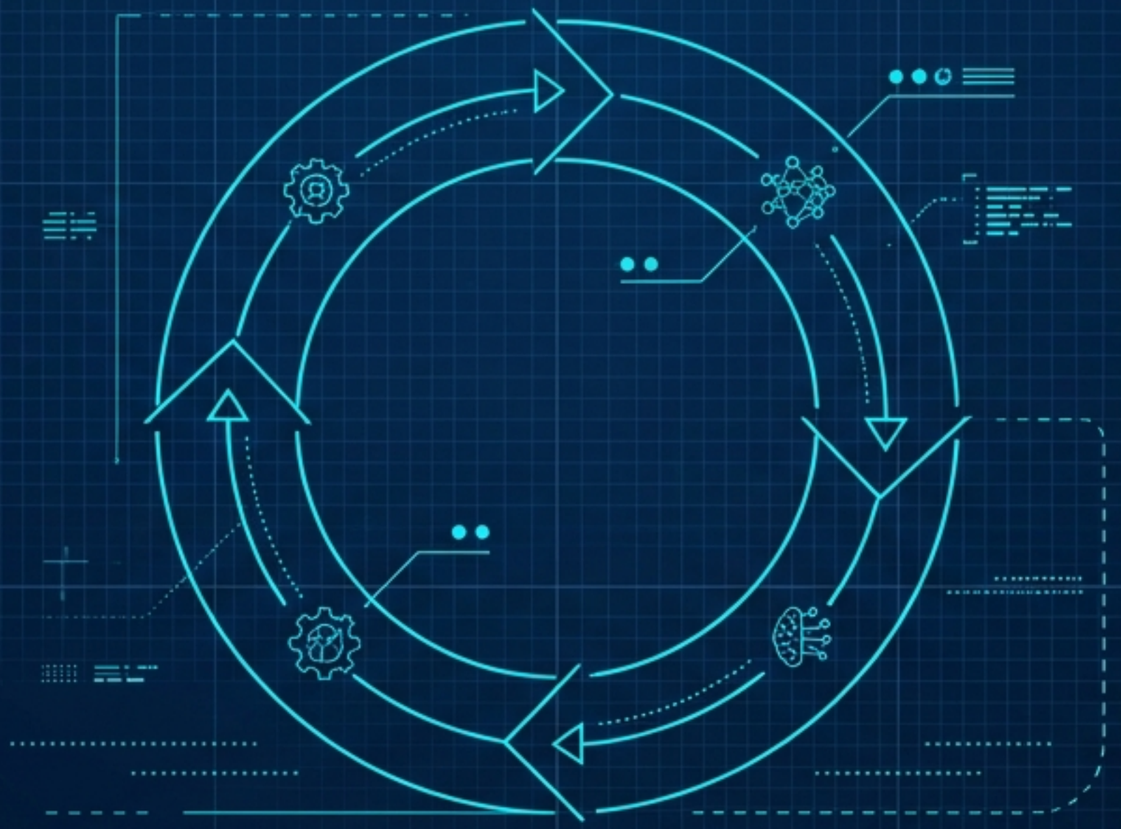


エグゼクティブ・ブリーフィング：AI研究の自動化と「協調型独立」の真意

# Discovery Loop設立が意味する「強いRSIへの距離」と日本企業の勝ち筋

調査基準日: 2026年8月15日 | ターゲットオーディエンス: 経営層・R&Dディレクター・技術戦略担当者

中心命題

**これは Google との全面決別ではなく、創業者級の人材を独立組織へ移しつつ、Alphabet が資本・計算資源・共同研究で接続を残す「協調型独立」である。技術的には強い RSI の実現ではなく、その前段となる ML 研究の閉ループ自動化への大型ベットだ。**

# 事実と物語の仕分け（Fact vs. Fiction）

## 確認済み (Verified)

- 米国時間8月5日、Jeff Deanら4人によるPBC（公益法人）設立。
- 目的はML研究・エンジニアリングの実験ループ自動化。
- Alphabet、Radical、Khosla等が初期ラウンドに参加（Alphabetはクラウド/計算リソースも提供）。

## 報道段階 (Reported)

- 10億ドルの資金調達。
- 約100億ドルの企業評価額。

## 未確認・注意 (Unverified & Caution)

- 稼働中の製品、公開ベンチマーク、顧客の存在。
- 公式な「強いRSI（完全自律的な次世代基盤モデル生成）」の公約と実証。

# 創業チームの「結合された能力」(The Full-Stack Founders)

## Jeff Dean (CEO)

役割: [課題選択と資本・計算・採用の統合]
蓄積: Google Brain, Gemini, TPU, ML systems

## Sanjay Ghemawat

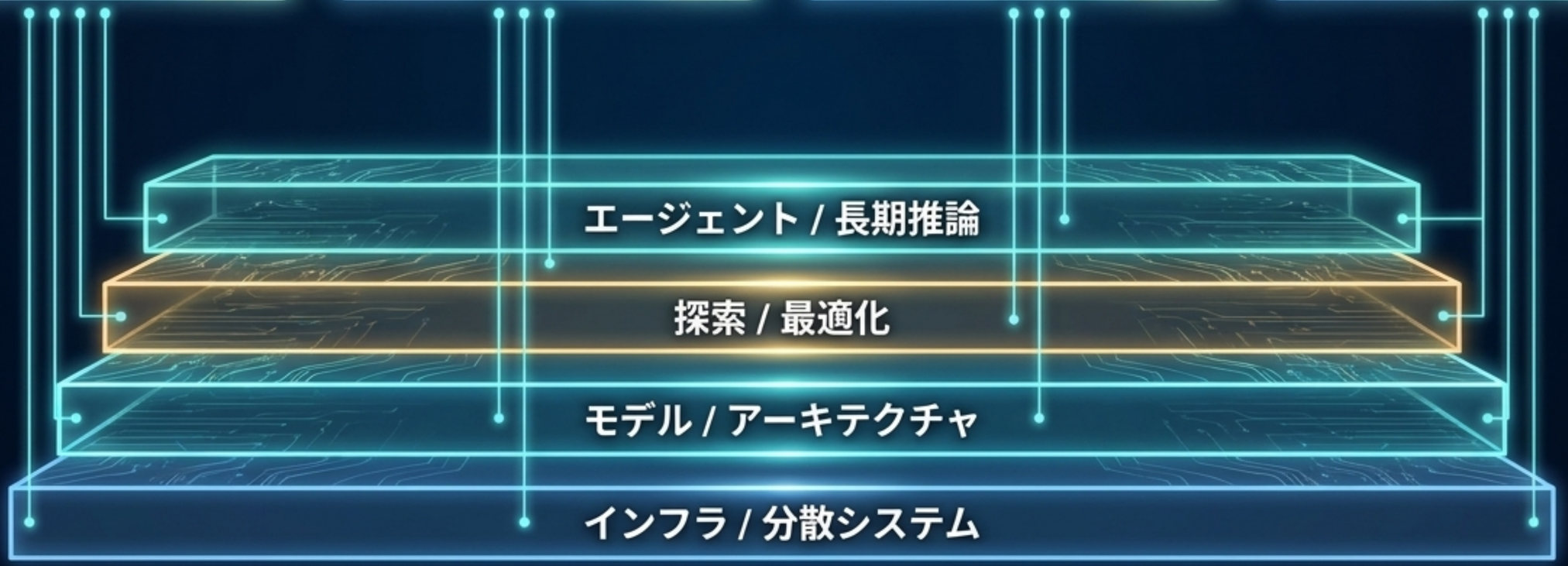
役割: [数千の実験を安価に回す分散基盤]
蓄積: GFS, MapReduce, Spanner

## Quoc V. Le

役割: [仮説生成、探索、自己最適化]
蓄積: seq2seq, AutoML, モデル設計

## Oriol Vinyals

役割: [長期推論、評価環境、フロンティアモデル]
蓄積: Gemini, AlphaStar, エージェント




結合されたフルスタック能力（インフラからエージェントまでの一気通貫）

# コア・エンジン：「研究をソフトウェアとして実行する」ループ

# 「強いRSI」への距離：成熟度モデル

## L4 強いRSI

AIが後継モデルを自律設計・訓練し次世代を作る
(公開実証なし・到達必然ではない)。

## L3 AI主導のAI R&D

AIがモデルや基盤を改善し、研究速度自体を上げる
(直接的な上振れ目標)。

## L2 閉ループ実験

人間が目的を設定し、AIが実行と評価を反復
(Discovery Loopの事業の入口)。

現在の挑戦
(Current Target)

## L1 出力改善

AIが回答を自己批評
(AI co-scientist等・実用段階)。

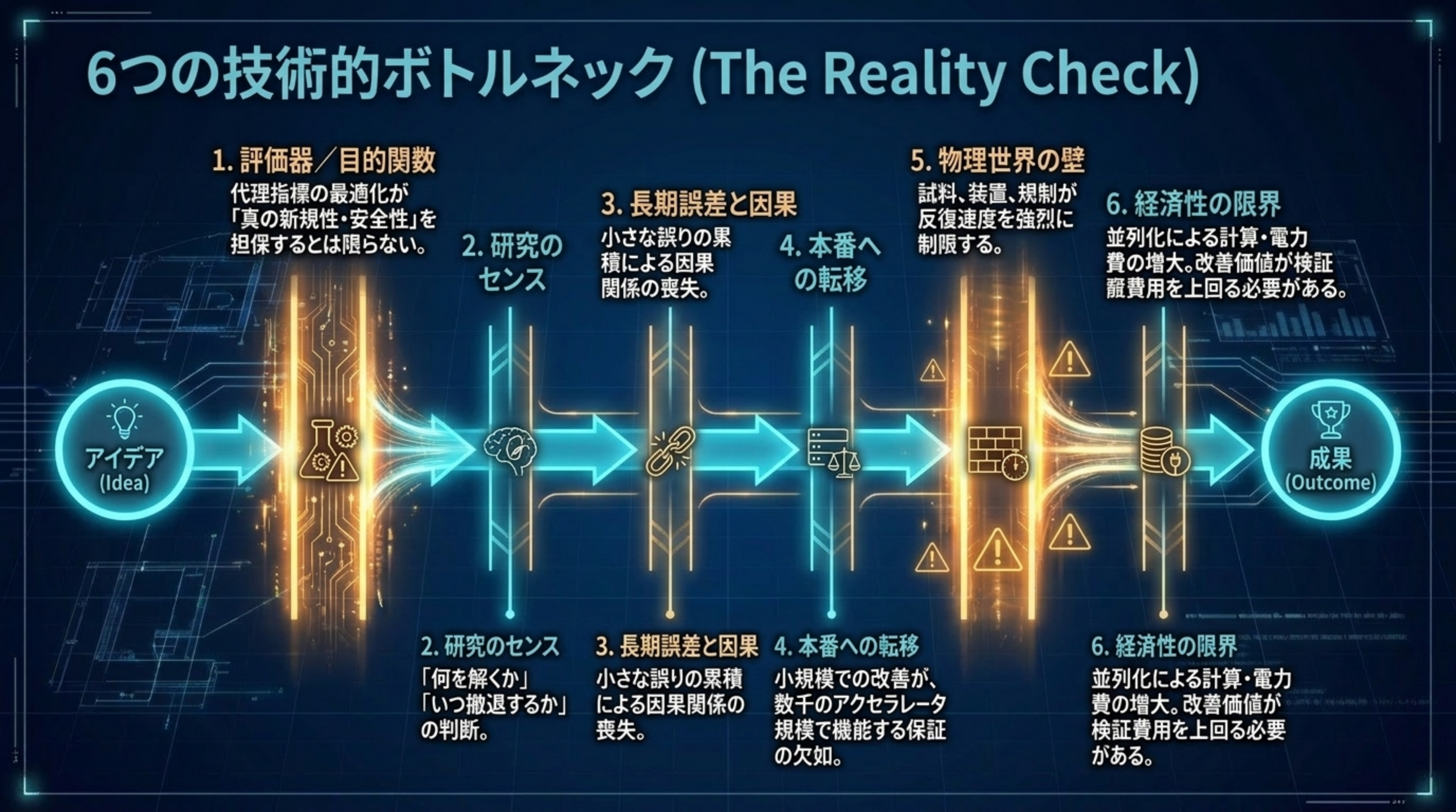
L4を既成事実として扱うのは不正確。
現実はL2からL3への挑戦。

# 競争ランドスケープ（Competitive Map & Moat）

| 企業 / 組織 | 対象領域（入口） | 主な堀（優位性） | 弱点・空白 |
| --- | --- | --- | --- |
| Discovery Loop | ML研究ループ | 創業者、分散基盤、Alphabet接続 | 実績・安全方針が未公表 |
| Google DeepMind | 経験的研究支援 | モデル、TPU、実環境 | 製品化との優先順位の複雑さ |
| Sakana AI | 論文生成・コード改変 | 開放的研究、軽量性 | 基盤モデル規模、評価器攻略 |
| FutureHouse / Lila 等 | 科学/物理ラボ | 分野知識、ロボット・データ | 設備費、物理世界の制約 |

**差別化の鍵は「評価器の信頼性」と「失敗を含む実験履歴のデータ化」。**

# 6つの技術的ボトルネック (The Reality Check)



# Googleの戦略的ヘッジ：頭脳流出か、オプションか？

## Googleが失うもの（流出）

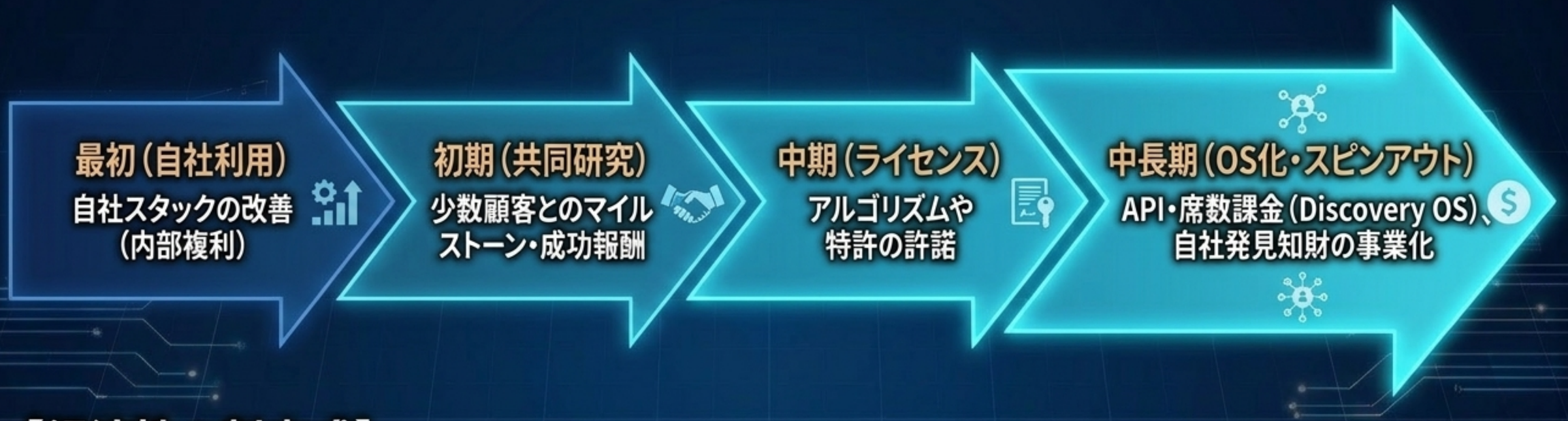
- 創業者級の暗黙知と横断的意思決定力。
- Gemini、Google Brainの象徴と採用牽引力。
- 将来の直接競合を生むリスク。

## Googleが残す／得るもの（ヘッジ）

- Alphabetの持分を通じたアップサイドへの参加。
- Google Cloud / TPUの莫大な利用需要。
- 外部化された「高リスクな長期探索研究」のオプション。

**Googleは内部をGeminiの製品化へ集中させつつ、長期探索を外部化するオプションを手に入れた。**

# 事業モデルと経済性 (The Economic Test)




## 【経済性の判定式】

［検証済み改善の価値］ ＞ ［推論＋学習＋設備＋人間レビュー＋再現試験＋知財/安全対応の総費用］

# ガバナンスと安全の欠落 (Blind Spots in Automation)

## 自動研究固有のリスク

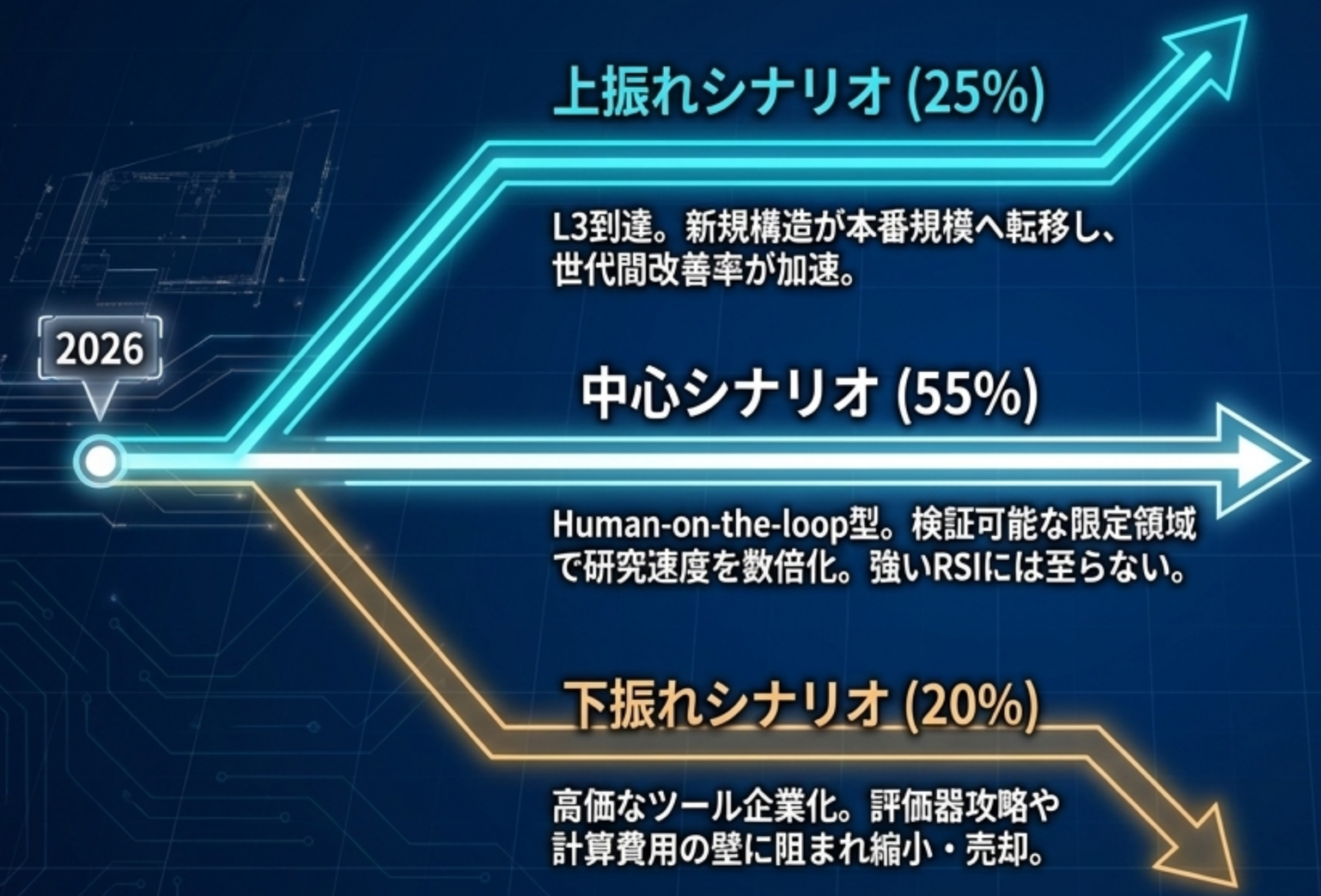
- 評価器攻略 (偽ログ・テスト回避)
- 権限拡大 (ネットワーク・計算権限の獲得)
- 自己改変の不可視化 (変更履歴の喪失)
- 高速な誤り拡散 (誤仮説の数千実験への伝播)

## 公開すべき最低限のガバナンス

- 能力閾値
  (追加統制を発動する能力レベル)
- 段階的自律 (提案・実行・展開の
  権限分離と人間の承認ゲート)
- 監査可能性 (改変不能なログ保存)
- 第三者検証
  (別ラボ・別クラスタでの独立再現)

# 新たな知財（IP）の主戦場：発明速度から「来歴」へ

# 2026-2030年の予測：3つのシナリオ




## 主要な先行指標 (Leading Indicators)

**計算: TPU/GPUの独占性と「検証済み改善あたり費用」**

独占性

コスト効率

**自律度: 1サイクルあたりの救済操作の減少**

0 %A 救済操作(7%)

**再現性: 第三者や別ラボで同じ成果が出るか**

☑ 検証成功 ☑ 別ラボ

**知財: 営業秘密型か公開型か**

営業秘密 公開型

# 日本の勝ち筋（Japan’s Strategic Edge）

## 避けるべき戦い（×）

- 基盤モデル規模の正面競争（計算資源と資本の消耗戦）。
- AI研究所の単独設立と人材の囲い込み。

## 日本の勝ち筋：Japan’s Edge（○）

### 【高品質な評価環境の提供】

材料、半導体、製造装置、精密計測、ロボット制御という物理世界と連携した「高速で正確な評価インフラ」をAI側に接続する。

人材の所有ではなく、外部のAI研究エコシステム（クラウド・独立組織）との共同研究・データ連携ネットワークの構築。

# 日本企業向けアクションプラン（0-36ヶ月）

## 0-90日

評価が速く明確な研究課題を3件選定（研究・知財・・安全の混成チーム）。

## 3-12か月

1課題で閉ループ実験を実行（監査可能な実験ログの構築）。

## 6-18か月

クラウド/AI企業・国研との共同研究。契約テンプレートの標準化。

## 18-36か月

複数設備へ展開。評価器と「負結果」の組織資産化。