

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、ネットワーク基盤とデータ基盤のさらなる進化を推し進めるとともに、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世紀の学術およ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を中心に SINET6 に至るネットワーク整備と、研究データ基盤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、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、さらに機械翻訳研究に端を発するAI基盤モデルにより論文やマルチメディアデータを高度に解釈することが可能となりつつあることから、ついに知識基盤の構築を本格的に目指すべき時代となった。

国立情報学研究所を中心に、ネットワーク基盤と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・運用の体制を整備し、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

本研究構想は、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に、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はかる。

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・

体系化を自動化する。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処理APIの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、国際協調に基づく非地上系ネットワーク構築を行う。さらに、ネットワーク状態の高度診断・障害予兆検知、クラウド・エッジサー

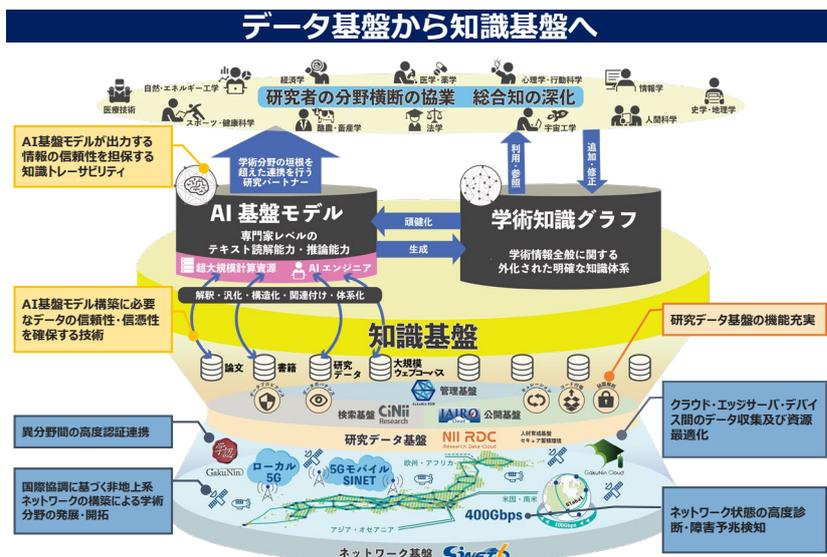


図1 本研究構想の概要

バ・デバイス間のデータ収集及び資源最適化、異分野間の高度認証連携の実現などを行う。

⑤ 学術的な意義

人間の専門家の専門性には限界があり、異分野の専門家とのコミュニケーションは必ずしも簡単ではない。一方で、AI 基盤モデルは多様な学術分野の論文で大規模な学習を行うことにより、医学にも物理学にも、そして法学にも経済学にも精通したモデルとなりうる。AI 基盤モデルはすでに類似性判定や因果関係認識で高いパフォーマンスを示している。さらに、類推（アナロジー）的な機能を備えることにより、分野横断の知識の関連付け、体系化を実現することができる。その対話能力を含めて、AI 基盤モデルが学術分野の垣根を超えた連携を行う研究パートナーとなり、学術の振興に大きく寄与することは想像に難くない。

さらに、現在の AI 基盤モデルは英語中心であり、学習コーパスに含まれる日本語テキストが 1%未満に過ぎないことから日本語の処理能力は限定的である。一方、人文・社会科学や臨床医療の現場では日本語の使用が中心であり、今後の世界における日本語の地位、日本文化の地位を守る意味でも、我が国における AI 基盤モデルの構築・運用体制の確立は大きな意味をもつ。

知識基盤を支える高速で広範囲のネットワーク基盤の整備は、国内外の大型研究施設（LHC、Spring-8、J-PARC、BelleII、スーパーカミオカンデ、KAGRA、eVLBI、LHD、ITER 等）、全国のスパコン（情報基盤センター、富岳、ABCi 等）、全国のクラウドの共同利用をさらに円滑化する。高エネルギー分野においてノーベル賞受賞を含む世界トップレベルの研究に寄与した実績に加え、地上だけでなく海・空・宇宙からのデータ収集基盤の実現により新しい学術分野の開拓を支援することができる。

⑥ 国内外の研究動向と当該構想の位置付け

OpenAI、Google、Meta などの独占状態である汎用的知識処理を目指す大規模な AI 基盤モデルにおいては、分野横断の学術研究利用などは始まっていない。

AI 基盤モデルを支える研究データ基盤については、欧州や豪州が先行しているが、知識の中身のレベルの構造化への取り組みではない。

ネットワーク整備については、米国の大学連合の Internet2 やエネルギー省の ESnet をはじめとして、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関しても強い関心を示している。

⑦ 社会的価値

今後、人間と AI の共存は必須である。本研究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ものであるが、加えて、AI 基盤モデルの開発・運用は経済的・産業的にも直接的な貢献が期待できる。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。

⑧ 実施計画等について

【実施計画】

R5－R8：検討・モデル開発・機能開発

R9－R11：検証・分析・改良

H12－R14：高安定化・性能最適化

【実施機関と実施体制】

国立情報学研究所（以下 NII）が主たる実施機関となり、主要な大学・研究機関等との協力関係を確立して進める。また、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、本研究構想に反映させる。

【総経費】 1,400 億円

⑨ 連絡先

黒橋 禎夫（国立情報学研究所）