

徹底比較レポート：GPT-5 Pro vs Gemini 2.5 Pro Deep Think



Genspark

Aug 10, 2025

徹底比較レポート：GPT-5 Pro（2025 年 8 月） vs Gemini 2.5 Pro Deep Think

（2025 年 5 月）

- 対象モデルへのナビゲーション: [OpenAI GPT-5](#)¹ | [Google DeepMind Gemini 2.5 Pro](#)² | [Gemini 2.5 Deep Think](#)³

要約（まず結論）

- 仕様と設計思想: GPT-5 は「統一システム」で内蔵推論を標準化した最新フラッグシップ。Gemini 2.5 Pro は「思考(Thinking)」を前提に設計されたマルチモーダル重視モデルで、Deep Think によりテスト時並列思考と RL で難問に強いモードを持つ。OpenAI¹ DeepMind⁴ Google Blog⁵
- コンテキスト長: GPT-5 は公式モデルカードで 40 万トークン入力・最大出力 12.8 万を明記（他メディアに 256K 報道もあり、開発者向けドキュメント優先が無難）。Gemini 2.5 Pro は Vertex AI で入力約 104 万・出力 6.5 万既定、2M 文脈は機能的に「近く/条件付き」で展開（低解像度モードなど）する文脈。OpenAI Models⁶ OpenAI gpt-5 製品頁⁷ Vertex AI Doc⁸ Google Developers Blog⁹ Google Blog（3 月）⁵
- ベンチマーク総観: GPT-5 は AIME 2025（数理）や GPQA（PhD 理工）・SWE-bench（実務コーディング）で強く、SWE-bench Verified 74.9%・GPQA 89.4%など複数 SOTA 級。Gemini 2.5 Pro は SWE-bench Verified 63.8%を公式公表、Deep Think で推論能力が顕著に底上げされるが、実務コーディングに関しては総じて GPT-5 が一歩先行の構図が多い（用途や設定に依存）。OpenAI 開発者向け発表¹⁰ OpenAI 発表¹ Vellum まとめ¹¹ Google Blog（3 月）⁵
- マルチモーダル: Gemini 2.5 Pro はネイティブに音声・画像・動画・PDF などを入

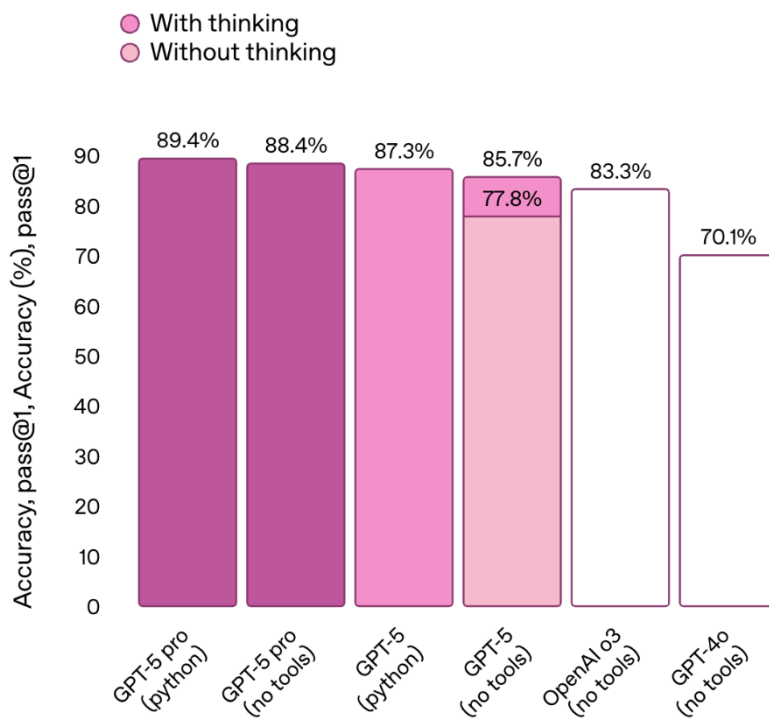
力でハンドリングし、動画理解機能（6 時間相当処理、時系列推論・モーメント検索など）が豊富。GPT-5 は画像・動画を含む理解・推論の強化を公式に謳うが、API 上のモダリティは「テキスト・画像入力／テキスト出力」が基本（音声や TTS は別エンドポイント）。動画理解の製品事例や SDK 整備は Gemini が先行感。DeepMind 製品頁 [2](#) Google Developers Blog（動画理解） [9](#) OpenAI 発表 [1](#) OpenAI Models [6](#)

- 料金・運用: API 単価は両者とも極めて競争的。OpenAI GPT-5 が入出力 \$1.25/\$10（100 万トークン）でキャッシュ入力 9 割引、Google の Gemini 2.5 Pro も同等水準の価格帯（200K 超入力は段階上がり）。Vertex/Google AI Studio 経由の運用・動画機能連携は Gemini の強み、Assistants/Responses/Function Calling などエージェント構築の総合 SDK は OpenAI が成熟。OpenAI Pricing [12](#) Gemini API Pricing [13](#) Vertex AI Generative AI Pricing [14](#)
- チューニング: GPT-5 は OpenAI 公式で Fine-tuning サポートあり。Gemini は Vertex AI 経由で SFT/LoRA 系の「Tuning」対応、Gemini API 直では Fine-tune 非対応（1.5 Flash-001 廃止以降）。OpenAI Models [6](#) Vertex AI Doc（2.5 Pro） [8](#) Vertex Tuning [15](#) Gemini API Model Tuning [16](#)
- レイテンシ・評判: 公式の明示数値は少ないが、実務者の報告では大規模入力時の Gemini 2.5 Pro で待ち時間が大きい局面がある一方、OpenAI も重推論モードでは初トークン遅延の指摘あり。メディアは GPT-5 の高性能・低価格を評価、Gemini は動画理解・Thinking 設計の優位を指摘。Google Cloud Community（遅延報告） [17](#) developers forum 遅延 [18](#) TechCrunch GPT-5 [19](#) IEEE Spectrum [20](#) The Verge Gemini 2.5 [21](#)

ビジュアルハイライト

- GPT-5 の GPQA/SWE-bench 可視化（出典 Vellum）

GPQA Diamond PhD-level science questions



Vellu

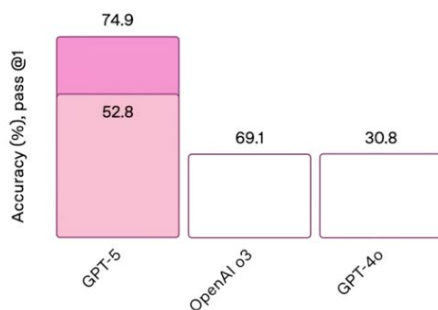
m.ai11

Academic

SWE-bench Verified

Software engineering

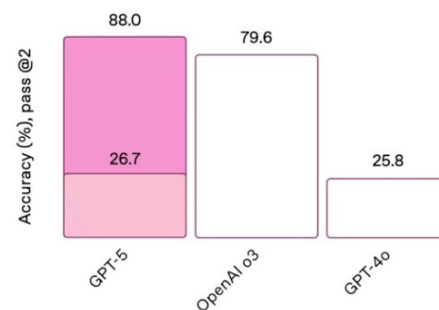
● Without thinking ● With thinking



Aider Polyglot

Multi-language code editing

● Without thinking ● With thinking



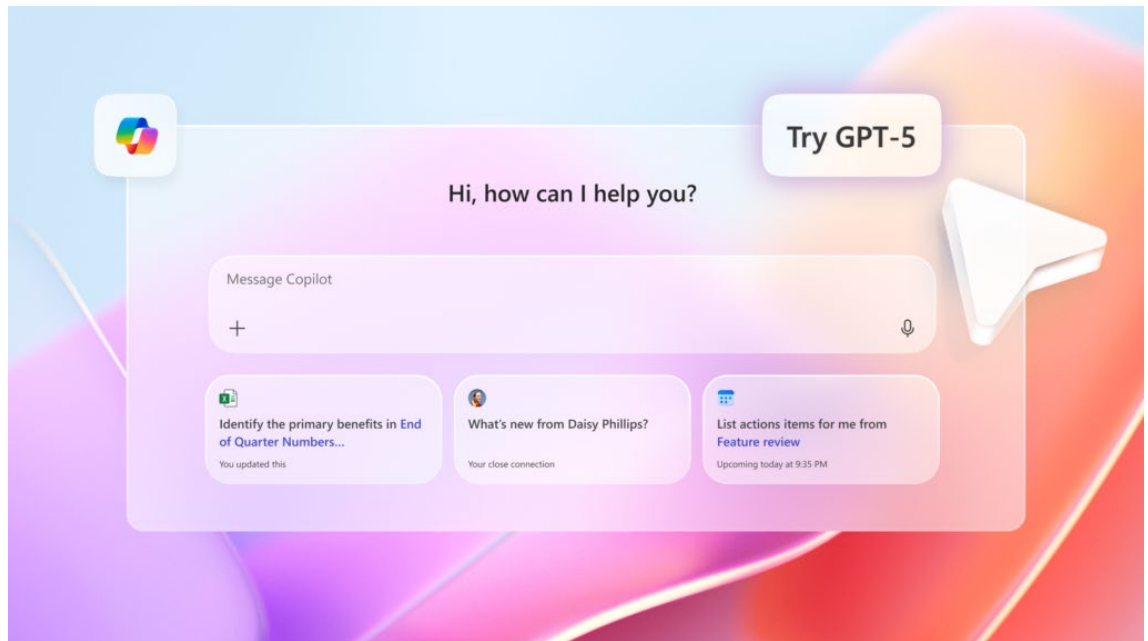
Vellum.ai11

- Gemini 2.5 の動画理解 (Dev Blog)

	Previous SoTA	Humans	Gemini 2.5 Flash Preview 04/17*	Gemini 2.5 Pro Preview 05/06*
EVALUATIONS WITH VISUAL INPUTS				
EgoTempo (test set) 0-shot open-ended VideoQA	40.3 (GPT 4.1*)	63.2	36.5	43.7
LVBench (test set) 0-shot 4-choice VideoQA	60.1 (GPT 4.1*)	94.4	60.9	68.2
Perception Test (test set) 0-shot 5-choice VideoQA	71.4 (Oryx)	91.4	71.2	77.3
QVHighlights (val set) 4-shot Video Moment Retrieval	76.1 (Mr BLIP)	–	70.2	72.6
VideoMMMU (test set) 0-shot 5-choice VideoQA	76.7 (Kimi-k1.6)	74.4	71.9	81.3
1H-VideoQA (test set) 0-shot 5-choice VideoQA	72.2 (Gemini 1.5 Pro)	–	64.3	76.2
EVALUATIONS WITH AUDIO-VISUAL INPUTS				
VideoMME (test set, long subset) 0-shot 4-choice VideoQA	72.0 (GPT 4.1)	–	77.8	82.0
YouCook2 Cap (val set) 4-shot Video Clip Captioning	198.8 (VAST)	–	185.3	198.0
YouCook2 DenseCap (val set) 4-shot Dense Video Captioning	67.2 (Vid2Seq)	–	67.6	69.3
EVALUATIONS WITH VISUAL-SUBTITLES INPUTS				
Minerva (test set) 0-shot 5-choice VideoQA	54.0 (GPT 4.1*)	92.5	61.9	63.5
Neptune (test set) 0-shot 5-choice VideoQA	85.1 (GPT 4.1*)	–	84.5	85.4
EVALUATIONS WITH AUDIO-VISUAL-SUBTITLES INPUTS				
VideoMME (test set) 0-shot 4-choice VideoQA	81.3 (Gemini 1.5 Pro)	–	79.3	85.2
Evaluation of Gemini 2.5 vs. prior models on video understanding benchmarks. Performance is measured by string-match accuracy for multiple-choice VideoQA, LLM-based accuracy for EgoTempo, R@0.5 for QVHighlights and CIDEr for YouCook2. *Videos were processed at 1fps and linearly subsampled to a maximum of 256 frames, except for 1H-VideoQA (7200 frames).				

Google Developers Blog

- Microsoft の GPT-5 採用発表イメージ



Microsoft News[22](#)

1. 公式仕様（アーキテクチャ、パラメータ数、学習データ、コンテキスト長ほか）
 - 公開範囲: 両社とも厳密なアーキテクチャ（層設計・パラメータ数・学習コーパス詳細）は未開示。OpenAI は「統一システム」「内蔵思考」を前面に、Google は「Thinking モデル」「Deep Think（並列思考+RL）」を前提に設計思想を明示。
OpenAI[1](#) DeepMind[4](#) Google Blog（Deep Think）[3](#)
 - モダリティ/入出力:
 - GPT-5（API のモデルカード）：入力はテキスト・画像、出力はテキスト。音声入出力は別 API（音声合成/転写）で提供。機能としてストリーミング/関数呼び出し/構造化出力/推論トークン/各種ツールが利用可。OpenAI Models[6](#)
 - Gemini 2.5 Pro: 入力は音声・画像・動画・テキスト・PDF、出力はテキスト。別バリエーションに TTS プレビュー（入力テキスト→出力音声）。
ai.google.dev モデル一覧 [23](#) DeepMind 製品頁 [2](#)
 - コンテキスト長:
 - GPT-5: 公式製品ページで入力 40 万トークン・最大出力 12.8 万。メディアには 256K 報道もあるため、SDK/モデルカードの数値を実務上の規範とするのが安全。OpenAI gpt-5 製品頁 [7](#) OpenAI Models[6](#) Ars Technica[24](#)
 - Gemini 2.5 Pro: Vertex AI で入力 1,048,576・出力 65,535 既定。Google は 2M コンテキスト（条件付き/設定依存で長尺動画処理など）を示唆。Vertex AI Doc[8](#) Google Developers Blog（動画理解）[9](#) Google Blog（3月）[5](#)
 - セーフティ/システムカード:

- GPT-5 System Card は統一システム構成と推論サブモデル、安全性の限界や評価観点を説明。Microsoft は Red Team 検証で「強固な安全プロファイル」を報告。OpenAI System Card PDF[25](#) Microsoft News[22](#)
 - Gemini 2.5/Deep Think のモデルカード/技術報告は推論強化の位置付け、能力・限界を概説。Gemini 2.5 Tech Report PDF[26](#) Deep Think Model Card PDF[27](#)
2. 主要ベンチマーク (MMLU、GSM8K、HumanEval 等)
- GPT-5 (公式・一次/二次情報)
 - SWE-bench Verified: 74.9% (Aider Polyglot 88%)。OpenAI 開発者向け発表 [10](#)
 - GPQA Diamond (PhD 理工) : 89.4% (Pro+ツールで)。Vellum まとめ [11](#)
 - AIME 2025 (数学) : 94.6% (ツール無し)。OpenAI 発表 [1](#)
 - MMLU: 集約サイトでは 0.871 (モデル粒度による)。一次発表に明記は乏しく、集約値は参考指標。Artificial Analysis (GPT-5) [28](#)
 - HumanEval: 公式値未確認。外部レビューは上振れ報告あるが、一貫した一次出所は限定的 (本レポートでは慎重に留保)。Vellum まとめ [11](#)
 - Gemini 2.5 Pro / Deep Think
 - SWE-bench Verified: 63.8% (Google 公式ブログ、カスタムエージェント設定)。Google Blog (3 月) [5](#)
 - MMLU: 集約サイトで 0.862 (2.5 Pro)。Artificial Analysis (Gemini 2.5 Pro) [29](#)
 - HumanEval/GSM8K: 一次資料で明示稀。第三者メディアで HumanEval pass@1≈82%、GSM8K で 2.5→Deep Think で顕著改善の記述あり (参考・非公式)。Medium (比較) [30](#) ApplyingAI (Deep Think 総括) [31](#)
 - 総括: 実務コーディング (SWE-bench Verified) と PhD 理工推論 (GPQA) で GPT-5 が優勢。マルチモーダルや動画系評価は Gemini 優位の報告が多く、指標ごとに強みが分かれる。OpenAI 発表 [1](#) Vellum まとめ [11](#) Google Developers Blog (動画理解) [9](#)
3. マルチモーダル能力 (テキスト/画像/音声/動画)
- GPT-5: 公式は「視覚/動画/空間/科学的推論を含むマルチモーダル強化」を強調。ただし API モデル定義上は「テキスト・画像入力／テキスト出力」。音声/TTS/画像生成は別エンドポイント・ツールと組み合わせで利用する設計 (Assistants/Responses で統合可能)。OpenAI 発表 [1](#) OpenAI Models[6](#)
 - Gemini 2.5 Pro: 音声・画像・動画・テキスト・PDF をネイティブ入力、TTS プレビューも提供。動画理解については、モーメント検索・時系列推論・長尺 (約 6 時間相当) 処理の実装例やパラメータ (低解像度モード) を公式に示し、開発ブログ

で豊富なユースケースを提示。DeepMind 製品頁 [2](#) Google Developers Blog (動画理解) [9](#) ai.google.dev モデル一覧 [23](#)

- まとめ: モダリティ統合の「使い勝手」と動画タスクの具体機能では Gemini が実装先行。GPT-5 はツール/API 群で機能統合が可能だが、モデル単体の I/O 仕様だけを見ると Gemini が広い。Google Developers Blog (動画理解) [9](#) OpenAI Models [6](#)
4. 推論 (論理・数学・常識) と「Deep Think」vs GPT-5 の対抗機能
- Deep Think の実像: 並列思考と新規 RL 手法で思考時間を拡張、難問に対する問題分解・仮説分岐・反証試行を強化する「拡張推論モード」。思考予算のキャリブレーション/制御/適応 (Calibrated/Controllable/Adaptive) という設計思想が示される。Google Blog (Deep Think) [3](#) DeepMind (Thinking/Deep Think) [4](#) Deep Think Model Card PDF [27](#)
 - GPT-5 の対抗: 「統一システム」に内蔵された Thinking で、タスク難易度に応じて深く考えるモードに移行。ChatGPT では推論ビュー (思考の進行表示) を提供。Microsoft は GPT-5 推論モデルの安全性検証を強調。OpenAI 発表 [1](#) Ars Technica [24](#) Microsoft News [22](#)
 - 実力差の含意: 純粋な数理・理工系難問 (AIME/GPQA) や実務コーディング (SWE-bench Verified) での一次/二次情報は GPT-5 優位の示唆が多い。他方、探索的・長時間文脈を要するマルチモーダル推論では Gemini の Deep Think/動画理解の実装優位が活きる局面が多い。OpenAI 発表 [1](#) Vellum まとめ [11](#) Google Developers Blog (動画理解) [9](#)
5. ソフトウェア開発支援 (コード生成・デバッグ・ドキュメント)
- GPT-5: SWE-bench Verified 74.9%、Aider Polyglot 88% で OpenAI 過去モデルを超える。エージェント/ツール連携 (コード実行・関数呼出・ファイル検索) と価格の両立で、広い現場に適用しやすい。OpenAI 開発者向け発表 [10](#) OpenAI Models [6](#)
 - Gemini 2.5 Pro/Deep Think: 公式 SWE-bench Verified 63.8% (設定依存)。第三者では Deep Think で実務系・競プロ系の改善報告もあるが、総合すると GPT-5 が一歩先行のシーンが多い。動画/音声/画像を含むエンジニアリング (例: 長尺動画からの要件抽出→設計→生成) には Gemini のネイティブ多モーダルが有利。Google Blog (3月) [5](#) Bind ブログ比較 (参考) [32](#) DeepMind 製品頁 [2](#)
6. 料金・レイテンシ・ファインチューニング
- 料金 (API, per 1M tokens)
 - OpenAI GPT-5: 入力\$1.25/出力\$10、キャッシュ入力\$0.125。ミニ/ナノはさらに低価格。OpenAI Pricing [12](#)
 - Gemini 2.5 Pro (Gemini API/Vertex AI) : 入力\$1.25 (200K 超は\$2.5)、出力\$10 (200K 超は\$15)。Batch 割引有。Gemini API Pricing [13](#) Vertex AI Generative AI Pricing [14](#)

- レイテンシ: 公式数値の開示は限定的。開発者コミュニティでは、Gemini 2.5 Pro の超大規模入力時に高レイテンシの報告が散見（平均数十秒〜p99 数分という投稿）。GPT-5 も重推論モードで初トークンが遅いという声あり。用途別にストリーミング/バッチや Thinking 予算調整が実務上重要。Google Cloud Community[17](#) developers forum[18](#) OpenAI 発表 [10](#)
 - ファインチューニング:
 - GPT-5: Fine-tuning 対応（API の Fine-tuning エンドポイント）。OpenAI Models[6](#)
 - Gemini 2.5 Pro: Vertex AI で Tuning (SFT/LoRA 系) が「サポート」。Gemini API 直は現行 FT 非対応（1.5 Flash-001 廃止）。Vertex AI Doc（2.5 Pro）[8](#) Vertex Tuning[15](#) Gemini API Model Tuning[16](#)
7. 専門家レビュー・評判（メディア、研究者、開発者コミュニティ）
- メディア総括:
 - TechCrunch は GPT-5 の公開で「わずかに並を出るが価格が決定打」とし、AIME/GPQA/SWE-bench の伸長を紹介。Google の Deep Think は「複数解を並列検討し性能向上」。TechCrunch GPT-5[19](#) TechCrunch Deep Think[33](#)
 - IEEE Spectrum は 2025 年前半、Gemini 2.5 Pro を「Thinking モデル」として台頭、価格優位を強調。IEEE Spectrum[20](#)
 - The Verge は Gemini 2.5 Pro のマルチモーダルと Thinking 強化、2M 文脈「coming soon」を報道。The Verge[21](#)
 - Ars Technica は GPT-5 のチャット階層（Free/Plus/Pro）や文脈長を解説。Ars Technica[24](#)
 - 開発者の声: 大規模文脈の Gemini で高レイテンシ報告が複数。GPT-5 の重推論は高精度だが応答待ちもあり、ユースケースごとの最適化（ストリーミング、思考予算、分割問い合わせ）が鍵。Google Cloud Community[17](#) developers forum[18](#)
 - 公式採用例: Microsoft は GPT-5 を広範な製品に統合、安全検証も強調。Microsoft News[22](#)

8. 総合比較表（性能・機能・コスト・ユースケース）

観点	GPT-5 Pro (OpenAI)	Gemini 2.5 Pro Deep Think (Google DeepMind)
設計思想	統一システム＋内蔵 Thinking、難問時は深い推論に移行 OpenAI 1	Thinking モデル＋Deep Think（並列思考と RL、思考予算の制御/適応） Google Blog 3 DeepMind 4
コンテキスト	入力 40 万/出力 12.8 万（モデルカード） OpenAI Models 6	入力約 104 万/出力 6.5 万（Vertex）、 2M 文脈の運用モードあり Vertex AI Doc 8 Dev Blog 9
ベンチマーク（抜粋）	SWE-bench Verified 74.9%、GPQA 89.4%、AIME 94.6% OpenAI Dev 10 Vellum 11 OpenAI 1	SWE-bench Verified 63.8%（設定依存）。第三者は Deep Think での改善も報告 Google Blog 5 Bind 比較（参考） 32
MMLU（集約指標）	≈0.871（参考）Artificial Analysis 28	≈0.862（参考）Artificial Analysis 29
モダリティ	入力：テキスト/画像、出力：テキスト（音声/TTS等は別 API で統合）OpenAI Models 6	入力：音声/画像/動画/テキスト/PDF、出力：テキスト。TTS プレビューも提供 ai.google.dev 23
動画理解	製品発表で強化言及（具体 API は別構成）OpenAI 1	長尺処理・モーメント検索・時系列推論・アプリ変換など豊富 Dev Blog 9
コーディング	実務系で強い（SWE-bench Verified 74.9%、Aider 88%）OpenAI Dev 10	マルチモーダル要件抽出との組合せが強み、実務スコアは GPT-5 に劣後ケースあり Google Blog 5
料金（API）	入力\$1.25／出力\$10、キャッシュ入力\$0.125 OpenAI Pricing 12	入力\$1.25 (>200K は\$2.5)／出力\$10 (>200K は\$15)、Batch 割引 Gemini API 13 Vertex AI 14
Fine-tuning	API で FT 対応 OpenAI Models 6	Vertex で Tuning 対応、Gemini API 直は非対応 Vertex Tuning 15 Gemini API 16
レイテンシ	重推論で初トークン遅延の指摘（非公式）OpenAI 発表 10	大規模入力で高遅延の報告が複数（コミュニティ）Cloud Community 17

最終評価とユースケース推奨

- 研究開発（数理・理工難問、エキスパート QA）：GPT-5 Pro 優位。AIME/GPQA

など高難度で強く、価格も有利。実務の品質・再現性・コストの総合バランスに秀でる。OpenAI 発表 [1](#) Vellum まとめ [11](#) OpenAI Pricing [12](#)

- 企業内エージェント/コーディング運用: SWE-bench Verified や Aider Polyglot のスコアから、既存リポジトリ修正・タスク駆動の自動化には GPT-5 が即戦力。Assistants/Responses/関数呼び出し等の SDK 成熟度も後押し。OpenAI 開発者向け発表 [10](#)
- マルチモーダル分析(特に動画・音声・長尺コンテンツ): Gemini 2.5 Pro Deep Think。動画からの知識抽出、時系列推論、長時間素材の要約→設計→生成の一貫パイプラインは Gemini が構築しやすい。Google Developers Blog (動画理解) [9](#) DeepMind 製品頁 [2](#)
- 超大規模コンテキストの統合参照: 設定次第で Gemini が 1M~2M 級の処理系を活用可能。超長文・複数媒体の照合では優位だが、遅延の実務影響は要注意。バッチ・低解像度・キャッシュなど運用チューニングを併用。Vertex AI Doc [8](#) Developers Blog [9](#)
- コスト・調達: いずれも低価格化が急進。純テキスト/コード中心なら GPT-5、マルチメディア実装・GCP 基盤統合なら Gemini が運用面で有利なことが多い。OpenAI Pricing [12](#) Vertex AI Pricing [14](#)

補足 (アクセスと提供形態)

- ChatGPT/Teams/Enterprise/EDU の層別提供 (Pro で推論ビュー/Pro バリエーション) と、API (gpt-5/mini/nano 等) での開発者アクセスが並行展開。Ars Technica [24](#) OpenAI Models [6](#)
- Gemini は Gemini アプリ/AI Studio/Gemini API/Vertex AI の多層で提供。Deep Think は段階的ロールアウト (AI Ultra/Advanced) から拡張。Google Blog (Deep Think) [3](#) Google Cloud Blog (GA) [34](#)

メディア・一次情報リンク (再掲)

- GPT-5: Introducing GPT-5 [1](#) / Developers 向け [10](#) / モデルカード/価格 [6](#)・価格 [12](#) / System Card [25](#)
- Gemini 2.5/Deep Think: 製品頁 [2](#) / 技術報告 PDF [26](#) / Deep Think 記事 [3](#)・モデルカード PDF [27](#) / AI Studio/モデル一覧/価格 [23](#)・価格 [13](#) / Vertex 仕様/価格 [8](#)・価格 [14](#)

Appendix: Supplementary Video Resources



【遂にキタ !! 】 超使える推論オープンモデル「GPT-OSS ...

2 days ago



GPT-5 Pro vs Gemini Deep Think — Head-to-Head Testing!

1 day ago



GPT 5 vs Gemini 2.5 Pro Comparison - Ultimate Coding Test

1 day ago

もっと詳しく

1

openai.com

2

deepmind.google

3

blog.google

4

deepmind.google

5

blog.google

6

platform.openai.com

7

openai.com

8

cloud.google.com

9

developers.googleblog.com

10

openai.com

11

www.vellum.ai

12

platform.openai.com

13

ai.google.dev

14

cloud.google.com

15

cloud.google.com

16

ai.google.dev

17

www.googlecloudcommunity.com

18

discuss.ai.google.dev

19

techcrunch.com

20

spectrum.ieee.org

21

www.theverge.com

22

news.microsoft.com

23

ai.google.dev

24

arstechnica.com

25

cdn.openai.com

26

storage.googleapis.com

27

storage.googleapis.com

28

artificialanalysis.ai

29

artificialanalysis.ai

30

medium.com

31

applyingai.com

32

blog.getbind.co

33

techcrunch.com

34

cloud.google.com