

JSAI2026 企画セッション KS-21 「人工知能とコンペティション」

# LA-Bench 2025

## 実験手順生成AIコンペティション

### 開催報告 & 表彰

主催：一般社団法人ラボラトリーオートメーション協会（LASA）AI分科会

支援：2025年度 人工知能学会コンペティション開催支援制度

2026年6月9日（火） Gメッセ群馬／オンライン

# 本日の内容

---

- LA-Bench とは — 背景と狙い
- コンペティション概要（主催・支援・賞金）
- タスク内容 — 実験手順生成
- 評価設計とメトリクス
- スケジュールと参加状況
- 採点結果と表彰
- 受賞者による手法発表 ／ 関連発表のご案内
- LA-Bench 2026 に向けて

# なぜ LA-Bench か — AI駆動科学のギャップ

---

## ■ AIによる科学研究（AI for Science）の自動化への期待が高まっている

- 「The AI Scientist」（Sakana AI, arXiv 2024.8 → Nature 2026.3） — 着想→実験→執筆を自動化

## ■ ただし「仮説→実験→解析」のループが自動で閉じるのは、計算機内で完結する研究が中心

- 「Robin」（FutureHouse, arXiv 2025.5 → Nature 2026.5） — 仮説・実験計画・解析を自律化し創薬候補を提案。ただし物理実験は人が実施

## ■ 従来のラボオートメーションは、ロボットを動かすため人が詳細な手順・プログラムを整備する必要があり手間が大きい

## ■ AIの計画と実機の実行をつなぐ「詳細手順への肉付け」が鍵 → LA-Bench を立ち上げ

# コンペティション概要

| 項目    | 内容                                    |
|-------|---------------------------------------|
| 正式名称  | LA-Bench 2025：実験手順生成AIコンペティション        |
| 主催    | 一般社団法人ラボラトリーオートメーション協会（LASA）AI分科会     |
| 運営責任者 | Science Aid株式会社 代表取締役CEO 山田 涼太        |
| 支援    | 2025年度 人工知能学会コンペティション開催支援制度           |
| 開催期間  | 2025年9月～2025年12月（表彰：2026年6月 JSAI2026） |
| 賞金総額  | 9社のスポンサー支援により 50万円 → 90万円 に増額         |

運営メンバー（50音順）：尾崎 遼（理研）／加藤 祥太（京大）／酒井 雄介（東大）／杉山 亜矢斗／西田 理彦（LASA）／山田 涼太（Science Aid）

# タスクの全体像 — 何を入れて、何を出すか

## 入力

実験指示 + 元プロトコル + 使用物品 + 参考文献

### 例（分子生物学）

「大腸菌K12の遺伝子 yexP を増幅し、発現ベクター構築用の断片と線形化ベクターを用意する。親ベクターは pET21b、制限酵素は EcoRI/XhoI、ポリメラーゼは KOD Plus neo を使用」



## 出力

実行可能な詳細手順（試薬量・温度・時間を明示）

### 例

「滅菌水 31μL／10×buffer 5μL／2mM dNTPs 5μL／25mM MgSO<sub>4</sub> 3μL ...、初期変性 94°C 2分 → （98°C 10秒／60°C 30秒／68°C 1分 30秒）×30サイクル → ...」

→ この精度の手順があれば、実験ロボットや不慣れな実験者でも再現実行できる = ラボ自動化・研究自律化の基盤

## 求める詳細度（イメージ：水に砂糖を溶かす）

### × 不十分な手順

1. コップに水を入れる
2. 10%になるよう砂糖を加える
3. かき混ぜる

### ○ 期待される手順

1. コップに100 mLの水（25℃）を入れる
2. 砂糖を10 g加える
3. スプーンで30秒間かき混ぜる

→ 分量・温度・時間・操作などのパラメータを明示化し、再現可能なレベルまで具体化することが鍵

# タスク内容 – 入出力の仕様

| 入力フィールド               | 内容                      |
|-----------------------|-------------------------|
| instruction           | 実験の目的と要求事項（実験指示）        |
| mandatory_objects     | 必ず使用する物品（機器・ラボウェア・サンプル） |
| source_protocol_steps | 参考となる元の実験手順             |
| expected_final_states | 期待される最終状態（生成物・中間生成物）    |
| references            | 関連する論文・資料               |

## 出力：procedure\_steps

- ・ 手順のリスト（id + text）
- ・ 要素数 50以下／1要素 10文以下
- ・ 学部レベルの訓練を受けた人が誤解なく実行できる詳細度
- ・ 暗黙知・省略・条件・数値をすべて明示化

# 評価設計 — 2段階フェーズ

## Development Phase（開発）

- 評価対象：Public Test（10問）
- 提出：1日5回まで
- LLM評価でリーダーボードに反映
- 練習用 Example（5問）も公開

## Test Phase（最終評価）

- 評価対象：Private Test（10問）
- 提出物3点：予測結果・実行コード・README
- 第1次：LLM評価（全チーム）
- 第2次：専門家評価（上位10チーム）

# 採点 — 評価基準と2段階評価

評価基準：10点満点（共通採点基準 5点 + 個別採点基準 5点）

- ・ 共通：指示パラメータ・物品・論理構造・最終状態・補完の妥当性
- ・ 個別：問題ごとの意図適合・安全性・コスト・効率・精度

|      | 第1次：LLM評価                  | 第2次：専門家評価          |
|------|----------------------------|--------------------|
| 対象   | 全21チーム                     | LLM評価 上位10チーム      |
| 採点者  | LLMを評価器に利用（LLM-as-a-judge） | 各問題に2名のドメイン専門家（匿名） |
| 位置づけ | 第1次スクリーニング                 | 最終評価               |

※ 評価設計（評価器・プロンプト等）の詳細は、関連発表 OS-8（加藤ほか／6月10日）で報告します。

※ Private\_test\_4 は問題の不備があったため採点対象から除外。

# スケジュール

| 日付         | 内容                                           |
|------------|----------------------------------------------|
| 2025-09-13 | 公式サイト公開・参加登録開始（ルール／Example／サンプルコード）          |
| 2025-10-10 | Public Test Dataset 公開（Development Phase 開始） |
| 2025-11-14 | Private Test Dataset 公開（Test Phase 開始）       |
| 2025-11-21 | 提出締切（23:59 JST）                              |
| 2025-12-20 | 採点結果 公開                                      |
| 2026-06-09 | JSAI2026 にて 表彰・成果発表（本セッション）                  |

## 参加状況 — 想定を大きく上回る反響

**57** チーム

参加登録

**21** チーム

Test Phase 提出

**9** 社

スポンサー企業

**90** 万円

賞金総額（増額後）

- ・ 大学・研究機関・民間企業など多様な組織から研究者・エンジニアが参加
- ・ 締切直前はアクセス集中でリーダーボードのサーバーが一時ダウンするほどの熱戦
- ・ 上位チームはLLM評価・専門家評価ともに僅差での接戦となった

RESULTS & AWARDS

---

# 採点結果と表彰

## 採点結果 ① LLM評価（上位10チーム）

| 順位 | チーム名          | スコア   |
|----|---------------|-------|
| 1  | cabbage patch | 9.289 |
| 2  | satoshihirose | 9.285 |
| 3  | NinNinNin     | 9.126 |
| 4  | O_AI          | 9.026 |
| 5  | Team SL       | 9.015 |
| 6  | moroqq82      | 8.889 |
| 7  | moomoo        | 8.796 |
| 8  | Chem3         | 8.719 |
| 9  | falcon        | 8.715 |
| 10 | JIYU-JIZAI    | 8.604 |

全21チームをLLMで評価。1位・2位は0.004差の大接戦

## 採点結果 ② 専門家評価（上位10チーム）

| 順位 | チーム名          | スコア |
|----|---------------|-----|
| 1  | moroqq82      | 8.8 |
| 2  | satoshihirose | 8.7 |
| 2  | O_AI          | 8.7 |
| 4  | Team SL       | 8.6 |
| 4  | falcon        | 8.6 |
| 6  | NinNinNin     | 8.4 |
| 7  | Chem3         | 7.4 |
| 8  | moomoo        | 7.2 |
| 9  | cabbage patch | 6.9 |
| 10 | JIYU-JIZAI    | 5.9 |

上位10チームを各問題2名の専門家が匿名で評価。LLM評価とは異なる順位傾向に



## LLM評価部門

### 最優秀賞

**cabbage patch**

賞金 35万円

### 優秀賞

**satoshihirose**

賞金 10万円

## 専門家評価部門

### 最優秀賞

**moroqq82**

賞金 35万円

### 優秀賞

**satoshihirose ・ O\_AI**

同点 → 各 10万円

受賞チームの皆様、誠におめでとうございます。（賞金の授与方法は後日個別にご連絡いたします）

# LA-Bench 2026 に向けて

■ LA-Bench はラボラトリーオートメーション実現に向けた継続的な取り組み（2026 開催を予定）

■ 拡充予定の評価タスク（構想）

- ・ 実験計画の最適化 / 実験結果からの異常検知 / マルチモーダル実験データ統合
- ・ ロボットへの作業指示生成 / 安全性評価とリスク予測

▶ 運営・協賛にご関心のある企業の皆様へ

セッション終了後に、ぜひお気軽にお声がけください。

# 関連発表のご案内

LA-Bench の評価設計と結果の詳細については、翌日のオーガナイズドセッションでも発表します。ぜひご参加ください。

| 項目    | 内容                                                                                    |
|-------|---------------------------------------------------------------------------------------|
| セッション | OS-8「生成AIと研究自動化が拓く医歯薬学・生命科学DX」（3F1-OS-8）                                              |
| 日時・会場 | 2026年6月10日（水）9:00～10:30／F会場（メインホールB）・オンライン併用                                          |
| 発表    | 9:30～9:45 [3F1-OS-8-03]<br>「LA-Bench 2025：実験指示から実行可能手順を生成するためのデータセット」                 |
| 発表者   | ○加藤 祥太（京都大学）、西田 理彦（LASA）、酒井 雄介（東京大学）、<br>尾崎 遼（理化学研究所）、杉山 亜矢斗（LASA）、山田 涼太（Science Aid） |
| キーワード | 実験手順生成／自動化／実験プロトコル／ベンチマークデータセット／大規模言語モデル                                              |

## SPONSORS

# スポンサー

9社のスポンサー企業のご支援により、賞金総額を 90万円 に増額することができました。心より御礼申し上げます。

The logo for DEEP(ORE) features the word "DEEP" in a bold, dark blue sans-serif font, followed by "(ORE)" in the same font and color.The logo for embrio consists of a blue circular icon made of a chain of small circles, followed by the word "embrio" in a blue sans-serif font.The logo for Epistra features the word "Epistra" in a blue serif font, with the tagline "Accelerate your R&D" in a smaller, blue sans-serif font below it.The logo for HERO IMPACT CAPITAL features a green circular icon with a white line through it, followed by the word "HERO" in a green sans-serif font, and "IMPACT CAPITAL" in a smaller, green sans-serif font to the right.The logo for inner Resource features a stylized geometric icon composed of red and blue shapes, followed by the words "inner Resource" in a black sans-serif font.The logo for ISP features a stylized black icon resembling a star or a burst, followed by the letters "ISP" in a bold, italicized black sans-serif font.The logo for Linkers OI Lab features a green icon of two overlapping squares, followed by the words "Linkers OI Lab" in a green sans-serif font.The logo for Polaris.AI features a blue geometric icon resembling a cube or a stylized letter 'P', followed by the text "Polaris.AI" in a blue sans-serif font.The logo for PRIMAL features a blue geometric icon resembling a stylized letter 'P', followed by the word "PRIMAL" in a bold, grey sans-serif font.

# 謝辞

ご参加いただいた皆様、ご支援いただいたスポンサー企業の皆様、  
ご協力いただいたドメイン専門家・運営メンバーの皆様に、心より御礼申し上げます。

主催：一般社団法人ラボラトリーオートメーション協会（LASA）AI分科会

支援：2025年度 人工知能学会コンペティション開催支援制度

# 受賞者による手法発表

ここからは受賞チームの皆様より、それぞれ約5分で手法をご紹介します。

|   | チーム           | 発表者（敬称略） | 所属                 | 受賞         |
|---|---------------|----------|--------------------|------------|
| 1 | cabbage patch | 田原-新井 悠也 | 筑波大学（オンライン登壇）      | LLM評価 最優秀  |
| 2 | satoshihirose | 廣瀬 智史    | 個人                 | LLM・専門家 優秀 |
| 3 | O_AI          | 日高 啓介    | 京都大学医学部附属病院 放射線診断科 | 専門家評価 優秀   |
| 4 | moroqq82      | 岡谷 基弘    | 個人                 | 専門家評価 最優秀  |

※ 田原-新井さんはオンライン登壇（Zoom画面共有）。続いて廣瀬さん・日高さん・岡谷さんに登壇いただきます。