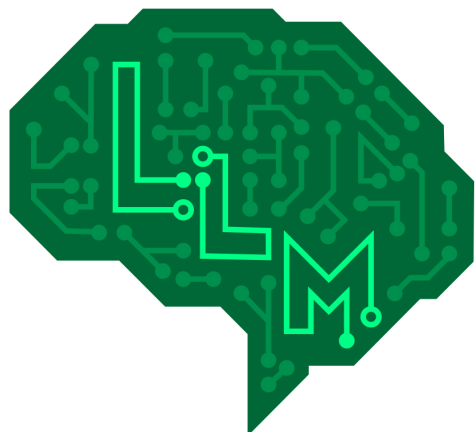


大規模言語モデル「tsuzumi」

IOWN Pick Up NTT版大規模言語モデル



背景

大規模言語モデル（LLM）は、その高度かつ柔軟な言語処理能力が注目される一方、1000億以上のパラメータのモデルを動作させる消費電力など運用コストに課題があります。また、モデルの追加学習や変更にも膨大な計算資源が必要であり、細かなカスタマイズが困難です。

成果の概要

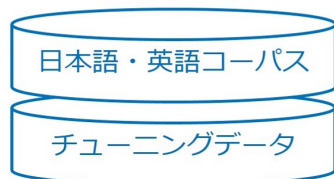
軽量で低消費電力のLLMを開発し、ローカル環境でGPU1枚でも動作が可能です。日本語の大量・高品質のデータの活用により日本語の理解・処理を得意としています。特定の業界や組織、個人特化のデータをアダプタへ追加学習することで、柔軟なカスタマイズが可能です。

特徴1:
軽量で日本語が得意

学習データに工夫を施し、パラメータ数70億のモデルにて、日本語でトップレベルの精度(*)を達成しました。

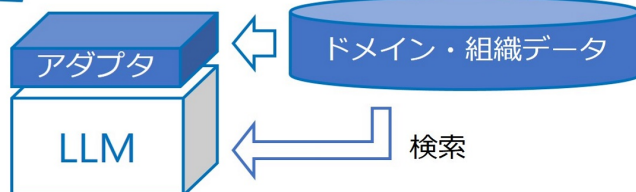
1兆を超える件数のテキスト

長年の研究開発の蓄積と
新規に作成したデータ

特徴2:
柔軟なカスタマイズ

汎用なLLMをお客様向けに「特化」して、独自のLLMとしてご利用いただくことをめざします。

チューニング



(*) Rakuda, Japanese Vicuna QA, JGLUEにて評価

技術のポイント

- 軽量なモデル（パラメータ数70億）にて、日本語のタスク評価でトップレベルの精度を達成
- NTTにおける40年以上の自然言語処理研究で蓄積したデータやノウハウを活用し独自のモデルを学習
- 追加で知識を学習できるアダプタにより、特定業界や組織に特化するなど、さまざまな利用用途によって柔軟に言語モデルをカスタマイズ

この研究がもたらす未来

LLMをビジネス領域やユーザに合わせてスペシャリスト化することで、人の能力を最大限に活かす存在とし、AIとヒトが協働し相互成長によって「持続可能な世界」の実現をめざします。

出展企業

日本電信電話株式会社

問い合わせ先

rdforum-exhibition@ml.ntt.com