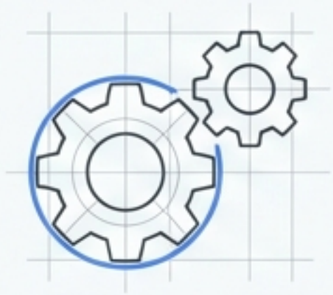
Gemini 3.5 Flashの衝撃：「対話」から「行動」へのシフト

[Strategic Executive Briefing: The Agentic Blueprint]

Executive Summary:
3つのパラダイムシフト



The Core Intent

「対話」から「行動」へ

チャットボットからエージェント実行への根本的転換。未公開の3.5 Proではなく、3.5 Flashこそが今回の実質的な主役。

The Unfair Advantage

圧倒的スピードと価格

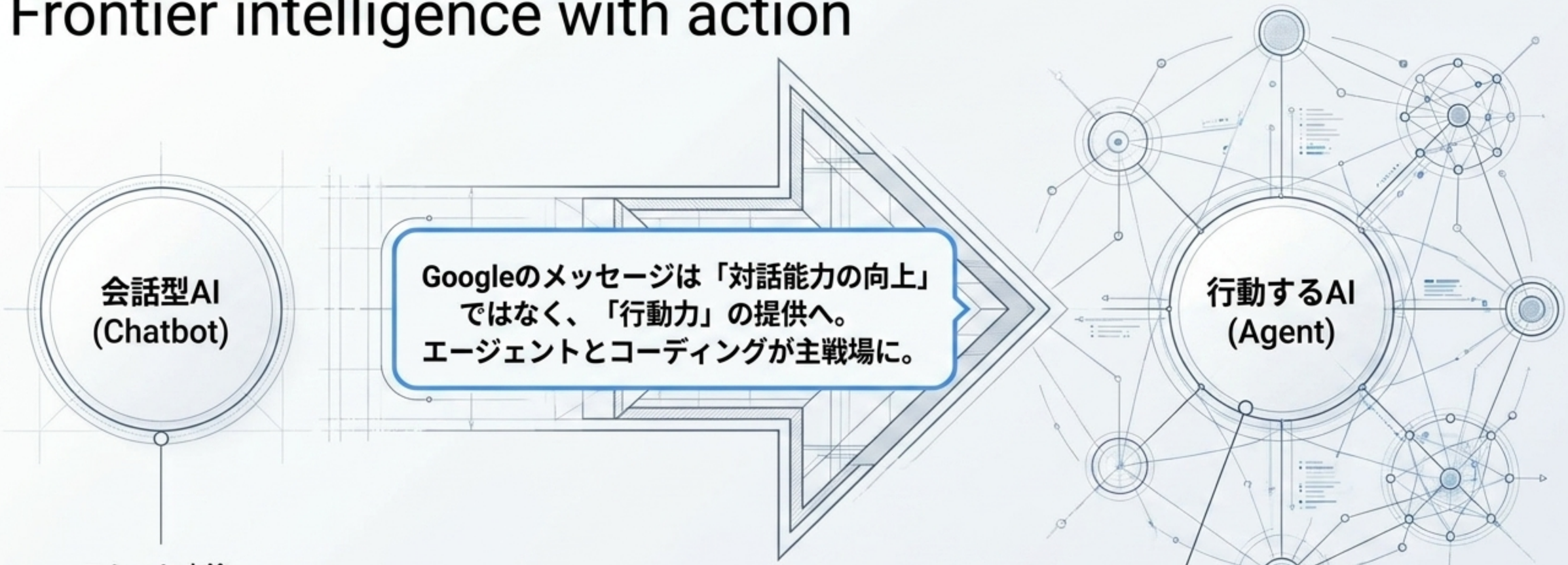
近Pro品質を維持しつつ、他の最先端モデルの出力速度の4倍、コストは半額未満というゲームチェンジ。

The Ubiquity

即日・横断的配備

発表当日にアプリ、検索、API、Enterpriseの全方位エコシステムへ同時にデプロイされた異例のスピード展開。

The Strategic Pivot:
Frontier intelligence with action



- テキスト応答
- 単発のタスク
- 人間の介在が必須

- マルチステップ業務自動化
- エージェント実行
- API連携による自律的行動

The Sweet Spot:
パレート境界の突破

「近Pro品質」を維持しつつ、比較対象の最先端モデルを圧倒する出力速度と低コストを実現。API標準料金は入力$1.50 / 出力$9.00。

The Engine:
100万トークン・マルチモーダル

1,000,000
トークン入力

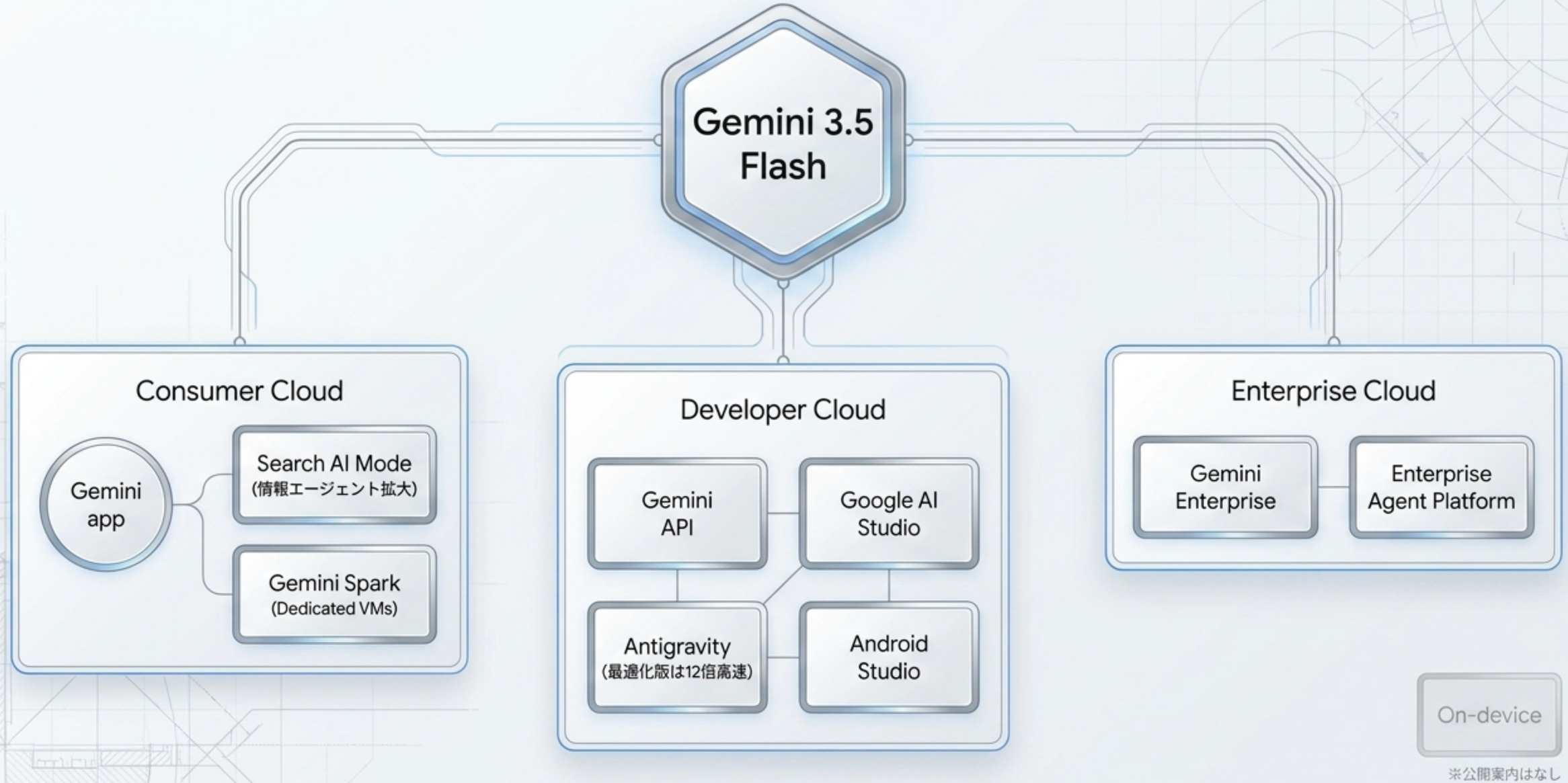
- Text
- Code
- Image
- Audio
- Video
- PDF

65,535
トークン出力

- Text

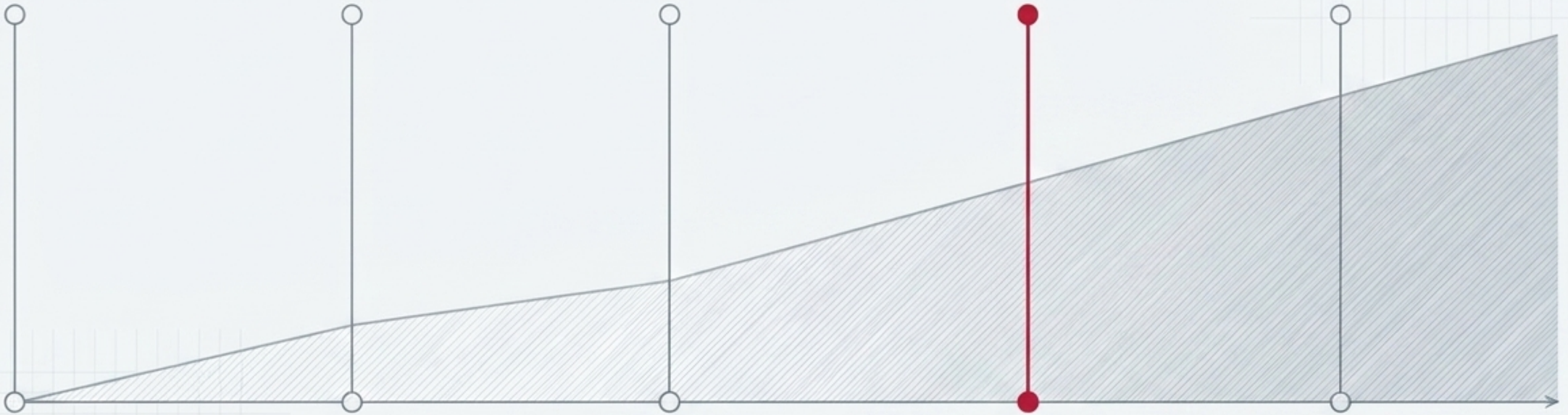
独立計測（Google AI Studio提供）で約140 tok/s、初回応答18.55秒。Function calling, structured output, code executionに対応。

Ecosystem Blueprint: 全方位への即日配備



※公開案内はなし

Rollout Timeline: 異例のスピード展開



5/19

3.5 Flash公開。app/Search/API/Enterpriseへ即日展開。

5/19週

SparkをTrusted Testersへ展開。

5/26週

Spark βを米Google AI Ultraへ提供。

6月予定

Gemini 3.5 Pro公開（詳細未公表）。

夏予定

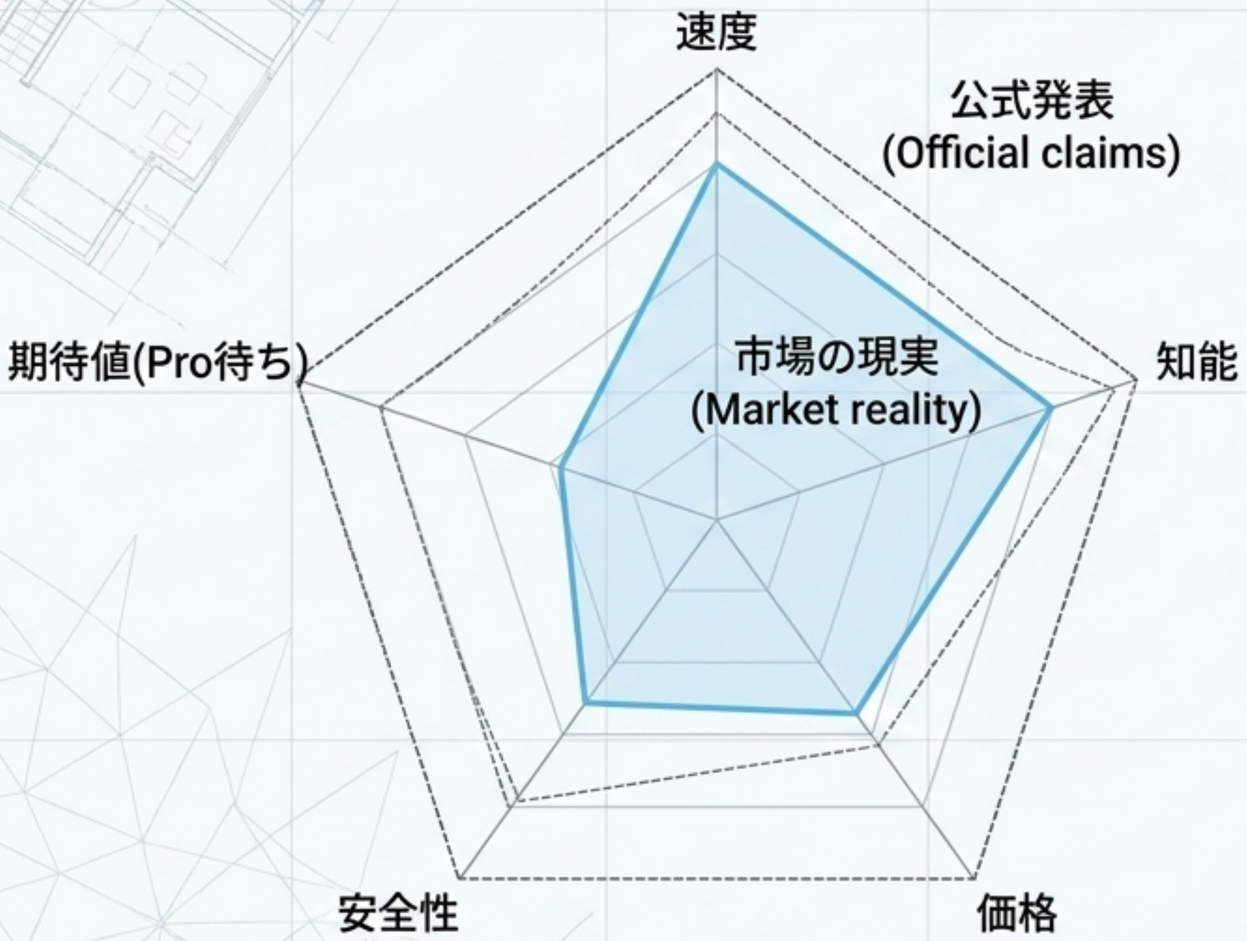
Searchの情報エージェントとGenerative UI拡大。

The Frontier Model Matrix: 競合比較と主戦場

Model	主戦場 (Main Battlefield)
Gemini 3.5 Flash	**エージェント・コーディング。**1M文脈・4倍速・$1.5/$9。
Gemini 3.5 Pro (未公表)	複雑問題・高度推論向け。
GPT-4o	最高クラスの総合知能と長文推論。約1M・高価($5/$30)。
Claude 3.5 Sonnet	企業ワークフロー・高効率エージェント。
Llama 3 Maverick	オープンソース・自己運用可能。

Insight: 3.5 Flashは万能首位ではなく「エージェント/コード/速度の交点」で最適化された特化型モデル。

Market Sentiment: 熱狂と現実のギャップ



Positive: 体感速度の圧倒的高さ、GA直行の実運用志向、コーディング/業務自動化での巻き返し布石。

Neutral: 速度×知能の境界では首位級だが、旧Flash比での価格上昇。

Negative: 本命「3.5 Pro」未公開への落胆、自動安全評価の微減、長期エージェント性能への懐疑。

Risks & Realities: 実運用における4つの壁

1. 安全性の揺らぎ

モデルカードにて、前世代比でtext safety、multilingual safetyの微悪化を記録。

2. 高度推論のムラ

長文・抽象推論（MRCR v k, ARC-AGI）では依然としてGPT-4oやClaude Opus等に及ばない領域が存在。

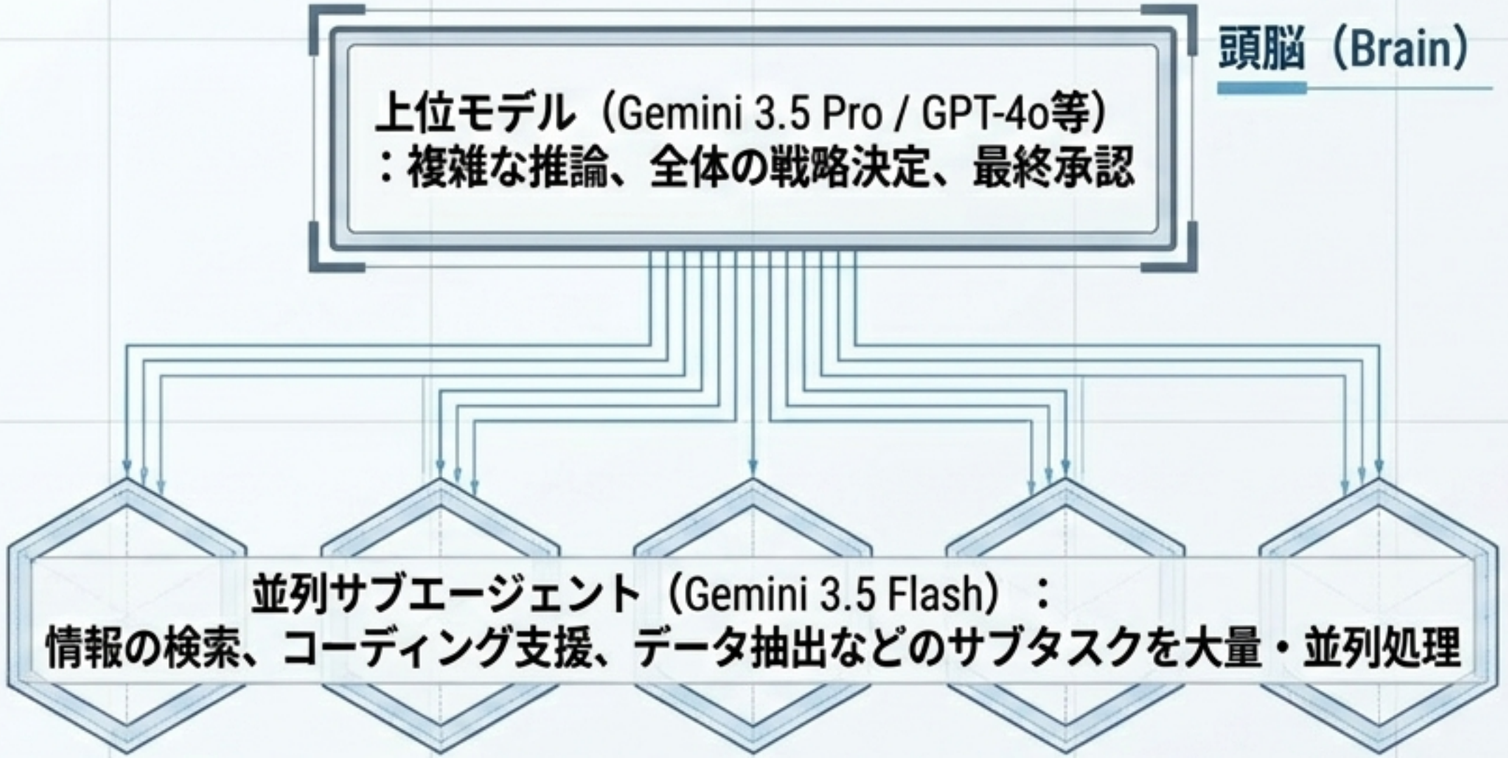
3. データ所在のガバナンス

リージョン内ML処理がグローバルエンドポイントでは保証されないリスク（専用VMのSpark等に依存）。

4. 利用ポリシーの分断

AI Studioの無料枠（製品改善に利用される）と有料枠（利用されない）でのデータガバナンスの違い。

The Architecture of Choice: 混成運用ベストプラクティス



全面置き換えではなく、適材適所の「エコシステムへの組み込み」が鍵。

Strategic Recommendations: 今すぐ取るべきアクション

開発者向け (Developers)

[利点]
高速・1M文脈

[難点]
Live API非対応

[Action]
サブエージェント・並列処理へ即時投入。最難タスクは上位モデルとAB比較。

企業向け (Enterprise)

[利点]
エコシステム一体運用

[難点]
データ所在とポリシー差

[Action]
Paid/Enterprise前提でのPoC。法務・情シスと連携したデータ分類と地域制約の先行設計。

研究者向け (Researchers)

[利点]
エージェント実行の進化

[難点]
アーキテクチャ非公開による再現性の限界

[Action]
公式ベンチに依存せず、日本語の業務タスクや安全性評価における独自の独立検証を実施。