AI Scientist for Large-Scale Biomedical Research

Eli Conlin[1], Prem Rajendran[1], Brooklyn Varela[2]

College of Arts and Sciences, Boston University, MA[1], College of Engineering, Boston University, MA[2]

Introduction

Our project builds a prototype neuro-symbolic AI Scientist for the biomedical domain. Rather than treating a large language model as a stand-alone chatbot, we decompose the research workflow into coordinated agents that read papers, extract candidate mechanisms, and design experiments. A separate reasoning layer then checks these ideas against the iKraph biomedical knowledge graph and simple simulation tools, so the system outputs hypotheses with explicit evidence paths and concrete experiment blueprints; moving from free-form text to something closer to a reproducible research plan.

Motivation

Modern LLMs can write, summarize, and reason over scientific text, but they still struggle with turning scientific ideas into complete, correct, and executable experiments. They hallucinate steps, violate physical constraints, omit crucial parameters, and cannot reliably validate their own work. At the same time, scientific research produces vast amounts of literature, making it increasingly difficult for humans to synthesize knowledge and design experiments efficiently.

Our project explores whether an "AI Scientist" can close this gap by integrating three capabilities:

1. reading and reasoning over existing scientific knowledge,
2. generating structured experimental protocols, and
3. validating those protocols through simulation and physical rules.

Research Objectives

Our work aims to address the following research questions:

- RQ_1: How does grounding AI-generated biomedical hypotheses in the iKraph knowledge graph affect the plausibility, safety, and novelty of proposed experiments compared to an ungrounded LLM baseline?
- RQ_2: Can robotic simulation serve as an intermediate layer toward autonomous experiment execution?

Knowledge Graph

Our AI Scientist is grounded in iKraph, a large-scale biomedical knowledge graph built from all PubMed abstracts and enriched with relations from more than 40 public databases and genomics sources, covering diseases, genes, drugs, pathways, and other key entities. iKraph thus serves as the symbolic constraint layer in our neuro-symbolic loop, filtering out biologically implausible or unsupported ideas before they proceed to simulation while providing structured, citable evidence for those that survive; we also log successful experiments into a project-specific subgraph, giving the agent an explicit, query-able memory of past findings. We query iKraph through a Neo4j graph database using Cypher, enabling the agent to run targeted analyses such as the example shown below.

Example iKraph Cypher Query

```
MATCH (s:Entity)-[r:Positive_Correlation]-(t:Entity)
WHERE
  (
    toLower(s.`official name`) = 'adenosine triphosphate' OR
    toLower(s.`common name`)  = 'atp' OR
    toLower(s.`official name`) = 'atp'
  )
  AND (
    toLower(t.`official name`) CONTAINS 'acetate kinase' OR
    toLower(t.`common name`)  CONTAINS 'acetate kinase'
  )
OPTIONAL MATCH
↪  (s)-[ms:MENTIONED_IN]->(p:PubMed)<-[mt:MENTIONED_IN]-(t)
RETURN DISTINCT
  s.id                                              AS source_id,
  coalesce(s.`official name`, s.`common name`)      AS source_name,
  s.type                                            AS source_type,
  t.id                                              AS target_id,
  coalesce(t.`official name`, t.`common name`)      AS target_name,
  t.type                                            AS target_type,
  type(r)                                           AS
  ↪  relationship_type,
  toFloat(coalesce(r.prob, ms.prob, mt.prob))       AS probability,
  p.pmid                                            AS pmid,
  CASE WHEN p.pmid IS NULL
       THEN NULL
       ELSE 'https://pubmed.ncbi.nlm.nih.gov/' + p.pmid
  END                                               AS url,
  p.date                                            AS
  ↪  publication_date,
  p.sentence                                        AS sentence
ORDER BY probability DESC, pmid;
```

source_id	source_name	source_type	target_id	target_name	target_type	relationship_type	probability	pmid	url	publication_date	sentence
92	ATP	Chemical	1767	Acetate Kinase	Gene	Positive_Correlation	0.9999	10816048	https://pubmed.ncbi.nlm	20000518	Enzyme activities for all reaction steps from glutaryl-CoA to butyryl-CoA, and ATP formation via acetate kinase were detected in cell-free extracts.
92	ATP	Chemical	1767	Acetate Kinase	Gene	Positive_Correlation	0.9999	11716215	https://pubmed.ncbi.nlm	20011122	In the metabolism of Lactobacillus sanfrancisconsis, the acetate kinase (AK) is a key enzyme and responsible for dephosphorylation of acetyl phosphate with the concomitant production of acetate and ATP.
92	ATP	Chemical	1767	Acetate Kinase	Gene	Positive_Correlation	0.9999	12450851	https://pubmed.ncbi.nlm	20021127	A similar benefit was obtained by inactivation of acetate kinase (ackA), reducing the production of acetate (and ATP) and sparing acetyl-CoA for biosynthetic needs.
92	ATP	Chemical	1767	Acetate Kinase	Gene	Positive_Correlation	0.9999	15516572	https://pubmed.ncbi.nlm	20041102	An ackA mutant, lacking the ability to generate ATP from acetyl-phosphate, also failed to grow in xylose minimal medium under anaerobic conditions, confirming the need for the ATP produced by acetate kinase for anaerobic growth on xylose.
92	ATP	Chemical	1767	Acetate Kinase	Gene	Positive_Correlation	0.9999	25979891	https://pubmed.ncbi.nlm	20150515	In the dental caries pathogen Streptococcus mutans, phosphotransacetylase (Pta) catalyzes the conversion of acetyl coenzyme A (acetyl-CoA) to acetyl phosphate (AcP), which can be converted to acetate by acetate kinase (AckA), with the concomitant generation of ATP.
92	ATP	Chemical	1767	Acetate Kinase	Gene	Positive_Correlation	0.9999	31007866	https://pubmed.ncbi.nlm	20190404	Characterization of the phosphotransacetylase-acetate kinase pathway for ATP production in Porphyromonas gingivalis. Acetyl phosphate (AcP) is generally produced from acetyl coenzyme A by phosphotransacetylase (Pta), and subsequent reaction with ADP, catalyzed by acetate kinase (AckA), produces ATP. ...ATP formation by AckA proceeds via a sequential mechanism.
92	ATP	Chemical	1767	Acetate Kinase	Gene	Positive_Correlation	0.9999	9484901	https://pubmed.ncbi.nlm	19980304	Double null mutants deficient in phosphotransacetylase (Pta) and acetate kinase (AckA) or AckA and TdcD were unable to metabolize threonine to propionate, indicating that propionyl-CoA and propionyl-phosphate are intermediates in the pathway and that ATP is generated during the conversion of propionyl-P to propionate by AckA or TdcD.

Figure 1: Result of the Cypher query exported as a csv file

Experimental Agent

The Experimental Planning Agent converts a mechanistic hypothesis about *E. coli* metabolism into a full fluorescence-plate experiment that discriminates between GIVEN causal confounds under fixed well/probe budgets. It behaves as a compiler: hypothesis → workspace → experiment plan.

2. Knowledge Graph Reporter Selection. KGraph paths from interventions to targets identify measurable entities. The agent selects a minimal reporter set maximizing path coverage and confound discrimination. Example: ATP optimization emphasized reporters tied to acetate overflow, NADH balance, and ATP maintenance—entities appearing across competing causal explanations.

3. Confound-Driven Experiment Design. Using the selected reporters, the agent builds experiments explicitly designed to separate GIVEN confounds:

- *Time series* for direct vs mediated vs compensatory dynamics.
- *CRISPRi/a dose–responses* for linear vs ultrasensitive effects.
- *Multi-node perturbations* (intervention + mediator/compensator KO).
- *State-variation* to expose state-dependent artifacts.

Each design includes expected patterns under competing mechanisms.

4. Well Allocation & Plate Layout. Experiment families are ranked by *discriminative power per well*. The agent fits the highest-value sets within the max-wells budget, trimming timepoints or doses when necessary. It outputs precise 96/384-well layouts annotated with intervention, reporter, timepoint, replicate, and confound tested.

5. Metabolic Validation After design, the agent attempts COBRApy simulations to validate its effects

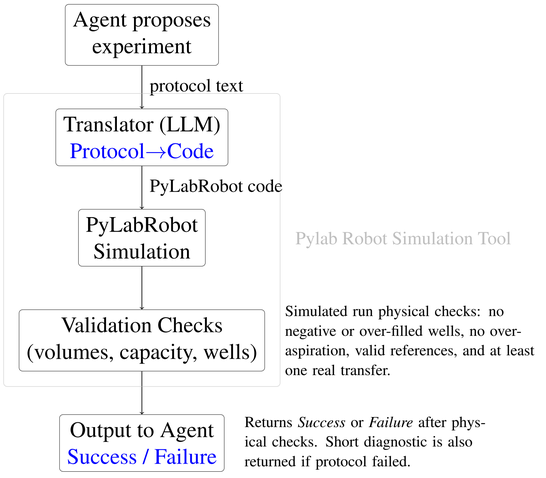
6. Output. The final JSON includes KGraph path summaries, reporter scoring, experiment families, prioritized selection, plate layout, budget summary, and (if available) metabolic validation results.

Demonstrated Behavior. In the ATP–growth run, the agent produced a complete experiment workspace, generated multi-KO designs with confound rationales, and maintained a coherent experimental plan—showing a functioning system rather than a hypothetical pipeline.

Robotic Validation with PyLabRobot

This robotics simulation acts as an external validation layer. After the agent proposes an experiment and the knowledge graph checks feasibility, the robotic layer verifies whether it can be executed as a real liquid-handling protocol. This forces the AI scientist to ground its ideas in experiments it can actually run.

Robotic Simulation Workflow



Natural Language to PyLabRobot Code

Example Translation

Input prompt

"Transfer 100 µL of buffer from A1 of the reservoir to B2 of the plate. Then mix the contents of well B2 three times with 50 µL."

Generated PyLabRobot code

```
# Transfer 100 µL from A1 → B2
await lh.pick_up_tips(tip_rack["A1"])
await lh.aspirate([reservoir["A1"][0]], vols=[100.0])
await lh.dispense([plate["B2"][0]], vols=[100.0])
await lh.return_tips()

# Mix B2 three times with 50 µL
for _ in range(3):
    await lh.pick_up_tips(tip_rack["A1"])
    await lh.aspirate([plate["B2"][0]], vols=[50.0])
    await lh.dispense([plate["B2"][0]], vols=[50.0])
    await lh.return_tips()
```

Conclusion

The integrated system demonstrated a full causal-to-execution pipeline: iKraph extracted mechanistic paths and confound structure; the Experimental Agent converted them into a budget-aware fluorescence experiments, and the PyLabRobot translator supplied the automation bridge for robotic execution.