国産生成 AI 基盤モデル「PLaMo 3.1 Prime」深掘り調査

―知財業務への利用可能性―

調査時点：2026 年 10 月 29 日

Claude Opus 5.5

1. 要旨

■結論（分析）　PLaMo 3.1 Prime は、秘密情報を社外に出さずに日本語の文献を読む・まとめる・調べる用途では、知財部門にとって有力な選択肢である。一方、クレーム解釈、拒絶理由への反論、FTO の結論のように法的判断や高度な推論が中心となる業務では、海外フロンティアモデルや特許専用ツールの代わりにはならない。

■新しい点（事実）　2026 年 10 月 27 日に提供開始。Web 検索と Deep Research（調査計画の立案から出典付きレポート作成までの自律実行）を搭載し、オンプレミスでは社内文書だけを対象にした閉域での調査もできる[1–4]。API 料金は Standard プランで 100 万トークンあたり入力 60 円・出力 250 円[3,4]。

■性能の位置づけ（事実）　PFN の自社評価では、安全性は比較対象の中で最高だが、STEM・コーディング・Web 検索を伴う回答は同クラスのオープンモデルを下回る[3,5]。GPT・Claude・Gemini の上位モデルとの直接比較は、今回確認した範囲では公表されていない。前世代の 3.0 Prime は日本の法令分野を弱点としていた[6]。

■知財分野での採用事例（事実）　今回の調査では確認できなかった。

凡例：本文中の上付き数字は文末の参考文献番号を示す。見出しの【事実】は出典で確認した内容、【分析】は公開情報に基づく筆者の評価・推測である。参考文献の※印は確認上の注記。

2. 調査方法と確認の限界

- 指定された一次ソース 2 件（PFN のニュースリリース[1]と技術ブログ[7]）は、サイト側が自動取得を拒否したため本文を直接読めなかった。記載内容は、検索結果に表示された抜粋と、NICT[2]・主要メディア[3,4,8]の報道で照合した。
- 3.1 Prime のパラメータ数、アーキテクチャ、コンテキスト長、個別ベンチマークの数値は確認できなかった。

- 性能値はすべて PFN の自社測定で、第三者による追試は確認できない[5]。
- 第 5 章の業務別適性は公開情報に基づく分析で、知財タスクで実測したものではない。

3. PLaMo 3.1 Prime の基本情報【事実】

3.1 リリースと提供形態

- 2026 年 10 月 27 日に提供開始。提供形態は PLaMo Chat・PLaMo API・オンプレミスの 3 つ[3,4,8]。
- フルスクラッチ開発の国産モデルで、NICT との共同研究で得た事前学習モデルをベースとする。安全性の開発・評価も NICT が支援した[2,8]。
- PLaMo シリーズは、クラウド API・オンプレミスのほか、Amazon Bedrock Marketplace、Snowflake、Microsoft Foundry 経由でも提供されている[9]。3.0 Prime と PLaMo 翻訳は 2026 年 10 月 6 日に Microsoft Foundry で提供が始まった[10]。3.1 Prime が各クラウドで使えるかは未確認。

3.2 主な改善点（PFN 発表）

- **事実性・忠実性**：知らないことは「分からない」と答え、与えた文脈を正しく読むよう事後学習を強化した。ハルシネーション関連の 8 ベンチマーク平均で、3.0 Prime 比約 15%改善[1,3,4]。8 指標は FaithEval、AA Omniscience、TruthfulQA、SimpleQA Verified、Japanese SimpleQA、ALCE、LongBench-Cite、TMG-CQA とされる[1]。技術ブログは、推論モードの事実性・忠実性カテゴリ平均が 6.4 ポイント向上したと説明している[7]。
- **ハイブリッド推論**：推論・非推論を 1 つのモデルで切り替える。非推論モードでは出力トークン数を約 90%削減[3,4]。
- **速度**：投機的デコーディングにより、3.0 Prime 比で最大 1.7 倍[3,5,8]。

3.3 Deep Research 機能

- Web を繰り返し検索し、調査計画の立案から出典付きレポートの作成までを数分〜数十分で自律的に行う[4,8]。

- オンプレミスでは社内文書の検索システムと連携し、社内文書だけ、または社内文書と Web を組み合わせた調査ができる[3,4]。
- PFN によれば、推論モードで Web 検索 QA が平均 2.4 ポイント、Web 検索レポートが平均 11.1 ポイント向上した（BrowseComp-Plus、SealQA、HealthBench、ResearchQA など）。評価の一部は、英語版 Wikipedia と PubMed を社内に構築した検索環境（OpenSearch）で行ったとされる[7]。
- 検索バックエンド、出典表記の形式、特許データベースとの連携可否は、公式情報で確認できなかった。

3.4 料金とデータの取扱い

表 1　PLaMo API のプラン別料金とデータの取扱い

プラン	料金	データの取扱い
Free	無料（利用上限・レートリミットあり）[3,5]	入出力を学習に利用するとの公式説明がある[11]。無償サービスの入出力はサービス改善・AI 開発に使われる場合があるとされる[12]
Standard	100 万トークンあたり 入力 60 円・出力 250 円[3–5]	学習に利用しない。ZDR（入力データ非保持）の対象（3.0 Prime 時点の情報）[6]
Provider・ 企業オンプレミス	個別契約（公開価格は未確認）	Provider プランも ZDR の対象（3.0 Prime 時点の情報）[6]

- PLaMo Chat では 2026 年 11 月 30 日まで、Deep Research を 1 日 2 回まで無料で試せる[3,4]。
- Web 検索の利用に別料金がかかるかは確認できなかった（調査過程で検索 1 回あたりの料金に触れた転載記事を見たが、出典を特定できなかった）。導入前に PFN への確認が必要。

3.5 開発経緯とソブリン AI としての位置づけ

- GENIAC 第 1 サイクルで PLaMo-100B を開発した[12]。PLaMo 翻訳も GENIAC の一環として開発され、設計から学習まで国内で完結している[13]。
- デジタル庁のガバメント AI では、PLaMo 2.0 Prime が 2026 年 3 月に試用モデルに選ばれ[14]、5 月に評価検証契約を結んだ 5 社の 1 つとなった[6,10]。PLaMo 翻訳は 2025 年 12 月から「源内」で使われている[13,15]。

- **民間の導入例**：日本総研は 2026 年 10 月から、PLaMo 3.0 Prime を自社のオンプレミス環境で全社利用している。自社データでの追加学習も検証中で、用途は設計書・ソースコードのレビュー支援、社内規程の検索、議事録作成、翻訳など。導入前に約半年の評価を行った[9]。
- PLaMo 2.0 Prime は 2025 年日経優秀製品・サービス賞の最優秀賞を受賞している[15]。

3.6 セキュリティ認証

- PLaMo 翻訳は国内サーバで処理し、入出力は応答後に破棄して学習に使わない。ISMS 認証を取得している[16]。
- PLaMo API について SOC2・ISMAP を未取得とする第三者記事がある[17]が、本文を再確認できなかった。調達要件に含まれる場合は PFN への直接確認が必要。

3.7 確認できなかった事項

- 3.1 Prime の総パラメータ数、アーキテクチャ（Dense／MoE）、コンテキスト長、最大出力トークン数。
- 参考として、前世代 3.0 Prime のコンテキスト長は 256K（Standard プランは 128K までとの記載あり）[6,18]、β 版は入力 64K・出力 20K だった[19]。PLaMo 3 系のベースモデルは 2B・8B・31B 規模で公開されている[18,20]。ただし、ここから 3.1 Prime の規模は推定できない。
- 個別ベンチマークの数値（公式発表では図のみで示されている）。

4. 他モデルとの比較【事実＋分析】

表 2　公開情報の範囲での比較

観点	PLaMo 3.1／3.0 Prime	海外フロンティア（GPT・Claude・Gemini 系）	他の国産 LLM
公表済みの比較	3.1：35 ベンチマーク・9 カテゴリで、gpt-oss-120b、Gemma 4 31B IT、Qwen3.8 27B など同クラスのオープンモデルと比較[3,5] 3.0：GPT-5.4 Mini、Claude Haiku 4.5 などと比較[6]	上位モデルと PLaMo の直接比較は確認できず	同一条件での第三者による横並び評価は確認できず

観点	PLaMo 3.1／3.0 Prime	海外フロンティア（GPT・Claude・Gemini 系）	他の国産 LLM
強み	安全性は比較対象中で最高（3.1）[3] 日本語に特化、オンプレミス対応	総合的な推論力、長文処理、エージェント性能（分析）	tsuzumi 2 は 30B の軽量モデル。Sarashina3 は約 30 兆トークンで学習[21]
弱み	STEM・コーディング・Web 検索を伴う回答で比較対象を下回る（3.1）[3,5] 3.0 は Web 探索、長文脈、数学・STEM、日本の法令が弱点[6]	データの越境・経済安全保障上の懸念（分析）	規模・機能は製品ごとに大きく異なる
価格	Standard：入力 60 円・出力 250 円／100 万トークン。3.0 時点で Claude Haiku 4.5・GPT-5.4 Mini と同価格帯とされる[6]	モデルにより大きく異なる	多くは法人契約

評価上の注意（分析）：3.1 Prime の比較相手は、いずれも同じ性能帯のオープンモデルである。「国産で最高水準」と「フロンティア級」は別物として扱うべきである。

5. 知財業務への利用

5.1 事例と関連製品【事実】

- PLaMo Prime を知財分野で採用・検証した事例や、PFN と知財関連企業の提携は確認できなかった。
- **競合陣営では知財特化の製品化が進む**：NTT ドコモビジネスは、tsuzumi を中核とする「知財文書作成エージェント」を 2025 年 6 月に提供開始した（発明発掘から明細書ドラフト生成まで、閉域網で運用）[22]。パテント・インテグレーションは「サマリア」に明細書作成支援機能を追加した（2026 年 3 月）[23]。知財向けの製品・事例を持たない点は、PLaMo の現状の弱点である。
- **PLaMo 翻訳（plamo-3-translate）**：2026 年 10 月 7 日発表。PLaMo 3 NICT をベースとし、最大 31B。日英・英日翻訳で比較対象のほぼすべての API を上回り、コストは PLaMo API 料金で試算するとおおむね 30 分の 1 と PFN は説明している。ベースモデルの学習データに占める日本語の割合は 30%近い[24]。

- PFN の技術ブログは、英文クレームを訳す際に文脈から特許文書と判断し、「前記」などの日本の特許文体で出力した例を示している。比較対象には Claude Fable 5.1、GPT-6.1 Sol、GPT-6 Astra、DeepL、Google 翻訳が含まれ、plamo-3-translate 31B を上回ったのは Claude Fable 5.1 のみで、コストはその約 85 分の 1 とされる[25]（本文未取得、検索結果の抜粋による）。
- **ローカル運用の選択肢**：前世代の plamo-2-translate は PLaMo Community License で配布され、個人と年商 10 億円未満の企業はオンプレミスで無償の商用利用ができる[26]。plamo-3-translate の重みも Hugging Face で公開予定（2026 年 10 月時点）。PLaMo 翻訳単体の API は未公開[16,24]。

5.2 業務別の適性【分析】

表 3　知財業務別の適性評価（◎適する　○条件付きで適する　△補助に限る　×使わない）

業務	適性	理由・使い方
特許公報の読解・要約	◎	日本語の指示追従に強く、非推論モードは安価で速い。大量の公報の一次要約・スクリーニングに向く
外国出願向け翻訳・ 外国公報の和訳	◎ （PLaMo 翻訳）	特許文体への適応が PFN の例で示されている。用語の統一と、クレームの限定関係の欠落・付け足しは人が必ず照合する
社内技術文書・ 発明提案書の横断調査	◎ （オンプレ）	閉域で社内文書に限定した調査ができる。未公開情報を扱う知財部門にとって最大の利点
IP ランドスケープ （非特許情報）	○	市場・競合・技術動向の調査に向く。特許統計は特許 DB から別途取得して与える
契約書レビュー	○ （機密重視時）	条項の抽出・比較はできる。法的評価は人が行う
先行技術調査	△	特許 DB を直接検索できるという公式情報はない。検索式・キーワードの案出しと、ヒット文献の対比表作成の補助に限る
明細書ドラフト作成	△	日本語の文章力は期待できるが、明細書での実績はない。下書き用途に限り、品質は人が担保する
中間処理 （拒絶理由への対応）	△	引用文献と本願の対比整理には役立つ。進歩性の反論ロジックは上位モデルと人が担う
クレーム解釈	△〜×	法令分野の弱さ（3.0）と推論力の差から、主たる判断には使わない

業務	適性	理由・使い方
FTO 調査	×	網羅性と法的判断が必要で、誤りの代償が大きい。補助的な読解に限る

5.3 Deep Research の有効性と限界【分析】

- PFN が社内に構築したコーパスで検索性能を評価している[7]ことから、検索対象を特定のコーパスに限定する設計は想定されていると考えられる。
- ただし、特許 DB との連携は公式に示されていない。J-PlatPat は検索フォーム型のため、一般的な Web 検索では公報本文を網羅的に参照しにくい。Google Patents の個別ページも網羅性は保証されない。このため、IPL や先行技術調査の母集団形成には使えない。
- 社内に特許 DB（商用 DB から出力したデータなど）の検索システムを置き、オンプレミスの Deep Research に接続する構成なら、閉域での技術動向調査は検討に値する。可否は PFN への確認事項である。
- Web 検索 QA は同クラスの競合を下回るため[3]、出典が実在するか、その出典が主張を本当に支えているかを人が全件確認する前提で使う必要がある。

5.4 オンプレミス対応の意味【分析】

- **守秘と営業秘密管理**：未公開発明、発明提案書、出願前の技術資料、係争・ライセンス交渉の資料を外部に送らずに済む。秘密管理性の説明や社内承認が通りやすくなる点は実務上大きい。
- **新規性喪失リスク**：守秘義務のある契約下のクラウド API に入力しても、直ちに「公然知られた」状態になるとは一般に解されていない。ただし、学習に利用される経路（Free プランなど）は避けるべきである。オンプレミスなら、この論点自体が生じにくい。
- **コスト**：GPU インフラと運用の負荷がかかる。部門単独での導入より、全社の生成 AI 基盤に乗る形（日本総研型[9]）が現実的である。

5.5 リスクと注意点

- **ハルシネーションと出典の誤り**：約 15%の改善は 3.0 Prime 比で、かつ自社評価である[1,5]。

- **最新の判例・法改正**：Web 検索で補えるが、検索を伴う回答は相対的に弱く、3.0 では日本の法令が弱点だった[6]。3.1 で改善したかは未確認。条文・審査基準・判例は一次資料での照合が必須。
- **化学・数値限定・パラメータ発明**：STEM の弱さ[3]から、構造式や数値範囲の理解には注意が要る。

6. 評価・反響【事実】

- 国内メディア（ITmedia AI＋、Impress Watch、窓の杜）は発表内容を中心に報じている。ITmedia は、安全性の高さと、STEM・コーディング・Web 検索での劣後を併記している[3,4,8]。窓の杜は、社内文書だけを対象とした閉域での調査を特徴として紹介している[4]。
- SmartScope は、数値がすべて PFN の自社測定で、第三者による追試がない点を指摘している[5]。
- 個人の論評でも、社内文書に限定した調査ができる点が注目されている[27]。
- 英語圏での独自の報道・レビューはほとんど見当たらなかった。

7. 知財部門への実務的示唆【分析】

1. **向いている業務**：オンプレミスまたは Standard 以上のプランでの、公報の大量要約・分類、社内技術文書の横断調査、外国公報の和訳（PLaMo 翻訳）、契約書の条項抽出、非特許情報による技術・市場動向の調査。
2. **向いていない業務**：クレーム解釈、進歩性の反論ロジック構築、FTO の結論出し、特許 DB 全体を対象とした網羅検索、数理・化学の深い理解を要する対比。
3. **使い分けの原則**：社外に出せない情報は PLaMo（オンプレミス）、社外に出せて高度な推論が必要な情報は学習に使われない契約の海外フロンティアモデル、母集団形成と網羅検索は特許専用ツール・商用 DB で扱う。この 3 系統の役割分担を社内ガイドラインに明記する。
4. **導入前の検証（PoC）**：公開済みの公報や応答済みの拒絶理由など、自社の過去案件でタスクを再現し、要約の欠落率と出典の正誤率を測る。

5. **PFN への書面確認事項**：Deep Research の検索バックエンドと社内特許 DB との接続可否、Web 検索の課金体系、オンプレミスの費用感、SOC2・ISMAP の取得状況、3.1 Prime のコンテキスト長、Free プランを含む各プランのデータ取扱い。
6. **Free プラン・無料キャンペーンの扱い**：入力が学習に使われるため、公開情報だけで試す用途に限る。

参考文献

閲覧日はいずれも 2026 年 10 月 29 日。※印は確認上の注記。

[1] 株式会社 Preferred Networks「国産生成 AI 基盤モデル PLaMo Prime に Web 検索・Deep Research 搭載 調査計画の立案から出典付きレポートまで自律的に実行、オンプレミスにも対応」ニュースリリース（2026 年 10 月 27 日）. https://www.preferred.jp/ja/news/pr20261027　※本文は自動取得不可（サイト側が拒否）。記載内容は検索結果の抜粋と報道で照合

[2] 国立研究開発法人情報通信研究機構（NICT）「Preferred Networks が国産生成 AI 基盤モデル「PLaMo 3.1 Prime」をリリース」トピックス（2026 年 10 月 27 日）. https://www.nict.go.jp/publicity/topics/2026/10/27-1.html

[3] ITmedia AI＋「PFN の国産フラグシップモデル「PLaMo 3.1 Prime」発表　オンプレで Deep Research 対応　API 無料プランも開始」（2026 年 10 月 28 日）. https://www.itmedia.co.jp/aiplus/article/2610/28/2000002652/

[4] 樽井秀人「国産の生成 AI 基盤モデル「PLaMo 3.1 Prime」が提供開始 ～Web 検索と Deep Research に対応」窓の杜（2026 年 10 月 28 日）. https://forest.watch.impress.co.jp/docs/news/2151522.html

[5] SmartScope「PLaMo 3.1 Prime とは？API 無料プランの条件・Deep Research・日本語性能」（2026 年 10 月 28 日）. https://smartscope.blog/blog/plamo-3-1-prime-free-api-plan-conditions-2026/

[6] AI 総合研究所「PLaMo 3.0 Prime とは？性能や料金、国産 LLM としての立ち位置を解説」（2026 年 7 月 1 日、監修：坂本将磨）. https://www.ai-souken.com/article/what-is-plamo-3-0-prime

[7] 株式会社 Preferred Networks「PLaMo 3.1 Prime をリリースしました」技術ブログ（公開日未確認）. https://www.preferred.jp/ja/blog/tech/plamo-3-1-prime-release　※本文は自動取得不可（サイト側が拒否）。記載内容は検索結果の抜粋と報道で照合

[8] 太田亮三「PFN、調査業務に対応した国産 AI 最新版「PLaMo 3.1 Prime」」Impress Watch（2026 年 10 月 27 日）. https://www.watch.impress.co.jp/docs/news/2151518.html

[9] 株式会社日本総合研究所・株式会社 Preferred Networks「日本総研と PFN、国産大規模言語モデル「PLaMo」のライセンス契約を締結 ～AI 活用の推進に向け日本総研オンプレミス環境での全社利用と追加学習の検証を開始～」ニュースリリース（2026 年 10 月 26 日）. https://www.jri.co.jp/company/release/2026/1026/

[10] SPIKE STUDIO「国産 LLM を徹底比較！各モデルの特徴と選び方【2026 年最新版】」（2026 年 10 月 9 日）. https://spikestudio.jp/blog/japan-llm-companies-2026　※検索結果上の表題（「【2026 年版】国産 LLM15 選｜日本企業の日本語モデルを商用利用の可否で比較」）と取得時の見出しが不一致

[11] 株式会社 Preferred Networks「PLaMo」公式サイト. https://plamo.preferredai.jp/ ※再取得時は本文の一部しか表示されず、データ取扱いの記述は照合できなかった

[12] 相坂ソウタ「国産 AI モデル「PLaMo」を無料で使う方法【初心者向け解説】」GMO 天秤 AI メディア（2025 年 11 月 17 日、2026 年 6 月 18 日更新）. https://tenbin.ai/media/ai_guide/plamo-free-guide

[13] 株式会社 Preferred Networks「PFN の PLaMo 翻訳、ガバメント AI「源内」で利用開始」ニュースリリース（2025 年 12 月 2 日）. https://www.preferred.jp/ja/news/pr20251202 ※本文は自動取得不可（サイト側が拒否）。記載内容は検索結果の抜粋と報道で照合

[14] 株式会社 Preferred Networks「PFN の国産大規模言語モデル「PLaMo」、デジタル庁のガバメント AI に試用モデルとして選定」ニュースリリース（2026 年 3 月 12 日、URL からの推定）. https://www.preferred.jp/ja/news/pr20260312 ※本文は自動取得不可（サイト側が拒否）。記載内容は検索結果の抜粋と報道で照合

[15] 株式会社 Preferred Networks「PFN の国産大規模言語モデル PLaMo 2.0 Prime、2025 年日経優秀製品・サービス賞にて最優秀賞を受賞」PR TIMES（2026 年 1 月 5 日）. https://prtimes.jp/main/html/rd/p/000000020.000156310.html

[16] 株式会社 Preferred Networks「PLaMo 翻訳 | 世界最高レベルの日本語性能の AI 翻訳サービス」公式サイト. https://translate.preferredai.jp/

[17] Admina by Money Forward「PLaMo 3.0 Prime とは？料金・性能・導入判断を解説」（公開日不明）. https://admina.moneyforward.com/jp/blog/plamo-3-prime-enterprise-adoption ※再確認時に本文を取得できず、SOC2・ISMAP に関する記述は未照合

[18] Oflight Inc.「PLaMo 3.0 Prime 徹底解説（PLaMo 3.0 Prime Deep Dive）」（2026 年 6 月 22 日）. https://www.oflight.co.jp/en/columns/plamo-3-0-prime-pfn-japanese-llm-2026-06

[19] 今村秀明・逸見榛一「国産生成 AI PLaMo を支える事後学習と推論最適化」Preferred Networks、AiTech Seminar 講演資料（2026 年 4 月 6 日）. https://speakerdeck.com/pfn/20260406_plamo_3_beta_posttrain_and_inference_opt

[20] STEAM/acc「【AI・5 行要約】PFN×NICT、日本語強化 LLM "PLaMo 3"の挑戦」note（2025 年 11 月 18 日）. https://note.com/steam0101/n/nf13b093bbc37

[21] AI 革命 編集部「Sarashina3 とは？SB Intuitions 開発の国産 LLM の 6 モデル・性能・料金・無料ベータ・tsuzumi 比較を整理【2026 年最新】」AI 革命メディア（2026 年 7 月 5 日、2026 年 10 月 27 日更新）. https://ai-revolution.co.jp/media/what-is-sarashina3/

[22] 萬秀憲「NTT ドコモビジネスの知財文書作成エージェント」よろず知財戦略コンサルティング ブログ（2025 年 12 月 9 日）. https://yorozuipsc.com/blog/ntt

[23] パテント・インテグレーション株式会社「生成 AI を活用した明細書作成支援機能をリリース～人間主体の特許実務を前提に、品質と効率の両立を支援～」PR TIMES（2026 年 3 月 19 日）. https://prtimes.jp/main/html/rd/p/000000020.000086119.html

[24] 樽井秀人「PFN、国産の翻訳特化モデル「plamo-3-translate」を発表。「PLaMo 翻訳」にも導入」窓の杜（2026 年 10 月 8 日）. https://forest.watch.impress.co.jp/docs/news/2146581.html

[25] 株式会社 Preferred Networks「plamo-3-translate: フロンティア級の翻訳特化 LLM」技術ブログ（公開日未確認）. https://www.preferred.jp/ja/blog/tech/plamo-3-translate-product-release ※本文は自動取得不可（サイト側が拒否）。記載内容は検索結果の抜粋と報道で照合

[26] 石山裕規「ローカル環境でも実行可能な翻訳特化型の国産フルスクラッチ LLM「PLaMo 翻訳」～個人は無償で商用可能」窓の杜（2025 年 5 月 28 日）.
https://forest.watch.impress.co.jp/docs/news/2017742.html

[27] ユニコ「PLaMo 3.1 Prime's Deep Research now enables research limited to internal documents」note（2026 年 10 月 27 日）. https://note.com/unikoukokun/n/nd45cc7bdd735 ※取得時の見出しは英語表記で、日本語の原題は未確認