definition and field overview	Chemistry and materials examples do not automatically transfer to biology
Distributed SDL	Knowledge graph coordination	Interoperability and protocol quality remain limiting

Theoretical

Theoretical capability includes biological self-driving labs that optimize cell engineering, protein expression, assay conditions, and perturbation screens under shared protocol standards. The idea is plausible where experiments are modular and measurements are reliable.

Beyond Current Capabilities

Beyond current capabilities includes autonomous laboratories that choose biomedical objectives, perform arbitrary experiments, and establish disease mechanisms without human governance. Objectives, constraints, and interpretation remain human responsibilities.

Practice Notes

- Version protocols, instruments, reagents, and model policies.
- Define objective functions with safety and cost constraints.
- Record failed runs and hardware faults as data.
- Use human review gates before experiments with safety, animal, human-subjects, or dual-use implications.

Robotic Lab Automation and Cloud Labs

Laboratory automation is the machinery beneath closed-loop discovery. The important design question is where human intent, protocol representation, instrument control, and data capture meet.

Learning Objectives

- Distinguish robotic execution from autonomous science.
- Identify protocol and metadata requirements for reproducibility.
- Use audit logs as part of experimental evidence.

TL;DR

Automation improves repeatability only when protocols, reagents, instruments, and data capture are explicit. A robot executing a vague protocol only scales ambiguity.

Introduction

ARPA-H's IGoR program explicitly names laboratory automation, robotics, protocol standardization, distributed systems, and agentic systems as part of a modern biomedical research ecosystem (ARPA-H IGoR, 2026). This confirms automation as a research infrastructure topic, not only a convenience layer.

Demonstrated

Demonstrated capability includes robotic execution of fixed workflows, cloud-lab style protocol execution, and machine-readable experimental logging. The mobile robotic chemist provides a concrete example of robotic execution within a defined workflow (Burger et al., 2020).

Evidence Anchor	What It Supports	Practical Constraint
ARPA-H IGoR	Biomedical research infrastructure with automation and protocol standards	Program goals still require teams and implementation
Mobile robotic chemist	Physical execution of laboratory tasks	Platform scope follows hardware design
Bridge2AI	Data and metadata practices	Automation without metadata does not create AI-ready data

Theoretical

Theoretical capability includes interoperable cloud-lab marketplaces where validated protocols run across qualified laboratories with comparable outputs. This requires shared protocol languages, instrument calibration, quality systems, and data standards.

Beyond Current Capabilities

Beyond current capabilities includes universal protocol portability across laboratories without local validation. Instruments, reagents, environmental conditions, and operator choices still affect results.

Practice Notes

- Treat protocol code, instrument logs, and reagent lots as experimental records.
- Require calibration and acceptance checks before model-guided runs.
- Separate execution errors from biological negative results.
- Use clear authorization boundaries for any remote biological work.

Synthetic Biology Design Tools

Synthetic biology design tools join sequence design, circuit design, strain engineering, assay planning, and manufacturing constraints. AI adds search and representation power, but biology still imposes context.

Learning Objectives

- Place AI inside design-build-test-learn cycles.
- Distinguish part design, pathway design, and system behavior.
- Use measurement plans before DNA synthesis or organism engineering.

TL;DR

Synthetic biology AI is useful when design output is tied to a build and test plan. Sequence novelty alone is not engineering progress.

Introduction

Evo demonstrated sequence modeling and design from molecular to genome scale, including examples relevant to CRISPR-Cas and transposon systems in the PubMed-indexed report (Nguyen et al., 2024). Protein design systems such as RFdiffusion and ProteinMPNN support a related but narrower layer of biological design (Watson et al., 2023; Dauparas et al., 2022).

Demonstrated

Demonstrated capability includes sequence design support, protein design support, and prioritization in design-build-test-learn loops. Evo demonstrated DNA, RNA, and protein sequence modeling across scales (Nguyen et al., 2024). RFdiffusion demonstrated protein structure and function design tasks (Watson et al., 2023).

Evidence Anchor	What It Supports	Practical Constraint
Evo	Genome-scale sequence modeling and design examples	Organism and assay constraints remain decisive
RFdiffusion and ProteinMPNN	Protein design workflow components	Part performance does not guarantee system behavior
Self-driving labs	Closed-loop optimization model	Measurement quality governs learning

Theoretical

Theoretical capability includes AI-guided design-build-test-learn systems for pathways, circuits, strains, and cell therapies. The field has pieces of this workflow, but system-level prediction remains hard.

Beyond Current Capabilities

Beyond current capabilities includes reliable design of complex living systems from high-level objectives alone. Evolution, regulation, burden, environmental context, and containment make such claims unsupported.

Practice Notes

- Define part, circuit, pathway, chassis, and assay separately.
- Review DNA synthesis, biosafety, and containment requirements before ordering constructs.
- Measure burden, stability, and off-target effects.
- Track every design choice through the build and test cycle.

Agentic Science Workflows

Agentic science workflows use software agents to plan, retrieve, write, reason over tools, and coordinate tasks. In life sciences, the central issue is not whether an agent sounds scientific. The issue is whether it preserves provenance and respects experimental limits.

Learning Objectives

- Define agentic workflows without overstating autonomy.
- Place human review gates around biological actions.
- Use provenance, tool logs, and citations as safety infrastructure.

TL;DR

Agentic systems are useful as research operating layers when tasks are bounded, sources are checked, and lab actions require authorization. They are risky when fluent plans are treated as validated science.

Introduction

ARPA-H IGoR names AI/ML orchestration and agentic systems alongside laboratory automation, protocol standardization, and distributed systems (ARPA-H IGoR, 2026). That framing puts agents inside a governed research infrastructure rather than outside it.

Demonstrated

Demonstrated capability includes literature triage, code execution, data cleaning, protocol draft support, and orchestration of bounded computational workflows. IGoR demonstrates that federal research programs now treat agentic systems as part of biomedical research infrastructure planning (ARPA-H IGoR, 2026).

Evidence Anchor	What It Supports	Practical Constraint
ARPA-H IGoR	Agentic systems in biomedical research infrastructure	Program ambition is not proof of deployed reliability
Bridge2AI	AI-ready data and workforce materials	Agents need high-quality inputs and human review
EMA and FDA materials	Lifecycle accountability for AI in regulated contexts	Regulatory use requires documentation

Theoretical

Theoretical capability includes agents that propose experiments, call analysis tools, update models, and prepare protocol-ready plans. This is plausible for bounded settings with source control, tool permissions, and human approval.

Beyond Current Capabilities

Beyond current capabilities includes unsupervised agents conducting open-ended biological research without human governance. Biological materials, safety controls, privacy, and scientific accountability require explicit human authority.

Practice Notes

- Require source links for literature-derived claims.
- Log tool calls, parameters, data versions, and outputs.
- Use separate permissions for reading, analysis, procurement, and laboratory execution.
- Block autonomous actions involving biological materials unless a human authorizes the exact protocol.

Part VI: Practice and Governance

Benchmarks for Bio AI

Benchmarks are social infrastructure for scientific claims. A good benchmark narrows the space of plausible claims; it does not settle all uses of a model.

Learning Objectives

- Match benchmarks to the biological decision under evaluation.
- Use failure-aware metrics in molecular and cellular tasks.
- Avoid benchmark overfitting and leakage.

TL;DR

Benchmarks matter when they are hard to game, close to the intended decision, and paired with failure analysis. A leaderboard is not a validation plan.

Introduction

CASP, CAMEO, MoleculeNet, and PoseBusters each illustrate a different benchmark role: blinded community assessment, continuous server evaluation, shared molecular datasets, and physical validity checks (Kryshtafovych et al., 2024; CAMEO, 2026; Wu et al., 2018; Buttenschoen et al., 2024).

Demonstrated

Demonstrated capability includes benchmark-driven progress in protein structure prediction and increasingly strict evaluation of molecular docking and generation. CASP15 documents categories beyond single-chain structure, including complexes, RNA, and ligand binding (Kryshtafovych et al., 2024). PoseBusters demonstrated that physically invalid poses can pass simpler docking metrics (Buttenschoen et al., 2024).

Evidence Anchor	What It Supports	Practical Constraint
CASP and CAMEO	Structure prediction assessment	Tasks evolve as methods improve
MoleculeNet	Molecular property benchmark tasks	Dataset splits shape conclusions
PoseBusters	Physical validity in docking evaluation	One metric can hide failure

Theoretical

Theoretical capability includes prospective discovery benchmarks where models choose experiments and are judged by cost-adjusted learning. This is the right direction for many life sciences tasks, but it is more expensive than static benchmark release.

Beyond Current Capabilities

Beyond current capabilities includes a universal biological benchmark that ranks all models. Biological tasks differ too much in ground truth, cost, and acceptable error.

Practice Notes

- Use benchmarks to reject claims, not only to support them.
- Prefer splits that reflect intended use.
- Report failure categories beside average metrics.
- Hold back prospective tests when the field is likely to overfit public leaderboards.

Reproducibility and Open Science

Reproducibility in AI-biology has two linked meanings: computational reproducibility and experimental reproducibility. A notebook that reruns is not enough if the assay cannot be repeated.

Learning Objectives

- Define reproducibility across code, data, model, and experiment.
- Use open science practices without ignoring dual-use and privacy concerns.
- Document enough context for independent review.

TL;DR

Open code and open weights help, but life sciences reproducibility also needs protocol detail, reagent provenance, data versioning, and assay records. Scientific openness and risk management must be handled together.

Introduction

Bridge2AI and IGoR both point toward data and protocol infrastructure as prerequisites for AI-supported science (NIH Common Fund Bridge2AI, 2026; ARPA-H IGoR, 2026). Open models such as Boltz-1 also show the pressure for transparent biomolecular modeling stacks (Boltz-1, 2024).

Demonstrated

Demonstrated capability includes public datasets, open-source model releases, protocol repositories, and reproducible benchmark scripts. AlphaFold DB and the ESM Metagenomic Atlas demonstrate the research value of large public predicted-structure resources (AlphaFold Protein Structure Database, 2026; ESM Metagenomic Atlas, 2026).

Evidence Anchor	What It Supports	Practical Constraint
AlphaFold DB	Public predicted protein structure resource	Predictions require confidence-aware use
Boltz-1	Open biomolecular interaction model ecosystem	Open release does not replace independent validation
Bridge2AI	Standards, tools, and training for AI-ready data	Ethics and quality are reproducibility requirements

Theoretical

Theoretical capability includes reproducible research networks where protocols, agents, data, and instruments share interoperable records. That goal requires common schemas, incentives, and institutional support.

Beyond Current Capabilities

Beyond current capabilities includes fully reproducible biology from computational artifacts alone. Physical experiments require materials, instruments, environments, and local expertise.

Practice Notes

- Release code, model version, data version, and filtering logic when possible.
- Archive protocol details, reagent lots, instrument settings, and failed runs.
- Use model cards and dataset cards for external readers.
- Use access controls when openness conflicts with privacy, contracts, or safety.

Information Hazards in Capability Research

Capability research can create knowledge that helps science and also lowers barriers to misuse. The practical task is responsible research communication, not broad policy writing.

Learning Objectives

- Identify information hazards in capability-focused AI-biology work.
- Use publication review without weakening reproducibility.
- Separate scientific detail needed for review from detail that increases misuse risk.

TL;DR

Responsible capability research keeps enough detail for verification while avoiding unnecessary operational detail that raises misuse risk. The standard is not secrecy by default. The standard is deliberate disclosure.

Introduction

AI-biology work now includes protein design, genome modeling, lab automation, and agentic workflows. ARPA-H IGoR's inclusion of agentic systems and laboratory automation shows why researcher practice needs explicit review gates (ARPA-H IGoR, 2026).

Demonstrated

Demonstrated capability includes models that design proteins, model biomolecular interactions, generate sequences, and help plan experiments in bounded settings. RFdiffusion, AlphaFold 3, Evo, and self-driving laboratory systems supply concrete examples of capability progress (Watson et al., 2023; Abramson et al., 2024; Nguyen et al., 2024; Abolhasani and Kumacheva, 2023).

Evidence Anchor	What It Supports	Practical Constraint
RFdiffusion and Evo	Capability gains in design and sequence modeling	Publication detail should match legitimate verification needs
ARPA-H IGoR	Agentic and automated research infrastructure	Human authorization and protocol standards matter
Bridge2AI	Ethical AI-ready data framing	Data and model release need governance

Theoretical

Theoretical capability includes publication norms that preserve reproducibility while reducing operational misuse. This requires journal policies, institutional review, and field-specific norms rather than ad hoc redaction.

Beyond Current Capabilities

Beyond current capabilities includes perfect separation of beneficial and harmful uses at publication time. Dual-use judgment remains contextual and imperfect.

Practice Notes

- Perform a disclosure review before releasing code, model weights, protocols, or datasets that materially increase biological capability.
- Keep reproducibility artifacts available to trusted reviewers when public release is not appropriate.
- Avoid step-by-step operational details when high-level scientific reporting is sufficient.
- Document the reason for any redaction or staged release.

Workforce, Compute, and Institutional Readiness

Life sciences AI is an institutional capability. Model access matters, but so do data engineering, biological review, wet-lab partnerships, compute governance, and decision accountability.

Learning Objectives

- Identify the team roles needed for credible life sciences AI work.
- Plan compute and data governance around the research question.
- Use procurement and access controls as scientific infrastructure.

TL;DR

The minimum viable team includes biological domain expertise, data engineering, machine learning, experimental validation, and governance. Compute without experimental judgment creates expensive noise.

Introduction

Bridge2AI explicitly includes workforce development, training materials, standards, and best practices as part of the AI-ready biomedical research agenda (NIH Common Fund Bridge2AI, 2026). FDA and EMA materials reinforce that regulated uses of AI require documentation and accountability (FDA, 2026; EMA, 2024).

Demonstrated

Demonstrated capability includes institutional programs that fund AI-ready data, training, and research infrastructure. Bridge2AI provides a current NIH example (NIH Common Fund Bridge2AI, 2026). ARPA-H IGoR provides a current example focused on protocolized, AI-supported biomedical research infrastructure (ARPA-H IGoR, 2026).

Evidence Anchor	What It Supports	Practical Constraint
Bridge2AI	Workforce and AI-ready data resources	Training and governance are part of deployment
ARPA-H IGoR	Research infrastructure and agentic systems	Large goals need operational standards
FDA and EMA	Regulatory lifecycle expectations	Documentation depends on context of use

Theoretical

Theoretical capability includes shared institutional platforms that let biologists run validated AI workflows without becoming infrastructure engineers. The barrier is not only software. It is governance, support, and trust.

Beyond Current Capabilities

Beyond current capabilities includes replacing cross-disciplinary teams with a single general model. Biology, engineering, regulation, and experimental work still require distinct expertise.

Practice Notes

- Assign owners for data, model, experiment, regulatory, and publication decisions.
- Budget compute together with storage, curation, and validation experiments.
- Set access controls for datasets, model weights, lab tools, and external APIs.
- Review institutional readiness before adopting high-cost platforms.

Consulting & Advisory

Bryan Tegomoh, MD, MPH advises research, biotechnology, health technology, and life sciences teams evaluating AI systems for biomedical discovery and translational research. Engagements are evidence-centered and non-promotional, with emphasis on model evaluation, biological validity, reproducibility, and responsible research practice.

Advisory Scope

- Model and dataset review for protein design, genomics, single-cell, imaging, or therapeutic discovery workflows.
- Evidence review for AI claims in biomedical discovery or biotechnology.
- Benchmark design and validation strategy for life sciences AI systems.
- Research governance, reproducibility, and information-hazard review.

Advisory work is independent. It does not imply endorsement of a product, organization, or public claim.

For advisory, speaking, or workshop inquiries, email bryan.tegomoh@gmail.com with the organization, decision context, and relevant materials.

The Physician AI Handbook | The Public Health AI Handbook | The Biosecurity Handbook | The Life Sciences AI Handbook

References

- Abolhasani and Kumacheva, 2023
- Abramson et al., 2024
- AlphaFold Protein Structure Database, 2026
- Arc Institute, 2025
- ARPA-H IGoR, 2026
- Avsec et al., 2026
- Boltz-1, 2024
- Bray et al., 2016
- Burger et al., 2020
- Buttenschoen et al., 2024
- CAMEO, 2026
- Cui et al., 2024
- Dauparas et al., 2022
- EMA, 2024
- ESM Metagenomic Atlas, 2026
- FDA, 2026
- Hayes et al., 2025
- Jumper et al., 2021
- Kim et al., 2023
- Kryshtafovych et al., 2024
- Lin et al., 2023
- Mehr et al., 2024
- Nguyen et al., 2024
- NIH Bridge2AI, 2022
- NIH Common Fund Bridge2AI, 2026
- Ochoa et al., 2021
- Rao et al., 2024
- RCSB PDB, 2026
- Roohani et al., 2024
- Theodoris et al., 2023
- Watson et al., 2023
- Wu et al., 2018
- wwPDB, 2026
- Zdrazil et al., 2024

TL;DR Compilation

AI for the Life Sciences

Life sciences AI is not one field. It is a set of modeling practices that share biological data constraints, experimental validation requirements, and high error costs. The first rule is to ask what biological object the model represents and what experiment would falsify the output.

Biological Data Infrastructure

Better models do not rescue poorly specified biological data. AI-ready data require provenance, assay context, versioning, licensing, and negative controls. The most useful model card is often a dataset card.

Foundation Models for Biology

A foundation model is useful when pretraining improves a downstream biological task under realistic validation. Model size, modality count, and dataset volume matter less than task transfer, assay fidelity, and external testing.

Evaluation Principles for Biomedical Discovery AI

The core evaluation question is not whether a model performs well on held-out rows. The core question is whether it improves a real experimental decision under the distribution where it will be used.

Protein Structure Prediction

Structure models are now routine inputs to biology, but they are not substitutes for experiments. Confidence, conformational state, ligand geometry, and biological context determine whether a predicted structure supports a downstream decision.

Protein Design and Engineering

Protein design is strongest when the target is structurally specified and the success assay is direct. Claims become weaker as design moves from fold, to binding, to catalysis, to cellular phenotype.

Antibody and Biologic Design

AI design methods help generate and prioritize binders. Therapeutic biologics still require assay cascades, liability screening, cell-based testing, and manufacturing review.

Nucleic Acid and Genome Models

DNA and RNA models are strongest when the output is tied to measured functional genomic assays. Variant interpretation remains difficult when disease mechanism, cell context, and long-range regulation are uncertain.

Variant Effect Prediction

Variant models help prioritize variants and hypotheses. They do not replace segregation evidence, functional assays, population frequency, disease mechanism, and clinical interpretation.

Target Identification and Prioritization

AI-assisted target selection is useful when it integrates evidence transparently. The winning target is not the top-ranked node. It is the target with a testable mechanism, feasible modality, safety rationale, and disease-relevant assay path.

Small Molecule Generation and ADMET

The useful output is not a molecule that looks novel. The useful output is a prioritized set of compounds with rationale, feasibility, assay plan, and acceptable risk across potency, selectivity, ADMET, and chemistry.

mRNA, RNA, and Vaccine Design

RNA and vaccine AI is strongest when the model output is tied to a measurable endpoint: expression, stability, antigenicity, manufacturability, or immune response. Program success still depends on delivery, dosing, safety, and clinical evidence.

Clinical Trial AI for Translational Research

AI in trials is safest when the context of use is explicit. Recruitment support, site selection, endpoint extraction, enrichment, and synthetic controls carry different evidentiary and regulatory burdens.

Translational Evidence and Failure Modes

A model that improves a proxy endpoint may still harm the program if the proxy is poorly linked to disease biology or developability. Failure analysis belongs near the start of the workflow, not after candidate nomination.

Single-Cell Foundation Models

Single-cell foundation models are useful representation systems, not general virtual cells. Evaluation must account for cell type, donor, batch, disease state, and perturbation split.

Spatial Omics and Tissue Models

Spatial AI is most useful when it links molecular signals to tissue structure with clear resolution limits. It is not enough to assign labels to spots or cells. The biological question is whether spatial organization changes mechanism or decision.

Cell Painting and Image-Based Phenotyping

Cell images are rich biological measurements, but morphology is not mechanism by itself. High-content imaging requires careful controls, segmentation quality checks, and orthogonal validation.

Perturbation Prediction and Virtual Cells

Virtual cell work is promising when framed as perturbation prediction for defined outputs. It becomes misleading when a transcriptomic forecast is treated as a full model of the cell.

Microbiome and Multi-Omics AI

Multi-omics models are useful when each modality has clear provenance and the validation endpoint is explicit. Integration can hide weak measurements if the workflow does not track missingness and batch effects.

Self-Driving Laboratories

A self-driving lab is an experimental system with a model in the loop. It needs reliable instruments, machine-readable protocols, calibration, error handling, and human review of objectives and stopping rules.

Robotic Lab Automation and Cloud Labs

Automation improves repeatability only when protocols, reagents, instruments, and data capture are explicit. A robot executing a vague protocol only scales ambiguity.

Synthetic Biology Design Tools

Synthetic biology AI is useful when design output is tied to a build and test plan. Sequence novelty alone is not engineering progress.

Agentic Science Workflows

Agentic systems are useful as research operating layers when tasks are bounded, sources are checked, and lab actions require authorization. They are risky when fluent plans are treated as validated science.

Benchmarks for Bio AI

Benchmarks matter when they are hard to game, close to the intended decision, and paired with failure analysis. A leaderboard is not a validation plan.

Reproducibility and Open Science

Open code and open weights help, but life sciences reproducibility also needs protocol detail, reagent provenance, data versioning, and assay records. Scientific openness and risk management must be handled together.

Information Hazards in Capability Research

Responsible capability research keeps enough detail for verification while avoiding unnecessary operational detail that raises misuse risk. The standard is not secrecy by default. The standard is deliberate disclosure.

Workforce, Compute, and Institutional Readiness

The minimum viable team includes biological domain expertise, data engineering, machine learning, experimental validation, and governance. Compute without experimental judgment creates expensive noise.

Case Studies

Case 1: Structure Prediction as a Research Input

A team without an experimental structure uses AlphaFold DB to generate hypotheses about a protein domain. The correct use is hypothesis generation, followed by confidence review, domain inspection, and assay planning (AlphaFold Protein Structure Database, 2026).

Case 2: Protein Binder Design

A protein engineering group uses RFdiffusion for backbone proposals and ProteinMPNN for sequence design. The design is not a success until expression, binding, specificity, and stability data support it (Watson et al., 2023; Dauparas et al., 2022).

Case 3: Docking Benchmark Failure

A docking method looks strong by RMSD but generates physically implausible poses. PoseBusters shows why chemical validity checks belong beside geometric metrics (Buttenschoen et al., 2024).

Case 4: Single-Cell Perturbation Forecasting

A group uses GEARS to rank perturbations for a follow-up experiment. The result supports prioritization, not general causal certainty (Roohani et al., 2024).

Case 5: Closed-Loop Experimentation

A self-driving lab chooses the next experiment based on prior measurements. The system needs protocol versioning, instrument logs, objective functions, and stopping rules (Abolhasani and Kumacheva, 2023; Mehr et al., 2024).

Case 6: Agentic Research Planning

A research agent drafts hypotheses and suggests protocols. The output requires source verification, human review, and authorization before any biological action (ARPA-H IGoR, 2026).

Glossary

ADMET

Absorption, distribution, metabolism, excretion, and toxicity. These properties shape whether a compound can become a drug.

AI-ready data

Data with sufficient metadata, provenance, structure, quality control, and governance for reliable machine learning use.

Backbone generation

Protein design task that proposes a protein fold or structural scaffold before sequence selection.

Benchmark leakage

Evaluation failure where training data or related examples contaminate a test set.

Cell Painting

High-content imaging assay that profiles cellular morphology using multiplexed fluorescent stains.

Closed-loop experimentation

Workflow where measurements update a model that selects future experiments.

Foundation model

Model pretrained on broad data so its representations transfer to downstream tasks.

Inverse folding

Protein design task that proposes a sequence for a target structure.

Perturbation prediction

Forecasting how a biological system changes after a genetic, chemical, or environmental intervention.

Virtual cell

Computational model intended to predict cell behavior, usually for defined outputs rather than complete cellular simulation.

Model and Dataset Index

Model or Dataset	Main Use	Evidence Tier	Source
AlphaFold 2	Protein structure prediction	Demonstrated	Jumper et al., 2021
AlphaFold 3	Biomolecular interaction structure prediction	Demonstrated	Abramson et al., 2024
AlphaFold DB	Predicted protein structure database	Demonstrated	AlphaFold Protein Structure Database, 2026
ESMFold	Protein language model structure prediction	Demonstrated	Lin et al., 2023
ESM3	Multimodal protein language model	Demonstrated for selected generation tasks	Hayes et al., 2025
RFdiffusion	Protein backbone generation and design	Demonstrated	Watson et al., 2023
ProteinMPNN	Protein sequence design for structures	Demonstrated	Dauparas et al., 2022
Boltz-1	Open biomolecular interaction modeling	Demonstrated as an open research system	Boltz-1, 2024
Evo	Genome-scale biological sequence modeling	Demonstrated in published tasks	Nguyen et al., 2024
AlphaGenome	Regulatory variant-effect prediction	Demonstrated	Avsec et al., 2026
Geneformer	Single-cell gene and cell representations	Demonstrated	Theodoris et al., 2023
scGPT	Single-cell multi-omics foundation model	Demonstrated	Cui et al., 2024
GEARS	Genetic perturbation response prediction	Demonstrated in selected settings	Roohani et al., 2024

Model or Dataset	Main Use	Evidence Tier	Source
PDB	Experimental macromolecular structure archive	Demonstrated infrastructure	wwPDB, 2026
ChEMBL	Curated bioactivity database	Demonstrated infrastructure	Zdrazil et al., 2024
PubChem	Chemical substance, compound, and assay database	Demonstrated infrastructure	Kim et al., 2023
MoleculeNet	Molecular ML benchmark suite	Demonstrated benchmark	Wu et al., 2018
PoseBusters	Docking pose validity checks	Demonstrated benchmark	Buttenschoen et al., 2024

How to Cite

Suggested Citation

Tegomoh, B. (2026). *The Life Sciences AI Handbook: AI for Biomedical Discovery, Biotechnology, and Translational Research*. DOI pending. URL: https://lifesciencesaihandbook.com

APA

Tegomoh, B. (2026). *The Life Sciences AI Handbook: AI for Biomedical Discovery, Biotechnology, and Translational Research*. https://lifesciencesaihandbook.com

BibTeX

```
@book{tegomoh2026lifesciencesai,
  title = {The Life Sciences AI Handbook: AI for Biomedical Discovery, Biotechnology, and Tra
  author = {Tegomoh, Bryan},
  year = {2026},
  url = {https://lifesciencesaihandbook.com},
  note = {DOI pending}
}
```

License

License