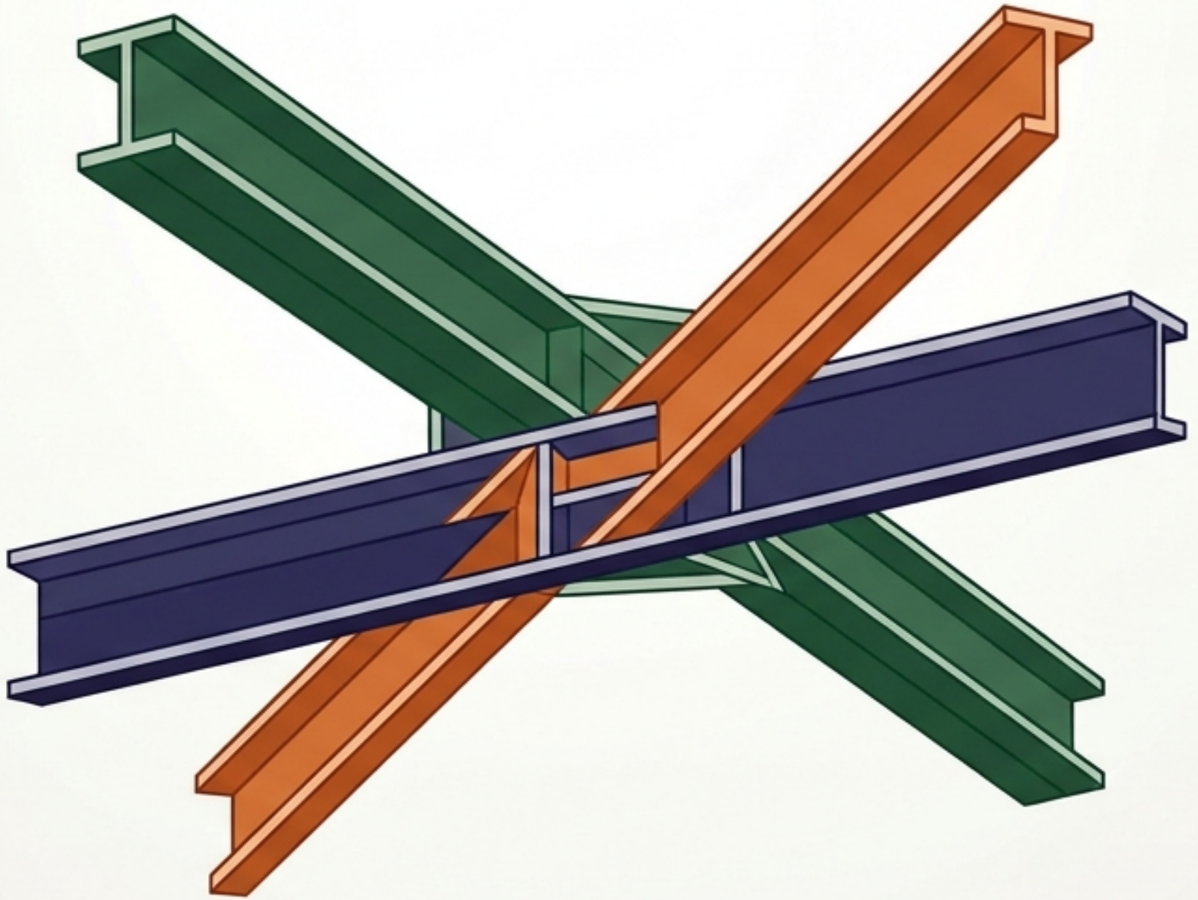


フロンティアLLM 実務適性・ROI比較と最適配置の設計図

GPT-6.2 Astra, Claude Fable 5.2, Gemini 4 Carbonの深層プロファイルとハイブリッド・アーキテクチャの構築

評価基準: Artificial Analysis Intelligence Index v5.0.1 (Nov 2026)

Gemini Notebook

総合スコアの僅差に隠された、実務ROIの決定的な非対称性

表面的な知能指標：トップ3の差はわずか5点



実運用への適用における3つのパラダイム

パレート最適の経済性

Astra: 極限の速度と低コストの自律実行

限界突破の論理と拡張思考

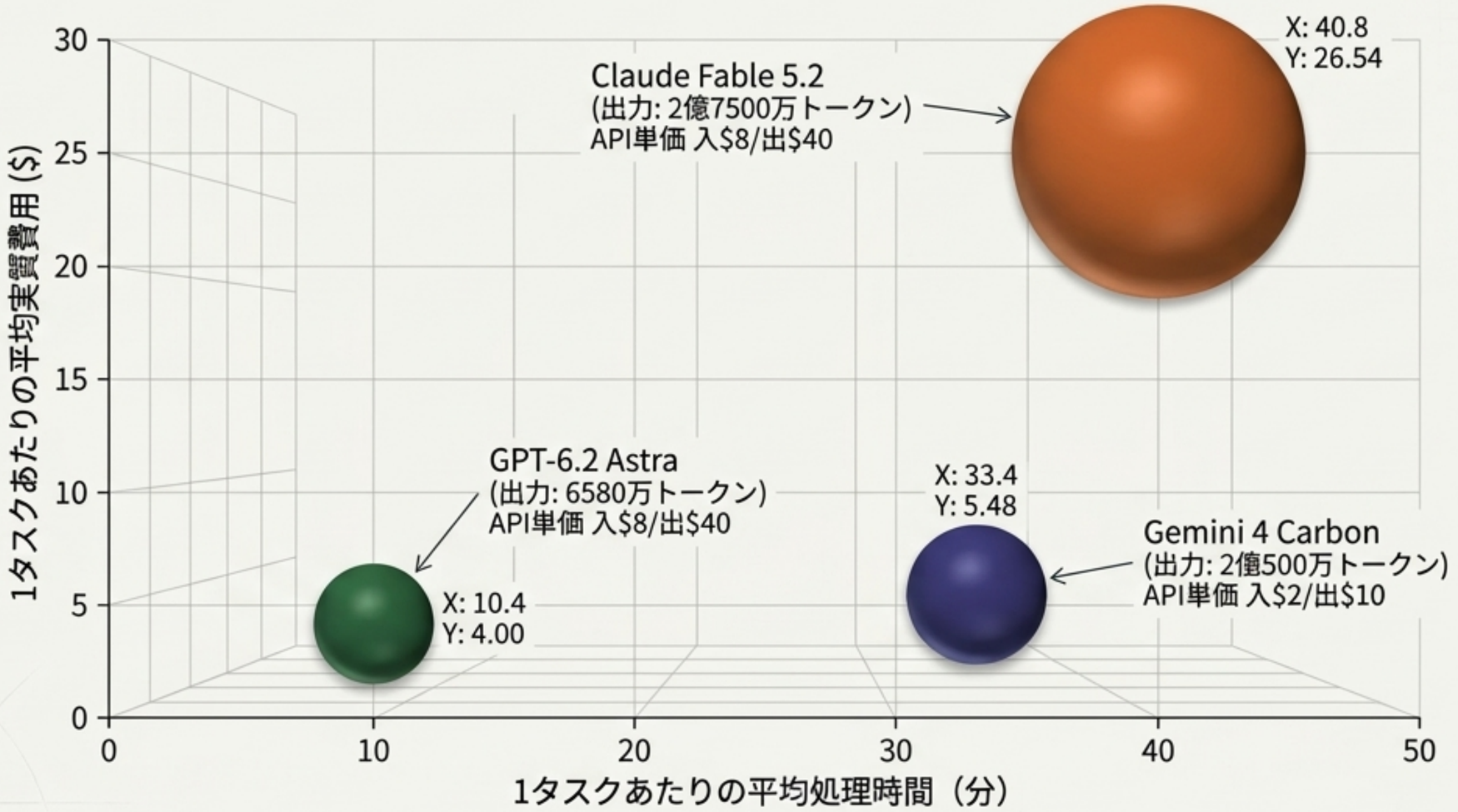
Fable: 最高峰の改修能力と時間・コストの犠牲

マルチモーダルとインフラ運用

Gemini: CLI制御と事実性の制御

Gemini Notebook

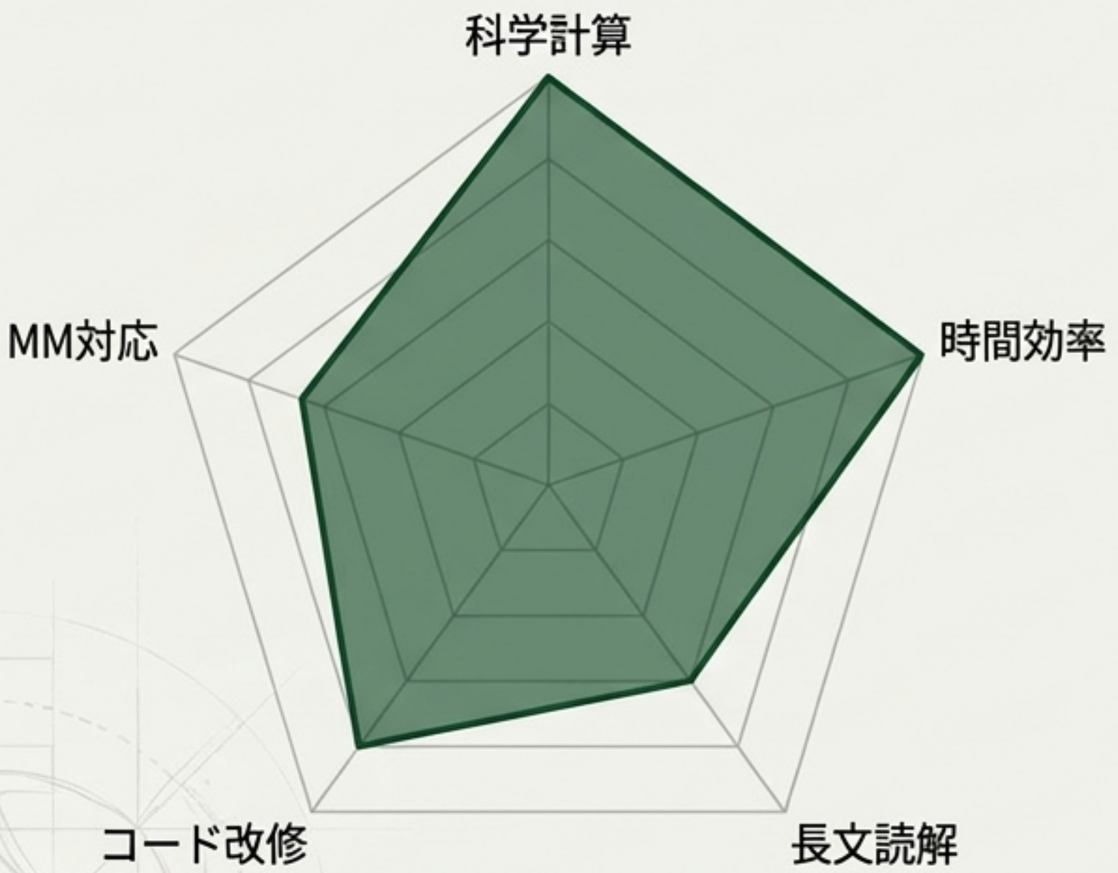
効率のパラドックス：API単価の一致と「トークン冗長性」によるコスト格差



インサイト: AstraとFableのAPIトークン単価は同一。しかし、Fableの極端な自己検証ループにより、実務におけるタスク単価には**6.6倍（$26.54 vs $4.00）**の劇的な格差が生まれる。

Gemini Notebook

GPT-6.2 Astra：圧倒的効率と自律科学計算の覇者



処理速度: **10.4分** / タスク (最速)

コスト: **$4.00** / タスク (最安)

強み

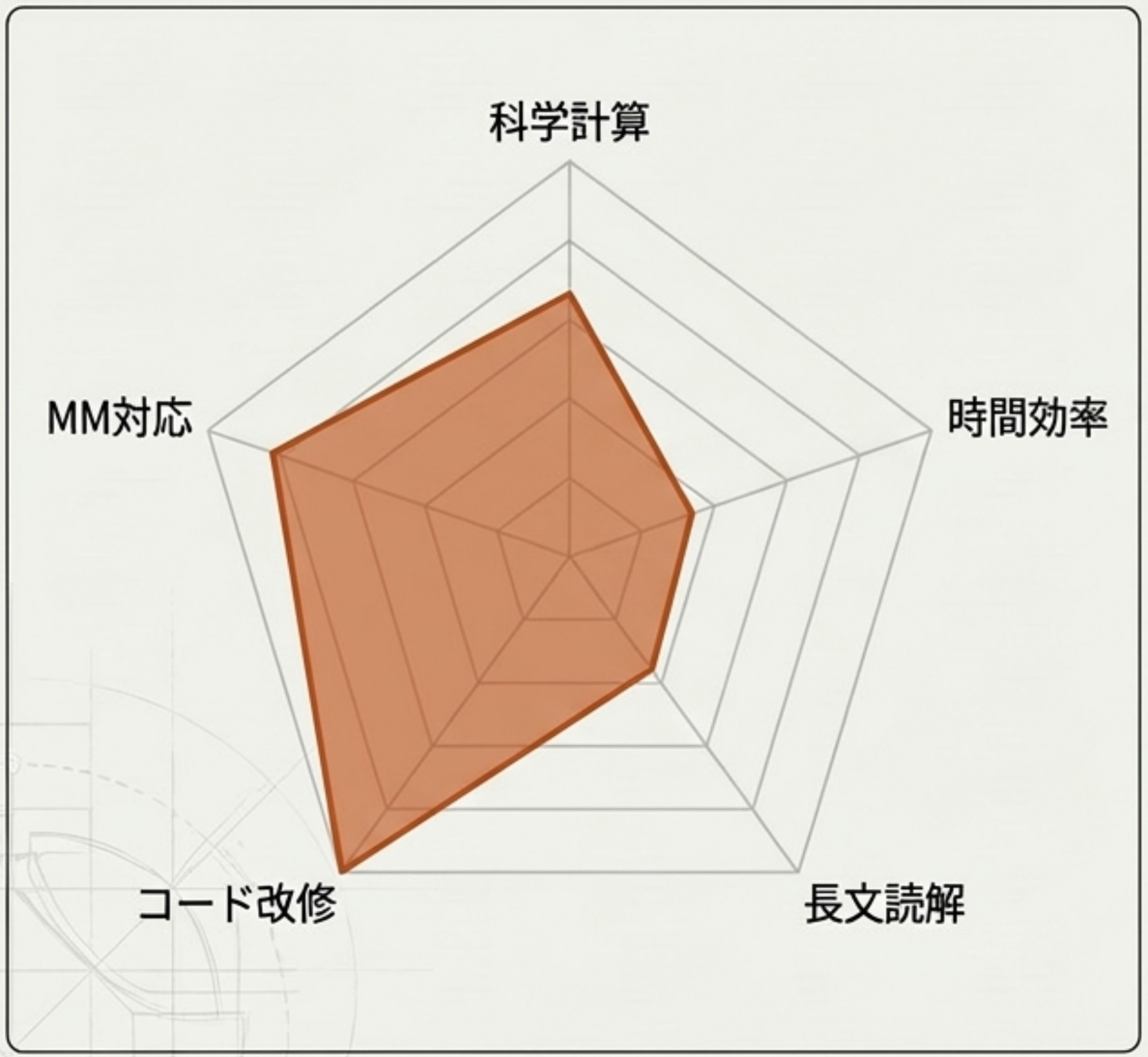
- Terminal-Bench-Science 0.1 首位
- 数理科学・実験パイプラインの自律完遂力
- パレート最適のROI

制約

- ネイティブ音声・動画入力に非対応
- 超高難度の純粋理論推論(CritPt v2)のピーク値で一歩譲る

Gemini Notebook

Claude Fable 5.2：限界突破の論理推論とリポジトリ改修の深思者



処理速度: 40.8分 / タスク (TTFT: 約221秒)

コスト: $26.54 / タスク

強み

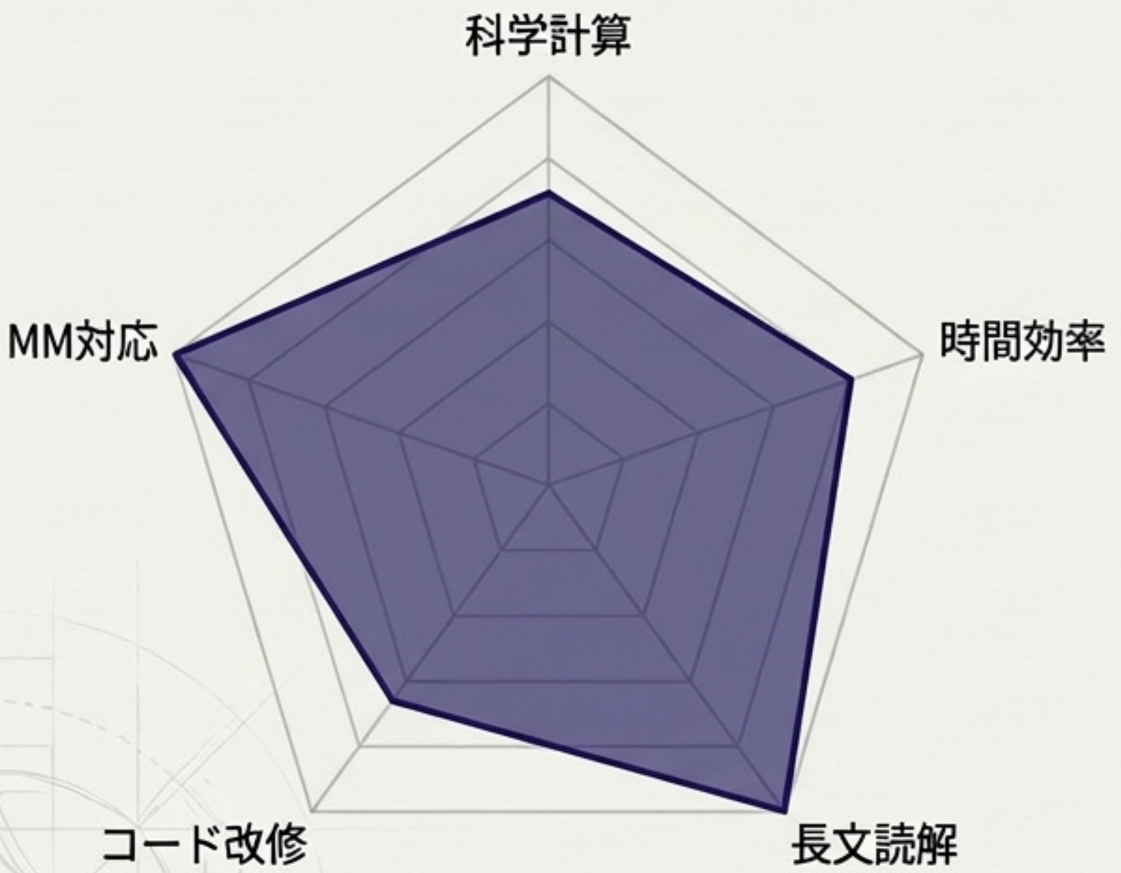
- FrontierCode-AA 2.0 (48.8%) 圧倒的首位
- 複数ファイルに跨る厳格なコード改修
- AA-Briefcase v1.1 (66.1) など知識労働のピーク性能

制約

- 日常業務には不向きな遅延と高コスト
- 複雑なレイアウトの長文専門PDF（GDP.pdf: 24.2%）での文脈抽出

Gemini Notebook

Gemini 4 Carbon：マルチモーダル基盤とインフラ自動化の操縦士



処理速度: **33.4分** / タスク

コスト: **$5.48** / タスク

強み

- Terminal-Bench 4.0 (65.2%) 総合首位を凌駕するCLI能力
- フルマルチモーダル（音声・動画）と長文PDFへの適性

制約

- 大規模コードベース改修（40.6%）での遅れ
- 推論設定High時のハルシネーション悪化リスク

Gemini Notebook

領域横断分析：「コーディング能力」は単一の指標ではない

レイヤー 1: ソフトウェア工学層（高抽象度）

大規模OSSリポジトリ改修、複数ファイル依存関係解決 (FrontierCode-AA 2.0)

✔ 勝者: Claude Fable 5.2 (48.8%) - 圧倒的優位

レイヤー 2: ミドルウェア・OS運用層（中抽象度）

CLI制御、インフラ構築、障害切り分けコマンド実行 (Terminal-Bench 4.0)

✔ 勝者: Gemini 4 Carbon (65.2%) - 逆転

レイヤー 3: 計算・実験パイプライン層（特定ドメイン）

科学計算スクリプトの自律実行、データ解析 (Terminal-Bench-Science 0.1)

✔ 勝者: GPT-6.2 Astra - 首位

インサイト: 開発対象のレイヤーにより最適モデルは完全に分化している。単一の『コーディングスコア』への依存は致命的な選定ミスを招く。

Gemini Notebook

長文ドキュメント処理における盲点と「事実性の制御」

Claude Fable 5.2の死角：複雑なPDFレイアウト



総合首位のFableだが、図表や複雑なレイアウトを含む長文専門PDF（GDP.pdf）では24.2%の抽出精度に留まる。大部な技術仕様書のそのままの投入にはリスクが伴う。

Gemini 4 Carbonの罠：過剰推論によるハルシネーション



事実性指標における逆転現象。ハルシネーション最小化を狙う調査業務では、推論設定「High」ではなく、あえて「Medium」に留めるチューニングが必須。過剰な推論が根拠のない推測を誘発する。

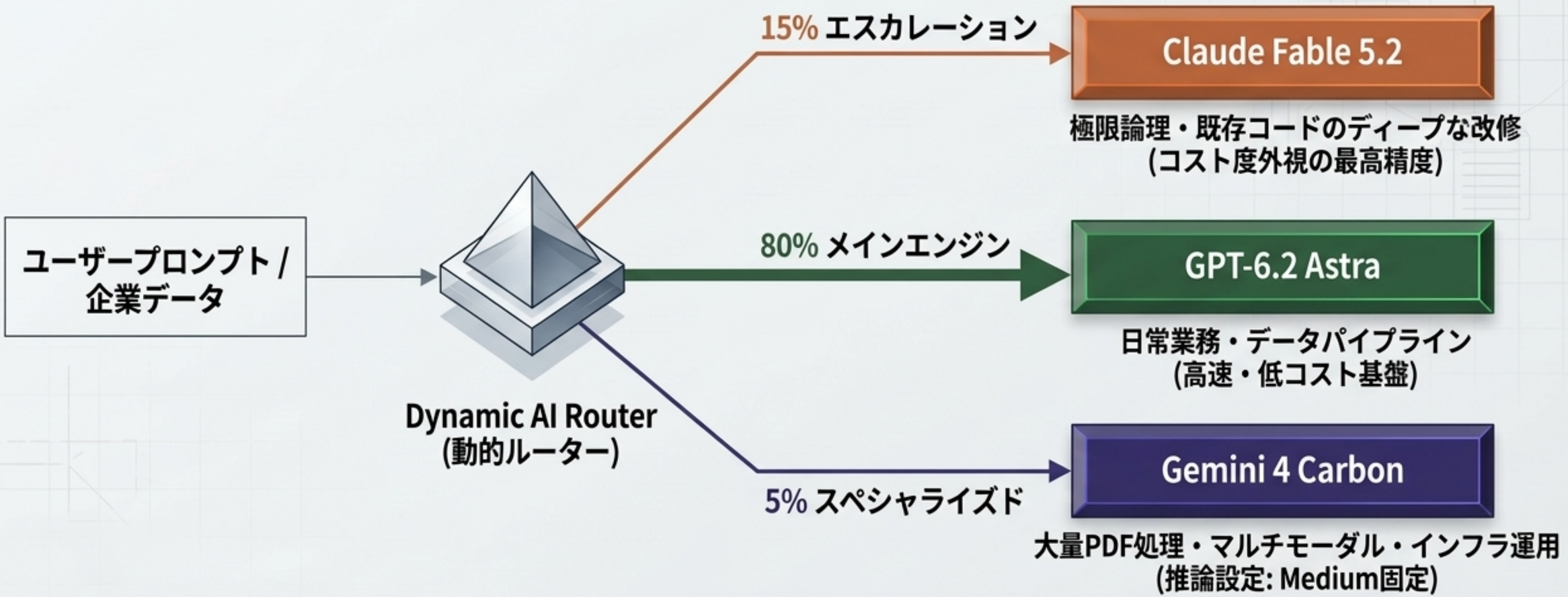
Gemini Notebook

エンタープライズ向け ユースケース別最適ルーティング・マトリックス

想定ユースケース	推奨モデル	選定根拠	運用上の留意点
日常的なAIエージェント基盤 科学研究・データ解析	GPT-6.2 Astra	Terminal-Bench-Science首位 1タスク10.4分・$4.00の圧倒的な速度と経済性	動画・音声の直接入力は不可。超高難度改修はFableを併用
大規模コードベース改修 難問アーキテクチャ設計・論理検証	Claude Fable 5.2	FrontierCode 2.0 (48.8%)、AA-Briefcaseにおける最高精度	平均40.8分・$26.54の高コスト。ここ一番の難所工程に限定投入する
大量の専門PDF精読 マルチモーダル解析 厳格なインフラ運用	Gemini 4 Carbon	GDP.pdf上位、フルマルチモーダル対応、Terminal-Bench 4.0首位	事実性重視の調査業務では、推論設定を「Medium」に固定すること

Gemini Notebook

最終形態：「ハイブリッドAIアーキテクチャ」の青写真



インサイト：単一モデルへの依存はコスト爆発か品質低下を招く。タスクの性質に応じたインテリジェント・ルーティングこそが、次世代エンタープライズAIの最適解である。

Gemini Notebook

エグゼクティブ・サマリー：戦略的アクション

01. 単一モデル依存からの脱却

「総合スコア」という幻想を捨て、タスクレイヤー（科学計算、大規模改修、インフラ運用）ごとに最適化されたマルチモデル基盤へ直ちに移行する。

02. 『隠れたコスト構造』の可視化と制御

API単価の比較ではなく、「出力トークンの冗長性」を含めた『実質タスク単価』と『遅延』をシステム設計のKPIとして組み込む。

03. インテリジェント・ルーティングの構築

Astraを日常の標準エンジンとし、難所にFable、非構造化データ・インフラ運用にGeminiを割り当てる動的ルーターをアーキテクチャの中核に実装する。

出典: Artificial Analysis Intelligence Index v5.0.1 Reports (Nov 2026)

Gemini Notebook