

2024年4月18日

東京工業大学

国立情報学研究所

## 東工大とNIIが日本語版の大規模言語モデルの 研究開発における連携協定を締結

ー生成AIモデルの透明性・信頼性の確保と社会実装の加速化への取り組みー

東京工業大学（以下、東工大）学術国際情報センター（以下、GSIC）と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（以下、<sup>エヌアイアイ</sup>NII、<sup>くろはし さだお</sup>所長：黒橋 禎夫、東京都千代田区）は、2024年4月16日、**大規模言語モデル**（LLM、用語1）の研究開発において連携・協力することを目的とした協定を締結しました。

### 【連携の背景】

LLMは**Society 5.0**（用語2）を駆動する基盤技術として社会全体のあり方を変える革新的な技術であり、その活用は既にあらゆる産業に波及しています。一方で、LLMは、個人情報の取扱い、著作権の保護、信頼性の確保等の観点から国民生活のさまざまな場面で課題が生じており、LLMのメカニズムの解明による透明性の向上や、安全で信頼できるAIの実現が喫緊の課題となっています。

このような社会背景の中、NIIでは2024年度、新たに大規模言語モデル研究開発センターを設置し、理解・生成の両面で日本語性能の高いLLMの構築と、生成AIの透明性・信頼性確保に向けた研究開発を実施する体制を整えました。

また、GSICでは2010年に世界で初めて**GPU**（用語3）をスーパーコンピュータに導入し、これまで多くの利用者に使いやすいGPU計算環境を整備してきました。2024年4月には、NVIDIA社製H100 GPU 960台からなる国内最大級のGPUスパコン**TSUBAME**（用語4）4.0の稼働を開始しました。

LLMの開発には膨大な計算資源の確保が必要です。今回の協定締結により、NIIにおいて実施するLLM構築に必要な学習データの収集や評価、モデル学習をスーパーコンピュータ上で安定的、効率的に実施することが可能となり、国産LLMの構築と社会実装の加速が期待されます。

### 【協定の内容】

今回、GSICとNIIが締結した協定で定める連携・協力の内容は以下の通りです。

- （1）スーパーコンピュータを用いた大規模言語モデルの効率的な学習に関する研究
- （2）日本語の大規模言語モデルの学習に必要なデータの収集・評価
- （3）日本語性能の高い大規模言語モデルの学習

本協定のもと、GSIC と NII はスーパーコンピュータ TSUBAME4.0 の活用により、LLM の研究を協力して推進し、オープンかつ日本語に強い LLM の構築を通じて、生成 AI の透明性・信頼性の向上と社会実装を加速化させる取り組みを行ってまいります。

#### 【用語説明】

- (1) **大規模言語モデル (LLM)**：テキストの現れやすさをニューラルネットによりモデル化したもので、与えられた文脈（問いかけ）に対して続くテキスト（応答）を予測・生成できる。
- (2) **Society 5.0**：サイバー空間（仮想空間）とフィジカル空間（現実空間）を高度に融合させたシステムにより、経済発展と社会的課題の解決を両立する、新たな未来社会のこと。
- (3) **GPU**：Graphics Processing Unit の略で、元は描画専用の演算装置として開発されていたが、近年は深層学習の高速化に用いられる。
- (4) **TSUBAME**：東工大のスーパーコンピュータシリーズに代々つけられている名称。元は Tokyo-tech Supercomputer and UBiquitously Accessible Mass-storage Environment の略。最新版は TSUBAME 4.0 であり、2024 年 4 月から稼働を開始した。

#### 【問い合わせ先】

（本協定書による教育・研究に関すること）

東京工業大学 学術国際情報センター

TEL: 03-5734-2087 FAX: 03-5734-3198

（報道に関すること）

東京工業大学 総務部 広報課

TEL: 03-5734-2975 FAX: 03-5734-3661

E-mail: media@jim.titech.ac.jp

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