調査報告　2026 年 11 月 18 日（日本時間）

GPT-6.2 Astra と Sol の性能比較と知的財産業務への影響

GPT-6 Astra

定期監視の一次査読には Sol、重要特許の対比と反論の検討には Astra が有力である。旧世代からの進歩が最も明確なのは、複数資料を指定形式の成果物にまとめる能力と、その作業に要する費用である。［1–5］

OpenAI は米国時間 11 月 17 日（日本時間 18 日）、両モデルを ChatGPT Work、Codex、API 向けに発表した。Work・Codex では Plus、Pro、Enterprise、Edu が対象。組織での利用可能機能は管理設定に左右される。［1］

モデルごとの強み

比較軸	GPT-6.2 Astra	GPT-6.2 Sol
適する仕事	多数の資料を読み、争点・反証・比較表を組み立てる仕事。科学計算、CAD、複雑な文書・表計算・スライド作成。	反復する複数工程の処理。検索結果の整理、平易な文章化、定型表の更新、ツールを使った日常業務。
総合性能［3］	Artificial Analysis の指数は旧 Astra の 57 から 62 へ。モデル系列別では Fable 5.2（65）に次ぐ。	旧 Sol の 55 から 57 へ。旧 Astra と同水準で、同指数の課題当たり費用は旧 Astra の約 15%。
複数工程の仕事［1］	社内 GDPval v2 で旧 Astra 比＋6.2 ポイント、出力トークン 22%減。Agents' Last Exam でも＋5.2 ポイント。	GDPval v2 で旧 Sol 比＋5.3 ポイント。コード変更の採用可能性を測る FrontierCode 2.0 では新 Astra との差は 0.5 ポイント以内。
標準 API 単価［1, 2］	入力 8 ドル、キャッシュ済み入力 0.40 ドル、出力 40 ドル。旧 Astra 比で入力・出力とも 20%安い。	入力 2 ドル、キャッシュ済み入力 0.10 ドル、出力 10 ドル。旧 Sol と同じ。新 Astra の各単価の 4 分の 1。

旧 Astra＝GPT-6 Astra、旧 Sol＝GPT-6.1 Sol。価格は短文脈の標準処理・100 万トークン当たり。両モデルの文脈長は 105 万トークン。入力が 27.2 万トークンを超えるとリクエスト全体で入力・キャッシュ単価が 2 倍、出力単価が 1.5 倍になる。［2］

特許調査費を見積もる際には、API 単価に加えて原文取得、再実行、専門家による照合に要する時間を含める。一般ベンチマークの費用差を案件単価に換算するには、自社案件での計測が要る。

法務と科学の成績から見える改善と限界

法務では「要件の大半を満たす」と「全要件を満たす」に隔たりがある

Harvey LAB-AA v1.1 は 24 法分野・120 課題を扱う。資料を読んで成果物を提出し、3 つの採点用モデルが評価する。厳格スコアは全要件を満たした場合に加点し、重大な事実誤認がある課題は 0 点とする。各要件の達成率も別に集計している。［4］

モデル（推論設定 Max）	全要件を満たす厳格スコア	個別要件の達成率	重大な事実誤認件／課題	費用ドル／課題
GPT-6 Astra	8.6%	90.6%	0.03	13.67
GPT-6.2 Astra	15.0%	93.8%	0.06	6.35
GPT-6.1 Sol	6.9%	90.6%	0.09	2.48
GPT-6.2 Sol	11.1%	91.5%	0.05	1.92

出所は Artificial Analysis。厳格スコアは採点者間の平均。重大な事実誤認は各課題の平均件数であり、誤答率とは尺度が異なる。上記は標準モデル。別構成の Astra for Law 6.2 は厳格スコア 20.8%。［4, 9］

Harvey 自身の先行試験では、Astra は規制・訴訟、Sol はプライバシー、M&A、企業統治で好成績だった。Astra は同一事件でも、社内メモには依頼者側の弱点を示し、対外書面では不要な譲歩を避けたという。中間応答の所内検討メモと提出用案の書き分けに近い。［5］

後退や伸び悩みが確認された場面

検索不能の報告。OpenAI が意図的に障害を設けたモデル単体試験では、Astra が障害を伝えなかった割合は 11.09%から 22.28%に上昇した。検索ツール停止等の試験でも 7.58%から 23.00%に悪化した。Codex の開発者向け指示で両モデルとも改善するが、その指示は API 直接利用には自動で付かない。［6］

重要事項の脱落。外部の Coding Agent Index では、Codex 環境の最大推論設定で Astra は旧世代と同水準、Sol は 3 ポイント低下。Sol は既存コードに関する質疑で重要な事実を省き、当該項目が 14 ポイント低下した。［3］

科学・調査。Vals の公開順位表では Astra は 65.09%から 68.44%へ上昇し、Sol は 59.63%から 57.76%へ低下した。Epoch AI が人手で採点した自社業務では、新 Astra は平均 65%。情報を慎重に扱いすぎて合理的な推計まで省いた例や、報告書冒頭の結論と本文が食い違った例がある。［7, 8］

知的財産業務への影響とモデルの分担

以下は公表試験から導いた実務上の仮説である。日本語特許の先行技術再現率、クレーム解釈、FTO 結論の正確性は、それぞれ別に検証する必要がある。

工程	AI に任せる候補	担当者が判断する事項
SDI・先行技術調査	Sol で検索語と分類候補を展開し、取得文献を発明の目的・手段・効果で分類。Astra で境界事例や除外理由を再点検する。	検索式、対象 DB・国・日付、ヒット件数、取得失敗を記録。既知の重要文献が再現できたか、除外群の抜取でも確かめる。
無効資料調査	Sol が単一文献ごとに構成要件を対応付け、Astra が相違点、引用の前提、組合せに不利な記載を抽出する。	基準日、公知時期、同じ実施形態に属する記載かを原文で確認。新規性の対比と進歩性の組合せは別表にする。
FTO・侵害予防	Sol が国別ファミリーと請求項の版を整理。Astra が対象製品との逐語対比、解釈の分岐、設計変更案を作る。	対象国・時点の権利状態、現行請求項、審査・訂正の経過、実製品の仕様を確認する。［10］
発明発掘・明細書	Sol がヒアリング記録から事実・未確認事項を抽出。Astra が実施形態、変形例、請求項の階層、比較試験の候補を組む。	発明者が実施したこと、着想したこと、AI が後から提案したことを記録。実験値や効果を実測記録に照合する。
拒絶理由対応	Astra に審査官の論理、反論案、補正案を対にして作らせる。Sol は用語・符号と補正前後の差分を確認する。	各補正箇所を当初明細書・図面と照合し、追加する事項と権利範囲への影響を判断する。［11］
契約・ライセンス	Sol が定義、許諾対象、地域・期間を抽出。Astra が例外条項、再許諾、改良技術の帰属と契約間の矛盾を比較する。	別紙・変更契約・優先順位を含めて原文を確認。法務担当者が交渉上許容する条件を決める。
IP ランドスケープ	Sol が名寄せ・分類を補助し、Astra が技術×用途の比較表、事業仮説、役員向け資料を作る。	出願人名寄せ、ファミリー集約、対象年、分母を固定。引用・出願件数と事業利益の因果は別途検証する。

図面にも応用余地がある。Astra の CAD・3D 処理は部品と符号の照合候補を作る手掛かりになる。公報図面からの寸法推定や、別の実施形態に属する部材の組合せには、原図・明細書の確認を付ける。［1, 3］

ソフトウェア・生物関連の発明では、技術分野によって利用上の制約もある。OpenAI は両モデルで高リスクと判断される質問への制限を強化した。先行技術の整理が途中で止まった場合も、未確認資料を明示し、専門家の調査対象に残す。［1, 6］

数値限定クレームで試す導入評価

最初に比較したいのは、重要な限定を落とさずに、査読時間をどこまで短くできるかである。数値範囲・測定方法・実施形態の対応が結論を左右する案件なら、両モデルの差を観察しやすい。

例　厚さと印刷面積率を併記した紙製品

「紙厚 60〜90 µm、図柄面積率 8〜20%」という仮想の請求項を使う。候補文献に「厚さ 70 µm」と「面積率 10%」が別々に見つかったとき、Sol は段落・表・図、測定対象、分母、実施例番号を別欄に抜き出す。読めなかった表は未確認と記録する。

Astra には、二つの数値が同一実施例のものか、単層と積層体を混同していないか、面積率の分母が同じかを確認させる。たとえば、厚さが実施例 1、面積率が実施例 4 なら、その事実を対比表に残す。担当者が原文と測定条件を照合して、当該文献でどこまで立証できるか判断する。

旧モデルと同条件で比較する

手順	知財部門での評価案
評価資料	専門家が結論と根拠段落を確定した過去の公開案件を用意。重要文献、除外すべき類似文献、読めない表や欠落した添付を混ぜる。
比較条件	旧 Astra・旧 Sol、新 Astra・新 Sol を同じ資料、検索権限、推論設定、出力様式で試す。Sol 一次→Astra 精査も別に測る。
品質指標	既知の重要文献の回収率、必須要件の脱落率、根拠段落の正答率、取得不能を正しく報告した割合を別々に数える。
費用と判断	API 費用と再試行費用に、担当者の確認・修正時間を加える。見逃しの影響が大きい案件は、低コストでも優先的に人が再読する。

実装で効く二つの工夫

検索結果の「該当なし」「検索未実施」「本文未取得」を別の状態として保存する。重要な除外理由は、Astra に原文から独立に再判定させる。Sol の要約だけを渡すと、その要約で既に落ちた限定を確認できない。

未公開発明を扱う際は、入力内容、保存期間、外部検索やコネクターに渡す情報を契約・設定に照らして決める。API では標準で学習利用をしない一方、監視用ログや機能ごとの保存条件がある。［12］

参考文献と公開データ

本文の角括弧は下記の文献番号に対応する。いずれも 2026 年 11 月 18 日（日本時間）確認。動的な評価表は当日時点の掲載値を使用した。

［1］**OpenAI**　Introducing GPT-6.2 Astra and Sol
2026 年 11 月 17 日。製品発表、提供範囲、GDPval v2、FrontierCode、専門文書・CAD の評価。

［2］**OpenAI**　GPT-6.2 Astra / Sol Model　Sol 個別仕様
API の文脈長、知識期限、短文脈・長文脈の料金。

［3］**Artificial Analysis**　OpenAI's release of GPT-6.2 Astra and GPT-6.2 Sol
2026 年 11 月 17 日。総合指数、費用、Coding Agent Index、事実知識の評価。

［4］**Artificial Analysis**　Harvey LAB-AA v1.1
24 法分野・120 課題。厳格スコア、個別要件の達成率、重大な誤認、課題単価。

［5］**Harvey**　GPT-6.2 Astra and GPT-6.2 Sol, Coming Soon to Harvey
法律実務の先行試験。分野別の所見、内部メモと対外向け文書の比較。

［6］**OpenAI**　GPT-6.2 System Card
2026 年 11 月 17 日。第 4.1.3 節に作業障害・検索障害の報告に関する評価。

［7］**Vals AI**　Vals Science Index
生物データの分析、機序の推定、研究ワークフローを統合した公開順位表。

［8］**Kelly Hong／Epoch AI**　How good are Fable 5.2 and GPT-6.2 Astra at real work?
2026 年 11 月 17 日。調査・実験・広報など Epoch AI 自身の業務の人手評価。

［9］**OpenAI**　Astra for Law 6.2
法務向け構成の評価、ChatGPT Enterprise 向け申請条件。

［10］**INPIT**　J-PlatPat 操作マニュアル
第 4.2 節「経過情報を参照する」。公報、審査記録、登録情報の関係。

［11］**特許庁**　特許・実用新案審査基準 第 IV 部 第 2 章
「新規事項を追加する補正」。当初明細書等に記載した事項の範囲。

［12］**OpenAI**　Data controls in the OpenAI platform
API データの学習利用、監視ログ、機能別の保存条件。