



2024 年（令和 6 年）12 月 24 日

完全にオープンな約 1,720 億パラメータ（GPT-3 級）の大規模言語モデル 「llm-jp-3-172b-instruct3」を一般公開

～GPT-3.5 を超える性能を達成～

大学共同利用機関法人 情報・システム研究機構 国立情報学研究所（エヌアイアイ N I I、くろはし 所長：黒橋 さだお 禎夫、東京都千代田区）の大規模言語モデル研究開発センター（LLMC）は、主宰する LLM 勉強会（LLM-jp）の活動の中で、約 1,720 億パラメータ^{(*)1}（GPT-3 のパラメータ数と同規模）の大規模言語モデル（LLM）を 2.1 兆トークンの学習データを用いてフルスクラッチで学習し、同モデルを「llm-jp-3-172b-instruct3」として一般公開しました。学習データまで含めすべてオープンなモデルとしては世界最大規模です。本モデルは言語モデルの日本語理解能力を測るベンチマーク「llm-jp-eval」および経済産業省・国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）による生成 AI 開発を支援するプロジェクトである GENIAC^{(*)2}で用いられた「llm-leaderboard」において GPT-3.5 を超える性能を達成しています。

本モデルは、データ活用社会創成プラットフォーム mdx^{(*)3} での 130 億パラメータの LLM の学習、国立研究開発法人産業技術総合研究所の第 2 回大規模言語モデル構築支援プログラムによる AI 橋渡しクラウド（ABCI）での 1,750 億パラメータ・モデルの学習トライアルの成果を踏まえ開発したものです。

LLMC では「llm-jp-3-172b-instruct3」を活用して LLM の透明性・信頼性の確保に向けた研究開発を進めていきます。

1. 今回公開した LLM の概要

(1) 利用計算資源

- 経済産業省・NEDO の GENIAC プロジェクトの支援によるクラウド計算資源（グーグル・クラウド・ジャパン）を利用して、約 0.4 兆トークンまでの事前学習を実施
- その後、文部科学省の補助金により調達したクラウド計算資源（さくらインターネット）を利用して、約 2.1 兆トークンまでの事前学習及びチューニングを実施

(2) モデル学習用コーパス^{(*)4}

- 以下に示すコーパス（約 2.1 兆トークン）を用いて事前学習を実施
 - 日本語：約 5,920 億トークン

- Web アーカイブ Common Crawl (CC) 全量から抽出・フィルタリングした日本語テキスト
 - 国立国会図書館インターネット資料収集保存事業 (WARP) で収集された Web サイトの URL (当該 URL リストは同館から提供) を基にクロールしたデータ
 - 日本語 Wikipedia
 - KAKEN (科学研究費助成事業データベース) における各研究課題の概要テキスト
- 英語 : 約 9,500 億トークン (Dolma 等)
 - 他言語 : 約 10 億トークン (中国語・韓国語)
 - プログラムコード : 約 1,140 億トークン
 - 以上の約 1.7 兆トークンに加え、日本語コーパスのうち約 0.4 兆トークンを 2 回学習することとし、合計約 2.1 兆トークン

(3) モデル

- パラメータ数 : 約 1,720 億個 (172B)
- モデルアーキテクチャ : Llama 2 ベース

(4) チューニング

- 日本語インストラクションデータおよび英語インストラクションデータの和訳データ 13 種類を用いてチューニングを実施

(5) 評価

- LLM-jp が開発している、既存の日本語言語資源に基づく 26 種類の評価データを用いて横断的な評価を行うフレームワーク「llm-jp-eval v1.4.1」を用いた評価で 0.613 を達成。GPT-3.5 の性能 0.590 を 0.023 ポイント上回った。
- GENIAC 事業にて性能評価に用いられるフレームワーク「llm-leaderboard (g-leaderboard ブランチ)」による評価で 0.669 を達成。GPT-3.5 の性能 0.653 を 0.016 ポイント上回った。

(6) 開発モデル・ツール・コーパスの公開 URL

<https://llm-jp.nii.ac.jp/release>

注：

- 本モデルは、現在の技術で可能な安全性のチューニングを行っていますが、完全な安全性を保証することは技術的に難しく、入力内容によっては必ずしも適切でない内容が出力される可能性があります。
- 安全性データ(AnswerCarefully v1 の評価データ)の内の 181 件で安全性評価を行った結果、7 件が安全性の基準に満たない回答でした。この評価結果は OpenAI の gpt-4-0613 など最先端システムを超える良い結果です。

2. 今後の予定

- LLM を社会で利活用していく上では LLM の透明性・信頼性の確保が必要であり、モデルの高度化に伴い、安全性の配慮もより重要となります。そのため、NII は、文部科学省の補助事業「生成 AI モデルの透明性・信頼性の確保に向けた研究開発拠点形成」（https://www.mext.go.jp/content/20240118-ope_dev03-000033586-11.pdf の p.7）の支援を受け 2024 年 4 月に大規模言語モデル研究開発センターを設置しました。
今回公開したモデルや、今後構築するモデルを活用してそれらの研究を進め、LLM 研究開発の促進に貢献します。
- また、今回のモデルは最終チェックポイント（292,812 訓練ステップ時点）以外に、そこに至るまでの 1,000 訓練ステップごとのチェックポイントのデータも保存しています。今後、それらのデータも提供予定です。

(参考 1) LLM 勉強会 (LLM-jp) の概要

1. NII が主宰する LLM-jp では、自然言語処理及び計算機システムの研究者を中心として、大学・企業等から 1,900 名以上（2024 年 12 月 24 日現在）が集まり、ハイブリッド会議、オンライン会議、Slack 等を活用して LLM の研究開発について情報共有を行うとともに、共同で LLM 構築等の研究開発を行っています。具体的には、以下の目的で活動しています。
 - オープンかつ日本語に強い LLM の構築とそれに関連する研究開発の推進
 - 上記に関心のある自然言語処理および関連分野の研究者によるモデル構築の知見や最近の研究の発展についての定期的な情報交換
 - データ・計算資源等の共有を前提とした組織横断的な研究者間の連携の促進
 - モデル・ツール・技術資料等の成果物の公開
2. 「コーパス構築 WG」「モデル構築 WG」「チューニング・評価 WG」「安全性 WG」「マルチモーダル WG」「実環境インタラクション WG」等を設置し、それぞれ、早稲田大学 河原大輔教授、東北大学 鈴木潤教授、東京大学 宮尾祐介教授、国立情報学研究所 関根聡特任教

授、東京科学大学 岡崎直観教授、早稲田大学 尾形哲也教授を中心に研究開発活動に取り組んでいます。このほか、東京大学 田浦健次朗教授、空閑洋平准教授（計算資源の利用技術）、東京科学大学 横田理央教授（並列計算手法等）を筆頭に、多数の方々の貢献により活動を進めています。

3. 詳細については、ホームページ <https://llm-jp.nii.ac.jp/> をご参照ください。

（参考 2）

この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構（N E D O）の助成事業及び文部科学省の補助事業の結果得られたものです。

〈メディアの皆様からのお問い合わせ先〉

大学共同利用機関法人 情報・システム研究機構 国立情報学研究所
総務部企画課 広報チーム

TEL : 03-4212-2164 E-mail : media@nii.ac.jp

(*1) パラメータ数：大規模言語モデルは言語を学習した大規模なニューラルネットワークで、パラメータはニューラルネットワークの規模を示す指標のひとつ。パラメータ数が多いほど高い性能であるといわれている。

(*2) GENIAC (Generative AI Accelerator Challenge)：国内の生成 AI の開発力強化を目的とし、経済産業省及び NEDO が協力して実施する事業。主に生成 AI のコア技術である基盤モデルの開発に対する計算資源の提供や、データや AI の利活用に向けた実証調査の支援等を行っている。

(*3) データ活用社会創成プラットフォーム mdx：9 大学 2 研究所が連合して共同運営する、データ活用に関心した高性能仮想化環境。研究環境を用途に合わせてオンデマンドで短時間に構築・拡張・融合できる、データ収集・集積・解析のためのプラットフォーム。

(*4) コーパス：自然言語の文章を構造化し大規模に集積したデータベース。