



NTT版LLM「tsuzumi 2」の性能と評価に関する徹底調査報告書

2025年10月20日、NTTは次世代大規模言語モデル（LLM）「tsuzumi 2」の提供を開始した。本報告書では、tsuzumi 2の技術仕様、性能評価、ベンチマーク結果、業界評価、導入事例、および今後の展望について包括的に分析する。^{[1] [2] [3]}

tsuzumi 2の基本仕様と開発背景

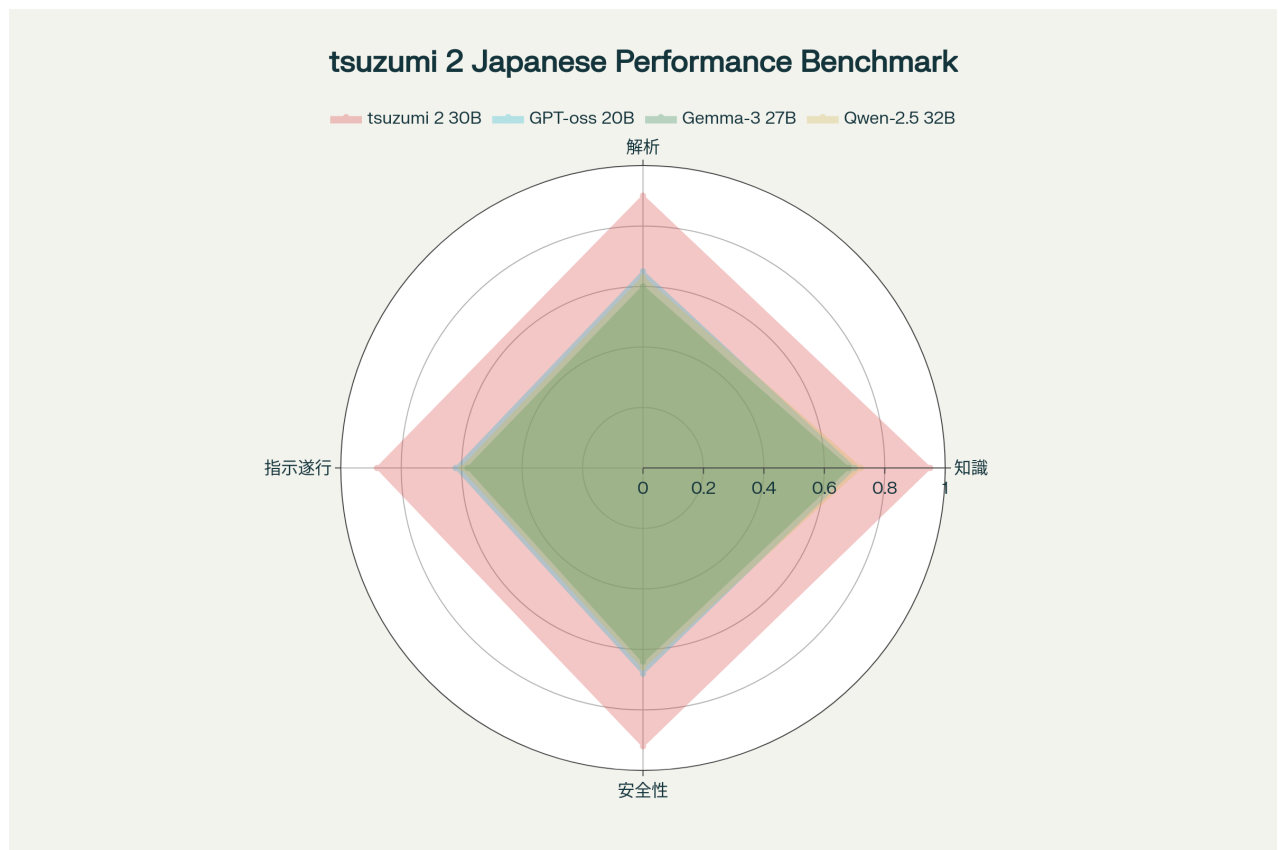
パラメータ数の進化

tsuzumi 2は、2023年11月に発表された初代tsuzumiの後継モデルである。初代tsuzumiのパラメータ数が70億（7B）であったのに対し、tsuzumi 2では300億（30B）へと約4.3倍に拡大された。この大幅な増加にもかかわらず、NTTは「1GPU（A100 40GB相当）で動作可能」という軽量性を維持しており、これがtsuzumi 2の最大の特徴の一つとなっている。^{[2] [3] [4] [5]}

純国産モデルとしての意義

tsuzumi 2はNTTがフルスクラッチでゼロから開発した純国産LLMである。多くの日本語特化LLMが海外製オープンモデルをベースに日本語データで追加学習する手法を採用する中、NTTは学習データ、開発プロセス、システム、運用の全てを自社でコントロールする戦略を選択した。これにより、著作権の完全管理、リリース・ライセンスの柔軟なコントロール、学習データの権利保護が実現されている。^{[3] [6] [2]}

島田明社長は発表会で、「安全保障・産業競争力強化の観点から、各国が自国開発AIを重視する動きが活発化している」と述べ、国産AI開発競争に対して「負けられない」との強い意気込みを示した。^{[4] [7] [3]}



tsuzumi 2の日本語性能ベンチマーク比較（同サイズ帯モデル）

日本語性能の評価とベンチマーク結果

同サイズ帯での圧倒的優位性

tsuzumi 2の最も顕著な特徴は、同サイズ帯のLLMと比較した際の日本語性能の高さである。NTTの評価によれば、知識、解析、指示遂行、安全性の4項目において、GPT-oss 20B、Gemma-3 27B、Qwen-2.5 32Bといった同規模モデルを大きく上回る性能を示した。^{[2] [4] [5]}

木下真吾執行役員は、「同サイズ帯のモデル（Gemma-3等）では世界トップクラスの日本語性能を達成した」と説明している。さらに注目すべきは、パラメータ数が2.3倍のLlama-3.3 70Bや4倍のGPT-oss 120Bといった大規模モデルと比較しても同等のスコアを示した点である。^{[5] [8] [2]}

MT-benchでの評価

代表的なベンチマークであるMT-bench（日本語版）において、tsuzumi 2はGPT-5と多くのタスクで同等レベルの性能を示した。MT-benchは、Writing（文章作成）、Roleplay（役割演技）、Reasoning（推論）、Math（数学）、Coding（コーディング）、Extraction（情報抽出）、STEM、Humanitiesといった多様なタスクで言語モデルの特性を評価するものであり、tsuzumi 2がこれらの幅広い分野で高い性能を発揮したことは、実用性の高さを示している。^[9]

安全性の評価

安全性面でも、tsuzumi 2は優れた評価を獲得している。バイアス・差別・ヘイト・反公序良俗、悪用、情報漏えい、誤情報、擬人化による心理的依存の5項目で評価した結果、同サイズ帯および数倍大きなモデルと比較して高いスコアを記録した。^{[2] [5]}

特化型モデル開発効率の大幅向上

RAG性能の実証

企業や自治体が保有する社内文書を活用する際に重要となるRAG（Retrieval-Augmented Generation：検索拡張生成）において、tsuzumi 2は世界トップクラスの性能を実現している。^[2]^[4]

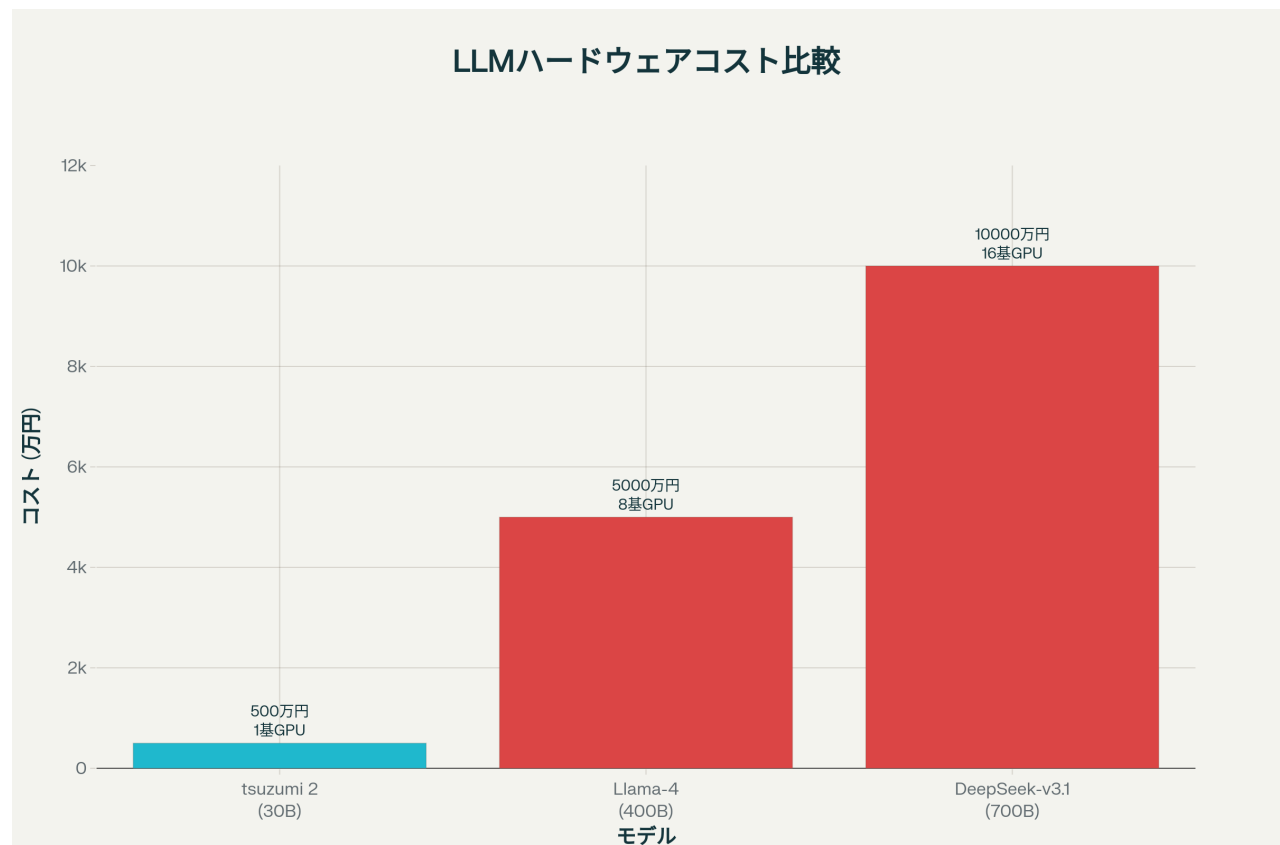
NTT社内での「財務システムに関する問い合わせ対応業務」における評価では、同サイズ帯のGPT-oss 20B、Gemma-3 27B、Qwen-2.5 32Bはもちろん、2.3倍のLlama-3.3 70Bや4倍のGPT-oss 120Bといった大規模モデルと比較しても同等以上の正答率を達成した。^[1]^[2]

ファインチューニング効率の革新

tsuzumi 2が特に高い評価を受けているのが、ファインチューニング（Fine Tuning）における学習効率の高さである。NTTは金融・医療・公共分野の知識を事前に強化しており、これらの分野での特化型モデル開発が大幅に効率化された。^[4]^[8]^[2]

具体的な実証として、ファイナンシャルプランニング技能検定2級試験での検証が行われた。Googleの同規模モデルGemma-2 27Bが合格基準（正解率60%以上）に到達するために1,900問の追加学習を必要としたのに対し、tsuzumi 2はわずか200問の追加学習で正解率64%を達成し、さらにチューニング後には70%に到達した。つまり、**約10分の1のデータ量で同等以上の性能を実現している**。^[3]^[8]^[2]

木下執行役員は、「元々tsuzumi 2のモデルが日本特有の金融データなどを使って開発されているため、追加学習なしでも54%の正答率を示せる」と説明しており、事前学習における専門知識の組み込みが効率化に寄与していることが分かる。^[5]



tsuzumi 2とLlama-4、DeepSeek-v3.1のハードウェアコスト比較

コストパフォーマンスと運用効率

圧倒的な低コスト運用

tsuzumi 2の最大の競争優位性の一つが、運用コストの低さである。300億パラメータでありながら1GPU（A100 40GB相当）で動作するため、推論時のハードウェアコストは約500万円に抑えられる。^{[2] [5] [8]}

これに対し、Llama-4（400億パラメータ）ではNVIDIA H100 80GB相当を8基必要とし、ハードウェアコストは約5,000万円、DeepSeek-v3.1（700億パラメータ）では16基必要で約1億円のコストがかかる。つまり、tsuzumi 2は大規模モデルと比較して**推論コストを10分の1から20分の1に削減**できる計算になる。^{[5] [2]}

オンプレミス環境での運用可能性

1GPUで動作するという特性は、コスト面だけでなくセキュリティ面でも大きなメリットをもたらす。オンプレミスやプライベートクラウドでの運用が可能であり、機密性の高い情報を外部に出すことなく安全に処理できる。^{[4] [10] [2]}

島田社長は、「GPUを増やせばパラメータ数を増やすことができるが、1個のGPUで一定程度の能力を出していくと値段を安くすることができる」と述べ、性能を上げながら価格を抑える戦略を強調している。^[6]

実用性の検証：デモンストレーション結果

契約書チェックと文書改善

発表会では、tsuzumi 2の実用性を示す複数のデモンストレーションが実施された。^[2]

契約書案のチェックでは、社内チェックリストと突き合わせて条項の過不足や問題点を的確に指摘し、表記揺れや全角半角の統一といった細かな点までカバーした改善案を表形式で表示した。^{[11] [2]}

また、企業のニュースリリースの改善提案では、簡潔な表現への修正や全角半角の混在といった重要な改善点を適切に指摘した。^[2]

指示遵守性の優位性

興味深いのは、他のLLMとの比較デモである。ニュースリリースの改善点を「リスト形式」で出力するよう指示した際、OpenAIのGPT-oss 20Bが指示を守らず表形式で出力したのに対し、tsuzumi 2はリスト形式で正確に出力した。このような指示遵守性の高さは、ビジネス現場での実用性を大きく左右する要素である。^[11]

情報抽出とフォーマット指定

JSONフォーマットでの情報抽出タスクでも、tsuzumi 2は高い精度を示した。NTT研究所採用ページのソースコードからリンク情報を抽出する際、プロンプトに「NTT」を「NT」と誤記した場合でも文意を正しく理解し、適切にJSON形式で出力した。^[9]

業界からの評価と専門家の見解

肯定的評価

情報通信研究機構（NICT）の鳥澤健太郎フェローは、初代tsuzumiについて「大したものだ」と性能を高く評価している。鳥澤氏はNICTのLLM開発責任者であり、業界では名の知れた人物である。同氏は「小型化というビジネス戦略も納得がいく。OpenAIと同じ路線を取れば、金銭的に体力のある方が勝つだけだ」とNTTの戦略を支持している。^[12]

島田社長は、「AIを導入するというのはコンサルティングとセットでやっていかないと、お客様の持たれているデータと正しい組み合わせをしないとなかなか正しい答えが出てこない」と指摘し、顧客に合った形でのAI導入を重視する姿勢を示している。^[6]

NTTのAI事業実績

NTTのAI関連事業は急速に成長している。2025年第1四半期（1Q）時点でのtsuzumi関連の受注件数は1,827件（国内521件、海外1,306件）に達し、受注実績は670億円を記録した。2025年度は前年度比244%増の1,500億円、2027年度には5,000億円超を見込んでいる。^{[3] [13]}

受注構成比では、国内では公共が約35%、金融が約20%を占めている。様々な業界、事業規模の顧客から受注しており、AIがITソリューションビジネスを加速させている状況が確認できる。^{[13] [3]}

批判的視点と課題認識

一方で、初代tsuzumiに対しては批判的な評価も存在する。個人のブログ記事では、GPT-4oが77%正解、GPT-4が53%正解、GPT-3.5 turboが12%正解したテストにおいて、初代tsuzumiが4%しか正解できなかったという検証結果が報告されている。^[14]

この評価者は、「入力プロンプトが増えれば増えるほど挙動が怪しくなる」「RAGのような大量の文章を入れるのはやめたほうがいい」と指摘し、実用性に疑問を呈している。ただし、この評価は初代tsuzumi（7B）に対するものであり、パラメータ数が4倍以上に増加し、複雑な文脈理解能力が大幅に向上したtsuzumi 2では状況が異なる可能性が高い。^[14]

鳥澤氏も「LLM開発はギャンブル。どんなLLMが勝ち残るのか、まだ誰もわからない」と述べており、国産LLMの将来性については不確実性が残る。^[12]

導入事例と実装状況

東京通信大学での全面導入

tsuzumi 2の提供開始に先駆けて、東京通信大学が教育機関として初めて導入を決定した。同大学は2024年9月にNTT東日本と連携協定を締結しており、生成AIの教育活用に積極的に取り組んでいる。^{[15] [16] [17] [18]}

東京通信大学がtsuzumi 2を選定した理由は、「学生・教職員のデータを学内に留める」という要件のもと、クラウド依存のない国産LLMを核に学内LLM基盤を整備できる点にある。tsuzumi 2により複雑な文脈理解や長文ドキュメント処理が安定し、複雑なタスクの複合工程においても実用水準に達しているとの判断から導入を決定した。^{[1] [16] [15]}

具体的な活用場面としては、以下が計画されている：^{[16] [17] [1] [15]}

- 授業Q&Aの高度化：学生からの質問に対し、即座に答えを提示するのではなく、対話を重ねながら最適な答えを導き出すコミュニケーションが可能となり、学生の思考力や主体性を引き出す学習サポートが実現される
- 教材・試験作成支援：教職員の業務負担を軽減しつつ、質の高い教育コンテンツの作成を支援
- 履修・進路相談のパーソナライズ：個々の学生のニーズに応じた細やかな相談対応が可能に
- レポート評価の効率化：400～500人分のレポート評価を効率化し、教育的な指導に時間を割けるようにする^[19]

東京通信大学の担当者は、「評価基準を決めてあげれば、AIの評価も結構なクオリティができそうな感触がある」と実用レベルでの活用可能性を評価している。^[19]

富士フイルムビジネスイノベーションとの連携

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、tsuzumi 2を活用した新たな生成AIソリューションの提供に向けた検討を開始した。^{[20] [1]}

富士フイルムビジネスイノベーションは、AI技術ブランド「REILI」のもと、企業内に散在する契約書や提案書など、形式の異なる文字・図・画像を含む非構造化データを構造化し、価値ある情報として活用可能にする技術を展開している。^{[21] [1]}

機密情報を安全に取り扱いながら高度な推論を行える「tsuzumi 2」と、富士フイルムビジネスイノベーションが持つデータの構造化技術を組み合わせることで、企業内の文書・画像データの分析や活用をさらに高度化する生成AIソリューションの提供をめざしている。^{[1] [20]}

技術的特徴と革新点

リーズニング機能の搭載

tsuzumi 2には、事前にタスクの詳細を精査してから回答を出力する「リーズニング機能」が搭載されている。これにより、生成AIがもっともらしい嘘をつく「ハルシネーション」を低減できるとされている。^[4]

マルチモーダル対応の計画

現時点ではテキスト中心の処理に特化しているが、将来的には視覚情報や聴覚情報、さらにはユーザーの状況を理解して応答を生成できる機能の追加が計画されている。^[22]

アダプタによる柔軟なチューニング

tsuzumiではアダプタと呼ばれる効率的なチューニング手法を採用している。モデル全体を再学習させることなく、業界固有の用語や特定の業務に迅速に最適化が可能であり、複数のアダプタを利用するマルチアダプタ機能により、組織ごとに異なるニーズにも柔軟に対応できるよう設計されている。^[22]

ソブリンAIとしての戦略的位置づけ

国家安全保障とAI主権

tsuzumi 2の開発で意識されているのが、データやシステムを国内インフラ内で完結させる「ソブリンAI」の概念である。^{[3][7]}

NTTの資料によれば、真に守るべきナレッジやインテリジェンスを扱うコア領域、専門性が高い業務、国・自治体・インダストリー共通的な汎用業務においては、「海外流出すると経済安全保障の観点でリスクに直結」するとされている。^[3]

各国の動向を見ると、日本政府はAI基本計画で「質の高い日本語を整備し、日本文化や慣習を理解した信頼できるAI開発」を掲げ、経済産業省はGENIACを推進している。米国は2025年7月に「Winning the AI Race: America's AI Action Plan」を発表し、AIイノベーションの加速、インフラ構築、AIにおける米国のリーダーシップ確立を目指している。^[3]

このような国際環境の中で、NTTは以下の点を強調している：^[3]

- 国産スクラッチ開発により、仕様・品質を自ら決定し、開発プロセスをフルコントロール
- リリース・ライセンスをコントロールして安定的に提供
- 学習データのコントロールにより権利保護（新聞等のデータを学習データから自主的に削除）
- 質の高い日本語学習データ、日本文化、慣習を理解
- データ・システム・運用の観点で自国・自社コントロール

権利保護への配慮

注目すべきは、米国での新聞社からの提訴の実態等を踏まえ、権利者の保護も意識し、新聞等のデータを学習データから自主的に削除した点である。これは、学習データの完全コントロールが可能な純国産モデルならではの対応と言える。^[3]

市場での位置づけと競合状況

国産LLMの中での評価

ZDNET Japanが2024年4月に実施した読者アンケート「注目している大規模言語モデル（LLM）」では、OpenAIの「GPT-4」が42.3%と最も多く、次点はGoogleの「Gemini」（17.4%）であった。以下は、NTTの「tsuzumi」（11.8%）、NECの「cotomi」（9.6%）、Metaの「Llama 2」（4.5%）という結果であった。^[23]

国産LLMとしてはtsuzumiがトップの評価を得ており、高い日本語性能へのニーズや国産LLMへの期待がうかがえる。^[23]

他の日本語LLMとの比較

日本国内では複数の日本語特化LLMが開発されている。主要なものとして、NTTのtsuzumi、NECのcotomi、サイバーエージェントのCyberAgentLM3、ELYZAのLlama-3-ELYZA-JP、楽天のRakuten AI 7Bなどがある。^[24]

tsuzumiの特徴は、パラメータサイズが約6億～70億（tsuzumi 2では300億）と比較的軽量でありながら、世界トップレベルの日本語処理能力を実現している点である。Rakudaベンチマークでの評価

では、tsuzumiはGPT-3.5と比較して81.3%の勝率を示した（2024年3月時点）。^{[25] [24]}

今後の展開と課題

事業拡大の見込み

NTTは、AI関連事業の受注額を2025年度に1,500億円、2027年には5,000億円超と予測している。2025年第1四半期の受注実績670億円からの急成長が期待されている。^{[13] [26]}

国内では公共や金融分野での採用が進んでおり、海外からもロボット活用に関する問い合わせがあるなど、海外市場での需要にも期待が寄せられている。^[26]

産業横断的な活用構想

NTTは、tsuzumi 2を企業単位の活用からインダストリー全体への展開を目指している。^[21]

例えば、トヨタとの協働による「モビリティAI基盤」では、交通事故ゼロ社会の実現に向けて、分散型計算基盤、インフラ協調AIモデル、インテリジェント通信基盤の三位一体でデータドリブンによる自動運転の高度化を進めている。^[21]

また、トライアルホールディングスとの連携では、製造・流通業界のサプライチェーン全体を最適化する「連鎖型AI」の開発を進めており、流通サプライチェーンにおける発注・棚割最適化により、在庫圧縮・廃棄削減、店舗作業コスト低減を実現している。^[21]

技術的課題と改善の方向性

初代tsuzumiに対する批判的評価が示すように、長文処理やRAG性能については引き続き改善の余地がある可能性がある。ただし、tsuzumi 2では複雑な文脈理解と長文処理能力が大幅に向上したとされており、これらの課題は相当程度改善されていると考えられる。^{[2] [4] [14]}

NTTは、お客様ニーズを研究開発へフィードバックし、基盤モデルをアップグレードする「開発－顧客のサイクル」を構築している。この継続的改善プロセスにより、実用性の更なる向上が期待される。^[3]

エコシステムの構築

NTTは、AIをサステナブルに実現するインフラとして、GPU as a Service、NaaS（Network as a Service）、データセンター、電力供給を統合的に提供する体制を整えている。^[21]

特に注目されるのが、米Gartner社がクラス最高の技術革新を推進する世界中のCSP（Communication Service Provider）を表彰する「Eye on Innovation Awards」のAPAC地域において、NTTドコモビジネスが日本に本社を置く企業として初めて"Winner"に選出されたことである（2025年10月発表）。これは、NTTのAIインフラが世界的に高い評価を受けていることを示している。^[21]

結論：tsuzumi 2の総合評価

tsuzumi 2は、以下の点で高い評価を受けている：

技術的優位性

- 同サイズ帯LLMで世界トップクラスの日本語性能

- 大規模モデルと同等の性能を1/10～1/20のコストで実現
- ファインチューニングに必要なデータ量を約1/10に削減
- オンプレミス環境での運用可能性

戦略的価値

- 純国産モデルとしてのソブリンAI実現
- 学習データ・開発プロセスの完全コントロール
- 金融・医療・公共分野での専門知識の強化

実用性の証明

- 東京通信大学での全面導入決定
- 富士フイルムビジネスイノベーションとの連携開始
- 受注件数1,827件、受注額670億円（2025年1Q）

一方で、以下の課題も認識されている：

- 海外の巨大LLMとの絶対的な性能差（特にGPT-4、GPT-5レベルとの比較）
- 初代tsuzumiに対する批判的評価の存在
- LLM開発競争の不確実性

しかし、NTTが目指すのは「何でも知っている1つの巨大なLLM」ではなく、「専門知識を持った小さなLLMを複数組み合わせる戦略」である。この「軽量・高性能・低コスト」という差別化戦略は、特に日本国内の企業や自治体、そして機密性の高い情報を扱う組織にとって、極めて魅力的な選択肢となっている。^[27]

tsuzumi 2は、日本語処理性能、コストパフォーマンス、セキュリティ、専門知識の効率的な学習という4つの要素をバランス良く実現した、実用的な国産LLMとして評価できる。今後の継続的な改善と実装事例の蓄積により、日本のAI活用を支える重要な基盤技術として発展していくことが期待される。



1. Geng-naruJin-Hua-woSui-ketaNTTBan-LLM-tsuzumi-2noTi-Gong-Kai-Shi-Ri-Ben-noQi-Ye-DX.pdf
2. 251020ac.pdf
3. 251020ab.pdf
4. <https://www.itmedia.co.jp/aipplus/articles/2510/20/news101.html>
5. <https://news.mynavi.jp/techplus/article/20251021-3569988/>
6. <https://japan.cnet.com/article/35236405/>
7. <https://news.yahoo.co.jp/articles/f23840bb4a8c08978639897a00e1c7e2e2d9e02f>
8. <https://aidiver.jp/article/detail/128>
9. NTTBan-Da-Gui-Mo-Yan-Yu-moteru-tsuzumi-2-_-NTT-R-D-Website.pdf
10. <https://k-tai.watch.impress.co.jp/docs/news/2056478.html>
11. <https://www.businessinsider.jp/article/2510-ntt-tsuzumi-2-unveiled/>

12. <https://business.nikkei.com/atcl/gen/19/00537/112200043/>
13. <https://businessnetwork.jp/article/30842/>
14. <https://pickerlab.net/2024/12/07/llm-tsuzumi/>
15. <https://ict-enews.net/2025/10/21internet/>
16. <https://www.agara.co.jp/article/554847>
17. <https://www.shijyukukai.jp/2025/10/29013>
18. <https://www.internet.ac.jp/news/news-110339/>
19. <https://aidiver.jp/article/detail/128?p=2>
20. <https://group.ntt.jp/newsrelease/2025/10/20/251020a.html>
21. 251020ad.pdf
22. <https://ai-market.jp/services/tsuzumi/>
23. <https://japan.zdnet.com/article/35218408/>
24. https://dify.tdse.jp/post_column/201/
25. <https://journal.ntt.co.jp/article/26659>
26. <https://www.sankei.com/article/20251020-MWJ4HXBOPJKDXJA3W5W4NHEXWY/>
27. <https://journal.ntt.co.jp/article/24521>
28. <https://group.ntt.jp/newsrelease/2025/10/20/pdf/251020aa.pdf>
29. <https://codezine.jp/news/detail/22414>
30. <https://news.yahoo.co.jp/articles/d54ce8019afdf695dc7c41fc028577aeab7d66a9>
31. <https://journal.ntt.co.jp/article/33345>
32. <https://x.com/hawkymisc/status/1980253012066218222>
33. <https://ai-market.jp/technology/llm-japanese/>
34. https://ledge.ai/articles/semianalysis_inferencemax_real_time_benchmark
35. https://note.com/wandb_jp/n/n4ee7d7b2bebe
36. https://zenn.dev/k_eclair39/articles/0a55081ef3d2c4
37. <https://prtimes.jp/main/html/rd/p/000000804.000011137.html>
38. <https://bizzone.jp/article/detail/12213>
39. <https://note.com/bostonmio/n/n4628f152497a>
40. https://www.rd.ntt/research/LLM_tsuzumi.html
41. <https://www.ntt.com/bizon/tsuzumi.html>
42. <https://digitalpr.jp/r/120589>
43. <https://news.yahoo.co.jp/articles/43ef0a2f4b8e9d90afc0dab0aaf19fdc382bfe44>
44. https://ledge.ai/articles/ntt_tsuzumi2_fullscratch_gpt5_level_japanese_llm
45. <https://japan.cnet.com/article/35239429/>
46. <https://note.com/happydigital/n/nbcfe4c14262f>
47. <https://www.watch.impress.co.jp/docs/series/nishida/1544492.html>
48. <https://hblab.co.jp/blog/chatgpt-gemini-claude-comparison/>
49. https://note.com/tank_ai/n/n62fb8a610106

50. <https://www.ntt.com/business/dx/smart/generative-ai/basic.html>
51. <https://www.nttdata.com/jp/ja/-/media/nttdatajapan/files/lineup/tsuzumi/tsuzumionazuremaasv10-2.pdf>
52. <https://www.youtube.com/watch?v=w0zOq2GZ4wA>
53. <https://hypervoice.jp/xJOXZ1ux>
54. https://journal.ntt.co.jp/wp-content/uploads/2023/12/JN202401_2.pdf
55. <https://news.livedoor.com/topics/detail/29820035/>
56. <https://www.tokyo-np.co.jp/article/443784>
57. <https://www.nikkei.com/article/DGXZQOUC066YX0W5A800C2000000/>
58. <https://news.livedoor.com/article/detail/29820035/>
59. https://www.soumu.go.jp/johotsusintokei/linkdata/r06_01_houkoku.pdf
60. <https://www.okinawatimes.co.jp/articles/-/1695249>
61. <https://group.ntt.jp/newsrelease/2025/10/20/pdf/251020ab.pdf>
62. https://subway.osakametro.co.jp/news/news_release/20250402_ugo_yumeshima.php
63. <https://www.ntt.com/bizon/d/00707.html>