「オープンモデル」で切り拓く日本の挑戦

国立情報学研究所　所長／大規模言語モデル研究開発センター　センター長　黒橋（くろはし）　禎夫（さだお）

生成AIの進化は、OpenAI社が2022年11月に公開したChatGPTをきっかけに日本の社会や産業の前提を大きく塗り替えた。特に大規模言語モデル（LLM）は情報検索、文章作成、プログラミング支援、データ分析などの日常業務の随所に組み込まれ、短期間のうちに社会の基盤的な技術へと位置付けを変えつつある。一方で、海外の巨大企業が主導するモデルに全面的に依存する状況が続けば、回答の出力のための学習に用いられるデータの英語への偏りや透明性の欠如、説明可能性の不足、さらには安全性・ガバナンスの面で課題が生じる可能性が指摘される。国内で安心してAIを利活用するためには、日本語や日本文化に根ざした基盤技術をどのように確保し社会に実装するかという、いわゆるソブリンAI（主権AI）の実現がこれまでになく重要性を増してきた。

生成AIの利用が急速に広がる中で、政策面では安全性や透明性の確保、人材育成、データの利活用ルール整備など、対応すべき領域が多岐にわたることも明らかになっている。国際的には「G7広島AIプロセス」をはじめ、安全性評価やガバナンスの枠組みづくりが進む一方、日本国内には計算資源の制約や日本語学習データの不足などの構造的な課題が存在する。社会で技術を安心して利用するためには、こうした基盤を自国で整備し、継続的に育てていく視点が欠かせない。

我々が国産LLMに取り組む背景には、このような課題認識と危機感がある。国産LLMは単なる技術開発ではなく、日本語を母国語とする社会全体のための知的基盤を構築する試みであり、教育研究に限らず産業全般を含めた未来のAI利活用の根幹を成す取り組みだと考えている。

1.「すべてをオープンに」――LLM勉強会の設立

2023年春、国内で生成AIへの関心が急速に高まる中、大学や研究機関に所属する自然言語処理の研究者たちが自主的な集まりとして始めたのが「LLM勉強会（LLM-jp）」である。海外ではLLM研究が急速に進む一方、日本では横断的な議論の場が乏しく、最先端の潮流から取り残されかねないという危機感があった。そこで特定の組織に縛られず研究者同士が気軽に知見を交換できるコミュニティを国立情報学研究所（以下NII）が中心となって立ち上げたのである。

LLM勉強会の最大の特徴は「すべてをオープンにするという姿勢」である。議論内容、メモ、ログ、実験過程、失敗例までも可能な限り共有し、参加者全員が学び合える環境を整えた。研究プロセスをオープンにする文化は国内ではまだ一般的ではなかったが、透明性を重視する姿勢は多くの研究者に支持された。週に複数回のオンライン会合が自然発生的に開かれ、計算資源の扱い方、データクリーニングのノウハウ、モデル評価の方法など、細かな知見が次々と蓄積されていった。

参加者は当初の30名ほどから急速に増加し、大学、企業、スタートアップ、学生、さらには海外在住の研究者まで含む幅広いコミュニティへと拡大し、国内最大級の研究者ネットワークとなり、現在2,500名を超える規模にまで成長している。この広がりは単なる人数の増加ではなく、日本のLLM研究文化そのものが「知識を囲い込むより共有する方が研究エコシステムは強くなる」という価値観への転換を示しているものと考える。

こうして形成された勉強会は、研究用データセット整備や評価基盤の構築、大規模モデル学習の実験といった組織横断の研究活動を支える基盤となった。その成果が130億パラメータの「LLM-jp-13B」の構築と公開である。

2023年7月から、データ活用社会創成プラットフォームmdxを活用して事前学習を進め、同年10月にモデルを公開した。学習には約3000億トークンのコーパス（日本語約1450億・英語約1450億・コード約100億）が用いられ、トークナイザーやウェブコーパスのフィルタリングツールも独自に整備した。

モデル構築にはMicrosoft DeepSpeed、ログ管理にはWeights & Biasesを採用するなど、最新の研究基盤を積極的に導入した。また「コーパス構築WG」「モデル構築WG」「チューニング・評価WG」など、大学横断の専門チームが連携して研究開発を進めた。

公開された「LLM-jp-13B」は研究初期段階のモデルではあるが、コーパス、モデル、ツール類をすべて公開するという姿勢は、国内の研究文化に大きなインパクトを与え

た。これにより、国内外の研究者が自由にモデルを試し、改善し、応用研究を行える土壌が整えられたと考えている。

2. 大規模言語モデル研究開発センター（LLMC）設立――政府支援で「研究基盤」を作る

この草の根の動きを政策として力強く後押ししたのが、文部科学省「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」事業である。同事業の下、2024年4月、NIIに「大規模言語モデル研究開発センター（Research and Development Center for Large Language Models、略称：LLMC）」が発足し、研究費7億円及び計算資源整備に約42億円が措置されるなど、基盤モデルを継続的に開発するための組織的・技術的基盤が整備された。LLMの開発には、多数のGPUや大容量ストレージ、広帯域ネットワーク、安定した電源・冷却設備からなる大規模な計算基盤が不可欠であり、これを単一の研究室や企業が整備するのは現実的ではない。LLMCは、ソフトウェアスタックを含むこうしたLLM構築のための計算基盤を全国の研究者が共有できる公共インフラとして整備し、研究コミュニティや産業界、行政機関と連携しながら、国全体の技術力向上を支える拠点として構想された。さらに、知識と経験を次世代に継承する人材育成の役割も担い、大学院生や若手研究者が先端的なLLM研究開発に参加できる環境づくりを進めている。

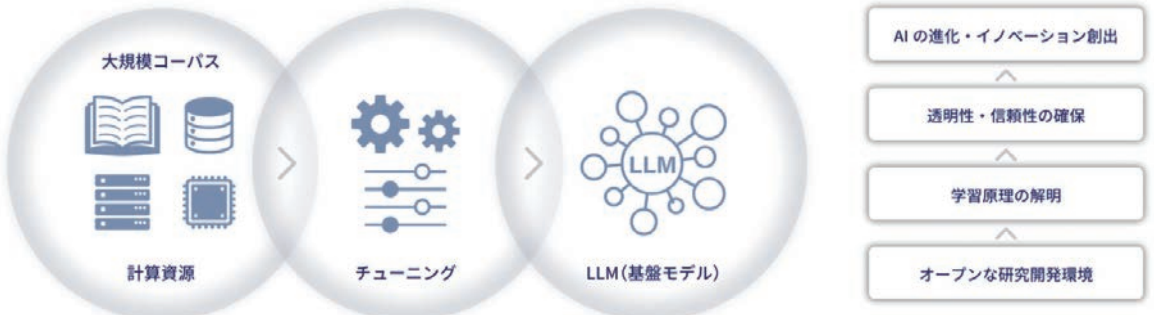
LLMCの研究課題は、大きく3つに整理される。第1は「研究開発用LLMの構築」であり、新規コーパスの整備やGPUを用いた大規模計算環境の構築、評価用ベンチマークの作成などを通じて、研究開発の基盤となるLLMを構築する。第2は「透明性・信頼性の確保に向けた研究開発」であり、生成AIの挙動原理の解明やデータバイアスの抑制、評価手法の整備などを通じて、生成AIの透明性と信頼性の向上を図る。第3は「高度化に向けた研究開発」であり、ドメイン適応やモデルの軽量化、新たなアーキテクチャの検討を通じて、生成AIモデルの性能向上と実用性の拡大を目指す。

特に公開性はLLMCの中核的価値であり、我々は「閉じて守る」よりも「開いてともに育てる」ことこそが、日本が国際的に存在感を発揮する戦略であると確信している（図）。

3. 172Bモデルの公開――世界最大級の「フルオープン」モデル公開

設立から半年後の2024年9月、我々は国産LLM開発の歴史における大きな節目を迎えた。これまでのデータ活用社会創成プラットフォームmdxでの130億パラメータモデルの学習、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によるAI橋渡しクラウド（ABCI）での1750億パラメータモデルの学習トライアルの成果を踏まえ、パラメータ数1720億（GPT-3相当）のモデル「LLM-jp-3 172B beta1」を公開したのである。本モデルの学習に用いたデータは、計約2.1兆トークンの大規模コーパスであり、日本語約5900億トークン、英語約9500億トークンのテキストに加え、コードデータ、専門文献、各種ドメインテキストを広範に収集して構成した。巨大なモデルの構築にとどまらず、日本語と多様な領域の知識を統合した総合的な基盤モデルを、日本から発信したことに大きな意義がある。

世界的に見ても、この規模のモデルで学習データまで公開した例は極めてまれである。モデルが大型化するほどデータの開示は困難になるが、我々はあえて「どのようなデータで学習し、どのような振る舞いが得られたのか」を誰もが検証可能な形にした。これは、科学研究の再現性を担保すると同時に、社会との信頼関係を築くために不可



■図．大規模言語モデル研究開発センターのミッション

欠な姿勢であり、公開型LLMを日本の戦略として位置付ける明確な意思表示でもあった。

モデル学習の前半では経済産業省・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）のGENIACプロジェクトの支援を受けたクラウド計算資源を、その後、文部科学省の補助金により整備されたさくらインターネットのクラウド計算資源を用いた。複数の省庁と民間企業が連携し、国内で1000億パラメータ級モデルの学習を実行できたことは、国産AI開発における基盤の成熟を示す象徴的な成果であったと言えよう。また、計算資源の確保、データ前処理、モデル設計、評価基盤の整備など、研究コミュニティとLLMCの協働が一体となって達成した点にも価値がある。

さらに、2024年12月には、フル学習版「llm-jp-3-172b-instruct3」を公開した。LLM-jp-evalによる日本語性能評価ではGPT-3.5を上回り、GENIACの評価ベンチマークでもGPT-3.5相当の性能を示した。公開モデルがこの水準に達したことは、国内外の研究者に強いインパクトを与え、日本においてもオープンな形で世界水準のLLMを開発できることを示した。

4. 安全性と透明性――日本が世界に提供できる知見

モデルの能力が高まれば安全性への懸念も一層大きくなる。LLMが誤った助言や有害な情報を生成するリスクは無視できず、社会が安心してAIを利用するために安全性の確保が性能向上以上に重要なテーマとなる。LLM-jpでは、安全性ワーキンググループを中心にモデルの出力を制御し、望ましくない挙動を抑制する研究を体系的に進めてきた。初期モデルでは、有害な質問に対して不適切な回答を返すケースが確認されていたが、244件、900件と段階的にインストラクションデータを拡充、1,800件規模のインストラクションデータを「AnswerCarefully」として体系化することで、有益性を保ちつつ安全性を高めるバランスを実現した。2025年には攻撃的なプロンプトを一般から募集するオンラインゲーム「Ailbreak（エイルブレイク）」を公開し、約1,200人の協力により58,000件もの攻撃データを収集した。これは「モデルを壊すためのデータ」をあえて集め、検証と改善につなげる試みである。多くの国や企業が脆弱性の公開に慎重な中、LLMCはあえて「開かれた安全性研究」を掲げ、攻撃データや改善手法を研究コミュニティと共有してきた。この姿勢は国際的にもまれであり、日本が安全性研究で存在感を示す重要な要因となっている。

こうした取り組みは、モデルの安全性向上にとどまらず「安全に使えるAIとは何か」という社会的理解を深める役割も担う。危険な出力の特定や発生要因、防止策といったプロセスを公開することで、透明性の高いガバナンスモデルを育てることができる。安全性の議論が組織内部に閉じれば、社会との信頼関係を築くことは難しい。だからこそLLMCは安全性を「開かれた研究領域」として位置付けているのである。

G7広島サミット以降、AIの安全性やガバナンスを巡る国際的議論は急速に加速した。日本政府もAIセーフティ・インスティテュート（AISI）を設立し、国産モデルの評価や国際連携を本格化させている。LLMCとしては、実際のモデル開発で得られた知見をこうした政策議論にフィードバックし、日本発の安全性研究を国際基準づくりに生かしていきたいと考えている。公開型LLMを推進する日本だからこそ示し得る安全性のアプローチがあるとの認識である。

5. 医療分野への応用――日本語で学んだモデルだからできること

国産LLMの強みが最も生きる領域の1つが医療である。LLMC副センター長の相澤彰子教授らが内閣府「戦略的イノベーション創造プログラム（SIP）」第3期「統合型ヘルスケアシステム」で開発した日本語医療LLMは、国内の医療文献や症例、ガイドラインなどで追加学習を行い、医師国家試験5年分で合格水準を超え、平均でGPT-4を上回る性能を示した。

医療現場では、退院サマリーの作成、紹介状の作成、検査計画の整理など日本語テキスト情報を扱う業務が多い。人手不足が深刻な医療現場において、こうした作業を国産LLMが一部担うことができれば、医療従事者が患者と向き合う時間の確保につながる。一方で、個人情報保護や医療機器認可といった法的・倫理的課題も多く、研究者だけでは解決できない。だからこそ医療者、法学者、技術者が同じ場で議論しながらモデルを育てる「共創」の枠組みが重要となる。

6. 計算資源と日本語データ――国産LLMを持続させるための条件

こうした成果が積み重なる一方で、国産LLM開発を持続的に進めるためには、いくつかの構造的課題に向き合う必要があると実感している。その1つが計算資源の安定確

保である。LLMの学習には、多数のGPUに加え、長期間の連続稼働を支える電力・冷却・ネットワーク・ストレージなどの計算基盤が不可欠だ。国内でも整備は進められているが、モデルの大型化に伴い、計算基盤の逼迫と安定運用は依然として大きな課題となっている。

もう1つの重要な課題は、日本語データの量と質の不足である。英語圏では研究利用可能な自然文データが豊富に公開されているのに対し、日本語では依然として利用可能なデータが限られている。十分なコーパスがなければ、言語表現の多様性を学習できず、モデル性能にも影響する。モデル自身が生成した文章を学習に取り入れる合成データの利用など技術的な改善は着実に進んでいるが、国産LLMを継続的に育てるためには、研究利用可能な日本語テキストを社会全体で拡充し、適切に保管・共有する仕組みが不可欠だと感じている。

著作権への不安も国産LLM開発において丁寧に向き合うべき重要なテーマである。我々の検証ではLLMが学習したテキスト原文をそのまま出力するケースは極めてまれであることが明らかになりつつあるが、技術者側の説明が不足すれば社会的な不安は解消されない。モデルの学習や生成の仕組みを明確に示し、社会と継続的に対話していく姿勢が求められる。

これらの課題は、国産LLMが成熟する過程で必然的に直面するものであり、長期的な協力の下で改善していくべき領域である。我々は計算資源、日本語データ、社会的理解という3つの基盤を整えることで、国産LLMがより強固な知的基盤へ成長すると考えている。

7. 国産LLMに期待するもの ――開かれた基盤としての役割

国産LLMの開発は、単に海外に追いつくことを目的とした取り組みではない。我々が目指すのは、日本語や日本社会・文化を深く理解し、それを世界に発信できる知の基盤を構築することである。言語は情報伝達の手段にとどまらず、価値観や歴史、社会の仕組みを内包している。国産LLMを育てることは、こうした文化的背景を含む日本語の全体性を次世代へ継承する試みでもある。モデルやデータを公開し、研究コミュニティ、企業、行政が共通基盤として活用できる環境を整えることが、日本のイノベーションを支える土台になると考えている。国産LLMを開かれた形で育てることで、多様な主体による創意工夫と社会実装が促される。

そのためには研究者だけでなく、企業や行政、利用者を含む多様な主体が共通の理解を持つことが不可欠である。具体的にはモデルの学習や出力生成の仕組みを透明に説明できる環境づくりが重要となる。こうした透明性と対話を基盤とした取り組みは、AIを社会で安心して利用するためのガバナンスの在り方を形づくる。

最後に1つ提案を述べたい。国産LLMを単なる「国産品」として完結するのではなく、世界に開かれた研究の結節点として育てていくことこそ、日本らしい貢献の形ではないだろうか。ソブリンAIに取り組む各国の知見やリソースを共有し、多言語・多文化の研究者が参加できる国際的なエコシステムを構築することで、国産LLMが国内外に開かれた形で成熟し、日本社会の共有財産として成熟していく。我々は今後も、技術を社会とともに育てる姿勢を大切にしながら、日本発のAI研究の未来に責任を持って取り組んでいきたい。