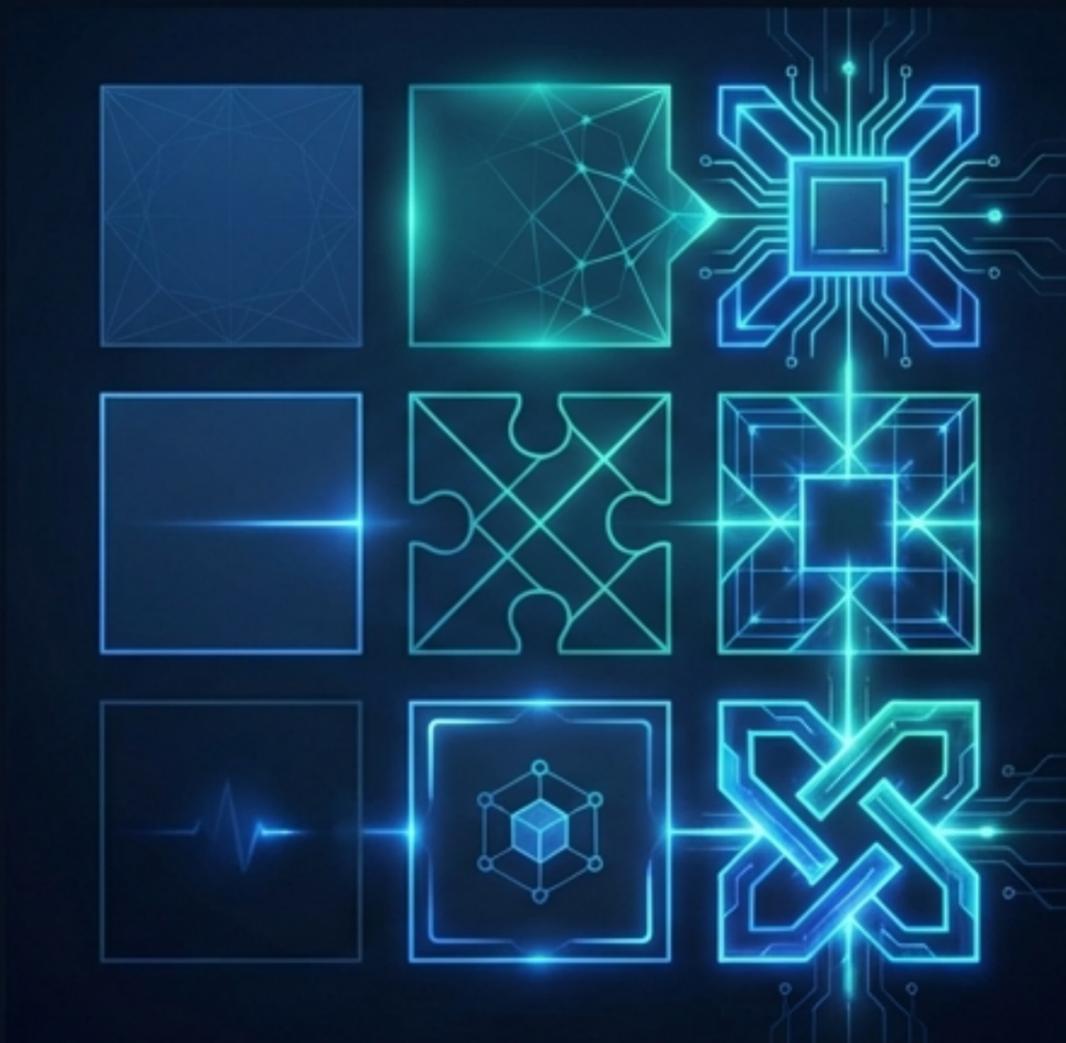


Gemini 3.1 Pro 深掘分析レポート

抽象推論の特異点と実務へのインパクト



Release: 2026年2月20日 | Source: Technical Documentation & Independent Benchmarks

エグゼクティブサマリー：“Ver.1” がもたらす非連続的な進化

Key Metrics

Release: 2026年2月20日

77.1%  +148%

26.3% 

市場位置付け & コスト

- **定義:** 単純な答えでは不十分なタスク特化型モデル
- **コスト優位性:** Claude Opus 4.6比で入力**-40%** / 出力**-48%**



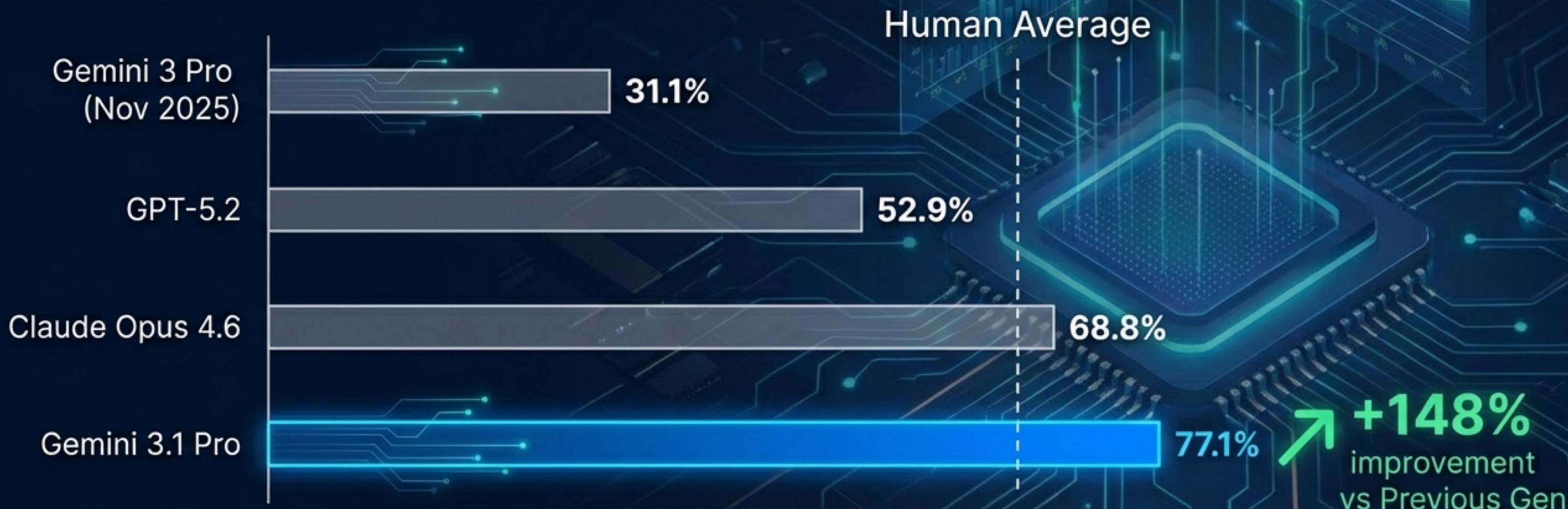
推奨 / 非推奨

 **推奨:** 高度な抽象推論、科学的知識 (GPQA)、Vibe Coding 

 **非推奨:** 大規模コードベースの一括ダンプ、100万トークン全量検索 

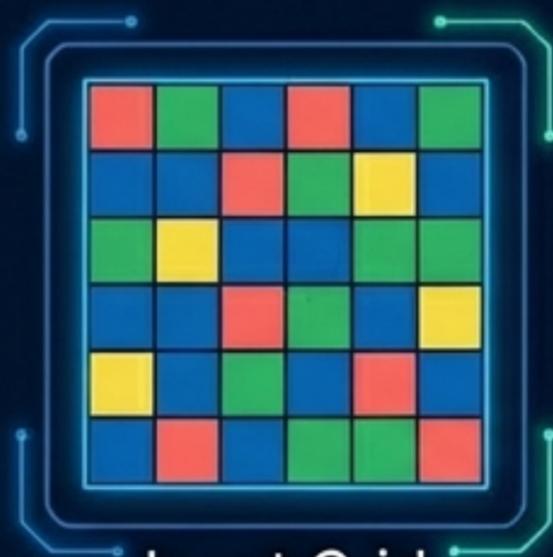
ARC-AGI-2：汎用人工知能（AGI）への登竜門

「暗記」を排除し、未知のタスクへの適応力を測定するベンチマークにおける特異点



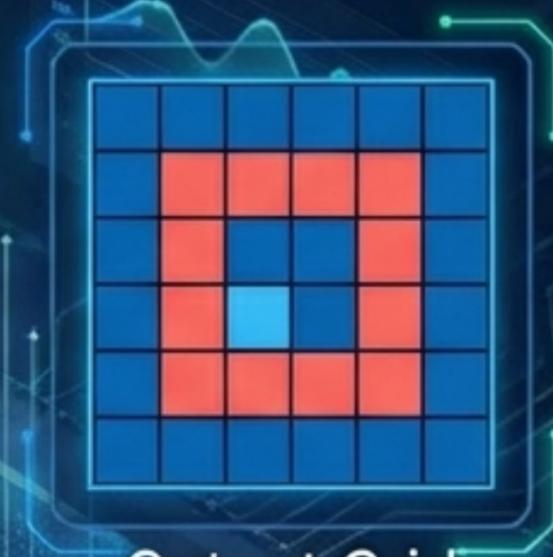
François Chollet (Keras創設者) は、これを「AGIへの進歩を厳格に測定する次世代ツール」と定義。

「パターンマッチング」から「推論」へ



Input Grid

Transformation Rule (Unknown)



Output Grid



1. 記号解釈

表面的な模様ではなく、記号が持つ意味（「囲み」「移動」など）を理解する能力。



2. 合成的推論

複数の異なるルール（色を変える+回転させる等）を同時に組み合わせ合わせて適用する能力。



3. 文脈依存的規則適用

グリッド内の状況に応じて、動的にルールの適用方法を変化させる能力。

モデル仕様と技術アーキテクチャ

Model ID

gemini-3.1-pro-preview

Context Window



Input: ~1,000,000 tokens (1,048,576 exact)

Output: ~64,000 tokens (65,536 exact)

Knowledge Cutoff

2025年1月



Modality



Input: Text, Image, Video, Audio, PDF

Output: Text only



Platform Availability



Vertex AI, Google AI Studio, Gemini API, Gemini App (Pro/Ultra). *Free tier excluded.

主要ベンチマーク比較：領域別の覇権争い

Benchmark 	Gemini 3.1 Pro	Claude Opus 4.6	GPT-5.2 
Abstract Reasoning (ARC-AGI-2)	77.1%	68.8%	52.9%
Science (GPQA Diamond) 	94.3%	91.3%	92.4%
Coding (SWE-Bench Verified) 	80.6%	80.8%	-
Terminal/Ops (Terminal-Bench)	68.5%	65.4%	77.3% GPT-5.3 Codex
Tool Use (HLE with tools) 	51.4%	53.1%	-

要約: 比較可能な13ベンチマーク中10項目でGeminiがClaude Opus 4.6を上回る。

“Vibe Coding” とクリエイティブ実装力

```
1 import python
2 import ss as sp
3
4 async def generate_starling_nurmuration(particles: int) -> VisualOutput:
5     """output""" = 0
6     def particles = VisualBist:
7         particles: = VisualEod1
8         for int in tocpatnpatzctes(particles: int):
9             particles.setints()
10            return particles.particles].stant()
11        }
12    particles[setViorfarct]
13 }
14 }
15 }
16 }
17 }
18 }
19 const renderSwirlingParticles = (data) => {
20     const renderSwirling(l), () => {
21         data.local.data => {
22             const.msangle.loæfn('oipi * oon xvb particles ), );
23         });
24     });
25 };
```

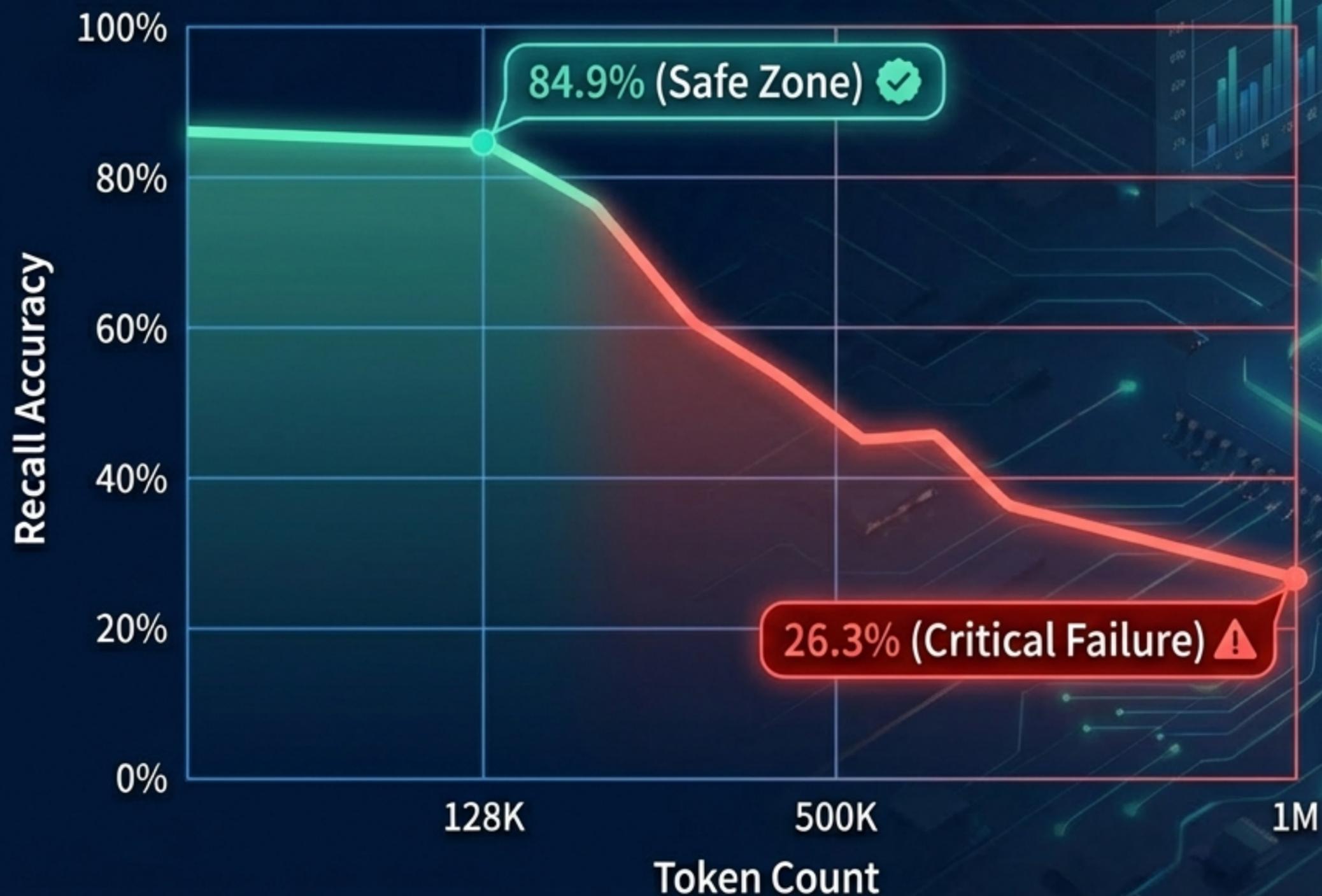
Logic & Instruction Following



Visual Execution

-  **Output Completion (出力完遂能力):** 55,000トークン超の完全な出力を生成可能。途切れずに複雑なアーティファクトを一発で実装。
-  **Understanding Intent:** 構文だけでなく、ユーザーの「意図 (Vibe)」を汲み取りUIデザインへ反映。

【警告】長文コンテキストにおける「針」の消失

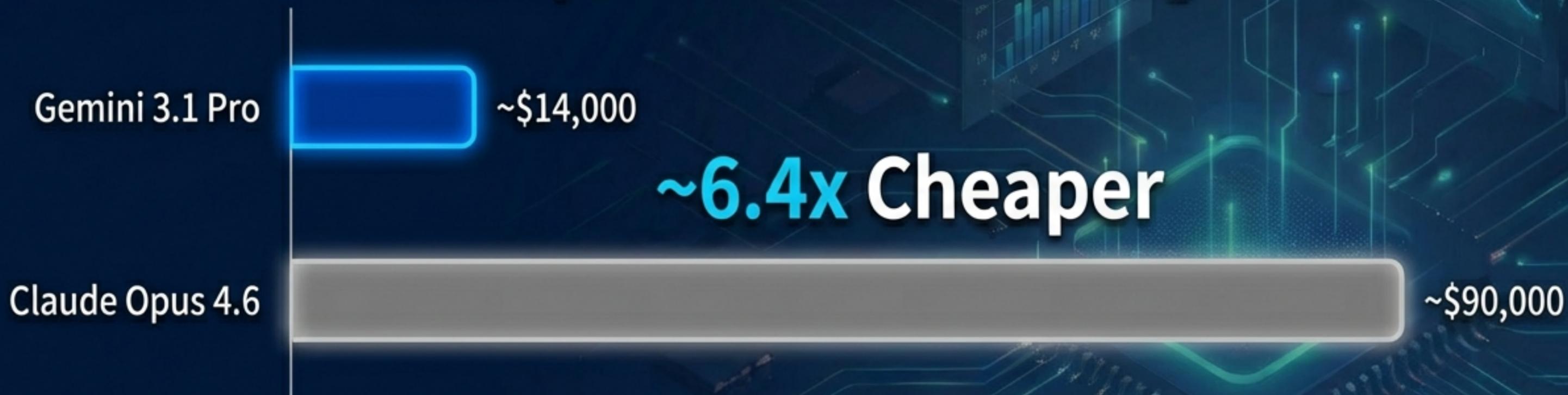


Operational Warning

仕様上の1M対応を過信してはならない。特許分析や大規模コードレビューでは、必ずRAG (Retrieval Augmented Generation) を併用すること。

コスト効率と経済性分析

Monthly Cost for 1B Token Processing



Price Breakdown (per 1M tokens)

- Input: Gemini (\$2.00) vs Opus (\$5.00)
- Output: Gemini (\$12.00) vs Opus (\$25.00)

*Note: Geminiの出力は冗長(verbose)になる傾向があり、実質コスト差は若干縮小する可能性がある。

実テストからの知見：UXと安定性



Visual Generation

SVGアニメーションや和モダン茶室UIの生成において、Claude Opusに匹敵または凌駕する品質。



Thinking Depth

哲学的な問いに対し、独自性と深みのある回答を提示。



Complex Logic Failures

3Dゲーム生成など、複雑なロジック構築でエラーが発生し動作しないケースを確認。



Platform Variability

Google AI Studio (API) と Web App版で出力品質に著しい差がある。

フロンティアモデルの勢力図 in Inter



「単一モデルの総合優位」は終了。タスクに応じたマルチモデル戦略が不可欠。

戦略的推奨事項と移行計画



For Developers: 論理・推論タスクにGeminiを採用。ただしコンテキストは **128K** 以内に抑える設計にすること。



For IP Strategy: ‘Vibe Coding’（自然言語アプリ生成）の特許出願増に備える。RAG特許の重要性は継続。



Migration Alert: Gemini 2.0 Flash / Lite は **2026年6月1日** にシャットダウン。Gemini 3 Flash Previewへの移行を計画せよ。



Safety: DeepMind Frontier Safety Frameworkに基づき、CBRNおよびサイバーリスク評価済み。

結論：記憶装置ではなく、「思考エンジン」として



Gemini 3.1 Proは、膨大なデータを記憶する「金庫」としては不完全だが、未知の解を導き出す「思考エンジン」としては史上最高峰にある。

過去のデータ検索ではなく、未来の解創出に活用せよ。