

GPT-5 Pro vs Gemini 2.5 Pro Deep Think 徹底比較調査 Claude Opus 4.1

GPT-5 Pro（2025年8月リリース）（Getpassionfruit +3）とGemini 2.5 Pro Deep Think（2025年5月リリース）（Helicone +2）は、両者ともAI推論能力の新たなマイルストーンを示している。**

（THE DECODER +2）GPT-5 Proはソフトウェア開発と実用的エンジニアリングタスクで優位性を示し、（OpenAI）（OpenAI）一方でGemini 2.5 Pro Deep Thinkは数学的推論と長文コンテキスト処理で卓越した性能を発揮している。**（9to5Google）（Google）両モデルとも\$1.25/\$10.00という同等の価格設定（OpenAI）ながら、それぞれ異なる強みを持つことが判明した。（VentureBeat +5）

この分析は、OpenAI（GPT-5 Pro）とGoogle DeepMind（Gemini 2.5 Pro Deep Think）の公式技術文書、業界標準ベンチマーク結果、技術系メディアのレビュー、開発者コミュニティの反応を総合的に検証したものである。両モデルとも前世代から大幅な進歩を遂げているが、選択は具体的な用途要件によって決まる状況となっている。

基本仕様の詳細比較

アーキテクチャと設計思想

GPT-5 Proは統一ルーター系アーキテクチャ（OpenAI）を採用し、高速処理用の（gpt-5-main）と深い推論用の（gpt-5-thinking）を組み合わせている。リアルタイム・インテリジェント・ルーターが動的にモデル選択を行い、（OpenAI）推論レベル（最小・低・中・高）を調整可能（Simon Willison）だ。（OpenAI +2）パラメータ数は非公開ながら、o1/o3シリーズの推論能力が内蔵されている。（Botpress）

Gemini 2.5 Pro Deep ThinkはSparse Mixture-of-Experts (MoE) Transformerを基盤とし、Gemini 2.0アーキテクチャの大幅な改良版である。パラレル思考技術を採用し、複数の仮説を同時検討する能力を持つ。強化学習により拡張推論パスが促進され、（Google）（AllThings）最大32,000トークンの思考バジェット設定が可能（Google Cloud）だ。（Google DeepMind）（Google）

コンテキストウィンドウとトークン制限

GPT-5 Pro：

- 入力トークン：272,000トークン（API）（OpenAI）（Simon Willison）
- 出力トークン：128,000トークン（（THE DECODER）不可視推論トークン含む）（OpenAI）（Simon Willison）
- **総コンテキスト：400,000トークン**（VentureBeat +2）
- ChatGPT：256,000トークン（DataCamp +2）

Gemini 2.5 Pro Deep Think：

- 入力コンテキスト：**1,000,000トークン**（（Google DeepMind）（Google Cloud）2,000,000トークンまで拡張予定）（Helicone +3）
- 出力トークン：64,000トークン
- **ネイティブマルチモーダル設計**（Google +2）

マルチモーダル能力の評価

GPT-5 Proのマルチモーダル対応は現在テキストと画像入力のみで、出力はテキストのみだ。しかし将来的にSORAとの統合により動画処理機能が追加予定（Botpress）である。

Gemini 2.5 Pro Deep Thinkはネイティブマルチモーダルとして設計され、テキスト・音声・画像・動画・完全なコードリポジトリの入力が可能（Google Cloud）だ。**（Leanware +2）24言語でのネイティブ音声出力**、（Google DeepMind）（Google DeepMind）感情検出、複数話者対応（VentureBeat）

[Google Cloud](#)を備えている。[Google DeepMind](#)[Google DeepMind](#) Video-MMEベンチマークで84.8%のスコアを記録し、動画理解で優位性を示している。[Helicone](#)[deepmind](#)

ベンチマーク性能の包括的比較

数学・科学推論における性能

ベンチマーク	GPT-5 Pro	Gemini 2.5 Pro Deep Think	優位性
AIME 2025 vellum OpenAI	94.6% OpenAI (ツール使用), 100% (Python使用) OpenAI +2	83.0%	GPT-5 Pro
GPQA Diamond TechCrunch OpenAI	89.4% TechCrunch (思考モード) TechCrunch +3	83.0%	GPT-5 Pro
Humanity's Last Exam TechCrunch Vellum	42.0% Vellum (ツール使用) TechCrunch +3	17.8% Medium +2	GPT-5 Pro

GPT-5 Proは**競技レベルの数学問題で圧倒的優位性**を示している。AIME 2025でのPython使用時100%達成は特筆すべき成果だ。[Vellum](#)[OpenAI](#)

プログラミング・コーディング性能

ベンチマーク	GPT-5 Pro	Gemini 2.5 Pro Deep Think	優位性
SWE-bench Verified TechCrunch +2	74.9% TechCrunch +4	63.2% DataCamp +2	GPT-5 Pro
Aider Polyglot OpenAI Vellum	88% OpenAI +2	76.5%	GPT-5 Pro
LiveCodeBench V6 Google 9to5Google	74.1%	87.6% Google +4	Gemini 2.5
WebDev Arena Google	強力な性能	ELOスコア1415 (Google DeepMind) ¹ Google +2	Gemini 2.5

GPT-5 Proは実世界のソフトウェアエンジニアリング ([OpenAI](#) SWE-bench) で優位性を持つ一方、**Gemini 2.5 Pro Deep Thinkは競技プログラミング** ([9to5Google](#)) で優秀な成績を示している。[OpenAI](#)

マルチモーダル理解

ベンチマーク	GPT-5 Pro	Gemini 2.5 Pro Deep Think	優位性
MMMU OpenAI	84.2% OpenAI DataCamp	79.6% (通常) / 84.0% (Google Deep Think) Google Google	GPT-5 Pro
Video-MME	83.3%	84.8% Helicone deepmind	Gemini 2.5
Long Context (MRCR 1M) DataCamp	利用不可	82.9% DataCamp Medium	Gemini 2.5

推論能力の詳細分析

GPT-5 Proの推論機能

GPT-5 Proは**o1/o3シリーズの推論能力を統合**し、調整可能な推論努力レベル（最小～高）を提供する。**Safe Completions** [OpenAI](#) アプローチにより、出力中心の安全性訓練を実装している。

[OpenAI](#) [Platformer](#) 幻覚率は4.8%まで削減され ([VentureBeat](#) 前世代の22%から大幅改善)、

[TechCrunch](#) [OpenAI](#) 偽装行動も2.1%に減少している。[TechCrunch +2](#)

Gemini 2.5 Pro Deep Thinkの「Deep Think」機能

パラレル思考技術により複数仮説を同時検討し、強化学習で拡張推論パスを促進 (Google) している。(Google)思考バジェット (最大32,000トークン) により推論の深度を制御可能 (Google Cloud) だ。**(Google) (AllThings)透明な思考要約**(Google Cloud) (Google)により、推論過程を可視化できる点が特徴的である。(Google DeepMind +4)

プログラミング能力の言語別比較

Python開発

GPT-5 ProはAIME 2025でPython使用時100%を達成し、(VentureBeat) (Vellum)Django、Flask、FastAPI等のフレームワークで卓越した性能を示している。**(OpenAI) (Apidog)Gemini 2.5 Pro Deep Think**はデータサイエンスとGoogle Cloudエコシステム統合で強みを持つ。

JavaScript/TypeScript

GPT-5 ProはReact、Vue、Angularで優秀で、(Apidog)単一プロンプトから完全なWebアプリケーション作成が可能だ。**(Tom's Guide)Gemini 2.5 Pro Deep Think**はマルチモーダル機能を活かしたフロントエンド開発で差別化を図っている。

システムプログラミング (C++、 Rust)

両モデルともシステムプログラミングで競争力を持つが、GPT-5 Proは最適化提案で、Gemini 2.5 Pro Deep Thinkはアルゴリズム的洗練度で優位性を示している。

コストとパフォーマンスの経済分析

API料金体系 (100万トークンあたり)

モデル	入力料金	出力料金	コンテキスト長
GPT-5	\$1.25	\$10.00	272k
GPT-5 Mini	\$0.25 (VentureBeat) (OpenAI)	\$2.00 (VentureBeat) (OpenAI)	272k
GPT-5 Nano	\$0.05 (VentureBeat) (OpenAI)	\$0.40	272k
Gemini 2.5 Pro (≤200k)	\$1.25	\$10.00	1M
Gemini 2.5 Pro (>200k)	\$2.50	\$15.00 (Leanware) (google)	1M
Gemini 2.5 Flash	\$0.30 (Vals AI +2)	\$2.50	1M

用途別コスト効率分析

高頻度シンプルタスク：Gemini 2.5 Flash-Lite (\$0.10/\$0.40) が最適 (google) 複雑推論・数学：両モデル同価格だが、Gemini 2.5 Pro Deep Thinkが長コンテキスト優位 (Google DeepMind) ソフトウェア開発：GPT-5 Proが74.9% SWE-bench性能で優位 (DataCamp +3) エンタープライズ長文解析：Gemini 2.5 Pro Deep Thinkの100万トークンコンテキストが決定的優位 (Helicone +2)

パフォーマンス指標

GPT-5 Pro：

- 推論トークンが出力トークンとしてカウント (実効コスト増)
- プロンプトキャッシングで最大90%割引 (VentureBeat) (Simon Willison)
- o3より22%少ない出力トークン、45%少ないツール呼び出し (OpenAI) (DataCamp)

Gemini 2.5 Pro Deep Think：

- 出力速度：144.3トークン/秒 (Artificial Analysis)

- 初回トークンまでの時間：37.43秒 (Artificial Analysis)
- バッチモード：50%割引 (Google AI) (google)
- コンテキストキャッシング：\$0.31-\$0.625/100万トークン (google)

専門家評価とマーケット反応

技術系メディアの評価

TechCrunchはGPT-5 Proを「統一AI モデル」として位置づけ、(OpenAI) (TechCrunch) SWE-bench Verifiedでの74.9%スコア ((OpenAI) Claude Opus 4.1の74.5%、Gemini 2.5 Proの59.6%を上回る) (TechCrunch) と大幅な幻覚率削減 (4.8%) (OpenAI) を評価した。 (TechCrunch +3)

MIT Technology ReviewはGPT-5を「技術的ブレークスルーではなく、洗練された製品」として位置づけ、AppleのRetinaディスプレイに例えた。 (MIT Technology Review)

開発者コミュニティの反応

GPT-5 Proの評価は二極化している。**Latent Space**は「世界最高のコーディングモデル」と絶賛し、ソフトウェアエンジニアリング自動化を65%から72%に向上させたと評価した。 (Substack) 一方、Reddit上では「GPT-5 is horrible」スレッドが4,600以上のアップvoteを獲得し、(Yahoo Finance) 人格の喪失と創作能力の低下が批判された。 (Tom's Guide) (TechRadar)

Gemini 2.5 Pro Deep ThinkはHacker Newsで慎重ながら好意的な反応を得ている。技術能力への評価は高いが、限定アクセスと価格（噂される\$250/月） (TechCrunch) への懸念が存在する。

エンタープライズ採用動向

OpenAIは積極的な価格戦略でマーケットシェア獲得を目指し、** (TechCrunch) Google**は「研究グレードの知性」と透明な思考プロセス (YouTube) でエンタープライズ市場を狙っている。 (Medium) 両社ともMicrosoft Azure AI、(OpenAI) (Microsoft Azure) Google Cloud TPUという強力なインフラを背景としている。

総合評価と推奨用途

GPT-5 Proの強みと推奨用途

強み：

- 実世界ソフトウェア開発で業界最高水準 ((OpenAI) SWE-bench 74.9%) (DataCamp +3)
- 競技レベル数学問題での圧倒的性能 ((OpenAI) AIME 100%) (Vellum) (OpenAI)
- 統合ルーター系による最適モデル自動選択 (OpenAI +2)
- 大幅改善された安全性 ((OpenAI) 幻覚率4.8%) (VentureBeat +4)
- 積極的価格戦略 (TechCrunch +2)

推奨用途：

- 複雑なソフトウェア開発プロジェクト
- 数学・科学計算が中心のアプリケーション
- 高い事実精度が要求されるシステム
- OpenAIエコシステム統合済み環境

Gemini 2.5 Pro Deep Thinkの強みと推奨用途

強み：

- 100万トークンの大規模コンテキスト処理 ((Google DeepMind) (Google Cloud) GPT-5 Proの2.5倍) (Helicone +3)

- **ネイティブマルチモーダル対応** (TechTarget) 音声・動画・画像 (Leanware +5)
- **透明な推論プロセス** (Google Cloud) (Google) デバッグ・監査に有効 (Google +2)
- **パラレル思考による複雑問題解決** (Google +2)
- **Google クラウドエコシステム統合** (Google +2)

推奨用途：

- 大規模文書解析・処理
- マルチモーダルアプリケーション
- 科学研究・学術応用
- エンタープライズ級データ分析
- Google Workspace統合環境

最終的な選択指針

ソフトウェア開発中心：GPT-5 Pro (SWE-bench性能、ツール統合) (OpenAI) (Vellum) **研究・学術用途：**Gemini 2.5 Pro Deep Think (長コンテキスト、マルチモーダル) (Google +2) **予算重視：**Gemini 2.5 Flash-Lite (\$0.10/\$0.40料金) (google) **エンタープライズ：**既存インフラとの親和性で判断 (OpenAI vs Google Cloud)

両モデルとも前世代から大幅な能力向上を実現しており、** (Google) AI推論技術の成熟期に入ったことを示している。選択は「どちらが優れているか」ではなく、「特定用途にどちらが適しているか」**という観点で決めるべき段階に達している。今後数ヶ月でさらなる改良版リリースが予想され、競争はますます激化すると予測される。