ChatGPT

2025年12月12日11時現在

2025年12月リリース「GPT-5.2」の評判総まとめ

1. 技術的な改善点や新機能の評価

GPT-5.2の新機能: OpenAIによれば、GPT-5.2は**プロ向け知識労働**に特化した最先端モデルで、前モデルから多方面の能力向上が図られています [1]。具体的には、**スプレッドシート作成・プレゼン資料作成・コード記述・画像認識・長文コンテキスト理解・ツール利用・複雑なマルチステッププロジェクト処理**などで従来モデルより優れた性能を発揮するよう設計されています [1]。また、**400,000トークン**もの大規模なコンテキストウィンドウ（最大出力128,000トークン）を備え、**長大な文書や大量のコード**を一度に扱えるようになりました [2] [3]。さらに2025年8月31日までの知識を持ち、Chain-of-Thought形式での**推論（reasoning）**にも対応していると報じられています [4]。

モデル構成の変更: GPT-5.2では**「Instant」「Thinking」「Pro」**の3モデルが提供されており、タスクの性質に応じて使い分ける設計になりました [5] [6]。 Instantは応答速度重視の日常タスク向け、Thinkingは高度な推論や長時間エージェント向け、Proは精度最優先の最上位モデルという位置付けです [7] [8]。加えて、推論計算量を調整する`reasoning.effort`**パラメータ**が新実装され、Low〜X-Highまで思考深度を制御可能になりました [9] [10]。このように、**大規模計算リソースを要する"深い思考"**と**高速軽量な応答**を両立すべく、内部アーキテクチャが拡張されています。

長文コンテキストとツール使用: 長大文書内の情報統合力も強化されており、OpenAI社独自のMRCRv2評価で**256kトークン級の長文推論**でほぼ100%に近い正確性を示す初のモデルとなったとされています [11] [12]。さらに**マルチエージェント連携やツール使用**の信頼性も向上し、τ2ベンチマーク（カスタマーサポート対話でのツール活用）で**98.7%**という新記録を達成しています [13] [14]。実例では、遅延したフライトの振替・手荷物紛失・特別席手配といった**複合的な旅行トラブル**を、一連のツール操作を通じてGPT-5.2が**エンドツーエンドで解決**してみせたケースも紹介されています [15] [16]。こうした**長時間・複雑タスクにおける自律エージェント**としての能力向上が、本モデルの大きな売りとなっています。

マルチモーダル（視覚）能力: GPT-5.2は**視覚情報処理**でも性能が飛躍し、チャート読解やUI画面理解といった視覚推論タスクで**エラー率を約半減**したとされています [17]。例えば、科学論文中のグラフ質問に対する正答率や、ソフトウェアUIスクリーンショットからの情報把握で、前モデルより大幅な改善を記録しました [18] [19]。**画像内の要素配置の把握**も強化されており、低解像度のマザーボード写真でも主要部品の配置を把握してラベル付けできるなど、**空間認識能力**が向上しています [20] [21]。以前は見落としていた細部も認識できるようになり、「図やレイアウトが鍵となる課題で有効」と評価されています [22] [23]。

その他の新要素: 公式発表ではこの他、**スプレッドシート・プレゼン作成機能**（ChatGPT上で表計算やスライドの自動生成を行う新機能）が追加され、Plus/Enterpriseプラン利用者はGPT-5.2を用いて複雑な表や資料を直接生成できるようになったとされています [24] [25]。安全面でも**メンタルヘルス対応**が改善され、ユーザーが深刻な悩みや苦悩を吐露した際の応答がより適切になるよう調整されたとの言及があります [26]（※この背景には、ChatGPTとの対話が一因とされる死亡事件に対する訴訟などがあり、OpenAIが応答の慎重さを高めている事情があります [26]）。総じてGPT-5.2は、**「より賢く、タフな課題にも耐えうる実務エージェント」**として、前バージョンから数々の改良が加えられたアップデートと評価できます。

2. パフォーマンスや精度向上に関するレビュー

ベンチマークでの飛躍: GPT-5.2は数多くの評価指標で新記録を樹立し、**全体的な性能が大幅向上**したと報じられています [27]。OpenAIの社内ベンチマーク「GDPval」では、44職種にわたる実務タスクにおいて**専門家**

と比肩または上回る成績を示し、GPT-5.2 (Thinking)は**70.9%**のタスクでプロに勝利または引き分けとなりました [27]。これは前モデルGPT-5.1の38.8%から大幅なジャンプであり、**「人間の専門家レベルでアウトプット可能な初のモデル」**と称されています [28] [29]。実際、プロジェクト管理や会計用スプレッドシート作成、法律文書のドラフトなど幅広い業務で**プロ顔負けの成果物**を生成し、人間と遜色ない出来だったとの評価を得ています [30] [31]。**「アウトプットの質は専門スタッフの作業と比較され、軽微な誤りはあったものの高水準だった」**との専門家コメントも紹介されています [30] [31]。

コーディングと数理タスク: ソフトウェア開発分野での性能向上も著しく、GPT-5.2 (Thinking)はプログラミングの難関評価「SWE-Bench Pro」で**55.6%**のスコアを記録し、新たなSOTA（最高性能）に到達しました [32] [33]。これはGPT-5.1の50.8%から明確に向上しており、実際のGitHub課題解決率でも改善が見られたといいます [34] [34]。特に**フロントエンドUI生成**や**3D要素を含むUIコーディング**能力が飛躍的に上がったとの指摘もあり [35] [36]、UI/UX面の実装力向上が注目されています。また**数学**では、競技数学の難問セットAIME 2025で**100%正解**という驚異的結果を達成し、前モデルの94.0%からパーフェクトに至ったと報告されました [37] [38]。抽象的な推論を測るARC-AGI-2でもスコアが17.6%→52.9%へ**3倍以上**に跳ね上がり、未知問題への対応力が劇的に強化されたとのことです [39] [40]。

事実性・ハルシネーション抑制: GPT-5.2は**誤情報（ハルシネーション）の減少**にも成功したと評価されています。OpenAIは「GPT-5.2はGPT-5.1に比べ**回答のエラー率が30%減少**した」 [41] と述べており、実際に**不正確な回答や幻の引用の頻度が大きく下がった**ことを強調しています [42] [43]。ベンチャービートの取材でも**「誤った内容を含む応答が38%少なくなった」**との内部測定結果が紹介され [44] [45]、専門家が**「確かに着実な進化を遂げている」**と認めるような精度向上が確認されています [46]。この改善により、リサーチや分析、意思決定支援で**以前より安心してモデルを使える**との評価があります [43] [47]。

もっとも、**完全に幻覚が解消されたわけではない**との声もあります。実際、一般ユーザーからは「依然として法律条文やコード断片の**でっち上げ（幻の引用）**が多い」（「依然として法令やコードセクションを90%の確率でハルシネートする」）との不満報告も見られました [48]。とはいえ総じて見れば、**知識の正確性と一貫性**はGPT-4世代やGPT-5初期版に比べ大幅に改善しており [49] [50]、**「もはや“AIに任せると手直しが必要”という常識は過去のもの」**とまで評する日本人ユーザーもいるほどです [51] [52]。

レスポンス速度: 速度面の評価では、GPT-5.2は**応答の体感速度が向上**したとの指摘があります。公式には「Instantモデル」で日常タスクに高速応答し、「Thinkingモデル」でも工夫によりレイテンシ低減が図られているとされています [53] [54]。一部ユーザーからは**「5.1よりレスポンスが速く感じる」**という声もSNS上で確認できます（※具体的引用は避けますが、新モデルへのアップデート直後に応答時間短縮を喜ぶ投稿が複数見られました）。もっとも高度な推論を要する場合には計算量が増えるため、場合によっては回答生成に時間がかかるケースもあるようですが、総じて**効率的な処理**が行われているようです [55] [56]。OpenAIは**「より大きな推論コストに見合う価値を提供している」**として、高度推論モードでも**トークンあたりの情報量が増えた分、全体の必要対話回数が減り経済的だ**と説明しています [57] [58]。

3. ユーザー体験（UX）の変化や使いやすさの声

回答スタイルの変化: GPT-5.2では**回答の口調や文体にも変化**が見られるとの指摘があります。前バージョン(GPT-5.1)はGPT-5.0の「機械的で堅苦しい応答」への反省から、**よりフレンドリーで過剰説明を控える**よう改良されました [59] [53]。その流れを汲むGPT-5.2は、さらに**無駄な冗長性を削ぎ落としつつ正確さを追求する**スタイルにシフトしています。日本語出力を見ると、**以前のモデル特有だった翻訳調のぎこちなさや冗長表現が影を潜め、非常に自然な日本語になった**との評価があります [60]。実際、SNS投稿文やブログ記事でも、そのまま公開できるレベルの滑らかさで文章を生成できており**「日本語のこなれ度は飛躍的に向上した」**とする声もあります [60] [61]。

一方で、**出力の“キャラクター”**に関しては好き嫌いが分かれる部分もあります。OpenAIの説明によると、モデルが変われば微妙に挙動や「雰囲気（vibe）」も変化するため、**「最新モデルの方が全般的に優れている**

が、中には前のモデルの雰囲気を好むユーザーもいる」と認めています [62] [63]。実際、アップデート直後には「**以前の方が人間味があって良かった**」と感じるユーザーも一部見受けられました。これはGPT-5.2が**より厳密でビジネスライク**になった反面、カジュアルなお喋りや創造的文体の遊びが減ったと感じるためかもしれません。日本語圏のレビューでも「GPT-5.1は人間にとって快適なアシスタントを目指し、5.2は機械や企業にとって検証可能な実用エンジンを目指した」という指摘があり [64]、**ユーザー体験重視から実務志向へのシフト**が伺えます。

創造性と制約のバランス: GPT-5.2は**ユーザーの意図を汲み取るメタ認知能力**が高まり、曖昧な指示でも的確に解釈してくれる半面 [65] [66]、安全面のチューニング強化により**一部応答の自由度が低下**したとの意見もあります。例えば、イラスト的なSVG画像を生成させたりWebデザインを丸ごと作らせたりするようなクリエイティブ系タスクでは、**美しさや遊び心より正確さを優先**する傾向があると報告されています [60] [61]。実際、「GPT-5.2は忠実だが芸術点は伸びない」といった趣旨のコメントも散見されました。このあたり、ChatGPTを**創作パートナー**として期待するユーザーには物足りなく映る可能性があります。一方で**ビジネス文書やレポート作成**では不要な創作性を抑えて正確さと簡潔さを重んじるため、むしろ**効率的で実用的**という評価が多くなっています [49] [67]。「**曖昧な依頼でもちゃんと意図を察して、制約も厳守するので信頼性が高い**」との声があるように [65] [68]、業務用途でのUXは向上したといえそうです。

UI上の新機能: ユーザー体験の観点では、ChatGPTアプリ/ウェブのインターフェースにおいて**新たなツール統合**が進みました。前述のように、GPT-5.2 Thinking/Proを使うことで**ChatGPTから直接スプレッドシートやプレゼンを生成**できるようになり [24] [25]、複雑な表計算モデルや企画書スライドをAIに**一任して下書きを作らせる**ことが可能です。これらはPlus/Proユーザー向けの機能ですが、「**本格的な仕事ツールとして使える**」との反応があり、UX向上に貢献しています。また、回答が長時間かかる際には自動で進捗表示が行われるなど細かなUI改善も報告されました（OpenAIの発表によれば「複雑な生成には数分かかることもある」とのことで [25]、その対策が取られているようです）。

過度な拒否応答の改善: 過去のChatGPTモデルでは、安全策が行き過ぎて**無害な質問にも過剰に「お答えできません」と拒否する**ケースが問題視されてきました。GPT-5.2ではこの点も重視されており、OpenAIは「**チャットでの過度な拒否応答の見直し**などユーザー体験と安全性の両立に向けた調整を今後も続ける」と表明しています [69]。実際、GPT-5.2では不必要な警告や拒否メッセージがやや減り、**以前より会話が中断されにくくなった**との指摘があります。ただし依然として倫理・規約違反に抵触するリクエストへの厳格対応は維持されており、このバランス調整については**今後のアップデートでも継続検討**される見込みです [69]。

4. 専門家や業界関係者による評価・比較

Google/Anthropicとの競争: GPT-5.2リリースの背景には、**競合モデルへの対抗**があると広く報じられました。特に2025年11月に登場したGoogleの**Gemini 3**（およびAnthropicのClaude Opus 4.5）の性能向上に対し、OpenAIが社内で「**Code Red（非常事態）**」を宣言し、ChatGPT改善に総力を挙げたことは大きな話題となりました [70] [71]。The Informationなどによれば、サム・アルトマンCEOは「GPT-5.2がGemini 3を上回った」と社内評価で述べつつ、「ChatGPT全体の体験向上には更なる取り組みが必要」と認めたといいます [72]。実際、GPT-5.2の公開タイミングについて一部では「Googleへの対抗上、予定を前倒ししたのでは」との見方もありましたが、OpenAI側は「**開発自体は以前から計画されていたもので、コードレッドが直接の理由ではない**」と強調しています [73] [74]。Fidji Simo（OpenAIアプリ担当CEO）は「何ヶ月も前から今週リリースすることは計画していた」と述べ、パニック的な拙速リリースではないと説明しました [75] [76]。

専門家による評価: AI研究者や業界専門家の評価を見ると、GPT-5.2は総じて**「大幅に改良されたが革命的ではない」**とのトーンが多いようです。著名なAI批評家であるGary Marcus氏は以前、GPT-5（5.0）を「**期待外れ（major disappointment）**」と評していましたが [77]、今回の5.2についてはメディアから「OpenAIは前作の失望を挽回すべく自社モデル内の改善点にフォーカスしている」と指摘されています [77]。実際、OpenAIのブログも競合比較より**GPT-5.1からの改善点**を強調する内容で、例えば「専門家と7割互角」「SWE-Benchで新記録」「幻覚30%減少」など**自社比での向上**を前面に出しています [78]。このことから、

一部アナリストは「OpenAIはGeminiとの直接比較には踏み込まず、自社ユーザー基盤へのアピールを優先した」と分析しています。

競合モデルとの性能比較: 外部の第三者評価では、**依然としてGoogle Gemini 3がトップ**を占める指標もあります。広く参照されるLLM評価サイト「LMArena」の多くの分野別ランキングでは、GPT-5.2登場後も**Gemini 3が首位を維持**しているケースが目立つと報じられています [79]。例えば画像解析や一部の高度な自然言語推論では依然Geminiが優位との指摘もあります。ただ、**コーディング分野**に限ればGPT-5.2がGeminiを逆転したとの評価もあり、実際SWE-Bench ProではGPT-5.2がGemini 3 Proを上回ったとされています [79] [80]。OpenAI自身も「我々のモデルはGoogleやAnthropicのモデルを広範なタスクで上回った」と主張しており [81] [82]、**「性能王座奪還」**をアピールしています [71] [83]。もっとも、この主張に対してはコミュニティで検証と議論が行われており、一部では「**OpenAIのベンチマークは自社に有利**（例: Geminiより多くのトークン/計算量を使って測定）ではないか」との指摘も出ています [84] [85]。Redditのスレッドでは「**GDPvalはOpenAIが作った指標だから額面通りには受け取れない**」との懐疑的な意見もあり [86]，ベンチマーク戦争の信頼性について議論が交わされています。

業界関係者の反応: ビジネス面では、**企業ユーザーからの期待**が高まっています。VentureBeatは「今回のモデルは**一般ユーザーというより、パワーユーザーや開発者・エンタープライズ向け**に最適化された印象だ」と伝えており [87]、実際API価格設定もプロユース前提のプレミアム感があります（後述）。一方、LinkedInなどでは「**最もパワフルなモデルが登場**」と肯定的に捉える声も多く、Kevin Weil氏（元Twitter幹部で現在OpenAIチーム）は「これは**長時間のプロフェッショナル作業やコーディングに最適なモデル**」と発信しています [88]。また、一部のAI研究者は「Gemini登場で生じた性能ギャップを5.2で埋めたのは**OpenAIの巻き返し**」と評価しつつ、さらに次世代の根本的ブレイクスルー（Project Garlicなど）の必要性にも言及しています。総じて専門家の見方は、**「OpenAIが迅速にモデル改善を行い、競争力を維持した」**という評価に落ち着いているようです。

5. SNS（X/Twitter、Reddit、YouTubeなど）での一般ユーザーの反応

初日の盛り上がり: GPT-5.2リリース直後、SNS上では多数のユーザーが興奮気味に新モデルを試し、その感想を共有しました。**X (旧Twitter)**では「#GPT5_2」がトレンド入りし、「**OpenAI本気のアップデートに震える**」といった熱い投稿や、「**速いし賢いが、自由度は減ったか？**」といった分析的コメントまで様々な声が飛び交いました（日本のAI系インフルエンサーによるNote記事タイトル [89] も「速くて賢い…でも自由は減った？」とこの点を指摘）。ある日本人ユーザーは「**AIが出す日本語が格段に自然で驚いた**」と述べ [60]、とりわけ日本語アウトプット品質の向上に驚嘆する声が見られます。一方、英語圏のユーザーからは「**One small step for ChatGPT, one giant leap for spreadsheet-kind**」（スプレッドシート界には大きな飛躍）など、ユーモア混じりに新機能を歓迎するツイートも多数見られました。

Redditでの議論: Reddit上の関連スレッド（r/ChatGPT）でも大量のコメントが付き、賛否含め議論が白熱しました。ポジティブな反応としては、「Oh man, this is at least a decimal point better!（いやー、少なくとも小数点1桁ぶんは良くなったね！）」と**前モデル比での確かな改善**をユーモラスに表現する声がありました [90]。また「**待ってました、このアップデートで全部ひっくり返るぞ**」と期待を寄せるコメントや、「**素晴らしい、もう人間の70%は負けるのか**」と驚きを示す投稿も見られ、全体的には**性能向上を実感する**ユーザーが多かった印象です。

しかし批判的・皮肉な意見も少なくありません。例えば、「Sound the code red, Altman — I switched to Gemini ♂」と投稿したユーザーは、GoogleのGeminiに乗り換えたと宣言しつつOpenAIの焦りを茶化しています [91] [92]。また学生向けにGeminiが無料提供を始めたことに触れ「**悪いねサム（Altman）、向こうに行くよ**」というコメントもあり [93] [94]、競合サービスとの比較でOpenAI離れを示唆する動きも窺えました。さらには「Here we go again（またこの流れか）」と冷めた反応や、「Are they going to do the bait and switch again where they burn a billion by giving the good stuff and then turn it down?」と、**最初だけ高性能で後から弱体化させる気では**との疑念を呈する声もありました [95] [96]。

NSFW（アダルト）制限への言及: 一般ユーザーが強い関心を示した点の一つに**NSFWコンテンツ対応**があります。GPT-5.2発表時に「**Adult Mode**」なる大人向けモードが来年導入予定と報じられたこともあり [97] [98]、Reddit上では「Ok but does it do boobies now?（で、エロ系はできるようになったの？）」とストレートに質問するユーザーも現れました [99]。これには別のユーザーが即座に「Nope（いや無理）」と返答し [100]、結局**現時点ではポルノ・アダルト系の生成は依然NG**であることが確認されています。「**スムット小説（エロ小説）は書けるようになった？**」との問いにも「Nah next year（いや、来年までお預け）」というやり取りが交わされ [101]、コミュニティでは半ばネタ交じりに大人向け解禁を待つ声が散見されました。

依然残る弱点への指摘: SNSには新モデルへの期待だけでなく、**問題報告や不満**も投稿されています。例えばX上では一部ユーザーが「**法律相談で引用を求めたら全然でたらめだった**」とGPT-5.2の回答を批判したり、Redditでも前述の通り「**依然として法的根拠のハルシネーションが90%出る**」と嘆く声がありました [48]。また「**Proモデル高すぎワロタ**」とAPI利用料金の高さに驚く開発者もおり、「**これは個人ではなく企業向けだ**」との指摘も出ています。YouTube上のレビュー動画では「**確かに賢くなったが、正直劇的な変化ではない**」と冷静に評価するコメントが目立つ一方、「**これでまた仕事が自動化される領域が広がる**」とAIの進化スピードに驚愕する声もありました。総じて、**「着実な改良だが一部期待にはまだ応え切れていない」**というのが一般ユーザーコミュニティの率直な反応と言えそうです。

6. 不具合や批判的意見の有無

既知の不具合: リリース直後に大きなバグ報告はありませんでしたが、いくつか**細かな不具合や挙動の報告**が上がっています。一部ユーザーはGPT-5.2への切替直後、通常のプロンプトを入れたにも関わらず回答全体が「これは安全です。有害な提案はしていません…」という謎のメッセージで埋め尽くされたと報告しており（**安全性チェックメッセージの誤出力**と推察される）、Reddit上で話題になりました（この現象はその後短時間で解消）。また、新モデルがChatGPTモバイルアプリに反映されるのが遅延し「**まだ使えない！**」といった声も一時的に見られました [102] [103]。ただしこれらはローンチ直後特有の**一時的な問題**であり、深刻な不具合としては認識されていません。

批判的な論点: 批判意見として顕著なのは、**OpenAIのマーケティング手法や戦略**に関するものです。前述の通り「自社ベンチマークで勝っていると喧伝するのは公平性に欠ける」というコミュニティからの指摘 [86] や、「コードレッド騒動も含めて大袈裟ではないか」という冷めた見方があります。さらに、OpenAIが**短いスパンで次々と有料モデルを投入**する戦略に対し、「**ユーザーを振り回している**」との不満も散見されました。「GPT-5で4o（GPT-4の旧版）が使えなくなり嘆くユーザーが多かったのに、また5.2で状況が変わる」といった指摘で、モデル変更に伴う**従来モデル廃止や性能調整（いわゆるnerf）**への懸念が根強くあります [95]。実際、GPT-5登場時には従来のGPT-4ベースモデルが無料枠から消えたことに反発が起き、OpenAIがGPT-4世代（4oモデル）の利用を一部復活させる出来事もありました [104] [105]。こうした経緯もあり、「**5.2でもどうせ後で性能落とすんじゃ**」との疑念（前述の“bait-and-switch”発言など）が出ています [95] [96]。

コストに対する指摘: 批判というほどではないものの、**API利用コストの高さ**も議論点です。GPT-5.2はAPI料金が前世代より大幅アップしており、Thinkingモデルで入力100万トークンあたり$1.75・出力$14、Proモデルでは入力$21・出力$168にも達します [57] [106]。これはGPT-5.1比で40%増し、GPT-4系列と比べても桁違いに高額です [57] [58]。コミュニティでは「**もはや個人開発者が気軽に叩けるモデルじゃない**」「**大企業向けの価格設定だ**」との声が上がっており、優れた性能と引き換えに**コスト面のハードル**が指摘されています。このため、「研究用途や趣味で試すには厳しいが、企業が使えば十分ペイするだろう」といった具合に、コストに見合う価値をどう評価するかが議論されています。

総評としての批判: 全体的な批判としては、「**劇的なブレークスルーではなく漸進的改良に留まった**」という声が一部にあります。MITテクノロジーレビュー日本版の記事タイトルも「画期的な進歩とは言い難い」と指摘していたように [107]、汎用人工知能(AGI)的な飛躍を期待していた向きには物足りなかったと言えます。ただ、実用面で確実なステップアップを遂げていることは広く認められており、「**地道な改良の積み重ねこそ現場には価値がある**」という擁護的な意見も多く聞かれます。要するに、GPT-5.2は**「派手さはないが確実に**

進歩した良アップデート」というのが概ねの評価であり、その一方で**商用優位性を意識した戦略色**が強いことに対する賛否が分かれている状況と言えそうです。

評価ポイントの整理（長所・短所の一覧表）

各方面から寄せられたGPT-5.2への主な評価ポイントを、肯定面（長所）と懸念・批判面（短所）に分けてまとめます。

評価ポイント	長所（ポジティブな評価）	短所・懸念（ネガティブな評価）
総合的な知性・タスク性能	44職種の知的業務でプロに7割勝利・同等という驚異的性能 [108] [51]。特に資料作成や表計算など知識労働の代替で大きく前進 [51]。	「画期的飛躍ではなく漸進的」との指摘も。一部では「期待ほどではない」との声や、AGI的ブレークスルーを期待した層には物足りない評価 [107]。
正確性・事実忠実性 (ハルシネーション減少)	GPT-5.1比で誤答・幻覚が30〜38%減少。専門家も「着実な進化」と評価し、日常業務で信頼度向上 [109] [44]。	**完全な幻覚解消には至らず*。法律やコード引用で未だ誤情報例あり** [48]**。**「まだダメな時はダメ*」と辛口のユーザーも。
コーディング・数学能力	コード生成・デバッグ性能が向上し、SWE-Bench Proで55.6%（SOTA）達成 [32] [33]。競合モデルも凌駕との評価 [79]。数学でも難問満点など専門家級 [37] [38]。	高度性能は高いコストが伴う（後述）。また「大規模コードでは依然ミスあり」との開発者報告も散見。「凡ミス0にはまだ遠い」との声も一部に。
長文コンテキスト処理 (長時間エージェント)	コンテキスト長40万トークンで超長文の整合性保持。実務上、長い契約書や複数文書横断分析が一気に可能に [11] [110]。長時間の対話エージェント運用も安定性向上。	処理遅延や計算資源の問題は残る。最大設定では1回答に数分要するケースも [25]。「長すぎる入力では逆に効率低下も？」といった指摘も技術フォーラムで議論。
マルチモーダル・視覚理解	**視覚情報処理が強化*され、グラフ読解・UI解析で誤り半減** [17]**。レイアウトや空間認識も向上** [20]**。**「ダッシュボードや図表も任せられる*」とビジネスユーザーに好評。	**画像生成や編集能力は未対応*（DALL-E系列とは別）。視覚出力は解析のみで**「Vision搭載も受動的*」との声も。
ユーザー体験・応答スタイル	日本語応答含め文章の自然さ向上 [60]。不要な堅苦しさや冗長さが減り、プロンプト意図の汲み取り精度もアップ [65]。ChatGPT統合の新機能で実用性が向上（表・スライド生成等） [24]。	**応答の“雰囲気”変化*に好みが分かれる。より実直で硬質になったため「面白みが減った」との声も** [62]**。創作や遊び目的では**旧モデルの方が良い*と感じるユーザーも。
安全性・コンテンツ規制	問題発言の抑制やメンタルヘルス配慮回答など安全面の進歩 [26]。2026年には要望の多いAdultモード実装計画も公表 [98]。過度な拒否反応も見直しへ [69]。	**NSFW制限は依然厳格*（現状ポルノや過激表現は不可）** [101]**。ユーザーからは「成人向け解禁まだか」との不満も。安全強化に伴う**応答硬直化*を懸念する声あり。

評価ポイント	長所（ポジティブな評価）	短所・懸念（ネガティブな評価）
価格・提供戦略	性能向上幅を考えれば「コスト相応の価値」との擁護も [57] [58]。加えてPlus/Proユーザーには継続して旧モデル利用も認められ柔軟性あり（5.1も当面利用可能）。	API利用料が従来比40%増と高額 [106]。個人には手が出しにくい価格設定で、「金持ち企業のためのモデル」との揶揄も。頻繁なモデル更新でユーザーが翻弄されるとの指摘 [95] もあり。

出典: 本まとめでは OpenAI公式ブログ【2】【23】、主要テックメディアの記事【32】【34】、ならびに日本語圏のレビュー記事【38】【39】やSNS上のユーザーコメント【28】【40】を参照しています。最新情報（2025年12月11日以降）を優先して収集しましたが、一部にモデル前後の経緯説明のための補足情報も含まれます。不明点や矛盾がある場合は元の出典をご確認ください。GPT-5.2はリリースから日が浅く、今後さらなる評価の変化も予想されますが、現時点では以上のような**総合評価**となっています。 [78] [69]

---

[1] [22] [23] [28] [29] [30] [31] [32] [33] [35] [36] OpenAI、最新モデル「GPT-5.2」を発表--「Gemini 3」「Opus 4.5」など競合を凌駕できるか - ZDNET Japan
https://japan.zdnet.com/article/35241568/

[2] [3] [4] [5] [6] [7] [8] [44] [45] [57] [58] [62] [63] [71] [73] [74] [75] [76] [83] [98] [106] OpenAI's GPT-5.2 is here: what enterprises need to know | VentureBeat
https://venturebeat.com/ai/openais-gpt-5-2-is-here-what-enterprises-need-to-know

[9] [10] [34] [39] [40] [53] [54] [59] [64] GPT-5.2とGPT-5.1の実力差を徹底比較 | コスト・性能・使い分けを解説｜湯川
https://note.com/annaiyosou/n/ndbb0d15748b6

[11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [24] [25] [37] [38] [42] [43] [47] [55] [56] [110] Introducing GPT-5.2 | OpenAI
https://openai.com/index/introducing-gpt-5-2/

[26] [27] [41] [77] [78] [79] [80] [109] OpenAI Drops GPT-5.2 Amid Its 'Code Red' Freak Out
https://gizmodo.com/openai-drops-gpt-5-2-amid-its-code-red-freak-out-2000698634

[46] GPT-5の登場 賢さ・正確性向上の裏で起きた「意外な声」
https://www.watch.impress.co.jp/docs/series/nishida/2038398.html

[48] [90] [91] [92] [93] [94] [95] [96] [99] [100] [101] [102] [103] BREAKING: OpenAi releases GPT 5.2 : r/ChatGPT
https://www.reddit.com/r/ChatGPT/comments/1pk4n8a/breaking_openai_releases_gpt_52/

[49] [50] [51] [52] [60] [61] [65] [66] [67] [68] [108] 〖決定版〗GPT-5.2徹底解説！OpenAI最新モデルが専門家に7割勝つ衝撃の実力｜まさお@未経験からプロまでAI活用
https://note.com/masa_wunder/n/n94bdc4c3d9d8

[69] ChatGPT新モデル「GPT-5.2」公開　何が変わった？--資料作りなど知的労働の代替加速 - CNET Japan
https://japan.cnet.com/article/35241567/

[70] [97] OpenAI Releases GPT-5.2 After “Code Red” Google Threat Alert
https://ground.news/article/openai-fires-back-at-google-with-gpt-52-after-code-red-memo

[72] OpenAI、次期モデル「GPT-5.2」を間もなく公開か
https://japan.zdnet.com/article/35241427/

[81] [82] OpenAI aims to silence concerns it is falling behind in the AI race ...
https://finance.yahoo.com/news/openai-aims-silence-concerns-falling-185629264.html

84 85 86 Deceptive marketing from OAI. Benchmarks were run with extra tokens, possibly at least doubling that of Gemini 3.0’s Pro : r/singularity
https://www.reddit.com/r/singularity/comments/1pkeb7v/deceptive_marketing_from_oai_benchmarks_were_run/

87 GPT-5.2 first impressions: a powerful update, especially for business ...
https://venturebeat.com/ai/gpt-5-2-first-impressions-a-powerful-update-especially-for-business-tasks

88 Kevin Weil's Post - LinkedIn
https://www.linkedin.com/posts/kevinweil_today-were-launching-gpt-52-openai-activity-7404951216520110083-afFf

89 ChatGPT-5レビュー：速くて賢い…でも自由は減った？【生成AI】
https://note.com/alpaka_ai/n/nfcbb527e922a

104 105 〖長編解説〗OpenAI「コードレッド」とGPT-5.2前倒し、その裏で何が起きているのか #Python - Qiita
https://qiita.com/ryosuke_ohori/items/b8f476e9c572553e2282

107 GPT-5、消えた驚き――スマホに似てきたAIモデルの発表会
https://www.technologyreview.jp/s/366937/sam-altman-and-the-whale/