先端技術キーワード解説

知っておきたい最新の動き

[大規模言語モデル tsuzumi]

NTT が、独自開発した大規模言語モデル（LLM：Large Language Models）「tsuzumi」を 2024 年 3 月から提供開始すると発表しました。tsuzumi は、軽量でありながら、世界トップクラスの日本語処理性能を持つと言われています。

１．大規模言語モデル（LLM）とは

「言語モデル」とは、文章の並び方に確率を割り当てる確率モデルです。

「大規模言語モデル（LLM）とは、巨大なデータセットとディープラーニング技術を用いて構築された言語モデルです。「大規模」とは、3 つの要素「計算量」「データ量」「モデルパラメータ数」が巨大化されていることを示します。

２．tsuzumi の概要

tsuzumi は、モデルパラメーター数が 6 億（超軽量版）、70 億（軽量版）と軽量でありながら、「世界トップクラス」の日本語処理性能を持つ LLM です。軽量なため、1 つの GPU や CPU で推論動作が可能で、学習やチューニングに必要な時間やコストを軽減できるそうです。

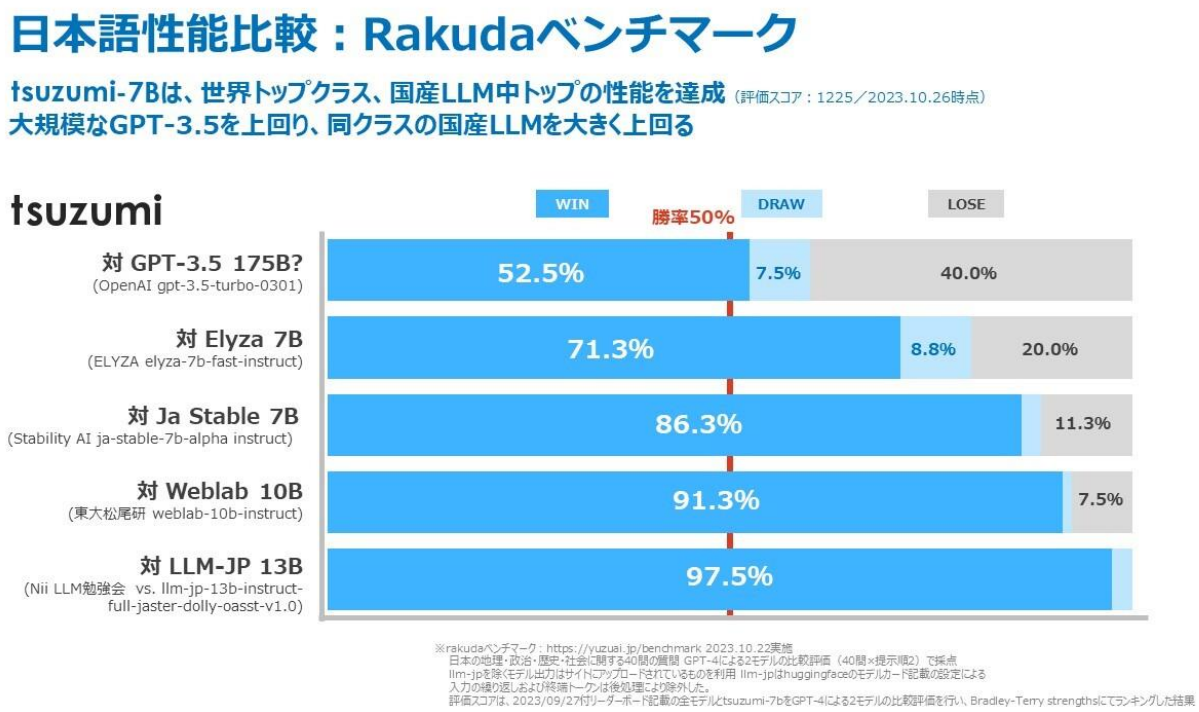
日本語／英語に対応する他、表が含まれる誓約書や契約書といった図表文書の視覚読解など、さまざまな形式にも対応できます。

３．特徴

主な特徴は以下の 4 つです。

（1）軽量な LLM

（2）日本語と英語に対応　～特に日本語が得意な LLM～



（3）柔軟なチューニング　～基盤モデル＋アダプタ～

チューニングは、モデルの振る舞いを特定のタスクや目的に合わせて調整するプロセスです。「tsuzumi」は、効率的に知識を学習させることのできるアダプタにより、チューニングを少ない追加学習量で実現します。

（4）マルチモーダル　～言語＋視覚・聴覚・ユーザー状況理解～

2024 年 3 月以降、言語化されていないグラフィカルな表示や音声のニュアンス、顔の表情、ユーザーのおかれている状況や、さらにはロボットが自分の身体感覚やヒトの身体的特徴を理解し、現実世界での人との協調作業も可能なモーダル拡張に対応予定とのことです。

４．現在の活用事例

（1）東京海上日動火災保険：事故対応部門において、専門的な通話を要約する仕組みをトライアル中です。オペレーターが通話後に行う事務作業の時間を 50%削減できる見込みとのことです。

（2）京都大学医学部附属病院：tsuzumi に電子カルテのデータを学習させることで、システムが電子カルテに記載された医療データを読解した上で、共通フォーマットに適切な形で整理します。
これにより、分析できる状態にすることを目指しているとのことです。

［参考文献］

1）NTT R&D Website：NTT 版大規模言語モデル「tsuzumi」
https://www.rd.ntt/research/LLM_tsuzumi.html

2）EE Times：1 つの GPU／CPU で推論可能な超軽量 LLM「tsuzumi」を 24 年 3 月から提供へ
https://eetimes.itmedia.co.jp/ee/articles/2401/16/news083.html

（注）

本解説は、執筆当時の状況に基づいて解説をしております。ご覧になる時には、状況が変わっている可能性がありますので、ご注意をお願いします。

技術・経営の戦略研究・トータルサポータ
ティー・エム研究所
工学博士
中小企業診断士
社会保険労務士（登録予定）
代表　芳賀　知
E-Mail：info_tm-lab@mbn.nifty.com　URL：http://tm-lab@a.la9.jp/