

データ基盤から知識基盤へ

① ビジョンの概要

多様な研究者の協働により学術研究が総合知として深化し、複合的な社会課題を解決していくためには、ネットワーク基盤及び研究データ基盤のさらなる進化を推し進めるとともに、AI を活用し、データの解釈、知識への汎化、知識の関係付け・体系化を自動化し、分野を横断する新たな知の創造を支援する知識基盤の構築が必要である。

② ビジョンの内容

21 世紀の学術及び社会の大きな潮流として、データの重要性が明確に認識された。さまざまな観測や計測からデータを作成し、デジタル化し、オープンにして議論・利活用することで学術的に大きな進展が起きている。このような状況下で、我が国では国立情報学研究所（以下 NII）を中心にネットワーク基盤（SINET6）と、研究データ基盤（NII RDC: NII Research Data Cloud）の整備が継続的に進められ、データ駆動による学術変革が強力に推進されてきた。

様々な科学技術の協働を進める上で問題となるのは、特定分野の専門家も他分野については素人であり、多様な分野を見通してデータを直に活用することは容易ではないという点である。これからの学術研究が総合知として深化し、複合的な社会課題を解決していくためには、AI を活用し、データの解釈、知識への汎化、知識の関係付け・体系化を自動化し、分野を横断する新たな知の創造を支援する知識基盤の構築が必要である。そのような基盤の必要性は 10 年以上前から指摘されてきたが、データ基盤が整い、データをオープンにすることの価値が認識され始め、AI の急速な進展が人間の知的活動の在り方を根底から変容させる激動時代を迎え、今後の科学研究の進展を支える AI 駆動の知識基盤を早急に実現することが求められている。

本構想では、NII を中心に、ネットワーク・研究データ基盤のさらなる進化を推し進めるとともに、大規模計算資源と AI エンジニアを確保して AI 基盤モデルの研究・開発・運用体制を整備し、学術の知識基盤を構築する。これによって多様な研究者の協働によるさらなる学術の深化と様々な社会課題の解決を実現する。

③ 学術研究構想の名称

データ基盤から知識基盤へ

④ 学術研究構想の概要

本構想は、学術研究を総合知として深化させ、複合的な社会課題を解決していくため、知識そのものを扱う知識基盤を構築し、これによって学術分野の横断的連携を加速度的に推進し、学術分野全体の振興を図る。

知識基盤のベースは深層学習に基づく AI 基盤モデルである。ウェブ等の超大規模コーパス、人文・社会科学分野を含む論文、生体分子配列を始めとするデータ配列等の様々なデータによって AI 基盤モデルを構築し、データの解釈、知識への汎化、知識の関係付け・体系化などを、知識基盤の中核機能として実装する。さらに、AI 基盤モデルから学術情報全般に関する外化された明確な知識体系として大規模学術知識グラフを生成し、AI 基盤モデルと学術知識グラフを統合した知識基盤サービスを多様な分野の研究者に提供する。

また、知識基盤があらゆる学術分野の発展と新しい領域の開拓に資するために、広範な研究分野から大量のデータを研究データ基盤に収集・管理し、AI 基盤モデルに提供可能な次世代研究データ基盤を構築する。さらに、大量のデータの流通に不可欠な従来の地上系ネットワーク（有線、モバイル網）の増強に加え、通信エリアを拡張する非地上系ネットワーク構築を行う。

⑤ 学術的な意義

個々の研究者が深く理解し得る知識範囲には限界があり、異分野の専門家とのコミュニケーションも難しい。一方で、AI 基盤モデルは様々な分野の論文の大規模学習で、医学にも物理学にも、法学にも経済学にも精通したモデルとして、分野横断知識の関連付け、体系化が可能である。その対話能力も含め、AI 基盤モデル



図1 本構想の概要

ルは既に学術分野の垣根を超えた連携を行う研究パートナーとして、学術の振興に大きく寄与し始めている。

さらに、人文・社会科学や臨床医療を初めとする学術分野、研究成果の社会実装や総合知による社会課題解決においては、日本の社会制度や文化に根差し、総合的にも世界モデルと大差ないソブリン AI 基盤モデルが不可欠である。日本の学術研究の主権を確保し、日本文化の地位を守る意味でも、我が国における AI 基盤モデル構築・運用体制確立は大きな意味をもつ。

NII は、2023 年から開かれた LLM 研究コミュニティ LLM-jp を運営し、2024 年には文部科学省の生成 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 LLM 研究開発センターを設置し、国産 LLM の研究開発を進めている。本構想は、これらの成果をさらに発展させるものである。

知識基盤を支えるネットワーク・研究データ基盤は、我が国の幅広い学術分野を支える安全・安定な基盤として、研究者の活動を支援し、研究を効率化している。ネットワーク基盤は、国内外の大型研究施設、全国のスパコン、クラウドの共同利用をさらに円滑化し、高エネルギー分野でノーベル賞等世界トップレベルの研究にも寄与している。これらの基盤のさらなる高度化と、広範な分野の研究データから AI 基盤モデルを構築し、知識基盤として提供することで、新しい学術分野の開拓を支援できる。さらに、データ収集エリアを海・空・宇宙へ広げる非地上系ネットワークは、例えば、地球温暖化等の地球規模問題の根本的対策検討に必須である。

⑥ 国内外の研究動向と当該構想の位置付け

既に分野固有データで学習した基盤モデルが AI for Science の中核的基盤として整備が進んでいるほか、汎用生成 AI においても、専門的な学術分野に関する高度な検索や推論が可能となり、それらを論文検索サービス等と結びつけた研究支援 AI が商用サービス提供されている。

欧州では、知識の構造化を通じてオープンサイエンスを実践し、科学的知識の発見、アクセス、相互運用、再利用（FAIR）の実現に取り組んでいる。人文・社会科学でも、都市空間を時系列を含む四次元で表現する「4D 都市モデル」と生成 AI/LLM を接続し、文化遺産データの知識化と AI 連携に挑んでいる。

ネットワークについては、米国の大学連合の Internet2 やエネルギー省の ESnet を初めとして、欧州、アジア、南米等が高速ネットワークの導入に積極的であり、非地上系ネットワークにも強い関心を示している。

こうした状況の中、我が国が国際的優位性を確立するため本構想の実現は不可欠である。

⑦ 社会的価値

今後、人間と AI の共存は必須である。本構想による総合知の深化と社会課題解決は国民の幸福に資するだけでなく、AI 基盤モデルの開発・運用は経済的・産業的にも直接的貢献が期待できる。また、知識基盤に関する情報の信頼性を担保するための技術は、AI により生成される多種多様なフェイクメディアがもたらす社会的脅威の解決に資するものである。

SINET や NII RDC 等の基盤の構築は、学術研究分野のみならず、社会的基盤、教育支援に対する効率的な投資であり、国民の理解を得られるものである。

⑧ 実施計画等について

【実施計画】

2026～2029 年度 知識基盤及び研究データ・ネットワーク基盤の新機能を開発し、運用開始

2030～2032 年度 知識基盤の機能充実、研究データ・ネットワーク基盤の性能最適化、高安定化

2033～2035 年度 知識基盤及び研究データ・ネットワーク基盤の高度化

【実施機関と実施体制】

NII を中心に主要大学・研究機関等と協力して進める。さらに、学術の中長期戦略 No. 31（生命科学クロスオーバー研究旗艦拠点の設立計画）とも連携する。また、学際大規模情報基盤共同利用・共同研究拠点（JHPCN）、国立大学法人情報系センター協議会、大学 ICT 推進協議会とも連携し、ユーザコミュニティの意見やニーズを幅広く吸い上げ、本構想に反映する。

【総経費】 1,400 億円

⑨ 連絡先

黒橋 禎夫（国立情報学研究所）