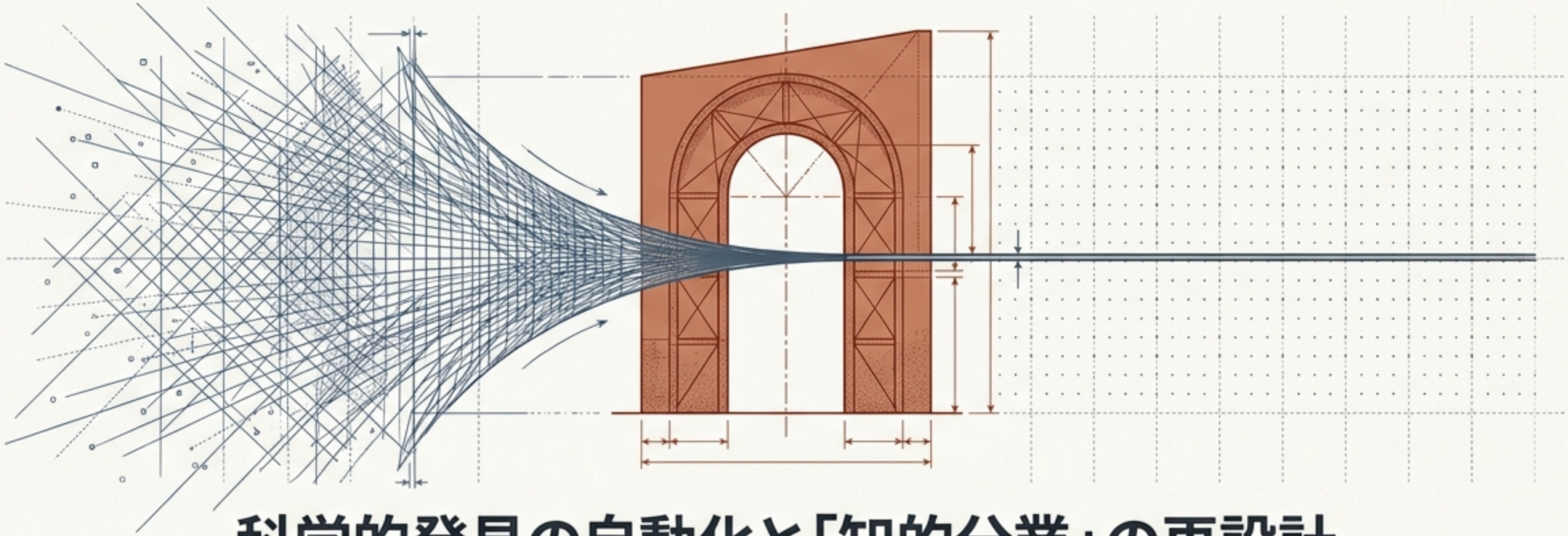
「仮説を作るAI」は科学者を不要にするのか？



科学的発見の自動化と「知的分業」の再設計

従来の脅威論

AIが未知の問い・仮説を
大量生成する

↓

科学者の「発想」の
価値が消滅する

↓

人間に残るのは
「価値判断」のみ

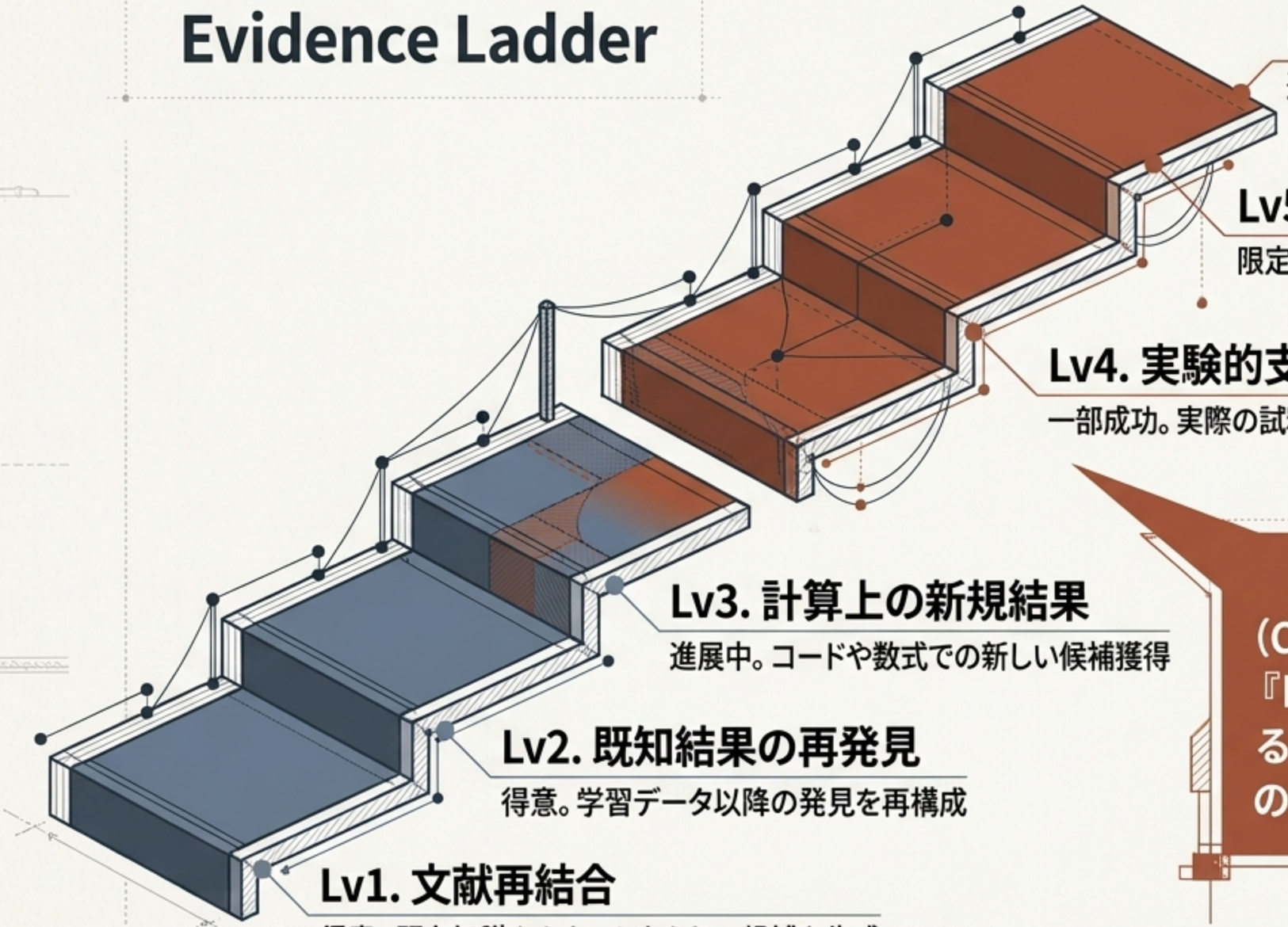
※80年来の数学問題解決や新規標的提案などにより誘発された神話

本稿の核心

「科学者の消滅」ではない。
「科学における“ボトルネック”の移動」である。

- AIは探索空間の走査コストを劇的に下げる。
- 結果、「検証」という新たな希少資源の価値が急騰する。
- 科学者はアイデアの生産者から、知識生成パイプラインの「設計者・監査者」へ移行する。

Evidence Ladder



Lv6. 臨床・社会的な知識
遠い。患者集団・現実環境での有効性と安全性

Lv5. 独立再現・外的妥当性
限定的。別条件での再現

Lv4. 実験的支持
一部成功。実際の試料での予測検証

Lv3. 計算上の新規結果
進展中。コードや数式での新しい候補獲得

Lv2. 既知結果の再発見
得意。学習データ以降の発見を再構成

Lv1. 文献再結合
得意。既存知識からもっともらしい候補を生成

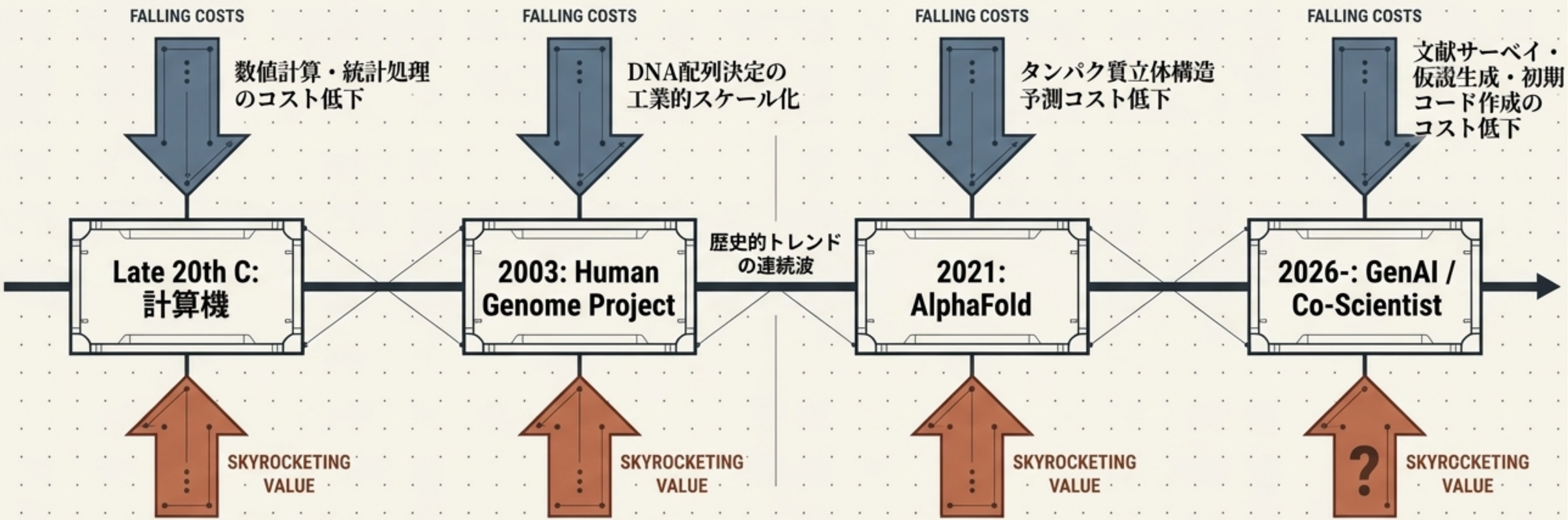
現在の最先端AI（Co-Scientist等）も、依然として『Lv4の入り口』にとどまっている。難易度が上がると正解仮説の回収率は38.8%に低下する。

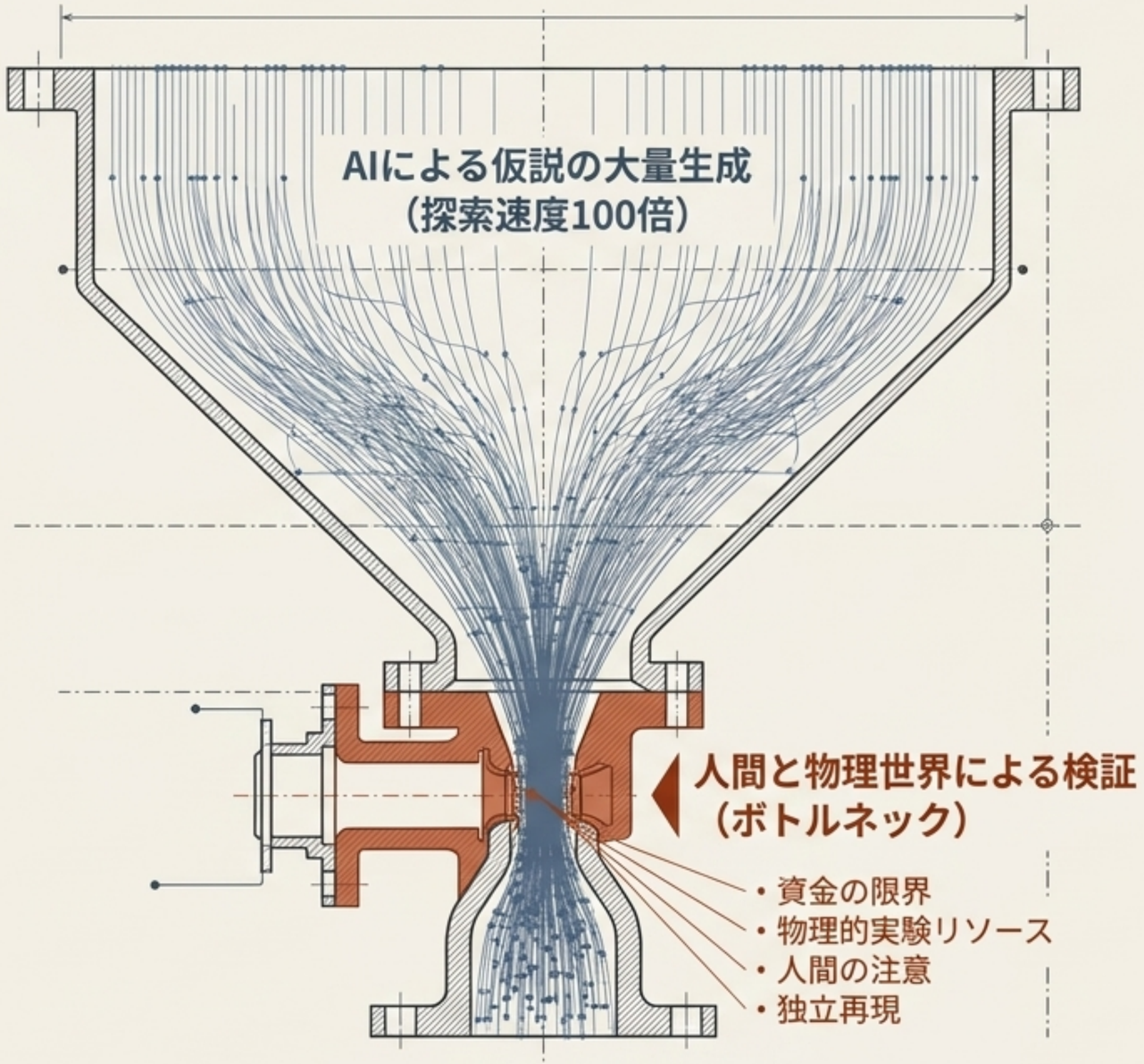
比較診断マトリクス

システム	技術方式	注目すべき実績	主な限界・リスク
The AI Scientist-v2	LLM + Agentic tree search	論文を自動生成。1本がICLRの平均採択基準を超過	開発者の内部評価では本会議基準未達。指標誤用や事後選択リスク
MOOSE-Chem	Multi-agent + 文献検索	学習期間後の未知の化学仮説を『再発見』	既存論文との類似度評価であり、真に前例のない未知の検証ではない
Google Co-Scientist	生成・批判エージェント群 + ツール	AMR機構を約2日で特定（同時期の人間と一致）。in vitroで成功	目標設定や実験優先度は人間が担当。負の結果の欠落や均質化リスク

結論: すべての有力システムは完全自律ではなく「Scientist-in-the-loop」を前提としている。

限界費用が下がると、その出力を入力とする「次の工程」の価値が急騰する





「検証負債 (Validation Debt)」の誕生

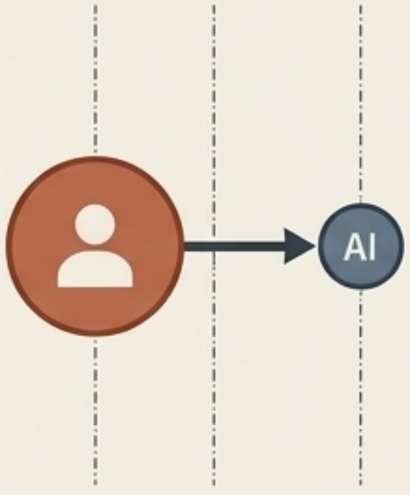
AIによる仮説生産速度が100倍になっても、検証能力が100倍にならなければ、「知識」が100倍になるのではない。
「検証待ちの仮説（偽陽性を含む）」が100倍になるだけである。

アイデア不足の時代から、偽陽性と検証負債が大量生産される時代へ。

人間とAIの知的分業の5モデル

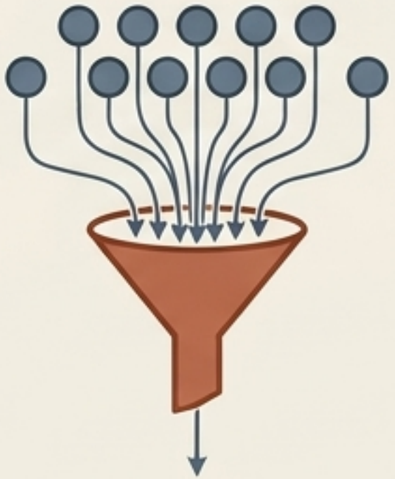
どの水準の証拠と説明責任が必要かで分業モデルを選ぶ

1. コパイロット型



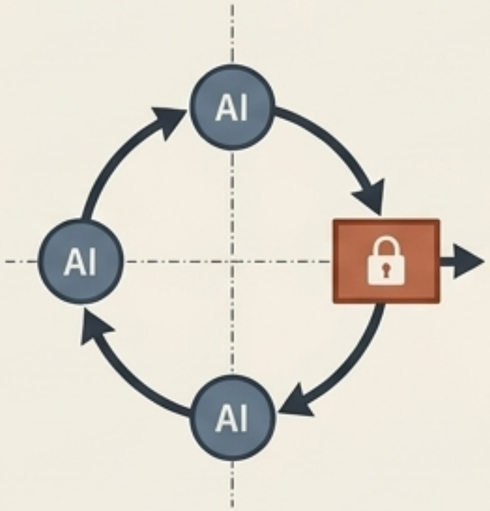
人間が主導、AIが補助。
高リスク臨床研究向け。

2. 仮説ポートフォリオ型
(★Co-Scientistモデル)



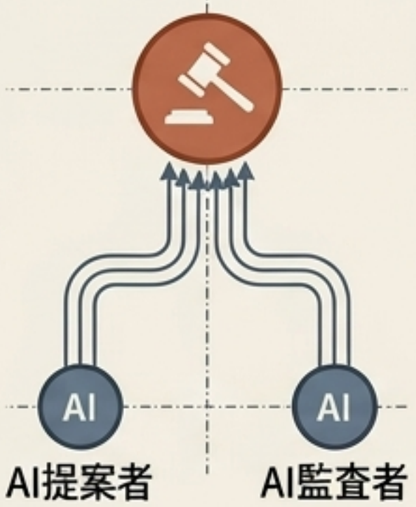
AIが数百の候補を生成し、人間が評価・選択。
創薬・材料向け。

3. ゲート付き閉ループ型



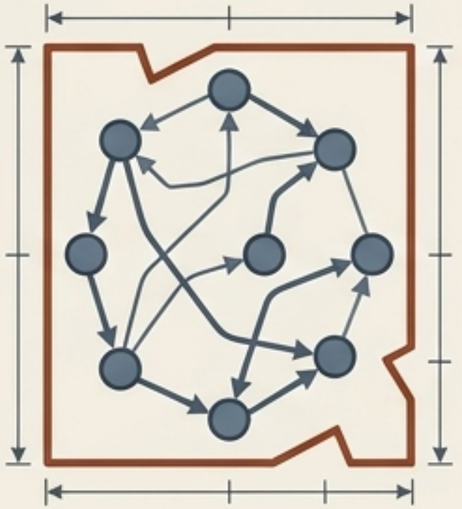
AIが高速反復し、人間は段階移行のゲートのみ承認。
標準化された実験向け。

4. 提案者対監査者型



AI同士が提案と反証を行い、人間が裁定。
AIのシステム的バイアス抑制向け。

5. 社会的ミッション型



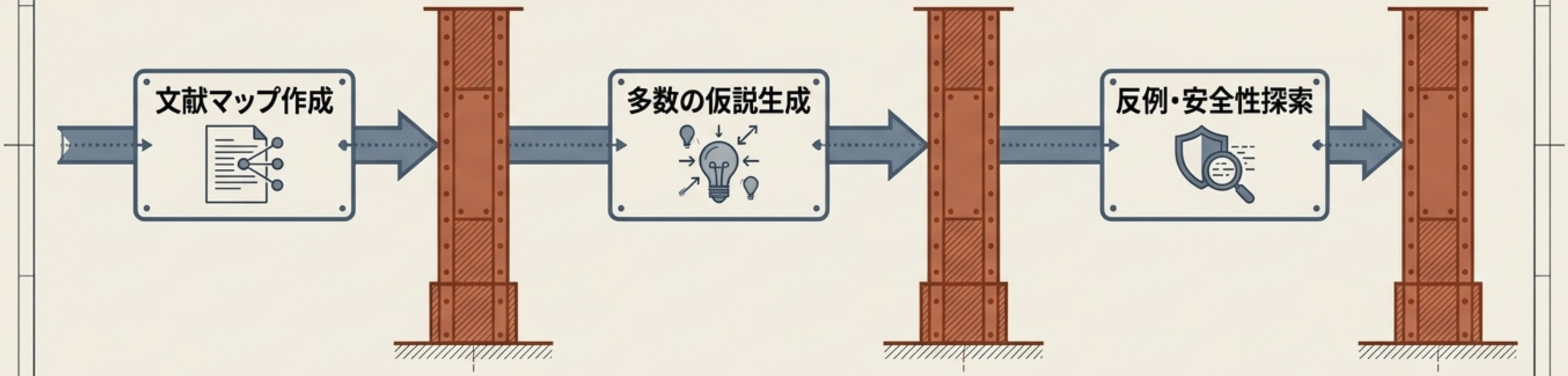
人間が公共的目標と境界を設定し、AIがその内部で最適化。
公衆衛生向け。

理想的な知的分業パイプライン

Gate 1:
研究する価値があるか？
(重要性・予算・倫理の審査)

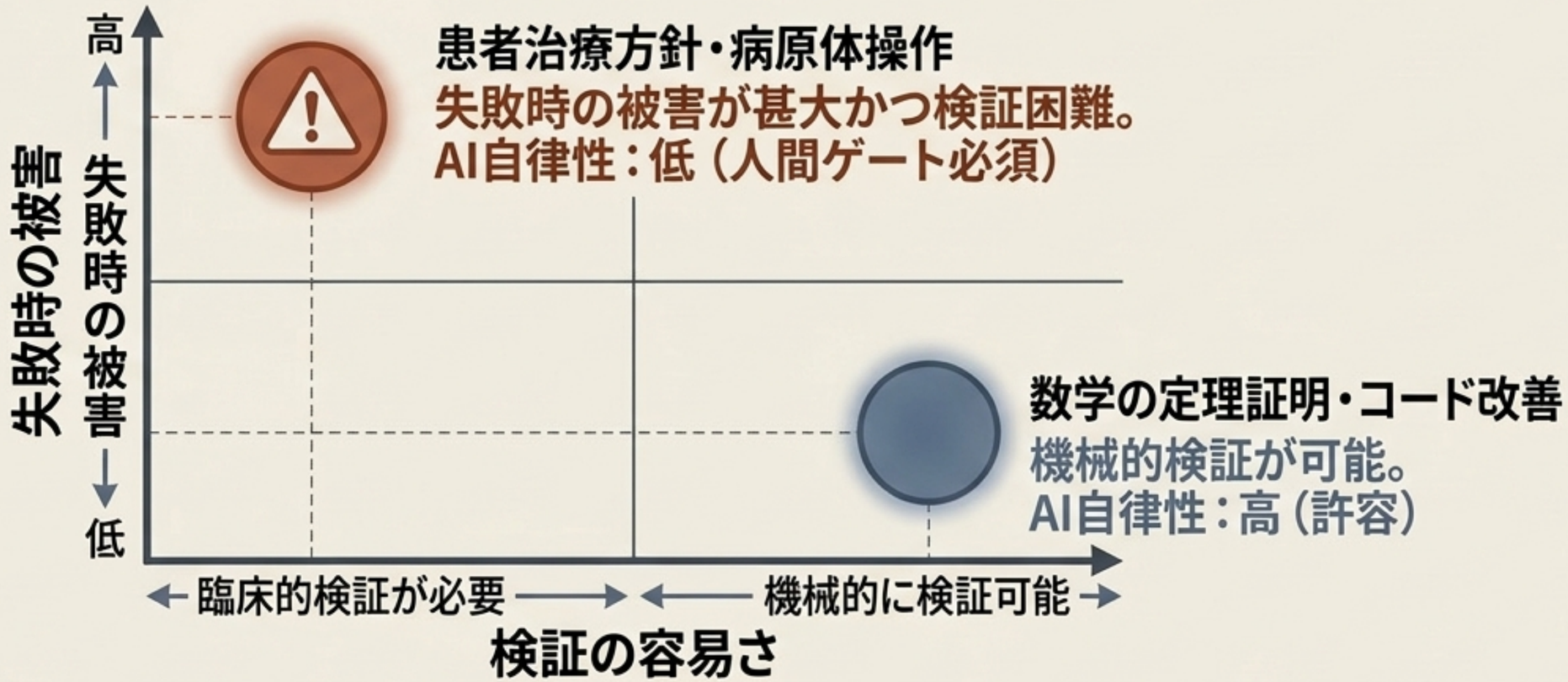
Gate 2:
実験する価値があるか？
(反証可能性・事前登録・安全性)

Gate 3:
知識として主張してよいか？
(独立再現・統計の監査・最終責任)



人間は全出力を読む必要はない。
「ゲート」を管理し、責任を負うことに集中する。

リスクマトリクス



許容できるAI自律性 ∝ (検証の容易さ) / (失敗時の被害)

自律性は「AIの賢さ」ではなく「監査の難易度」で決めるべきである。

制度再設計: 資金配分

Before: 従来の科学経済

従来: 「派手な新規性」「アイデアの数」の評価

結果: AI時代にこれを続けると、検証負債と幻覚論文の大量生産を招く。

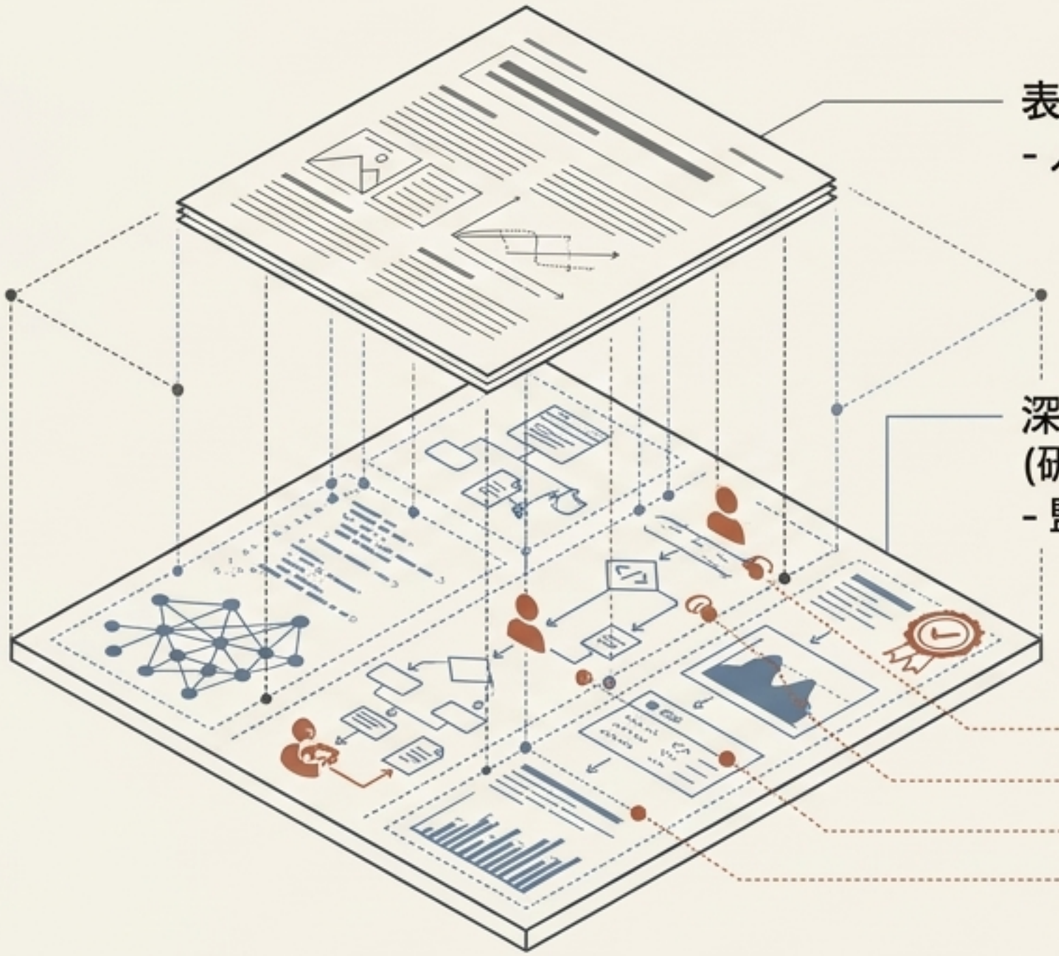
After: AI時代の科学経済

AI時代:
「情報価値の高い検証」
の評価へ資金をシフトする

- ✓ 独立再現実験への予算配分
- ✓ Negative Results（失敗データの公共財化）の支援
- ✓ 高品質なデータセットと共通ベンチマークの構築

※制約事項: NIH方針（NOT-OD-23-149）に準じ、未公開の申請書を公開LLMへ投入することは機密保持違反。評価基盤は閉鎖環境が必須。

学術記録のレイヤー構造



表層: 論文（The Paper）
- 人間が読むナラティブと主張

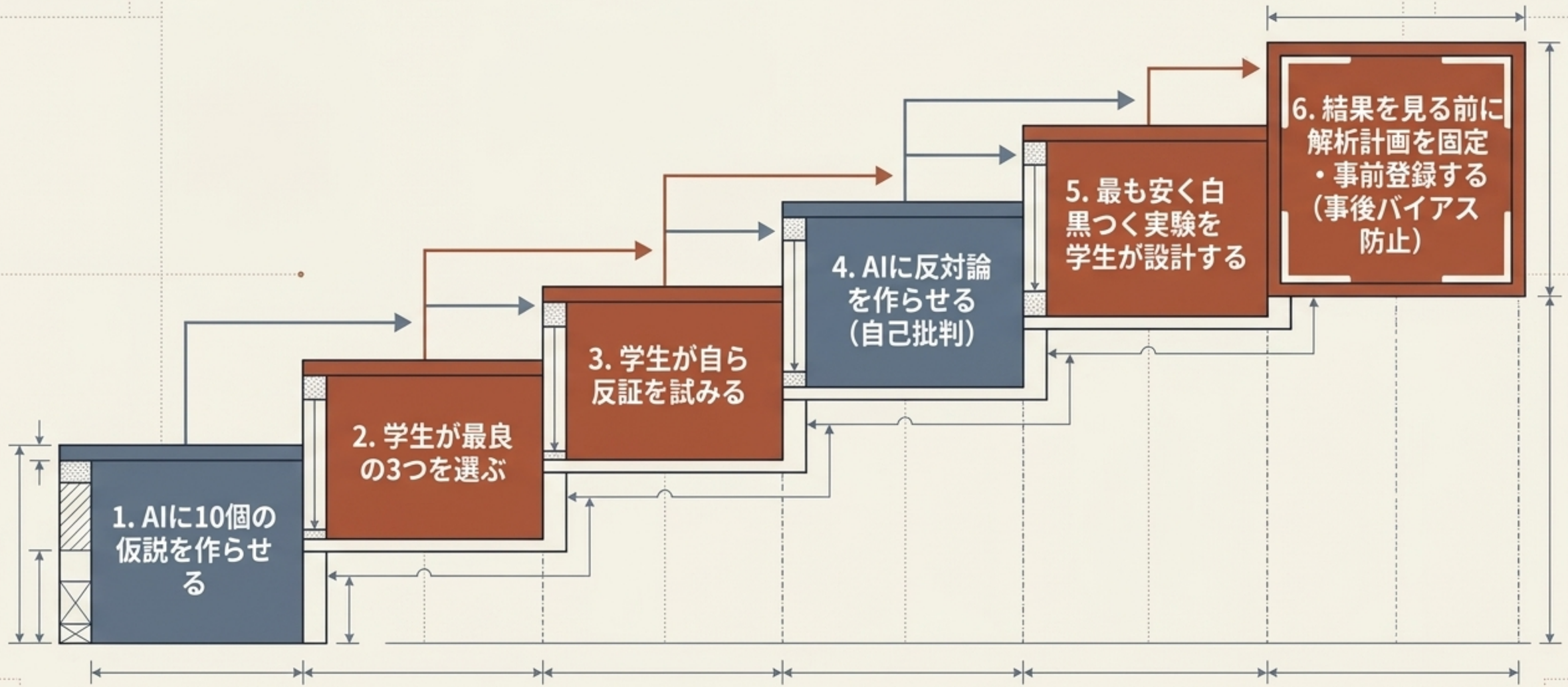
深層: AI Methods Statement
(研究トレース / 実行ログ)
- 監査可能なプロセスの証明

使用モデル / バージョン
初期プロンプトと目標設定
アクセスした文献リスト
棄却された数百の仮説候補
AI実行の完全なログ

Nature誌方針：「学術的判断と説明責任は人間から移転できない」。
最終論文の体裁だけではAIのデータリークや指標誤用は見抜けない。

教育のシフト: 専門知識の役割は「記憶」から「検証」へ

「AIを使うな」ではなく、「AIが最も間違いやすい実験を設計させよ」



アクション・マトリクス

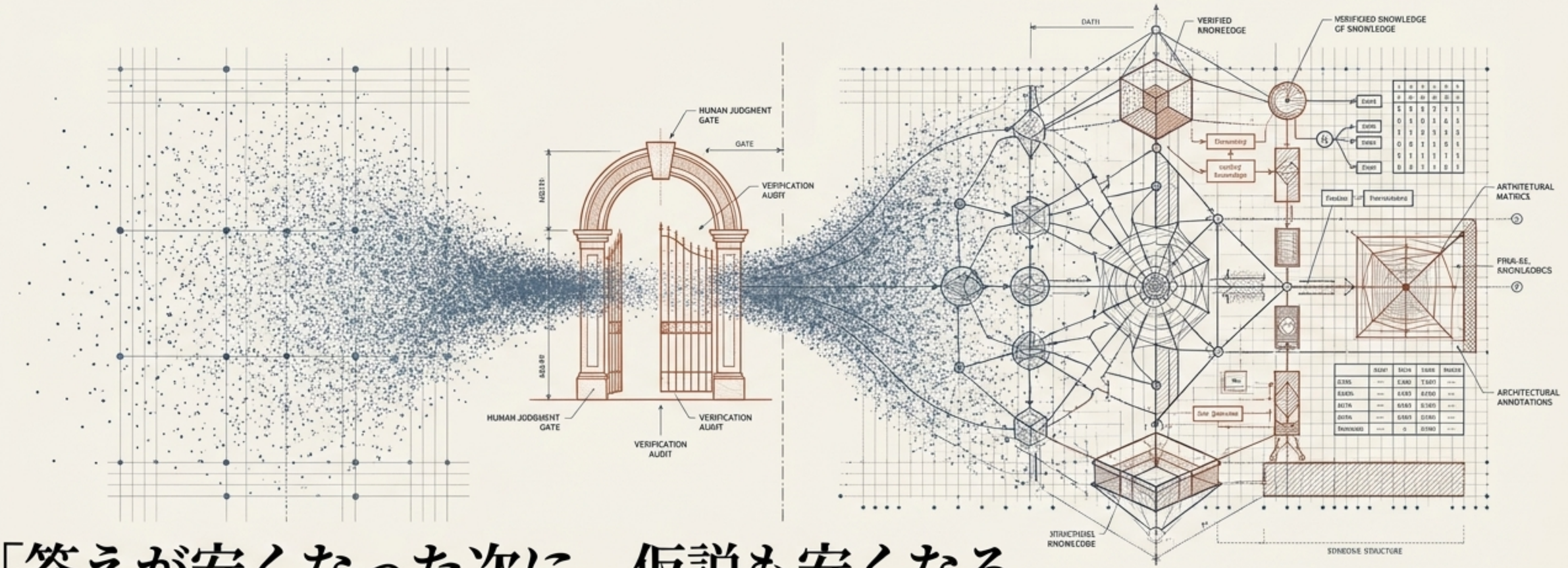
ステークホルダー	直ちに始めるべきこと (DO)	避けるべきこと (DON’T)
研究者	AI仮説をポートフォリオとして扱い、反証・事前登録を重視する	AIのランキング上位＝真実と妄信する
大学・研究所	承認済みAI環境の整備、モデル版と介入ログの保存	AIの一律禁止。または非公開データの公開AIへの無承認投入
資金配分機関	再現研究、Negative results、高品品質データ基盤を支援する	AIが生成した「派手な新規性」の量だけで評価する
学術誌・学会	AI利用申告と、実行ログ(Provenance)の提出を要求する	AI生成論文を、従来の人間向け書類審査のみで通過させる

~~優秀な科学者 ＝ 思いついた仮説の質~~

優秀な科学者 ＝ 重要な問題の選択
×探索の多様性×反証力
× 測定品質 × 測定品質
× 再現可能性 × 社会的正統性

競争優位は「何を考えつくか」から、
「世界に何を尋ねることができるか（独自の測定系・データ）」へ移行する。

「仮説を作るAI」は科学者を終わらせない。科学者を「アイデアの生産者」から「知識生成プロセスの設計・監査・責任者」へ進化させる。



「答えが安くなった次に、仮説も安くなる。
だからこれから高価になるのは、問いそのものだけではなく、
“何を信じるに値する知識として残すか”を決める能力である。」