

GPT-5、GPT-5 Thinking、GPT-5 Pro



Genspark

Aug 08, 2025

エグゼクティブサマリー (要約)

- 存在と基本像: OpenAI は GPT-5 を「統合システム」と位置づけ、「高速・高スループット (gpt-5-main)」と「深い推論 (gpt-5-thinking)」をリアルタイムでルーティング。ChatGPT では、さらに並列テスト時計算で精度を上げた「GPT-5 Thinking Pro」(UI 上の名称「GPT-5 Pro」) も提供されます。OpenAI¹、OpenAI²
- 技術の核: モデル選択は自動(難易度やユーザー行動・正答率シグナルに基づく)または有料プランの手動で制御可能。API には応答長の `verbosity`、推論深さの `reasoning_effort` (minimal~high) など開発者向けノブが実装。OpenAI Help Center³、OpenAI⁴
- ベンチマーク: 公開システムカードは多言語 MMLU(0-shot) の表を掲載する一方、HumanEval や GSM8K は直接数値未掲載。SWE-bench Verified は図示(数値は図表で、本文に明示値なし)。Thinking/Pro は長手の推論・並列計算で難題に強み。cdn.openai.com (PDF) ⁵、OpenAI²
- 用途と対象: GPT-5 は日常~専門までの総合的用途。Thinking/Pro は高度推論・曖昧課題・多段ツール実行・安全性重視の現場向け(法務・研究・医療・エンジニアリング等)。OpenAI¹、VentureBeat⁶
- アクセスと料金: ChatGPT は Free/Plus/Team/Pro で提供(Enterprise/Edu は後追い)。Free は上限後に mini へフォールバック。Plus/Team は手動で Thinking 選択可、週 200 件の上限(自動切替は別カウント)。Pro/Team は Thinking Pro にアクセス可能。API は gpt-5/mini/nano のトークン単価を公式提示(Thinking/Pro の API 提供は現時点の公式価格表では未記載)。OpenAI Help Center³、OpenAI Platform⁷
- 総合比較: 速さと手軽さの GPT-5、難題に強い GPT-5 Thinking、さらに推論時間と並列計算で精度を押し上げる GPT-5 Pro、という階層。大量実運用の開発では API のコスト最適(キャッシュ割引、処理ティア)や推論ノブの活用が要点。OpenAI⁴、OpenAI Platform⁷

1. 公式発表・論文・技術ブログによる存在確認と概要

- GPT-5（統合システム）：OpenAI は GPT-5 を「これまでで最も高度な統合 AI システム」として発表。コーディング、数学、ライティング、ヘルス、視覚などで SOTA 級を謳い、会話の種類・複雑性・ツール需要・明示指示（例: “think hard”）に応じて、高速モデルと深い推論モデルを即時に切り替えるリアルタイム・ルーターで動作します。OpenAI¹、OpenAI²
- GPT-5 Thinking（深い推論）：難題に対して長めの推論を施す構成。o3 世代より少ない出力トークンで同等以上の性能を引き出す効率改善が強調され、視覚推論・エージェント型コーディング・大学院レベルの科学問題解決で有意な改善が記載されています。OpenAI¹、OpenAI²
- GPT-5 Pro（並列テスト時計算を伴う Thinking の強化運用）：ChatGPT では「GPT-5 Thinking Pro」として提供。内部は Thinking と同系ですが、推論時に並列に計算パスを走らせて回答品質を底上げする運用で、複雑・多段の課題で好まれやすいとされます（専門家の選好が Thinking に対し高率、という社内評価が紹介）。OpenAI¹、OpenAI²

2. 技術仕様（アーキテクチャ、パラメータ数、コンテキスト長、マルチモーダル）

- アーキテクチャ/運用構成：「高速・高スループット (gpt-5-main)」 「深い推論 (gpt-5-thinking)」 「リアルタイム・ルーター」の統合。ChatGPT では Thinking を並列テスト時計算で強化した Pro を UI から選択可能（Pro/Team）。API は thinking/mini/nano 系列の直アクセスと推論深さノブを提供。OpenAI²、OpenAI⁴
- パラメータ数：公式未公表（本調査時点）。OpenAI の公開資料ではパラメータ規模の明示は見つかりませんでした。OpenAI²
- コンテキスト（入出力上限）：公式ページでの明示は確認できず。著名開発者の実測・整理として「入力約 272k トークン、出力約 128k（不可視推論トークン含む）」という非公式観測がありますが、これは OpenAI の公式仕様ではありません（参考情報として要留意）。Simon Willison’s Weblog⁸
- マルチモーダル：テキスト・コードに加えて画像・動画・空間・科学図表などの理解を含むベンチマークでの強化が示唆。API 入出力の標準はテキスト中心ですが、視覚・動画系の理解評価での改善がシステムカードに言及されています。OpenAI¹、OpenAI²
- 開発者向けノブ：verbosity（short/medium/high 相当）で応答長、reasoning_effort（minimal/low/medium/high）で推論時間と精度をトレードオフ。独自ツール呼び

出しのプレーンテキスト化と CFG 制約も導入。OpenAI⁴
図版（公式システムカードより、SWE-bench Verified の図示）

- [cdn.openai.com](#) (PDF) ⁵

3. ベンチマーク（MMLU、HumanEval、GSM8K など）

- 公式システムカードの要点: 多言語 MMLU(0-shot)の言語別表を掲載(例: Japanese 0.898 など)。総合平均は表記なし。HumanEval や GSM8K の直接数値は未掲載。SWE-bench Verified は図示で示されるものの本文に明示値はなし。[cdn.openai.com](#) (PDF) ⁵
- 外部報道・所感: Wired や VentureBeat は Thinking/Pro が健康関連質問 (HealthBench)、長鎖推論・コーディング系 (SWE 系) で優位と報じます。ただし一部数値比較 (例: HealthBench Hard の具体値) に整合性の疑義がある記述も見られるため、一次の OpenAI 資料での確認が無い項目は参考扱いが妥当です。WIRED⁹、VentureBeat⁶
- 現時点の結論: MMLU では Thinking が main より高水準。HumanEval/GSM8K の公式数値は未提示のため、定量比較は保留。SWE-bench Verified は図示上の優位性が示唆されるが数値は PDF 図表参照。実運用指標（事実性・指示追従・へつらい低減など）での改善が強調されています。OpenAI²、[cdn.openai.com](#) (PDF) ⁵

4. ターゲットユーザー層・主用途の違い

- GPT-5（既定）: 迅速回答を優先しつつ、必要に応じて自動で深い推論へ切替。日常の Q&A、ドラフト作成、一般的なコーディング、情報要約・変換、軽量のエージェント的タスクなど、幅広いカバーを想定。OpenAI Help Center³、OpenAI¹
- GPT-5 Thinking: 難易度が高い推論問題（数理・科学的分析・複雑な設計・多段ツール実行）で品質優先。自動ルーティングのほか、Plus/Team/Pro では手動選択で確実に深い推論を使えるのが強み。OpenAI Help Center³、OpenAI²
- GPT-5 Pro (Thinking Pro): 並列テスト時計算でさらに精度・網羅性を狙う。曖昧課題、長手の合議・比較・反証、医療・法務・研究開発の高信頼ユースケース、企業意思決定支援・高度自動化などでの採用を想定。VentureBeat⁶、OpenAI¹

デモ・所感（メディア/研究者）

- Wired/ZDNET の取材では、単一プロンプトから数十秒で洗練されたインタラクティブ Web アプリを生成するデモが示され、実務体験でのスピード・使い勝手の進化が強調されました。WIRED⁹、ZDNET¹⁰
- Ethan Mollick 氏は、既定の自動ルーティングでは「簡単」と判定されると弱いモデ

ルで即応し、「難しい」と判定されると Reasoner に切替わる、と実例で解説。Thinking を明示選択すると長めの思考で質が大きく上がる一方、望外の追加演出（SVG アニメ効果等）が混入するなど制御上の注意点も報告しています。One Useful Thing¹¹

図版（ChatGPT の Thinking UI イメージ）

Thinking longer for a better answer
Get a quick answer >

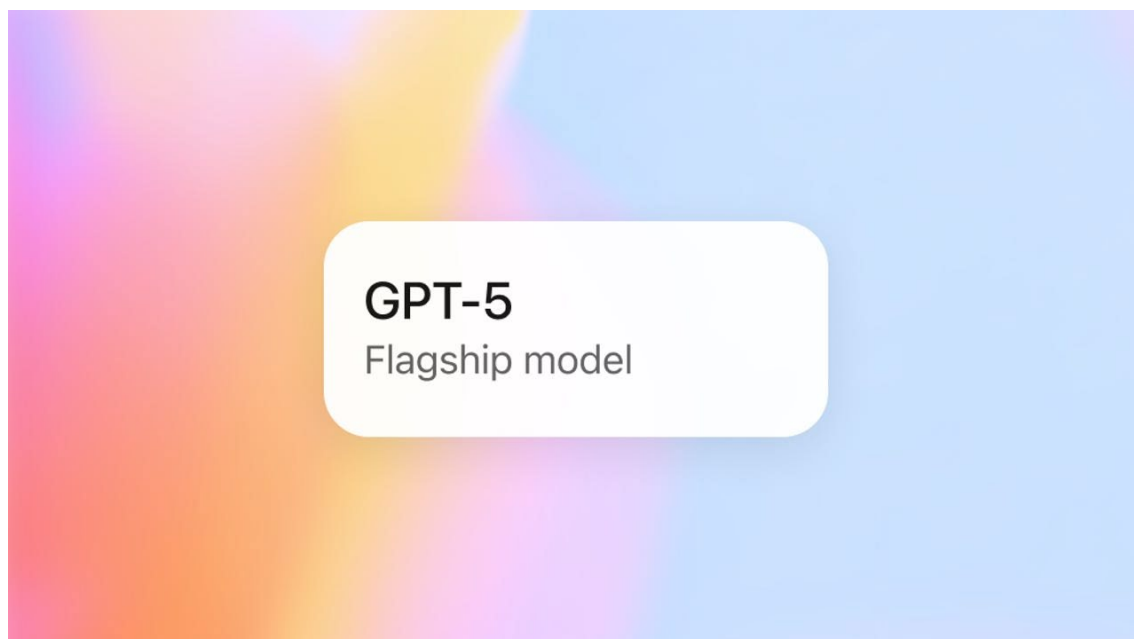
-

OpenAI Help Center³

5. 機能面の差異（速度、推論品質、創造性、特定タスクの精度）

- 速度と効率: GPT-5 は「考える」前に最小限の推論で即時回答も可能。Thinking は o3 比で 50～80%少ない出力トークンでも同等以上の性能を出しうる効率化が示され、プロンプトと履歴のシグナルにより自動的に深い推論へ移行します。開発者は reasoning_effort=minimal～high で明示制御可。OpenAI¹、OpenAI⁴
- 推論品質と事実性: 事実誤り・幻覚・へつらいの低減がシステムカードで強調。Thinking/Pro は曖昧・多段課題で選好され、より堅牢な回答を導きやすい設計です。OpenAI²、OpenAI¹
- 創造性・エージェント性: コーディング協業、長鎖ツール呼び出し、説明付きの行動計画提示などが強化。デモでは複雑なインタラクティブ Web アプリの生成も実演されています。WIRED⁹、ZDNET¹⁰
- ユーザー反応の揺らぎ: 初期ローンチでは、Plus 層が Thinking 週 200 件上限や旧モデル縮小に不満という報道・SNS が散見。速度・品質の体感は個人差が大きく、正式ベンチの公開待ちの領域もあります。TechRadar¹²

図版（メディア掲載のヒーロー画像）



•

ZDNET[10](#)

6. アクセス方法（API/製品）と料金体系

- ChatGPT（Free/Plus/Team/Pro、Enterprise/Edu は順次）：
 - Free: GPT-5 にアクセス。上限到達後は自動で mini へ切替。Thinking は 1 日 1 メッセージ枠（ヘルプ記載）。OpenAI Help Center[3](#)
 - Plus: GPT-5 を既定で利用、モデルピッカーで Thinking を手動選択可。Thinking の手動利用は週 200 メッセージ上限（自動切替はこの上限に含まれない）。OpenAI Help Center[3](#)
 - Team/Pro: 無制限に近い扱い（乱用防止ガードレールは適用）。Thinking Pro（UI 上「GPT-5 Pro」）にアクセス可能。Pro は月額\$200 と報じられ、最上位モデルの利用権含むと解説されています。OpenAI Help Center[3](#)、WIRED[9](#)、ZDNET[10](#)
- API（開発者向け）：
 - 公式価格（百万トークン単価）: gpt-5 入力\$1.25/出力\$10、gpt-5-mini 入力\$0.25/出力\$2、gpt-5-nano 入力\$0.05/出力\$0.40。キャッシュ割引や処理ティア（priority/flex）も提供。OpenAI Platform[7](#)
 - Thinking/Pro の API 露出: 現行の公式価格表は gpt-5/mini/nano 中心で、ChatGPT UI での Thinking Pro に相当する API SKU の明記は確認できません(将来の更新可能性はありうるため、ドキュメントの随時確認が推奨)。OpenAI Platform[7](#)

7. 3 モデルの違いを明確化（性能・機能・コスト・最適用途）

- GPT-5（既定）：
 - 性能/機能: もっとも広い汎用性。高速応答で「まず答え」を返し、必要時には自動で Thinking へ。事実性・指示追従・へつらい低減など実運用品質が底上げ。OpenAI[2](#)
 - コスト/アクセス: ChatGPT の全層で提供（使用量は層により異なる）。API では\$1.25/\$10（入/出力）と攻めた価格設定。OpenAI Help Center[3](#)、OpenAI Platform[7](#)
 - 最適用途: 一般対話、要約、ドラフト、ライトなコーディング、軽量の自動化・エージェント。OpenAI[1](#)
- GPT-5 Thinking:
 - 性能/機能: 難題・曖昧課題・多段推論・長尺タスクに強い。o3 比でトークン効率改善の示唆。手動選択で確実に深い推論を走らせられる（Plus/Team/Pro）。OpenAI[1](#)、OpenAI Help Center[3](#)
 - コスト/アクセス: ChatGPT で週 200 メッセージの上限（手動選択分）。自動切替は別カウント。API では推論ノブ（reasoning_effort）で Thinking 相当の挙動を調整可能。OpenAI Help Center[3](#)、OpenAI[4](#)
 - 最適用途: 数学・科学・設計・方針立案・レビュー・長鎖ツール計画など、正確性と説明性を要する課題。OpenAI[2](#)
- GPT-5 Pro（Thinking Pro）：
 - 性能/機能: Thinking と同系だが、並列テスト時計算でさらに品質・網羅性を引き上げる運用。専門家選好でも優位とされ、実務での確度・堅牢性を重視する用途に合致。OpenAI[2](#)、OpenAI[1](#)
 - コスト/アクセス: ChatGPT の Pro/Team（および順次 Enterprise/Edu）で提供。API の公開価格 SKU としては未記載（将来変更の可能性あり）。OpenAI Help Center[3](#)、VentureBeat[6](#)
 - 最適用途: 企業の意思決定、医療・法務・研究の高リスク領域、長手の合議・反証・網羅照合が必要な案件、SWE/エージェント運用での落ち戻り低減。VentureBeat[6](#)

簡易比較表（要点抜粋）

観点	GPT-5	GPT-5 Thinking	GPT-5 Pro (Thinking Pro)
概要	汎用・高速。自動で Thinking へ切替	長めの推論で難題に強み	並列テスト時計算で Thinking を強化
公式ベンチ指標	多言語 MMLU 表 (0-shot) など。HumanEval/GSM8K は未揭示	同左。MMLU で main より高水準	同左（運用面の強化）
コスト / 提供	ChatGPT 全層、API \$1.25/\$10	ChatGPT で手動選択 (Plus/Team/Pro、週 200 件上限)、API は推論ノブ	ChatGPT Pro/Team (順次 Enterprise/Edu)
主用途	日常対話、要約、軽コーディング	数学・科学・長鎖ツール・曖昧課題	企業・研究の高信頼タスク、長手検討

OpenAI[1](#)、OpenAI[2](#)、OpenAI Help Center[3](#)、OpenAI Platform[7](#)

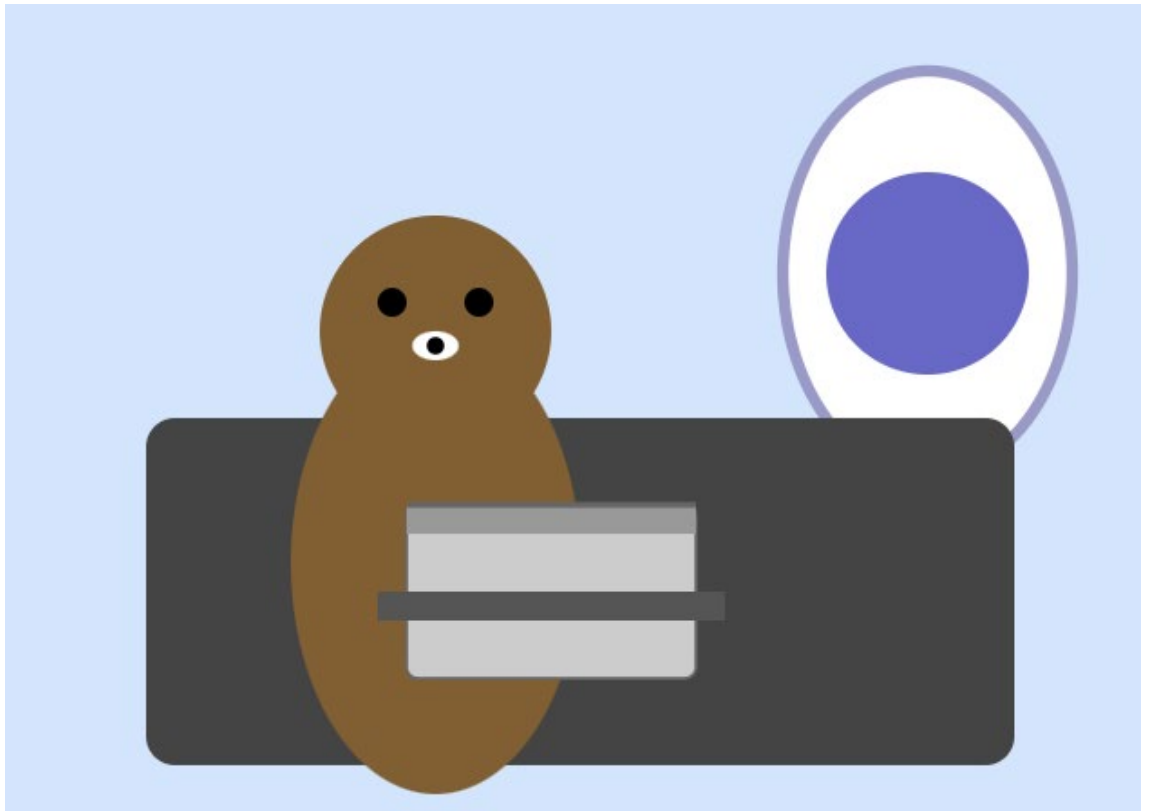
実務導入の指針

- まずは GPT-5 既定＋自動ルーティニングを基本に、品質基準を満たせない領域で Thinking（手動）を併用。重要判断・高リスクでは Pro（Thinking Pro）を選択し、レビュー工程を標準化。OpenAI Help Center[3](#)
- API では reasoning_effort をタスク別に最適化。高頻度・繰返しタスクは入力キャッシュ割引や Batch/flex ティアで徹底的にコスト最適化。OpenAI[4](#)、OpenAI Platform[7](#)

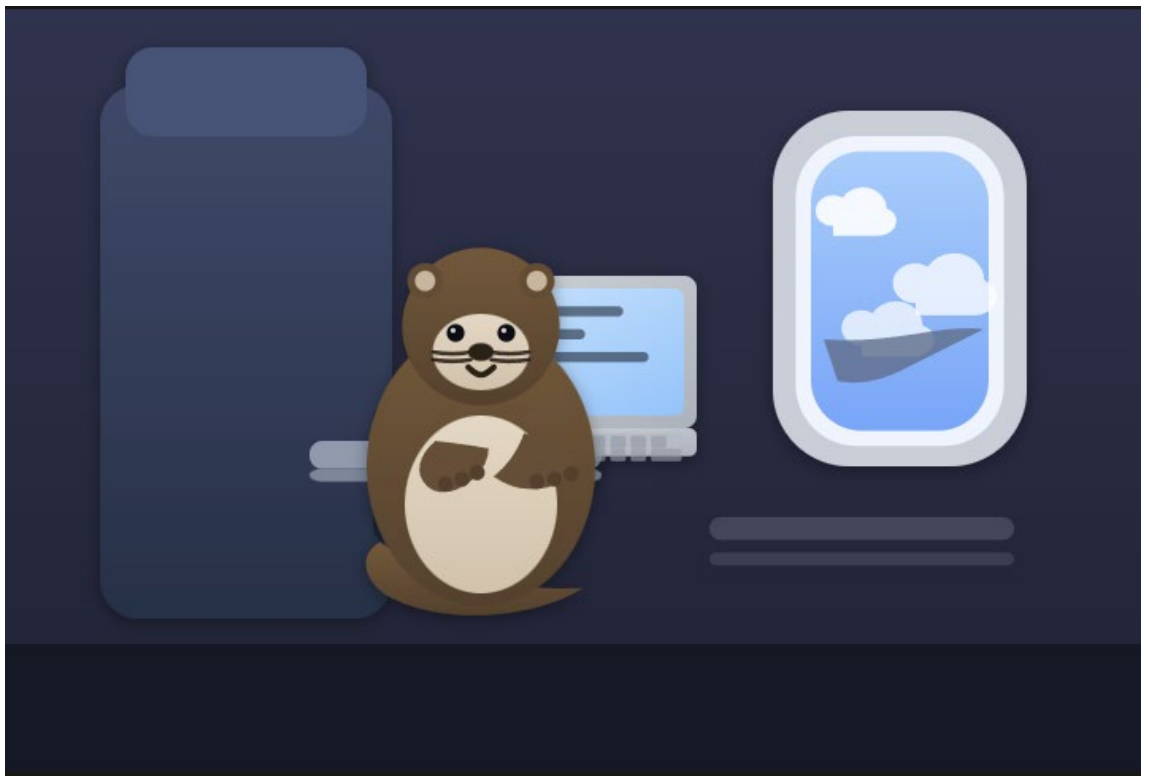
既知の制約・未確定事項

- パラメータ数は未公表。コンテキスト長も公式明記なし（非公式観測は参考扱い）。OpenAI[2](#)、Simon Willison’s Weblog[8](#)
- HumanEval/GSM8K 等の公式スコアが未揭示のため、当該ベンチでの厳密な横比較は時期尚早。MMLU や SWE-bench Verified など公開情報の範囲で評価すべき。cdn.openai.com (PDF) [5](#)
- 初期ユーザー反応には不満の声もあり（Thinking の週上限、旧モデル縮小、体感品質の揺らぎ）。運用安定化・UI 改善の継続が見込まれます。TechRadar[12](#)

参考（追加のマルチメディア）

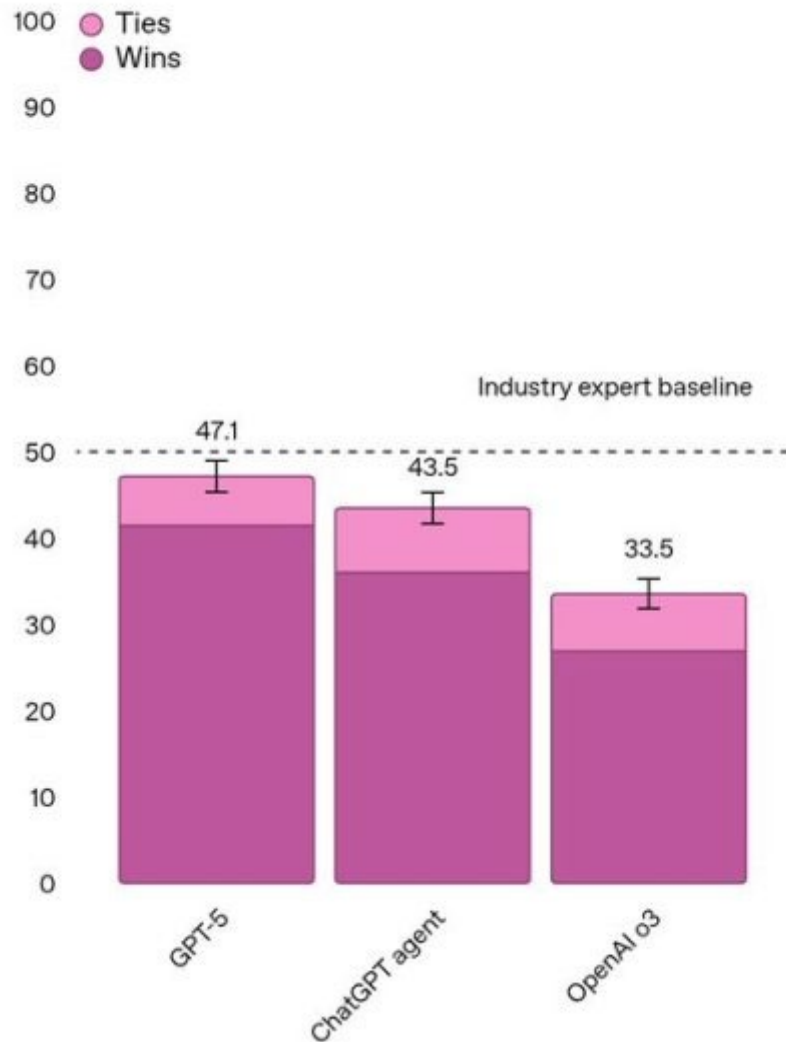


/



One Useful Thing[11](#)

Economically important tasks



•

Vent

ureBeat[6](#)

エンティティ (ナビ用リンク)

- OpenAI: Introducing GPT-5[1](#)
- OpenAI: GPT-5 System Card[2](#)
- OpenAI Help: GPT-5 in ChatGPT[3](#)
- OpenAI Platform Pricing[7](#)
- Introducing GPT-5 for Developers[4](#)

補足 (調査設計)

- 一次：OpenAI 公式発表、システムカード PDF、ヘルプセンター、API 価格表・開発者ブログを精読し、仕様・提供形態・安全設計・評価方法を確認。OpenAI[1](#)、

OpenAI^{[2](#)}、cdn.openai.com (PDF) ^{[5](#)}、OpenAI Help Center^{[3](#)}、OpenAI^{[4](#)}、OpenAI Platform^{[7](#)}

- 二次：Wired、VentureBeat、ZDNETなどを参照し、用途・体感・価格まとめ・導入事例を補強。ユーザー反応はTechRadarで俯瞰、開発者視点はSimon Willison氏の記事でAPI挙動・非公式観測を参照（非公式は明示）。WIRED^{[9](#)}、VentureBeat^{[6](#)}、ZDNET^{[10](#)}、TechRadar^{[12](#)}、Simon Willison's Weblog^{[8](#)}

もっと詳しく

^{[1](#)}
openai.com

^{[2](#)}
openai.com

^{[3](#)}
help.openai.com

^{[4](#)}
openai.com

^{[5](#)}
cdn.openai.com

^{[6](#)}
venturebeat.com

^{[7](#)}
platform.openai.com

^{[8](#)}
simonwillison.net

^{[9](#)}
www.wired.com

10

www.zdnet.com

11

www.oneusefulthing.org

12

www.techradar.com