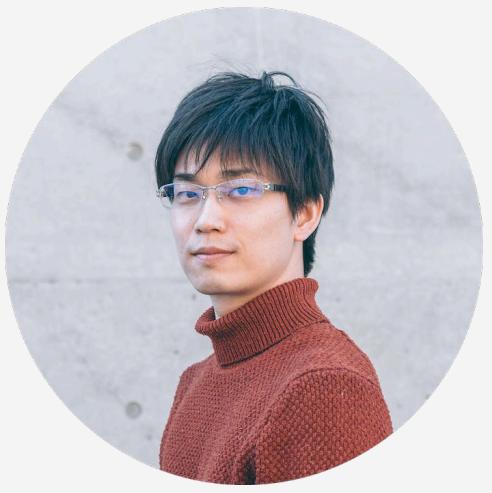


# 研究特化型AIエージェントの 構築と活用

Science Aid株式会社 代表取締役 CEO 山田涼太



# 山田 涼太

Science Aid株式会社 代表取締役 CEO

✉ ryota.yamada@science-aid.com

𝕏 [@roy29fuku](https://twitter.com/roy29fuku)

2010-04 東京大学農学部獣医学専修

2017-04 東京大学工学部システム創成学科  
知能社会システムコース

2018-03 Science Aid株式会社 創業  
(旧社名: fuku株式会社)

一般社団法人ラボラトリーオートメーション協会  
AI分科会 オーガナイザー

- ・月例勉強会 (LA\_Study)
- ・年次学会 (LADEC)
- ・コンペティション (LA-Bench)



<https://lasa.or.jp/en/>

## AI・ロボットによる科学研究の自動化

# サイエンスを開き、 人類の可能性を広げる。

科学情報は、世界中で圧倒的な量と複雑さで存在しています。

私たちの使命は、技術を活用して人類がこの膨大な知識を導き、活用するのを支援することです。

これらの技術的ソリューションを提供することにより、私たちは人類が新たな可能性を開き、科学の進歩を促進するのをサポートします。



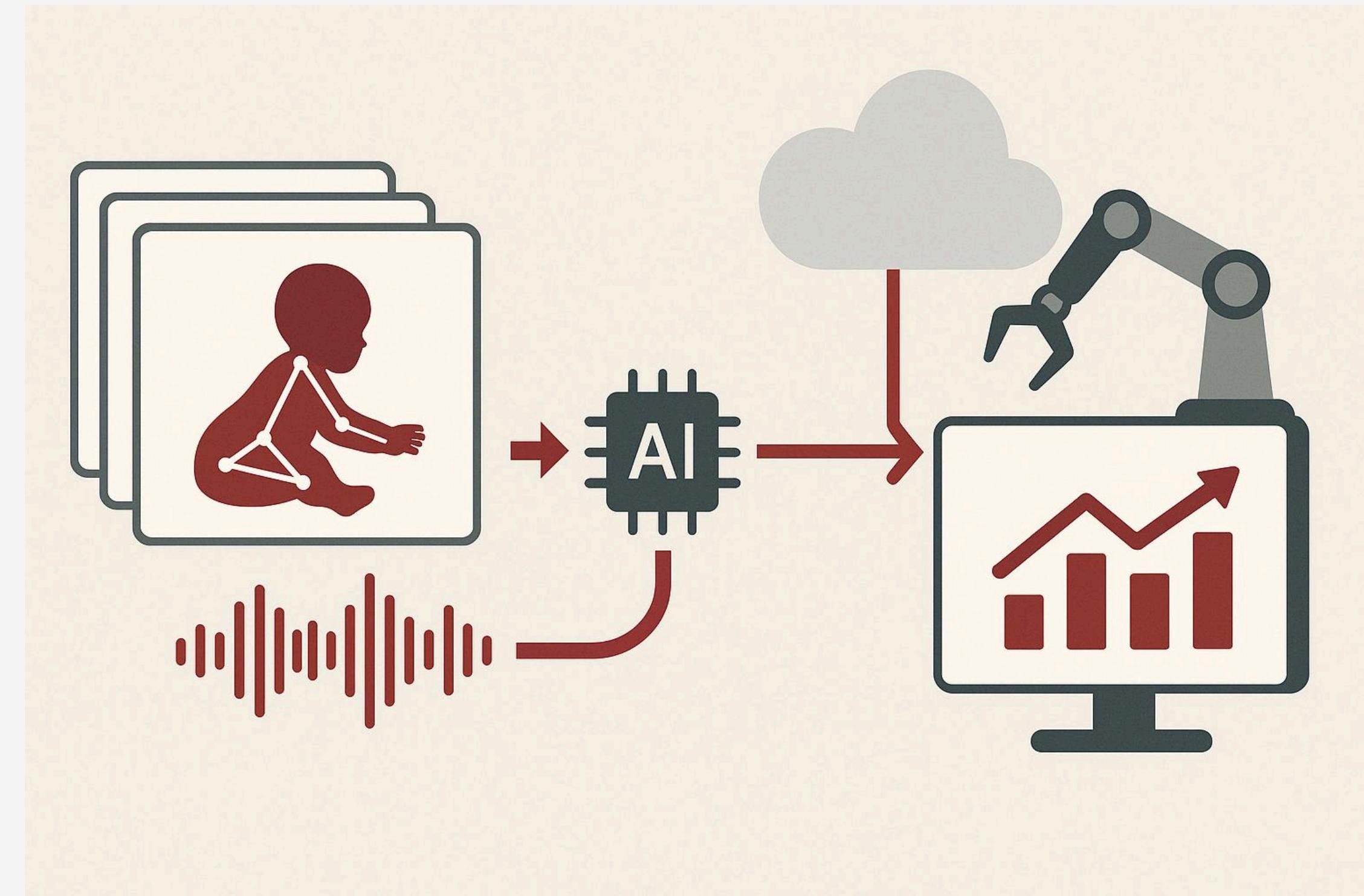
# これまでの取り組み

## 【多数】論文解析



Generated by ChatGPT

## 【FVital】乳幼児音声動画解析



Generated by ChatGPT

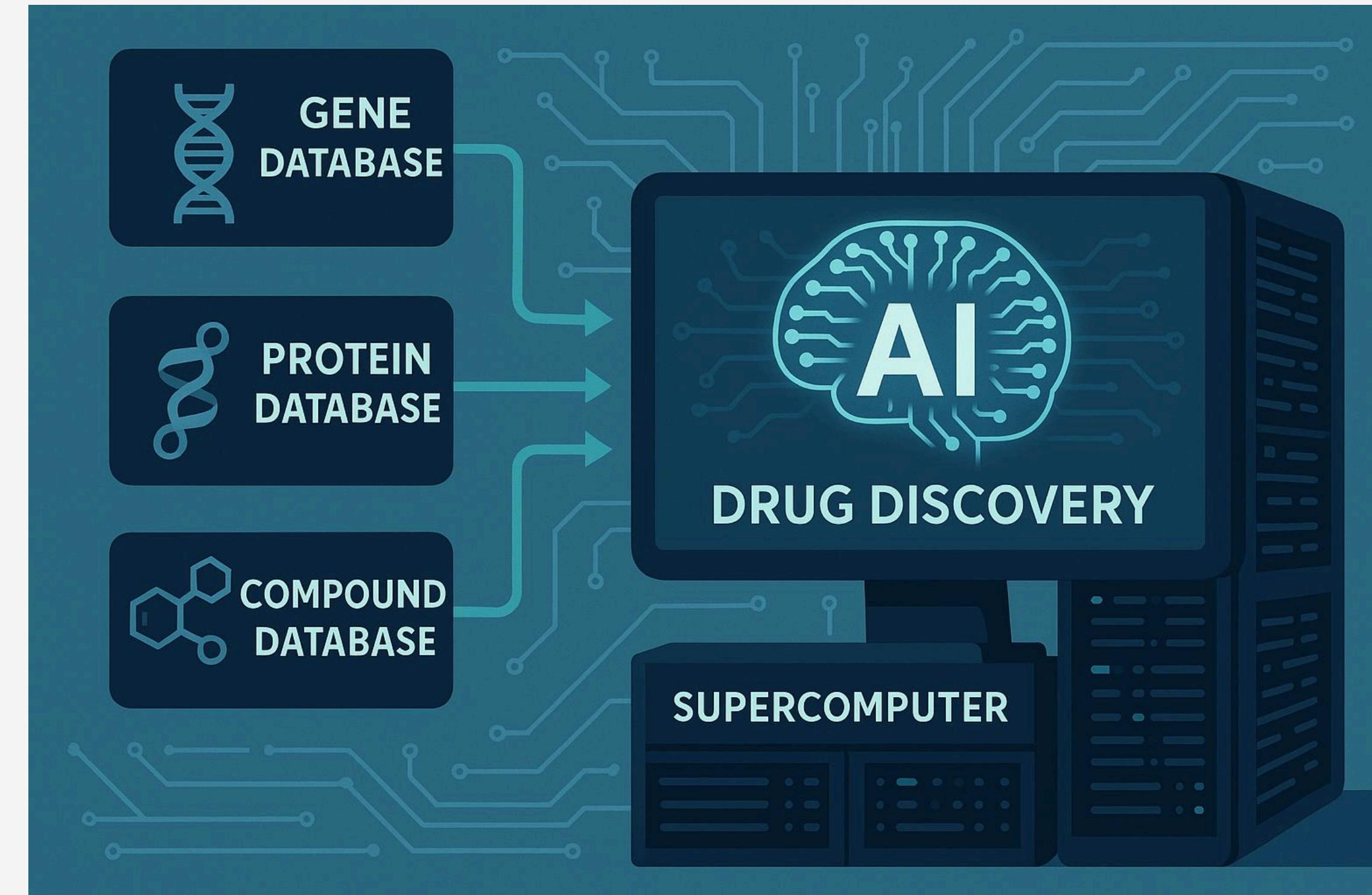
# これまでの取り組み

## 【理研】ロボット実験施設 バックエンド開発



Generated by ChatGPT

## 【京大】創薬AI データベース構築



Generated by ChatGPT

最近は生成AI、AIエージェントの相談が増加

特に科学研究を行うAIエージェントに  
に関する問い合わせが急増

# AIエージェントに関する取り組み

**Science Aid株式会社、京都大学医学研究科 奥野研究室と共同で創薬AIエージェント開発を開始**

— 標的サイト情報抽出に特化した次世代AIで創薬研究の効率化を支援 —

Science Aid株式会社 2025年9月30日 10時00分



<http://prtmes.jp/main/html/rd/p/00000004.000059896.html>

# AIエージェントに関する取り組み

飲料メーカー (PoC検証中)

製薬企業 (PoC開始予定)

製薬企業 (構想段階)

自動車メーカー (構想段階)

理研・筑波大学・東京科学大学 (構想段階)

# 今日のテーマ

## ライフサイエンス研究への AIエージェント活用の鍵

# AIエージェントとは

# AIエージェントとは

「世界を観察し、自由に使えるツールを使って行動することで、目標を達成しようとするアプリケーション」

## Agents

Authors: Julia Wiesinger, Patrick Marlow and Vladimir Vuskovic



Google

### What is an agent?

In its most fundamental form, a Generative AI agent can be defined as an application that attempts to achieve a goal by observing the world and acting upon it using the tools that it has at its disposal. Agents are autonomous and can act independently of human intervention, especially when provided with proper goals or objectives they are meant to achieve. Agents can also be proactive in their approach to reaching their goals. Even in the absence of explicit instruction sets from a human, an agent can reason about what it should do next to achieve its ultimate goal. While the notion of agents in AI is quite general and powerful, this whitepaper focuses on the specific types of agents that Generative AI models are capable of building at the time of publication.

In order to understand the inner workings of an agent, let's first introduce the foundational components that drive the agent's behavior, actions, and decision making. The combination of these components can be described as a cognitive architecture, and there are many such architectures that can be achieved by the mixing and matching of these components. Focusing on the core functionalities, there are three essential components in an agent's cognitive architecture as shown in Figure 1.

<https://www.kaggle.com/whitepaper-agents>

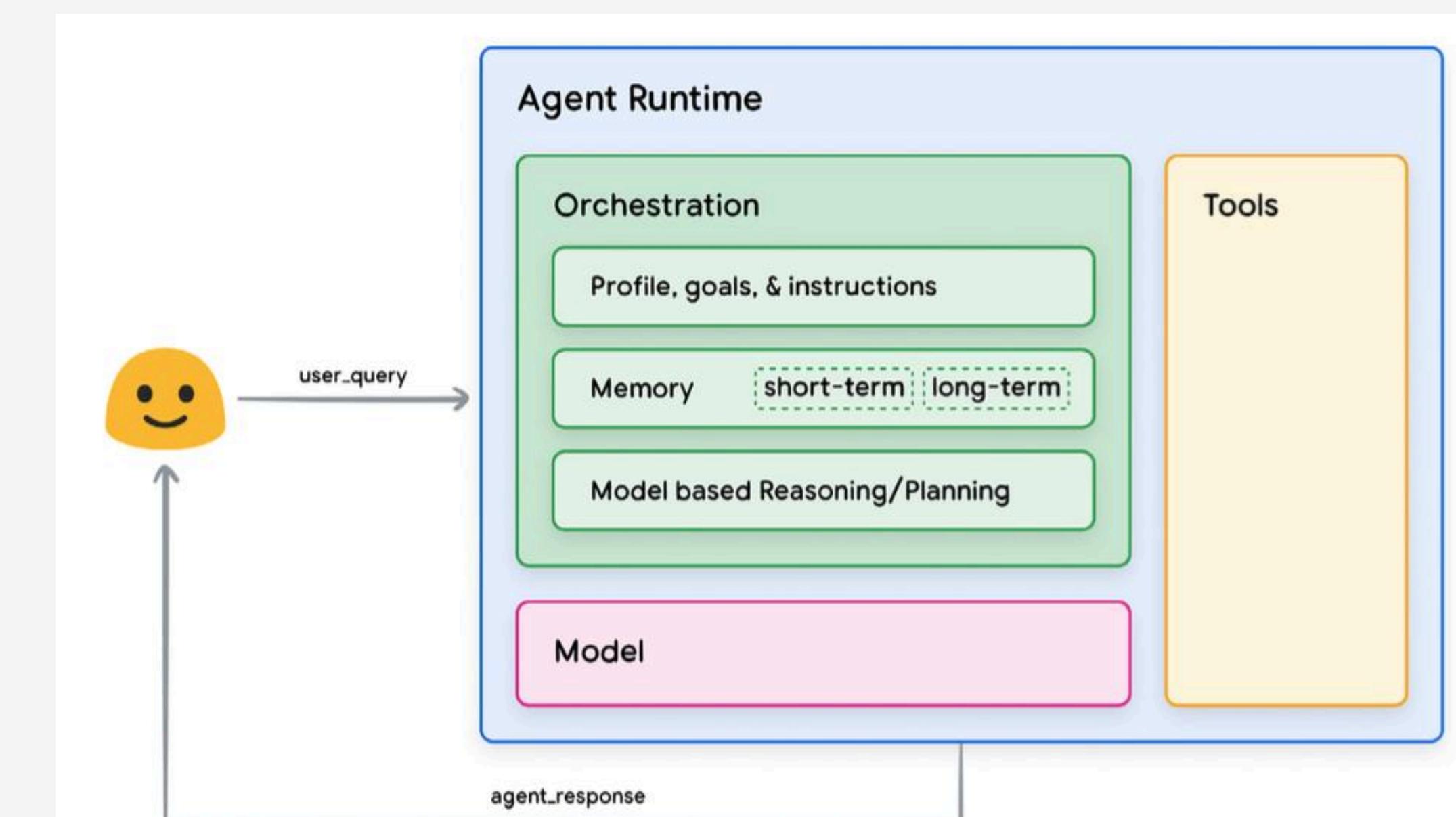
# 大規模言語モデル (LLM: Large Language Model) との違い

## LLM

- テキストの理解と生成に特化
- 入力に対して単一の応答を返す
- 知識は学習時点までのデータに基づく

## AIエージェント

- **LLMを中核として、それに追加機能を組み合わせたシステム**
- 目標達成のために複数のステップを計画・実行
- **外部ツール**（検索、計算、データベースなど）を使用可能
- 環境からのフィードバックを受けて行動を調整できる
- メモリ機能により過去のやり取りを記憶・活用できる



<https://www.kaggle.com/whitepaper-agents>

# AIエージェント in Computer Science

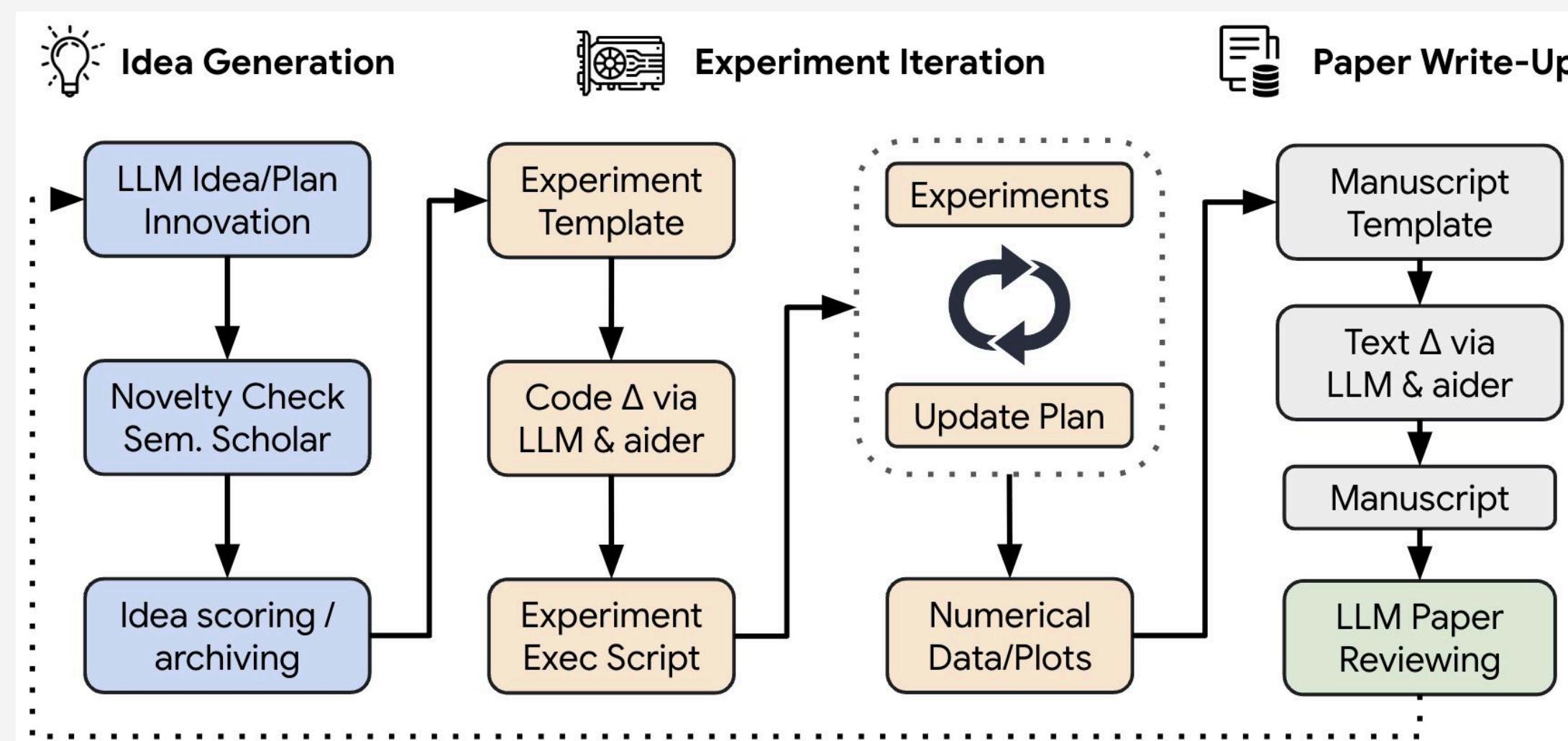
# The AI Scientist (Sakana AI)

[Lu 2024, arXiv:2408.06292; Yamada+ 2025, arXiv:2504.08066]

2024-08-12 v1

“the first comprehensive framework  
for fully automatic scientific discovery”

一定レベルで機械学習研究の自動化を達成



- 先行研究の調査
- アイディアの創発
- 実験（コードの生成・実行）
- 実験結果の分析
  - 執筆
  - レビュー

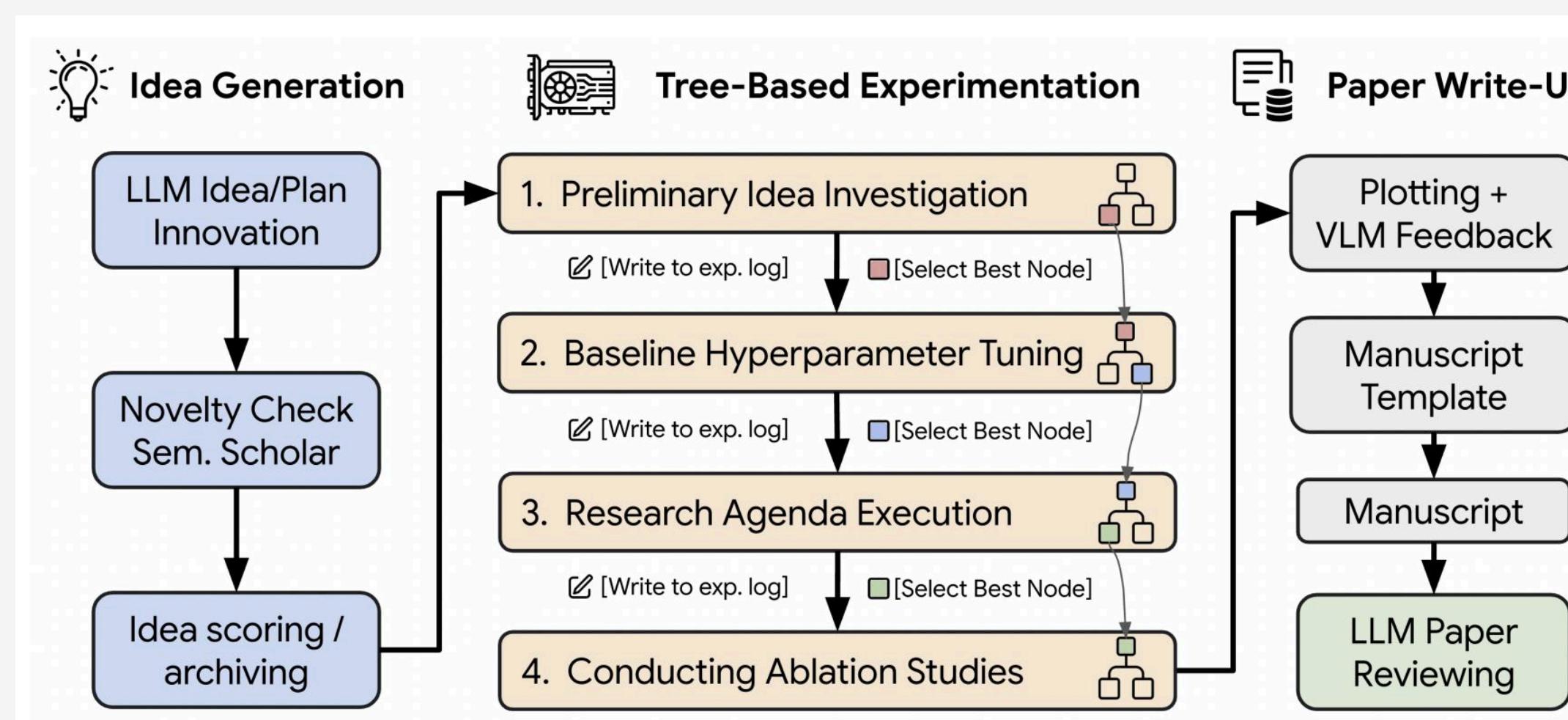
# The AI Scientist (Sakana AI)

[Lu 2024, arXiv:2408.06292; Yamada+ 2025, arXiv:2504.08066]

2025-03-13 v2

“The AI Scientist Generates  
its First Peer-Reviewed Scientific Publication”

国際会議のWSレベルを達成



3本のAI生成論文を提出し、1本が採択  
提出論文のtop 45%に位置するスコア

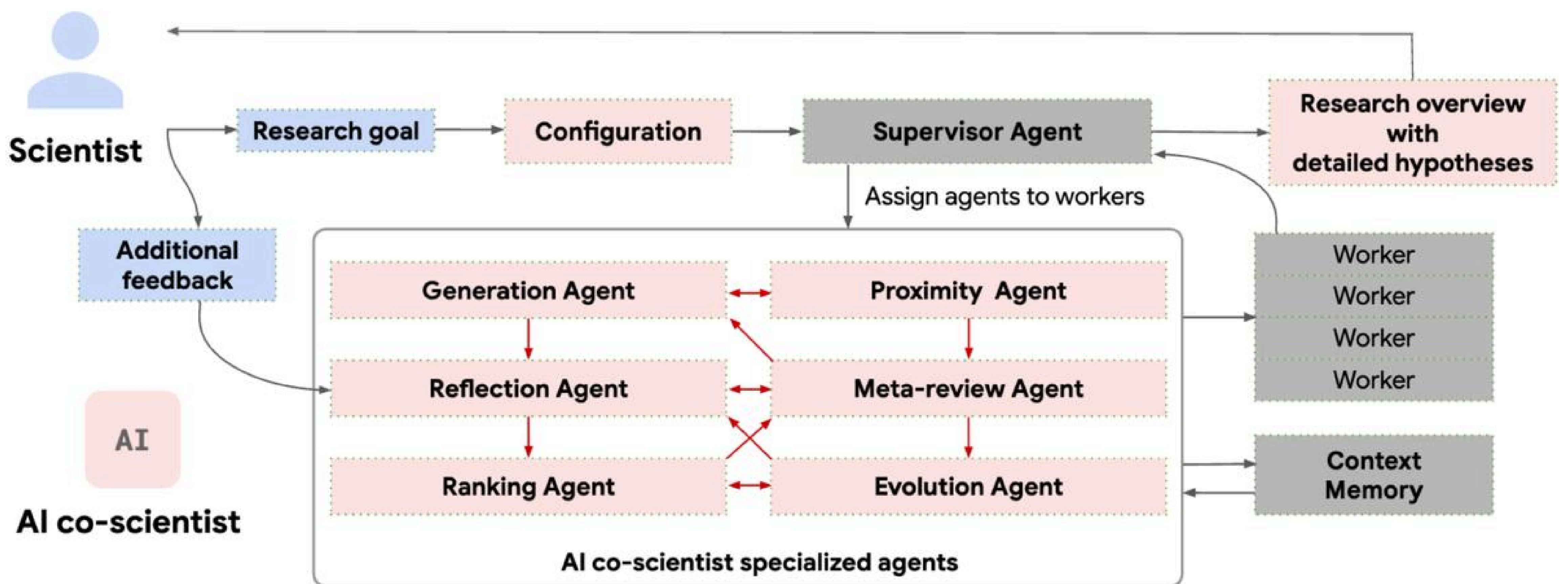
査読者にはAI生成論文が  
含まれることのみ通達  
(どのがAI製かは明かされず)

# AIエージェント in Life Science

# AI co-scientist (Google)

[Gottweis+, arXiv:2502.18864]

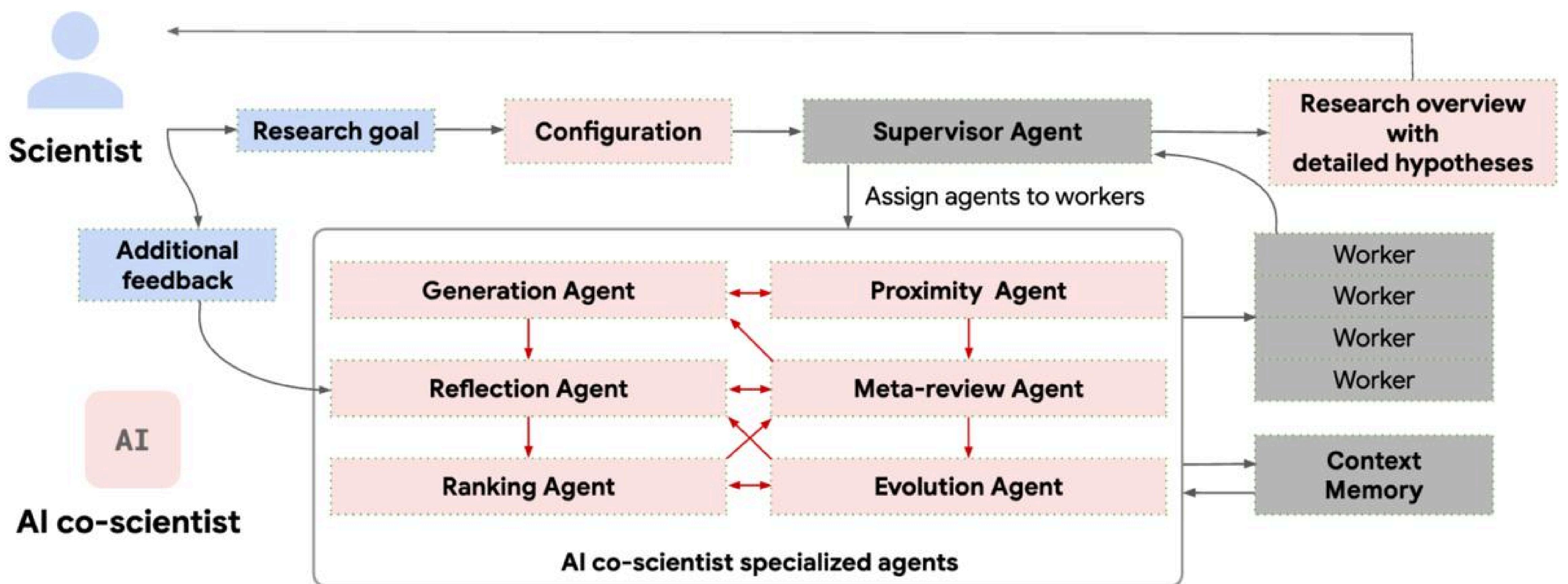
“virtual scientific collaborator”



# AI co-scientist (Google)

[Gottweis+, arXiv:2502.18864]

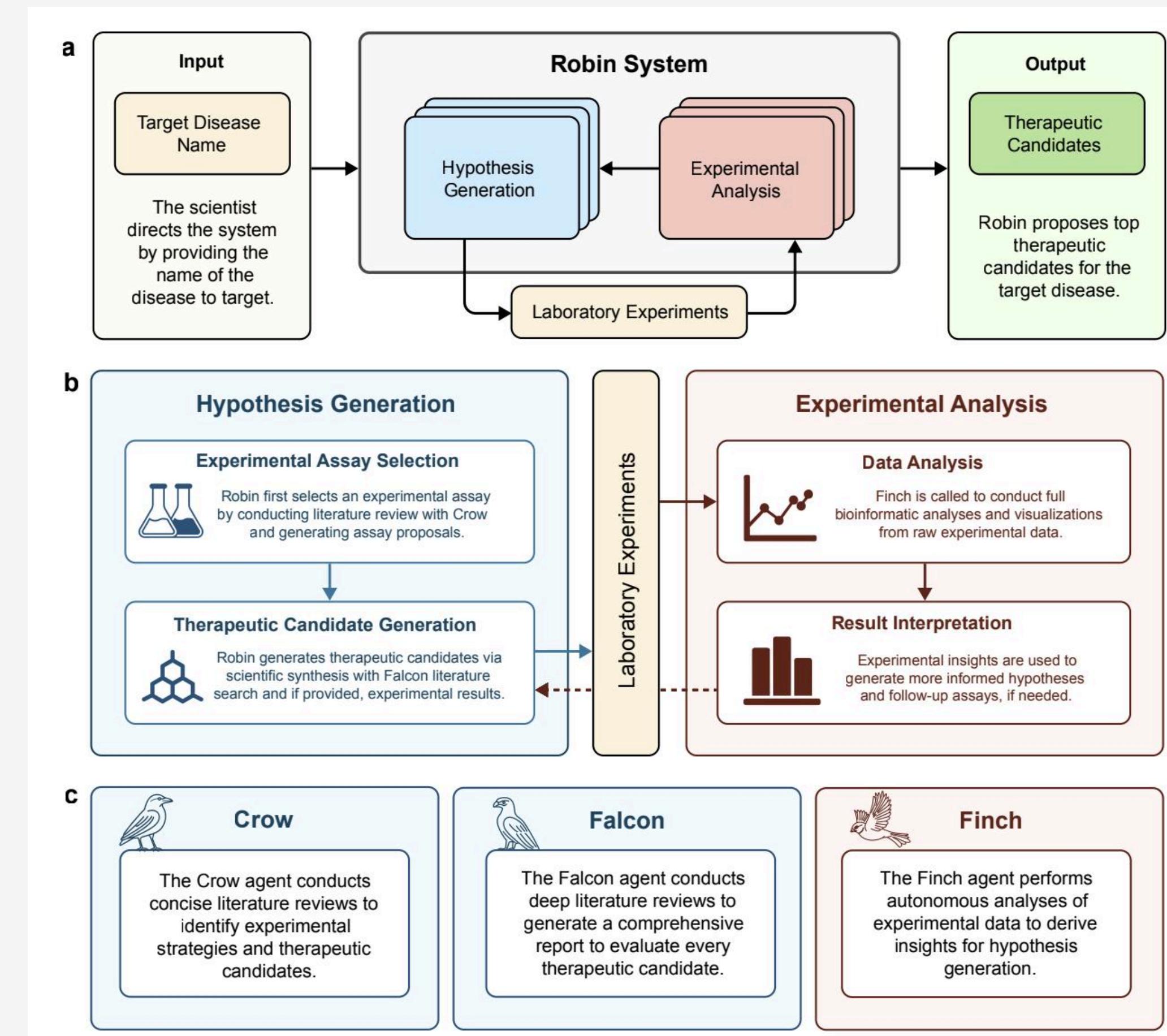
“virtual scientific collaborator”



# Robin (FutureHouse)

[Ghareeb+, arXiv:2505.13400]

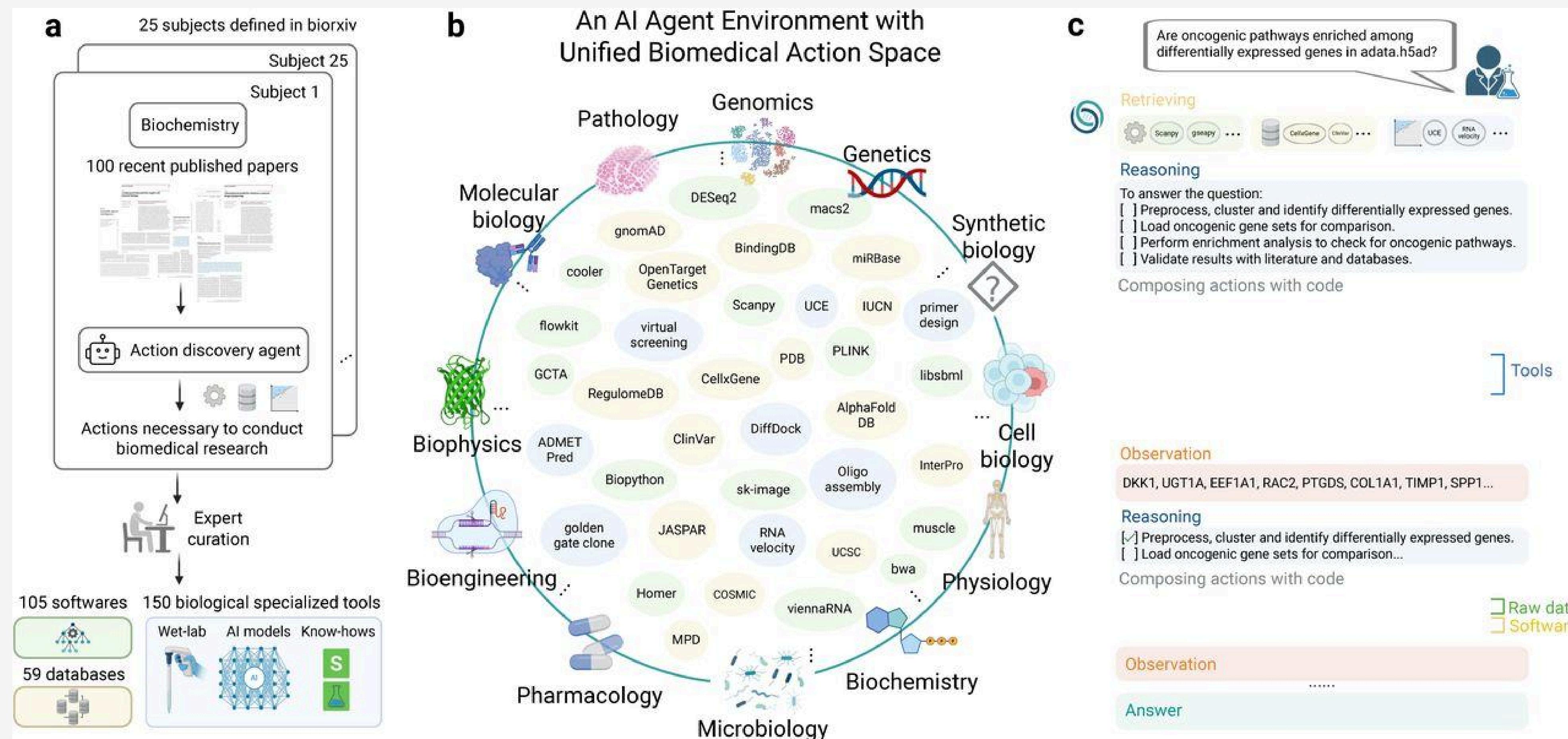
“Multi-agent system for automating scientific discovery”



# Biomni (Stanford University)

[Huang+, bioRxiv:2025.05.30.656746]

# “A General-Purpose Biomedical AI Agent”

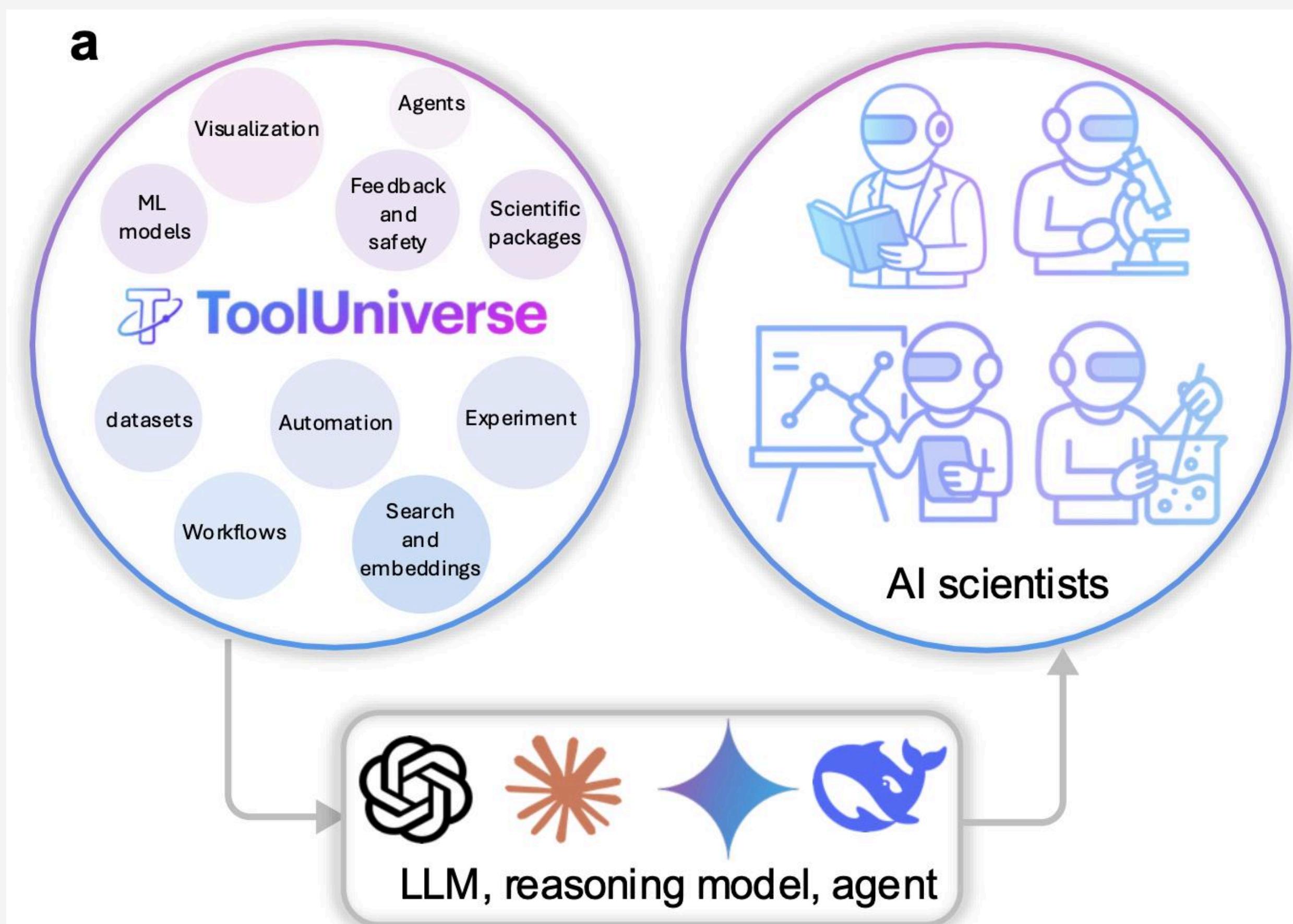


# 汎用AI Scientist

# ToolUniverse (Harvard University)

[Gao+, arXiv:2509.23426]

“ecosystem for building AI scientists”



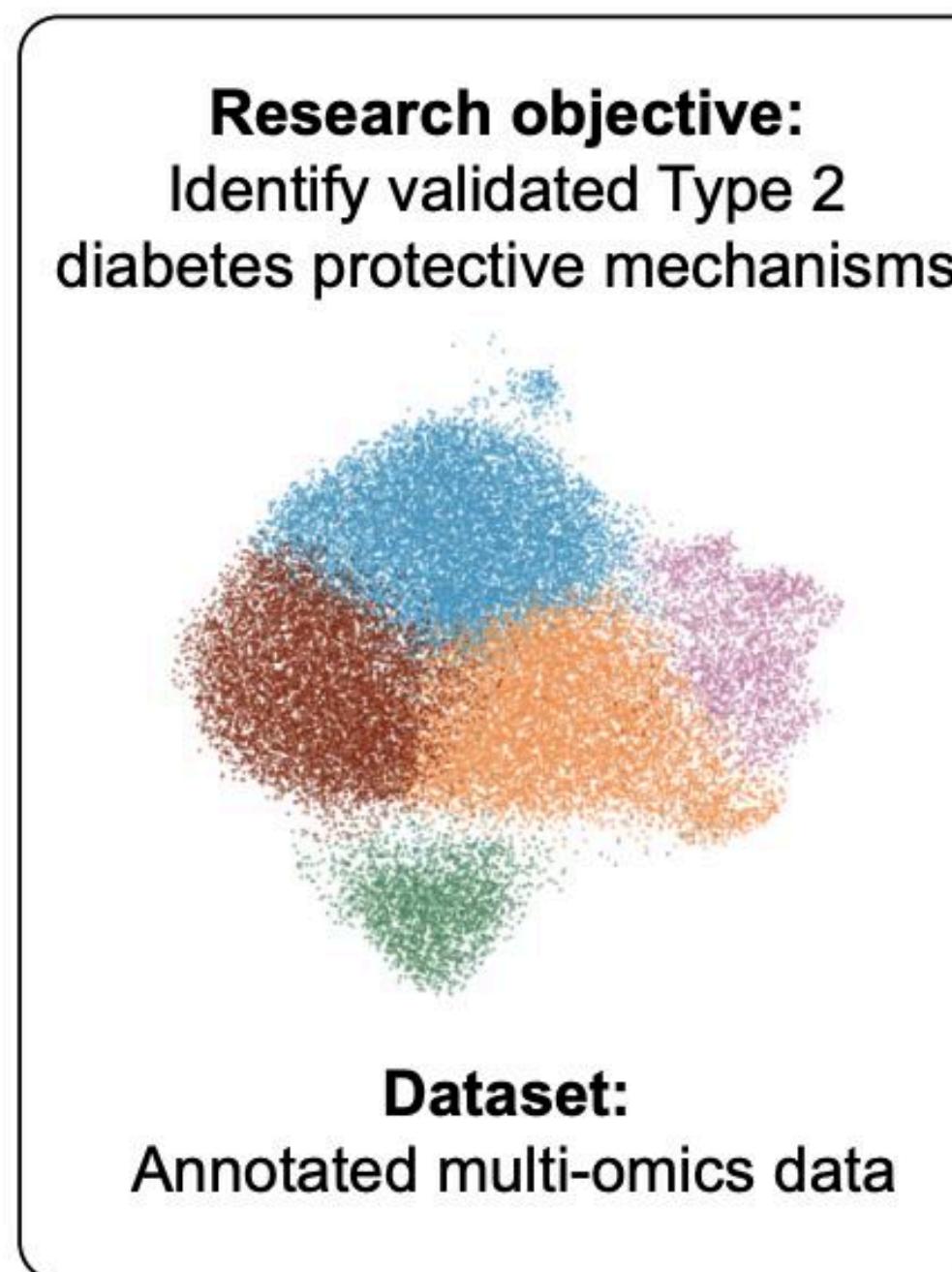
# Kosmos (Edison Scientific\*) \* FutureHouseが営利目的の活動のために立ち上げたスピノアウト

[Mitchener+, arXiv:2511.02824]

“An AI Scientist for Autonomous Discovery”

a

Input



Kosmos World Model



Output

## The T2D protective variant rs9379084-A tags a candidate cis-acting regulatory sequence that modulates levels of SSR1 via altered ATF3 binding

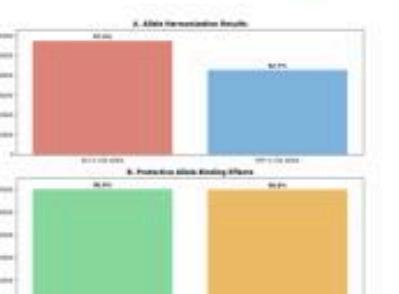
### Summary

The missense variant rs9379084-A at the RREB1 locus emerges as a regulatory hotspot for type 2 diabetes protection, with convergent multi-omic evidence indicating that the protective allele modulates transcription factor occupancy, chromatin accessibility, and downstream transcript levels in a cell-type-specific manner. Among the six genes influenced, SSR1 is a leading candidate target, supported by increased ATF3 motif affinity, islet-selective ATAC-QTL effects, and ranking within the top downregulating protective mechanisms.

ability of causality of 1.0, and integrative visualization confirms signal dominance in the local region [49]. Ranking of protective mechanisms that favor expression-lowering effects across available QTL streams places this locus among the most validated: six of the top ten protective mechanisms arise from rs9379084, including RREB1 (#2; score 6.000) and SSR1 (#4; score 5.000), with additional targets RIOK1, DSP, and SNRNP48, indicating a multi-gene regulatory hotspot consistent with coordinated cis-regulatory control [49].

### Background

Human genetic studies of type 2 diabetes increasingly point to noncoding and regulatory mechanisms that act through transcription factor binding, chromatin accessibility, and context-specific gene expression to rewire islet and metabolic tissues. Integrative frameworks that harmonize allele directionality, evaluate directional concordance between genome-wide association signals and quantitative trait loci across molecular layers, and prioritize mechanisms with multi-source support are needed to move from association to function. In this context, regulatory hotspots in which single variants orchestrate multi-gene programs have particular importance for identifying protective pathways and therapeutic entry points.



**Figure 1:** Allele harmonization establishes risk allele dominance and reveals direction of protective allele binding effects. (A) Distribution of SNP-transcription factor (TF) pairs where the risk allele is the alternate (ALT) versus the reference (REF) allele. (B) Proportions of protective alleles that cause a gain-of-function (GOF) or loss-of-function (LOF) in TF binding. The equal distribution of GOF and LOF effects for protective alleles minimizes systematic bias in inferring downstream regulatory mechanisms. (Source: [49])

### Results & Discussion

This work establishes rs9379084-A as a high-confidence protective variant and a coordinated regulatory driver at the RREB1 locus. Harmonization of allele definitions across datasets produced an analysis-ready resource (100,116 SNP-TF pairs; 1,276 SNPs; 389 TFs) with 100% consistency in risk/protective allele direction, enabling robust assessment of concordance between GWAS and QTL effects; notably, protective-allele TF binding effects were evenly split between gain and loss of function, minimizing systematic bias in mechanism inference [41]. At the locus level, rs9379084 shows strong protective association ( $\beta = -0.085$ ,  $P = 9.51 \times 10^{-37}$ ) with posterior prob-

The mechanistic chain for SSR1 is coherent across molecular layers. At the level of sequence-to-factor binding, rs9379084-A increases motif affinity for ATF3 ( $\Delta$ binding  $+1.39$ ), while also modulating a broader set of metabolic and stress-responsive factors including increased binding for NR112 ( $+1.62$ ), NR113 ( $+1.74$ ), VDR ( $-1.17$ ), and MYBL2 ( $+1.13$ ), and decreased binding for PPARA ( $-0.76$ ), EGR4 ( $-1.15$ ), KLF6 ( $-1.17$ ),

どういう仕組み？  
どれくらい信頼できる？

# AI エージェント in Life Science - Biomniを題材に深掘り -

## Biomniとは？

スタンフォード大学が開発した  
汎用バイオメディカルAIエージェント

150の専門的ツール、59のデータベース、  
105のソフトウェアを活用可能

無料のウェブアプリケーション  
Apache 2.0のソースコード

## どんなことができるか？（Biomniウェブサイト内事例より）

- Variant interpretation (変異解釈)
- Perturb-seq data analysis (Perturb-seq データ解析)
- ADMET property prediction (ADMET特性予測)
- Cloning protocol design (クローニングプロトコル設計)
- CRISPR screen design (CRISPRスクリーニング設計)
- Rare disease diagnosis (希少疾患診断)
- Molecular docking (分子ドッキング)
- scRNA-seq annotation (シングルセルRNA-seqアノテーション)

etc.



# Biomni

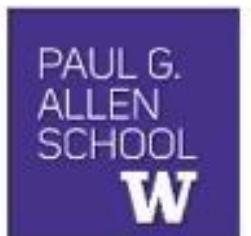


**Genentech**  
A Member of the Roche Group

**Arc Institute**



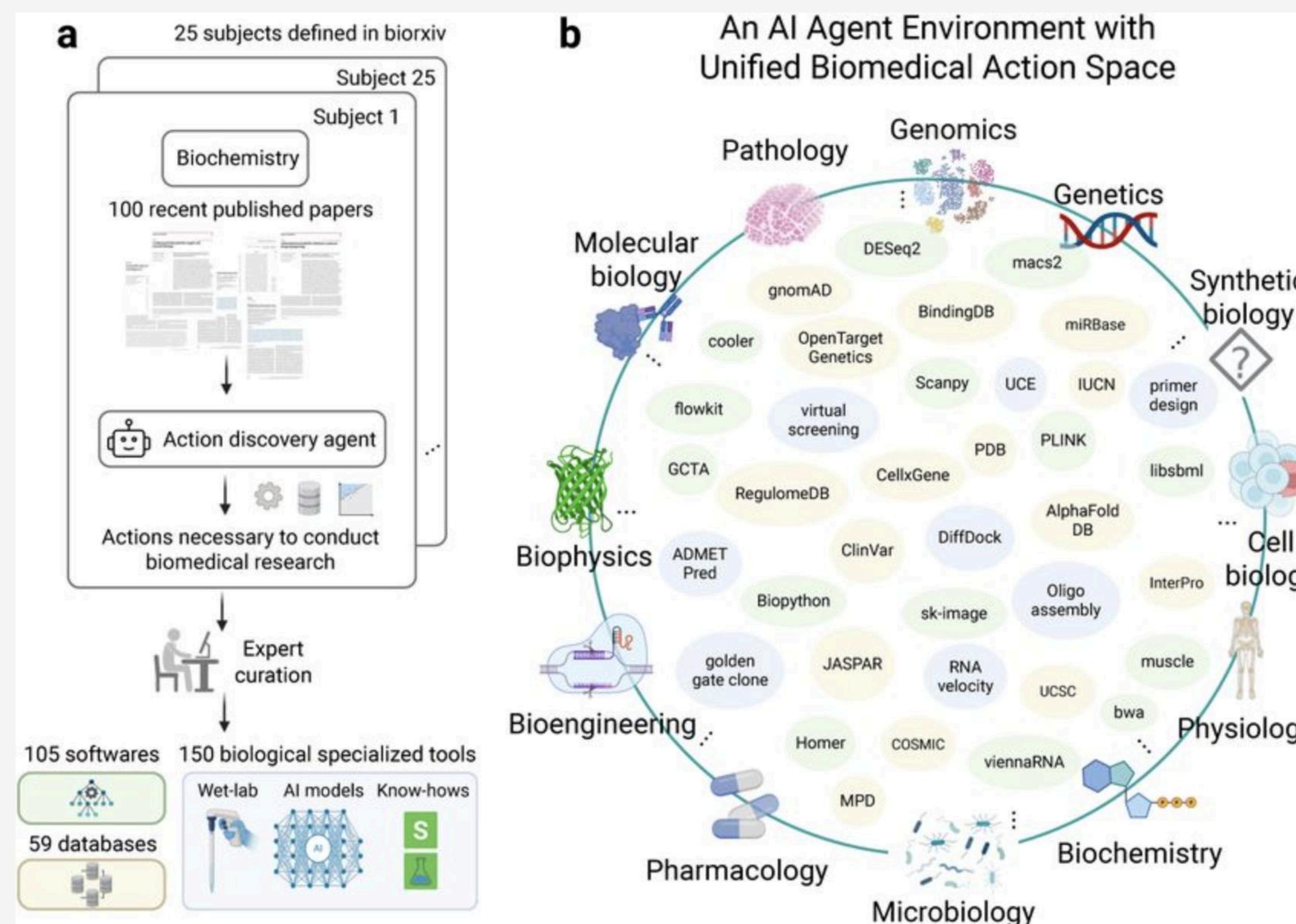
**PRINCETON**  
UNIVERSITY



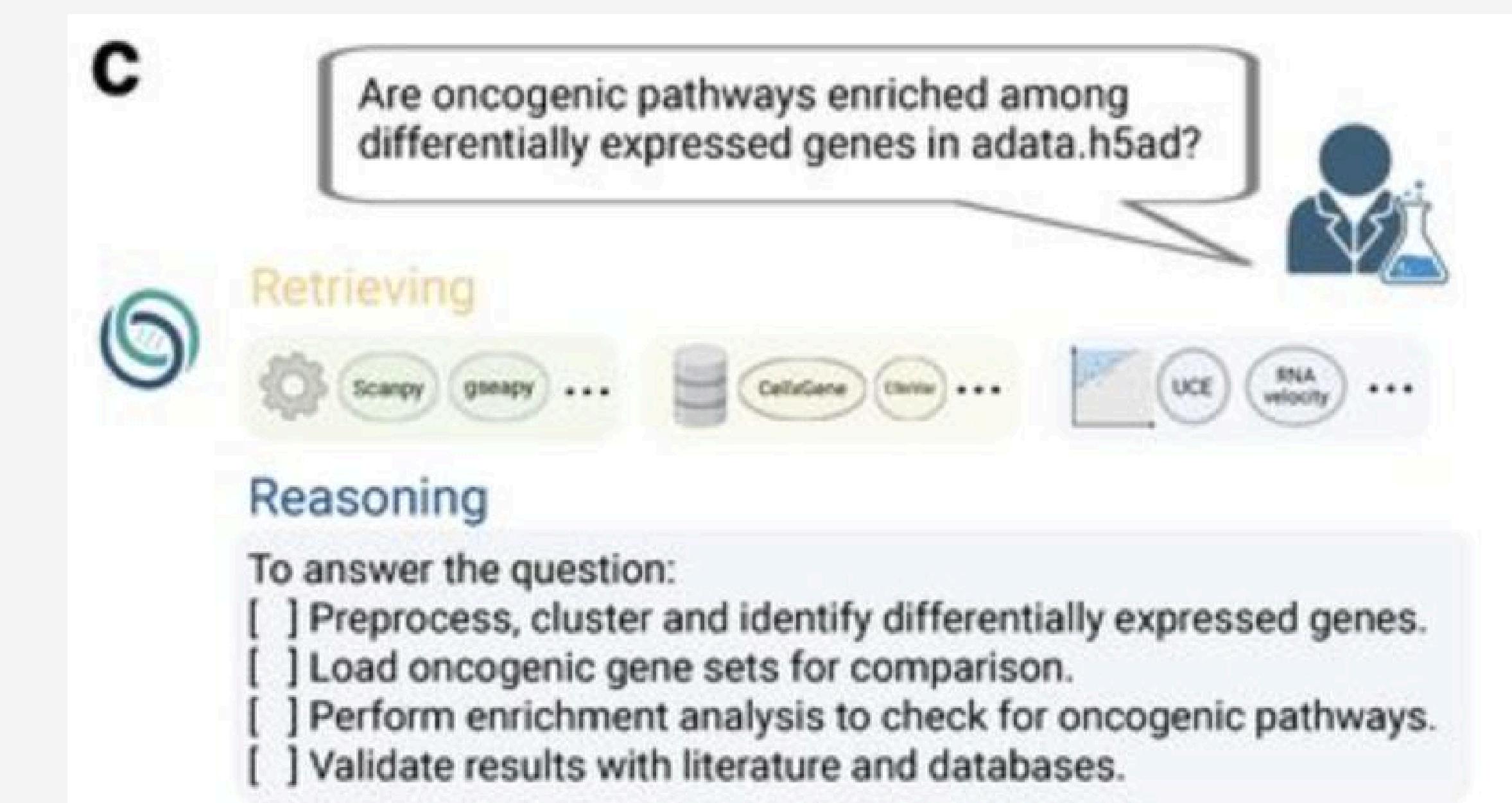
**[biomni.stanford.edu](http://biomni.stanford.edu)**

# Key components

**Biomni-E1:** バイオメディカルエージェントが利用可能な統一された環境。さまざまなデータベースやツールへのアクセスが可能な環境。

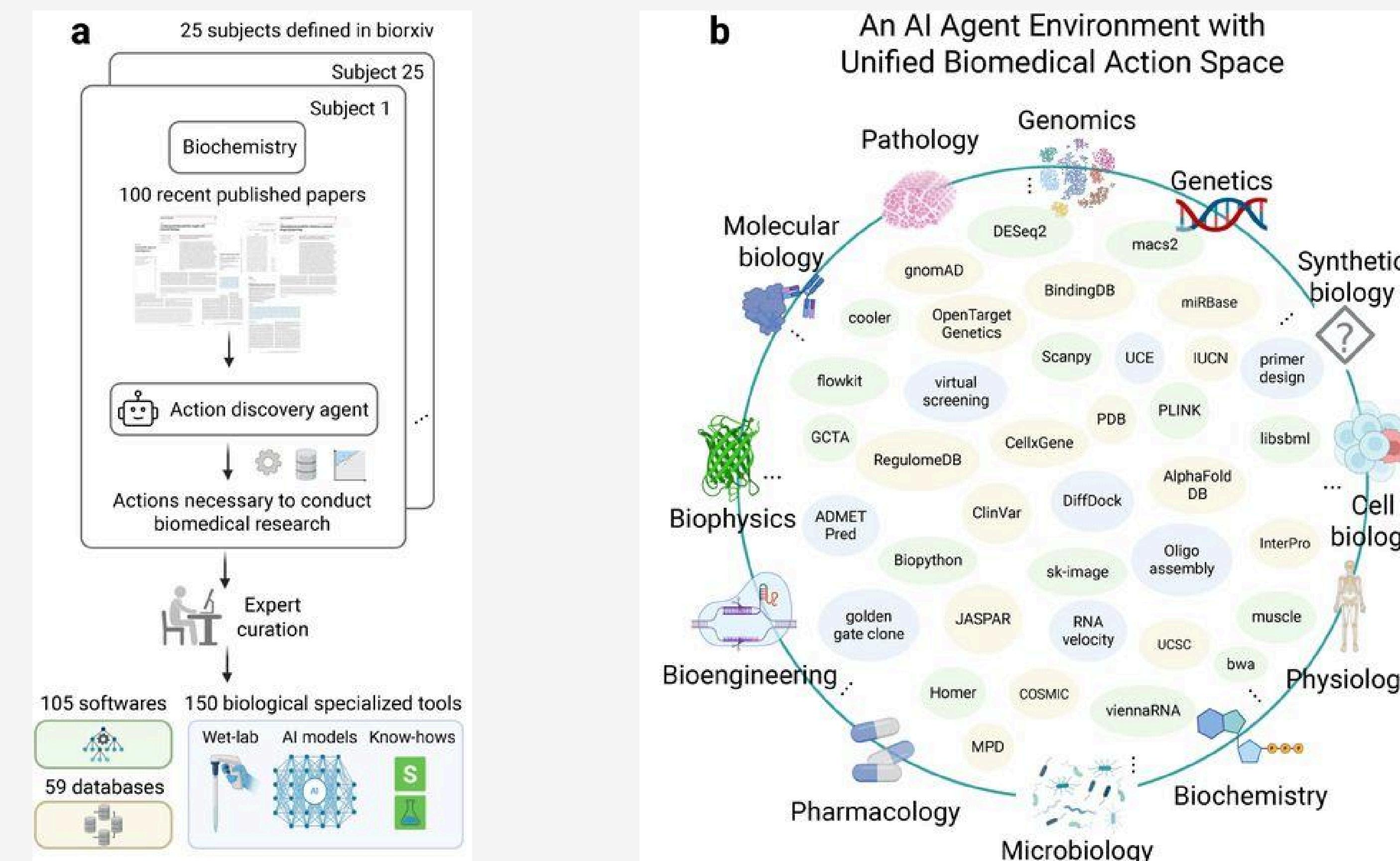


**Biomni-A1:** 広範なバイオメディカル領域にまたがるタスクを（あらかじめ定められたテンプレートなしに）遂行可能な汎用的なエージェント。



特徴① 大量のツール、データベース、ソフトウェアを利用可能

25分野の文献からよく使われているツール、データベース、ソフトウェアを抽出  
エージェントが利用可能なアクション空間（conda環境）Biomni-E1を作成



# 特徴① 大量のツール、データベース、ソフトウェアを利用可能

## ツール (Tools) - 150個

- ・ 特定の生物医学タスクを実行するための専門的な機能やアルゴリズム
- ・ 具体例：
  - ・ BLAST検索を実行する機能
  - ・ タンパク質の3D構造を予測するAlphaFold
  - ・ 遺伝子オントロジー (GO) 解析ツール
  - ・ 系統樹作成アルゴリズム
  - ・ プライマー設計ツール
- ・ 特徴：
  - ・ 特定の目的に特化
  - ・ 入力→処理→出力が明確
  - ・ Biomniが直接呼び出して使用

# 特徴① 大量のツール、データベース、ソフトウェアを利用可能

## データベース (Databases) - 59個

- 生物医学情報を保存・提供するデータリポジトリ
- 具体例：
  - PDB：タンパク質3D構造データ
  - GenBank：DNA配列データ
  - ClinVar：臨床的変異情報
  - KEGG：代謝経路情報
  - DrugBank：薬物情報
- 特徴：
  - 情報の保存と検索が主目的
  - 定期的に更新される
  - クエリを送って情報を取得

# 特徴① 大量のツール、データベース、ソフトウェアを利用可能

## ソフトウェア (Software) - 105個

- より広範な機能を持つ汎用的なプログラム環境
- 具体例：
  - R : 統計解析言語環境
  - Bioconductor : Rの生物情報学パッケージ群
  - samtools : 配列データ処理ソフト
  - GATK : ゲノム解析ツールキット
  - ImageJ : 画像解析ソフト
- 特徴：
  - 複数の機能を含む統合環境
  - コマンドラインやスクリプトで操作
  - 他のツールの基盤となることが多い

## 特徴② 動的なツール選定に基づくプロンプト更新

1. 全リソースの収集: 利用可能な全ツール、データレイク項目、ライブラリを収集
2. プロンプトベースの選別: ユーザーのクエリに関連するリソースのみを選択
3. システムプロンプトの更新: 選別されたリソースのみでシステムプロンプトを更新

### 初期プロンプト

#### Function Dictionary:

In your code, you will need to import the function location using the following dictionary of functions:

biomni.tool.genomics: [tool1, tool2, tool3, ...]

biomni.tool.pharmacology: [tool4, tool5, tool6, ...]

...

### ツール選定後のプロンプト

#### Function Dictionary:

Based on your query, I've identified the following most relevant functions that you can use in your code:

biomni.tool.genomics: [tool1, tool3] # クエリに関連するもののみ

IMPORTANT: When using any function, you MUST first import it from its module.

# 特徴③ CodeAct [Wang+, arXiv:2402.01030] の採用

エージェントにコードを書かせて実行させる

Instruction: Determine the most cost-effective country to purchase the smartphone model "CodeAct 1". The countries to consider are the USA, Japan, Germany, and India.

Available APIs

- [1] `lookup_rates(country: str) -> (float, float)`
- [2] `convert_and_tax(price: float, exchange_rate: float, tax_rate: float) -> float`
- [3] `estimate_final_price(converted_price: float, shipping_cost: float) -> float`
- [4] `lookup_phone_price(model: str, country: str) -> float`
- [5] `estimate_shipping_cost(destination_country: str) -> float`

**LLM Agent using [Text/JSON] as Action**

**CodeAct: LLM Agent using [Code] as Action**

**Fewer Actions Required!**

**Control & Data Flow of Code**  
**Simplifies Complex Operations**

**Re-use 'min' Function from Existing Software Infrastructures (Python library)**

Think I should calculate the phone price in USD for each country, then find the most cost-effective country.

Action Text: `lookup_rates, Germany`  
JSON: `{"tool": "lookup_rates", "country": "Germany"}`

Environment 1.1, 0.19

Action Text: `lookup_phone_price, CodeAct 1, Germany`  
JSON: `{"tool": "lookup_phone_price", "model": "CodeAct 1", "country": "Germany"}`

Environment 700

Action Text: `convert_and_tax, 700, 1.1, 0.19`  
JSON: `{"tool": "convert_and_tax", "price": 700, "exchange_rate": 1.1, "tax_rate": 0.19}`

Environment 916.3

[... interactions omitted (look up shipping cost and calculate final price) ...]

Action Text: `lookup_rates, Japan`  
JSON: `{"tool": "lookup_rates", "country": "Japan"}`

[... interactions omitted (calculate final price for all other countries) ...]

Response The most cost-effective country to purchase the smartphone model is Japan with price 904.00 in USD.

Think I should calculate the phone price in USD for each country, then find the most cost-effective country.

Action countries = ['USA', 'Japan', 'Germany', 'India']  
final\_prices = {}  
  
for country in countries:  
 exchange\_rate, tax\_rate = lookup\_rates(country)  
 local\_price = lookup\_phone\_price("CodeAct 1", country)  
 converted\_price = convert\_and\_tax(  
 local\_price, exchange\_rate, tax\_rate  
 )  
 shipping\_cost = estimate\_shipping\_cost(country)  
 final\_price = estimate\_final\_price(converted\_price, shipping\_cost)  
 final\_prices[country] = final\_price  
  
most\_cost\_effective\_country = min(final\_prices, key=final\_prices.get)  
most\_cost\_effective\_price = final\_prices[most\_cost\_effective\_country]  
print(most\_cost\_effective\_country, most\_cost\_effective\_price)

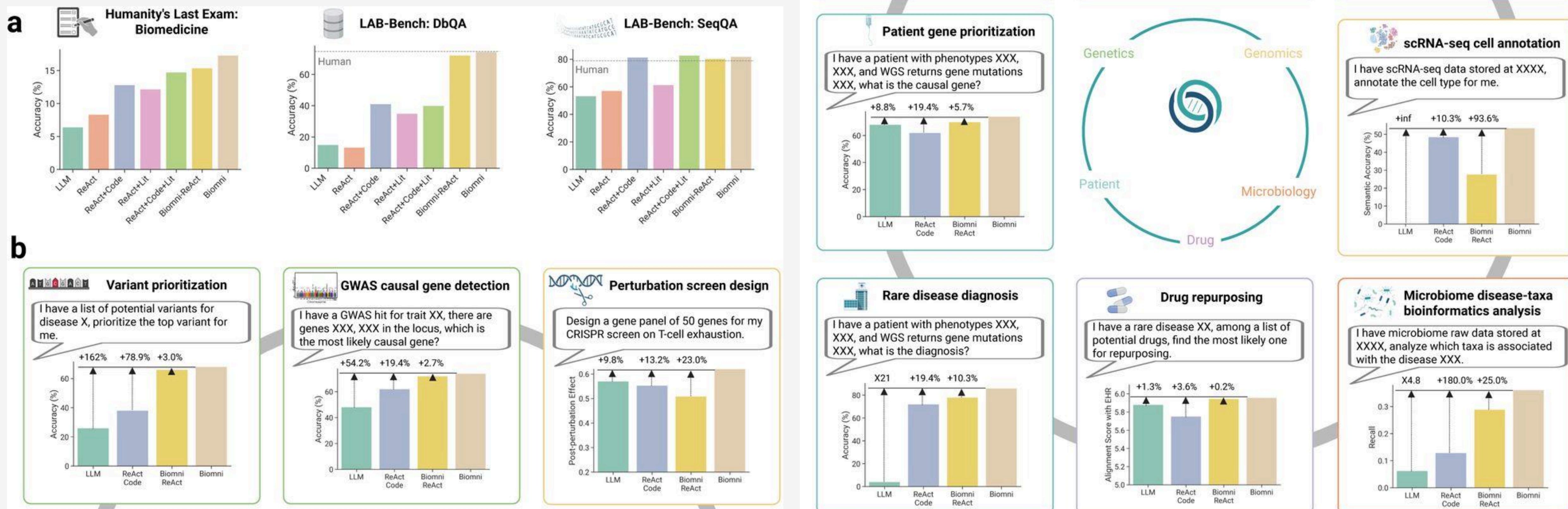
Environment 1.1, 0.19

Response The most cost-effective country to purchase the smartphone model is Japan with price 904.00 in USD.

# 評価

## a. 主要なベンチマーク

## b. 新たに作成した「実世界の研究目標を反映するような」ベンチマーク



# AIエージェントをどう評価するか

ライフサイエンスにおける  
エージェント研究は  
3つの段階で評価

1. 既存ベンチマークでの評価
2. 独自ベンチマークでの評価
3. ウェット実験で評価

# GPQA: A Graduate-Level Google-Proof Q&A Benchmark

## Chemistry (general)

A reaction of a liquid organic compound, which molecules consist of carbon and hydrogen atoms, is performed at 80 centigrade and 20 bar for 24 hours. In the proton nuclear magnetic resonance spectrum, the signals with the highest chemical shift of the reactant are replaced by a signal of the product that is observed about three to four units downfield. Compounds from which position in the periodic system of the elements, which are also used in the corresponding large-scale industrial process, have been mostly likely initially added in small amounts?

- A) A metal compound from the fifth period.
- B) A metal compound from the fifth period and a non-metal compound from the third period.
- C) A metal compound from the fourth period.
- D) A metal compound from the fourth period and a non-metal compound from the second period.

## Organic Chemistry

Methylcyclopentadiene (which exists as a fluxional mixture of isomers) was allowed to react with methyl isoamyl ketone and a catalytic amount of pyrrolidine. A bright yellow, cross-conjugated polyalkenyl hydrocarbon product formed (as a mixture of isomers), with water as a side product. These products are derivatives of fulvene. This product was then allowed to react with ethyl acrylate in a 1:1 ratio. Upon completion of the reaction, the bright yellow color had disappeared. How many chemically distinct isomers make up the final product (not counting stereoisomers)?

- A) 2
- B) 16
- C) 8
- D) 4

## Genetics

If a sperm from species A is injected into an egg from species B and both species have the same number of chromosomes, what would be the main cause of the resulting zygote mortality?

- A) Species specific zona pellucida proteins on the egg cannot bind sperms from a different species.
- B) Epistatic interactions between the genes of different species
- C) Chromosomal incompatibilities will cause failure of meiosis leading to death of zygote.
- D) Chromosomal recombination will not occur in different species.

## Molecular Biology

A scientist studies the stress response of barley to increased temperatures and finds a protein which contributes to heat tolerance through the stabilisation of cell membrane. The scientist is very happy and wants to create a heat-tolerant cultivar of diploid wheat. Using databases, they find a heat tolerance protein homologue and start analysing its accumulation under heat stress. Soon enough, the scientist discovers this protein is not synthesised in the wheat cultivar they study. There are many possible reasons for such behaviour, including:

- A) A miRNA targets the protein, which makes exonucleases cut it immediately after the end of translation and before processing in ER
- B) Trimethylation of lysine of H3 histone in position 27 at the promoter of the gene encoding the target protein
- C) A stop-codon occurs in the 5'-UTR region of the gene encoding the target protein
- D) The proteolysis process disrupts a quaternary structure of the protein, preserving only a tertiary structure

## Astrophysics

Astronomers are studying a star with a Teff of approximately 6000 K. They are interested in spectroscopically determining the surface gravity of the star using spectral lines ( $EW < 100$  mA) of two chemical elements, El1 and El2. Given the atmospheric temperature of the star, El1 is mostly in the neutral phase, while El2 is mostly ionized. Which lines are the most sensitive to surface gravity for the astronomers to consider?

大学院レベルの科学問題を集めた高難度ベンチマーク

形式：多肢選択式

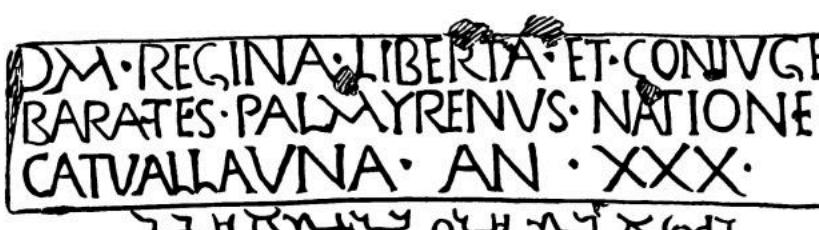
対象分野：生物学・物理学・化学

問題数：448問

# HLE: Humanity's Last Exam

**Classics**

Question:



Here is a representation of a Roman inscription, originally found on a tombstone. Provide a translation for the Palmyrene script.

A transliteration of the text is provided: RGYN BT HRY BR CT HBL

👤 Henry T  
Merton College, Oxford

**Ecology**

Question:

Hummingbirds within Apodiformes uniquely have a bilaterally paired oval bone, a sesamoid embedded in the caudolateral portion of the expanded, cruciate aponeurosis of insertion of m. depressor caudae. How many paired tendons are supported by this sesamoid bone? Answer with a number.

👤 Edward V  
Massachusetts Institute of Technology

**Mathematics**

Question:

The set of natural transformations between two functors  $F, G : C \rightarrow D$  can be expressed as the end

$$Nat(F, G) \cong \int_A Hom_D(F(A), G(A))$$

Define set of natural cotransformations from  $F$

$$CoNat(F, G) \cong \int_A Hom_D(F(A), G(A))$$

Let:

- $F = B_*(\Sigma_4)_*$  be the under  $\infty$ -category of delooping of the symmetric group  $\Sigma_4$  on 4 letter  $\infty$ -simplex  $*$  of  $B_*\Sigma_4$ .
- $G = B_*(\Sigma_7)_*$  be the under  $\infty$ -category of the symmetric group  $\Sigma_7$  on 7 letters under  $t$  of  $B_*\Sigma_7$ .

How many natural cotransformations are there

👤 Emily S  
University of São Paulo

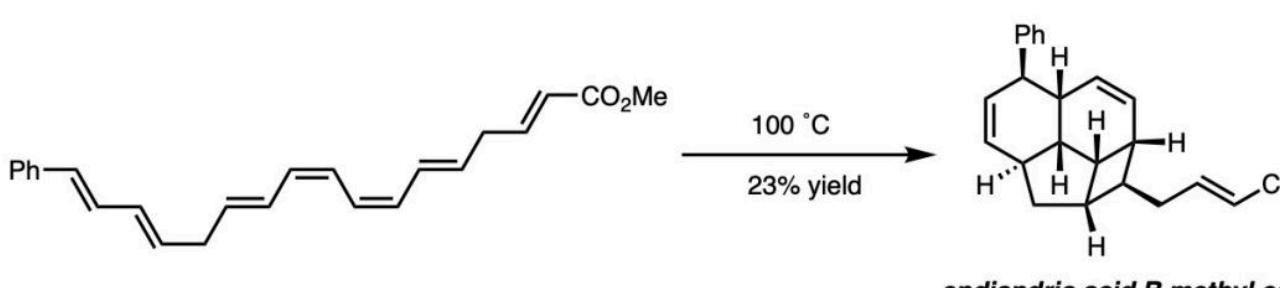
**Computer Science**

Question:

Let  $G$  be a graph. An edge-indicator of  $G$  is a function  $a : \{0, 1\} \rightarrow$

**Chemistry**

Question:



The reaction shown is a thermal pericyclic cascade that converts the starting heptaene into endiandric acid B methyl ester. The cascade involves three steps: two electrocyclizations followed by a cycloaddition. What types of electrocyclizations are involved in step 1 and step 2, and what type of cycloaddition is involved in step 3?

Provide your answer for the electrocyclizations in the form of  $[n\pi]$ -con or  $[n\pi]$ -dis (where  $n$  is the number of  $\pi$  electrons involved, and whether it is conrotatory or disrotatory), and your answer for the cycloaddition in the form of  $[m+n]$  (where  $m$  and  $n$  are the number of atoms on each component).

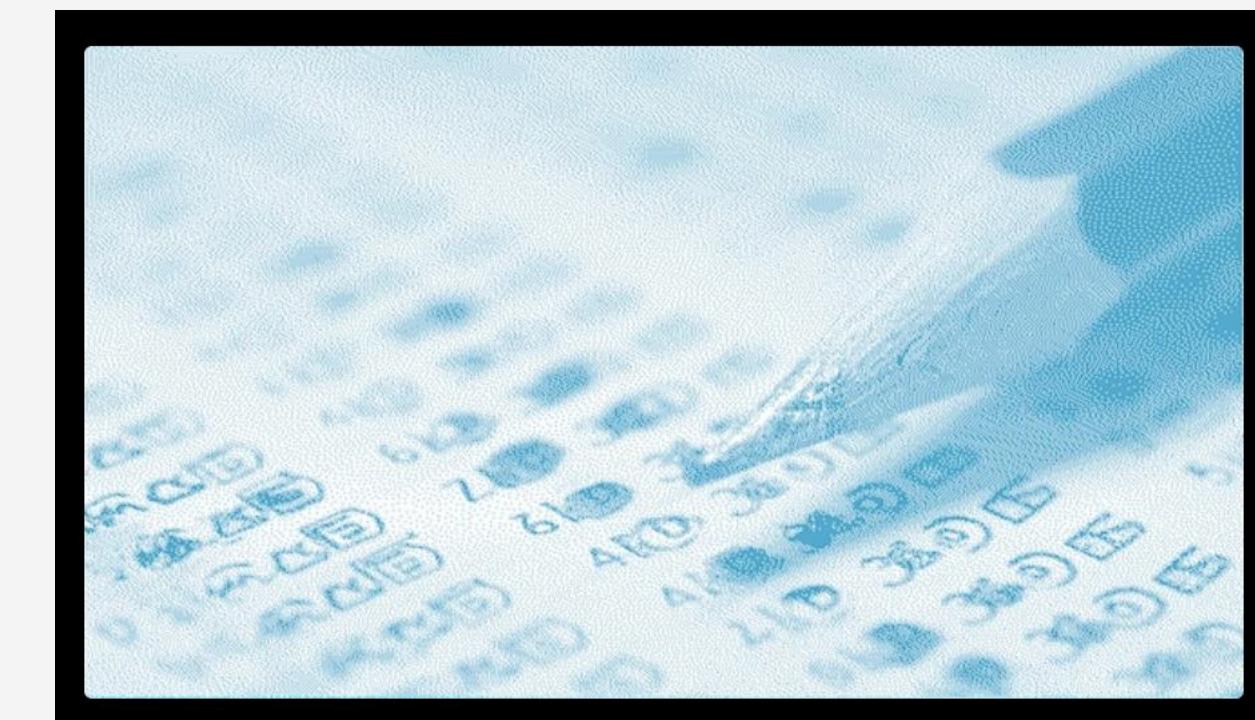
👤 Noah B  
Stanford University

LLMの限界に挑むために設計された  
マルチモーダルな学術ベンチマーク

形式：多肢選択式 + 短答式

対象分野：数学、人文学、自然科学など、幅広い学問分野

問題数：2,500問

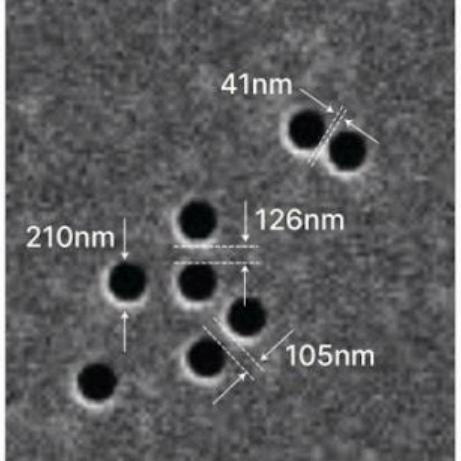


Research July 23, 2025

About 30% of Humanity's Last Exam chemistry/biology answers are likely wrong

# LAB-Bench: Language Agent Biology Benchmark

**FigQA**



Query: What is the most likely width of the entire image?  
A) 4.5 uM B) 0.5 uM  
C) 2.75 uM D) 1.03 uM

Answer: "Based on the observations...[ANSWER]B[/ANSWER]"

**TableQA**

Annotation	Type	In_group	Total	Fraction	Fisher's test
intergenic	Common	9686	30681	0.316	n.s.
	LowFreq	412	1228	0.336	n.s.
missense_variant	Common	5297	30681	0.173	n.s.
	LowFreq	204	1228	0.166	n.s.
synonymous_variant	Common	15299	30681	0.499	n.s.
	LowFreq	594	1228	0.484	n.s.
All variants					
intergenic	Common	332	918	0.362	n.s.
	LowFreq	40	104	0.385	n.s.
missense_variant	Common	188	918	0.205	n.s.
	LowFreq	20	104	0.192	n.s.
synonymous_variant	Common	383	918	0.417	n.s.
	LowFreq	43	104	0.413	n.s.
SNP-eQTLs					

Query: Which annotation in the SNP-eQTLs section has the lowest fraction of 'In\_group' occurrences?  
A) missense\_variant, LowFreq B) intergenic, Common C) synonymous\_variant, Common D) missense\_variant, Common

Answer: "In the SNP-eQTLs section... [ANSWER]D[/ANSWER]"

**DbQA**

Query: Which of the following genes is a computationally predicted human gene target of the miRNA MIR6810\_3P according to miRDB v6.0? A) RBM3 B) RABGEF1 C) DCANP1 D) Insufficient... E) WIPF2

Ans: "[ANSWER]D[/ANSWER]"

**ProtocolQA**

Protocol: Generation of Sequencing Libraries for ...  
Query: After running the first round of PCR in Step 38, you notice very low RNA quantity. What could you do to improve the quantity?  
A) in Step 38 Use Primer LibAmp\_F (1 uM) 1.25uL  
B) In Step 38 Use Primer LibAmp\_F (10 uM) 12.5uL  
C) In Step 26c dry the RNA pellet for 60 minutes  
D) Insufficient...  
E) In Step 26c dry the RNA pellet for 10 min

Ans: "Given the...would be:  
[ANSWER]B[/ANSWER]"

**SuppQA**

Title: Development and validation of AI/ML derived splice-switching oligonucleotides  
Query: How many data files for breast cancer RNA-seq from TCGA were used to assembly mRNA contigs? A) 1,522 B) Insufficient... C) 1,131 D) 1,743 E) 1,252

Ans: "[ANSWER]B[/ANSWER]"

**SeqQA**

Query: What is the AA encoded at position 15 in the longest ORF contained within the sequence <REDACTED>? A) Insufficient... B) Pro C) Asparagine D) H E) Leucine

Ans: "To find the correct...[ANSWER]E[/ANSWER]"

**LitQA**

Query: Active olfactory receptor genes increase their contacts with greek island regions by what factor in mouse olfactory neurons? A) 2.0 fold B) 1.7 fold C) 3.0 fold D) 2.7 fold E) Insufficient... F) 27 fold

Ans: "[ANSWER]B[/ANSWER]"

**Cloning Scenarios**

Query: I have three plasmids with sequences pLAB-CTU: <REDACTED>, pLAB-gTU2E: <REDACTED>, pLAB-CH3: <REDACTED>. I combined all three plasmids together in a Golden Gate cloning reaction with Esp3I. The resulting plasmid expresses Cas9 protein as well as a targeting gRNA. What gene does the gRNA target?  
A) Insufficient...  
B) Yeast SCL1 C) Human PRC3 D) Human SCL1  
E) Yeast PRC3

Ans: "To determine the gene...answer is:  
[ANSWER]D[/ANSWER]"

科学研究の現場でLLMが本当に使えるかを問う  
実務直結型の生物学ベンチマーク

形式：多肢選択式

対象分野：主に分子生物学および生命科学

問題数：2,457問

8種の情報の取り扱いを評価

- FigQA: 図
- TableQA : 表データ
- DbQA: DB
- SuppQA: 補足資料
- ProtocolQA: 実験プロトコル
- SeqQA: DNAやタンパク質などの配列
- LitQA: 論文

# AIエージェントをどう評価するか

ライフサイエンスにおける  
エージェント研究は  
3つの段階で評価

1. 既存ベンチマークでの評価
2. 独自ベンチマークでの評価
3. ウェット実験で評価

# AIエージェントをどう評価するか

ライフサイエンスにおける  
エージェント研究は  
3つの段階で評価

1. 既存ベンチマークでの評価
2. 独自ベンチマークでの評価
3. ウェット実験で評価

## Biomni (Stanford)

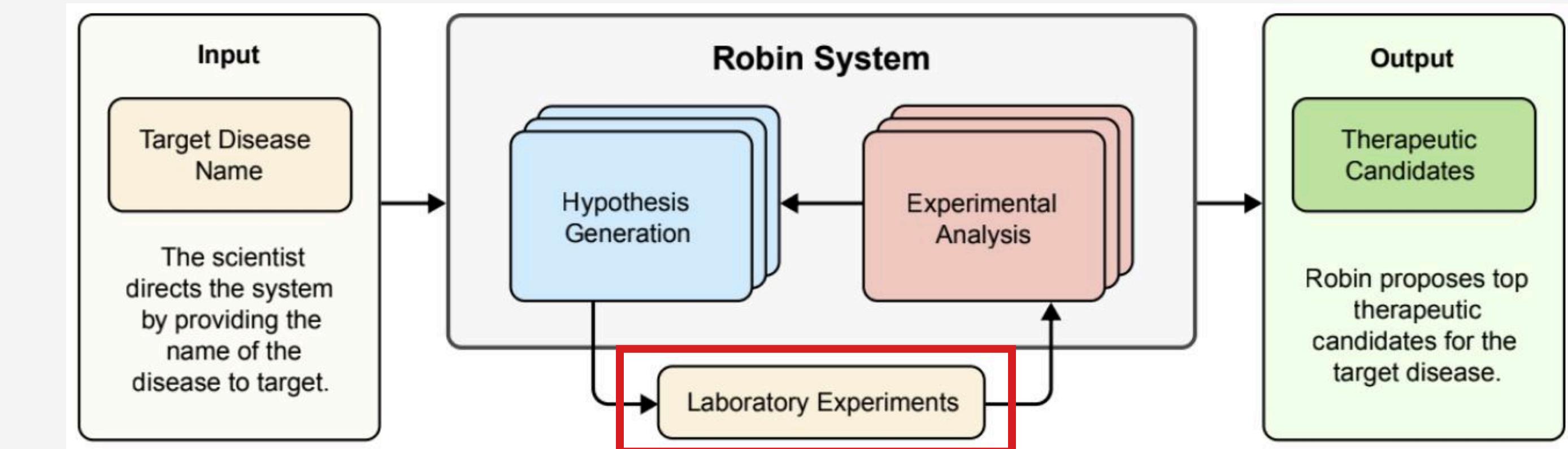
現実的な研究タスクにおける汎化性能を評価  
するため、遺伝学、ゲノミクス、微生物学、薬  
理学、臨床医学にまたがる8つの新しいバイオ  
メディカルベンチマークを作成

## Robin (FutureHouse)

2024年6月にLAB-Benchを公開  
それを使ってエージェント開発

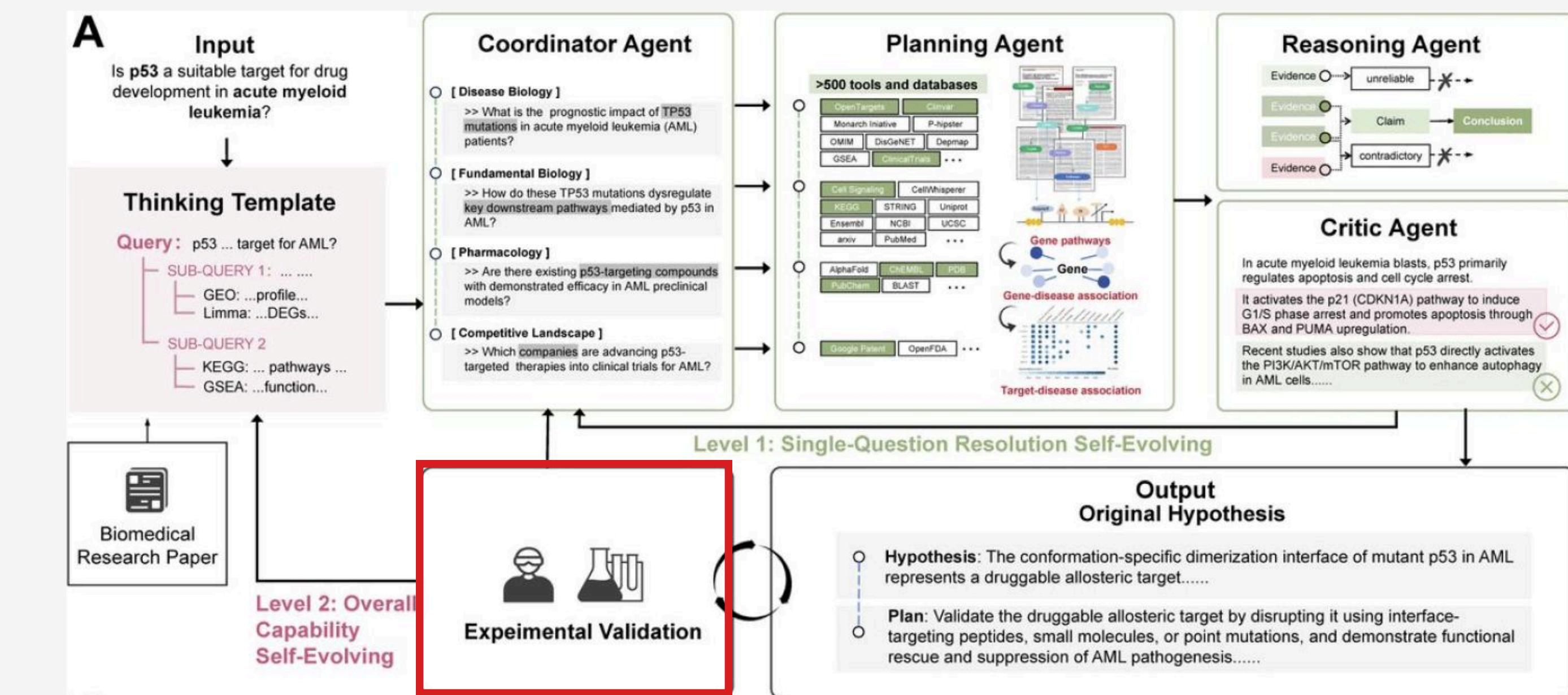
# AIエージェントをどう評価するか

ライフサイエンスにおける  
エージェント研究は  
3つの段階で評価



Robin (FutureHouse) [Ghareeb+, arXiv:2505.13400]

1. 既存ベンチマークでの評価
2. 独自ベンチマークでの評価
3. ウェット実験で評価



OriGene (Lingang Lab) [Zhang+, bioRxiv:2025.06.03.657658]

最後は人間が実験をして  
「有望なものもあった」で終わる論文が多い

有望な仮説・実験条件の生成能力を  
直接的かつ定量的に示した論文は見たことがない

人間レベルには遠く及んでいないと推察

# まとめ

Biomniは汎用バイオメディカルエージェント

汎用性の源は150の専門的ツール、59のデータベース、105のソフトウェアと  
多数のツールが実装されたアクション空間 **Biomni-E1**

多数のツールをエージェントが混乱せず使える秘訣は

- 動的なツール選定によるシステムプロンプトの更新
- CodeActアーキテクチャ

OSSかつツールの追加が容易なため、拡張性が高い

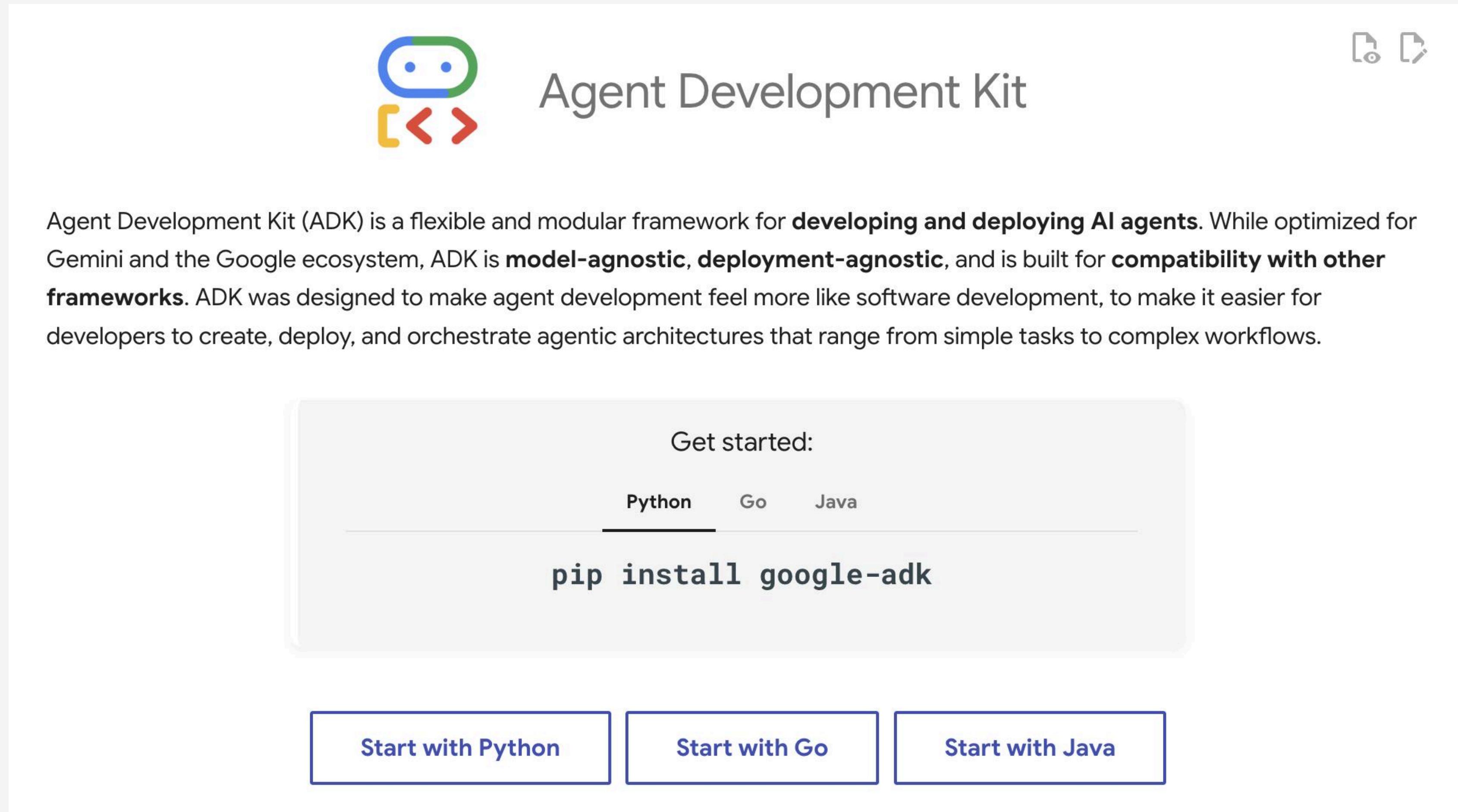
→ 弊社でもクライアントワークや社内研究にカスタマイズしたものを活用

# AI エージェントを開発するには

# ハンズオン

Google Agent Development Kitを使って、文献検索をするエージェントを開発

投影するコードおよびドキュメントはアンケート回答者に配布



The screenshot shows the official documentation for the Agent Development Kit (ADK). The page features a logo with a blue and green rounded rectangle and red brackets (<>). The title "Agent Development Kit" is displayed prominently. A detailed description of ADK is provided, stating it is a flexible and modular framework for developing and deploying AI agents, optimized for Gemini and the Google ecosystem, and is model-agnostic, deployment-agnostic, and compatible with other frameworks. Below this, a "Get started:" section offers links for Python, Go, and Java, with the Python link currently selected. A command-line instruction "pip install google-adk" is shown. At the bottom, three buttons labeled "Start with Python", "Start with Go", and "Start with Java" are available for each respective programming language.

Agent Development Kit

Agent Development Kit (ADK) is a flexible and modular framework for **developing and deploying AI agents**. While optimized for Gemini and the Google ecosystem, ADK is **model-agnostic, deployment-agnostic**, and is built for **compatibility with other frameworks**. ADK was designed to make agent development feel more like software development, to make it easier for developers to create, deploy, and orchestrate agentic architectures that range from simple tasks to complex workflows.

Get started:

Python   Go   Java

`pip install google-adk`

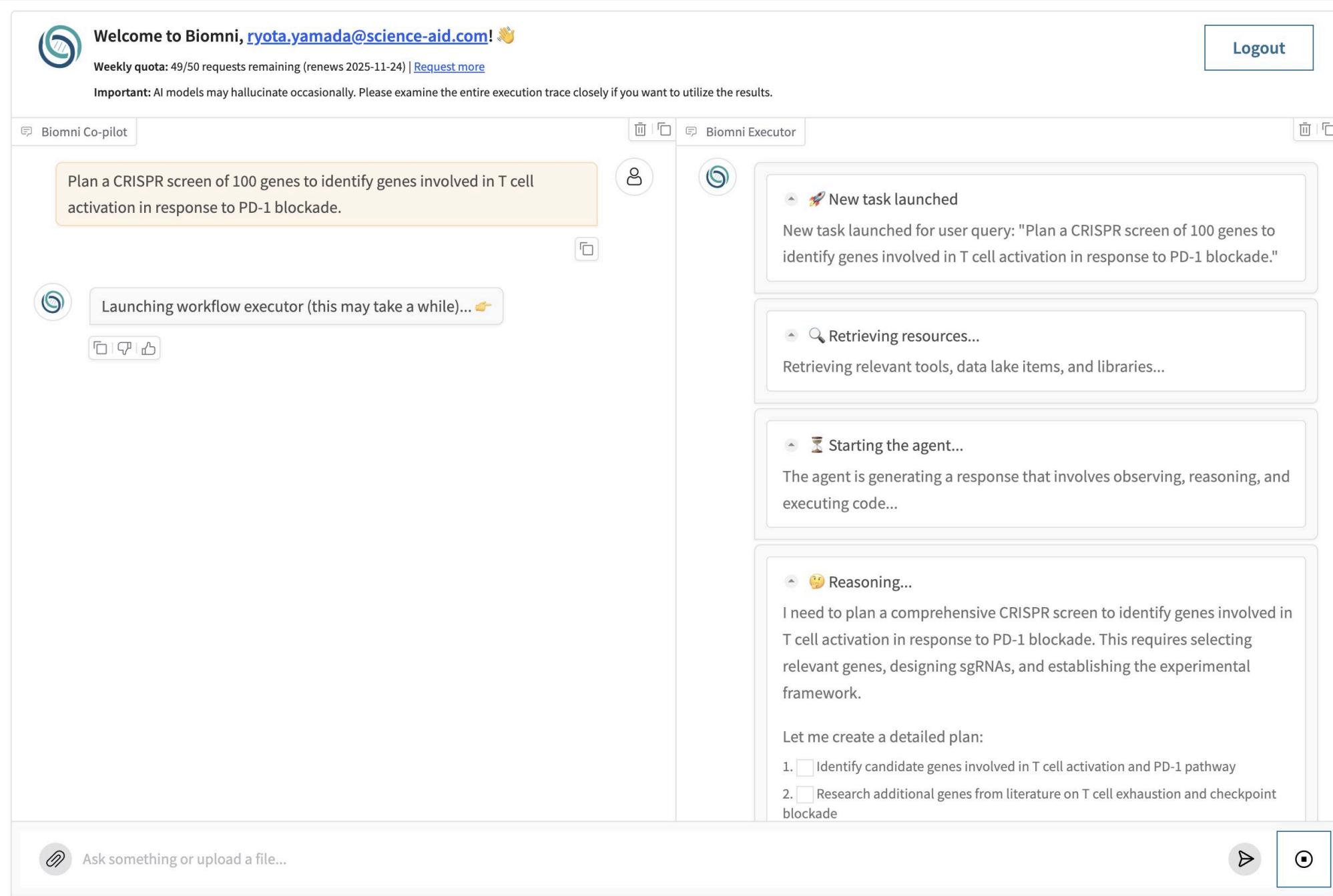
Start with Python   Start with Go   Start with Java

おすすめのNext Step

# Next Step: Lv. 1

研究特化型のエージェントサービスを利用して、仮説生成、データ分析などを試してみる

## Biomni



Welcome to Biomni, ryota.yamada@science-aid.com! 🎉

Weekly quota: 49/50 requests remaining (renews 2025-11-24) | Request more

Important: AI models may hallucinate occasionally. Please examine the entire execution trace closely if you want to utilize the results.

Biomni Co-pilot      Biomni Executor

Logout

Plan a CRISPR screen of 100 genes to identify genes involved in T cell activation in response to PD-1 blockade.

Launching workflow executor (this may take a while)... 🚧

New task launched

New task launched for user query: "Plan a CRISPR screen of 100 genes to identify genes involved in T cell activation in response to PD-1 blockade."

Retrieving resources...

Retrieving relevant tools, data lake items, and libraries...

Starting the agent...

The agent is generating a response that involves observing, reasoning, and executing code...

Reasoning...

I need to plan a comprehensive CRISPR screen to identify genes involved in T cell activation in response to PD-1 blockade. This requires selecting relevant genes, designing sgRNAs, and establishing the experimental framework.

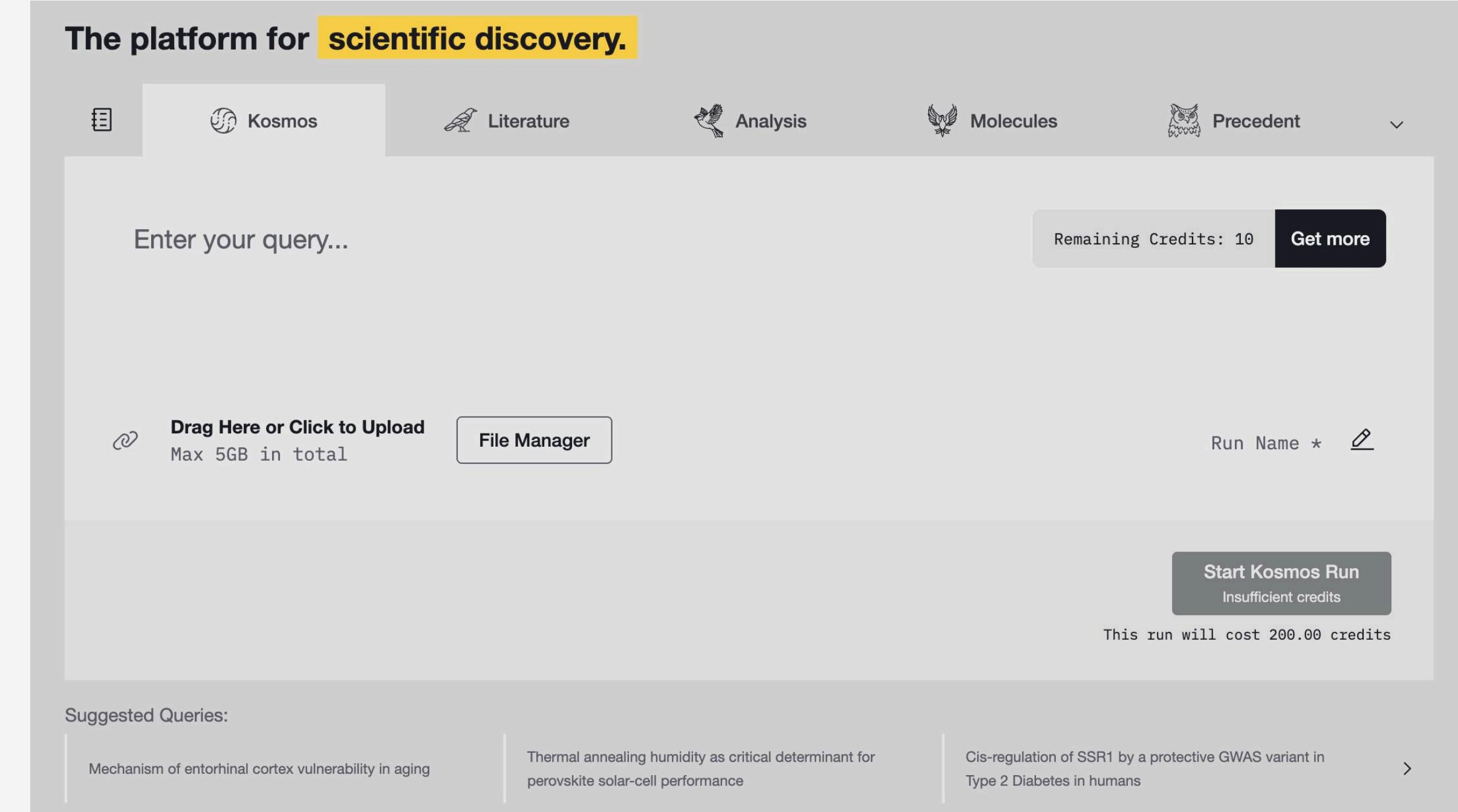
Let me create a detailed plan:

- Identify candidate genes involved in T cell activation and PD-1 pathway
- Research additional genes from literature on T cell exhaustion and checkpoint blockade

Ask something or upload a file...

<https://app.biomni.stanford.edu/>

## Kosmos



The platform for **scientific discovery.**

Kosmos      Literature      Analysis      Molecules      Precedent

Enter your query...

Remaining Credits: 10      Get more

Drag Here or Click to Upload      Max 5GB in total      File Manager

Run Name \*  Run

Start Kosmos Run      Insufficient credits

This run will cost 200.00 credits

Suggested Queries:

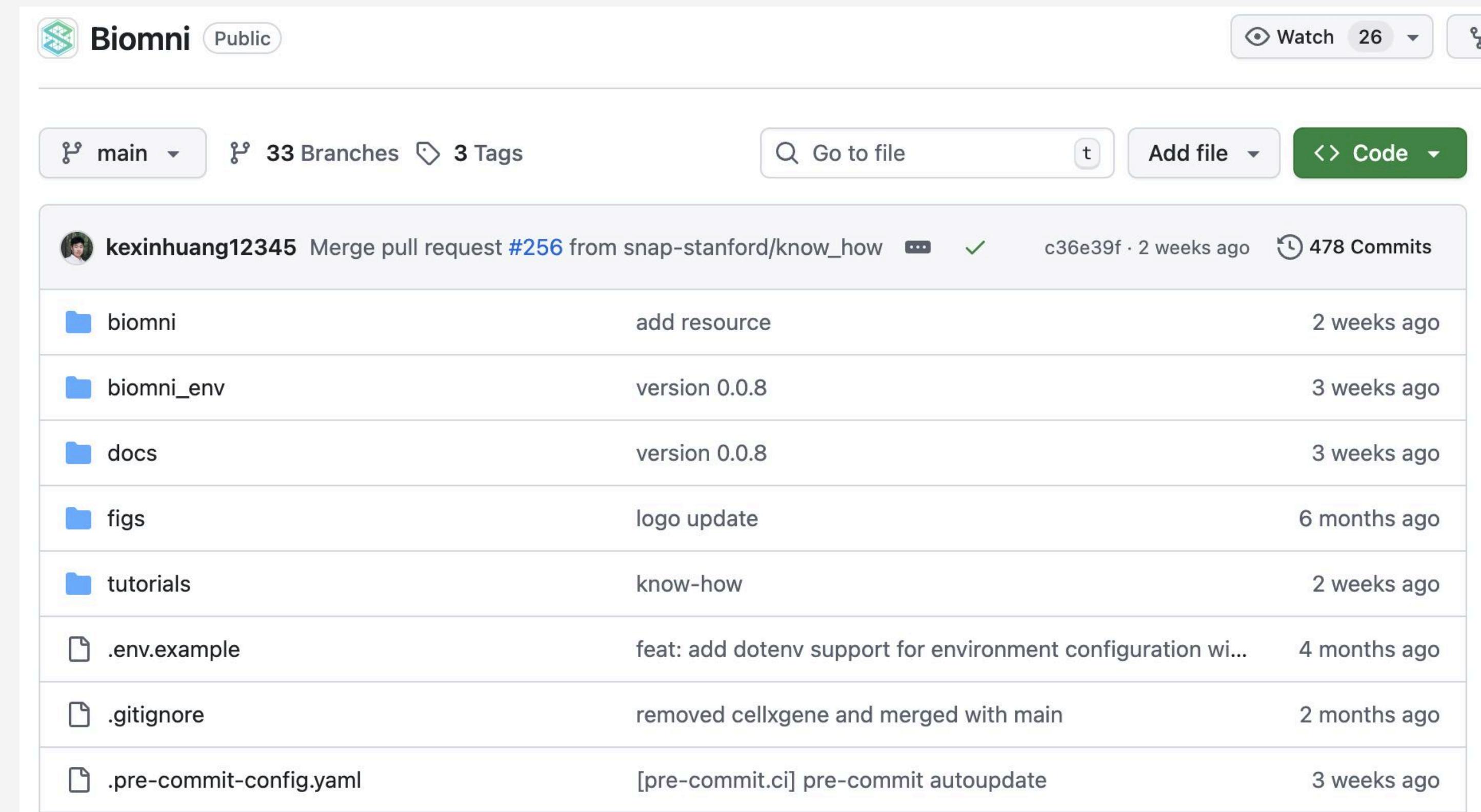
- Mechanism of entorhinal cortex vulnerability in aging
- Thermal annealing humidity as critical determinant for perovskite solar-cell performance
- Cis-regulation of SSR1 by a protective GWAS variant in Type 2 Diabetes in humans

<https://platform.edisonscientific.com/>

# Next Step: Lv. 2

OSSのツールやシステムプロンプトをカスタマイズ

## Biomni



The screenshot shows the GitHub repository for 'Biomni'. The repository is public and has 33 branches and 3 tags. The main branch is selected. The commit history is as follows:

Commit	Message	Time
biomni	add resource	2 weeks ago
biomni_env	version 0.0.8	3 weeks ago
docs	version 0.0.8	3 weeks ago
figs	logo update	6 months ago
tutorials	know-how	2 weeks ago
.env.example	feat: add dotenv support for environment configuration wi...	4 months ago
.gitignore	removed cellxgene and merged with main	2 months ago
.pre-commit-config.yaml	[pre-commit.ci] pre-commit autoupdate	3 weeks ago

<https://github.com/snap-stanford/Biomni>

# Next Step: Lv. 3

Google ADKなどのフレームワークを使ってオリジナルのエージェントを開発

## AI Agent Tutorial (Science Aid)

The screenshot shows a dark-themed AI interface. On the left, a code editor displays the following JSON-like code for an AI agent:

```
Event 1 of 3
SESSION ID e68f9a8f-8ea0-428e-95bd-cf05484d21e0
Event Request Response
literature_survey_agent → pubmed_search
content:
  parts:
    0:
      functionCall:
        id: "call_5usIhi4Rh8v112lVoCNxP1yv"
        args:
          query: "cow milk allergy review 2023 2024 2025"
          max_results: 3
          name: "pubmed_search"
      role: "model"
  partial: false
  usageMetadata:
    candidatesTokenCount: 234
    promptTokenCount: 381
    totalTokenCount: 615
  invocationId: "e-82de153c-e19b-46e1-a99b-50b60a59986a"
  author: "literature_survey_agent"
  actions:
    stateDelta:
    artifactDelta:
    requestedAuthConfigs:
    requestedToolConfirmations:
  longRunningToolIds:
    id: "9ebc6faa-1aa6-46c6-b1a9-6bf5f14a66d1"
    timestamp: 1763439249.427303
    title: "functionCall:pubmed_search"
```

On the right, a message history shows two "pubmed\_search" messages. A large yellow box highlights the text "アンケート回答者に配布" (Distribute to survey respondents). The interface also includes a "Type a Message..." input field and a toolbar with icons for file, message, and video.

最後に

# Science Aidのサポート領域

調査、教育、コンサル、開発まで幅広く支援します

## 研究支援

研究の企画～実験の実施～データの解析まで、研究活動の全フェーズをご支援致します。

## 研究データを活用した 関連業務支援

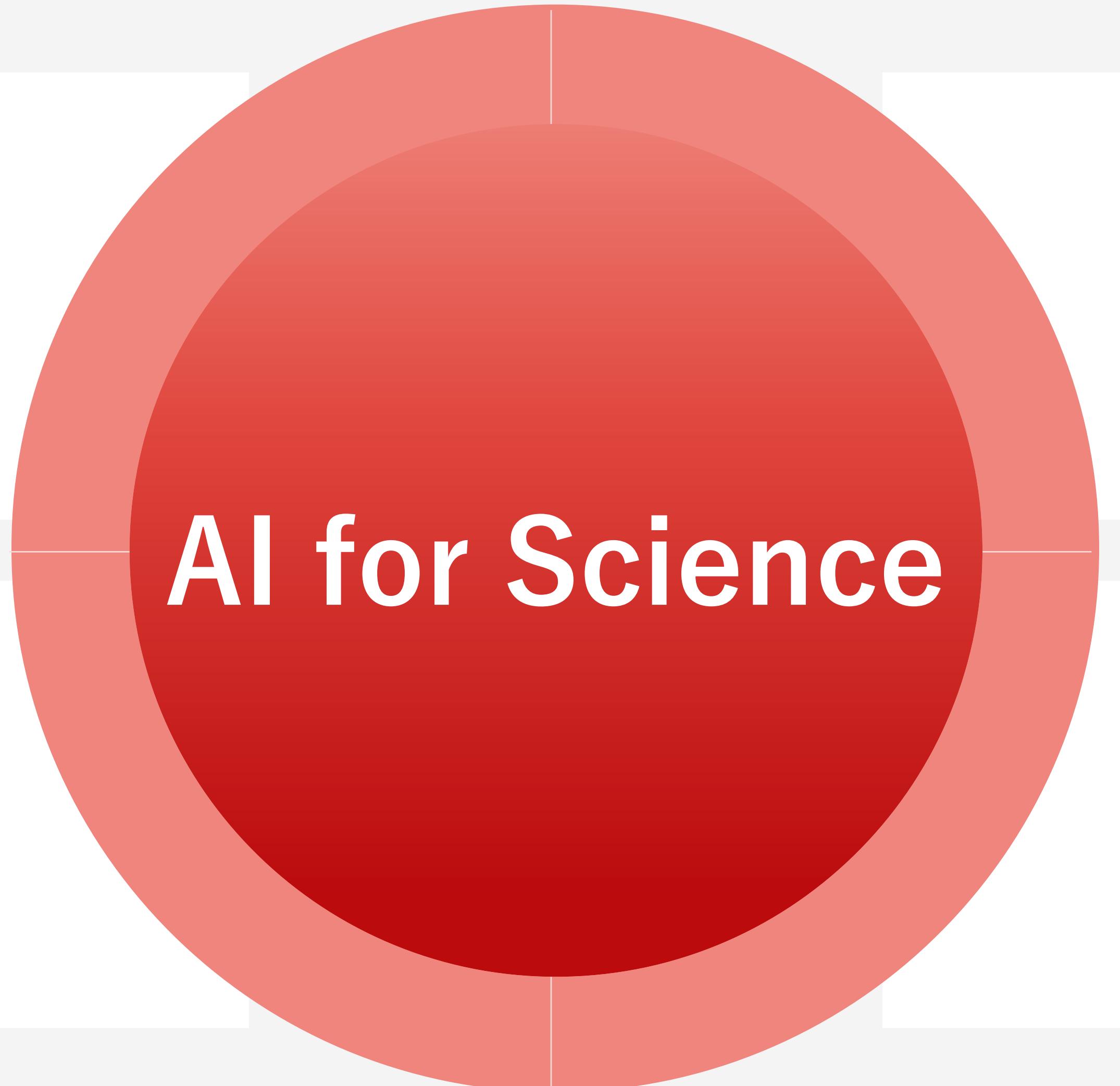
規制対応や各種申請書類の作成など、研究データを用いた関連業務をご支援致します。

## 研究成果の製品化支援

試作・スケールアップから製造・品質管理まで、研究成果の円滑な製品化をご支援致します。

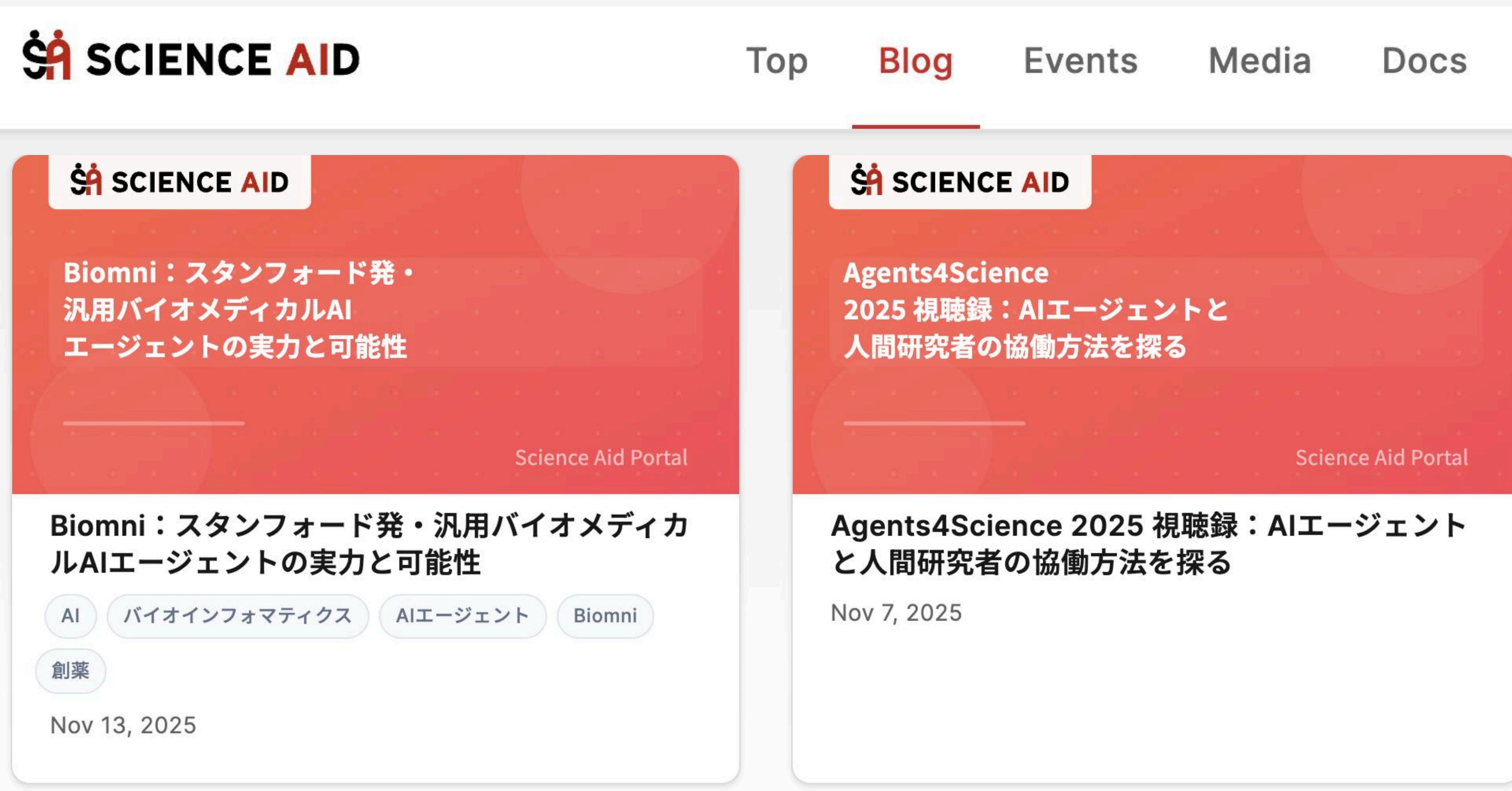
## AI活用研修・ ワークショップ

カスタマイズされた研修プログラムや業務推進のためのワークショップをご提供致します。



AI for Science

# AI for Scienceの情報をまとめたポータルサイトを立ち上げました



The screenshot shows the homepage of the Science Aid portal. The header features the "SCIENCE AID" logo with a red "SA" icon. The navigation menu includes "Top", "Blog" (which is highlighted in red), "Events", "Media", and "Docs".

**Left Column (Blog Post 1):**

- Header:** SCIENCE AID
- Title:** Biomni : スタンフォード発・汎用バイオメディカルAIエージェントの実力と可能性
- Text:** Biomni : スタンフォード発・汎用バイオメディカルAIエージェントの実力と可能性
- Tags:** AI, バイオインフォマティクス, AIエージェント, Biomni, 創薬
- Date:** Nov 13, 2025

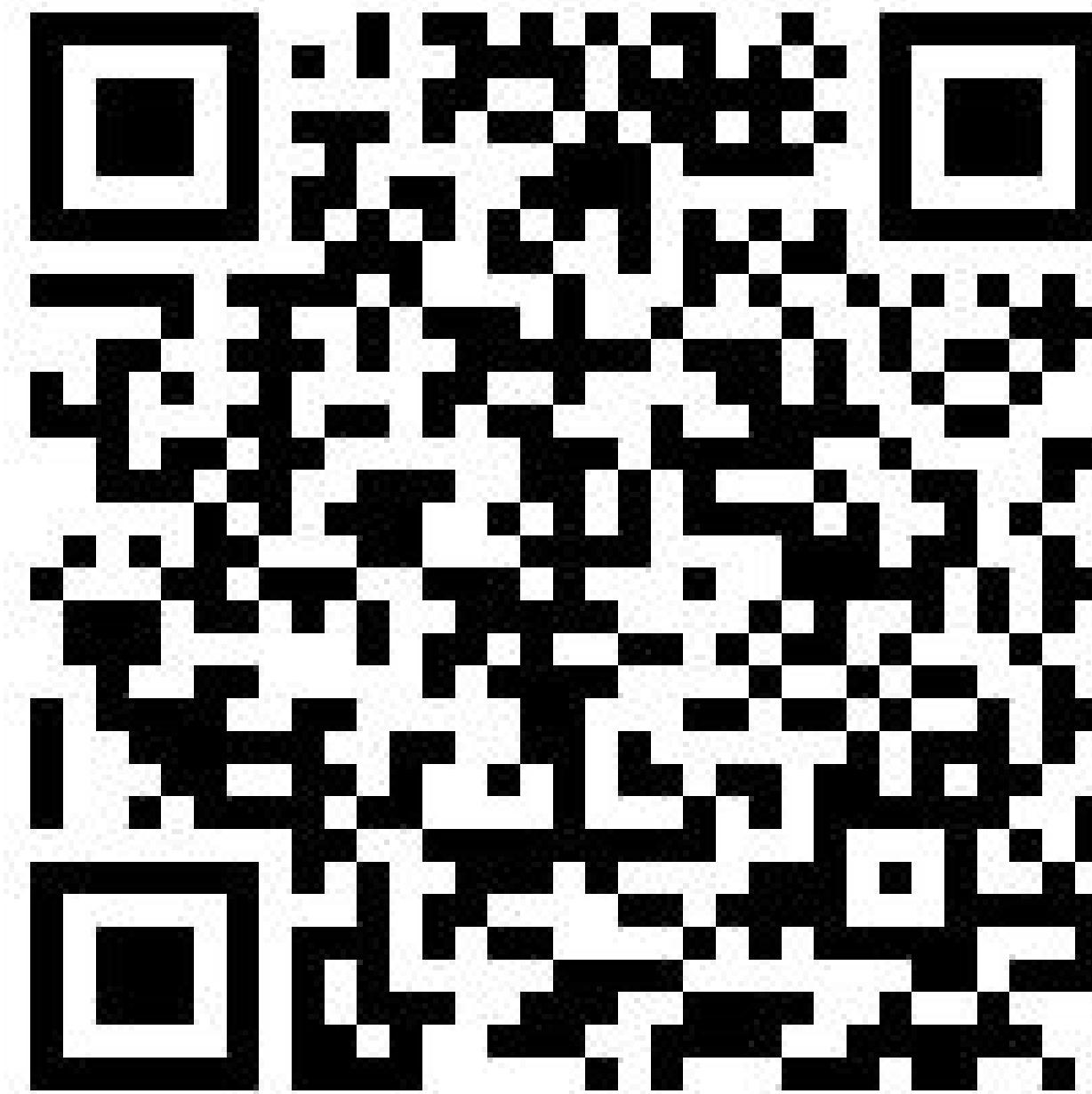
**Right Column (Blog Post 2):**

- Header:** SCIENCE AID
- Title:** Agents4Science 2025 視聴録：AIエージェントと人間研究者の協働方法を探る
- Text:** Agents4Science 2025 視聴録：AIエージェントと人間研究者の協働方法を探る
- Date:** Nov 7, 2025

<https://portal.science-aid.com/>

## アンケート & QA

アンケート回答者にはハンズオン資料を配布します  
ご協力お願ひいたします



[https://rcjbd.share-na2.hsforms.com/2rBEi4\\_xjSX6pN0mb\\_QwGyg](https://rcjbd.share-na2.hsforms.com/2rBEi4_xjSX6pN0mb_QwGyg)