生成AIモデルの透明性・信頼性の確保に向けた研究開発

大規模言語モデル研究開発センター (LLM研究開発センター: LLMC)

どんな研究？

ChatGPTなどの大規模言語モデル(LLM)が社会基盤として浸透するうえで多くの課題があります。オープンかつ日本語に強いLLMを開発し、その原理解明やモデルの公開を通して、LLMの透明性や安全性の確保などの最重要課題に取り組みます。

何がわかる？

学習用データセットの拡大や、学習手法の改善によりLLMの性能は着実に向上しています。さらにマルチモーダル化や対話、学術分野への適合など、高度化・多様化するモデルとその知見を広く社会と共有して、生成AIの受容と利用に貢献します。

研究活動

研究開発用LLM構築
コーパス整備、計算環境整備、評価用ベンチマーク作成などを行うとともに研究開発用のLLMを構築します

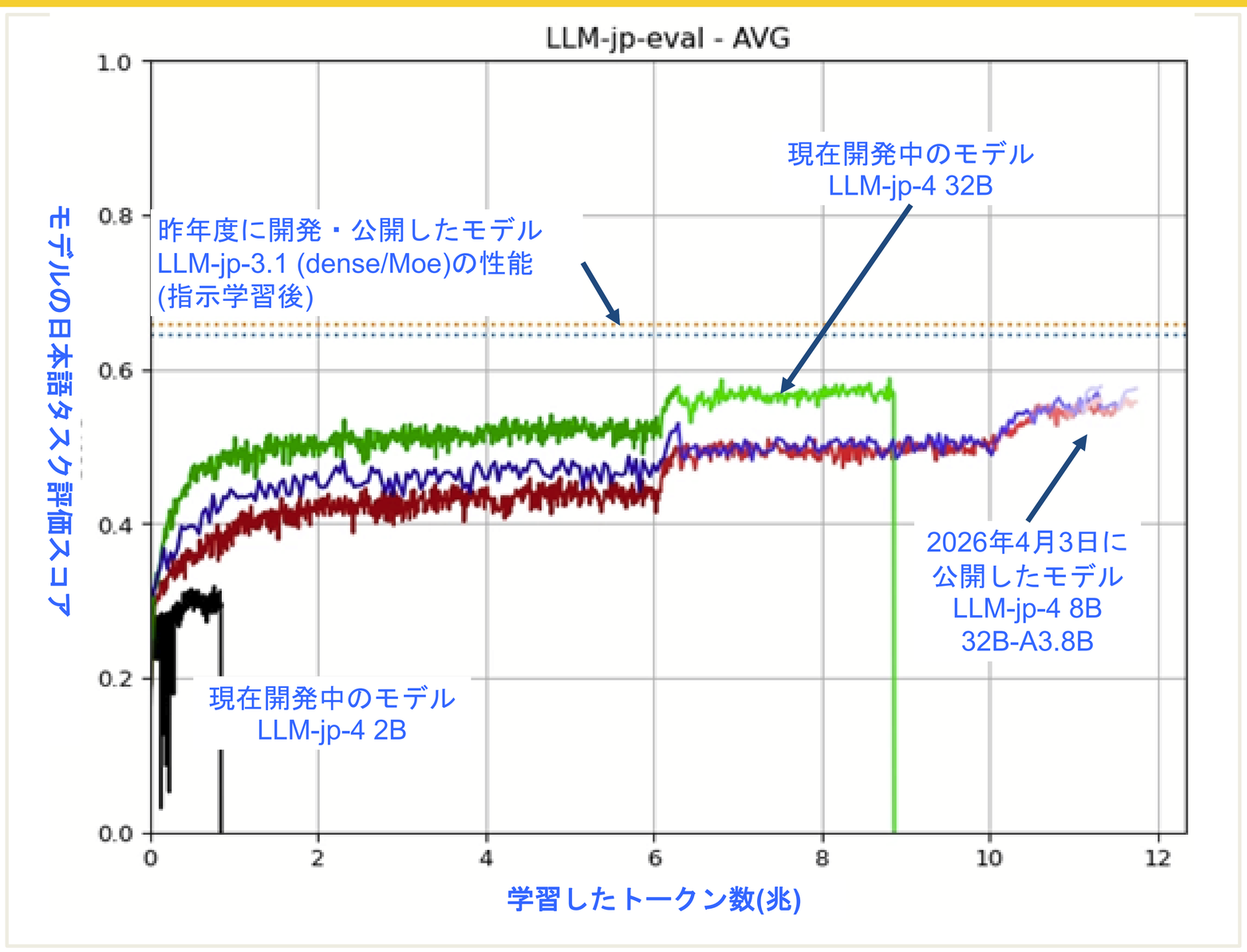
透明性・信頼性確保に向けた研究開発
生成AIの挙動原理を解明すること、またデータ改変やデータバイアス等の影響を抑制する技術を開発することなどにより、生成AIの透明性・信頼性を確保します

LLMの高度化に向けた研究開発
ドメイン適応、モデル自体の軽量化など、生成AIモデルの高度化に資する研究開発を行います

LLM-jp (LLM研究開発コミュニティ)
生成AIに関連する分野の研究者・開発者・実務家が、WGを通した研究開発に従事したり、定期的に情報共有を行う(2,700名以上が参画)

- オープンかつ日本語に強い大規模モデル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

これまでに公開したモデルや開発中のモデル



- 8B denseおよび32B-A3B MoEモデルのbase/thinkingの計4モデルを公開
- 日本語 MT-Bench、MT-Benchにおいて、GPT-4oやQwen3-8Bを上回る性能を達成
- 商用利用も可能なオープンソースライセンス(Apache 2.0)で提供
- 「LLM-jp-3.1」シリーズと比較して約6倍の規模となる学習コーパス(約19.5兆トークン)を構築し、うち合計約10.5兆トークンを事前学習に使用してフルスクラッチで学習
- 中間学習を実施し、事前学習コーパスに指示事前学習データを含むLLM による合成データを加えた、合計1.3兆トークンの学習コーパスを使用。文脈長は約6万5千トークン

モデル、データ、ツールを**LLM-jp**のサイトで公開中

LLM-jp: 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2,700名以上(*)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

(*)2026年4月1日時点

連絡先：国立情報学研究所 大規模言語モデル研究開発センター
URL：https://www.nii.ac.jp/research/centers/llmc/ Email：llm-admin@nii.ac.jp
国立情報学研究所
大規模言語モデル研究開発センター
Research and Development Center for Large Language Models

LLM研究開発センター(LLMC)の研究紹介

小田 悠介、劉 超然、小林 和馬、清丸 寛一、磯沼 大、児玉 貴志、中山 功太、劉 倩瑩、Yang Zhishen、Su Myat Noe、橘 秀幸、江 俊鋒、Ho Thi Xanh、Iffat Maab、Michal Štefánik、高橋 諒、Wan Zhen、Wu Yun-Ang、Yang Zhengdong (LLMC 特任研究員)

磯沼 大

劉 倩瑩

清丸 寛一

Scaling Data-Constrained Language Models with Synthetic Data (EACL'26)

小田 悠介

児玉 貴志

劉 超然

Language Models Learn Universal Representations of Numbers and Here's Why You Should Care (ACL '26)

Michal Štefánik

中山 功太

Multimodal Safety Evaluation of Vision-Language Models in Japanese (NLP'26)

Su Myat Noe

Iffat Maab

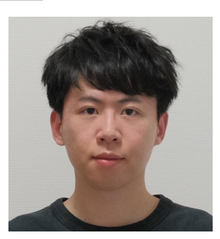
Yang Zhishen

Bis reasoning 1.0: The first large-scale Japanese benchmark for belief-inconsistent syllogistic reasoning (LREC'26)

橘 秀幸

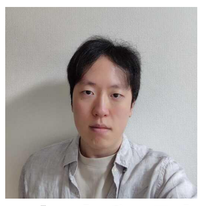
Zhengdong Yang

小林 和馬



江 俊鋒

Wu Yun-Ang



高橋 諒

SciClaimEval: Cross-modal Claim Verification in Scientific Papers

Ho Thi Xanh

GPT-RE: In-context learning for relation extraction using large language models

Wan Zhen

連絡先：国立情報学研究所 大規模言語モデル研究開発センター
URL：https://www.nii.ac.jp/research/centers/llmc/ Email：llm-admin@nii.ac.jp