

GPT-6 Sol と GPT-6 Luna の特徴と評判

調査レポート 参考文献付

調査基準日 2026 年 9 月 23 日

GPT-6 Astra

OpenAI の公式発表・API 資料、第三者ベンチマーク、報道および公開直後の利用者の反応を整理し、知財実務への適用を考察する。本文中の番号は文末の参考文献に対応する。数値・提供範囲は調査時点の情報である。

1 調査概要と名称の確認

2026 年 9 月 22 日に OpenAI が公開した正式名称は「GPT-6 Sol」と「GPT-6 Luna」である。当初の依頼にある「GPT-5 Sol Luna」は、公開日が一致するこの 2 モデルを指すものとして調査した。[1, 2]

今回の大きな特徴は、前世代に近い総合能力を大幅に低い費用で利用できる点にある。事実性や業務自動化には改善が見られるが、第三者評価では性能が低下した項目も確認された。用途別に能力・費用・確認負担を比較することが適切である。[3]

公開から約 1 日の段階であり、長期運用の評判は固まっていない。本稿は公開資料の調査で、独自のモデル比較実験は実施していない。日本語の特許実務に関する十分な独立検証は今回の調査で確認できなかった。第 11 項の実務適用は、公開評価に基づく考察として示す。

2 モデルの位置づけと主な仕様

GPT-6 シリーズの最上位は Astra である。Sol は難度の高い推論や複数工程の業務を担い、Luna は範囲を絞った反復的な大量処理に向くモデルと位置づけられる。公式ガイドも、課題に必要な推論能力、待ち時間、費用を基準に選ぶことを勧めている。[4]

表 1 モデルの位置づけと主な仕様

項目	GPT-6 Sol	GPT-6 Luna
主な用途	複雑なコーディング 複数工程の業務	範囲を絞った 大量・反復処理
API モデル名	gpt-6-sol	gpt-6-luna
公称コンテキスト長	105 万トークン	105 万トークン
最大出力	12.8 万トークン	12.8 万トークン
入出力	テキスト・画像入力 テキスト出力	テキスト・画像入力 テキスト出力
知識の基準日	2026 年 4 月 20 日	2026 年 5 月 18 日

出典：公式 API モデル資料。[5, 6] コンテキスト長は公式公称値を採用した。第三者サイトの Sol の表示値には差があるため、実際の利用環境の制限は別途確認する。

両モデルの推論設定は **none**、**low**、**medium**、**high**、**xhigh**、**max** である。音声・動画の直接処理は API 仕様ではサポートされていない。また、長文を入力できることと、すべての記載を正確に参照できることは別の評価事項である。公称容量だけで長い特許明細書や大量文献の処理品質を判断することはできない。[5, 6]

3 価格と業務費用への影響

通常処理で入力に 27.2 万トークン以下の場合、Sol の入力・出力料金は前世代の半額となった。Luna も入力は 50%、出力は約 58% 低下している。比較対象の GPT-5.6 Sol は公表されているプロモーション料金である。[1, 7, 8, 9]

表 2 前世代と新モデルの API 料金

モデル	入力	出力	前世代からの変化
GPT-5.6 Sol	\$4.00	\$20.00	比較基準
GPT-6 Sol	\$2.00	\$10.00	入出力とも 50% 低下
GPT-5.6 Luna	\$0.20	\$1.20	比較基準
GPT-6 Luna	\$0.10	\$0.50	入力 50% 出力約 58% 低下

単位：100 万トークン当たりの米ドル。新料金は公式料金表、旧料金はモデル別資料等に基づく。[7, 8, 9]

Astra の同条件の料金は入力 10 ドル、出力 50 ドルである。同じトークン数なら Sol は Astra の 5 分の 1、Luna は 100 分の 1 となる。ただし、実際の業務費用は推論量、再試行、外部ツール利用、人による確認・修正にも左右される。[7]

単純な試算として、1 件当たり「入力 1 万トークン＋課金対象の出力 2,000 トークン」を 1 万回処理する場合、入力は合計 1 億トークン、出力は 2,000 万トークンとなる。料金表からの計算では Sol は 400 ドル、Luna は 20 ドルである。キャッシュ割引、外部ツール料金、想定を超える推論出力などは含めていない。[7]

入力が 27.2 万トークンを超えると、リクエスト全体に長文料金が適用され、入力・キャッシュ料金は 2 倍、出力料金は 1.5 倍となる。Batch・Flex は通常料金の 50% である。大量処理では、モデル選択だけでなく、文献の分割単位や処理方式も費用を左右する。[5, 6]

4 業務自動化とコーディングの性能

複数のアプリやツールを使う業務を評価する Zapier の AutomationBench 1.0.6 では、Sol が比較的低い費用で一定の成績を示している。調査時点の公開値は表 3 のとおりである。[10]

表 3 AutomationBench 1.0.6 の公開結果

モデルと推論設定	スコア	1 課題当たり費用
GPT-6 Astra max	41.4%	\$1.73
Claude Opus 5.5 max	40.0%	\$1.28
GPT-6 Sol xhigh	33.2%	\$0.27
GPT-6 Sol max	32.0%	\$0.34

出典：Zapier。スコアと費用は同一ベンチマークの公表値で、信頼区間や課題構成も考慮して解釈する。[10]

Sol は、この評価で最上位のスコアには届かない一方、課題当たり費用が低い。この測定では max より xhigh の方が高得点・低費用であり、推論設定を上げれば常に結果が改善するとは限らない。なお、ベンチマークの成功率を、そのまま自社業務の完遂率と読み替えることはできない。

コーディングについて、OpenAI は DeepSWE v1.1 で Sol の max 設定が 68.8%、Luna の max 設定が 66.6%と報告した。これは OpenAI が公表した評価条件での数値である。複雑な開発課題でも Luna が一定の能力を持つことを示すが、すべての開発作業で Sol と同等という意味ではない。[1]

5 第三者評価から見た前世代との差

Artificial Analysis のモデル別ページを同時点で比較すると、総合能力の伸びは小さい一方、課題当たり費用と出力速度の改善が目立つ。表 4 はすべて max 設定の結果である。[8, 9, 11, 12]

表 4 Artificial Analysis による同時点の比較

モデル	Intelligence Index	費用 米ドル／課題	出力速度 トークン／秒
GPT-5.6 Sol	47	\$1.99	82.0
GPT-6 Sol	48	\$1.06	131.2
GPT-5.6 Luna	37	\$0.18	142.9
GPT-6 Luna	37	\$0.07	162.4

出典：Artificial Analysis の各モデルページ。[8, 9, 11, 12] 調査時点の Intelligence Index v4.3.2。費用は同指数の評価課題当たりであり、特許業務の処理単価ではない。

過去の別バージョンの Intelligence Index とは直接比較できない。出力速度は生成開始後の速度で、回答前に考える時間や業務全体の所要時間を表すものでもない。公開値は更新されるため、導入時には同じ評価条件で再確認する必要がある。[11, 12]

別指標である Coding Agent Index では、Sol は 55 から 57 へ上昇し、Luna は 43 から 41 へ低下した。Luna の低価格化は明確であるが、難しいコード作業まで一律に改善したとは評価できない。[3]

同日に公開された Claude Opus 5.5 は、Artificial Analysis の Intelligence Index で 58 を記録している。少なくともこの指標では Sol の 48 との間に差があり、費用の低さと能力差を合わせて考える必要がある。[11, 13]

6 事実性と回答を控える傾向

OpenAI は、ユーザーが過去に誤りを指摘した会話に基づく内部評価で、Sol の誤りが前世代のおよそ半分になったと説明している。ただし、誤りを誘発しやすい事例を集めた評価であり、日常利用全体の誤答率を示すものではない。[1]

第三者の AA-Omniscience でも改善が見られるが、回答を控える頻度の増加と、正答率の変化を分けて読む必要がある。[3]

表 5 AA-Omniscience における Sol の変化

指標	GPT-5.6 Sol	GPT-6 Sol
AA 独自のハルシネーション率	92%	60%
回答を試みた割合	99%	83%
全問題に対する正答率	59%	54%

出典：Artificial Analysis。いずれも max 設定。Luna のハルシネーション指標も 93% から 77% に低下した。[3]

このハルシネーション率は、「誤答数を、誤答・部分回答・未回答の合計数で割った割合」と定義されている。「日常の回答の 60% が誤り」という意味ではない。AA は、知らない場合に回答を控える能力も評価している。[14]

Sol では、回答を控える頻度が増えることで誤答が減る一方、全問題に対する正答率は下がっている。知財実務では、存在しない公報や判例をもっともらしく示すことを減らせるなら有益である。ただし、回答保留を「関連文献なし」「該当構成なし」と取り違えず、再確認へ回す仕組みが必要になる。この実務上の示唆は、上記評価からの考察である。

7 文書作成における簡潔さと網羅性

OpenAI は、Sol と Luna の応答について、専門用語や不要な説明を減らし、やや簡潔にする方向で改善したと説明している。読みやすさには利点がある一方、必要事項の網羅性は別に評価すべきである。[1]

Artificial Analysis は、知識労働の成果物を評価する GDPval-AA v2.1 で、Sol が約 100 Elo、Luna が約 75 Elo 低下したと報告した。出力の目視確認では、提示品質の低下や、採点上必要な要素の欠落が後退の要因として挙げられている。[3]

特許調査報告書、拒絶理由対応案、判決評釈では、文章の自然さだけでは評価が足りない。例えば、対象請求項、対比根拠、相違点、反対解釈、留保事項、出典を先に指定し、各項目の充足を確認する運用が考えられる。簡潔な文章でも、判断に必要な事実と論点が残っているかを確認することが重要である。

8 キャッシュと継続業務の支援機能

GPT-6 では、繰り返し使う指示や資料の先頭部分を再利用するキャッシュが改善され、条件を満たす入力の読み出し料金が 90%割引になる。共有部分を 30 分以内に再利用する仕組みに加え、キャッシュの利用状況や再利用できなかった原因を調べる機能も提供されている。[15]

会話途中で推論の強さを変えたり、利用可能なツールを調整したりしても、過去の文脈のキャッシュを保ちやすくなった。GPT-6 向け API には、ツールの処理を待つ間に別の処理を進める機能や、作業途中でユーザーが追加指示を入れる機能もある。[4, 15]

実務上は、共通の評価基準や社内資料を参照しながら、多数の文献を順番に処理する仕組みとの相性がよいと考えられる。知財ツール事業者にとっても、複数回の照合・検証を組み込む費用を抑えられる可能性がある。ただし、こうした製品側の費用削減が利用料金にどの程度反映されるかは、各サービスの設計と価格方針に依存する。

9 指示遵守と安全性の評価

システムカードの社内 Codex 業務を再現した 50,319 件の評価では、Sol の深刻度 3 以上の問題行動の検出件数が、旧 Sol の 66 件から 42 件へ約 36%減少した。これは社内業務のシミュレーションであり、外部利用の事故率を直接示すものではない。[16]

Luna については、安全性評価での高い拒否率の一部が、正当な依頼まで広く拒否する傾向に由来する可能性も公式に記載されている。安全性の改善と、必要な業務を完遂する能力の双方を確認する必要がある。[16]

10 公開直後の評判と提供範囲

報道では、Reuters が今回の発表を、高性能モデルの運用費用を下げる競争の一環として取り上げている。The Decoder は、価格低下に比べて性能面の前進は限定的という論調で報じた。これらは報道・論評であり、媒体自身が条件をそろえて実施した比較実験とは区別する。[17, 18]

Reddit の初期投稿には、安さや使い勝手を歓迎し旧世代に戻らないという意見がある一方、プログラミング能力は旧世代と大きく変わらず、Astra との差が縮まった実感はないという意見もある。また、新モデルを歓迎しながら、Plus の利用枠が依然として早く消費されるとする投稿も見られた。[19, 20]

これらは公開直後の少数の自己申告であり、作業内容や推論設定は統一されていない。利用者全体の満足度を表す統計ではなく、確認すべき論点を把握する材料として読むのが適切である。API 単価の低下と、定額プランでの利用体験も分けて評価する必要がある。

表 6 発表時点の提供範囲

提供先	対象
ChatGPT Work・Codex	Plus、Pro、Business、Enterprise、Edu で Sol・Luna
Free・Go	デスクトップアプリで Luna

提供先	対象
OpenAI API	Sol ・ Luna
通常の Chat	発表時点では未提供

出典：OpenAI の公開発表。段階的な展開であり、通常の Chat と ChatGPT Work を区別する必要がある。[1]

GitHub Copilot でも同日から提供が始まっている。Sol は Pro+ ・ Max ・ Business ・ Enterprise、Luna はそれらに加えて Pro が対象とされる。利用可能なプランや管理者設定は提供先ごとに確認する必要がある。[21]

11 知財実務への適用と検証方法

以下は、公開評価に基づく導入候補の整理であり、特許業務で実証済みの性能評価ではない。大量処理と難しい判断を分け、必要に応じて上位モデルや専門家による再確認を組み込む運用が考えられる。

表 7 知財業務で試す候補と確認事項

業務	試す候補	確認すべき点
公報の書誌情報・技術要素の抽出	Luna	抽出漏れ、原文との対応
大量公報の分類・一次スクリーニング	Luna を中心に Sol で再確認	関連文献の見落とし、判断保留の扱い
請求項分解・構成要件対比表の下案	Sol	要件の脱落、引用段落の正確さ
調査報告書・技術動向レポート	Sol	必須項目の網羅、出典、反対材料
複雑な FTO 論点・拒絶理由対応の検討	Astra 等も 併用して比較	解釈の妥当性、根拠、専門家による確認

本表は分析・提案。特許実務に固有のベンチマークで検証された性能を示すものではない。

Luna の価格は、文献ごとの要約や分類を数千・数万件規模で行う用途で意味を持つ。一方、低価格化により処理件数を増やしても、スクリーニングで関連文献を落とせば実務上の価値は損なわれる。前世代や現在の業務手順を基準に、同じ文献群と正解データで比較することが望ましい。

評価では、関連文献の再現率、構成要件の抽出漏れ、根拠箇所の正確さ、未回答の扱い、人による修正時間を測る。今回のモデルでは、回答が簡潔になっても必要な論点が残っているか、回答を控えた箇所が適切に再確認へ回るかを特に確認したい。

参考文献

閲覧日はすべて 2026 年 9 月 23 日。公開日を確認できない動的ページは「公開日表示なし」とした。公式発表と独立評価では実行環境・推論設定が異なる場合がある。掲示板の投稿は評判の把握に用い、性能の確定根拠とはしていない。

[1] OpenAI. Introducing GPT-6 Sol and Luna.

2026 年 9 月 22 日。参照内容：公式発表。公開日、価格比較、OpenAI 公表ベンチマーク、事実性、応答スタイル、提供範囲。

<https://openai.com/index/introducing-gpt-6-sol-and-luna/>

[2] OpenAI. Changelog — OpenAI API.

2026 年 9 月 22 日の更新項目。参照内容：GPT-6 Sol・Luna の API 公開とモデル識別子。

<https://developers.openai.com/api/docs/changelog>

[3] Artificial Analysis. GPT-6 Sol and Luna push the cost efficiency frontier.

2026 年 9 月 22 日。参照内容：前世代比較、Coding Agent Index、事実性、知識労働の成果物評価。

<https://artificialanalysis.ai/articles/gpt-6-sol-and-luna-push-the-cost-efficiency-frontier>

[4] OpenAI. Model guidance — Using GPT-6.

公開日表示なし 随時更新。参照内容：モデル選択、非同期ツール呼び出し、作業途中の追加指示など。

<https://developers.openai.com/api/docs/guides/latest-model>

[5] OpenAI. GPT-6 Sol Model — OpenAI API.

公開日表示なし 随時更新。参照内容：モデル仕様、知識の基準日、料金条件、対応入出力。

<https://developers.openai.com/api/docs/models/gpt-6-sol>

[6] OpenAI. GPT-6 Luna Model — OpenAI API.

公開日表示なし 随時更新。参照内容：モデル仕様、知識の基準日、料金条件、対応入出力。

<https://developers.openai.com/api/docs/models/gpt-6-luna>

[7] OpenAI. Pricing — OpenAI API.

公開日表示なし 随時更新。参照内容：Standard 処理の Short context 料金を参照。

<https://developers.openai.com/api/docs/pricing>

[8] Artificial Analysis. GPT-5.6 Sol (max) — Intelligence, Performance & Price Analysis.

公開日表示なし 随時更新。参照内容：旧 Sol の料金、Intelligence Index、課題当たり費用、出力速度。

<https://artificialanalysis.ai/models/gpt-5-6-sol>

[9] Artificial Analysis. GPT-5.6 Luna (max) — Intelligence, Performance & Price Analysis.

公開日表示なし 随時更新。参照内容：旧 Luna の料金、Intelligence Index、課題当たり費用、出力速度。

<https://artificialanalysis.ai/models/gpt-5-6-luna>

[10] Zapier. AutomationBench — AI Agent Benchmarks.

バージョン 1.0.6 調査時点の公開値。参照内容：複数アプリを用いる業務自動化のスコアと課題当たり費用。

<https://zapier.com/benchmarks>

[11] Artificial Analysis. GPT-6 Sol (max) — Intelligence, Performance & Price Analysis.

公開日表示なし 随時更新。参照内容：新 Sol の Intelligence Index、課題当たり費用、出力速度と測定定義。

<https://artificialanalysis.ai/models/gpt-6-sol>

[12] Artificial Analysis. GPT-6 Luna (max) — Intelligence, Performance & Price Analysis.

公開日表示なし 随時更新。参照内容：新 Luna の Intelligence Index、課題当たり費用、出力速度。

<https://artificialanalysis.ai/models/gpt-6-luna>

[13] Artificial Analysis. Claude Opus 5.5 takes the top spot on the Artificial Analysis Intelligence Index.

2026 年 9 月 22 日。参照内容：同日公開モデルとの総合指標上の比較。

<https://artificialanalysis.ai/articles/claude-opus-5-5>

[14] Artificial Analysis. AA-Omniscience — Knowledge and Hallucination Benchmark.

公開日表示なし 随時更新。参照内容：正答率・ハルシネーション率の定義、Intelligence Index の構成。

<https://artificialanalysis.ai/evaluations/omniscience>

[15] OpenAI. Better prompt caching for GPT-6.

2026 年 9 月 22 日。参照内容：キャッシュの再利用条件、割引、診断機能、設定変更時の再利用。

<https://openai.com/index/better-prompt-caching-for-gpt-6/>

[16] OpenAI. GPT-6 Astra System Card — Section 11 GPT-6 Sol, GPT-6 Luna.

初版 2026 年 9 月 3 日 Sol・Luna の追補は 9 月 22 日。参照内容：第 11 項、特に 11.3.1 および 11.6.5 の安全性・指示遵守評価。

<https://deploymentsafety.openai.com/gpt-6-astra>

[17] Harshita Mary Varghese and Deepa Seetharaman / Reuters. OpenAI expands GPT-6 lineup with cheaper Sol and Luna models.

2026 年 9 月 22 日。参照内容：報道。低費用モデルの提供と運用費用をめぐる競争。

<https://www.reuters.com/technology/openai-expands-gpt-6-lineup-with-cheaper-sol-luna-models-2026-09-22/>

[18] The Decoder. OpenAI's GPT-6 Sol and Luna cut prices in half but barely move the needle on performance.

2026 年 9 月 23 日閲覧。参照内容：報道・論評。価格改善と性能向上の大きさに関する評価。

<https://the-decoder.com/openais-gpt-6-sol-and-luna-cut-prices-in-half-but-barely-move-the-needle-on-performance/>

[19] Reddit r/codex / 投稿者 k8vinn およびコメント参加者. How does GPT-6 Sol fare?.

公開直後の投稿 2026 年 9 月 23 日閲覧。参照内容：利用者の自己申告。比較条件が統制されていない初期反応。

https://www.reddit.com/r/codex/comments/1wnocd6/how_does_gpt6_sol_fare/

[20] Reddit r/OpenaiCodex / 投稿者 Over-Necessary-4774 およびコメント参加者. Introducing GPT-6 Sol and Luna.

公開直後の投稿 2026 年 9 月 23 日閲覧。参照内容：利用者の自己申告。新モデルへの反応と利用枠に関する意見。

https://www.reddit.com/r/OpenaiCodex/comments/1wnh100/introducing_gpt6_sol_and_luna/

[21] GitHub. OpenAI's GPT-6 Sol and GPT-6 Luna now available.

2026 年 9 月 22 日。参照内容：GitHub Copilot の提供プランと段階的展開。

<https://github.blog/changelog/2026-09-22-openais-gpt-6-sol-and-gpt-6-luna-now-available/>