

2025年4月3日

リコー、モデルマージによって GPT-4o と同等の 高性能な日本語 LLM(700 億パラメータ)を開発

～省コスト・省リソースを実現し、お客様のオンプレミス環境でのプライベート LLM 導入を加速～

株式会社リコー(社長執行役員:大山 晃)は、米 Meta Platforms 社が提供する「Llama-3.3-70B-Instruct」の日本語性能を向上させた「Llama-3.3-Swallow-70B-v0.4^{*1}」をベースモデルに、同社の Instruct モデルから抽出した Chat Vector^{*2}と、リコー製の Chat Vector^{*3}を独自のノウハウでマージすることで、リコーが開発・提供する 700 億パラメータの日本語大規模言語モデル(LLM^{*4})の性能を向上させました。ベンチマークツールによる検証の結果、米 OpenAI が開発した GPT-4o と同等レベルの優れた性能が確認されました(2025 年 3 月 25 日時点)。高性能でありながら、省コスト・省リソースを実現し、オンプレミス環境でお客様情報の追加学習が可能なプライベート LLM として、企業内での導入を支援します。

生成 AI の普及に伴い、企業が業務で活用できる高性能な LLM へのニーズが高まっています。しかし、LLM の追加学習には、高コスト、かつ長時間を要するという課題があります。こうした課題に対し、複数のモデルを組み合わせて、高性能なモデルを構築する「モデルマージ^{*5}」は効率的な開発手法として注目されています。

リコーは、モデルマージのノウハウと LLM 開発の知見を活かし、今回、新たな LLM を開発しました。本技術は、企業独自のプライベート LLM や特定業務向けの高性能 LLM を効率的に開発するための一助になります。リコーは、自社製 LLM の開発にとどまらず、お客様の用途や環境に最適な LLM を低コスト・短納期で提供するために、多様で効率的な手法・技術の研究開発を推進してまいります。

【評価結果】

複雑な指示・タスクを含む代表的な日本語ベンチマーク「ELYZA-tasks-100」と日本語のマルチターンの対話能力を評価する「Japanese MT-Bench」の 2 つのベンチマークツールを組み合わせて、平均スコアを比較しました。今回リコーがモデルマージの手法で開発した LLM は GPT-4o (gpt-4o-2024-08-06)と同等レベルの高いスコアを示しました(評価には「gpt-4o-2024-08-06」を使用)。

株式会社リコー <https://jp.ricoh.com/>

報道関係のお問い合わせ先 広報室 TEL : 050-3814-2806 (直通) E-mail : koho@ricoh.co.jp

お客様の問い合わせ先 仕事のAI お問合せフォーム

https://www.secure.rc-club.ricoh.co.jp/shigoto-no-ai_inq?

No.	企業/組織	モデル	Japanese MT-Bench								平均	ELYZA-tasks-100 ^{*6}	平均スコア
			コーディング	情報抽出	人文科学	数学	推論	ロールプレイ	科学・技術・工学・数学	ライティング			
1	OpenAI	gpt-4o-2024-08-06	9.35	8.55	9.00	8.40	7.60	8.70	8.45	8.80	8.61	9.12	8.86
2	OpenAI	gpt-4o-mini-2024-07-18	8.35	8.55	8.70	7.90	6.45	8.80	8.30	8.90	8.24	8.54	8.39
3	Meta	Llama-3.3-70B-Instruct	7.40	8.75	7.30	6.75	8.05	8.40	7.45	8.25	7.79	8.12	7.96
4	Tokyotech	Llama-3.3-Swallow-70B-Instruct-v0.4	7.00	7.90	8.60	7.35	6.75	8.65	8.45	8.80	7.94	8.38	8.16
5	Ricoh	Llama-3.3-Ricoh-70B-20250325	8.15	9.00	8.93	9.23	7.93	8.85	7.95	8.90	8.62	9.06	8.84

ベンチマークツールにおける他モデルとの比較結果(リコーは最下段)

■リコーの AI 開発について

リコーは、1990 年代に AI 開発を開始し、2015 年からは画像認識技術を活かした深層学習 AI の開発を進め、外観検査や振動モニタリングなどへの適用を行ってきました。2020 年からは自然言語処理技術を活用し、オフィス内の文書やコールセンターに寄せられた顧客の声 (VOC) などを分析して業務効率化や顧客対応に活かす「仕事の AI」の提供を開始しました。

さらに、2022 年からはいち早く、大規模言語モデル (LLM) の研究・開発に着目し、2023 年 3 月にはリコー独自の LLM を発表。その後も、700 億パラメータの大規模ながらオンプレミスでも導入可能な LLM (日英中 3 言語に対応) を開発するなど、お客様のニーズに応じて提供可能なさまざまな AI の基盤開発を行っています。また、画像認識や、自然言語処理に加え、音声認識 AI の研究開発も推進し、音声対話機能を備えた AI エージェントの提供も開始しています。

*1 Llama-3.3-Swallow-70B: 東京科学大学情報理工学院の岡崎研究室と横田研究室、国立研究開発法人産業技術総合研究所の研究チームで開発された日本語 LLM モデル。2025 年 3 月 10 日に公開。

*2 Chat Vector: 指示追従能力を持つモデルからベースモデルのウェイトを差し引き、指示追従能力のみを抽出したベクトル。

*3 リコー製の Chat Vector: Meta 社のベースモデル「Meta-Llama-3-70B」に対し、リコー独自開発を含む約 1 万 6 千件のインストラクションチューニングデータで追加学習した Instruct モデルから抽出した Chat Vector。

*4 Large Language Model (大規模言語モデル): 人間が話したり書いたりする言葉 (自然言語) に存在する曖昧性やゆらぎを、文章の中で離れた単語間の関係までを把握し「文脈」を考慮した処理を可能にしているのが特徴。「自然文の質問への回答」や「文書の要約」といった処理を人間並みの精度で実行でき、学習も容易にできる技術。

*5 モデルマージ: 複数の学習済みの LLM モデルを組み合わせて、より性能の高いモデルを作る新たな方法のこと。GPU のような大規模な計算リソースが不要で、より手軽にモデル開発ができるとして、近年注目されています。

*6 ELYZA-tasks-100 のスコア: Japanese MT-Bench との平均スコアを算出するため、ELYZA-tasks-100 のスコア (5 点満点) を 2 倍に算出し、比較しています。

■関連ニュース

リコー、モデルマージによって GPT-4 と同等の高性能な日本語 LLM(700 億パラメータ)を開発

https://jp.ricoh.com/release/2024/0930_1

リコー、日英中 3 言語に対応した 700 億パラメータの大規模言語モデル(LLM)を開発、お客様のプライベート LLM 構築支援を強化

https://jp.ricoh.com/release/2024/0821_1

インストラクションチューニング済みの 130 億パラメータの日本語 LLM を開発

https://jp.ricoh.com/release/2024/0603_1

日本語精度が高い 130 億パラメータの大規模言語モデル(LLM)を開発

https://jp.ricoh.com/release/2024/0131_1

※社名、製品名は、各社の商標または登録商標です。

| リコーグループについて |

リコーグループは、お客様のDXを支援し、そのビジネスを成功に導くデジタルサービス、印刷および画像ソリューションなどを世界約200の国と地域で提供しています(2024年3月期グループ連結売上高2兆3,489億円)。

“はたらく”に歓びを 創業以来85年以上にわたり、お客様の“はたらく”に寄り添ってきた私たちは、これからもリーディングカンパニーとして、“はたらく”の未来を想像し、ワークプレイスの変革を通じて、人ならではの創造力の発揮を支え、さらには持続可能な社会の実現に貢献してまいります。

詳しい情報は、こちらをご覧ください。<https://jp.ricoh.com/>