

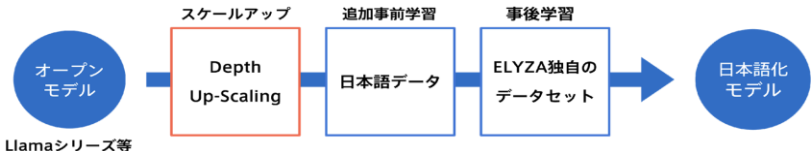
モデル組み合わせによる日本語大規模基盤モデル 開発と半自動データ作成フレームの構築

実施者 株式会社ELYZA

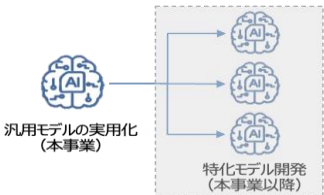
事業概要

日本語処理能力の高い汎用基盤モデルの開発

グローバルモデルの日本語継続学習を行う際に、LLMを深さ方向に拡張するDepth Up-Scaling (DUS) を実施。パラメータ数の増加による性能向上の影響について検証を行った。



社会実装イメージ



開発した日本語モデルを自社プロダクト・APIで展開しつつ、領域特化モデルの開発・展開に繋げていく。まず、日本の法令や行政手続といった特定の業界やタスクに特化したモデルの開発を行い、企業利用を促進する。

※本資料に掲載する製品名等は、各社の商標または登録商標です。

事業成果

Meta社のLlama-3 70BをベースにDUSを行い、1220億パラメータのモデルを作成。追加事前学習 (150Bトークン)、事後学習を行なったLlama-3-ELYZA-JP-120Bは、3つのベンチマークでGPT-4を超える性能を達成。

モデル名	Japanese MT-Bench	ELYZA-Tasks-100	Nejumi-v3
Llama-3-ELYZA-JP-70B	9.08	4.07	0.6693
Llama-3-ELYZA-JP-120B	9.24	4.09	0.6898
Llama-3-Swallow-70B-Instruct-v0.1	8.27	3.65	0.5802
GPT-4 (0613)	9.01	4.03	0.6883
GPT-4o (2024-08-06)	9.40	4.49	0.7433

日本の法令や行政手続きに関するデータを独自収集・加工し、特化学習を実施。知識の正確性、理解度を測るテストにおいて、GPT-4を上回るスコアを達成。

モデル名	スコア	スコア (contextあり)
Llama-3-ELYZA-JP-120B	2.90	4.08
Llama-3-ELYZA-JP-120B + 特化学習	3.65	3.88
GPT-4 (0613)	2.92	4.27
GPT-3.5 Turbo (0125)	2.20	3.17

※ 評価は、「GPT-4」による自動評価を実施
※ 「contextあり」は、必要な情報が含まれるテキストを、回答時のpromptに挿入した場合