

2024 年（令和 6 年）4 月 1 日

国立情報学研究所に「大規模言語モデル研究開発センター」新設

～国産 LLM を構築し、生成 AI モデルの透明性・信頼性を確保する研究開発を加速～

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、所長：くろはし さだお 黒橋 禎夫、東京都千代田区）は、2024 年（令和 6 年）4 月 1 日、新たに大規模言語モデル（Large Language Models、以下「LLM」）の研究開発を行う「大規模言語モデル研究開発センター（センター長：黒橋 禎夫 NII 所長、以下「LLM 研究開発センター」）」（英語名：Research and Development Center for Large Language Models、略称：LLMC）を開設しました。

NII では 2023 年 5 月、国内の研究機関のみならず民間企業などの幅広い人材が参加する LLM 勉強会（LLM-jp）を立ち上げ、オープンな生成 AI の研究開発を推進してきましたが、今般、文部科学省の「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」事業を実施する拠点として、研究所内に新たに専門のセンターを設立し、新進気鋭の AI 研究者が LLM の研究開発に邁進できる体制を構築しました。

LLM 研究開発センターでは、まずは今年夏ごろの完成を目指して 1750 億パラメータ（GPT-3 級）の LLM を構築するとともに、LLM の透明性・信頼性の確保に向けた先端的な研究開発など、オープンかつ日本語に強い大規模モデルの構築に関連する研究開発を推進します。そして、このような活動を通じ、一連の知識と経験を蓄積し、AI の進化、ひいては将来に渡る革新的なイノベーションの創出に貢献します。

大規模言語モデル（LLM）は全産業に波及しつつあり、基盤モデルとしてこれまでの産業基盤を抜本的に変革する可能性を秘めているとともに、幅広い科学技術研究に必須の知識基盤になると期待されています。しかし、現状では強いモデルの学習コーパス・データは非公開で、モデルの振舞いを含めて全体がブラックボックスとなっており、ハルシネーションやバイアス等の課題が山積しています。また、有力な基盤モデルにおいて学習に用いられるデータの言語は英語が中心であり、日本語の理解・生成能力は相対的に劣っています。現状、国内では、1000 億パラメータ規模の学習が長期間行われた例が限られており、LLM 開発における知見の獲得が遅れています。

国立情報学研究所（NII）が主宰する LLM 勉強会（LLM-jp）では、2023 年 10 月に 130 億パラメータの LLM を構築、公開し、国内の LLM 開発に貢献してきました。今般、文部科学省の「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」事業を実施する拠点として、NII 内

National Institute of InformaticsWeb: <https://www.nii.ac.jp>

X(旧 Twitter): @jouhouken

facebook: <https://www.facebook.com/jouhouken>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

〒101-8430 東京都千代田区一ツ橋 2-1-2

直通：03-4212-2164 E-Mail : media@nii.ac.jp

に「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)を設置し、生成 AI モデルの透明性・信頼性の確保に向けた研究開発を推進する体制を整えました(図 1)。LLM-jp における LLM 開発の知見を踏まえ、研究者と技術者が一体的に研究開発を行う知の拠点を構築し、生成 AI モデルに関する研究力・開発力醸成のための環境を整備します。

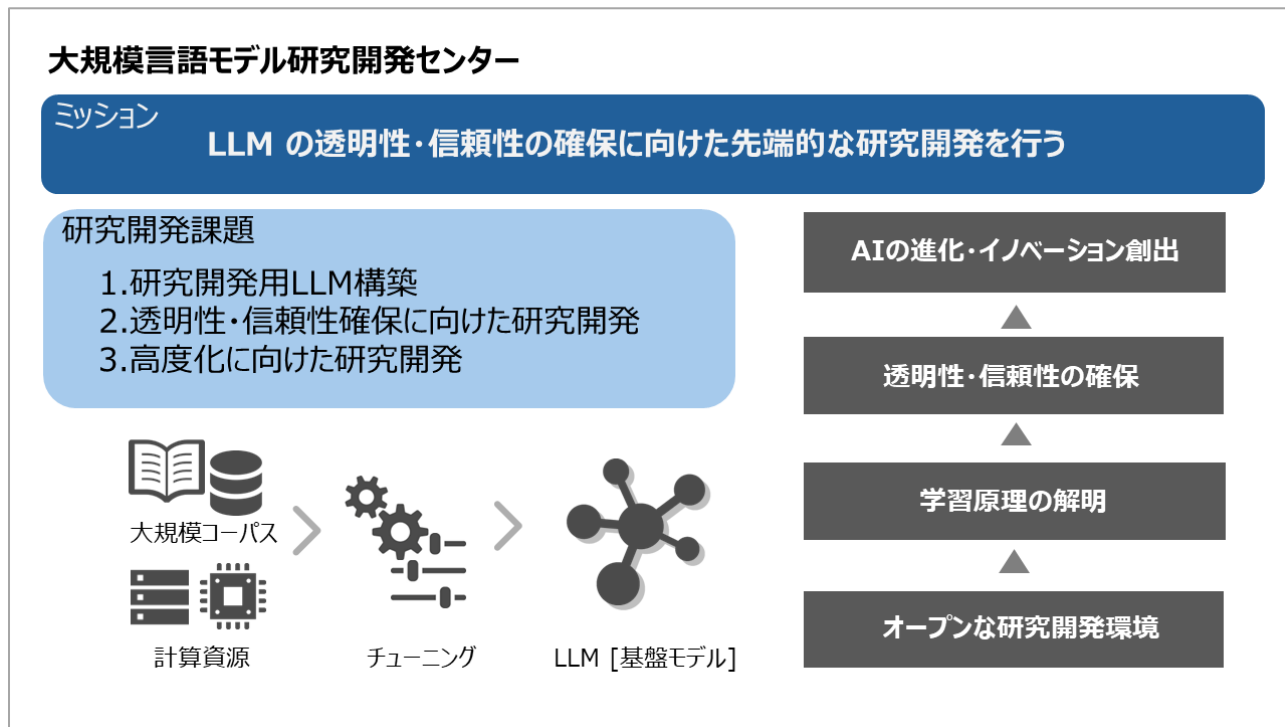


図 1 大規模言語モデル研究開発センター 概要

LLM 研究開発センターでは以下の研究開発を実施します。

1. 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 を構築する

2. 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・信頼性を確保する

3. 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う

LLM-jp における研究開発をさらに発展させ、まずは今年夏ごろの完成を目指して 1750 億パラメータ規模 (GPT-3 級) の LLM を構築します。研究成果のモデルへの適用や試行錯誤を通じて、透明性・信頼性を確保した生成 AI モデル構築手法の確立を目指すとともに、一連の知識と経験を蓄積し、AI の進化、ひいては将来に渡る革新的なイノベーションの創出に貢献します。

【センターの概要】

センター名：大規模言語モデル研究開発センター

センター長：黒橋 禎夫（NII 所長／京都大学特定教授）

副センター長：相澤 彰子（NII 副所長／コンテンツ科学研究系 教授）

武田 浩一（NII 特任教授）

本センターの研究開発には所属研究員の他、LLM-jp との連携等を通じて、大学、企業などから研究者が多数参加します。また、今後も研究体制の増強を図る予定です。詳細はセンターweb サイトをご参照下さい。

センターweb サイト：<https://www.nii.ac.jp/research/centers/llmc/>

【黒橋 禎夫 センター長コメント】

「国立情報学研究所では、2023 年 5 月から、趣旨に賛同した人なら誰でも参加できる LLM-jp を立ち上げ、モデル・データ・ツール・技術資料等を議論の過程・失敗を含めすべて公開する取り組みを行ってきました。大学、企業から多くの参加を得て、現時点で 1000 名を超える規模まで成長しています。このような取り組みが認められ、LLM 研究開発センターを設置することになりました。必要な計算資源を整備し、組織横断で生成 AI の原理解明と構築手法の確立に挑戦したいと考えています。優秀な若い方に集まって頂き、日本の LLM 研究開発のハブとなり、また国際的な連携も進めていきます。」

〈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