

NTT純国産AI tsuzumi 2 と tsuzumi 2 Vision の最新状況と知財業務適用性評価

エグゼクティブサマリ

情報取得日：2026年6月7日（JST）。本調査で確認できた公開情報を総合すると、**tsuzumi 2**は「30B・1GPU動作可能・オンプレミス適性が高い・日本語に強い」ことを中核に据えた純国産LLMであり、2025年10月に提供開始、2026年3月にはデジタル庁のガバメントAI「源内」での試用対象に選定され、2026年5月には図表入り日本語ビジネス文書の理解と数値処理を含む論理的思考力を強化したアップデートが公表されました。公開面では、Microsoft Foundry上の **tsuzumi2** は **Version 2 / GA / Chat completion / text input-output** として確認できる一方、**vision対応版の独立した一般公開APIは本日時点で明確には確認できず**、NTT公式本文でも「tsuzumi 2アップデート」として順次提供予定とされています。したがって、**本レポートでは、ユーザーの呼称に合わせ、この2026年5月の視覚拡張版を便宜上「tsuzumi 2 Vision」と表記します。**

1

知財業務への適用性は、**要約・翻訳、発明者インタビュー支援、出願ドラフト支援、検索結果の整理・論点抽出、契約レビューの初期論点洗い出し**では高く、**侵害調査・FTO・新規性/進歩性の最終判断**では限定的です。理由は、公開資料上の強みが**日本語処理、RAG、ビジネス文書理解、低コスト運用**に集中している一方、知財実務で決定的に重要な**引用根拠の完全性、特許クレーム解釈の一貫性、図面・化学構造式・数式・CAD図の厳密読解**については、十分な公開検証がまだ示されていないためです。日本弁理士会の2025年ガイドラインも、生成AIの成果物の正確性は保証されず、**最終責任は弁理士側にある**ことを強調しています。 2

導入方針としては、**単体チャット導入ではなく、閉域環境・RAG・証拠トレース・人手監査・ログ統制を組み込んだ「業務特化AIコパイロット」構成が必須**です。個人情報保護委員会は、生成AIに個人データを入力する際、**機械学習への利用が行われないこと等を十分確認**することを求めています。また総務省・経産省のAI事業者ガイドライン第1.2版は、AIのライフサイクル全体での**リスクベースアプローチ**を求めています。知財部門で安全に使うには、**入力データ分類、学習不使用保証、アクセス権限、出力根拠の保存、モデル評価、差戻しフロー**を前提にすべきです。 3

結論を先に言えば、**tsuzumi 2**は知財部門向けの「**日本語重視・機密重視・コスト重視**」基盤として有力です。特に、**社内限定の出願支援・要約・契約レビュー補助・発明ヒアリング補助**には適します。他方で、**「tsuzumi 2 Vision」を前提にした図面/表/帳票の本格利用は、公開API・ベンチマーク・評価手順がもう一段明確になるまで段階導入が妥当**です。短期はテキスト中心のクローズドPoC、中期はVisionとRAGの統合、長期は知財専用アダプタやAIエージェント化、という三段階が実務上もっとも現実的です。 4

調査前提と最新状況

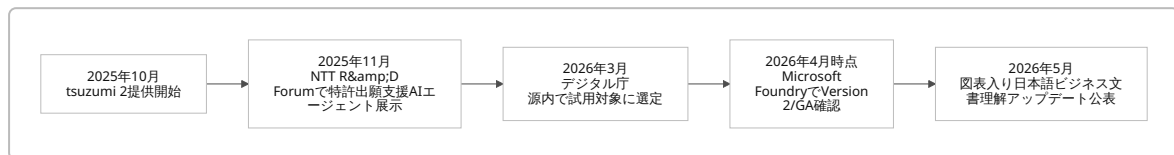
まず、**最新の公開状況を時系列で整理**すると、NTTは2025年10月20日に「tsuzumi 2」を提供開始し、純国産・高性能・高セキュア・低コストを前面に出しました。続いて、2026年3月にはデジタル庁が政府職員向け生成AI環境「源内」で試用する国内LLMとして **tsuzumi 2** を選定しています。さらに2026年5月19日、NTTは**図表・グラフ・チャートを含む日本語ビジネス文書理解と数値処理・論理思考の強化**を公表し、これを「tsuzumi 2アップデート」として順次サービス提供するとしました。 5

この2026年5月アップデートについて、公式本文の重要な一文は次の通りです。

ここで重要なのは、ユーザーが言う「tsuzumi 2 Vision」に相当する能力向上は公開確認できる一方、公式日本語資料では主に独立製品名ではなく「tsuzumi 2アップデート」として記述されている点です。したがって、商談・調達・PoC企画では、「Vision機能がどの提供チャンネルで、どの契約形態で、いつ利用可能か」を個別確認する必要があります。これは特に、知財部門のように帳票画像、図表、契約書PDF、特許図面を扱うユースケースでは重要です。 7

公開チャンネルを見ると、Microsoft Foundryの tsuzumi2 カタログには Version 2、GA、Chat completion、text input / text output、managed computing、custom license が記載されています。また、「本モデルではチューニング機能の利用はできない」と明記されています。つまり、2026年6月7日時点で一般公開面から確認できるtsuzumi 2の主なAPI経路はテキスト中心であり、Vision機能は少なくとも公開カタログ上では別オファーとしては確認できませんでした。 8

この点は、知財導入判断に直結します。なぜなら、テキスト中心の初期PoCと、画像/表/レイアウト理解を含む本番設計では、必要な評価セットも監査ポイントもまったく異なるからです。特許要約や日英翻訳の補助であればテキスト版から始められますが、請求項と図面の対応確認、表を含む契約比較、クレームチャートの読解まで狙うなら、Vision相当機能の可用性と品質確認が前提になります。 9



上図のとおり、知財実務に直接関係する公開マイルストーンは、提供開始、知財向けエージェント展示、政府採用、Vision系アップデートの四点です。特に、R&D Forumで特許出願支援AIエージェントが既に示されていることは、知財領域がNTT側の想定用途に入っていることを示します。ただし、これは公開資料上 R&D フェーズ/ビジネス展開としての展示であり、量産導入事例そのものではありません。 10

技術概要

技術面で公開情報を厳密に整理すると、tsuzumi 2 の公開確定情報はかなり多い一方、層構成・コンテキスト長・推論TPS・ライセンス条項全文などは未公表です。以下の表は、公開確認できた範囲と、未確認点を分けて整理したものです。 11

項目	tsuzumi 2	tsuzumi 2 Vision
公式上の位置付け	NTTが2025年10月に提供開始した純国産LLM。30Bで1GPU動作可能と説明される。 12	2026年5月公表の図表入り日本語ビジネス文書理解アップデート。本レポートでは便宜上「Vision」と表記。公式本文は「tsuzumi 2アップデート」。 6
モデルサイズ	30B。PDF資料では「tsuzumi 2 30B」と明記。 13	ベースはtsuzumi 2 30Bと読むのが自然だが、Vision側アダプタ等を含む総パラメータは未公表。 14

項目	tsuzumi 2	tsuzumi 2 Vision
アーキテクチャ	public documentation上は詳細未公表。ただしtsuzumi系列の公開モデルカードでは Transformer architecture / SFT が確認できる。tsuzumi 2も同系列である蓋然性は高いが、 tsuzumi 2固有の層構成は未確認 。 15	tsuzumi系の視覚読解研究は 画像エンコーダ+OCR+アダプタ+LLM で、画像エンコーダとLLMは固定、アダプタのみ学習する構成。2026年アップデートはこの研究系譜の実装と読むのが妥当だが、tsuzumi 2版の最終実装詳細は未公表。 16
学習データ	金融・自治体・医療分野に関する知識を幅広く学習と記載。 具体的なコーパス、トークン量、データ比率は未公表 。 17	独自収集データセットと、AAAI 2024の InstructDoc 系知見を活用した視覚読解学習の系譜が確認できる。InstructDoc自体は30の公開VDUデータセット、12タスクを統合。 18
推論性能	同サイズ帯で世界トップクラスの日本語性能、数倍以上大きいモデルにも引けを取らないとNTTが説明。RAGでも「実システムへの適用評価において世界トップクラス」と主張。R&DページではMT-bench日本語でGPT-5と同等程度の高い数値と紹介。 19	2026年5月リリースでは、 図表入り日本語ビジネス文書理解で同サイズ帯比世界トップレベル と主張。帳票・フロー図・グラフ理解の例示あり。 ただし公開テキストで数値一覧は十分確認できず 。 6
マルチモデル	公開Foundryオファーではtext input / text output。テキスト用途の一般公開経路が中心。 20	図表・グラフ・チャートを含む文書画像理解、帳票抽出、原因判断フロー理解が強調される。写真認識やJSON構造化の先行技術も確認できる。 21
API / SDK	Microsoft FoundryでChat completionとしてGA。Managed computing採用。 tsuzumi固有SDKは公開確認できず 。開発はFoundry標準手段を使う構成と理解するのが妥当。 22	独立した一般公開APIは未確認 。NTTグループ経由で順次提供予定。 6
ライセンス	Foundryカタログ上は Custom license 。公開ウェイト配布型OSSライセンスは確認できず。 23	同左。Vision版の独立ライセンス文書は未確認。 24
オンプレ/クラウド	1GPU動作可能、A100 40GB相当1基で推論可能という試算。オンプレ/プライベートクラウド適性が強く強調される。過去のtsuzumi商用説明ではオンプレ・プライベートクラウド・パブリッククラウドの3形態が示されている。 25	2026年5月リリースでも、機微情報を扱うオンプレ/プライベートクラウド用途に言及。Vision機能もその延長線上で位置づけられている。 6
セキュリティ / プライバシー	モデル自体の安全性比較で高スコアを示すとNTTが主張。加えてNTTグループはAI憲章・AIガバナンスポリシー・生成AI利用ガイドラインを整備。 26	文書画像内の機微情報も画像として処理しうる点が価値。逆に、入力データ統制とログ管理の重要性は高まる。個人情報保護委員会は、機械学習不使用等の確認を求める。 27

技術的にとくに注目すべきなのは、「**30Bにあえて留め、1GPU・オンプレ適性を確保する**」という設計思想です。NTTはPDF資料で、30Bのtsuzumi 2が**A100 40GB相当1基・約500万円**の推論ハードウェアで済み、大規模クラス比で**推論コストを約10~20分の1に低減可能**と試算しています。知財部門では、外部API従量課金よりも**閉域・定額・可監査**を優先したいケースが多いため、この思想は実務に合います。 28

Vision系の技術は、先行研究に照らすとOCRの文字位置情報と画像情報をLLMへ橋渡しするアダプタ学習がコアです。NTT技術ジャーナルでは、**画像エンコーダとLLMのパラメータを固定し、アダプタだけを学習**することで推論能力を保つと説明されています。これは、知財向けに将来**特許明細書PDF、契約スキャン、表入り実験報告書、審査通知書**へ適用するとき、**本体モデルを大きく再学習せずに拡張**できることを示唆します。

29

ただし、ここは重要な留保です。公開事例は**帳票、グラフ、表、フロー図、写真、JSON構造化**が中心で、**化学構造式、回路図、CAD図面、特許図面の細粒度比較、数式を多く含む学術先行技術**に対する性能は公開確認できませんでした。したがって、知財実務でVision機能を使う場合、「**文書レイアウト理解**」には期待してよいが、「**専門図面の意味解釈**」までは現時点で過信しないという線引きが必要です。 21

公開資料レビュー

以下は、**優先ソースを公式・原著・政府・業界ガイドライン**中心で整理した公開資料一覧です。なお、ユーザー指定のYouTube URLは入力文中に存在しなかったため、**公式YouTubeチャンネル/プレイリスト**を基準に調査しました。

区分	資料	日本語要約	重要引用	知財実務に との意味
公式リリース	NTTニュースリリース 「更なる進化を遂げた NTT版LLM tsuzumi 2の 提供開始」 (2025-10-20) 30	tsuzumi 2の正式ローンチ資料。純国産、高性能、高セキュア、低コストを打ち出し、大学導入や文書・画像データ分析連携の検討も掲載。 31	「高性能・高セキュア・低コストな純国産LLM」 32	知財部門に必要な 閉域性と日本語品質 の位置づけが最も明確。
公式リリース	NTTニュースリリース 「tsuzumi 2アップデート」 (2026-05-19) 6	図表入り日本語ビジネス文書理解と論理思考強化を公表。Vision相当機能の最新公式情報。 6	「図表入り日本語ビジネス文書処理性能を1GPU環境で実現」 6	契約書レビュー、帳票抽出、表付きレポート要約に直結。
公式技術資料	NTT R&D Website 「NTT版大規模言語モデル tsuzumi 2」 33	1GPU・40GB未満メモリGPUを想定した軽量設計や、GPT-5比較を含む性能訴求をまとめた研究紹介。 33	「1GPU動作可能な軽量モデル」 34	オンプレ前提の知財システム設計で最重要。
公式PDF	NTT資料「tsuzumi 2 進化のポイント」 (2025-10-20 PDF) 35	30B、同サイズ帯世界トップクラス、RAG性能、金融特化性能、1GPU・A100 40GB試算、安全性比較を掲載。 35	「Why 30B? Only 1 GPU On-premise」 36	導入コスト試算と評価観点づくりに有用。
原著論文	AAAI 2024 「InstructDoc」 37	30の公開VDUデータセットを統合した大規模視覚指示データセット。初見タスクへのゼロショット一般化を狙う。 37	「30 publicly available VDU datasets」 38	Vision系の汎化性能の理論的背景。

区分	資料	日本語要約	重要引用	知財実務に としての意味
原著論文	EMNLP 2023 「DueT」 39	画像・テキストエンコーダを凍結し、dual-adapterのみ学習する効率的転移学習を提案。 39	「アダプタのみを学習」 40	将来の特許・契約特化拡張のコスト感を考えるうえで重要。
技術ジャーナル	NTT技術ジャーナル「グラフィカルな文書を理解できる tsuzumi」 (2024) 41	OCR・画像エンコーダ・アダプタ・LLMによる視覚読解の実装概念を解説。 29	「アダプタのパラメータだけを学習対象」 42	Visionの内部理解に最重要。
公式展示資料	NTT R&D Forum 2025「tsuzumi 2を活用した特許出願業務支援AIエージェント」 43	特許出願プロセスを、対話エージェント・検討エージェント・執筆エージェントで支援する構想を提示。Human in the Loopを明示。 43	「特許出願件数を増やし、知財戦略を一段と強化」 44	知財実務への直接的な公式コースケース。
YouTube	NTT公式YouTube「tsuzumi 2が描く、AIビジネスの新地図」およびR&D Forum 2025ブレイリスト 45	tsuzumi 2の技術セミナー動画が公式チャンネルで公開。ロードマップと日本発LLMの戦略を把握できる。 45	「tsuzumi 2が描く、AIビジネスの新地図」 46	記事より話者の意図や製品戦略が把握しやすい。
リリースノート相 当	Microsoft Foundry Model Catalog tsuzumi2 23	Version 2、GA、text入出力、managed computing、custom license、チューニング不可を確認できる。独立した詳細changelogは今回確認できず。 23	「Version: 2」 「Lifecycle Generally available (GA)」 20	実際にどのAPI面が公開されているかを確認する一次資料。

引用ベースでみると、公開資料群は一貫して「日本語性能」「1GPU」「オンプレ/機微情報」「業務特化」を強調しています。反対に、**コンテキスト長、レイテンシ、トークン単価、ベクトルAPI、視覚版の別SKU、評価データの法務・知財特化度**は公開不足です。知財導入では、この「強く分かっている部分」と「まだ分からない部分」を切り分ける必要があります。 47

知財業務への適用可能性

NTT自身がR&D Forumで**特許出願業務支援AIエージェント**を展示しているため、知財領域は想定ユースケースに入っています。ただし、その展示も**Human in the Loop**を明示しており、知財プロセスを完全自動化する思想ではありません。これは極めて妥当です。日本弁理士会も、生成AIを使って得られた成果物をそのまま提供するのではなく、**最終的には専門家が責任を持って確認・提供する必要**を示しています。 48

「最終的には弁理士が責任をもって提供」 49

以下では、各業務ごとに**適合度、期待効果、必要な前処理・後処理、運用フロー、人的監査の要点、追加開発コスト概算**を整理します。コストは**公開価格ではなく推定**であり、NTT公表の1GPU推論用ハードウェア試

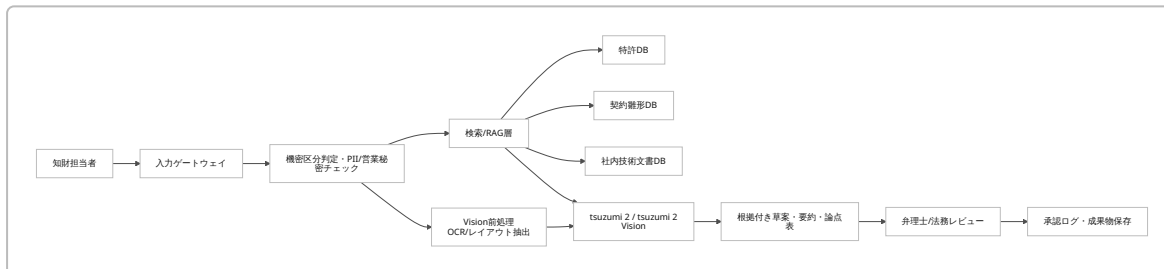
算（A100 40GB相当1基≒約500万円）をベースに、RAG・監査UI・連携開発・評価運用を加味した国内企業向けの概算レンジです。ベンダー見積りではありません。²⁸

業務	適合度	期待される効果	必要な前処理・後処理	推奨運用フロー	人的監査の要点	追加開発コスト概算
特許調査	中	検索式候補生成、検索結果のタグ付け、要約、論点マップ生成。日本語特許・技術文書の整理に強み。 ⁵⁰	前処理：特許公報・社内技術文書・分類コードをRAG化。後処理：引用文献リストと根拠段落を必ず付す。 ⁵¹	検索式案生成 → 特許DB検索 → 上位ヒット要約 → 類似群分類 → 調査員確認。	抽出漏れ、請求項の過小読解、異表記・同義語の取りこぼし。	200万 ～800 万円
先行技術調査	中	学術文献・報告書・社内資料横断の読解補助。Vision相当機能が使えれば表・図入りレポートの把握に有効。 ⁵²	前処理：PDF OCR整形、表・図・本文の位置情報保持。後処理：AI要約ではなく根拠箇所付きサマリに変換。 ⁵³	文献収集 → OCR/レイアウト抽出 → tsuzumi要約 → 人が必須文献を再読。	図面・数式・化学式は未検証。AIだけで公知性判断をしない。	300万 ～ 1,000 万円
特許出願支援	高	発明の論点整理、課題・作用効果整理、実施例の骨子化、クレーム観点のたたき台生成。NTTも直接展示。 ⁴³	前処理：発明メモ、実験結果、従来技術、社内テンプレートを整備。後処理：専門家がクレーム語彙・サポート要件を手修正。	発明ヒアリング → 先行技術の概観 → 出願ポイント抽出 → ドラフト生成 → 弁理士レビュー。	用語の広狭、サポート要件、実施可能要件、過剰一般化。	300万 ～ 1,200 万円
侵害調査・FTO	低～中	製品仕様と請求項の 初期論点整理 やclaim chartの叩き台には有用。 最終結論には不適 。 ⁵⁴	前処理：製品仕様、クレーム、図面、証拠資料を構造化。後処理：法の見解部分は必ず人が再作成。	クレーム分解 → 製品仕様との対応候補生成 → 根拠証拠紐付け → 弁理士/弁護士判断。	クレーム解釈、均等論、黙示限定、裁判例反映。ハルシネーションは致命的。	500万 ～ 2,000 万円
要約・翻訳	高	日英・英日ドラフト、審査経過、契約書、先行技術文献の要約に有効。tsuzumi 2は日本語・英語性能向上を訴求。 ⁵⁵	前処理：文書種別ごとのスタイル指示。後処理：用語集・定義語・請求項番号整合性をチェック。	文書投入 → 要約/翻訳ドラフト → 用語集照合 → 人の最終修文。	法的定義語、否定表現、クレーム番号・参照番号のズレ。	100万 ～500 万円

業務	適合度	期待される効果	必要な前処理・後処理	推奨運用フロー	人的監査の要点	追加開発コスト概算
契約書レビュー	中～高	チェックリストベースの不足条項検出、修正文案のたたき台、比較レビュー。NTT資料でも契約書案レビュー例あり。 35	前処理：契約種別別のチェックリスト、標準条項、リスクルール。後処理：法務・知財が最終確認。	契約投入 → 条項抽出 → チェックリスト照合 → 修正文案 → 法務確認。	準拠法、責任制限、知財帰属、再委託、秘密保持の例外。	300万～1,500万円
発明者インタビュー支援	高	質問設計、抜け漏れ確認、作用効果の深掘り、発明メモの構造化。日本語対話品質と低コスト運用が相性良い。 56	前処理：業界用語・製品名辞書。後処理：面談録の事実確認と機密マスキング。	面談前質問設計 → 面談中リアルタイム補助 → 面談後サマリ生成 → 発明者確認。	誘導質問、事実誤認、営業秘密の混入。	150万～600万円

上の表から分かる通り、tsuzumi 2 の最も強い適用先は「出願支援」「要約・翻訳」「発明者インタビュー支援」です。これは、公開資料の強みである日本語処理、業務特化、軽量オンプレ、RAG適性が、そのまま効くからです。逆に、侵害調査・FTOは、求められる説明責任の重さに対し、公開検証がまだ不足しています。そこでは、tsuzumi 2 は判断者ではなく、論点整理者・証拠整理者・草案作成者として使うのが妥当です。
57

知財向けに組むべき運用フローは、次のような証拠中心・人手中心の形です。



この構成の狙いは、モデルに「考えさせる」前に、根拠データを限定し、出力後に必ず専門家が監査することです。NTTの知財展示資料がHuman in the Loopを明記していること、日本弁理士会が専門家責任を強調していること、METIが契約時のデータ利用範囲・生成物利用条件・想定外利用をチェックポイントとしていることと整合します。
58

制約・リスクと推奨導入アーキテクチャ

実運用で最初に直面する制約は、精度と説明責任のギャップです。個人情報保護委員会は、生成AIの応答結果には不正確な内容が含まれると明示し、個人情報を入力する場合は、学習利用されないこと等を十分確認するよう求めています。日本弁理士会も、もっともらしい虚偽情報、いわゆるハルシネーションの存在を前提に、AI生成結果を十分吟味する必要があると述べています。知財実務でこれは、誤訳・誤要約よりむしろ、“それらしく見える誤った引用関係”が危険だという意味です。
59

次に大きいのが、**機密データ取り扱い**です。tsuzumi 2 はオンプレ/プライベートクラウド適性が強く、NTT も機微情報ユースケースでの利用を強調しています。これは知財部門にとって追い風です。しかし、**公開API で本当に学習不使用か、ログ保持がどうなっているか、再委託先はどこか、保存期間はどうか**は契約レイヤで確認しなければなりません。METIの契約チェックリストがいう「提供データの利用範囲」「生成物の利用条件」「想定外利用の回避」は、そのまま知財導入に当てはまります。⁶⁰

三つ目は、**ライセンスと提供形態の制約**です。公開Foundryオファーは **Custom license** で、さらに**チューニング機能が利用不可**と明示されています。したがって、知財部門が欲しい**社内案件特化・テンプレート特化・審査官応答スタイル特化**を本格実装したい場合、公開オファーだけでは足りず、**NTTグループとの個別契約、オンプレ提供、あるいは別メニューの確認**が必要です。ここは調達・法務・セキュリティの三者で詰めるべき論点です。⁶¹

四つ目は、**評価指標の偏り**です。NTTの公開評価は、日本語知識、解析、指示遂行、安全性、RAG、金融タスクには強い一方で、**特許新規性判定、クレームチャート整合性、契約差分比較の正確率**のような知財固有指標はまだ公開されていません。したがって、導入前には**社内ベンチマークの自前作成**が不可欠です。最低でも、**出願ドラフト、先行技術サマリ、契約赤入れ、翻訳、インタビュー要約**の五系統で、**正確性、引用率、再現率、修正率、レビュー時間短縮率**を測定すべきです。⁶²

推奨アーキテクチャは、次の四層です。第一に**データ保護層**。ここでは案件ごとに「公開・社外秘・極秘・訴訟準備・営業秘密」を分類し、極秘以上はオンプレまたは専有環境に限定します。第二に**モデル実行層**。テキスト業務は tsuzumi 2、図表や帳票を扱う業務のみ Vision 相当ルートを使い分けます。第三に**監査層**。プロンプト、参照文書ID、モデル版、出力案、レビュー差分をログとして残します。第四に**評価層**。毎月の正答率・レビュー差戻し率・リーガルリスク事案を監視し、再学習やプロンプト修正に反映させます。これは、NTTのAIガバナンス規程類が採る**リスクベース管理**と整合的です。⁶³

推奨される検証プロセスは、PoC前に**禁止入力の定義**、PoC中に**教師なし自由利用を認めないこと**、本番前に**レッドチーム評価と誤答パターン収集**、本番後に**四半期ごとの再認証**です。特に知財部門では、**AIが出した結論ではなく、AIが出した根拠候補**をレビューする設計にしないと、説明可能性を確保できません。⁶⁴

ベンチマーク・比較と導入事例

比較を行う際に最も注意すべきなのは、**tsuzumi 2 の強みは「日本語×軽量×閉域」であり、必ずしも汎用フロンティアモデルの絶対性能競争とは土俵が違う**ことです。したがって、知財部門での比較軸は、①日本語品質、②閉域導入性、③文書理解、④チューニング/カスタマイズ自由度、⑤コスト可視性、の五つに置くべきです。⁶⁵

モデル	開発主体	性格	モダリティ	日本語・業務適性	提供形態	知財部門視点の評価
tsuzumi 2	NTT / NTT DATA	純国産、30B、1GPU、オンプレ適性強	テキスト中心の公開提供を 確認 ²⁰	日本語性能とRAGで強い訴求。金融・自治体・医療知識も学習。 ⁶⁶	NTTグループ経由、Foundry GA、managed compute、custom license。 ²⁰	機密重視の知財部門に最有力。

モデル	開発主体	性格	モダリティ	日本語・業務適性	提供形態	知財部門視点の評価
tsuzumi 2 Vision	NTT	2026-05視覚拡張アップデート	文書 画像・ 図表理解 <small>6</small>	図表入り日本語ビジネス文書に強い。特許図面の厳密理解は未検証。 <small>52</small>	順次提供予定。一般公開APIは未確認。 <small>6</small>	表・帳票・レビュー用途で有望、ただし段階導入向き。
cotomi v3	NEC	国内開発LLM、軽量・高速・安全環境対応	テキスト中心	NECは高い日本語性能、軽量高速、業界特化、AIエージェント対応を訴求。デジタル庁選定あり。 <small>67</small>	専用HW、データセンター、API等。 <small>68</small>	tsuzumi 2の強い国内対抗。比較対象として必須。
Takane 32B	Fujitsu + Cohere	企業向け日本語強化LLM。純国産スクラッチではない	テキスト中心	JGLUE平均0.92、公私分離されたプライベート環境を強調。 <small>69</small>	Fujitsu Kozuchi / DI PaaS、オンプレ系対応。 <small>70</small>	契約・法務・行政用途の比較対象。
PLaMo 2.0 Prime	Preferred Networks	フルスクラッチ国産	テキスト中心	31B系で高い日本語性能、公開価格も比較的明瞭。クラウドAPI、Bedrock、オンプレ、Snowflake。 <small>71</small>	API、オンプレ等。価格は入力60円/出力250円/1M tokens (2025-05時点)。 <small>72</small>	コスト比較の透明性が高く、ベンチマーク比較相手として優秀。
GPT-5.5	OpenAI	汎用フロンティア	マルチ用途	1M context、入力\$5/出力\$30 per MTok。高度推論・コーディング向け。 <small>73</small>	API中心	絶対性能の上限比較には有用だが、閉域・主権・契約条件は別問題。
Llama 4 Maverick	Meta	オープン系・ネイティブマルチモーダル	画像対応	MetaはChartQA 90.0、DocVQA 94.4等を掲示。MoEで効率化。 <small>74</small>	オープンウェイト系エコシステム	Visionベンチの外部比較には参考。ただし企業統制は自前負担。

導入事例・近接事例としては、第一にデジタル庁「源内」試用対象に選ばれたことが大きいです。これは、行政実務での適合性、安全性、ガバナンス・セキュリティ・日本語適合性のシグナルになります。75

第二に、2025年10月の公式リリースでは、**東京通信大学が学内データを学内に留める要件のもと tsuzumi 2 導入を決定した**ことが紹介されています。また、NTTドコモビジネスと富士フイルムビジネスイノベーションは、**契約書や提案書など文字・図・画像を含む非構造化データの分析・活用高度化**に向けた検討を公表しています。後者は、契約レビューや文書アーカイブ知財管理にかなり近いユースケースです。 ³¹

第三に、知財そのものの近接事例として、NTT R&D Forum 2025 の「**tsuzumi 2を活用した特許出願業務支援AIエージェント**」が挙げられます。これはまだ公開事例というより展示・構想段階ですが、**アイデア創出 → 先行技術調査 → 出願ポイント検討 → 明細書執筆 → 出願**というプロセス分解を示しており、知財部門への導入設計図として有用です。 ⁴³

一方で、**公開ベンチマークの不足**は正直に指摘すべきです。tsuzumi 2 は NTT 自身の資料で、同サイズ帯世界トップクラス、RAG世界トップクラス、金融タスクで高性能、GPT-5と多くのタスクで同等程度といった訴求がありますが、**知財固有ベンチマーク**は未公開です。したがって比較表はあくまで**導入可能性の整理**であり、最終判断は自社データでの評価が不可欠です。 ¹⁹

結論と実務向け推奨

総括すると、tsuzumi 2 は「**知財部門のための国産・閉域・日本語重視LLM**」として十分に検討に値する一方、tsuzumi 2 Vision は“**非常に有望だが、API可用性と知財特化検証がもう一段必要**”というのが現時点の厳密な結論です。公開情報の範囲では、tsuzumi 2 は30B・1GPU・オンプレ運用・日本語性能・RAG性能・安全性訴求が明確で、知財と親和性が高いです。他方で、Vision系は文書画像理解に強いものの、**特許の専門図面・化学式・回路図の厳密読解**まで保証する公開データは見当たりません。 ⁷⁶

実務的な推奨は、次の三段階です。

期間	推奨方針	具体策
短期	テキスト中心の閉域PoC	要約・翻訳、発明者インタビュー支援、契約レビューのチェックリスト照合、検索結果要約から着手。公開Foundryオファーまたは閉域提供で、 学習不使用・ログ・権限 を契約確認したうえで導入。 ⁷⁷
中期	RAG統合と部門内ワークフロー化	特許DB、社内過去案件、契約雛形、技術報告書をRAG連携し、 根拠付き出力 を標準にする。出願支援AIエージェントのようなHuman in the Loop型に寄せる。 ⁷⁸
長期	Vision・知財特化・監査自動化	tsuzumi 2 Vision の提供チャンネルが明確になり次第、帳票・表・スキャンPDF対応を追加。並行して知財専用評価セットを構築し、 専門家レビュー差分 から業務ガバナンスを継続改善する。 ⁷⁹

最終的な一文でまとめれば、**NTT tsuzumi 2 は「知財業務を自動化するAI」ではなく、「知財専門家の作業密度を下げ、機密を外に出さず、日本語で伴走するAI」として導入するのが正解**です。短期的には十分実用、中期的にはRAGとガバナンスが勝負、長期的にはVisionと知財特化評価が差別化要素になります。現時点で最も避けるべきなのは、**検索・FTO・契約最終判断をAI単独に委ねること**です。そこを守る限り、tsuzumi 2 は日本の知財実務にかなり現実的な選択肢です。 ⁸⁰

¹ ⁶ ⁷ ¹⁴ ²¹ ²⁷ ⁵² ⁶⁰ ⁷⁹ <https://group.ntt.jp/newsrelease/2026/05/19/260519a.html>
<https://group.ntt.jp/newsrelease/2026/05/19/260519a.html>

2 10 43 44 48 56 58 78 <https://www.rd.ntt/forum/2025/doc/A04-j.pdf>
<https://www.rd.ntt/forum/2025/doc/A04-j.pdf>

3 59 https://www.ppc.go.jp/files/pdf/230602_alert_generative_AI_service.pdf
https://www.ppc.go.jp/files/pdf/230602_alert_generative_AI_service.pdf

4 8 9 11 20 22 23 24 47 55 61 77 <https://ai.azure.com/catalog/models/tsuzumi2>
<https://ai.azure.com/catalog/models/tsuzumi2>

5 12 30 31 32 <https://group.ntt.jp/newsrelease/2025/10/20/251020a.html>
<https://group.ntt.jp/newsrelease/2025/10/20/251020a.html>

13 19 26 28 35 36 50 51 57 62 65 76 <https://group.ntt.jp/newsrelease/2025/10/20/pdf/251020ac.pdf>
<https://group.ntt.jp/newsrelease/2025/10/20/pdf/251020ac.pdf>

15 <https://ai.azure.com/catalog/models/tsuzumi-7b>
<https://ai.azure.com/catalog/models/tsuzumi-7b>

16 18 29 41 42 53 https://journal.ntt.co.jp/wp-content/uploads/2024/05/nttjnl5002_20240601.pdf
https://journal.ntt.co.jp/wp-content/uploads/2024/05/nttjnl5002_20240601.pdf

17 66 <https://www.nttdata.com/jp/ja/lineup/tsuzumi/>
<https://www.nttdata.com/jp/ja/lineup/tsuzumi/>

25 33 34 https://www.rd.ntt/research/LLM_tsuzumi.html
https://www.rd.ntt/research/LLM_tsuzumi.html

37 38 <https://ojs.aaai.org/index.php/AAAI/article/view/29874>
<https://ojs.aaai.org/index.php/AAAI/article/view/29874>

39 <https://aclanthology.org/2023.emnlp-main.839/>
<https://aclanthology.org/2023.emnlp-main.839/>

40 https://www.anlp.jp/proceedings/annual_meeting/2023/pdf_dir/B7-3.pdf
https://www.anlp.jp/proceedings/annual_meeting/2023/pdf_dir/B7-3.pdf

45 46 https://www.youtube.com/watch?v=8ttLPiXQb_Y
https://www.youtube.com/watch?v=8ttLPiXQb_Y

49 54 80 <https://www.jpaa.or.jp/cms/wp-content/uploads/2025/04/Alservices-guideline.pdf>
<https://www.jpaa.or.jp/cms/wp-content/uploads/2025/04/Alservices-guideline.pdf>

63 64 <https://group.ntt.jp/newsrelease/2024/06/07/240607a.html>
<https://group.ntt.jp/newsrelease/2024/06/07/240607a.html>

67 68 <https://group.nec.jp/ja/solutions/ai/llm/>
<https://group.nec.jp/ja/solutions/ai/llm/>

69 70 <https://pr.fujitsu.com/jp/news/2024/09/30.html>
<https://pr.fujitsu.com/jp/news/2024/09/30.html>

71 <https://tech.preferred.jp/ja/blog/plamo-2-31b/>
<https://tech.preferred.jp/ja/blog/plamo-2-31b/>

72 <https://tech.preferred.jp/ja/blog/plamo-2-prime-release/>
<https://tech.preferred.jp/ja/blog/plamo-2-prime-release/>

73 <https://developers.openai.com/api/docs/models>
<https://developers.openai.com/api/docs/models>

⁷⁴ <https://www.llama.com/>

<https://www.llama.com/>

⁷⁵ <https://www.digital.go.jp/news/10d55c63-b3e1-42b9-9cc5-93a06943ae0e>

<https://www.digital.go.jp/news/10d55c63-b3e1-42b9-9cc5-93a06943ae0e>