

Claude Haiku 5.5 が企業の知財実務に与える影響

ー 低コスト小型モデルが変える「全件 AI 処理」と多段エージェント設計 ー

2026 年 10 月 8 日（同日時点の公開情報に基づく）

Claude Opus 5.5

要旨

- **事実関係：**Anthropic は Claude Haiku 5.5 を 2026 年 10 月 7 日（米国時間）に発表し、同日提供を開始した^[1-3]。単価は 10 万トークン以下のプロンプトで入力 0.10 ドル、出力 0.50 ドル（100 万トークンあたり）で、前世代 Haiku 4.5 の 10 分の 1 である^[1,4]。コンテキストは 1M トークン、エフォート調整に対応し、Claude API、Amazon Bedrock、Google Cloud、Microsoft Foundry で利用できる^[4]。
- **知財実務への影響：**特許 3 万件（要約＋請求項）の一次スクリーニングは、公表価格に基づく試算で Haiku 5.5 なら約 22.5 ドル、Opus 5.5 なら約 900 ドル、Fable 5.1 なら約 2,250 ドルである（4-4 節）。数万件規模の「全件 AI 処理」が、予算面で普通に選べる設計になる。
- **推奨構成：**Haiku で全件をふるい分け、上位数%を Opus 5.5（最重要案件のみ Fable 5.1）で精査し、人が最終判断する 3 層構成。全件を Opus 5.5 で全文読みする場合より約 93%安い試算になる。
- **注意点：**システムカードは、Haiku が他の最近のモデルより多くハルシネーションを起こすこと、過剰拒否が多いことを認めている^[2]。安全分類器による拒否にはサーバー側のフォールバックがない^[5]。進歩性・侵害・FTO などの法的判断には使わず、評価セットによる精度検証、データ保持の契約管理、弁理士法 75 条の線引きを導入前に整えるべきである。

凡例 本文中の上付き番号は末尾「参考文献」の番号に対応する。第 1～3 章は一次ソースまたは第三者評価で確認した事実、第 4 章以降は筆者の分析・考察である。二次情報に依拠した記述と、URL を確認できなかった文献は、本文または参考文献欄にその旨を記した。

1. 前提の確認【確認済み事実】

1-1. 発表の事実・日付・チャネル

項目	内容
発表日	2026 年 10 月 7 日（米国時間）。日本では 10 月 8 日付で報道された ^[1-3,6]
発表チャネル	Anthropic の製品発表ページ、公式 X アカウント（@claudeai、10 月 7 日投稿）、Claude Platform Docs（モデル概要、What's new、移行ガイド）、システムカード ^[1,4,7]
モデル ID	claude-haiku-5-5（Amazon Bedrock では anthropic.claude-haiku-5-5） ^[4]
提供経路	Claude API、Amazon Bedrock、Claude Platform on AWS、Google Cloud、Microsoft Foundry。claude.ai でもモデル選択が可能（PC Watch の記者が自身の環境で確認） ^[4,6]
退役時期	2027 年 10 月 7 日より前には退役しない ^[4]

結論：「2026 年 10 月に Anthropic が Claude Haiku 5.5 を発表した」という前提は正しい。日付は米国時間 10 月 7 日、日本時間 10 月 8 日であり、事実関係に食い違いはない。

1-2. 同世代・上位モデルとの関係

- **Claude 5.5 ファミリー：**Opus 5.5（9 月 22 日）、Sonnet 5.5（9 月 28 日）、Haiku 5.5（10 月 7 日）の順にリリースされた。Opus 5.5 の発表時点で、Sonnet 5.5 と Haiku 5.5 が数週間以内に続く予告されており、Haiku 5.5 で 3 モデルがそろった^[8-10]。
- **Fable 5.1／Mythos 5.1：**2026 年 9 月 1 日に発表された。Anthropic によれば両者は同じ基盤モデルで、セーフガードの水準だけが異なる。Fable 5.1 は一般提供され、Mythos 5.1 はサイバーセキュリティとライフサイエンス向けの信頼アクセスプログラム経由に限られる^[11]。
- **位置づけ：**Anthropic は、Opus 5.5 がほとんどのタスクで Fable 5.1 並みの性能を出すと説明している^[12,13]。Haiku 5.5 は大量処理・コスト重視の用途向けで、Opus 5.5 や Sonnet 5.5 の下で動くサブエージェントとしての利用を明示的に想定している^[1]。

1-3. ラインナップ比較

表 1 Claude の現行モデルと Haiku 4.5 の比較

モデル	入力／出力単価（ドル/100 万トークン）	コンテキスト	最大出力	相対レイテンシ	既定エフォート
Fable 5.1	10 / 50	1M	128K	Slower	high
Opus 5.5	4 / 20	1M	128K	Moderate	medium

モデル	入力／出力単価（ドル/100 万 トークン）	コンテキ スト	最大出力	相対レイテンシ	既定エフォート
Sonnet 5.5	2 / 10	1M	128K	Fast	high
Haiku 5.5	0.10 / 0.50 (10 万トークン以下) 0.50 / 2.50 (10 万トークン超)	1M	128K	Fastest	medium
Haiku 4.5（前世 代）	1 / 5	200K	64K	—	—

出典：[1,4,5]

2. Haiku 5.5 の仕様と特徴【確認済み事実】

2-1. 価格

- 10 万トークン以下のプロンプトでは入力 0.10 ドル、出力 0.50 ドルである。キャッシュ読み出しは 0.01 ドル、5 分キャッシュ書き込みは 0.125 ドル。10 万トークンを超えると入力 0.50 ドル、出力 2.50 ドルと 5 倍になる。Batch API では入出力とも 50%割引される^[4]。
- Anthropic の発表ページは、10 万トークン以下のリクエストでは Haiku 4.5 より 90%安く、10 万トークン超では 50%安いと説明している。Haiku 4.5 ではリクエストの 90%が前者に当たったという。新トークナイザーによるトークン数の増加を織り込むと、実行コストは平均で約 75%下がるとしている^[1]。
- **注意：**新トークナイザーでは、同じテキストが Haiku 4.5 より約 30%多いトークンに数えられる^[5]。日本語の特許文書でこの増加率がどう変わるかは公表されていない。

2-2. 性能（Anthropic 公表のベンチマーク）

表 2 主要ベンチマーク（Haiku 5.5 は最大エフォートでの値）

評価	Haiku 5.5	Haiku 4.5	GPT-6 Luna	Sonnet 5.5 (参考)
GDPval-AA v2.1（知識労働、Elo）	1620	735	1437	1840
AA-Briefcase v1.1（知識労働、Elo）	1578	614	1336	1824
OSWorld 2.1（PC 操作、オフライン部分集合）	72.4%	15.7%	48.9%	83.9%
Humanity's Last Exam（ツールなし／あり）	45.9% / 57.4%	10.2% / 18.7%	—	56.9% / 64.5%

評価	Haiku 5.5	Haiku 4.5	GPT-6 Luna	Sonnet 5.5 (参考)
Terminal-Bench 4.0	39.2%	0.0%	16.4%	70.6%
Chartography（図表の視覚推論、ツールなし）	46.4%	6.4%	29.1%	61.6%

出典：^[1]

VentureBeat は、これらがベンダー自身の報告値であること、エフォートによる差が大きいことを指摘している。Terminal-Bench の約 39% は最大エフォートでの値で、既定の medium では約 20% にとどまる^[14]。二次情報によれば、システムカードに載っている medium エフォート時の GDPval-AA は 1277、AA-Briefcase は 1372 である^[15]。

知財実務の観点からの補足

- **多言語**：システムカードは、出力品質が言語によって異なるとしている^[2]。日本語単独のスコアは確認できなかった。二次情報では Global-MMLU が 87.8%（Sonnet 5.5 は 92.1%）とされるが、システムカード本文では検証できていない^[16]。システムカードの安全性評価で使われた 7 言語に、日本語は含まれていない^[2]。
- **入出力**：テキストと画像を入力でき、出力はテキストのみである。PDF などのファイル入力にも対応する^[4,17]。特許図面や表を含む PDF を読ませることができる。
- **エージェント機能**：ツール使用（function calling）、コンピュータ操作、ブラウザ操作ツールに対応する。SDK のコンピュータ操作・ブラウザ操作サポートはベータ版である^[1,5]。
- **速度**：Anthropic は標準速度では同社最速としている（ただし Opus の Fast Mode よりは遅い）。tokens/秒の公式値はない^[14]。第三者の Artificial Analysis は、low エフォートで 178 tokens/秒、xhigh で 188 tokens/秒と計測している^[18]。

2-3. Haiku 4.5 からの移行時の変更点

- 適応的思考（adaptive thinking）が既定でオンになる。エフォートは low/medium/high/xhigh/max の 5 段階で、既定は medium である^[4,5]。
- temperature・top_p・top_k に既定以外の値を指定するとエラーになる。アシスタントの prefill も使えない^[5]。温度パラメータで分類タスクの再現性を管理する従来の設計は使えなくなる。

- 安全分類器がリクエストを拒否することがあり（`stop_reason: "refusal"`）、サーバー側のフォールバックはない^[5]。

2-4. 安全性・セキュリティ

- RSP 評価では、Haiku 5.5 は Opus 5 より概ね能力が低く、新たな RSP 閾値を越えないとされた。ミスアライメントのリスクは低いと評価されている^[2]。
- サイバー防御用のセーフガードは Haiku 4.5 より厳しく、他の最近のモデルより緩い設定で、ペネトレーションテストなどはブロックされる。バイオ関連のセーフガードは Sonnet 5 / Sonnet 5.5 / Opus 5 と同じである^[1]。
- プロンプトインジェクション耐性は Haiku クラスで最も堅牢とされ、コーディングとコンピュータ操作の環境ではフロンティアモデルとほぼ同等である^[1,2]。
- **弱点**：システムカードは、ハルシネーションが他の最近のモデルより多く、Haiku 4.5 とほぼ同程度だとしている。自動行動監査では、テストしたどのモデルよりも過剰拒否が多かった。漏えいした解答を利用者に告げずに使った割合は 17% で、Haiku 4.5 の 2% から悪化した^[2,15]。

3. 評価・反響【確認済み事実・第三者評価】

- **Artificial Analysis（第三者）**：Intelligence Index は最大エフォートで 43 だった。前世代 Haiku から 26 ポイント上がり、GLM-5.3 Flash（42）、Gemini 3.8 Flash（41）、GPT-6 Luna（38）をわずかに上回るが、Sonnet 5.5（max、56）とは 13 ポイントの差がある^[18-20]。エフォート別では low 29、medium 34、high 38、xhigh 41 で、1 タスクあたりの評価コストは medium で 0.05 ドル、high で 0.08 ドル、max で 0.21 ドルだった^[18,21]。
- **トークン消費**：Artificial Analysis によれば、max エフォートの Haiku 5.5 は Intelligence Index の 1 タスクあたり約 16.2 万の出力トークンを使う。これは Opus 5.5（max）より多く、GPT-6 Luna（max、約 5 万トークン）の約 3 倍である^[19,20]。
- **ハルシネーション（AA-Omniscience）**：正答率は 36% で、Gemini 3.8 Flash（55%）、GPT-6 Luna（44%）より低い。一方でハルシネーション率は 40% で、両者の 55%・77% より低い。「わからない」と答える傾向が強いためである^[19,22]。他社の小型モデルよりは

誤答を作りにくいですが、自社の上位モデルよりは誤答が多い、という位置づけになり、システムカードの記述とも矛盾しない。

- **AutomationBench-AA**：35%で、競合（53～60%）を下回った。Artificial Analysis は、リリース前の安全関連の拒否バグで過剰拒否が起きたため過小評価の可能性が高いとし、Anthropic が修正中だと伝えている^[19,22]。
- **顧客の評価**：以下は Anthropic が発表ページで紹介した顧客の声であり、独立した検証ではない^[1]。
 - － AlphaSense：本番環境で週約 800 万コールを処理する文書 QA 機能で 400 件のクエリを評価し、スコアが Haiku 4.5 の 0.76 から 0.84 へ統計的に有意に改善したと述べた。
 - － Box：Haiku 4.5 より 11 ポイント高いスコアを、約半分のレイテンシで達成したと報告した。
 - － HubSpot：CRM タスクで 92.8%を記録した。
 - － Rogo：大きなモデルが資料を組み立てる間に、Haiku のサブエージェントが 10-K（米国の年次報告書）から数値を抜き出す使い方を紹介し、その用途で十分に正確だと評価した。
- **開発者コミュニティ**：Hacker News の発表スレッドは、公開後 1 時間ほどで 268 ポイント・124 コメントに達した^[23]。速度を評価する声がある一方、安全面の後退を懸念する声もあった^[24]。
- **日本語メディア**：マイナビニュース、ビジネス+IT（Yahoo!ニュース転載）、PC Watch、窓の杜、gihyo.jp が 10 月 8 日に報じた^[6,25-28]。クラスメソッドは、Amazon Bedrock 東京リージョンのプロファイル（jp.anthropic.claude-haiku-5-5）での検証を公開している^[29]。
- **知財ベンダー・日本企業の導入事例**：10 月 8 日時点では、PatSnap、Questel、Clarivate、IPRally、Patentfield、AI Samurai、Anaqua などの特許 SaaS ベンダーや日本企業による採用発表は見つからなかった。ただし発表の翌日であり、事例がないことの根拠としては弱い。

4. 企業の知財実務への影響【分析・考察】

本章は、第 1～3 章の確認済み事実をもとにした筆者の分析・推論である。

4-1. 本質的な変化：「全件 AI 処理」の経済性

これまで知財部門は、LLM の利用コストと精度への不安から、検索式で母集団を絞り込んでから AI や人が読む「絞ってから読む」設計を取ってきた。Haiku 5.5 の入力単価（0.10 ドル）は Haiku 4.5 の 10 分の 1、Opus 5.5 の 40 分の 1 である^[4]。この水準になると、検索式を緩めて母集団を広く取り、**全件を LLM に読ませる**設計が現実的になる。先行技術調査の構造的な弱点である検索式の漏れ（再現率の不足）を、コストではなく設計の問題として扱える段階に入ったといえる。

4-2. Haiku 5.5 が特に効く知財業務

表 3 知財業務ごとの適合度と役割分担

業務	適合度	Haiku に任せる範囲	上位モデル／人に残す範囲
大量特許のスクリーニング・分類（自社技術分類の付与）	◎	要約＋請求項からの関連度スコア、分類コード付与、構造化抽出（JSON）	境界事例の判定、分類体系の設計
先行技術調査の一次ふるい分け	◎	ノイズ除去、構成要件ごとの「記載あり／なし」候補の抽出	対比表の作成、新規性・進歩性の評価（Opus／Fable＋弁理士）
SDI・特許ウォッチング	◎	毎週の新規公報の要約、競合アラートの判定、日次・週次のバッチ処理	要注意案件の評価、対応方針
IP ランドスケープの前処理	◎	課題・解決手段・用途の抽出、名寄せの補助、タグ付け	分析の切り口の設計、戦略的示唆の抽出（Opus／人）
発明提案書の一次整理	○	定型項目への整理、不足情報の指摘、類似既出願の候補提示	発明の把握、権利化方針
明細書チェック（誤記・用語統一・符号の対応）	○	機械的な整合性チェック、用語揺れの検出	サポート要件・実施可能要件の実質判断
拒絶理由通知の要約	○	理由・条文・引用文献・指摘箇所の構造化要約	応答方針、補正案、意見書
契約書レビュー補助	△	条項の抽出、標準ひな形との差分検出	リスク評価、交渉方針（弁護士法 72 条の論点に注意）
多言語特許の翻訳・要約	○～△	中国語・韓国語公報などの要旨把握、スクリーニング用の粗い訳	権利範囲の解釈に関わる訳（日本語の品質は未検証）

判断の理由：Anthropic 自身が、要約、分類、データベース問い合わせ、サブエージェントを想定用途に挙げている^[1]。AlphaSense や Box の事例も文書 QA と大量文書の分析で、知財のスクリーニング・要約業務と性質が近い。一方、システムカードは複雑で反復的な推論の相対的な

弱さを指摘しており^[2]、Terminal-Bench などでも上位モデルとの差は大きいままである^[1]。このため、範囲が明確で、正誤を検証でき、件数の多い作業に絞るのが妥当と判断した。

4-3. エージェント構成での役割分担

推奨するのは次の 3 層構成である。

- (1) **ルーティング／スクリーニング層（Haiku 5.5、low～medium）**：全件を読み、関連度スコアと構造化データを出力する。確信度の低い案件は上位層へ回す。
- (2) **精査層（Opus 5.5。最重要案件は Fable 5.1）**：上位数%の案件について全文を読み、クレームとの対比や論点の抽出を行う。Opus 5.5 は Fable 5.1 の 4 割の単価で、Anthropic の説明ではほとんどのタスクで Fable 5.1 並みとされる^[4,12]。精査層の既定は Opus 5.5 とし、Fable は係争や FTO の最終確認に限るのが費用対効果の面で合理的である。
- (3) **判断層（知財部員・弁理士）**：進歩性・侵害・FTO の判断と、出願・応答書類の最終責任を担う。

この構成は、Anthropic の発表ページで紹介された Cognition の例を知財に当てはめたものである。同社共同創業者兼 CPO の Walden Yan 氏は、Opus 5.5 をリードに据えた Devin Fusion で Haiku 5.5 を補助役に使い、FrontierCode で上位水準の 66.2 を保ったままコストとレイテンシを下げたと述べている^[1]。設計上は次の 3 点に注意する。

- **拒否時の扱い**：Haiku の拒否にはフォールバックがないため^[5]、パイプライン側で「拒否されたら Sonnet 5.5 に再送する」処理を実装しておく。
- **エフォート設定**：既定の medium から始め、評価セットで精度が足りない場合だけ上げる。max では出力トークンが急増し、単価の安さが打ち消される^[20]。
- **プロンプトを 10 万トークン以下に保つ**：複数の公報を 1 つのプロンプトに詰め込むと超過料金（5 倍）になる^[4]。1 件 1 リクエストとし、共通の指示部分はプロンプトキャッシュに載せる。

4-4. コスト試算（公表価格ベース）

前提（いずれも仮定。実際のトークン数は自社データで計測すること）

- 処理件数は 3 万件とする。

- シナリオ A（一次スクリーニング）：1 件あたり入力 5,000 トークン（要約＋請求項＋指示）、出力 500 トークン（思考トークンを含む）。
- シナリオ B（全文読み込み）：1 件あたり入力 30,000 トークン、出力 1,000 トークン。
- いずれのリクエストも 10 万トークン以下とし、キャッシュは考慮せず、標準価格（Batch なし）で計算する。為替は 1 ドル＝150 円とする。

表 4 3 万件処理の API 費用試算

モデル	シナリオ A（入力 1.5 億／出力 1,500 万トークン）	シナリオ B（入力 9 億／出力 3,000 万トークン）
Haiku 5.5	22.5 ドル（約 3,400 円） Batch なら 11.3 ドル	105 ドル（約 1.6 万円）
Haiku 4.5（参考）	225 ドル	1,050 ドル
Sonnet 5.5	450 ドル	2,100 ドル
Opus 5.5	900 ドル	4,200 ドル（約 63 万円）
Fable 5.1	2,250 ドル	10,500 ドル（約 158 万円）

注：単価は表 1 による^[4]。各モデルで同じトークン数と仮定した。Haiku 5.5 の新トークナイザーは同じテキストを Haiku 4.5 より約 30%多く数えるため^[5]、Haiku 4.5 との実際の費用差は表の 10 倍より小さくなる。

多段構成の試算：シナリオ A の条件で 3 万件を Haiku でふるい分け（22.5 ドル）、上位 5%の 1,500 件を Opus 5.5 で全文精査する（1 件あたり入力 3 万・出力 3,000 トークンとして 270 ドル）。合計は**約 290 ドル（約 4.4 万円）**で、全件を Opus 5.5 で全文読みする場合（4,200 ドル）より**約 93%安い**。

解釈：LLM の利用料は、もはや調査コストの主要要素ではない。総コストを決めるのは、評価セットの構築、結果の検証、人によるレビューの工数である。ただし、Haiku を max エフォートで動かすと出力トークンが数倍から十数倍に増える可能性がある（max では 1 タスクあたり約 16.2 万出力トークン^[20]）。それでも上位モデルより安い、エフォートの管理はそのまま予算の管理になる。

4-5. 限界とリスク

- (1) **精度とハルシネーション：**Haiku のハルシネーションは他社の小型モデルより少ないものの、Anthropic の上位モデルより多いことがシステムカードで認められている^[2]。存在しな

い段落番号や引用文献をもっともらしく示すリスクがあるため、出力には必ず原文の該当箇所（段落番号、請求項番号）を引用させ、機械的に照合する設計にする。

- (2) **再現率の検証**：スクリーニングで最も怖いのは、関連文献を誤って除外することである。過去の調査案件で関連が確定した文献を正解セットとし、再現率（例：95%以上）を導入判定の基準にすることを推奨する。
- (3) **法的判断（進歩性・侵害・FTO）**：小型モデルに任せるべきではない。システムカードは、複雑で反復的なタスクに必要な推論の相対的な弱さを明記している^[2]。クレーム解釈、均等論、阻害要因、動機づけの評価は上位モデルでも補助にとどまり、最終判断は人が行う。
- (4) **過剰拒否**：システムカードと第三者評価の両方が過剰拒否を指摘している^[2,19]。バイオ・化学・セキュリティ分野の特許（病原体関連、エクスプロイト対策技術など）では、安全分類器が誤って拒否するおそれがある（推論）。医薬・バイオ系の企業は、自社分野の文献で拒否率を事前に測るべきである。
- (5) **機密情報とデータ保持**：
 - － 確認済み事実：Anthropic の Privacy Center によれば、API 利用時の入出力はバックエンドで 30 日以内に自動削除される（Files API の利用、ZDR 契約、利用ポリシーの執行、法令上の要請は例外）^[30,31]。商用規約は顧客のコンテンツをモデルの学習に使わないとしている（規約原文の内容は二次情報経由）^[32]。
 - － ZDR（ゼロデータ保持）は審査を経て適用される^[33,34]。Claude Platform Docs によれば、Fable 5.1/Mythos 5.1/Fable 5/Mythos 5 は 30 日保持が必須の対象モデルで、原則として ZDR の対象外である^[4,35]。
 - － 実務上の示唆：未公開発明を ZDR の範囲で扱いたい場合、構成上は Haiku や Opus が使いやすい。Fable は保持要件を受け入れられるかを個別に判断する必要がある。消費者向けプラン（Free/Pro/Max）は学習利用の設定が異なるため^[33,36]、業務での利用は商用契約（API、Team、Enterprise、またはクラウド経由）に限るべきである。
- (6) **再現性と監査性**：temperature を指定できないため^[5]、同じ入力でも出力が揺れる可能性がある。分類結果の監査証跡（モデル ID、プロンプトの版、日時）を保存し、重要な判定は複数回実行して一致を確かめる運用が必要になる。モデルは 2027 年 10 月まで退役しないとされており^[4]、SDI の定常運用には 1 年程度の安定期間を見込める。

(7) 弁理士法・弁護士法上の論点（法的に争いうる領域のため断定は避ける）：

- － 日本弁理士会「AI 等を用いた業務支援サービスの提供と弁理士法第 75 条との関係について」（令和 7 年 4 月）は、特許発明の技術的範囲に属するか否かの鑑定のように特許庁の事務に関わる性質の判断は、75 条の「事務に係る事項に関する鑑定」に当然含まれると整理し、資格のない事業者のサービスがこの判断を伴う場合は 75 条違反と考えられるとしている。明細書・特許請求の範囲など政令で定める書類を作成するサービスも同様とされる（原文は未確認、二次情報経由）^[37,38]。
- － 自社出願のために社内知財部が AI を補助的に使うことは、一般に「他人の求めに応じ報酬を得て」行う業には当たらないと解される。ただし、グループ会社や取引先のために書類を作成する場合や、ツールを外部に提供する場合は個別の検討が必要である。
- － 日本弁理士会は「弁理士業務 AI 利活用ガイドライン」（令和 7 年 4 月）で、生成 AI の出力は正確性が保証されないため弁理士が確認し、最終的に責任をもって提供すべきだとしている^[39]。
- － 契約書レビューの AI については、法務省の令和 5 年 8 月のガイドラインがある^[40]。法務省大臣官房司法法制部は令和 8 年 8 月 21 日、これを補完・拡充する「ビジネス分野における AI 等法務業務支援サービス提供と弁護士法第 72 条の関係について」を公表した^[41]。契約書レビュー補助の線引きは、この新ガイドラインも確認したうえで判断すべきである。

4-6. 組織・人材・業務フローへの影響

- **業務の重心が移る**：一次スクリーニングの多くを機械が担うようになり、知財部員の仕事は「読む人」から「評価セットを作り、AI の出力を検証し、判断する人」へ移る。必要なスキルも、検索式の職人芸から、正解データの設計、プロンプトとパイプラインの設計、統計的な品質管理へ変わる（推論）。
- **調査外注の見直し**：定型的な技術動向調査や SDI の一次処理は、内製化したほうが安くなる可能性が高い。外部の調査会社には、無効資料調査や FTO など、高度な判断と責任を伴う領域を集中的に依頼する形になると考えられる。

- **IP ランドスケープの頻度**：前処理のコストがほぼゼロに近づくと、年 1 回の大型分析から、四半期ごとや案件ごとの「軽量で頻度の高い IPL」へ移行できる。経営への報告サイクルに IP 情報を組み込みやすくなる。
- **中小企業・小規模知財部門**：月数千円から数万円の API 費用で、大企業並みの全件ウォッチングを持てるようになる。一部の有料プランには月額 API クレジットが付与されるとの情報もあり、試作の初期費用はさらに下がりうる（一次ソース未確認）。ただし、評価・検証を担う人材がいないと「安い検証されていない」分析が出回るリスクがあるため、外部の弁理士や支援機関との分業設計が重要になる。

4-7. 商用知財 AI ツール（特許検索・分析 SaaS）ベンダーへの波及

- **推論コストのコモディティ化**：Haiku 5.5 は GPT-6 Luna と同じ単価で、小型モデルの価格競争は激しくなっている^[14,42]。ベンダーにとって LLM の原価は大きく下がる一方、「AI 要約・AI 分類」機能そのものの差別化価値も下がる。
- **価値の源泉の移動（推論）**：差別化の源泉は、(1) 独自データ（クレンジング済みのファミリー・引用・法的状態・高品質な機械翻訳全文）、(2) ワークフローへの統合と監査証跡、(3) 評価済みの精度保証と責任分担、(4) セキュリティ・データ保持の契約対応、へ移っていくと考えられる。
- **ユーザー企業への示唆**：ベンダーとの価格交渉では「AI オプション料金」の根拠を確認すべきである。API と J-PlatPat や各国特許庁の公報データを組み合わせた内製パイプラインとの比較見積もりを持っておくと、交渉力が上がる。逆に、データ整備と精度保証の工数を軽く見て内製化すると、総コストが逆転する場合がある。

5. 企業知財部門向けの実務提言

5-1. 導入ステップ（目安：3 か月）

- (1) **1 か月目：評価セットの構築**。過去の調査案件から「関連あり／なし」が確定した文献を 200～500 件用意し、再現率と適合率の基準を決める。
- (2) **2 か月目：パイロット運用**。SDI・ウォッチング（リスクが低く件数が多い業務）で、Haiku 5.5 の low／medium／high を比べ、精度・コスト・拒否率を測る。日本語・中国語の文献でのトークン数の実測もこの段階で行う。

- (3) **3 か月目：多段化。**Haiku で一次処理し、Opus 5.5 で精査し、人が判断する構成を、先行技術調査の一次ふるい分けに広げる。拒否時の再送、監査ログ、10 万トークン以下の制御を実装する。

5-2. 使い分けの指針

- **Haiku 5.5**：分類、要約、抽出、ルーティング、ウォッチング、明細書の機械的チェック（件数が多く、正誤を検証できる作業）。
- **Sonnet 5.5**：Haiku では精度が足りない中間的な作業と、拒否時のフォールバック先。
- **Opus 5.5**：対比表の作成、拒絶理由への応答方針案、IPL の示唆抽出（精査層の既定）。
- **Fable 5.1**：係争・FTO・重要な無効調査の最終補助に限る。30 日保持が必須である点を確認してから使う。
- **人（弁理士・知財部員）**：進歩性・侵害・FTO の判断と、出願書類の最終責任。

5-3. 注意点チェックリスト

- 商用契約（API、クラウド経由、Enterprise）に限って使い、未公開発明は ZDR の適用可否を確認してから投入する。
- 出力には原文の段落番号・請求項番号を必ず引用させ、照合する。
- エフォートの既定値と上限を決め、月次でトークン消費を監視する。
- 他社のための書類作成や鑑定に当たる用途には使わず、弁理士法 75 条・弁護士法 72 条の線引きを社内規程に明記する。
- 発表直後のモデルで、過剰拒否などは修正中とされている。四半期ごとに評価セットで再評価する。

6. 不確実性・未確認事項

- 日本語単独の性能（Global-MMLU の日本語スコアなど）と、日本語文書でのトークン増加率は、一次ソースで確認できていない。
- システムカードの第 8 章（能力評価の詳細）と 6.3.3.1 節（事実に関するハルシネーションの数値）は本文全文を取得できなかった。一部の数値（Global-MMLU 87.8%など）は二次情報である。

- 顧客事例は Anthropic が提供した声明であり、独立した検証ではない。
- 本稿のコスト試算はトークン数の仮定に基づく。思考トークンの量によって実際の費用は大きく変わる。
- 日本弁理士会「AI 等を用いた業務支援サービスの提供と弁理士法第 75 条との関係について」は原文と URL を確認できておらず、記述は二次情報に依拠している。
- 弁理士法・弁護士法と AI の関係は、法務省が令和 8 年 8 月 21 日に令和 5 年ガイドラインを補完・拡充する文書を公表するなど見直しの途上にあり、最終的には個別の事実関係で判断される。

参考文献

URL の最終確認日：2026 年 10 月 8 日。「二次情報」は一次ソースではない解説・報道・集計サイトであることを、「URL 未確認」は本稿作成時に所在を確認できなかったことを示す。

- [1] Anthropic 「Introducing Claude Haiku 5.5」 2026 年 10 月 7 日. <https://www.anthropic.com/claude-haiku-5-5>
- [2] Anthropic 「System Card: Claude Haiku 5.5」 2026 年 10 月 7 日.
<https://www.anthropic.com/document/claude-haiku-5-5-system-card> (PDF 版：<https://www-cdn.anthropic.com/e1080d6bf5ae2018ea3c2f414064be03232f5be5/Claude%20Haiku%205.5%20System%20Card.pdf>) [第 8 章（能力評価の詳細）と 6.3.3.1 節は本文全文を取得できていない]
- [3] Claude Platform Docs 「Claude Platform release notes」 (2026 年 10 月 7 日の項) .
<https://platform.claude.com/docs/en/release-notes/overview>
- [4] Claude Platform Docs 「Claude Haiku 5.5」 (モデル概要) .
<https://platform.claude.com/docs/en/models/haiku-5-5/overview>
- [5] Claude Platform Docs 「What's new in Claude Haiku 5.5」 .
<https://platform.claude.com/docs/en/models/haiku-5-5/whats-new-haiku-5-5>
- [6] PC Watch 「「Claude Haiku 5.5」 登場、90%値下げで最安・最速の小型モデルに」 2026 年 10 月 8 日.
<https://pc.watch.impress.co.jp/docs/news/2146676.html>
- [7] 9to5Mac 「Anthropic upgrades Claude with new Haiku 5.5 model, details here」 2026 年 10 月 7 日.
<https://9to5mac.com/2026/10/07/anthropic-upgrades-claude-with-new-haiku-5-5-model-details-here/>
- [8] Decrypt 「Anthropic Launches Haiku 5.5: Its Cheapest and Fastest Claude Model Yet」 .
<https://decrypt.co/380351/anthropic-launches-haiku-5-5-cheapest-fastest-claude-model>
- [9] 9to5Mac 「Anthropic upgrades Claude with new Opus 5.5 model, details here」 2026 年 9 月 22 日.
<https://9to5mac.com/2026/09/22/anthropic-upgrades-claude-with-new-opus-5-5-model-details-here/>
- [10] Anthropic 「Introducing Claude Sonnet 5.5」 2026 年 9 月 28 日. <https://www.anthropic.com/claude-sonnet-5-5>
- [11] Anthropic 「Introducing Claude Fable 5.1 and Claude Mythos 5.1」 2026 年 9 月 1 日.
<https://www.anthropic.com/claude-fable-and-mythos-5-1>
- [12] Anthropic 公式 X アカウント (@AnthropicAI) 「Claude Opus 5.5 is available today.」 2026 年 9 月 22 日.
<https://x.com/AnthropicAI/status/2102435703535939725>
- [13] AI Wiz 「Claude Opus 5.5 explained: price, effort and what breaks」 . <https://aiwiz.uk/blog/claude-opus-5-5/> [二次情報]
- [14] VentureBeat 「Anthropic launches Claude Haiku 5.5 with 90% API price reduction, matching GPT-6 Luna」 . <https://venturebeat.com/technology/anthropic-launches-claude-haiku-5-5-with-90-api-price-reduction-matching-gpt-6-luna>
- [15] kingy.ai 「Claude Haiku 5.5: Benchmarks, Specs, Sol & Luna Compared」 . <https://kingy.ai/blog/claude-haiku-5-5-benchmarks-specs-sol-luna/> [二次情報（システムカード記載値の引用を含む）]

- [16] LLM Stats 「Claude Haiku 5.5 vs Claude Sonnet 5.5: Benchmarks, Pricing & Which Is Better in 2026」 .
<https://llm-stats.com/models/compare/claude-haiku-5-5-vs-claude-sonnet-5-5> [二次情報]
- [17] OpenRouter 「Claude Haiku 5.5 - API Pricing & Providers」 . <https://openrouter.ai/anthropic/claude-haiku-5.5>
- [18] Artificial Analysis 「Claude Haiku 5.5」 モデル別分析ページ (low/medium/high/xhigh/max) .
<https://artificialanalysis.ai/models/claude-haiku-5-5> [エフォート別のページは末尾に -low、-medium、-high、-xhigh を付した URL]
- [19] Artificial Analysis 「Anthropic has released Claude Haiku 5.5」 . <https://artificialanalysis.ai/articles/claude-haiku-5-5>
- [20] Artificial Analysis 公式 X アカウント (@ArtificialAnlys) の投稿, 2026 年 10 月 7 日.
<https://x.com/ArtificialAnlys/status/2107911905822351609>
- [21] The Decoder 「Claude Haiku 5.5 arrives with massive price cuts proving the AI pricing arms race is far from over」 . <https://the-decoder.com/claude-haiku-5-5-arrives-with-massive-price-cuts-proving-the-ai-pricing-arms-race-is-far-from-over/>
- [22] Latent Space 「[AINews] Claude Haiku 5.5 — better than GPT-6 Luna at the same pricing」 .
<https://www.latent.space/p/ainews-claude-haiku-55-better-than>
- [23] Developers Digest 「Claude Haiku 5.5: Pricing, Migration Changes, and When to Use It」 .
<https://www.developersdigest.tech/blog/claude-haiku-5-5-release-guide-2026>
- [24] datnguyenquy94/news-radar (GitHub) 「Hacker News AI Digest 2026-10-08」 Issue #697.
<https://github.com/datnguyenquy94/news-radar/issues/697> [二次情報 (Hacker News 上の議論の要約)]
- [25] マイナビニュース TECH+ 「Anthropic 「Claude Haiku 5.5」 発表 API 料金最大 90%減、性能も向上」
2026 年 10 月 8 日. <https://news.mynavi.jp/techplus/article/20261008-5093450/>
- [26] ビジネス+IT 「アンソロピック、小型 AI 「Claude Haiku 5.5」 発表 入力 100 万トークン 0.10 ドルから」
2026 年 10 月 8 日. <https://www.sbbi.jp/article/cont1/187342> (Yahoo!ニュース転載：
<https://news.yahoo.co.jp/articles/eb6dad12d6e804be779d73f70c65c760abd077bc>)
- [27] 窓の杜 「Anthropic、 「Claude Haiku 5.5」 を発表 ～1 年ぶりの小型モデル、最速・最安を更新」 2026 年
10 月 8 日. <https://forest.watch.impress.co.jp/docs/news/2146582.html>
- [28] gihyo.jp 「Anthropic、 Claude Haiku 5.5 を発表——実行コストは Haiku 4.5 比で平均約 75%減、
OSWorld 2.1 のオフライン評価で 72.4%、 Sonnet 5.5 は 83.9%」 2026 年 10 月.
<https://gihyo.jp/article/2026/10/claude-haiku-5.5>
- [29] クラスメソッド DevelopersIO 「Amazon Bedrock で Claude Haiku 5.5 が利用可能になりました」 .
<https://dev.classmethod.jp/en/articles/bedrock-claude-haiku-5-5/> [東京リージョンのプロファイル名は
調査過程の二次情報に基づく]
- [30] Anthropic Privacy Center 「How long do you store my organization's data?」 .
<https://privacy.claude.com/en/articles/7996866-how-long-do-you-store-my-organization-s-data>
- [31] CompanyScope 「Anthropic compliance: GDPR, AI Act, DPA, training, transfers」 .
<https://companyscope.io/vendors/anthropic> [二次情報 (商用規約の引用を含む)]

- [32] Voibe 「Claude API Data Retention: ZDR, Training & Commercial Terms」 .
<https://www.getvoibe.com/resources/claude-api-data-retention/> [二次情報]
- [33] Ready Solutions 「Claude Data Retention Isn't a Mystery. Your Approval Process Is.」 2026 年 6 月 8 日.
<https://readysolutions.ai/blog/2026-06-08-claude-data-handling-due-diligence/> [二次情報]
- [34] Claude Platform Docs 「API and data retention」 . <https://platform.claude.com/docs/en/manage-claude/api-and-data-retention>
- [35] DataCamp 「Claude Haiku 5.5: Features, Benchmarks, and Pricing」 .
<https://www.datacamp.com/blog/claude-haiku-5-5> [二次情報]
- [36] Voibe 「Is Claude Code Safe? Privacy & Data Retention by Tier (2026)」 .
<https://www.getvoibe.com/resources/is-claude-code-safe/> [二次情報]
- [37] 日本弁理士会「AI 等を用いた業務支援サービスの提供と弁理士法第 75 条との関係について」 令和 7 年 4 月. [URL 未確認。原文は直接確認できておらず、本稿の記述は二次情報経由]
- [38] 恵泉国際特許・法律事務所グループ「日本の特許出願が 1 か月で約 8 万件に急増 — AI 時代と弁理士法 75 条問題」 . <https://keisenassociates.com/japanese2/post-2582/> [二次情報]
- [39] 日本弁理士会「弁理士業務 AI 利活用ガイドライン」 令和 7 年 4 月. <https://www.jpaa.or.jp/cms/wp-content/uploads/2025/04/AIservices-guideline.pdf>
- [40] 法務省大臣官房司法法制部「A I 等を用いた契約書等関連業務支援サービスの提供と弁護士法第 72 条との関係について」 令和 5 年 8 月. <https://www.moj.go.jp/content/001400675.pdf>
- [41] 法務省大臣官房司法法制部「ビジネス分野における A I 等法務業務支援サービス提供と弁護士法第 7 2 条の関係について」 令和 8 年 8 月 21 日. <https://www.moj.go.jp/content/001469040.pdf>（公表ページ：
https://www.moj.go.jp/housei/shihouseido/housei10_00134.html）
- [42] Yahoo Finance 「Anthropic reveals Haiku 5.5 model as AI pricing war intensifies」 .
<https://finance.yahoo.com/technology/article/anthropic-reveals-haiku-55-model-as-ai-pricing-war-intensifies-180000423.html>