



2026-02-19版「Gemini 3.1 Pro」が企業の知財業務にもたらす新機能と運用変革

エグゼクティブサマリー

2026-02-19にプレビューとして提供開始された「Gemini 3.1 Pro」は、同じGemini 3世代の「Gemini 3 Pro」と比べて、(a)推論・エージェント性能の底上げ、(b)思考深度の制御粒度（`thinking_level`に`medium`が追加）、(c)カスタムツール優先の専用モデルID（`...-customtools`）という3点で、知財（IP）業務の“実装可能な自動化の幅”を実務レベルで広げます。¹

一方で、コンテキスト長（入力1M/出力64k）・知識カットオフ（2025年1月）・テキスト入出力の単価は、**3 Proと同等**であり、いわゆる「桁違いの長文投入」や「モデル単価の大幅低下」が差分の中心ではありません。差分の本丸は、“長い・複雑・多段”の仕事を、より確実に、より安定したコストで回すための改善です。²

知財業務に当てはめると、従来から可能だった要約・抽出を超えて、**先行技術調査→クレーム要件分解→対比表（クレームチャート）→拒絶理由対応（OA応答）ドラフト→品質ゲート（根拠検証・改変履歴・監査ログ）**のような“案件ワークフロー全体”を、ツール連携込みで半自動運転に近づけられます。特に、**長期タスク（APEX-Agents）や多段ツールワークフロー（MCP Atlas）、検索・閲覧を伴うエージェント（BrowseComp）**で大幅にスコアが伸びている点は、知財実務（調査・根拠収集・証拠性）との親和性が高い根拠になります。³

ただしプレビュー提供であり、**プレビュー版はレート制限が厳しめになり得る、少なくとも2週間前通知で廃止（deprecate）され得る**ため、企業導入は「本番一括置換」よりも「優先パイロット→統制→段階展開」が合理的です。⁴

事実ベースの差分

仕様・モデルカード・提供形態

Gemini 3.1 Proは、モデルカード上「Gemini 3 Proをベースにした次イテレーション」と位置づけられ、入力はテキスト/画像/音声/動画などのマルチモーダル、最大1Mトークン文脈、出力はテキスト最大64Kトークンとされています。⁵

同じくGemini 3 Proのモデルカードでも、入力が最大1M、出力が64K、知識カットオフが2025年1月であること、一般的な基盤モデルとしての限界（幻覚、遅延/タイムアウトが起り得る）が明記されています。⁶

企業利用（クラウド側の提供）としては、Vertex側のモデルページで、3.1 Proプレビューは**公開プレビュー、リリース日2026-02-19、知識カットオフ2025年1月、最大入出力トークン（1,048,576/65,536）**が記載されています。⁷

“事実差分”一覧（差分がある項目と、同等項目）

事項	Gemini 3.1 Pro（プレビュー）	Gemini 3 Pro（プレビュー）	企業知財への意味
リリース日	2026-02-19（プレビュー）	2025年11月（モデルカード）／Vertex側は2025-11-18リリース	3.1は改善サイクルが速い前提。プレビュー運用の統制が重要。
モデルID	gemini-3.1-pro-preview + gemini-3.1-pro-preview-customtools	gemini-3-pro-preview	ツール優先の別IDがあるかが最大の機能差。
入出力上限	入力1,048,576 / 出力65,536	入力1,048,576 / 出力65,536	“長文の上限”は同等。差は“長文での精度と段取り”。
知識カットオフ	2025年1月	2025年1月	最新情報は検索グラウンディング等の併用が前提。
思考深度制御	thinking_level : low/medium/high (highが既定)	thinking_level : low/high (mediumなし)	大量案件の“ほどほど推論”が実装可能（コスト・遅延の管理がしやすい）。
価格（Developer API）	入力\$2/1M（≤200k）・出力\$12/1M（≤200k）等（3 Proと同額）	同左	単価差ではなく、推論トークンの節約が差を生む。
価格（Vertex）	入力\$2/1M（≤200k）・出力\$12/1M（≤200k）等（3 Proと同額）	同左	企業はZDR/データ統制/スループット確保とセットで評価。

出典（モデルID・トークン上限・知識カットオフ・コスト・思考深度差・カスタムツール専用ID）： 8

API/エンドポイント（差分の核）

Gemini APIの主要エンドポイントは、標準生成（generateContent）、ストリーミング（streamGenerateContent）、バッチ（batchGenerateContent）、埋め込み（embedContent）などで構成されます。 9

3.1 Proのリリースノートには、通常の gemini-3.1-pro-preview に加えて、gemini-3.1-pro-preview-customtools という別エンドポイント（モデルID）が追加され、カスタムツールを優先しやすい旨が明記されています。 10

この ...-customtools は、知財領域で言えば「社内特許DB」「自社出願・契約書リポジトリ」「知財管理システム（IPMS）」「製品BOM/仕様DB」といった社内ツールを確実に呼ばせる設計に寄与します（後述のワークフローで具体化します）。 11

思考深度 (thinking_level) の差分

Gemini 3シリーズはデフォルトで動的推論 (high) を行い、thinking_level で内部推論の最大深度を制御できます。lowは低遅延・低コスト寄り、highは推論深度最大で初回トークンまで遅くなることがある、と説明されています。¹²

この表で重要なのは、medium が「3.1 ProではSupportedだが、3 ProではNot supported」である点です。¹²

知財実務は、(1)大量のスクリーニング (低～中推論) と、(2)意思決定に近い最終分析 (高推論) が混在します。mediumの追加は、“全件high”のコスト爆発や遅延を避けつつ、lowでは精度不足になりやすい中間帯を作れるのが実務上の効能です。¹³

ベンチマーク差分 (知財ワークフローに効く指標を中心に)

モデルカードに掲載された主要ベンチマークのうち、知財業務に近い性質 (長期タスク、検索・閲覧、ツールワークフロー、長文文脈) に寄せて要点をまとめます。³

ベンチマーク (性質)	3.1 Pro (High)	3 Pro (High)	知財への含意 (要旨)
ARC-AGI-2 (未知パターン推論)	77.1%	31.1%	“想定外の組合せ”でも筋の良い仮説を立てやすい=設計回避/無効理由の構成に寄与し得る。
BrowseComp (検索+Python+閲覧)	85.9%	59.2%	公開情報を根拠付きで集めるタスクの適性向上=先行技術調査・証拠集めの安定化。
MCP Atlas (多段ワークフロー)	69.2%	54.1%	多段の手順実行が得意=「抽出→整形→システム投入」系が壊れにくい。
APEX-Agents (長期業務タスク)	33.5%	18.4%	“案件として完了させる”方向に改善=OA応答やFTOの一連処理で効く。
MRCR v2 (長文文脈: 128k平均)	84.9%	77.0%	大量資料からの針探し性能が改善=長い明細書・多数文献の横断参照が安定。
MRCR v2 (1M pointwise)	26.3%	26.3%	1Mを入れれば自動的に良くなるわけではない=“入れ方”設計 (チャンク・RAG・キャッシュ) が重要。

出典: ³

上のうち、特にBrowseComp / MCP Atlas / APEX-Agentsの伸びは、知財業務 (検索→比較→根拠整理→文書化→レビュー) を“エージェント化”する時の成功率に直結しやすい指標です。¹⁴

ベンチマークの可視化 (代表例)

```
xychart-beta
title "知財業務に近いベンチマークの伸び (High) "
x-axis ["ARC-AGI-2", "BrowseComp", "MCP Atlas", "APEX-Agents", "MRCR v2 (128k)"]
y-axis "score" 0 --> 100
bar [31.1, 59.2, 54.1, 18.4, 77.0]
```

```
bar [77.1,85.9,69.2,33.5,84.9]
legend ["3 Pro","3.1 Pro"]
```

(数値出典：モデルカード表) 3

知財ユースケース別：3 Proで詰まりやすかった点と、3.1 Proで現実的になる点

知財業務は、「根拠のある判断」と「大量の反復」が同居します。例えば先行技術調査では、調査時間が増えるほど探索範囲（カバー率）は上がるが、実務上は時間制約がある、という構造が示されています。¹⁵ また、先行技術スクリーニングでは「新規性・進歩性を否定できる文献か」等の観点での選別が中心になります。¹⁶

以下では、各ユースケースで「3 Proだと運用が難しい/コストが合わない/破綻しやすい」ポイントを明確化し、それが3.1 Proのどの差分で緩和されるかを整理します（※法的判断の最終責任は必ず有資格者/社内規程に従う前提）。

先行技術調査（Prior-art search）

3 Proで詰まりやすい点（実装面）

長文（多数公報・非特許文献）からの要件分解と対比表作成は可能でも、(1)推論深度をhigh/lowしか選べず、スクリーニングでhighを使い続けるとコストと待ち時間が読みにくい、(2)社内DB/商用DBの検索APIを確実に呼ばせる“ツール選好”が案件によりブレる、(3)検索・閲覧・抽出・整形という多段プロセスで「途中の破綻（手戻り）」が生じやすい、という運用課題が残りがちです。¹⁷

3.1 Proで現実的になること

- 1) `thinking_level=medium` で、“粗スクリーニング（medium）→上位候補だけhighで精査”の二段階設計が可能になります。¹²
- 2) 検索・閲覧・ツールを伴う指標（BrowseComp、MCP Atlas）が大きく伸びており、**根拠収集→要約→構造化出力**のエラー率低下が期待できます（ベンチマークは直接の知財タスクではないため“期待”である点は留保）。¹⁴
- 3) 社内検索APIや文献管理ツールなど**カスタムツールを強く優先させたい場合に**、`gemini-3.1-pro-preview-customtools` という選択肢が増えます。¹⁰

発明抽出（Invention mining）

3 Proで詰まりやすい点

発明抽出は、実験ノート/設計資料/議事録/コード差分などの“雑多な一次情報”をまとめ、特許性の核（課題・手段・効果）を抽出し、クレーム候補に落とし込むプロセスです。文書理解自体は可能でも、資料が多いほど「まとめが浅い（low）」か「コストが高い（high）」に寄りやすい問題が出ます。¹⁸

3.1 Proで現実的になること

文書（PDF）を“テキスト抽出以上”に理解し、図表も含めて構造化抽出できることを前提に、medium思考で大量の社内資料から“発明の種”を広く拾い、highで少数精査に寄せられます。¹⁸

OA応答（拒絶理由通知対応）

3 Proで詰まりやすい点

OA応答は、多段で「拒絶理由の論理構造を分解→引用文献の該当箇所特定→相違点→主張（新規性/進歩性/記載要件等）→補正案→整合性チェック」を回します。新規性判断が“全要素開示”に依存するなど、要素分解・対比表が核になります。¹⁹

3 Proでもドラフトは作れますが、長期タスク完遂（APEX-Agents）や多段ワークフロー（MCP Atlas）での性能が3.1に劣るため、「ドラフトは出るが、検証・整合性チェックで手戻りが多い」運用になりがちです。

²⁰

3.1 Proで現実的になること

APEX-AgentsやMCP Atlasの伸びを根拠として、“**応答パッケージ（主張骨子+補正案+根拠引用+リスク注記）**”を、より一貫した形で出す期待が合理的になります（ただし、最終確認は人が行う前提）。¹⁴

また、構造化出力で「拒絶理由ごとの論点」「引用文献ごとの根拠箇所」「提案補正（条文制約・サポート要件チェック）」をJSON化し、レビュー工程を機械化できます。²¹

FTO（Freedom to Operate：実施自由度）

3 Proで詰まりやすい点

FTOは「製品/プロセスを計画通り使っても第三者権利侵害の法的責任を負わないか」という観点で、特に“権利範囲（クレーム）”と“有効・存続・地域性”に依存します。²²

運用上は、製品要件→クレーム要件の対応付け→国別の権利状態→リスクマトリクス→設計回避/ライセンス方針、という多段タスクになり、検索・閲覧・整理の高い成功率が求められます。²³

3.1 Proで現実的になること

BrowseComp（検索+閲覧+Python）や多段ワークフロー系の改善を踏まえ、“**候補権利のスクリーニングと根拠整理（ただし最終リーガル判断は人）**”をより広い範囲で自動化しやすくなります。³

さらにmedium思考により、FTOの最初の大量スクリーニングを「コスト・遅延を抑えた中推論」で回し、最終候補だけhighで深掘り可能です。¹²

無効資料調査（Invalidity search）

3 Proで詰まりやすい点

無効資料調査は、先行技術調査よりも“狙い撃ち”で、特定クレーム要件を潰しに行く証拠探索であり、検索→検証→対比表→論理構成を高速で反復します（時間をかければカバー率が上がるが、実務は時間制約がある）。²⁴

3 Proでも可能ですが、探索の反復が多いほど「high思考のコスト」「手戻り」が重くなりがちです。²⁵

3.1 Proで現実的になること

長文文脈での針探し（MRCR v2 128k）改善を背景に、“**引用可能箇所の特定→構造化抽出→対比表生成**”の安定性を上げやすいです。²⁶

契約レビュー（契約・共同研究・ライセンス）

3 Proで詰まりやすい点

契約レビューは、条項抽出・リスク判定・修正案提示の“定型+例外”で、構造化出力が有効です。一方で、企業導入では「秘匿情報を外部に出さない」「ログ・保持・学習利用の統制」が最優先になります。²⁷

3.1 Proで現実的になること

モデル性能差よりも、**データ保持ゼロ（ZDR）や学習制限、キャッシュ制御**を含む統制設計を、Vertex側の企業統制機能と組み合わせて進める価値が上がります（後述）。²⁸

ポートフォリオ管理（棚卸し・分類・費用対効果）

3 Proで詰まりやすい点

大量の特許/案件を、技術領域・製品ライン・競合・維持費・引用関係で俯瞰するには、バッチ処理と構造化出力が必要です。²⁹

3.1 Proで現実的になること

単価は同等でも、(a) medium思考で大量分類の過剰推論を抑える、(b) Batch APIで単価を半分に寄せる、(c) 構造化出力で“IPMSにそのまま入る形”にする、という最適化により、「**人が作っていた棚卸し資料**」から「**システムが常時更新する棚卸し**」へ移行しやすくなります。³⁰

推奨アーキテクチャと構造化出力サンプル

知財エージェントの基本設計

Gemini APIは、関数呼び出し（Function calling）で外部ツール/DBに接続し、自然言語と実世界のデータ・アクションの橋渡しを行えます。³¹

また、構造化出力（JSON Schema）により、抽出結果を“型付き”にできます。³²

この2つを組み合わせると、知財の典型ワーク（抽出→照合→記録→レビュー）を、**エージェント・パイプライン**として実装できます。

ただし、ツールには組合せ制約があります。たとえばファイル検索（File Search）は他ツール（検索グラウンディングやURLコンテキスト等）と併用できないと明記されています。³³

また、（少なくとも現時点のガイドでは）組込みツールとFunction callingの併用が未サポートとされており、ツール設計は**多段呼び出し（オーケストレーション）**前提が安全です。³⁴

一方で、構造化出力自体は、検索グラウンディングやURLコンテキスト等の組込みツールと併用できるとされています。³⁵

このため実装では「同一呼び出し内で全部やる」よりも、「段階ごとにツールを切り替え、各段階の出力をJSONで受け渡す」設計が堅牢です。³⁶

推奨ワークフロー（例：先行技術調査→クレームチャート）

flowchart TD

A[入力: 発明要約/仕様/クレーム案] --> B[段階1: 検索戦略生成 (thinking=medium)]

B --> C{検索手段}

C --> |社内/商用DB| D[Function calling: PatentSearch()]

C --> |Web一般| E[Search grounding + URL context]

D --> F[段階2: 上位候補の要点抽出 (structured output)]

E --> F

F --> G[段階3: クレーム要件分解→対比表生成 (thinking=high)]

G --> H[段階4: QA/根拠検証(引用箇所の再チェック)]

H --> I[出力: クレームチャートJSON/CSV + レビュー用HTML]

（この設計が必要になる背景：調査は時間制約があり、スクリーニングと精査が分かれる）³⁷

構造化出力のサンプル (JSON Schema)

以下は「クレーム要件分解と対比表」の出力例です。Geminiの構造化出力は、JSON Schemaに準拠するJSONを生成するよう設定できます。 [32](#)

```
{
  "title": "ClaimChart",
  "type": "object",
  "required": ["matter_id", "claim_id", "limitations", "overall_assessment"],
  "properties": {
    "matter_id": {"type": "string", "description": "案件ID (IPMSのキー) "},
    "claim_id": {"type": "string", "description": "対象クレーム番号 (例: 'Claim 1') "},
    "limitations": {
      "type": "array",
      "description": "要件分解 (クレームの要素) ",
      "items": {
        "type": "object",
        "required": ["limitation_id", "limitation_text", "evidence"],
        "properties": {
          "limitation_id": {"type": "string", "description": "要件ID (例: L1, L2...) "},
          "limitation_text": {"type": "string", "description": "要件の日本語化 (必要なら原文も含む) "},
          "evidence": {
            "type": "array",
            "description": "候補文献ごとの根拠",
            "items": {
              "type": "object",
              "required": ["doc_id", "support_level", "citation"],
              "properties": {
                "doc_id": {"type": "string", "description": "公報/論文/URLの識別子"},
                "support_level": {
                  "type": "string",
                  "description": "根拠の強さ",
                  "enum": ["explicit", "implicit", "not_found", "uncertain"]
                },
                "citation": {
                  "type": "string",
                  "description": "根拠箇所 (段落番号、図表番号、ページ等)。後で人が検証できる粒度。"
                },
                "notes": {"type": ["string", "null"], "description": "補足 (翻訳揺れ、用語同定の仮説等) "}
              }
            }
          }
        }
      }
    },
    "overall_assessment": {
      "type": "object",
      "required": ["novelty_risk", "inventive_step_risk", "top_docs"],
      "properties": {
        "novelty_risk": {"type": "string", "enum": ["low", "medium", "high"]},

```

```

    "inventive_step_risk": {"type": "string", "enum": ["low", "medium", "high"]},
    "top_docs": {"type": "array", "items": {"type": "string"}},
    "assumptions": {"type": "array", "items": {"type": "string"}}
  }
}
}
}

```

新規性の判断は「クレームの全要素が先行技術に明示/黙示されるか」が核であり、要素分解と根拠紐づけを構造化する価値が高いです。³⁸

CSV例（FTOリスクマトリクス）

FTOは、第三者の保護されたIPを無断で実施することで法的責任（侵害）が生じ得るかを見ます。WIPOのツール説明では、FTOは計画通りの利用が法的責任なく行えるか、無断利用が侵害となること、そしてクレームが権利範囲を定義することが説明されています。²²

```

matter_id,jurisdiction,patent_family_id,doc_id,status,claim_scope_match,evidence_citation,risk_level,recommended_
M-2026-001,US,FAM-123,US1234567,granted,high,"Claim 1 / col.3 l.10-35",high,"Design-around or
License inquiry"
M-2026-001,EP,FAM-123,EP2345678,lapsed,medium,"[0032]-[0040]",low,"Monitor only (inactive
right)"
M-2026-001,JP,FAM-987,JP2020-111111,pending,uncertain,"p.12 Fig.2 + para[0045]",medium,"Watch
+ continue search"

```

組織運用への影響

体制・役割の変化（“作業”から“統制と判断”へ）

3.1 Proは、複雑問題への推論強化と、検索・閲覧・ツールを伴うエージェント性能の改善が示されています。³⁹

その結果、知財組織の仕事は、次の方向にシフトしやすくなります（ここは将来予測であり、各社の成熟度に依存）。

- **パラリーガル/アナリストの定型作業（要約・抜粋・対比表の素案）**は自動化比率が上がり、代わりに「根拠監査」「例外処理」「テンプレ統制」の比重が増える。根拠の手元検証が可能な粒度での抽出は、文書理解+構造化出力が前提になります。⁴⁰
- 弁理士/知財責任者は、ドラフト作成よりも「論点設計」「最終判断」「リスク説明（経営向け）」に時間を再配分しやすい（APEX-Agents/MCP Atlasの改善は“案件完遂”方向の改善根拠）。³

KPIの再設計（例）

従来の「作成物の量」ではなく、**意思決定品質とリードタイム**に寄せたKPIが整合的です。

- 先行技術調査：上位N件の**根拠付き要件充足率**、再調査率、スクリーニング時間。調査は時間制約があるため、スクリーニング品質がKPIになり得ます。²⁴
- OA応答：ドラフト初版までの時間、レビュー差分（修正量）、却下（手戻り）率。

- FTO：国別の“アクティブ権利”同定率、リスクマトリクス更新頻度、設計変更の早期発見率（WIPO資料は早期段階での利用と更新を推奨）。⁴¹

ガバナンス・データセキュリティ（導入可否を左右する部分）

企業導入では「モデル性能」よりも「データ統制」がボトルネックになりがちです。ポイントだけ事実ベースで整理します。

- **学習制限（Training restriction）**：Vertex側では、サービス規約で「顧客の事前の許可/指示なく、顧客データを学習/微調整に使わない」旨が説明されています。⁴²
- **ゼロデータ保持（ZDR）の現実条件**：Developer API側のZDR説明では、(a)Paid Servicesではプロンプト/レスポンスを製品改善に使わない、(b)ただし特定機能（検索グラウンディング等）では30日保持があり無効化できない等、具体的な条件が列挙されています。⁴³
- **検索グラウンディング利用時の保持**：Vertex側のZDRガイドでも、検索グラウンディングは30日保持があり、無効化できない（ゼロ保持が必要なら別手段推奨）とされています。⁴²
- **ログの保持と共有**：Developer APIのログはデフォルト55日で失効し、データセット化すると保持期限が固定されない一方、共有すると“Unpaid Services”の条件でモデル改善に使われ得るため、機密情報を含めない注意が明記されています。⁴⁴
- **インメモリキャッシュ**：Vertex側では、低遅延のためのインメモリキャッシュ（24時間TTL、プロジェクト分離）があり、プロジェクト単位で無効化できるとされています。⁴²

この一連は、知財（契約書・未公開発明・共同研究情報）を扱う場合の「利用可/不可」を決める要件になります。

加えて、企業向け生成AIは暗号化・鍵管理・VPCサービスコントロール・データレジデンシー等のセキュリティ制御が重要、という観点が公式ブログでも述べられています。⁴⁵

導入ロードマップ

プレビュー版は、モデルの更新やレート制限の変化、少なくとも2週間前通知での廃止可能性があるため、初期段階は「置換」より「並走」が安全です。⁴

優先パイロット（推奨順）

1. **契約レビュー（機密度高だが社内文書完結）**：構造化出力で条項抽出→リスクフラグ→修正案。まずはZDR/ログ/保持設計を固める目的。²⁷
2. **発明抽出（社内一次情報→出願候補の提案）**：文書理解+構造化抽出で“発明の種”の拾い漏れを減らす。⁴⁶
3. **先行技術調査のスクリーニング（外部DB/社内DB活用）**：mediumで大量スクリーニング→highで上位精査。検索・閲覧が絡むため、ベンチマーク改善の恩恵が出やすい。¹⁷
4. **OA応答支援（ドラフト+根拠パッケージ）**：最終判断は人。多段ワークフロー性能が効く。⁴⁷
5. **FTO/無効調査（最も高リスク領域）**：最後に。FTOは権利状態・地域性・クレーム同定が難しく、誤りのコストが高い。²²

タイムライン例（12週間で“統制→価値→拡張”）

```
gantt
title 知財向け導入タイムライン（例）
dateFormat YYYY-MM-DD
```

section 統制設計

データ分類・ZDR方針・ログ/保持設計 :a1, 2026-02-24, 14d

section パイロット

契約レビューPoC (構造化抽出+レビューUI) :a2, after a1, 14d

発明抽出PoC (medium→high二段階) :a3, after a2, 14d

section 本格化

先行技術スクリーニング (社内DBツール連携) :a4, after a3, 21d

OA応答支援 (根拠パッケージ化) :a5, after a4, 21d

(プレビュー運用の注意：プレビューはレート制限が厳しめになり得る/廃止通知があり得る) 48

必要データ・必要ツール (最低限)

- ・社内文書ストア (契約、発明資料、過去OA、過去出願の拒絶/応答履歴)
- ・IPMS連携 (案件ID、期限、担当、ステータス、成果物格納)
- ・検索手段：社内/商用特許DBのAPI (Function callingでツール化) 31
- ・構造化出力のスキーマ (IPMSに直結するJSON) 32
- ・(必要に応じて) 検索グラウンディング/URLコンテキスト：最新情報やWeb根拠の補完。ただし保持条件と課金を理解して使用。 49

失敗モードと緩和策

- ・幻覚 (根拠の捏造)：基盤モデルの一般的限界として明記。対策は“根拠箇所 (段落/ページ) を必須フィールド化し、人が検証できる形で出す”。 50
- ・遅延/タイムアウト：3 Proモデルカードで言及。対策はthinking levelの段階設計 (medium→high) と、長文はRAG/キャッシュ分割。 51
- ・ツール組合せ制約で詰まる：File Searchは他ツールと併用不可。設計は“段階分割”で回避。 33
- ・データ保持要件違反：検索グラウンディング等の保持条件によりZDRにならないケースがある。事前に“使ってよい機能”を決める。 52
- ・コストが読めない：出力価格にはthinkingトークンが含まれるため、思考深度と出力制御を設計する。 53

前提・未確定事項・限界

主要前提 (Assumptions)

- ・本レポートは「3.1 Proの改善=知財タスクそのものの正答率が必ず上がる」と断言せず、エージェント性能指標の改善 (BrowseComp/MCP Atlas/APEX-Agents等) と、思考深度制御・カスタムツール優先IDという機能差から、知財ワークフローの成功率が上がり得る、という推論に基づきます。 54
- ・価格は、Developer APIおよびVertexの公開価格表を前提に記載しています (将来変動し得る)。 55
- ・検索グラウンディングは、1プロンプトが複数検索クエリに展開され得て、その分課金され得ることを前提に、コスト設計が必要です。 56

未確定事項 (本レポート時点で公式に明記されない/企業契約で決まる)

- ・SLA、サポート範囲、補償 (インデムニティ) 条件は、契約形態・提供形態に依存しやすく、本レポートでは断定しません (プレビューは“as is / サポート限定”になり得る旨は明記)。 57
- ・gemini-3.1-pro-preview-customtools の“どの程度ツール優先が改善するか”の定量は、公式には提示されていないため、パイロットで検証が必要です。 58

限界 (Limitations)

- モデルの知識カットオフは2025年1月であり、最新の法改正・審査動向・競合の新出願などは、検索グラウンディング等で補完する必要があります。 ⁵⁹
- 1M文脈を投入しても長文性能が自動で最大化するわけではなく（MRCR 1Mが同値）、RAG/分割/キャッシュの設計が成功要因になります。 ⁶⁰
- 知財判断（侵害・有効性・補正方針等）は、最終的に人の責任領域であり、モデルは“根拠整理とドラフトの生産性”を押し上げる位置づけが安全です（基盤モデルの幻覚リスクはモデルカードでも言及）。 ⁶¹

¹ ⁸ ¹⁰ <https://ai.google.dev/gemini-api/docs/changelog>

<https://ai.google.dev/gemini-api/docs/changelog>

² ¹² ¹³ ¹⁷ ¹⁸ ²⁵ ³⁰ ³⁴ ⁵⁹ <https://ai.google.dev/gemini-api/docs/gemini-3>

<https://ai.google.dev/gemini-api/docs/gemini-3>

³ ⁵ ¹⁴ ²⁰ ²⁶ ³⁹ ⁴⁷ ⁵⁴ ⁶⁰ <https://deepmind.google/models/model-cards/gemini-3-1-pro/>

<https://deepmind.google/models/model-cards/gemini-3-1-pro/>

⁴ ⁴⁸ <https://ai.google.dev/gemini-api/docs/models>

<https://ai.google.dev/gemini-api/docs/models>

⁶ ⁵⁰ ⁵¹ ⁶¹ <https://deepmind.google/models/model-cards/gemini-3-pro>

<https://deepmind.google/models/model-cards/gemini-3-pro>

⁷ ⁵⁷ <https://docs.cloud.google.com/vertex-ai/generative-ai/docs/models/gemini/3-1-pro?hl=ja>

<https://docs.cloud.google.com/vertex-ai/generative-ai/docs/models/gemini/3-1-pro?hl=ja>

⁹ <https://ai.google.dev/api>

<https://ai.google.dev/api>

¹¹ ⁵⁸ <https://ai.google.dev/gemini-api/docs/models/gemini-3.1-pro-preview?hl=ja>

<https://ai.google.dev/gemini-api/docs/models/gemini-3.1-pro-preview?hl=ja>

¹⁵ ¹⁶ ²⁴ ³⁷ <https://www.jpo.go.jp/e/news/kokusai/developing/training/e-learning/document/2016PAS/m-all.pdf>

<https://www.jpo.go.jp/e/news/kokusai/developing/training/e-learning/document/2016PAS/m-all.pdf>

¹⁹ ³⁸ https://www.wipo.int/en/web/pct-system/texts/ispe/13_01_02

https://www.wipo.int/en/web/pct-system/texts/ispe/13_01_02

²¹ ²⁷ ²⁹ ³² ³⁵ <https://ai.google.dev/gemini-api/docs/structured-output>

<https://ai.google.dev/gemini-api/docs/structured-output>

²² ²³ ⁴¹ <https://www.wipo.int/documents/d/tisc/docs-en-tisc-toolkit-freedom-to-operate-description.pdf>

<https://www.wipo.int/documents/d/tisc/docs-en-tisc-toolkit-freedom-to-operate-description.pdf>

²⁸ ⁴² <https://docs.cloud.google.com/vertex-ai/docs/generative-ai/data-governance>

<https://docs.cloud.google.com/vertex-ai/docs/generative-ai/data-governance>

³¹ <https://ai.google.dev/gemini-api/docs/function-calling>

<https://ai.google.dev/gemini-api/docs/function-calling>

³³ ³⁶ <https://ai.google.dev/gemini-api/docs/file-search>

<https://ai.google.dev/gemini-api/docs/file-search>

40 46 <https://ai.google.dev/gemini-api/docs/document-processing>

<https://ai.google.dev/gemini-api/docs/document-processing>

43 49 52 <https://ai.google.dev/gemini-api/docs/zdr>

<https://ai.google.dev/gemini-api/docs/zdr>

44 <https://ai.google.dev/gemini-api/docs/logs-policy>

<https://ai.google.dev/gemini-api/docs/logs-policy>

45 <https://cloud.google.com/transform/spotlighting-shadow-ai-how-to-protect-against-risky-ai-practices>

<https://cloud.google.com/transform/spotlighting-shadow-ai-how-to-protect-against-risky-ai-practices>

53 55 <https://ai.google.dev/gemini-api/docs/pricing>

<https://ai.google.dev/gemini-api/docs/pricing>

56 <https://ai.google.dev/gemini-api/docs/google-search>

<https://ai.google.dev/gemini-api/docs/google-search>