

# 科学の第5パラダイム： AIがR&Dにもたらす 変革的インパクト

段階的自動化とガバナンスを軸と  
した、実践的導入ための設計図







OECD・JST報告に基づく最新の実証的アセスメント

# R&DにおけるAIの現在地：CopilotからAgentic Systemへ

## 「Copilot」モデル（周辺作業の補助）

- 役割: 読む・探す・書く
- 適用: 文献探索、要約、文章作成
- 特徴: 人間が常にタスクを牽引し、AIは一問一答の補佐を行う。広く実務化済み。

## 「Agentic System」モデル（閉ループの限定実行）

- 役割:  考える・ 試す・ 評価する
- 適用:  仮説生成、 実験計画、 自律ラボ制御
- 特徴: 限定された問題設定の中で、AIが複数ステップを自律的に回す。分野限定で急速に実証中。

AIの影響は「文章作成補助」から、探索・実験・解析を含む「研究プロセス全体の自動化」へ移行した。

# 探索から実行へ：2015年から2026年への歴史的タイムライン

2015–2019  
[AI支援検索・読解]

論文探索がキーワードから「意味検索」へ移行。（例: Semantic Scholar 2.3億本超の検索）

実験前の候補絞り込みと構造仮説形成が劇的に高速化。（例: AlphaFold 2 原子精度級予測）

2020–2023  
[自律実験の台頭]

データ取得そのものがアルゴリズム化。（例: A-Lab, モバイル・ロボット化学者）

AIが「装置を動かす・複数モダリティを統合する」段階へ。（例: Coscientist 6タスク実証）

2024–2025  
[研究エージェント化]

探索・統合・提案のフロントエンドがエージェント化。（例: PaperQA2 超人的性能）

アイデア生成から原稿・自己査読までの自動化。（例: The AI Scientist）

論文探索がキーワードから「意味検索」へ移行。（例: Semantic Scholar 2.3億本超の検索）

2020–2021  
[予測科学の実用化]

実験前の候補絞り込みと構造仮説形成が劇的に高速化。（例: AlphaFold 2 原子精度級予測）

データ取得そのものがアルゴリズム化。（例: A-Lab, モバイル・ロボット化学者）

2023–2024  
[基盤モデルの研究工程侵入]

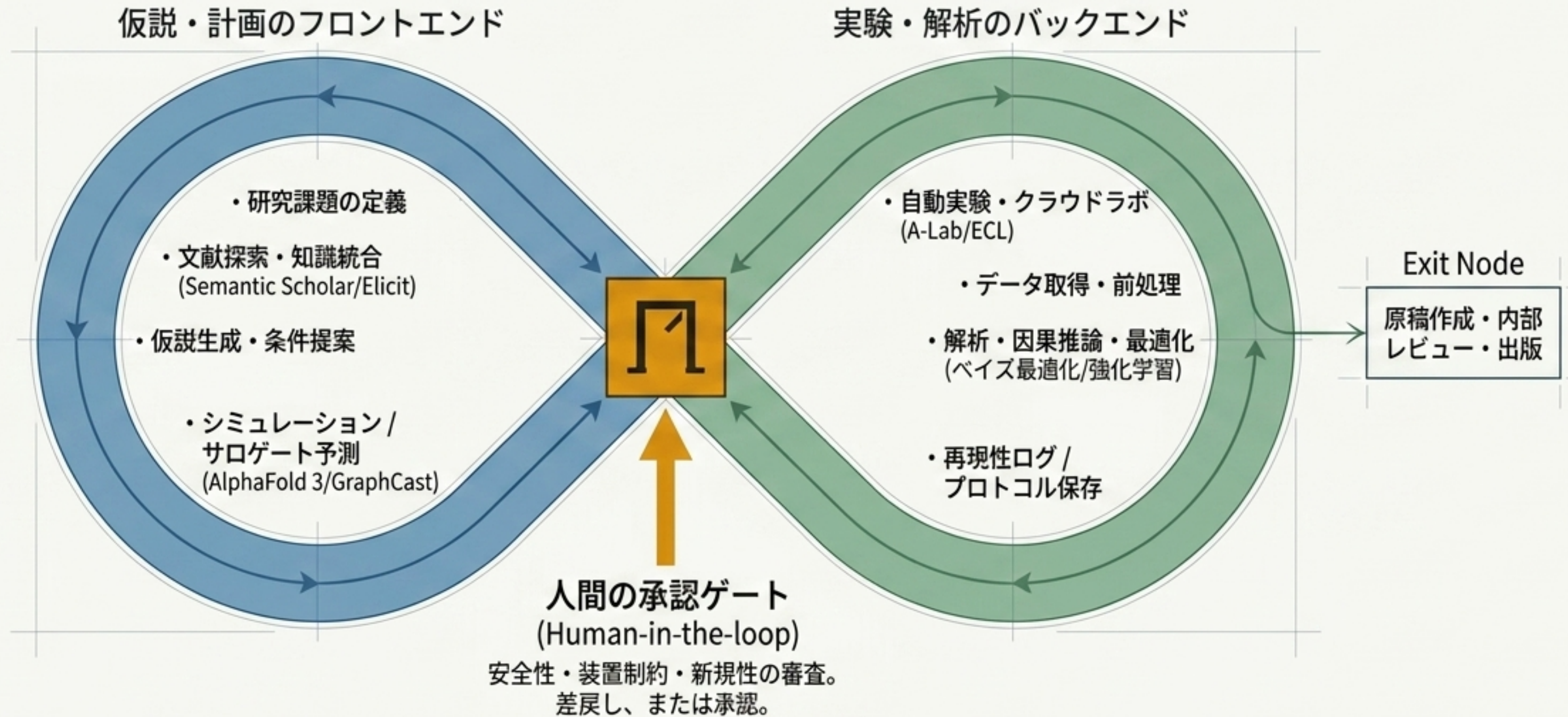
AIが「装置を動かす・複数モダリティを統合する」段階へ。（例: Coscientist 6タスク実証）

探索・統合・提案のフロントエンドがエージェント化。（例: PaperQA2 超人的性能）

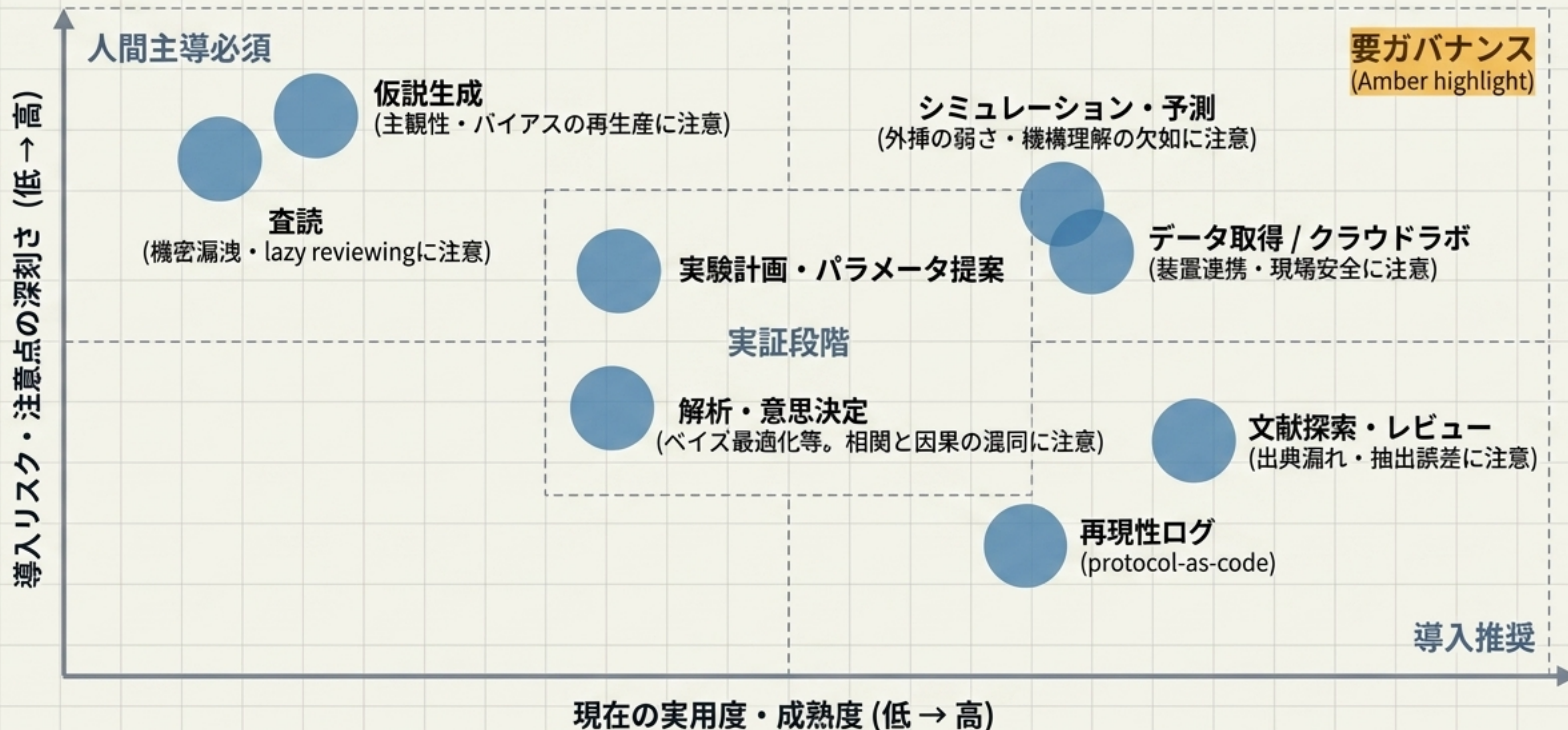
2025–2026  
[限定領域のAI科学者]

アイデア生成から原稿・自己査読までの自動化。（例: The AI Scientist）

# AI駆動型R&Dの「閉ループ」モデル

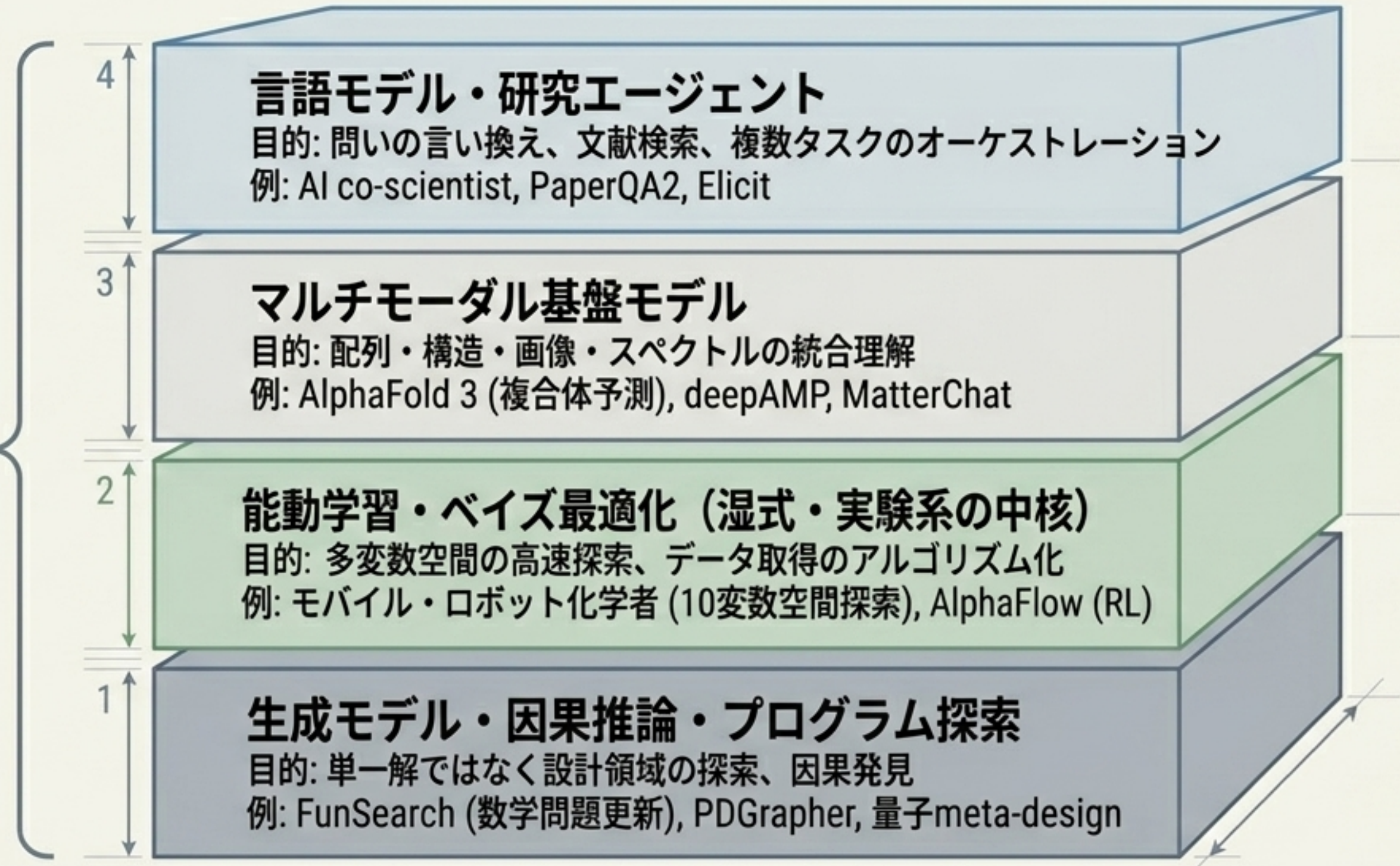


# 研究プロセス別・AI成熟度とリスクの診断マトリクス



# 万能モデルの幻想を捨てる：「連携スタック」としての科学AI

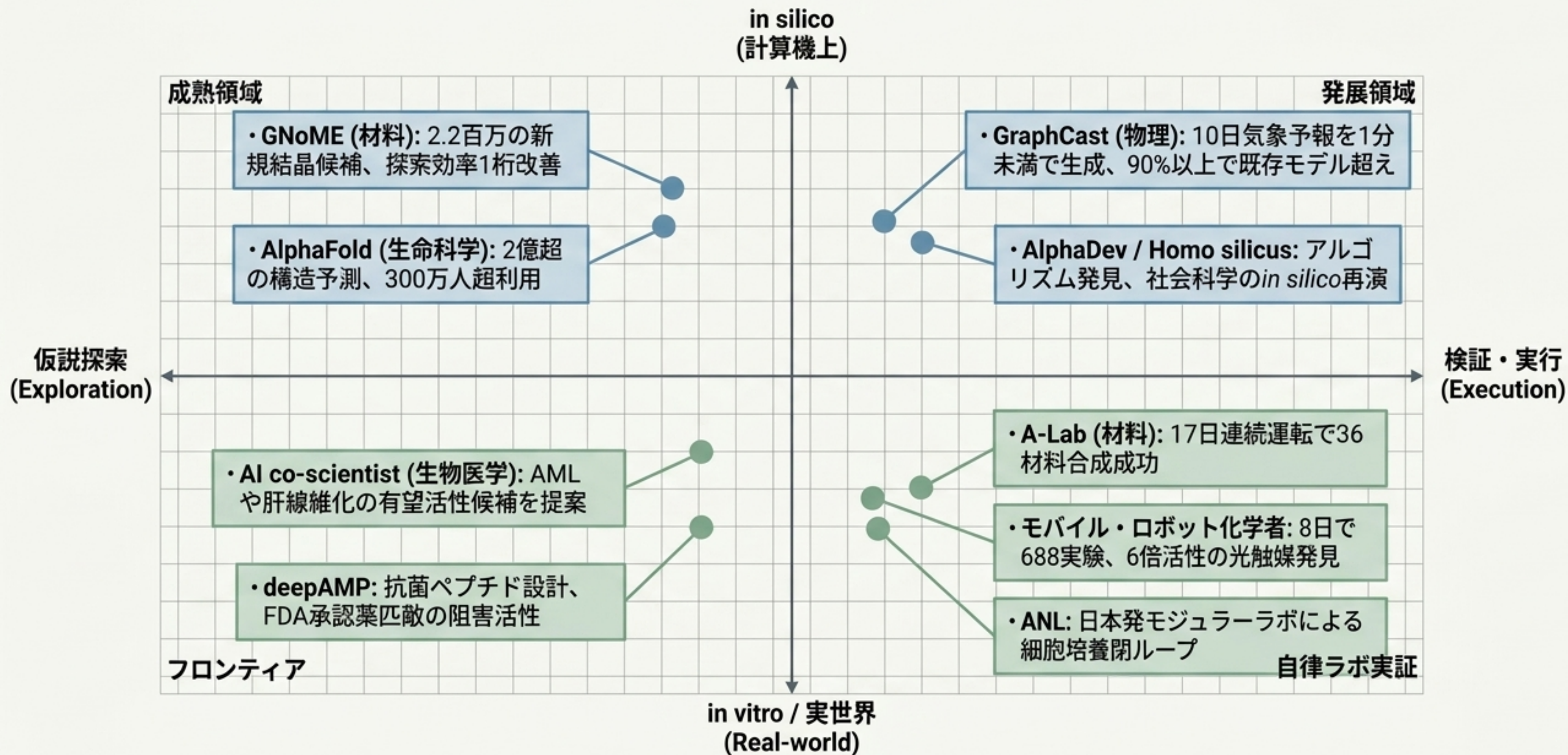
単一のLLMではなく、  
特化型モデル群の  
連携によって機能する



# 主要ツールとエージェントの現在地：能力と制約

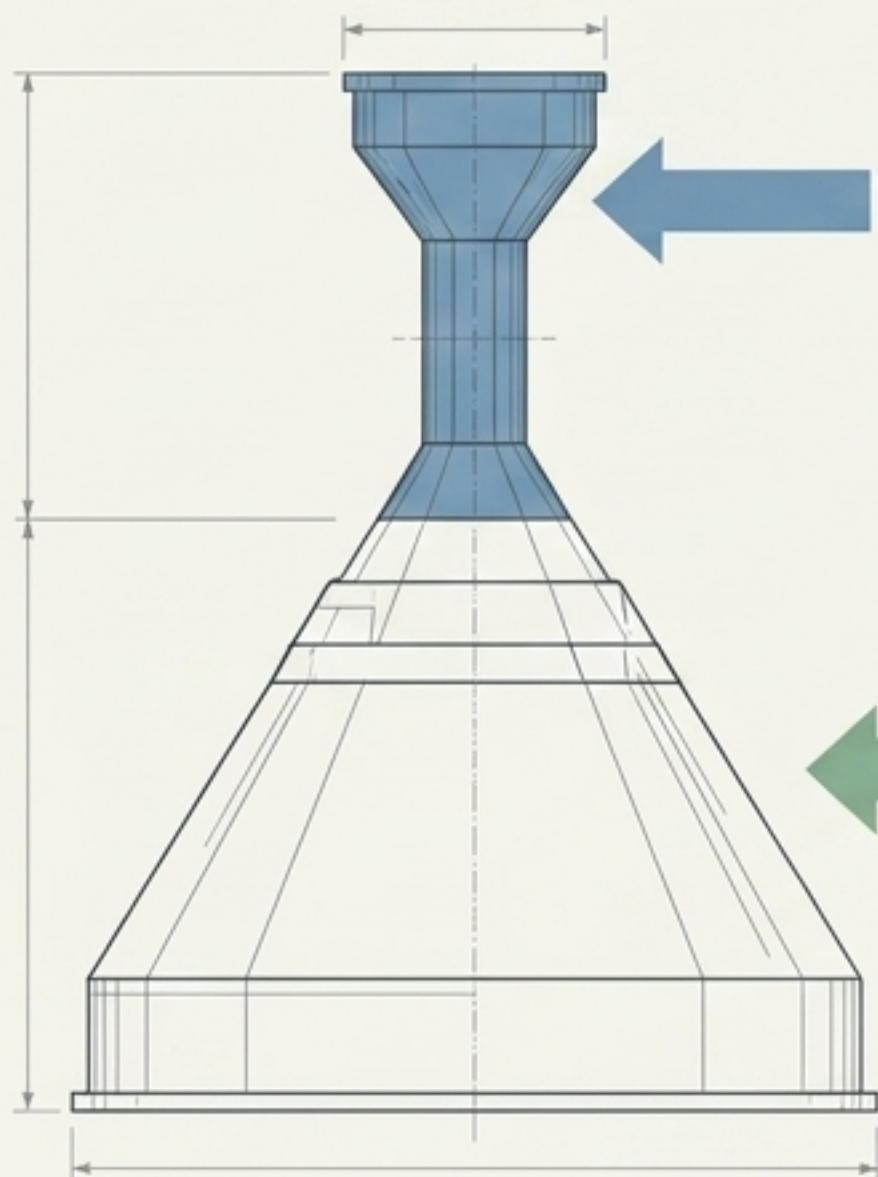
| 機能領域         | ツール / 企業                              | 主な能力・価値         | 現在の制約と注意点                  |
|--------------|---------------------------------------|-----------------|----------------------------|
| in silico 構造 | AlphaFold Server<br>(Google DeepMind) | タンパク質相互作用予測     | 実験検証は依然必要、外挿限界あり           |
| 文献探索         | Semantic Scholar<br>(AI2)             | 大規模意味検索         | 検索再現率は問いの立て方に依存            |
| 文献レビュー       | Elicit                                | 質問起点の論文抽出・比較    | 抽出誤差の人手確認が必要               |
| 科学エージェント     | FutureHouse Platform                  | 文献探索・統合専門       | 公称性能の独立検証は今後の課題            |
| 自律研究プロトタイプ   | The AI Scientist<br>(Sakana AI)       | end-to-endのML研究 | 現状は計算機上で完結する領域中心。安全な実行環境必須 |
| クラウドラボ       | ECL<br>(Emerald Cloud Lab)            | リモート実験・自動計測     | 装置適合性、ワークフロー制約             |

# 学術・産業横断の実証ケース：限界突破の現在地



# 人間の新たな役割：「律速段階（ボトルネック）」の劇的な移動

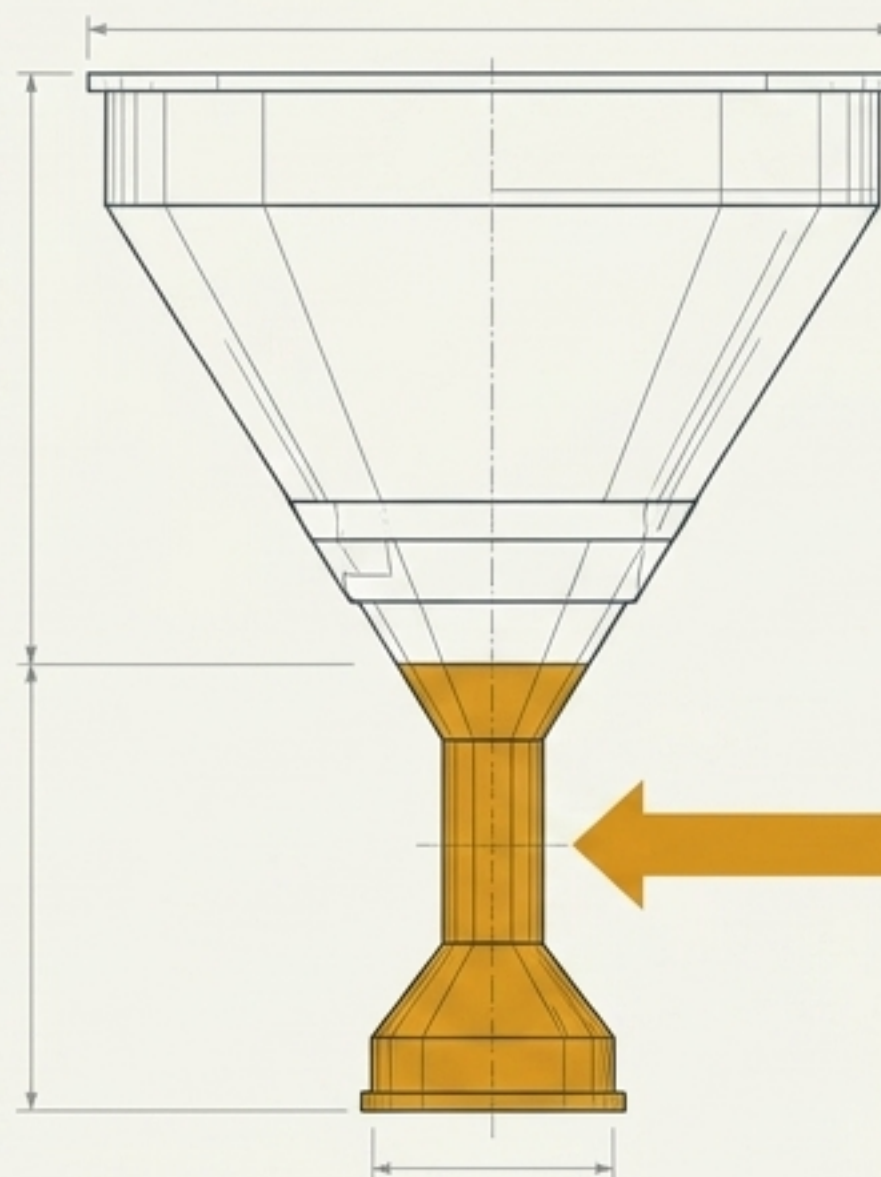
過去のR&D



**細い上部（律速段階）**  
論文の発見、読解、仮説構築  
（人間が手作業で行うため極めて遅い）

**太い下部**  
実験の実行と検証  
（人員と時間をかければスケールする）

AI時代のR&D



**極太の上部**  
AIによる膨大な論文処理  
と仮説の大量生成  
（速度・幅が無限大に広がる）

**細い下部  
（新たな律速段階）**  
生成された膨大な仮説の中から、  
「真の新規性・再現性・安全性」  
を検証・評価する作業

AIは研究者を置き換えるのではなく、律速段階を別の場所に移す装置である。  
今後の人間の主戦場は「生成」ではなく「審査とガバナンス」になる。

# AIがもたらす4つの便益と「科学のモノカルチャー化」のパラドックス

## 圧倒的な便益と生産性

- 1. 速度: 計算圧縮 (GraphCast)、24時間稼働 (クラウドラボ)。
- 2. 探索幅: 人間では到達困難な候補母集団の拡大 (GNoME)。
- 3. コスト: 安価なパイロット研究の実現 (Homo silicus)。
- 4. 再現性: プロトコル逸脱や暗黙知依存の排除。

**Data:** AI活用科学者は論文数3.02倍、引用数4.84倍 (Nature 2026)

## 「モノカルチャー化」の逆説

- AIは研究トピックの集合的多様性を4.63%縮小し、研究者同士の関与を22%減少させる。
- 「わかった気にさせる」危険性: 問いや方法論が単一化し、知的多様性が低下するパラドックス。
- 生産性向上と引き換えに、科学の画一化が進むリスク。

# 組織が直面する6つのリスク・ディメンション

## ⚠️ 1. ハルシネーションと真実性

Confabulation検出には semantic entropy などの統計的方法が不可欠。

## ⚠️ 2. バイアス

base modelに見られる「内集団びいき」と「外集団敵/敵視」の再生産。

## ⚠️ 3. 物理的安全性 (EHS)

AIによる病原体操作や化学条件設定の誤りが、現場の致死事故に直結する危険。

## ⚠️ 4. 便益の不均等配分

分野や研究者の属性によるAI恩恵の偏り。

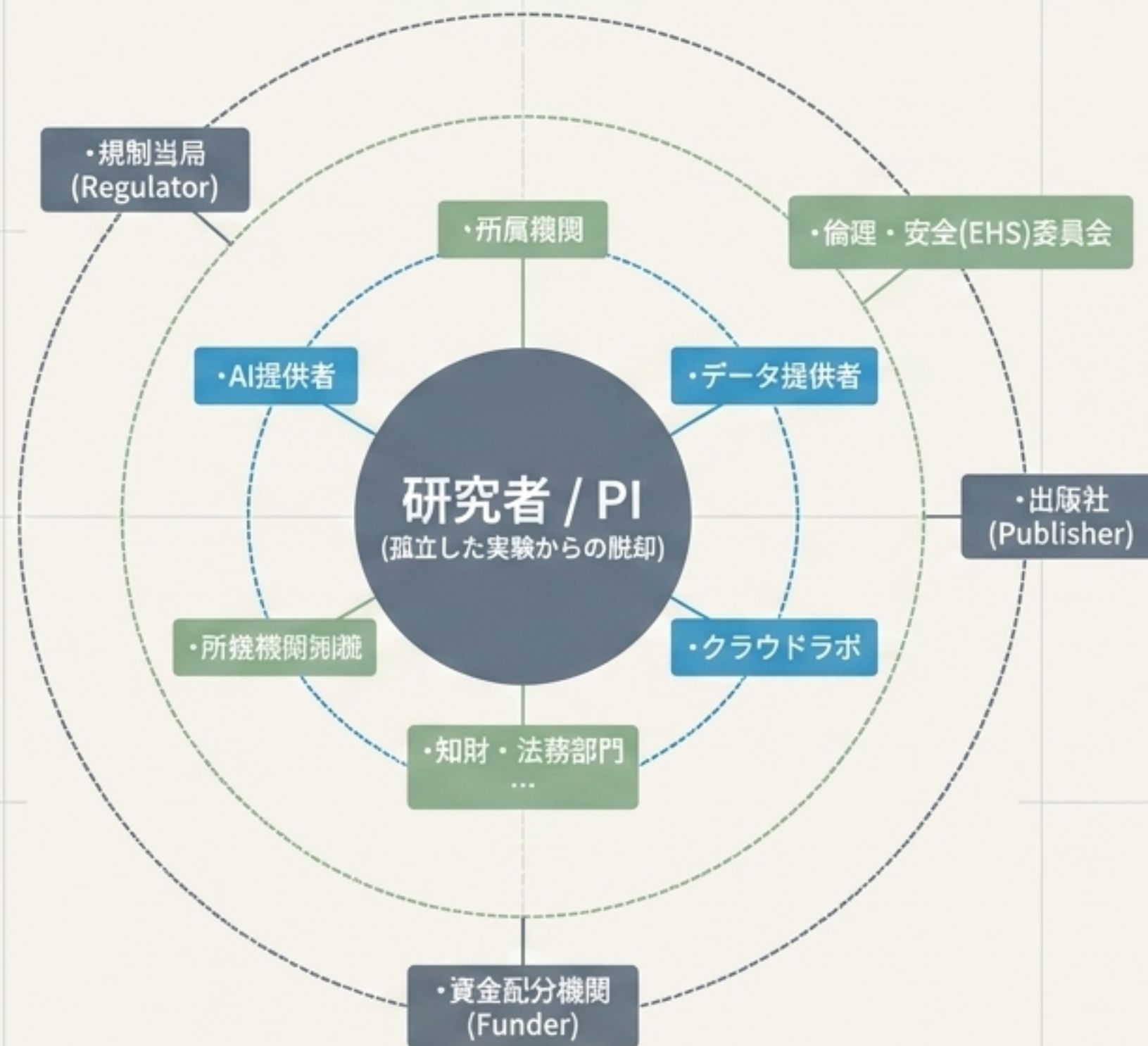
## ⚠️ 5. 知財・データ統治

EU AI Actや経産省ガイドラインへの対応。モデル調達は法務と不可分。

## ⚠️ 6. 査読・出版の制度的負荷

査読プロセス外への機密漏洩リスク。AIの「著者性」否定方針 (ICMJE/COPE)。

# エコシステムとガバナンス：R&Dを取り巻くステークホルダー相関図



Key Insight: AI導入は「ツールの追加」ではなく、多重ステークホルダー間の「運用設計・契約・審査フローの再構築」である。

# 実践のプレイブック：成功する組織の「段階的自動化」

## 1. 出典接地 (Grounding)

文献タスクはRAG（検索拡張生成）と引用付き回答を必須化。根拠不十分なら「不明」と出力する設計。

## 2. 不確実性の管理 (Uncertainty Mgmt)

Semantic entropy等による不確実性検知。モデル間照合と、シミュレーション・実験の「二重化」の実装。

## 3. 来歴管理 (Provenance)

プロンプト、モデル版、データ版、装置ログを全て記録。研究公正と製品化のための完全な再現性を確保。

## 4. サンドボックス化 (Sandboxing)

コード実行や装置制御は「最小権限」と「ネットワーク制限」を徹底。危険操作のホワイトリスト化。

### 組織横断アプローチ

PI、研究DX、法務/IP、EHSを統合したクロスファンクショナルチームによる一元的な公開判断プロセス。

# ガバナンスモデルの3層構造 (3-Layer Autonomy)

## Tier 3: Delegated Autonomy / 限定的委任 (高リスク・最先端)

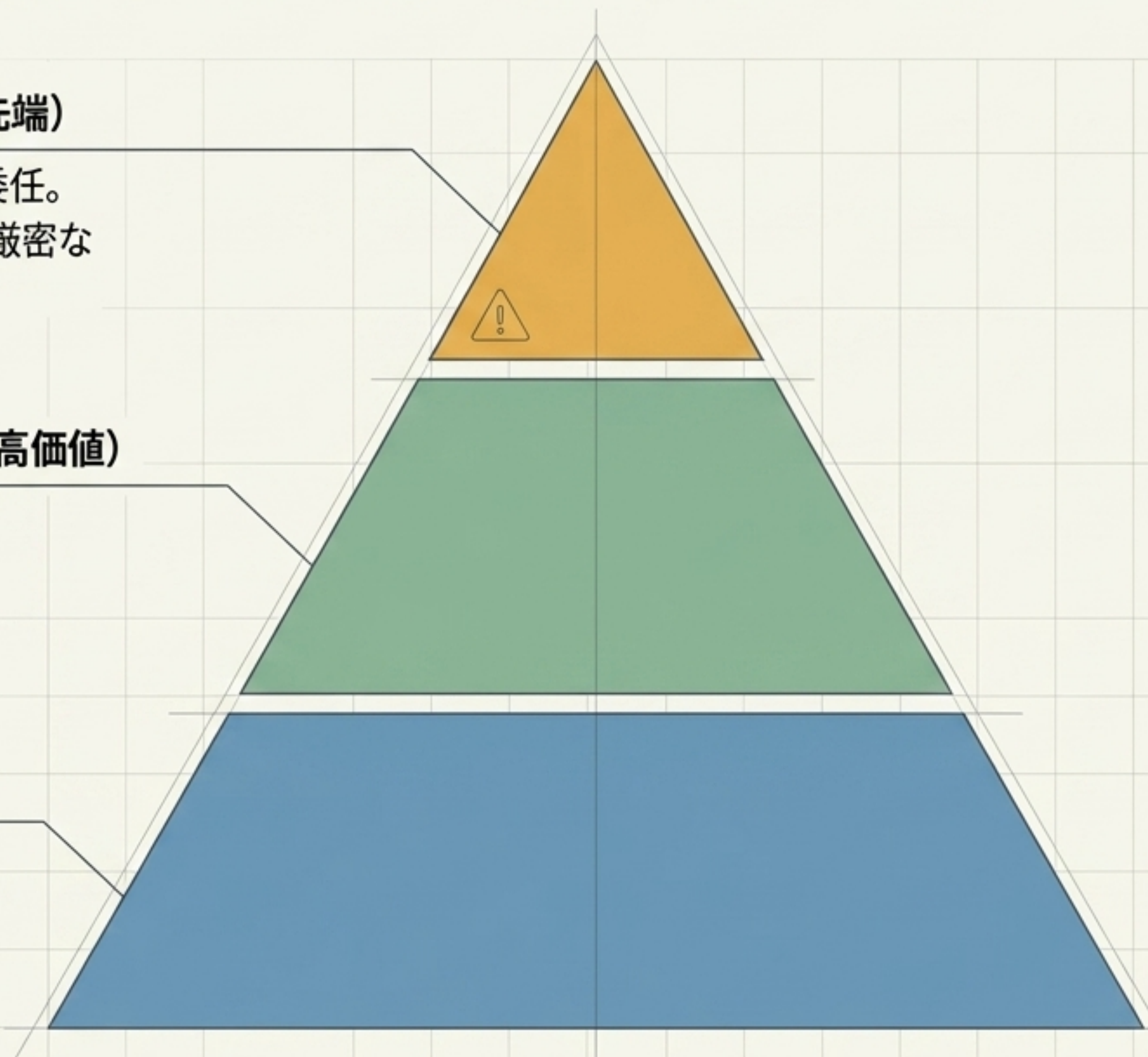
内容: 企画から実験までの完全自律。明示的な責任分界を伴う委任。  
要件: EU AI Act に準じたシステムリスク評価、自己報告機能、厳密なサンドボックス。

## Tier 2: Bounded Autonomy / 制約付き自律化 (中リスク・高価値)

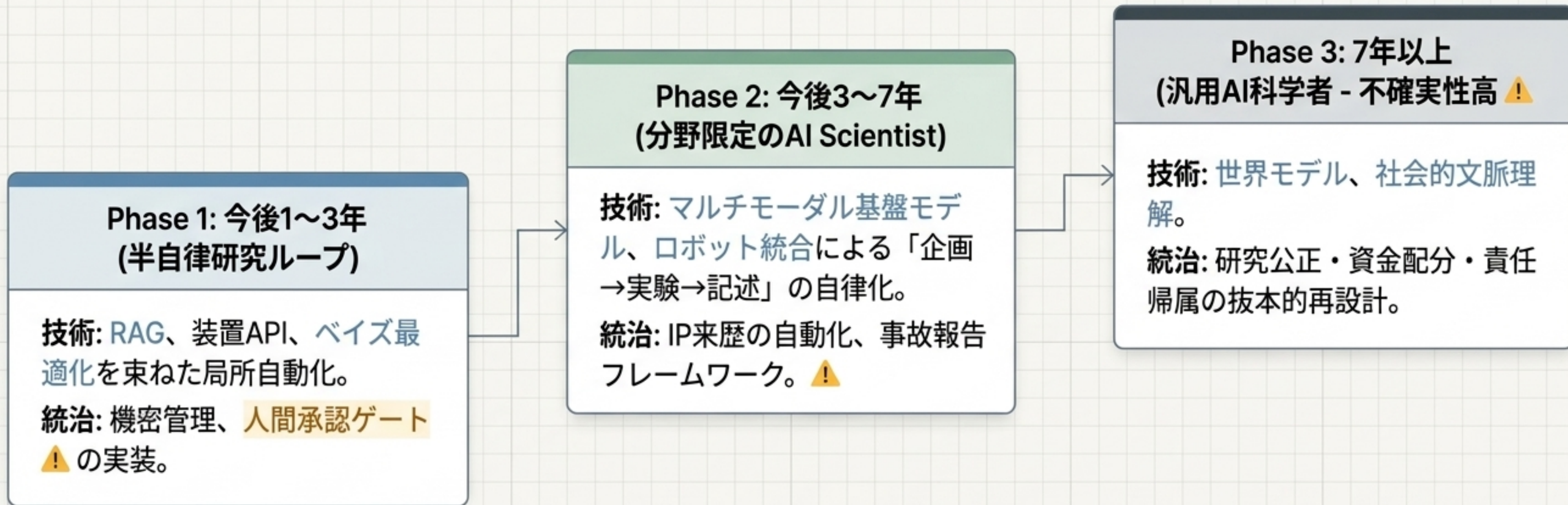
内容: パラメータの条件探索や標準化された測定の自律運用。  
要件: 評価関数の厳密定義、異常検知、人間の承認ゲートの必須化。※現在の主戦場。

## Tier 1: Copilot (低リスク・広範適用)

内容: 文献・記述・初期分析の支援。  
要件: 内部ポリシーの明文化、査読時の透明な申告 (Nature/ICMJE準拠)。



# 未来図：自律的AI科学者の発展ロードマップ



「AIは研究者を置き換えるのではなく、律速段階を別の場所に移す装置である。その速度増加を、発見率・安全性・知的多様性の向上へと変換できた組織のみが、真に次世代の研究力を獲得する。」