（報道発表資料）

2026.11.12

NTT 株式会社

# 少量データから特定の専門性を有した LLM にチューニングする技術を確立

## ～業務固有の知識を持った個社オリジナルの tsuzumi をもっと容易に提供～

発表のポイント：

- ◆ LLM に業務全般にかかる知識を覚え込ませるには超膨大なデータが必要であるのに対し、各社が所有するドキュメント量は学習データとして圧倒的に不足しており、チューニング時の障壁となってきた点を解決
- ◆ 業務マニュアル等の知識を LLM に定着させるのに、LLM 自身が多様な観点から異なる学習データを自動生成する方法を用いて実現するチューニング技術を開発
- ◆ 業務固有の知識を有した個社オリジナルの LLM を、超膨大なデータを収集することなく作成することが可能となる

NTT 株式会社（本社：東京都千代田区、代表取締役社長：島田 明、以下「NTT」）は、業務マニュアルの知識を LLM に定着させるための新たなチューニング技術を開発しました。知識の定着には、知識に関する極めて膨大なデータを用意する必要があり、業務特化の LLM を構築するための障壁となっています。本技術では、LLM 自身が LLM 向けの学習教材を作成するかのように業務マニュアルから学習データを生成し、それを通して学ぶことで知識を高い精度で定着させます。これにより、特定の業務知識を有した専用の LLM を従来よりも少ない準備で構築できるようになります。本技術は NTT が開発する tsuzumi を始め、他の LLM にも適用可能です。本研究成果は、2026 年 11 月 18 日～25 日に開催される NTT R&D FORUM 2026 [※1] に展示予定です。

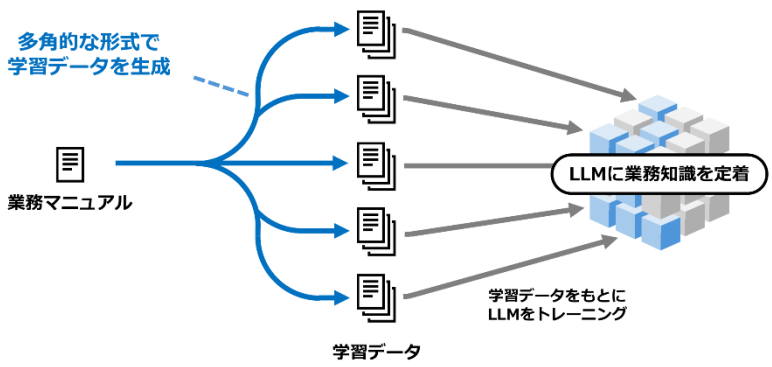



図 1. 技術概要：人間が知識をより深く定着させるために多様な教材を通して学ぶように、知識の活用のされ方を想定した様々な形式の学習データを生成し、LLM チューニングを実施

## 1. 背景

既存の LLM（大規模言語モデル）を特定の用途やタスクに合わせて調整する作業のことをチューニングと呼びます。チューニングには、出力の形式を整える、特定タスクの精度を向上させる、といった幾つかの観点があり、目的に応じた手法が用いられます。

その中でも特に難易度が高いのが、LLM に新たな専門知識を定着させ、業界特有の知識に詳しいモデルを作ることです。例えば、業務マニュアルや専門書の内容をモデルに覚えさせたい場合がこれにあたります。

知識の付与には、モデル全体を再学習させる手法が一般的ですが、これには極めて膨大な量のデータが必要です。A4 用紙のテキスト1ページが数キロバイトである中、数十ギガバイト（1千万倍）を超えるようなデータを用います。

実際の業務現場で、これほど大量の業務マニュアルや資料を用意するのは困難です。インターネット上から類似情報を集めようとしても、業務と無関係な情報が混入しやすく、高品質なデータセットを構築するのは容易ではありません。そのため、各社が所有する限られた業務資料のみから、LLM に効率的に知識を定着させるためのチューニング手法が求められています。

## 2. 技術のポイント

本技術は、お客様問い合わせマニュアルなどの業務知識を記載したドキュメントを活用し、LLM がその内容を深く理解できるように設計されています。そのポイントは多角的な学びです。例えば学校の国語の授業では、文章を読むことに加え、著者の意図を理解するための設問を用意したり、生徒間で感想を述べ合ったりと様々な様式で学びを深めます。本技術はこのような着想に基づいて、LLM 自身に業務マニュアルを学ぶための教材（学習データ）を多角的に構築させ、自らの学びに活用させます。業務マニュアルから、仕様書、FAQ といった多様な表現形式のデータを生成するなどします。こうして生成されたデータを用いて学習を行うことで、従来よりも高精度にマニュアルに記載された知識を獲得することができるようになりました。

## 3. 実験の概要

NTT の実業務マニュアルを与え、その知識の定着度合いを計測しました。知識の定着率は、この業務マニュアルに記載範囲内の問題を作り、LLM に解かせて、その正答率を計測しました。LLM には、tsuzumi 2 と Llama 3（3B）を用いて、それぞれ学習前後の正答率を比較しました。

tsuzumi 2 において、学習前（業務マニュアルを与える前）の正答率は 65.3%でした。先行研究では最大 73.3%まで向上しましたが、本技術を用いることで 83.1%と、さらに 10%近く正答率を向上させることができました。

特化モデルは、基本的な性能が低くなりがちな軽量モデルほど実施の必要性が高いことを鑑みて、tsuzumi 2 よりも軽量なモデルである Llama 3（3B）でも評価を行いました。同様に先行研究よりも優れていることを確認しています。

[比較手法]

・先行研究※2：本技術とは異なる観点で学習データを生成する手法

・従来手法：オリジナルデータ（業務マニュアル）を繰り返し利用する手法

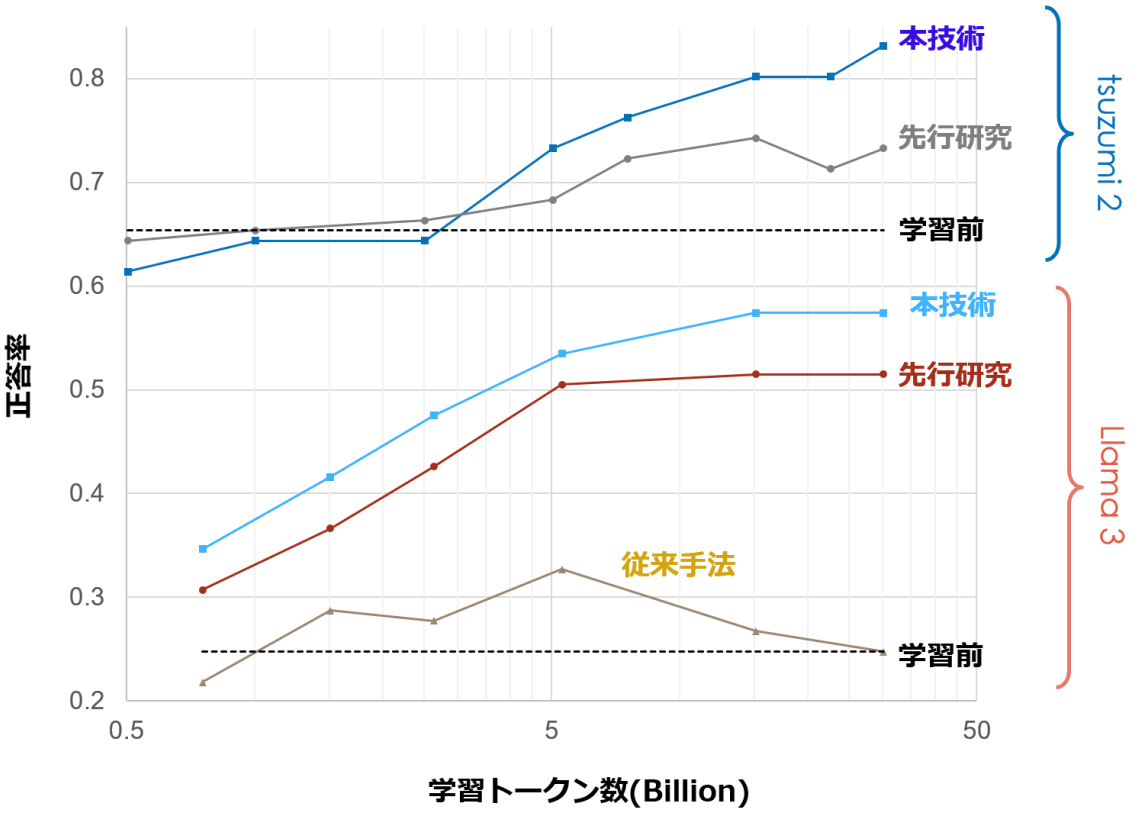



図 2. 業務マニュアルの知識を問う問題の正答率。横軸はチューニング時に与えたデータ量。

## 4. 今後の展開

本技術は NTT グループ各社に導入される予定です。チューニング観点は多様ですので、様々な手法を組み合わせることになりますが、その1つ手法となります。なお、NTT R&D フォーラム 2026（11 月開催）にて本技術の紹介を行います。展示では、本技術の説明に加え、tsuzumi の用途拡大の軌跡や、AI エージェントが能動的に業務を覚えるための技術開発についても紹介する予定です。

## 5. 関連する過去の報道発表

・2026 年 5 月 19 日「NTT 版 LLM tsuzumi 2 アップデート」

・2025 年 10 月 20 日「更なる進化を遂げた NTT 版 LLM tsuzumi 2 の提供開始」

【用語解説】

※1 NTT R&D FORUM 2026 https://www.rd.ntt/forum/2026/

Mehr Licht!
More Photonics for Intelligence

※2 数学問題・FAQ（よくある質問）・表問題・手順書（チュートリアル）形式に言い換える手法 FinePhrase [Niklaus+, 2026]および、Nemotron-Personas-Japan 等の外部データセットからランダムに選んだペルソナ（読者像）に応じて言い換える手法。

■本件に関する報道機関からのお問い合わせ先
NTT 株式会社
サービスイノベーション総合研究所
広報担当
問い合わせフォームへ