

MiniMax M3.1 Flash

日本企業の知財業務への影響

公開情報に基づく性能評価と導入方針

調査基準日 2026 年 10 月 20 日 日本時間

GPT-6 Astra

要旨

MiniMax は 2026 年 10 月 19 日、M3.1 Flash を正式発表し、API、M Plan、MiniMax Code で提供を開始した。日本企業の知財部門にとっての主な意味は、公開特許などの大量の文書を読み、分類・比較し、周辺のデータ処理までつなぐ費用が下がる可能性にある。[1]

導入候補として優先度が高いのは、SDI による新着特許の定期監視、競合動向の整理、検索後の一次スクリーニング、知財データの整備である。一方、FTO の結論、権利範囲の判断、未公開発明を含む明細書作成への適用は、根拠の照合や契約条件の確認を組み込んで判断する必要がある。

本稿が推奨する進め方は、公開情報を扱う一つの業務で試行し、重要文献の見逃し、修正時間、出典の正確さを現在の方法と比較することである。採否はトークン単価だけで決めず、確認を終えて採用できる成果物 1 件当たりの総費用と所要時間で評価する。

調査基準日は 2026 年 10 月 20 日朝の日本時間とする。本文の番号は文末の参考文献に対応する。公式仕様、企業公表のベンチマーク、第三者評価と、本稿による業務上の考察を区別して記す。公開情報を調べた範囲では、日本語特許に特化した M3.1 Flash の精度を直接実証する評価は確認できておらず、本稿でも知財案件を用いたモデル実行試験は行っていない。

目次

- 1 公開内容と技術仕様
 - 2 公表された性能評価と知財実務への読み替え
 - 3 料金と費用構造への影響
 - 4 業務別の導入優先度
 - 5 優先業務の具体的な設計
 - 6 データ連携とシステムの構成
 - 7 機密情報と利用規約の影響
 - 8 重み公開と社内運用の可能性
 - 9 知財組織と競争環境への中期的な影響
 - 10 導入を判断するための実証計画
 - 11 日本企業への導入提案
- 参考文献

1 公開内容と技術仕様

M3.1 Flash は、匿名の先行版「Space Bunny」として評価されていたモデルの正式版である。モデルの提供開始と重みの公開は別の日程で、公式 X は重みを 10 月 23 日に公開予定としている。調査時点で、正式版を社内運用できる状態になったと判断することはできない。[1][2]

表 1 公式資料で確認できる主な仕様

項目	確認内容
モデル ID	MiniMax-M3.1-Flash
コンテキスト上限	1,000,000 トークン
モデルの入力	テキスト、画像、動画
主な用途	推論、ツール利用、コーディング、長い文脈を使うエージェント処理
思考量の設定	low / medium / high / xhigh / max の 5 段階。既定値は max
思考の無効化	無効化できない
互換インターフェース	Anthropic 互換および OpenAI 互換。Responses API も提供
生成長の上限	524,288 トークン。推論トークンも生成予算を使用する

出典は Model Invocation、Create Message、Create Response。コンテキスト上限は、長文の全域を同じ精度で参照できることや、特許調査の網羅性を保証する数値ではない。また、生成長の上限すべてを最終回答に使えるという意味でもない。[3][4][5]

入力形式は、モデル自体の能力と、実際に使うアプリ・SDK の対応を分けて確認する必要がある。例えば公式 AI SDK では、PDF 入力は既定の minimax プロバイダーである Anthropic 互換経路に限定され、動画入力は同 SDK では対応していない。Word や Excel のファイルをアプリで扱っても、そのままモデルのネイティブ入力仕様だとは限らない。[6]

知財業務では、公報本文、図面、拒絶理由通知、引用文献、表形式の出願一覧を一つの作業に組み込めるかが重要になる。採用する接続経路について、読めるファイル形式、図面の解像度、OCR の誤り、長文の切り分け方を個別に確認すべきである。

2 公表された性能評価と知財実務への読み替え

企業公表値が示す範囲

MiniMax の発表は、回答文を生成する能力に加え、ツールやコンピューターを使って複数段階の仕事を進める能力を重視している。主な数値は表 2 のとおりである。これらは MiniMax 自身の公表値であり、日本語特許の検索、請求項解釈、進歩性判断を直接評価したものではない。[1]

表 2 MiniMax が公表した主なベンチマーク

評価	公表値	主な評価対象
DeepSWE v1.1	73.7%	ソフトウェア開発の課題
OSWorld V2	61.7 点	コンピューターの操作
ALE	29.6 点	専門的で複雑な業務
CyberGym	82%	脆弱性の発見と検証

公式ドキュメントの **Interleaved Thinking** は、ツールの実行結果を受け取り、次の操作の前に推論を行う仕組みである。知財への応用としては、文献を読む、足りない情報を特定する、許可されたデータ源から追加情報を取得する、分類・比較・集計する、根拠付きの報告を作る、という処理が考えられる。ただし、特許データへのアクセスや各ツールの接続は別途必要になる。[7]

第三者評価と回答保留の問題

Artificial Analysis の比較ページでは、**Intelligence Index** は M3 の 29 に対し M3.1 Flash が 42 である。同 Index のタスク当たり加重平均費用は 0.51 米ドルから 0.22 米ドルとなり、この数値同士の比較では約 57%低い。この費用は複数の評価課題を所定の重みで集計した指標であり、特許公報 1 件の処理費用ではない。[8]

同社の公式投稿では、M3.1 Flash の AA-Omniscience スコアは 16、同評価の幻覚率は 93%である。この 93%は、全回答や特許業務における誤答率を表す数値ではない。[9]

AA-Omniscience の幻覚率は、正答を分母に含めず、「誤答数 ÷ (誤答数 + 回答保留数)」で算定される。したがって、難しい知識問題で正答できなかった場面において、保留よりも誤った回答を出しやすかったことを示す。日本語特許での実務精度をこの値から算出することはできない。[10]

知財の試行では、実在しない段落番号や実験結果を示す、公開時と登録時の請求項を混ぜる、製品情報の不足を要件不充足と判断する、文献にない効果を補う、数値の測定条件を落とす、といった誤りを個別に測る必要がある。根拠が不足する場合に「不明」として人に回せることも、採用に必要な能力として評価する。

3 料金と費用構造への影響

トークン単価の比較

公式の従量料金は表 3 のとおりである。M3 との比較は、入力が 512K トークン以下の標準料金を基準とする。M3.1 Flash の入力・出力単価はこの基準の 50%低く、キャッシュ読み取り単価は 75%低い。[11]

表 3 100 万トークン当たりの料金 米ドル

モデルと条件	通常入力	出力	キャッシュ読取
M3.1 Flash	0.15	0.60	0.015
M3 標準 入力 512K 以下	0.30	1.20	0.060
上記 M3 比の低下率	50%	50%	75%

発表記事の利用コスト「70%削減」は、企業が示すエージェント業務全体の比較である。全トークンの単価が一律 70%下がったという意味ではなく、自社の知財業務で同じ低下率になるとも限らない。処理速度についての「約 3 倍」も企業公表の比較であり、調査完了までの時間がそのまま 3 分の 1 になる保証ではない。[1]

公開特許を読む場合の試算

ここでは、1 文献につき入力 20,000 トークン、推論を含む総出力 3,000 トークン、1 回だけ処理し、キャッシュを使わないと仮定する。この仮定に表 3 の M3.1 Flash の単価を掛けると、表 4 の計算になる。実測値や特許文献の標準的な長さを示すものではない。

表 4 仮定に基づく API 費用の試算

処理文献数	入力費用	出力費用	合計
1,000 件	20 × 0.15 ドル = 3.00 ドル	3 × 0.60 ドル = 1.80 ドル	4.80 ドル
10,000 件	200 × 0.15 ドル = 30.00 ドル	30 × 0.60 ドル = 18.00 ドル	48.00 ドル

表中の 20、3、200、30 は、対象トークン数を 100 万で割った値。入力費用はその値に入力単価を掛け、出力費用も同様に計算した。検索、データベース利用、OCR、実装、再実行、人による確認、税などの費用は含めていない。

この試算が示すのは、単純な一括処理では API 費用よりも、検索・整備・確認の費用が大きな論点になり得ることである。例えば 1,000 件の結果を 1 件 1 分で確認するだけでも約 16.7 時間かかる。思考の既定値が max であることや、文献の長さ、再試行による追加出力も実費を左右する。[3]

したがって、「1,000 件をいくらで生成したか」と併せて、重要文献をどれだけ拾えたか、何件の修正が必要か、採用できる成果物まで何時間かかったかを記録する必要がある。人の確認が増える構成では、低い API 単価だけを導入理由にできない。

4 業務別の導入優先度

表 5 は、公表された機能と知財業務の性質を踏まえた本稿の提案である。優先度は導入効果が実証された順位ではなく、公開情報から試しやすいことと、人が確認できる形に出力を限定しやすいことを重視している。

表 5 知財業務ごとの活用候補と評価点

業務	優先度	任せる候補	主要な評価点
SDI と競合監視	高	課題・手段・効果・対象製品の抽出、新着の変化点整理	重要文献の見逃し、分類の一貫性
検索後の一次選別	高	関連・要確認・除外の仮分類、根拠箇所の提示	誤った除外、判断保留の適切さ
知財データの整備	高	番号の正規化、欠損の確認、表の加工、処理コードの作成	データ消失、誤同定、ファミリーの扱い
IP ランドスケープ	中～高	課題と技術手段の分類、競合の変化の整理	母集団、名寄せ、重複、公開時差
FTO の候補分析	中	請求項の要素分解、製品との対比表の下書き	現行の請求項、権利状態、製品の証拠
明細書と拒絶理由対応	条件付き	用語・記載根拠・引用文献・論理の確認	機密性、新規事項、事実の補作
発明検討と出願方針	条件付き	代替案、追加実験、検討すべき問いの提示	技術的な検証、人の検討経緯の記録
ライセンス契約	条件付き	条項、義務、期限の抽出と比較	機密性、条項間の関係、準拠法

商標や意匠でも特徴の言語化や候補整理は考えられる。ただし、今回の公表値から商標の類否、意匠の類似、登録可能性の判断性能までは確認できない。こうした判断を対象にする場合は、それぞれに合った評価資料が必要になる。

5 優先業務の具体的な設計

SDI と競合技術の継続監視

SDI は、関心のある分野の新着文献を継続的に選び出す運用である。M3.1 Flash の活用では、公報ごとに技術課題、解決手段、記載された効果、対象製品・用途、請求項の特徴、根拠箇所を抽出する。これを過去の文献と同じ分類軸で蓄積すれば、新しい文献が何を変えたのかを見やすくできる。

例えば、同じ課題に対する技術手段が変わったか、新たな用途が加わったか、部品の構成から製造工程へ保護対象が広がったか、といった問いを設定する。週次の出力には、従来との差分、根拠文献、担当者が確認すべき理由を付けると、単なる要約より次の行動につながりやすい。

ここでは、公報に記載された技術と、実際の製品採用や事業上の成果を分けて記録する。特許の記載だけを根拠に「競合が量産を開始した」といった推定を事実として登録してはいけない。また、分類軸を毎回作り直すと時系列比較が崩れるため、承認済みの分類定義を維持し、追加や変更は提案として扱う設計が適している。

FTO の候補分析と対比表

FTO は、事業を実施するうえで他者の特許権などに関する支障を調べる活動である。M3.1 Flash には、候補文献から請求項の要素を取り出し、製品の資料と対応づける下書きを任せる使い方が考えられる。出力には、請求項番号、要素の原文、対応する製品の証拠、対応づけの理由、未確認情報、次の確認事項を保持する。

仮の判定は「充足を支持する根拠あり」「不充足を支持する根拠あり」「情報不足」などに分ける。製品の数値が見つからないことは、その要件を満たさないことの証拠にはならない。測定条件や対象部位が異なると比較できない場合もあるため、数値だけで結論を出さないようにする。

法的状態や請求項の版は、文章の類似度とは別に確認する。INPIT の説明では、J-PlatPat の経過情報には更新の遅れがあり、表示される存続期間満了日は年金の納付状況にかかわらない最長の満了日である。リアルタイムでより正確な登録状況の確認が必要な場合、INPIT は原簿の確認を案内している。対象国、判断時点、現行の請求項、権利状態を記録したうえで対比表を使用する。[12]

文献と製品の文章が似ているだけで侵害の結論は得られない。確認可能な事実と推定を対比表で分け、担当者や専門家が原文と法的条件を照合する工程を残すことが、ここでの実務上の位置づけになる。また、将来の権利化に備える実務上の対応として、公開済みで審査係属中の出願も、請求項の変更や分割出願の状況を継続して確認する対象に含めることを提案する。

明細書と拒絶理由対応の確認

明細書や意見書への応用では、請求項の各要素と明細書の記載根拠、用語・符号・数値の一貫性、拒絶理由と引用文献の対応、補正案の根拠、主張と証拠の整合性を確認する用途が考えられる。これらは、未確認の技術的な事実を新しく書かせる場合より、原文との照合を設計しやすい。

補正案を出す場合には、提案する表現、出願当初の根拠箇所、そこから読み取れる技術的事項、一般化した部分を明示させる。特許庁の審査基準では、当初明細書等の記載との関係で新たな技術的事項を導入するかが、新規事項の判断に関わる。文章として自然な補正でも、この確認を省略できない。[13]

モデルは根拠付きの候補を提示し、担当者が採否を決める構成が適する。未公開の発明内容、出願方針、実験データを使う場合は、第 7 節の契約・データの扱いを先に整理する必要がある。

6 データ連携とシステムの構成

知財業務全体の品質は、モデル単体の性能では決まらない。表 6 のように、データ取得、処理、検証、判断の役割を分けると、改善すべき点と責任を明確にしやすい。これは本稿が提案する構成であり、MiniMax が特許専用システムとして提供している機能の一覧ではない。

表 6 知財業務を構成する四つの役割

役割	扱うもの	主な管理点
データ取得	公報、図面、審査経過、ファミリー、法的状態	利用権限、収録範囲、更新日、版
モデルによる処理	抽出、分類、比較、要約、コード作成	指示、出力形式、思考設定、失敗時の処理
結果の検証	原文、出願番号、件数、計算、引用箇所	実在性、支持関係、欠損、重複
業務上の判断	重要度、追加調査、出願、事業への対応	担当者の評価と判断理由

モデルの低料金が直接効くのは主に処理の部分である。データの収録漏れや出願人名の名寄せの誤りは、モデルを置き換えるだけでは解消しない。公報や請求項の版を固定し、どの文書からどの結果を作ったか追跡できるようにする必要がある。

データ取得では利用条件を守る。INPIT は、J-PlatPat の利用上の案内で、一般の利用を妨げる可能性のある、データの単純収集を目的とする大量ダウンロードや、プログラムによる定期的な自動収集等を禁止している。業務で一括処理する場合は、利用を認められたデータベースの API やエクスポート、適切な提供データなどを選ぶべきである。[14]

既存の特許データベースと、低料金の抽出・整理処理を組み合わせる構成は考えられる。ただし、これをもって Patentfield や Summaria などの特定サービスとの連携済み、あるいは利用料金の値下げが確定したと解釈することはできない。既存契約のデータ利用範囲と実際の接続機能を確認して設計する。

7 機密情報と利用規約の影響

標準 API 規約の機密情報条項

知財部門にとって、料金と同じ程度に重要なのが入力情報の扱いである。2026 年 3 月 30 日発効の MiniMax API の利用規約は、機密情報条項の第 8 項で、アルゴリズム改善またはサービス向上を目的とする機密情報の使用・開示を、機密保持義務違反に当たらないとする例外を置いている。[15]

この条項があることから、すべての入力を実際に学習へ利用されていると断定することはできない。一方、機密保持条項が存在するという事実だけで、未公開発明の利用条件が十分だとも判断できない。契約上、何のために、どの範囲で利用できるかという問題として評価する必要がある。

未公開発明を扱う契約を検討する際には、学習・改善への利用範囲とその除外、前記例外条項の適用、保存期間と削除、ログ・キャッシュ・バックアップ、処理・保管・保守対応の国や地域、再委託先を確認する。この確認事項は、本稿が提案する契約レビューの観点である。

公開資料の調査では、標準条件における包括的な学習利用除外やゼロデータ保持を一律に確約する仕組みまで確認できていない。これは個別の企業契約で利用条件を交渉できないという意味ではない。実際に締結する契約とサービス構成に基づいて判断する。

API と Agent の条件の違い

統合された利用規約には、Open Platform と Agent など、製品に応じた条項がある。したがって、API の説明を Agent の利用条件にそのまま当てはめることはできない。利用する製品の契約を特定して読む必要がある。[16]

Agent のアプリ・Web 利用規約は、入力と出力について、無償、永続的、撤回不能、全世界的な利用許諾を含む広い条項を置いている。入力データの権利や秘密保持の要件と照らして、適用範囲を確認すべきである。有料利用であることだけを根拠に機密性が確保されるとは判断しない。[17]

また、デスクトップアプリから使う場合でも、推論が自社の端末内で完結するとは限らない。利用するアプリがどの通信先に何を送るか、モデルを自社環境で実行するかは、別に確認すべき事項である。

保管地域と公開情報の扱い

API のプライバシーポリシーには、個人データを米国のデータセンターに保管する旨がある。ただし、その説明から全プロンプト、ログ、保守アクセスなどの処理経路を一括して確定することはできない。保持期間も、目的や法令上の必要性などに基づく記載であり、すべてのデータに共通する固定日数の保証として読むべきではない。[18]

公開特許を使う場合にも、検索式や評価軸に未発表の製品計画が含まれることがある。公報が公開済みであることと、それに添える調査目的が公開情報であることは同じではない。試行の初期段階では、外部サービスに渡す処理を公報からの一般的な特徴抽出に絞り、自社製品との対応づけや出願方針の検討を承認済みの環境で行う構成が考えられる。

8 重み公開と社内運用の可能性

公式 X では重みを 10 月 23 日に公開する予定とされているため、公開後には閉じた環境で推論できる可能性を改めて評価する価値がある。ただし、10 月 20 日時点で、正式版の最終ライセンス、利用条件、実行要件がすべて確定したものとして導入を決めることはできない。[2]

MiniMax の Preview 紹介ページの FAQ には、総パラメーター数 428B、推論時に活性化するパラメーター数が約 23B という MoE 構成の説明がある。これは Preview の紹介として確認できる情報であり、正式版の確定したモデルカードの値として無条件に転記すべきではない。[19]

「Flash」という名称や活性パラメーター数だけでは、128GB メモリーの PC で実用的に動くかは判断できない。総パラメーターの保持に必要なメモリー、量子化、長文の処理に伴うメモリー、画像・動画の処理、同時利用、運用速度を含めて評価する必要がある。

表 7 重み公開後に確認する項目

確認項目	導入判断に必要な確認
ライセンス	商用利用、改変、再配布、利用分野などの条件
必要な設備	正式版の総パラメーター数、量子化、GPU とメモリーの要件
業務上の速度	対象文書の長さ、同時処理数、待ち時間
提供機能の一致	画像・動画・長文・ツール利用が API と同じ範囲で動くか
運用の費用	設備、保守、監視、更新、障害対応を含めた総費用

機密の発明情報を社内で扱える構成には価値がある。しかし、重みの入手だけで運用が完成するわけではない。正式な公開物と条件を確認した後、実際の知財文書で処理品質と費用を比べることが適切な判断順序になる。

9 知財組織と競争環境への中期的な影響

以下は、公表仕様と費用構造から導く本稿の考察である。第一に、分類の定義を維持できれば、競合分析の更新間隔を短くしやすくなる。年次や四半期ごとにまとめていた観点をより頻繁に更新し、重要な変化だけを担当者へ渡す運用が考えられる。更新回数を増やすこと自体ではなく、どの変化を早く知ると意思決定が変わるかを先に定めるべきである。

第二に、人数の少ない知財部門でも、同じ文献を用途、課題、技術手段、競合など複数の軸で整理する余地が広がる。ただし、モデルが作る項目を増やすほど確認の負担も増える。実際に調査や出願方針の検討へ回せる時間が増えたかを測る必要がある。

第三に、企業が継続して蓄積すべきものは、自社に合う分類体系、確認済みの重要文献、根拠の記録方法、修正履歴、意思決定の履歴である。これらを特定モデルの会話の中だけに置かず、更新やモデル変更後も使える形で管理すれば、業務の品質を改善しやすい。

第四に、競合企業も公開情報を安く分析できるようになる。自社が公開した特許、論文、製品情報を組み合わせると、開発上の重点や技術の変化をどこまで推定できるかを点検する意義が増す。これは M3.1 Flash だけに固有の現象ではないが、処理費用の低下が公開情報の組合せ分析を広げる可能性がある。

10 導入を判断するための実証計画

30 日から 60 日の進め方

最初の対象には、自社分野の公開特許 300～500 件を使った SDI または一次スクリーニングを提案する。表 8 の日数と文献数は導入計画の例であり、精度の保証や統計的に十分な評価規模を意味しない。現在使っている方法を基準に、同じ資料と評価基準で比較する。

表 8 公開情報を用いた実証計画の例

期間	実施内容	確認する成果
1～10 日目	対象分野と資料を選び、関連・要確認・除外の基準を作る。現行業務の時間を測る	評価用資料、判断基準、現在の品質と時間
11～20 日目	同じ資料・指示・出力形式で、現行方法と M3.1 Flash を比較する	抽出結果、根拠箇所、処理量、API 費用
21～30 日目	担当者による修正時間を測り、除外文献も抽出して確認する	重要文献の見逃し、修正内容、総所要時間
31～60 日目	対象を限定して並行運用し、失敗時の復旧や件数の整合を確認する	安定性、再実行の費用、監査可能な記録

指示や分類を調整する資料と、最終評価に使う資料は分ける。同じ特許ファミリーの近い内容が両方へ入ると、未知の文献への対応を過大評価しやすいため、ファミリー単位で分ける設計が望ましい。化学、機械、電気、ソフトウェアなどでは誤りの現れ方が異なるため、自社の分野で評価する。

日本語特許で測る品質

表 9 日本語特許に即した評価項目

項目	確認例
論理関係	及び／又は、少なくとも、例外、否定を正しく扱うか
従属請求項	引用する請求項の限定を引き継ぎ、要素を落とさないか
数値と測定条件	端点、単位、重量比とモル比、測定条件を維持するか
請求項の版	公開時、補正後、登録時などの内容を混在させないか
根拠の正確さ	文献・段落が実在し、示した根拠が結論を支持するか
回答の保留	情報不足を特定し、無理に判定せず次の確認を示せるか
図面の理解	符号、接続、対応箇所を読み取れるか。読めない箇所を明示するか

合格条件は試行前に決める。例えば、重要文献の再現率が現在の方法を下回らないこと、重大な論理・数値の誤りを検証工程で検出できること、修正を含む総作業時間が減ることを条件にする。数百件で見逃しがなかったとしても、全体の調査漏れがない証明にはならないため、運用後も除外結果の抜取確認を続ける。

設定とモデル変更の管理

思考量の既定値は max である。単純な項目抽出と複雑な対比で同じ設定を使う必要はなく、設定ごとに品質、出力トークン数、待ち時間を比べる。低い思考設定でも必要な根拠が得られる処理を切り分ければ、総費用の調整余地がある。[3]

M Plan の FAQ は、個人の対話的な開発利用と本番処理を区別し、本番用途には従量 API の利用を勧めている。定額プランを前提に継続的な大量処理の費用を計算せず、実際の使用条件と契約に合う経路を選ぶ必要がある。[20]

公式 Model List の移行案内では、M Plan および Token Plan で M2.5 や M2.7 を指定したリクエストは、10 月 19 日から M3.1 Flash へ自動的に転送される。従量 API の既存利用者については 11 月 18 日まで利用でき、その後停止するという別の日程が示されている。両者の扱いを混同しないようにする。[21]

このため、リクエストのモデル名だけで品質が固定され则认为ない。指定したモデル名、応答に含まれるモデル名、実行日時、思考設定、入力文書の版、指示、結果を保存し、更新後には代表的な評価資料を再実行する。これらの記録は比較に役立つが、同じ出力が常に再現されることまで保証するものではない。

11 日本企業への導入提案

表 10 利用範囲ごとの判断方針

利用範囲	当面の判断
公開情報の SDI と分類	優先的に試行し、重要文献の捕捉と総作業時間で採否を判断する
FTO と権利化の準備	根拠の整理と対比の下書きに使い、原文・版・法的条件の確認を残す
未公開発明と機密情報	利用契約と処理環境を確認し、承認された範囲で利用する
社内でのモデル運用	10 月 23 日の重み公開予定後、正式な公開物・ライセンス・実行要件を評価する

M3.1 Flash の公開を受けて知財部門が検討すべきなのは、低料金で処理できる量を増やし、その結果を確認可能な証拠へつなぎ、重要な判断を早く行えるかである。公開情報を扱う一つの業務から始め、実際の品質と確認時間の改善を確かめたうえで対象を広げる進め方を推奨する。

参考文献

本文中の番号順に掲載する。閲覧日はすべて 2026 年 10 月 20 日。随時更新される API 資料、料金、規約、評価ページは調査時点の内容を参照した。日付の記載がない資料は、その旨を示す。X の投稿タイトルがないものには内容を示す資料名を付した。

- [1] MiniMax. MiniMax M3.1 Flash: Redefining Agent Performance and Cost. 2026 年 10 月 19 日公開. 参照箇所: 正式発表、提供経路、企業公表の性能比較と利用コスト.
<https://www.minimax.io/blog/minimax-m31f> (閲覧日 2026 年 10 月 20 日)
- [2] MiniMax 公式 X アカウント @MiniMax_AI. M3.1 Flash 発表投稿. 2026 年 10 月 19 日投稿. 参照箇所: モデル重みの 10 月 23 日公開予定.
https://x.com/MiniMax_AI/status/2112150867868541221 (閲覧日 2026 年 10 月 20 日)
- [3] MiniMax API Docs. Model Invocation. 公開日記載なし. 参照箇所: モデル ID、入力形式、コンテキスト、思考設定、互換 API.
<https://platform.minimax.io/docs/guides/text-generation> (閲覧日 2026 年 10 月 20 日)
- [4] MiniMax API Docs. Create Message. 公開日記載なし. 参照箇所: 生成トークン上限と API パラメーター.
<https://platform.minimax.io/docs/api-reference/text-chat-anthropic> (閲覧日 2026 年 10 月 20 日)
- [5] MiniMax API Docs. Create Response. 公開日記載なし. 参照箇所: 推論トークンを含む出力予算.
<https://platform.minimax.io/docs/api-reference/responses-create> (閲覧日 2026 年 10 月 20 日)
- [6] MiniMax API Docs. AI SDK. 公開日記載なし. 参照箇所: SDK 経路による PDF 入力と動画入力の制約.
<https://platform.minimax.io/docs/api-reference/text-ai-sdk> (閲覧日 2026 年 10 月 20 日)
- [7] MiniMax API Docs. Tool Use & Interleaved Thinking. 公開日記載なし. 参照箇所: ツール利用と処理の途中に行う推論.
<https://platform.minimax.io/docs/guides/text-m3-function-call> (閲覧日 2026 年 10 月 20 日)
- [8] Artificial Analysis. MiniMax Models: Intelligence, Performance & Price. 随時更新. 参照箇所: Intelligence Index と同 Index のタスク当たり加重平均費用.
<https://artificialanalysis.ai/models/creators/minimax> (閲覧日 2026 年 10 月 20 日)
- [9] Artificial Analysis 公式 X アカウント @ArtificialAnlys. MiniMax-M3.1-Flash scores 42 on the Artificial Analysis Intelligence Index. 2026 年 10 月 19 日投稿. 参照箇所: M3.1 Flash の評価結果、AA-Omniscience のスコアと幻覚率.
<https://x.com/ArtificialAnlys/status/2112211865128608002> (閲覧日 2026 年 10 月 20 日)
- [10] Artificial Analysis. AA-Omniscience: Knowledge and Hallucination Benchmark. 2025 年 11 月 16 日公開. 参照箇所: 幻覚率の分母と回答保留の評価方法.
<https://artificialanalysis.ai/articles/aa-omniscience-knowledge-hallucination-benchmark> (閲覧日 2026 年 10 月 20 日)
- [11] MiniMax API Docs. Pricing Overview. 公開日記載なし. 参照箇所: M3.1 Flash および M3 の従量課金単価.
<https://platform.minimax.io/docs/pricing/overview> (閲覧日 2026 年 10 月 20 日)
- [12] 独立行政法人工業所有権情報・研修館 INPIT. J-PlatPat 講習会での質問と回答. 公開日記載なし. 参照箇所: 共通 Q1 および Q7、経過情報の更新と存続期間満了日.
https://www.inpit.go.jp/j-platpat_info/lecture/qa.html (閲覧日 2026 年 10 月 20 日)

- [13] 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. 公開日記載なし. 参照箇所: 新たな技術的事項を導入する補正の判断.
https://www.jpo.go.jp/system/laws/rule/guideline/patent/tukujitu_kijun/ht/04_0200.html (閲覧日 2026 年 10 月 20 日)
- [14] 独立行政法人工業所有権情報・研修館 INPIT. 利用上のご案内. 2026 年 4 月 1 日最終更新. 参照箇所: 第 2 項、J-PlatPat の大量ダウンロードとロボットによる自動収集.
https://www.inpit.go.jp/j-platpat_info/guide/j-platpat_notice.html (閲覧日 2026 年 10 月 20 日)
- [15] MiniMax. MiniMax Terms of Service. 2026 年 3 月 30 日発効. 参照箇所: 機密情報条項の第 8 項ほか、API の機密情報の扱い.
<https://platform.minimax.io/protocol/terms-of-service> (閲覧日 2026 年 10 月 20 日)
- [16] MiniMax. Terms of Service. 2026 年 4 月 15 日更新. 参照箇所: Open Platform と Agent の製品別条項の適用.
<https://www.minimax.io/terms-of-service-v2.html?show=platform,agent> (閲覧日 2026 年 10 月 20 日)
- [17] MiniMax. MiniMax AI App and Web Terms of Service. 2026 年 1 月 19 日更新. 参照箇所: Agent の入力および出力についての利用許諾.
<https://agent.minimax.io/doc/en/terms-of-service.html> (閲覧日 2026 年 10 月 20 日)
- [18] MiniMax. MiniMax API Privacy Policy. 2026 年 3 月 30 日更新・発効. 参照箇所: 個人データの保管とデータ保持に関する条項.
<https://platform.minimax.io/protocol/privacy-policy> (閲覧日 2026 年 10 月 20 日)
- [19] MiniMax Agent Tools. M3.1-Flash-Preview for Everyday Expert Work. 公開日記載なし. 参照箇所: Preview の紹介と FAQ に記載されたパラメーター数.
<https://agent.minimax.io/tools/m3-1-flash-preview> (閲覧日 2026 年 10 月 20 日)
- [20] MiniMax API Docs. FAQ. 公開日記載なし. 参照箇所: M Plan の利用用途、個人の対話利用と本番処理.
<https://platform.minimax.io/docs/m-plan/faq> (閲覧日 2026 年 10 月 20 日)
- [21] MiniMax API Docs. Model List. 公開日記載なし. 参照箇所: M Plan および Token Plan のモデル転送と従量 API の移行日程.
<https://platform.minimax.io/docs/faq/models> (閲覧日 2026 年 10 月 20 日)