

ARC-AGI-2 & The Future Future of of IP Strategy

流動性知能の突破と「知財業務の信頼性閾値」に関する分析



AIの進化は「知識」から「推論」へ移行し、特許実務の核心に到達した



01. The Breakthrough (技術的突破)

2025年3月の「0%」から、
2026年2月には「84.6%」へ。
Gemini 3 Deep Thinkが人間平均
均（60%）を大幅に超過。



02. The Application (実務への適用)

「抽象推論」能力の向上により、
キーワード検索を超えた「概念的
先行技術調査」と「進歩性判断」
が可能に。単純なLLM呼び出しで
はなく、構造化された推論
(MARG等) が鍵。



03. The Verdict (結論)

推論能力は「あれば便利」ではな
く、IP業務適用の「必須要件」。
ただし、EPO（欧州特許庁）が警
告するように、ハルシネーション
リスクへの専門家による検証は依
然として不可欠。

ARC-AGI-2は、暗記不可能な「流動性知能」を測定する唯一の指標である

結晶性知能 (Crystallized Intelligence)



- 記憶された事実 (Memorized Facts)
- 事前学習データへの依存
- 例：司法試験の合格、条文の暗記

流動性知能 (Fluid Intelligence)



- 未知の問題解決 (Solving Unknown Problems)
- パターンの抽象化
- 例：特許の新規性論証、発明の概念化

定義: François Cholletによって提唱された、事前知識に依存せず未知の問題を解決する能力の測定。

IP業務との関連: 過去の判例（データ）を暗記していることと、新しい発明（未知のパターン）の論理構造を理解することは全く異なる能力である。

2025年初頭まで、AIは論理的推論において「全滅」状態だった

The "0%" Reality:

2025年3月時点で、推論モード非搭載の純粋なLLM (GPT-4o, GPT-4.5) はスコア **0%** を記録。



0%

「歯が立たない」

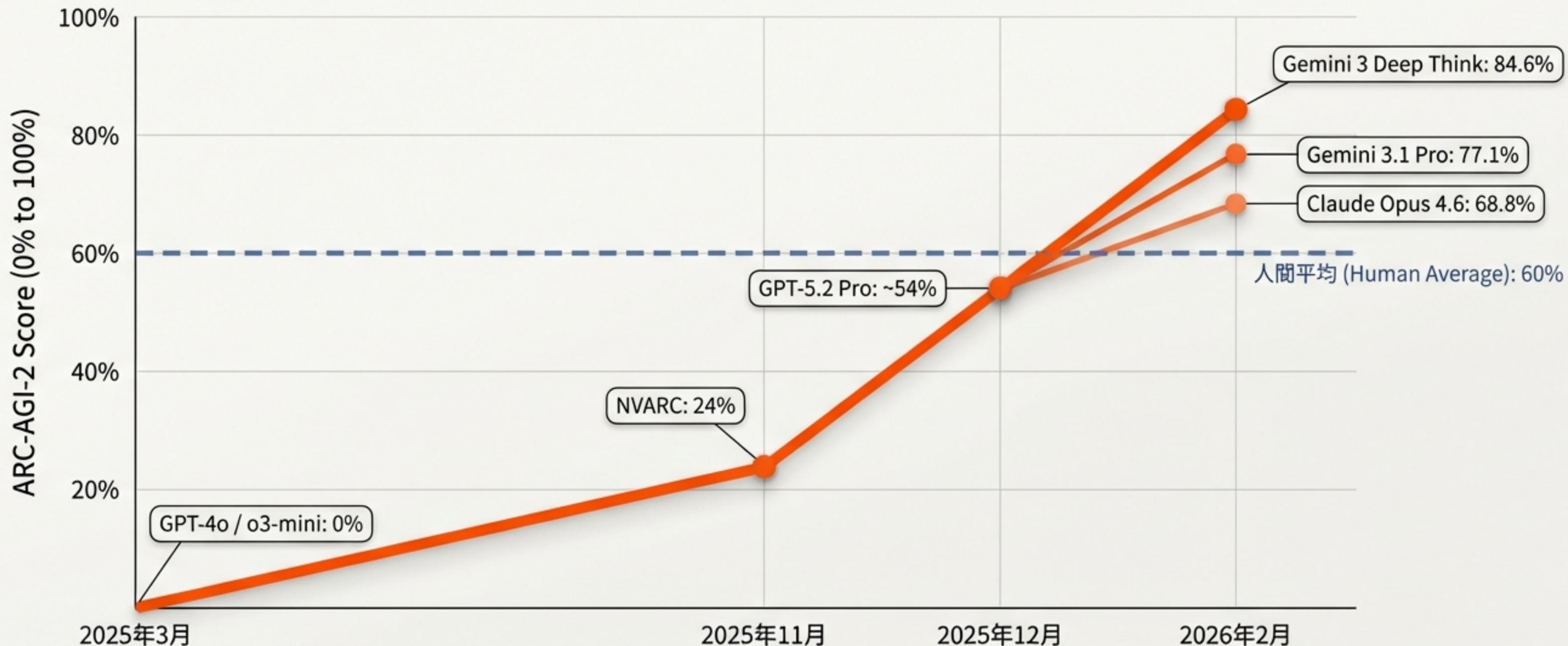
— TechnoEdge

「全滅」

— Nazology

単純なパターンマッチング（システム1）だけでは、複雑な規則性の抽出や適用が不可能であり、この段階のAIは特許の「構成的推論」に不適合であった。

2026年、Gemini 3 Deep Thinkが「人間平均」の壁を突破した



わずか11ヶ月で、AIは「不能」から「超人」の領域へ到達した。

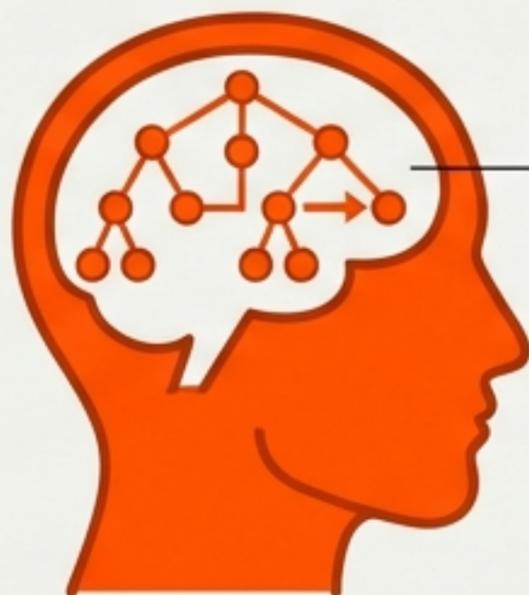
「システム2（熟考）」の実装が、推論能力の飛躍を生んだ



System 1 (Old AI)

直感的・高速・パターン認識

ARCスコア 0%



System 2 (New AI)

論理的・低速・推論拡張
(Test-time compute)

ARCスコア 84.6%

ARC-AGI-2の要求能力：

1. シンボルの意味解釈
2. 多段階の構成的推論
3. 少数事例からの一般化

Cost Efficiency: GPT-5.2はo3モデルと比較してコスト効率が390倍改善。実務利用可能なコストで「熟考」が可能になった。

抽象推論能力 (ARC) は、特許実務の質を決定する 重要スキルと直結する



シンボルの意味解釈
(Symbol Interpretation)



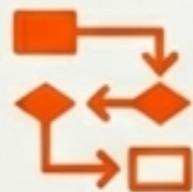
キーワードを超えた「概念的」
先行技術調査
(Semantic Prior Art Search)



少数事例からの一般化
(Generalization)



異なる技術分野からの
転用・類推
(Analogical Reasoning for Obviousness)

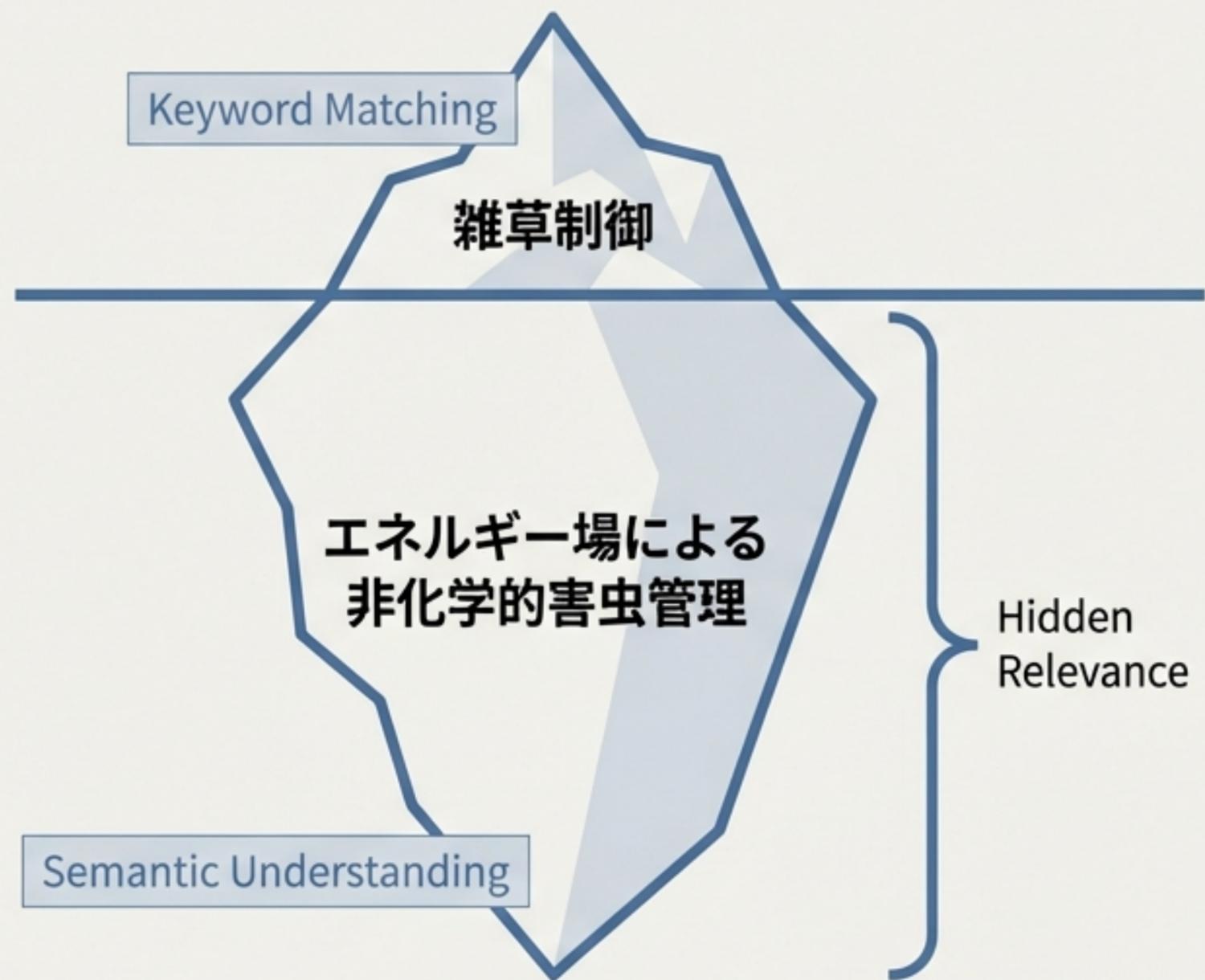


多段階の構成的推論
(Compositional Reasoning)



請求項 (クレーム) の
構成要件解釈と侵害判断
(Claim Construction & Infringement)

先行技術調査：表面的な「キーワード」から、本質的な「意味理解」へ



Deep Dive 1: Prior Art

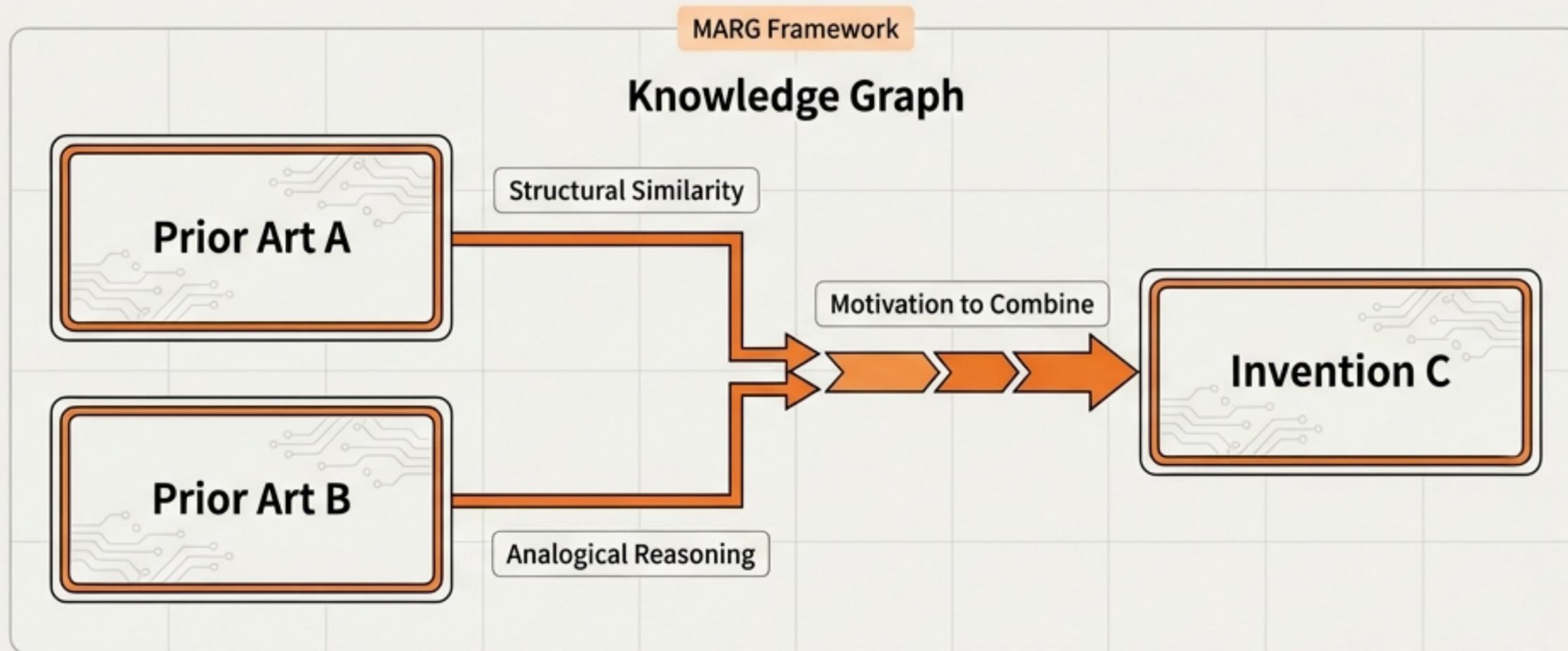
Key Study: Ikoma & Mitamura (Carnegie Mellon Univ, 2025).

Problem: 従来の検索では「電磁パルスによる雑草制御」と「エネルギー場による非化学害虫管理」の関連性を見抜けない。

Solution: 生成型LLMは、特許と先行技術の間の「抽象'な概念の対応関係」を理解し、合理的な精度で新規性を予測可能。

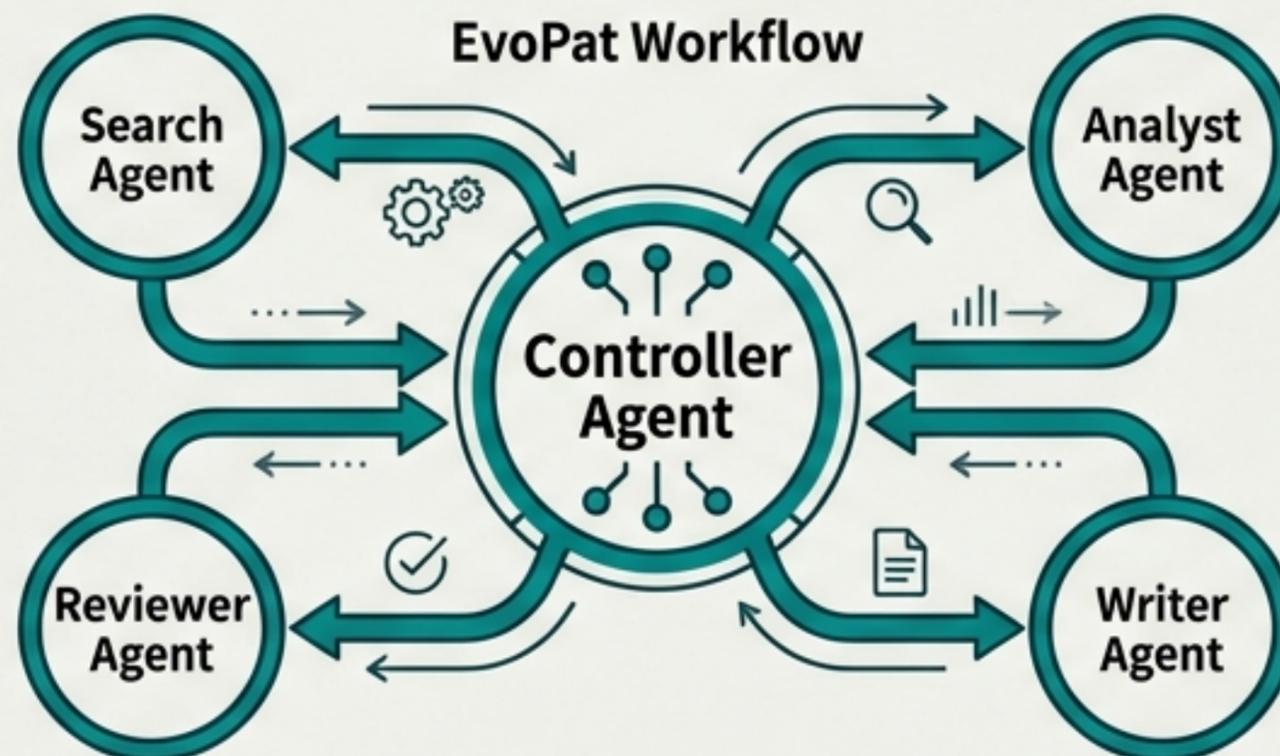
Evidence: 専門家による評価と高い相関を示し、特許の新規性判断において十分な説明を生成できることが実証された。

進歩性判断：「当業者」の思考プロセスを構造化された推論で模倣する



- Key Study: PatentMind (2025) & MARG Framework.
- The Challenge: 進歩性（非自明性）の判断は、引用文献の組み合わせが「自明」かどうかを評価する高度な類推推論。
- The Result: 多面的推論グラフ（MARG）を使用し、専門家の評価とピアソン相関 **0.938** を達成。
- Implication: ARC-AGI-2で測定される「未見の文脈への規則適用」能力は、この進歩性判断プロセスと構造的に同型（Isomorphic）である。

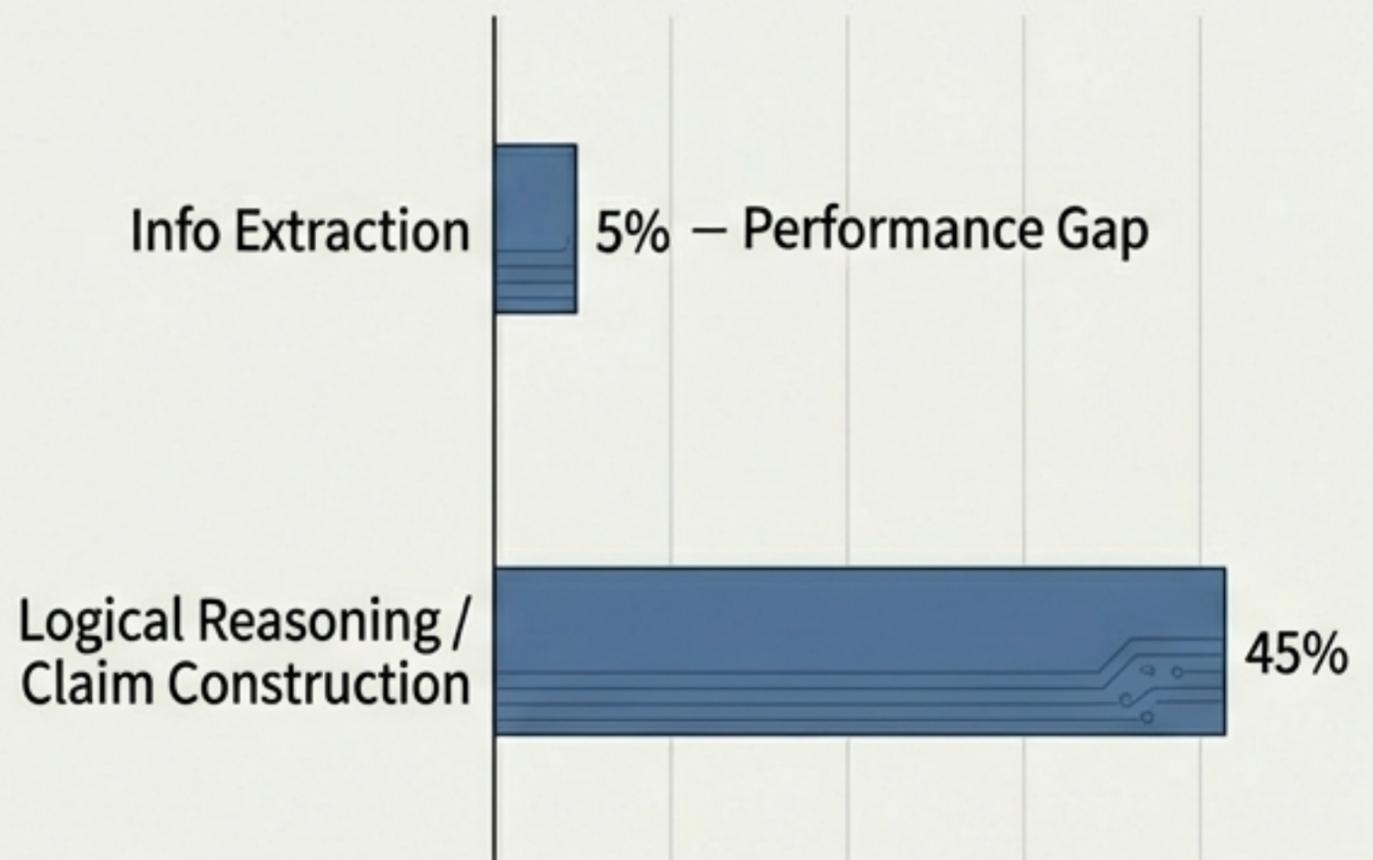
IPランドスケープ：マルチエージェント化が戦略分析の精度を最大化する



- **Key Study: EvoPat (2024).**
- **Evolution:** 単一のGPT-4よりも、役割分担されたマルチエージェントシステム（RAG活用）が、要約・比較分析・技術評価のすべてで優れた性能を発揮。
- **Capability:** 数千件の特許から競合の技術戦略をパターン認識し、「ホワイトスペース（空白地帯）」を発見する。
- **Future:** ARC-AGI-3（2026予定）は「対話型推論（探索・計画）」をテストするため、エージェント型IP業務への移行と方向性が一致する。

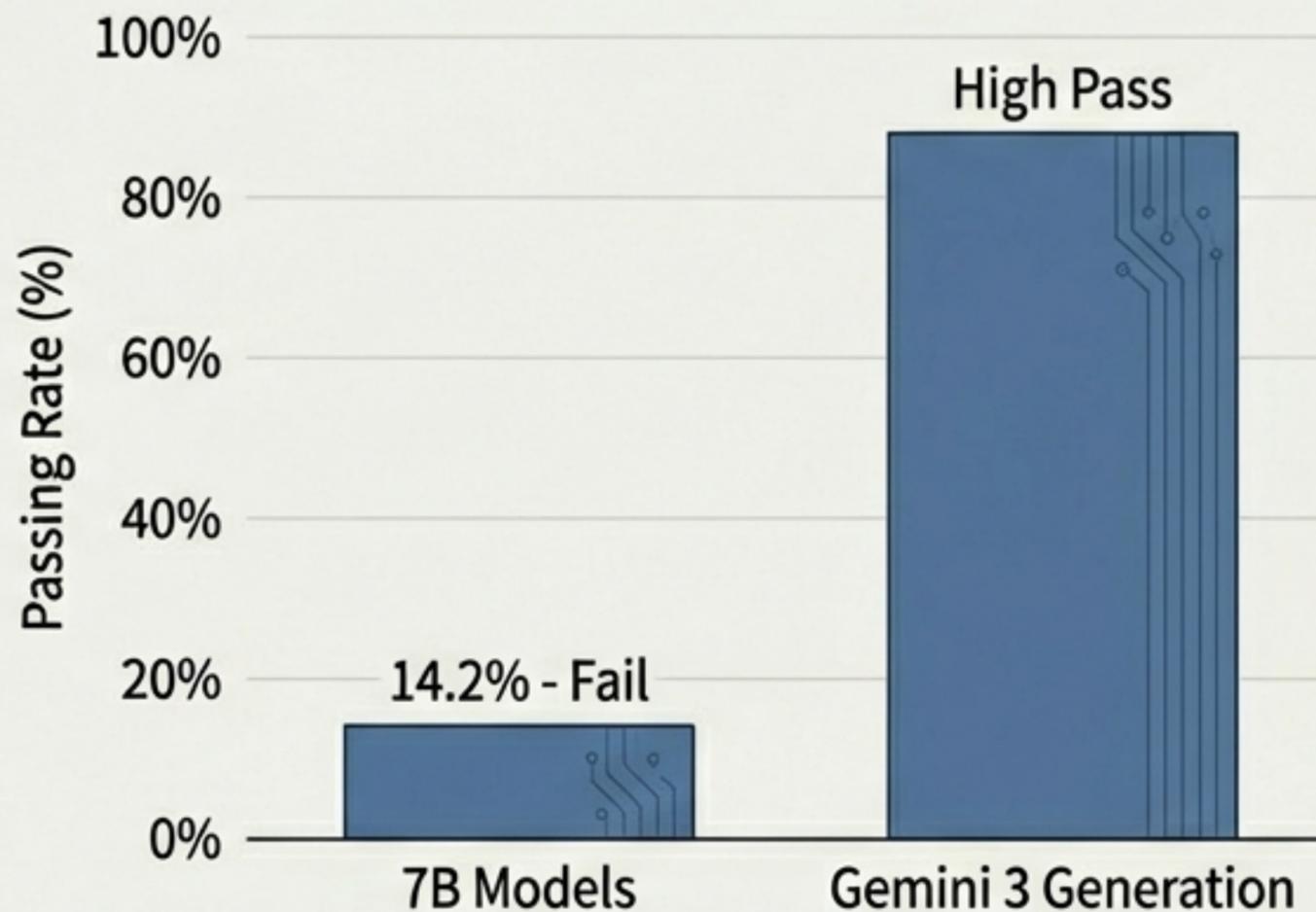
「推論力の差」は、高次の知財タスクにおける性能差に直結している

IPBench (2025) - Task Complexity



高次タスクほど、高性能モデルの優位性が拡大する。

MOZIP (2024) - Passing Rate



古いモデルはランダム推測以下だったが、Gemini 3世代で専門的推論の壁を突破。

日本のIP業界におけるAI実装は、官民双方で加速している



日本特許庁 (JPO): AI関連特許出願約11,400件 (2023年度)。「AI時代の知的財産権検討会」中間報告 (2024年5月)。



Tokkyo.Ai: 対話型検索「ChatTokkyo」や思考過程可視化「Deep Agent」を展開。



Emuni (Univ of Tokyo): Meta Llama-3-70B 微調整モデルが、特許翻訳でGPT-4o/DeepLを上回る精度(BLEU/RIBES)を達成。



Community: Reinforz Insight等の研究者が、ARC-AGI-2の「演繹と帰納」の違いを鋭く分析。

リスク管理：推論モデルは「ハルシネーション」を低減するが、排除はしない



Survey Data

特許弁護士の最大の懸念は「精度とハルシネーション」(42%) (HGF Webinar, 2025)

EPO Warning

欧州特許庁 T1193/23 は、追加的証拠なしにLLMのクレーム解釈を信頼することに警告。

推論能力の高いモデルほどハルシネーションは少ないが、実務においては「AI の出力」を専門家が検証するプロセスが必須であり、完全自動化は時期尚早である。

次なる進化：ARC-AGI-3と「エージェント型AI」への移行

 Current State

Chat Interfaces
(対話型)

2026 Milestone

ARC-AGI-3
Release.

 Future State

Agent Workflows
(自律型)

- テスト項目: 対話型推論、探索、計画、記憶。

- 段階的な調査戦略の立案。
- 仮説検証の反復ループ。
- 「チャット」から「ワーク (自律的な仕事)」へ。

現在の「推論」モデルは、将来の「自律」エージェントへの通過点に過ぎない。

結論：「信頼性閾値」を超えたAIを、 いかに実務に統合するか

- **Capability:** ARC-AGI-2 (84.6%) の達成により、AIは特許実務に必要な「抽象推論力」を獲得した。
- **Validation:** 高次推論タスク（進歩性評価・クレーム解釈）において、モデルの性能差は業務品質に直結する。
- **Action:** AI導入はもはや「効率化」の問題ではなく、競争に対する「品質優位性」の維持に関わる戦略的必須事項である。

ARCが測定する「真の推論力」と、IP業務が求める「法的正確性」の交差点に、次世代の知財戦略がある。