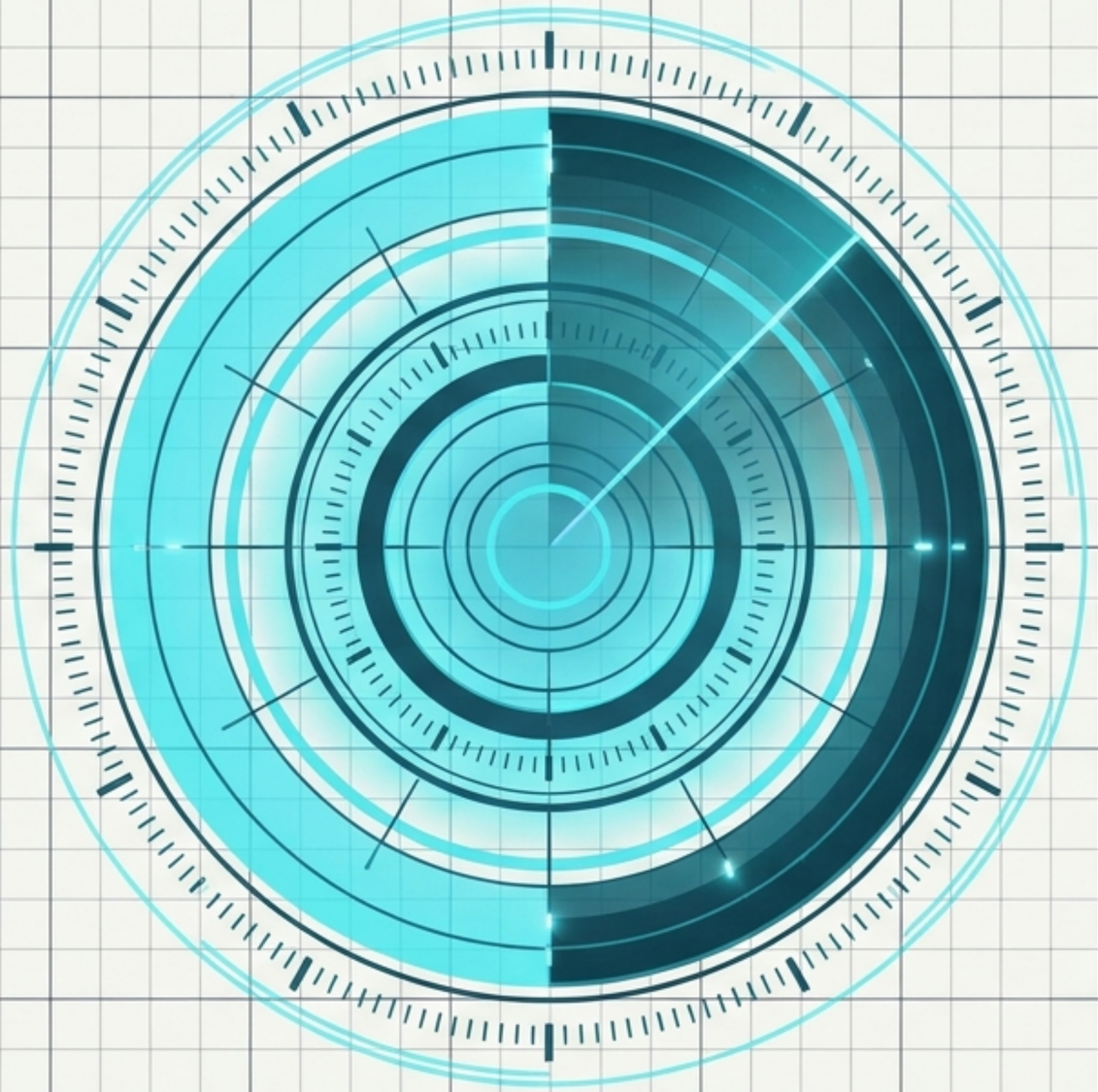


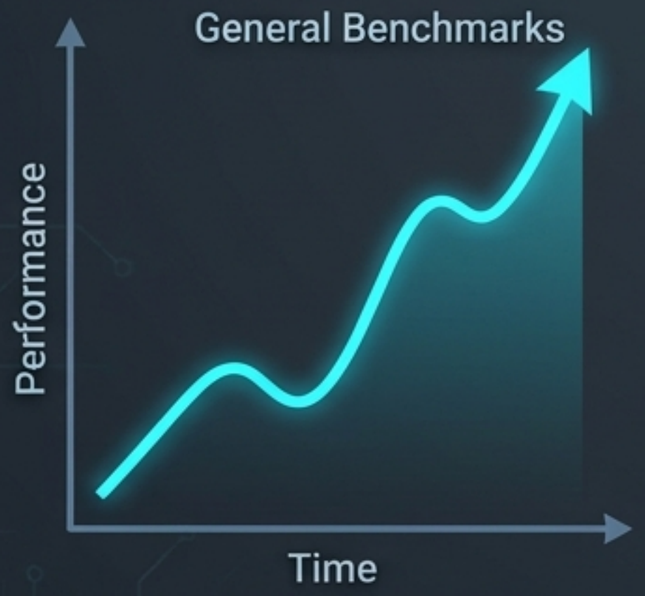
Claude Fable 5.2 導入の戦略的評価

「総合首位」の誇大広告を超えて：知財・研究開発部門のための費用対効果と実務導入ガイド

Target Audience: Corporate Strategy, R&D, IP & Legal
Assessment Date: 2026-11-10
[CONFIDENTIAL / INTERNAL DISTRIBUTION ONLY]

Gemini Notebook

状況 (Situation)



Fable 5.2は複雑な推論と長時間のエージェント作業において大きな進歩を示し、各種独立ベンチマークで総合首位を獲得。

課題 (Complication)



しかし「Max設定」での運用は費用・待ち時間が激増し、長文PDF処理では他社モデルに劣後。能力に強い「凹凸」が存在する。

解決策 (Resolution)



総合スコアではなく、タスクの複雑性に応じた推論設定（High/Medium）の最適化と、特許・知財実務に特化した独自の検証体制が必須。

Gemini Notebook

製品仕様とコスト構造：長期エージェント作業への特化



Specs (仕様)		API 料金体系 (100万トークンあたり)	
コンテキスト窓	100万入力 / 12.8万出力トークン	通常API	入力 $8 / 出力 $40
推奨用途	数日単位の開発、多数のエージェント統括		
日常業務への Anthropic公式見解	「まずはOpus 5.5からの開始を推奨」	高速モード	入力 $16 / 出力 $80

Gemini Notebook

ベンチマークの錯覚：推論設定（Inference Settings）の罠

【警告】Max設定はスコアを3ポイント上げるために、費用を2.9倍（High比）に増大させ、1課題に約46分を要する。日常的なチャット用途には不適。



スコア 56 | 費用 $3.80

スコア 62 | 費用 $9.08

スコア 65 | 費用 $26.54
処理時間 約46分/課題

Gemini Notebook

フロンティアモデル性能比較：特定ドメインにおける「強みと死角」

評価項目	Fable 5.2 Max	Gemini 4 Carbon High	GPT-6 Astra Max
科学研究 (Terminal-Bench-Science)	73.3%	64.8%	63.3%
ターミナル課題 (Terminal-Bench)	59.6%	65.2%	59.1%
コーディング (FrontierCode 2.0)	48.8%	40.6%	42.4%
長文PDF処理 (GDP.pdf)	24.2%	28.4%	31.0%

Strategic Insight: 総合首位は「全ての作業での首位」を意味しない。長文PDFを多用する知財・法務実務では、GPT-6 Astraが優位性を持つ領域が存在する。

Gemini Notebook

実証されたR&D能力：即時応答から「長時間自律作業」へのシフト

理論物理学 (Theoretical Physics)

ブラックホール関連の未解決問題2件を解決（Anthropic発表）。

電子機器の完全設計 (Hardware Design)

スピーカーフォン等の回路・基板・筐体・FWを設計。人間による回路修正なしで試作品が動作。

数學の形式化 (Math Formalization)

数學の形式化 (Math Formalization)

1,000のエージェントが約29時間稼働し、フェルマーの最終定理 (Leanによる証明) の形式化を再現。

Seconds (秒) — Hours (時間) — Days (日)

※既存定理の検証や設計工程の自動化であり、無からの新規発明を意味するものではない。長期耐久性等は未確認。

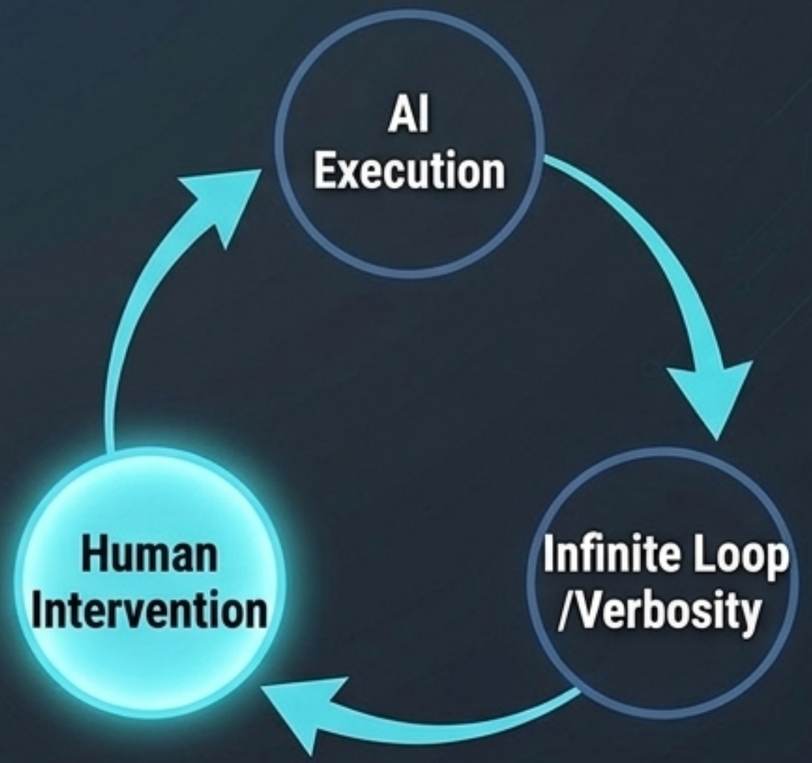
Gemini Notebook

先行レビューが示す「Jagged AGI（能力の凹凸）」と運用リスク



コンテキストの枯渇 (Simon Willison & Spring Prompt)

Max設定での創作・コーディング課題において、思考上限（12万8,000トークン）を使い切り、成果物を出力せずに失敗する事象が多発（29件中8件）。推論を下げることで完了。



監視の必要性 (Ethan Mollick)

幅広い課題で専門家に近い能力を示す一方、冗長化や過剰なルール作りを自己認識できず、人間の介入（Human-in-the-loop）が不可欠。

Gemini Notebook

タスク適合性スペクトラム：実践者による信頼度マッピング

アンケート分析と広告予算運用
(Spring Prompt)

Webデザインの単独設計
(Claire Vo)

得意領域 / 高い自律性
(High Reliability)

要注意領域 / 単独実行不可
(Requires Human Intervention)

動画制作、家具の組み立て支援、ゲーム制作 (Claire Vo)

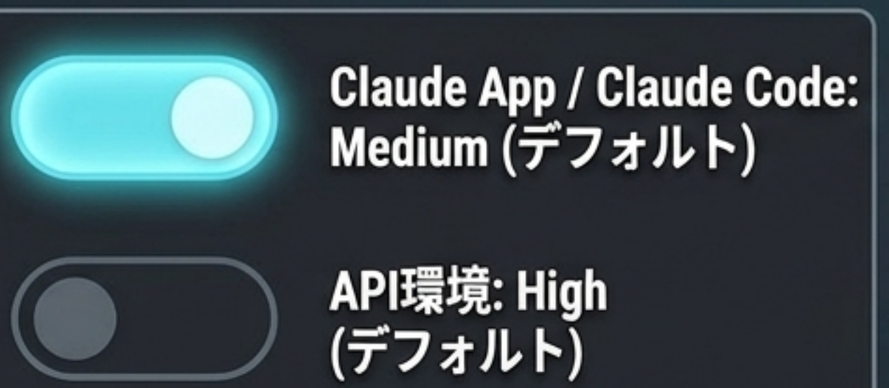
成果物の「長さ」「退屈さ」の自己評価と修正 (Mollick)

結論: 単発のプロンプトで全てを解決するのではなく、
エージェントワークフローの一部として得意領域に限定して組み込むべき。

Gemini Notebook

運用上の保護措置：初期設定と知財データへのリスク

環境別・初期設定の不一致



- Claude App / Claude Code: Medium（デフォルト）
- API環境: High（デフォルト）

※モデル名だけでなく「推論設定の記録」が比較検証において必須。

【重要】データ保持ポリシー (IP Risk)



- APIは原則「30日間のデータ保持要件」が存在。
- ゼロデータ保持はAnthropicの明示的許可が必要。
- 未公開発明や顧客資料を扱う際、既存モデル（Opus等）と同じ契約条件と盲信せず、適用条件の再確認が必須。

Gemini Notebook

法律ベンチマークのパラドックス：評価指標の解離

Artificial Analysis Index (Task-Based)

首位 (63点)

結果: Fable 5.2 Max設定

評価基準: 法律知識、推論、文書処理の組み合わせ



Vals Legal Index (Workflow/Revenue-Based)

20位 (69.17%)

結果: Fable 5.2 High設定
※37モデル中

評価基準: 米国法律事務所の業務別収入に基づくワークフロー重み付け

Core Insight: タスク単体の推論力は極めて高いが、実際の法律事務所の収益化ワークフローに当てはめると順位が急落する。総合スコアから特許実務への適性を直接判断することは不可能。

Gemini Notebook

知財・特許実務への導入戦略と独自検証プロトコル

Step 1: 適用対象の絞り込み

推奨領域: 複数文献を突き合わせる技術分析、複雑なクレーム対比、実験データの解析、長時間の調査工程の統括。

Step 2: 比較ベースラインの構築

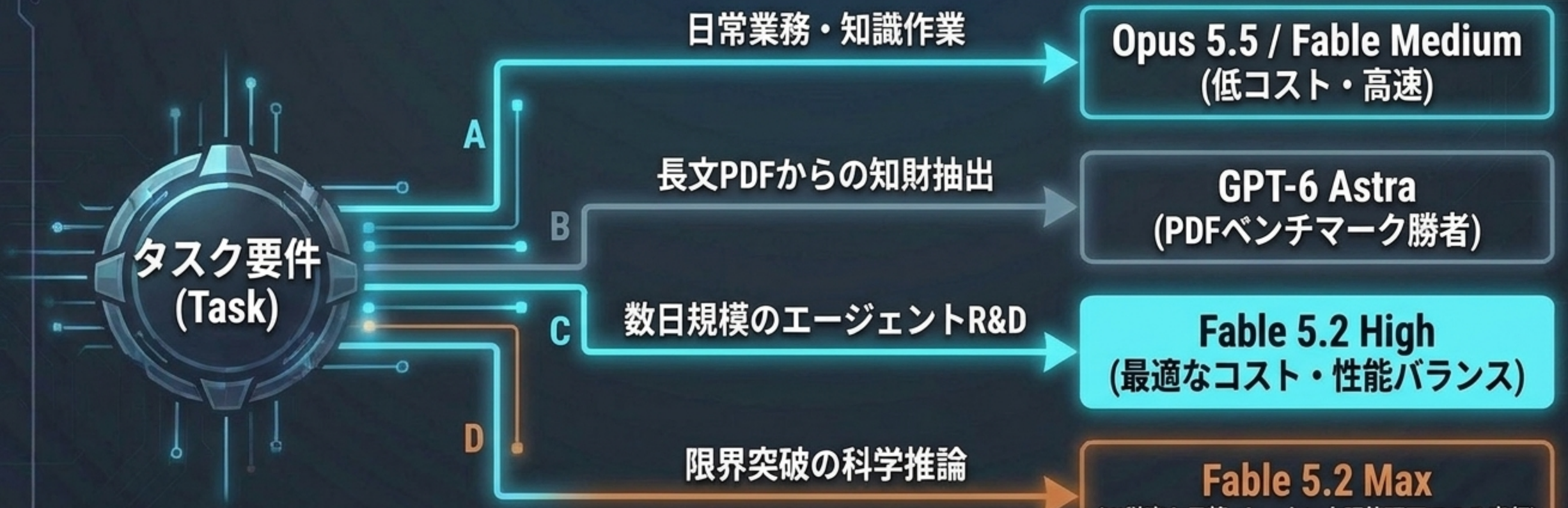
Opus 5.5などの既存モデルに対し、同じ公開済み実案件データと指示を与え、並行テストを実施。

Step 3: 知財特化型・評価軸の策定

1. 引用箇所の正確さ
2. 重要事項の見落とし
3. 技術的な読み違い
4. 修正に要した人間の時間
5. 処理ごとの総費用

Gemini Notebook

タスク複雑性に基づくモデル選定マトリックス



結論：Fable 5.2は万能薬ではなく、高度な専門性を備えた「精密機械」である。
用途に応じた推論設定の使い分けがROI最大化の鍵となる。

Gemini Notebook