



# tsuzumi: 国産大規模言語モデルの挑戦

膨大な計算リソースやセキュリティリスクといった課題を抱えるLLM (Large Language Models) に対し、NTTは軽量かつ高性能な日本語LLM「tsuzumi」の研究開発に取り組んでいます。本稿では、日本語に最適化したトークナイザやアラインメント学習、ソフトウェア開発能力の向上、および視覚能力の拡張といった技術的な取り組みと今後の展望について紹介します。

キーワード：#生成AI、#大規模言語モデル、#ソブリンAI

にしだ こうすけ / すぎやま ひろあき  
 西田 光甫 / 杉山 弘晃  
 かざと ひろし / はせがわ たく  
 風戸 広史 / 長谷川 拓  
 にしだ きょうすけ  
 西田 京介

NTT 人間情報研究所

## 国産ソブリンAIの挑戦

近年、ChatGPTをはじめとするLLM (Large Language Models) への注目が高まる一方で、膨大な学習データと計算リソースを必要とする従来のLLMは、電力消費や運用コストの増大、機密情報の取り扱いにおけるセキュリティリスクといった課題を抱えています。NTTは、これらの課題に対応すべく、軽量でありながら優れた日本語処理性能を持つ「tsuzumi」を2023年に発表しました。国内企業・自治体等でのAI (人工知能) の普及を推進する中で、実際のビジネス現場でのAI活用において、特に企業・自治体が保有する複雑なドキュメントへの理解や専門的な知識への対応力強化等のご要望を多数いただきました。このようなニーズを研究開発へフィードバックし、次世代モデル「tsuzumi 2」を開発し、2025年10月に提供開始しました。

特に、tsuzumi 2は、日本語性能においては、同サイズ帯のモデルと比較して世界トップクラスの性能を実現しています。ビジネス領域で主に重視される知識、解析、指示遂行、安全性の基本性能では、数倍以上大きなフラッグシップモデルに匹敵するレベルを達成し、コストパフォーマンスに優れています (図1)。

NTTでは、お客さまからのニーズにおこたえできる性能に加えて環境負荷とコストを抑え、企業・自治体等のAI活用を加速できるように進めています。本稿では、tsuzumiプロジェクトに関する取り組みについて紹介します。

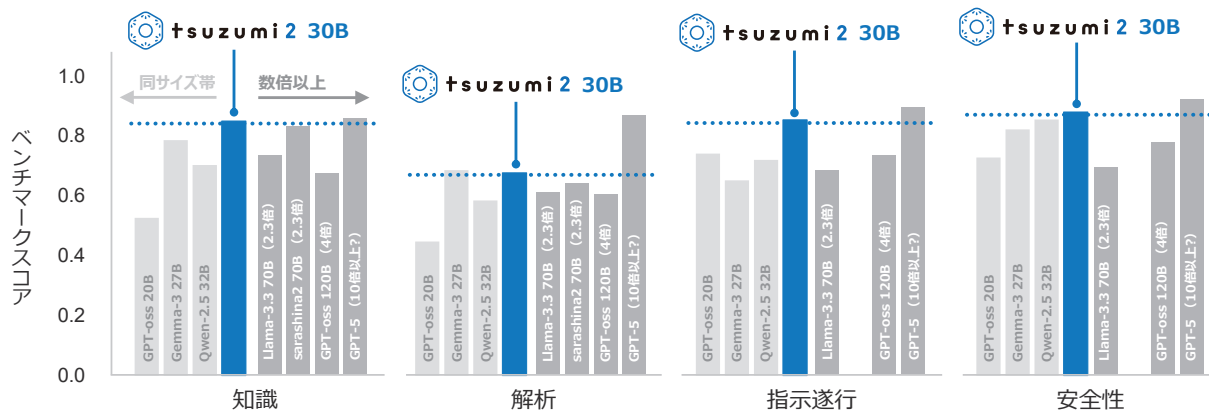
## 日本語に強いトークナイザと事前学習

LLMが自然言語を処理する際、テキストをそのまま扱うのではなく、「トークン

と呼ばれる最小単位に分割するトークナイザが必要となります。トークナイザの設計は、モデルの言語理解・生成能力や推論時の処理速度に直接的な影響を及ぼします。

特に日本語は、英語のような空白区切りが存在せず、漢字・ひらがな・カタカナが混在する複雑な構造を持つため、日本語について考慮していないトークナイザではトークンが文字やバイトに分割されたり、日本語の構造に基づいていないトークンに分割されたりする課題があります (図2)。

そこで私たちは、日本語の言語特性に最適化したトークナイザの構築に取り組んでいます。日本語の文法・語彙や頻出する複合語に基づいた語彙に基づく独自のトークナイザを構築することで、日本語におけるトークン効率 (1トークン当りの文字数) を向上させ、より少ないトークン数で深い意味理解を可能にする工夫を凝らしています。私たちのトークナイザでは、日本語は



※評価条件  
 ・知識: llm-jp-eval (JAQKET (日本語QAタスク), NIIIC (質問応答), JCommonsenseQA (常識推論能力)), pfgen-bench (PFN日本知識ベンチ) の平均値  
 ・解析: llm-jp-eval (Wikipedia読み推定, Wikipedia係り受け解析) の平均値  
 ・指示遂行: M-IFEval\_Ja  
 ・安全性: AnswerCarefully (v0.2)  
 ・知識と解析においては、事前学習モデルを使用。ただし、GPT-oss 20B/120B、GPT-5は事前学習モデルが公開されていないため事後学習モデル (Reasoningモデル) を使用  
 ・指示遂行と安全性において、すべて事後学習モデルを使用。sarashina2 70Bは事後学習モデルが公開されていないため対象外

図1 日本語ベンチマークの性能

翌日は何事もなかったかのように  
彼はいつものカフェでコーヒーを注文した。

### GPT-5（語彙数200k）で分割

翌日は何事もなかったかのように  
彼はいつものカフェでコー ■ ■ を注文した。

### tsuzumi 2（語彙数128k）で分割

翌日は何事もなかったかのように  
彼はいつものカフェでコーヒーを注文した。

NTTの形態素解析ノウハウを用いて  
日本語文法を考慮した分割

日・英・24言語平均の1トークン当りの文字数

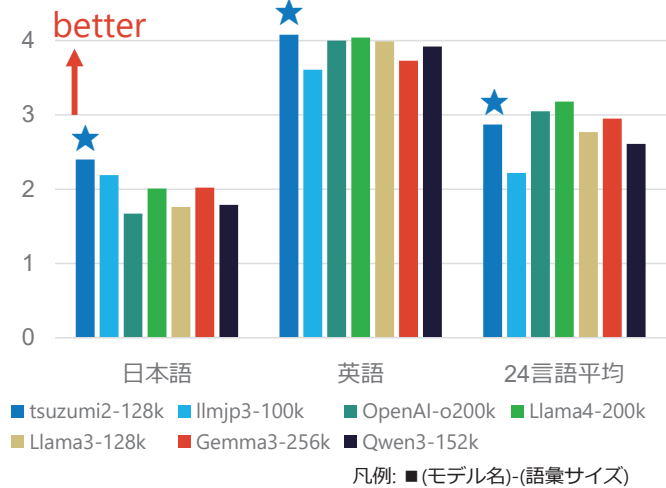


図2 トークナイズ例および性能

もちろん、多言語に対しても優れたトークン効率を実現しました。

そして、LLM構築の最初のステップとして、このトークナイズを用いて学習コーパスを前処理して事前学習を行いました。私たちは、Web上の膨大なコーパスから、高品質な日本語テキストを抽出するために独自のフィルタリングアルゴリズムを開発し、ノイズの除去とデータの純度向上を徹底しています。このように、データの量だけでなく、日本語としての「質」と「量」の両方にこだわった事前学習を行うことで、日本に関する知識を正確に身に付け、日本語の言語処理について高い能力を持たせることに成功しました。

## アラインメント学習

事前学習を通じて知識を溜め込んだtsuzumiに、「知識をどう人間の役に立たせて表現するか」を教えるのがアラインメントの役割です。アラインメントをどう設計するかは、LLMのできることで使い勝手に直結するため、学習における重要な要素といえます。

tsuzumiのアラインメントは特に、①日

本語が得意なLLM、②小型かつ高性能なLLM、③ビジネスで使えるLLM、の3つをめざして設計しています。日本語が得意なLLMについては、日本の知識についてよく答えられることはもちろん、特に日本語における指示追従性に注力しました。例えば日本語の特徴的な点の1つに、漢字・平仮名・片仮名と同じことばを複数の方法で表記できる点があります。どの表記を使うべきかは状況によって異なるため、ユーザが指示したとおりに制御する必要があります。このような日本語ならではの問題に注力することによって、日本のユーザに使いやすいと感じてもらえるLLMをめざしています。

またLLMの性能は、ユーザへ応答するためにどこまで広く深く考えられるか、つまりLLMの思考能力を決定づけるといえます。私たちは、小型であっても高い思考能力を持つLLMをめざしてアラインメントを行いました。具体的には、思考能力が重要となる分野として代表的な数学タスクに注目し、数学能力の向上に取り組んでいます。さらに、LLMの重要なユースケースである、検索システムと併用するRAG (Retrieval-Augmented Generation: 検索拡

張生成)についてはアラインメントの中で特に重視しています。今後も、実サービスで使われる中で得られたフィードバックを基にして継続的な改善を行い、tsuzumiがより皆様の役に立つことをめざしていきます。

## ソフトウェア開発能力の向上

LLMが処理できる言語やタスクは、日本語や英語などの自然言語を対象としたものだけではありません。大量のソースコードや開発ドキュメントを学習することにより、LLMはソフトウェアを開発できる能力を獲得します。私たちは、AI分野の研究開発やベンチマークでよく用いられるPythonに加えて、Java、C、C++、C#、JavaScript、TypeScriptといった国内の商用ソフトウェア開発でよく利用されるプログラミング言語や、レガシーシステムで利用されモダン化に期待の集まるCOBOLなどの学習を行っています。LLMのコーディング能力を高めるためには学習データの量とともに品質が重要であり、NTT独自の事前処理やフィルタリングを行って学習データの品質を高める工夫をしています。

ヒトのように思考を言語化するだけでなく、計算機で実行可能なコードやアルゴリズムとして形式的に表現し、それを実際に実行することで、LLMは現実世界の電子データやICTサービスと相互作用することが可能です。会話する相手はヒトから開発ツールや外部システムに変わり、LLMは一定の開発目的を達成するまで、自律的に思考や環境との相互作用を繰り返します。このような活用形態はソフトウェア開発エージェントと呼ばれており、開発環境に統合されるかたちでソフトウェア開発の現場で急速に普及しています。私たちは、LLMのプログラミング能力や推論能力を継続的に向上させていくことで、外部ネットワークから隔離された開発環境や機微情報を扱うプロジェクトにおいてもソフトウェア開発エージェントを通じて高度なタスクを自動化し、飛躍的な生産性向上をめざします。

### 評価体制の構築

LLMの性能を正しく評価することは、LLMの性能改善の方針を定めたり、ビジネス上の意思決定を行ううえで、非常に重要な営みです。しかし、LLMが持つ能力は多岐にわたり、列挙しきることは容易ではありません。またLLMの能力は日々飛躍的に高まっており、それらに対応するため日々新たなベンチマークが提案されています。比較すべきモデルも非常に多く、特に登場直後のモデルを評価する場合、コミュニティ全体に知見が存在しないバグに悩まされることもあります。

そうした難しい環境において、私たちはLLM評価の安定性・再現性を高めるため、多様なベンチマーク・モデル群を、統一的な形式で評価するためのソフトウェアを開発しています。個々のベンチマークの違いを吸収するアダプタ部と、共通処理を実行するバックボーン部を分離する設計とし、新規ベンチマークに対しても半日足らずで導入することができます。モデルごとの違

いについても、市中のOSSライブラリを積極的に利用して対応稼働を低減しつつ、それらに対応しきれない新規・マイナーなモデルを動かすフォールバック系も用意することで、即日での対応が可能です。また、ユーザが求めるLLMの能力の中には、評価ベンチマークが十分には整備されていない領域も多々残っています。私たちは実際の事業で要望の高い領域について、新たなベンチマークを開発し正しく評価できる体制を整えることで、LLMの性能改善やビジネス上の意思決定に貢献していきたいと考えています。

### LLMの視覚拡張

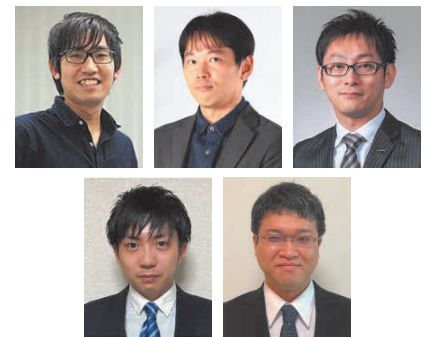
LLMは通常、自然言語だけを入力として受け取り、自然言語で応答しますが、視覚エンコードとプロジェクトと呼ばれるモジュールを追加することで、画像も入力として扱えるようになります。視覚エンコードは、画像に何が映っているかという視覚情報を処理する役割を担います。具体的には、入力された画像を数値演算に適したベクトル（数値列）に変換します。さらに、プロジェクトと呼ばれるモジュールを用いて、このベクトルをLLMのトークンに対応するベクトルへと変換することで、画像の情報をLLMが理解できるようになります。このように画像を入力として扱えるLLMは、一般にLVLM (Large Vision-Language Model) と呼ばれます。LVLMは自然画像だけでなく、文書画像や医療画像など、さまざまな画像を扱うことが期待されています。

私たちは特に、日本語の文字や図表を含む文書画像やスライド画像を扱う能力の向上を目的として、それらを含む幅広い学習データを構築し、学習を行っています。その結果、文字の読み取りにとどまらず、図の幾何的な意味や表の行・列といった概念についても理解できる能力を獲得しつつあります。今後はさらなるデータの収集と学

習を進めることで、オフィス文書の構造化やスライドの要約など、ビジネスのあらゆるシーンで利用可能なLVLMをめざしていきます。

### 今後の展開

私たちは今後もLLMの研究開発を深化させ、NTTグループが展開する多様なエージェントAIサービスへと積極的に適用していきます。実社会での運用を通じて得られる実用上の課題を、迅速に技術改善へとフィードバックするサイクルを回すことで、理論と実践の両輪から真のイノベーションを創造していきます。人間とAIが自然に、そして高度にコミュニケーションし合える世界の実現に向け、NTTはこれからも最先端の研究開発に挑戦し続けます。



(上段左から)西田 光甫/ 杉山 弘晃/  
風戸 広史

(下段左から)長谷川 拓/ 西田 京介

競争の激しい分野ですが、tsuzumiの研究開発を通じて日本の産業、学術を支える大規模言語モデルの実現に挑戦します!

#### ◆問い合わせ先

NTT人間情報研究所  
思考処理プロジェクト