

2.1.6 AI・データ駆動型問題解決

(1) 研究開発領域の定義

人工知能（Artificial Intelligence：AI）・ビッグデータ解析が可能にする大規模複雑タスクの自動実行や膨大な選択肢の網羅的検証などによる、問題解決手段の質的变化、産業構造・社会システム・科学研究などの変革を生み出す研究開発領域である。本節では、「AI 駆動」「データ駆動」を冠して呼ばれることが多い、さまざまな問題解決に共通的な考え方やフレームワーク・基盤技術を中心に俯瞰し、個別的・具体的なアプリケーションは最近の注目トピックのみ取り上げる。

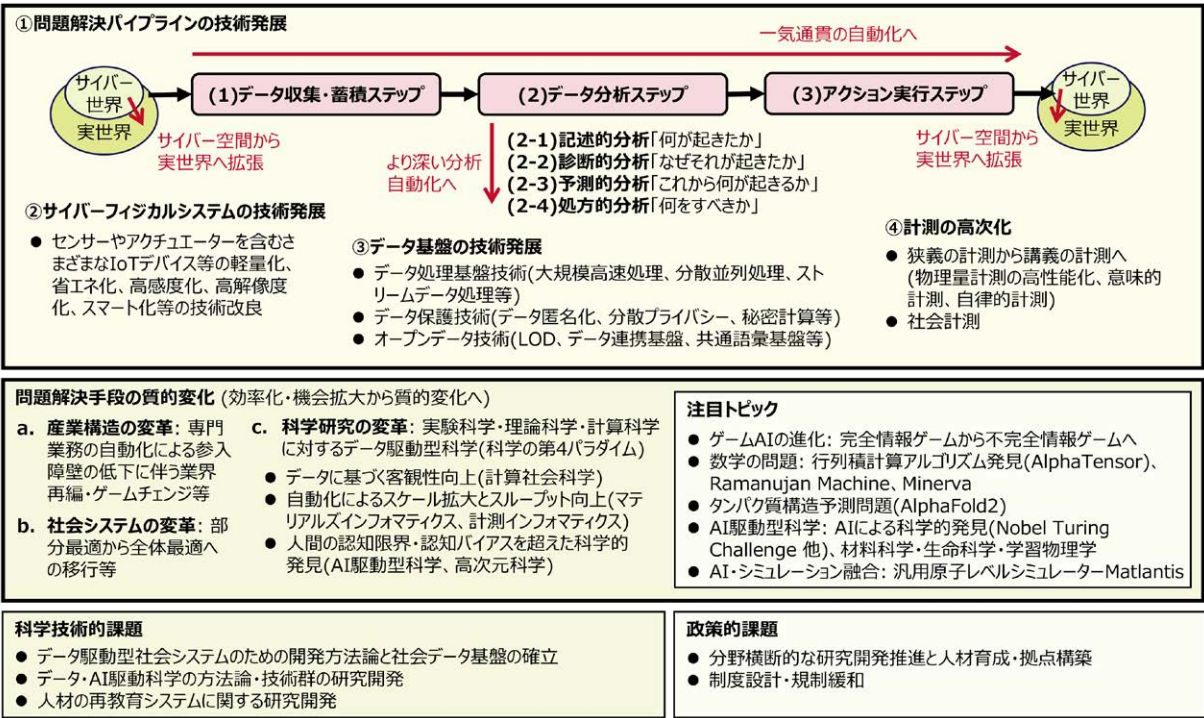


図2-1-9 領域俯瞰：AI・データ駆動型問題解決

(2) キーワード

ビッグデータ、Cyber Physical Systems（CPS）、IoT（Internet of Things）、データサイエンス、オープンデータ、データ連携基盤、データ駆動、AI 駆動、デジタルトランスフォーメーション（DX）、計測、ゲームAI、AI・シミュレーション融合、AI 駆動型科学

(3) 研究開発領域の概要

[本領域の意義]

ビッグデータ（Big Data）^{1), 2)} は、元来は膨大な量のデータそのものを指す言葉だが、その収集・蓄積・解析技術は、大規模性だけでなくヘテロ性・不確実性・時系列性・リアルタイム性などにも対応できる技術として発展している。また、センサー、IoT（Internet of Things）デバイスの高度化と普及によって、さまざまな場面で実世界ビッグデータが得られるようになり、その収集・解析技術は、実世界で起きる現象・活動の状況を精緻かつリアルタイムに把握・予測するための技術としても期待されている。今日、さまざまな社会課題が人間の手に負えないほどに大規模複雑化しており、実世界ビッグデータの収集・解析による状況の把握・予測は、そのような課題の解決に共通的に貢献し得る有効な手段になる。ここにさらにAI 技術が加わ

り、AI技術とビッグデータ（データそのもの、および、処理技術）が深く関係し合いながら発展している。すなわち、ビッグデータが集められることでAI技術（特に機械学習技術）は高度化し、精度を高め、そのAI技術を用いて実世界のビッグデータを解析することで、実世界の現象・活動のより深く正確な状況把握・予測が可能になってきた。

具体的なアプリケーションは、当初、Googleなどのサーチエンジンにおける検索連動型広告や、Amazonなどのショッピングサイトにおける商品レコメンデーションのように、インターネット上のサービスに集まるビッグデータを売り上げ向上に活用するものが中心であった。しかし、現在は、実世界から集まるビッグデータを活用した社会課題解決へと広がってきており^{1), 2)}、その社会的価値はますます高まっている。例えば、電力・エネルギーの需要を予測して最適に制御したり、実店舗のさまざまな商品の品ぞろえや仕入れを最適化したり、防犯のため不審な人や振る舞いを検知・通知したり、病気の疑いや機器の異常を早期に検知したりといった実世界のアプリケーションが広がっている。

わが国がビジョンとして掲げる「Society 5.0」は、内閣府によると、「サイバー空間とフィジカル空間（現実）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、人間中心の社会（Society）」であり、サイバー空間とフィジカル空間の高度な融合は「フィジカル（現実）空間からセンサーとIoTを通じてあらゆる情報が集積（ビッグデータ）、人工知能（AI）がビッグデータを解析し、高付加価値を現実空間にフィードバック」によって実現するとされている。これにより、交通、医療・介護、ものづくり、農業、食品、防災、エネルギーなど、さまざまな分野で新たな価値創出が目指されている。昨今、産業界を中心にデジタルトランスフォーメーション（DX）の推進が叫ばれているが、Society 5.0と方向性を同じくする動きであり、産業構造、社会システム、科学研究などに変革をもたらす。そして、これらをドライブするのが、本節で述べるAI・ビッグデータを活用した「AI駆動型」「データ駆動型」と呼ばれるアプローチである。

【研究開発の動向】

AI・データ駆動型の問題解決の基本的な枠組みを踏まえて、その研究開発の動向を、①問題解決パイプラインの技術発展、②サイバーフィジカルシステムの技術発展、③データ基盤の技術発展、④計測の高次化という4面から述べる。次に、このような発展が生み出す問題解決手段の質的变化が、産業構造の変革、社会システムの変革、科学研究の変革をもたらす可能性とその状況について述べる。

① 問題解決パイプラインの技術発展

AI・データ駆動型問題解決の基本的な処理の流れは、（1）データ収集・蓄積ステップ、（2）データ分析ステップ、（3）アクション実行ステップ、という順に進む。ここでは、これを「問題解決パイプライン」と呼ぶことにする。（2）のデータ分析ステップは、さらに、データ分析の深さによって段階がある。米国の調査・アドバイサリー企業であるGartnerは、データ分析の段階を、（2-1）記述的分析（Descriptive：何が起きたか）、（2-2）診断的分析（Diagnostic：なぜそれが起きたか）、（2-3）予測的分析（Predictive：これから何が起きるか）、（2-4）処方的分析（Prescriptive：何をすべきか）という4段階としている³⁾。（2-4）によってアクションが計画され、（3）のアクション実行が可能になる。

問題解決パイプラインで、（1）→（2-1）→（2-2）→（2-3）→（2-4）→（3）とステップを深めるほど、問題の解決に近づき、社会価値・ビジネス価値が高くなる。つまり、例えば（1）（2-1）しか自動化されなければ、（2-2）以降は人間が行うことになるが、（1）から（3）まで一気通貫で自動化されれば、人間は実行状況をモニタリングしていればよいことになる。電力マネジメントの例で具体的に説明するならば、前者のケースは、電力消費状況の計測・可視化までが自動化され、その状況に基づいて人間が今後の必要量を判断し、アクションを考えることになる。後者のケースは、電力消費状況を自動計測し、今後の必要量を自動予測し、最適な状況になるように自動制御も行われる（人間はその様子を見ていけばよい）。

このような（1）から（3）まで一気通貫での自動化を可能にする方向で、技術開発が進められている。

2.1

俯瞰区分と研究開発領域 人工知能・ビッグデータ

そのために使われる具体的な技術としては「2.1.1 知覚・運動系のAI技術」に記載されている機械学習・パターン認識・運動生成などの技術が挙げられる。

② サイバーフィジカルシステムの技術発展

前述したように、問題解決パイプラインは、当初、インターネット上のサービスに集まるデータを収集・解析し、そのサービスを改良・強化するために使われた。つまり、サイバー空間に閉じたパイプラインであった。しかし、現在は実世界（フィジカル空間）からデータを収集し、その解析結果に基づいて、実世界のシステムにフィードバックをかけるような応用へも広がっている。つまり、サイバーフィジカルシステム（Cyber Physical Systems：CPS）としての問題解決パイプラインへと拡張されている。

この拡張は、問題解決パイプラインにおける、（1）データ収集・蓄積ステップと（3）アクション実行ステップが、サイバー空間から実世界に広がったということである。そのために、センサーやアクチュエーターを含むさまざまなIoTデバイス、あるいは、ロボットが（1）や（3）に導入されるようになった。軽量化、省エネ化、高感度化、高解像度化、スマート化などの技術改良が進められている。

③ データ基盤の技術発展

上記①②のような問題解決パイプラインを支える技術として、データ基盤の研究開発も進められている。ここでいうデータ基盤は、a.データ処理基盤技術、b.データ保護技術、c.オープンデータ技術を含む。

a. データ処理基盤技術：

大規模なデータを高速に処理するための技術群である。ますます大規模化するデータを、より高速に処理するという要求が高まり、分散並列処理技術、圧縮データ処理技術、ストリームデータ処理技術などが発展している。その具体的な技術内容は「2.2.3 AIを支えるコンピューティングアーキテクチャー」に記載した。

b. データ保護技術：

分析対象となるデータの保護のための技術群である。暗号化などのセキュリティ技術に加えて、分析対象データが個人属性や行動履歴のようなパーソナルデータである場合に、そのプライバシーを保護するための技術、さらには、データの分析と保護を両立させるプライバシー保護データマイニング技術が開発されている。匿名化、差分プライバシー、秘匿計算などの技術がある。それらの詳細は「2.1.9 社会におけるAI」に記載した。

c. オープンデータ（Open Data）：

最小限の制約のみで誰でも自由に利用、加工、再配布ができるデータのことであり、さまざまな問題解決にデータ利用が促進され、また、他のデータと組み合わせた新しい価値創出・サービス創出が活性化される。そのために、共通的なデータ形式や付加的な情報（メタデータなど）の記述形式などがデザインされている。特に、セマンティックWeb分野で開発・標準化された技術を用いたリンクトオープンデータ（Linked Open Data：LOD）⁴⁾がよく知られている。さらに、分野・組織をまたいだデータの連携を容易にするため、共通語彙の設定も含むデータ連携基盤の構築が推進されている。米国では2005年にNIEM（National Information Exchange Model）、欧州では2011年にSEMIC（Semantic Interoperability Community）がデータ連携標準の取り組みとして始まった。わが国でも「未来投資戦略2018」で描くデータ駆動型社会の共通インフラとしてデータ連携基盤の構築が掲げられ、共通語彙基盤（Infrastructure for Multilayer Interoperability：IMI）が構築された⁵⁾。IMI、NIEM、SEMICの間の国際的な相互運用性も検討されている。

2.1

俯瞰区分と研究開発領域 人工知能・ビッグデータ

4 計測の高次化

計測は「科学の母」(Mother of Science) と言われ、さまざまな科学研究を支えている。また、現在の状況を計測 (センシング) することは問題解決 (ソリューション) の出発点であり、計測技術はさまざまなソリューションビジネスを左右する。今日、①で述べた問題解決パイプラインに沿って、計測は「狭義の計測」から「広義の計測」へと概念を広げ、これがさまざまな研究やビジネスに波及している。

「狭義の計測」は物理量計測である。従来の計測機器は、温度・重量などの物理量を直接計測して出力するものであった。しかし、いまでは、計測機器 (あるいは計測システム) の中で、物理量データの統計処理・データ分析処理などの情報処理 (AI・ビッグデータ技術の適用を含む) まで行い、その結果を計測結果として出力するものが増えている。そのような情報処理を加えることで、a.物理量計測の高性能化、b.意味的計測、c.自律的計測が可能になってきた。ここで、意味的計測と自律的計測が「広義の計測」に相当する。以下、これらa・b・cについて簡単に説明する (詳細は調査報告書⁶⁾ にまとめた)。

- a. 物理量計測の高性能化は、従来と同様に物理量を計測結果として出力するが、情報処理を加えることで、精度や効率を高めるものである。例えば、カメラ画像の超解像 (画像処理によって解像度を高める) などがある。
- b. 意味的計測は、計測結果として得られた物理量データを分析することで、その計測結果に意味を与える (上位概念に変換する) ものである。位置の計測データ (座標) の住所・ランドマークへの変換や、指紋認証・顔認証などのバイオメトリクス認証機器・システムがその一例である。
- c. 自律的計測は、物理量データの分析結果に基づいて、次のアクションの決定・実行まで行うものである。例えば、現在の計測結果に基づいて、次に何を計測するかを決定するような、ロボットやドローンをベースとした自律的な計測システムがこれに該当する。

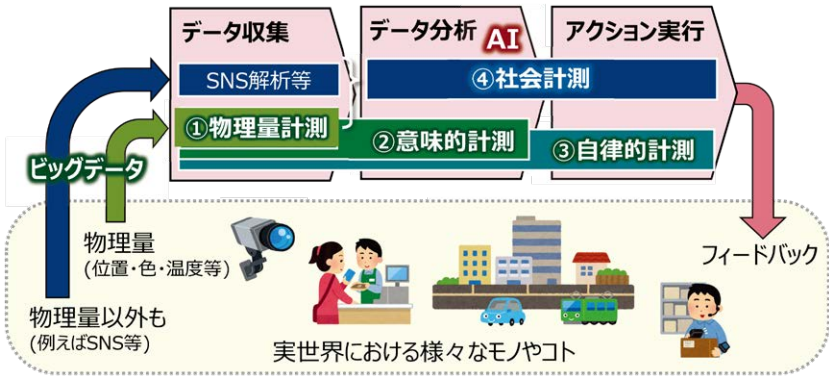


図2-1-10 計測の高次化と問題解決パイプラインとの対応

また、上記a・b・cは、物理量計測を出発点とした計測の高次化であるが、物理量計測だけでなく、人々がSNS (Social Networking Service) やCGM (Consumer Generated Media) で発信する情報も集めて、人々の行動や社会現象を把握しようというアプローチ (「社会計測」とも呼ばれる) も生まれている。さらに、「広義の計測」や「社会計測」で得られた人間や社会に関するビッグデータを分析して、人間の行動や社会の現象を定量的に理解しようとする計算社会科学も、近年、取り組みが活発になっている。このような計測の高次化と合わせて、その自動化によって、規模の大きな現象・活動のリアルタイムな計測という方向への発展も進んでいる。

5 問題解決手段の質的变化

①②③④で述べたようなAI・データ駆動型問題解決の枠組みの発展によって、大規模複雑タスクの自動実行や膨大な選択肢の網羅的検証などが可能になり、問題解決手段の質的变化が起き、産業構造の変革、社会システムの変革、科学研究の変革にもつながる。

AI・データ駆動型問題解決の枠組みは、既に多くの業種・分野に広がっており、さまざまな種類、多数の事例が生まれている⁷⁾。それらの事例では、従来人手で行っていた作業を自動化することで効率化が進んだり、自動化に加えて、膨大なデータを精緻に観察・分析することによる精度向上によって適用場面（ビジネス機会）が拡大したりと、効率化・機会拡大の効果がまずは見られる。しかし、それにとどまらず、産業構造・社会システムなどの転換を引き起こすような質的变化も起こる。効率化（コスト削減など）と機会拡大（売り上げ拡大など）は従来の土俵の上での競争だが、この質的变化は土俵を変える（ゲームチェンジが起きる）。このゲームチェンジに備えるための打ち手、さらには、ゲームチェンジを主導するための打ち手が、技術開発と制度整備の両面から求められる。以下、a.産業構造の変革、b.社会システムの変革、c.科学研究の変革という三つの面で、質的变化の可能性に着目する。

a. 産業構造の変革

産業構造の変革については、人工知能技術戦略会議が2017年3月に公開した「産業化ロードマップ」が参考になる。このロードマップは、主要分野として、生産性分野、健康・医療・介護分野、空間の移動分野という3分野を取り上げ、各分野でのAI利活用の進展を三つのフェーズで捉えている。フェーズ1では、各領域においてデータ駆動型のAI利活用が進み、フェーズ2では、個別の領域の枠を越えてAI・データの一般利用が進み、フェーズ3では、各領域が複合的につながり合ってエコシステムが構築されるとしている。フェーズ1・2はおおむね効率化と機会拡大に相当し、フェーズ3はゲームチェンジが起こり得る質的变化の段階に相当する。AI技術（特に機械学習技術）によって各業界の専門的業務が自動化され得る。これはその業界にとって業務効率化だが、業界外から見れば参入障壁の低下になるため、業界構造が変わり、ゲームチェンジが起こり得る。例えば、動画の効果的な加工はかつて専門業者に依頼するしかなかったが、いまでは個人がスマートフォンで簡単に加工操作でき（その裏ではAI的な技術が使われている）、それを用いた新たなサービスも生まれている。また、UberやLyftに代表されるような自家用車によるオンデマンドのライドシェアも、従来の業界構造を変えた事例だが、このゲームチェンジを可能にしたのは、一般ドライバー（自家用車の所有者）と車で移動したい一般ユーザー（乗客）とを、リアルタイムに把握して最適マッチングする仕組みが、AI・ビッグデータ活用によって実現されたからである。

なお、さまざまな産業分野の変革に影響を与えるようなAI技術については、新エネルギー・産業技術総合開発機構（NEDO）が2022年2月に発表した「人工知能（AI）技術分野における大局的なアクションプラン」が参考になる。

b. 社会システムの変革

AI・ビッグデータの活用によって、社会システムの部分最適から全体最適への移行が考えられる。IoT技術の進化と普及によって、社会のさまざまな事象がビッグデータとして精緻かつリアルタイムに観測できるようになる。その一方で、さまざまな社会システムが相互に接続し合ったり、影響を与え合ったりするようになり、大規模で複雑な系をAI技術で全体最適化する方向が考えられる。大規模複雑なシステムを個別で精緻な観測に基づきながら全体最適化を行うことは、人間には困難であり、質的变化が生まれると考えられる。何を最適と考えるかという価値観は国・地域・文化や個人個人によって異なるため、電力・水道・交通などのライフライン系は共通的な方針のもとでの全体最適な供給制御が考えられるが、より個人の生活スタイルに関わる部分は各自の価値観に任せるべきものとなる。

2.1

俯瞰区分と研究開発領域 人工知能・ビッグデータ

c. 科学研究の変革

科学には四つのパラダイムがあると言われる⁸⁾。第1のパラダイムは実験科学（あるいは経験科学）、第2のパラダイムは理論科学、第3のパラダイムは計算科学（あるいはシミュレーション科学）、第4のパラダイムはデータ駆動型科学（あるいはEサイエンス）と呼ばれる。データ駆動型科学は、データに基づいて科学的な知見や社会的に有益な知見を導き出そうとするアプローチを取る。これはAI・データ駆動型問題解決の発展によって生まれた新しい科学のパラダイムだと言える。以下、データ駆動型科学によってもたらされ得る科学の質的变化として考えられる点を挙げる。

第1点として、さまざまな現象・事象についてビッグデータが取得できると、従来は人間の主観や限られた観察に強く依存していたタイプの学問や施策設計が、データに基づく客観性の高い分析・検証を行えるようになる。計算社会科学のような研究分野が立ち上がっているのがその一例である。

第2点として、AI技術とロボット・IoT機器などを活用した高度な自動化によって、人間には不可能なスケールとスループット、すなわち組み合わせ的に膨大な数の条件・ケースに対して高速な実験・仮説検証の繰り返しが可能になる。例えば、マテリアルズインフォマティクス⁹⁾や計測インフォマティクス^{6), 10)}と呼ばれる取り組みでは、このような面が生かされている。

第3点として、人間の認知限界・認知バイアスを越えた科学的発見がもたらされる可能性がある。科学研究においても、自分の研究に関連したすべての論文を読むことは不可能であり、自分の仮説に合うデータのみに着目したり、想定に合わなかったケースのみ厳しくチェックしたりといった認知限界や認知バイアスがあり、それが科学的発見の可能性を狭めているという指摘がたびたびなされている^{11), 12), 14)}。AI技術を活用すれば、このような限界・バイアスを越えた仮説探索・検証が可能になり、これまでと質的に異なる科学的発見が生まれるかもしれない。

この第3点（および第2点を含めることもある）を強調して「AI駆動型科学」¹⁵⁾や「高次元科学」¹⁴⁾という呼び方がされることもある。また、計算科学（シミュレーション科学）にAI技術を組み合わせた「AI・シミュレーション融合」も活発な取り組みと発展が見られる。これらについては〔新展開・技術トピックス〕**②**で、最近の注目トピックを紹介する。

〔論文や特許の動向〕

研究開発領域別の論文・特許調査の結果において、本領域は、論文数・特許数とも2018年前後から急増している。論文数・論文数シェアとも、国別では中国と米国が二強状態になっている。欧州についても、直近では中国・米国より若干下回る。論文数・論文シェアとも2012年当初から中国が米国を上回っている。本領域はAI関連領域の中で最も応用色の強い領域であり、AI応用において中国が勢いのあることが表れたものと考えられる。その一方、2021年はかなり競った状況になったものの、Top1%論文数・Top10%論文数は米国優位が続いてきており、中核技術では米国が先導してきたものとみられる。

(4) 注目動向

〔新展開・技術トピックス〕

① ゲームAIの進化

チェス、将棋、囲碁のようなゲームは、問題を定式化しやすいことから、AI研究の対象として早い時期から取り上げられてきた。最近ではより複雑な問題設定も扱えるようになり、さまざまな現実の問題解決につながりつつある。

ゲームAI分野では、モンテカルロ木探索による膨大な先読みで機械学習が組み合わせることが行われ、チェスは1997年に、将棋は2015年に、人間のレベルを上回ったとみなされた後、囲碁はさらに10年かかると言われていたところ、Google DeepMindのAlphaGoは2016年～2017年に世界トップランクプロに圧勝した。AlphaGoは、モンテカルロ木探索に組み合わせて、膨大なプロの棋譜を訓練データとした

教師あり学習と、膨大な回数の自己対戦による深層強化学習を用いて訓練された¹⁶⁾。その後、AlphaGoはAlphaGo Zero¹⁷⁾、AlphaZero¹⁸⁾、MuZero¹⁹⁾へと進化した。AlphaZeroは、訓練データを必要とせず、自己対戦だけで成長することができ、囲碁だけでなく、チェスや将棋でも世界チャンピオンプログラムに勝利した。さらにMuZeroは、ゲームのルールすら与えられていない状態から学習し、AlphaZeroに匹敵する強さに至った。

以上は完全情報ゲーム¹⁾として定式化されるタイプの問題を扱ったものだったが、現在では不完全情報ゲームも扱われている。DeepMindはリアルタイムストラテジーと呼ばれるジャンルのゲームStarcraft IIでも、マルチエージェント強化学習などを用いたAlphaStar²⁰⁾というソフトウェアを開発し、2019年にプロのゲーマーに勝利した。また、ポーカーにおいて、米国カーネギーメロン大学のLibratus²¹⁾というソフトウェアが2017年に2人制ゲームで人間に勝利し、さらに、それを発展させたPluribus²²⁾というソフトウェアが2019年に6人制ゲームでもプロに圧勝した(「2.1.3 エージェント技術」で関連技術を紹介している)。MetaとDeepMindは2022年にそれぞれ、Diplomacyという交渉ゲーム(複数プレイヤー間で外交交渉をしながら領土拡大を目指すゲーム)で人間並みのプレーをするAIを発表した^{23), 24)}。

このような技術は、ゲームにおいて効果を示した後、科学研究における探索問題、ビジネスや軍事などの戦略立案、マーケットなどにおける交渉問題などの実問題への展開が進んでいる。

タンパク質構造予測問題では、DeepMindのAlphaFold2が驚異的なスコアを出した(これについては[注目すべき国内外のプロジェクト] ❶で取り上げる)。DeepMindからは他にも、行列の積を計算するアルゴリズムという数学の問題について、従来知られていたものよりも高速なアルゴリズムを発見したAlphaTensor²⁵⁾が発表された。このAlphaTensorは、深層強化学習を用いた前述のAlphaZeroをベースとして「行列の積を計算する最適な方法を求める」というゲームを実行させたものである。同様にAlphaZeroをベースとしたAlphaDevは、ソートアルゴリズムを高速化する方法を発見した⁴³⁾。また、現状は数学定数に関する連分数式の形に限定されるが、膨大な組み合わせの探索によって、インドの天才的な数学者Srinivasa Ramanujanのように公式の候補を生成するRamanujan Machine²⁶⁾が、イスラエル工科大学の研究チームによって開発された。より幅広く、異なる数学的対象の間に潜むパターンや関係性を見つけて、人間の直感を導くことで、結び目理論における新たな定理発見につなげたという成果⁵⁴⁾がDeepMindから発表された。

なお、「2.1.1 知覚・運動系のAI技術」や「2.1.2 言語・知識系のAI技術」において、大規模学習による基盤モデル(Foundation Model)がさまざまなタスクに対応できるようになったことを述べたが、数学・物理学で扱うような定量的推論が苦手だと言われる。Googleの基盤モデル(大規模言語モデル)PaLMで、米国の高校レベルの数学問題を集めたMATH Datasetを解くと正答率は8.8%にとどまる。これに対して、PaLMに科学論文や数式が含まれる文書を大量に学習させたMinerva²⁷⁾では、正答率が50.3%まで向上したということである。しかし、数学問題の意味を理解して定量的推論をしたわけではなく、確率的な出力の当たる率が高まっただけである。

❷ AI駆動型科学

上記❶で述べた通り、囲碁の世界において、AI技術を用いたAlphaGoが世界トップクラスの棋士に圧勝した。その際、AlphaGoが行っていた膨大な可能性の探索から導出された打ち手は、人間の棋士には思いもよらなかった手を含んでいたが、それはその後、新手として人間の棋士も取り入れるようになった。同様のことは、今後、科学的発見においても起こり得る。

1 完全情報ゲームとは、すべての意思決定ポイントにおいて、これまでの行動や状態に関する情報がすべて得られるタイプの展開型ゲーム(ゲーム木の形式で表現できるタイプのゲーム)である。

このような科学的発見の可能性の拡大に向けてキーとなる技術チャレンジは、a. 人間の認知能力を超えた仮説生成・探索のための技術開発と、b. 仮説評価・検証のハイスループット化と考えられる³⁰⁾。aに関しては、超多次元の現象（非常に多くのパラメーターで記述される現象）から規則性を見いだすことは人間には困難だが、深層学習を用いれば¹⁴⁾、それが可能になりつつある。また、複数の異なる専門分野の知識をつなぎ合わせた推論による仮説の生成・探索は人間には困難だが、論理推論の枠組み²⁸⁾を分野横断で実行すれば、それが可能になるかもしれない。bに関しては、ロボットなどによる物理的な実験の自動化技術²⁹⁾も含め、科学的発見プロセスを構成するさまざまな技術を一つのプラットフォーム上に統合^{12), 15)}するとともに、計算量や物理的操作を抑える効率の良い処理フローや絞り込みアルゴリズムが必要になる。

このような方向のグランドチャレンジとして、「2050年までに生理学・医学分野でノーベル賞級の科学的発見をできるAIシステムを作る」ことを目標に掲げたNobel Turing Challengeが2016年に提唱された^{12), 13)}。このAIを活用した科学的発見のためのエンジンを作るというグランドチャレンジは国際的な目標になりつつあり、英国のアラン・チューリング研究所（The Alan Turing Institute）では、The Turing AI Scientist Grand Challengeプロジェクトを2021年1月にスタートさせた（グランドチャレンジ提唱者の北野もメンバーとして招聘されている）。日本国内でも、科学技術振興機構（JST）の未来社会創造事業で本格研究フェーズに移行した「ロボティックバイオロジーによる生命科学の加速」（研究開発代表者：高橋恒一）、ムーンショット型研究開発事業のムーンショット目標3に採択された「人とAIロボットの創造的共進化によるサイエンス開拓」（プロジェクトマネージャー：原田香奈子）などが推進されている。

材料科学（マテリアルズインフォマティクス）や生命科学・創薬の分野では、膨大な可能性から探索・絞り込みをするためにAIを活用して、科学的発見を実現することが試みられている。具体的には、計算コストが高いシミュレーションを機械学習で高速化し、目的とする新化合物を探索・絞り込みするバーチャルスクリーニングの取り組みが活発である。一方、物理学分野では、深層学習モデル、ニューラルネットワークを物理現象の理解のために用いるという取り組みが活発化し^{31), 32), 33)}、物性物理学や重力理論などの分野で成功事例が報告されている。国内では、学術変革領域研究（A）に「学習物理学の創成」（領域代表：橋本幸士、研究期間：2022年6月～2027年3月）が採択された。ここでは、機械学習と物理学を融合して基礎物理学を変革し、新法則の発見や新物質の開拓につなげることを目指している。また、ボルツマンマシンや拡散モデルなど、物理学のモデルが深層学習の発展につながる流れも見られる。

一方で、人間の認知能力を超えた超多次元の規則性の発見は、もし人間に理解できないとしたら、それを科学として許容してよいのかという議論も起きている²⁾。科学とは何かという基本的な問題や、科学コミュニティや社会による受容の問題も併せて考えていくことが必要である。

また、AI駆動型科学とシミュレーション科学が重なるAI・シミュレーション融合の開発・応用も取り組みが進んでいる。科学分野では、物理シミュレーションの結果を学習データに用いることで、従来の数値シミュレーションを機械学習で置き換え、処理の高速化・時間短縮や、調整の容易化が行われている。[注目すべき国内外のプロジェクト] ②に取り上げるMatlantisはその顕著な事例であるとともに、材料科学の進展に大きく寄与し得る取り組みである。また、複雑系や社会科学系では、ミクロ・マクロループへの対応、個人データ利用回避（プライバシー保護）、希少事例対応（データ合成）など、機械学習で生じる課題に対してシミュレーションを組み合わせることで対処するアプローチが有効である（「2.1.3 エージェント技術」に関連動向を記載）。

2 超多次元の現象に規則性を見いだすことは人間に困難であっても、発見されてしまえば、その規則性を人間は理解し得るのかもしれないという見方や、直感的理解が困難でも、その規則性に反するものが見つからなければ受容し得るのかもしれないという見方もある。いずれにせよ、研究コミュニティや社会による受容という面から継続的に考えていく必要がある。また、この受容性を高めるための補助的な枠組み（モデルの解釈性、数学的な枠組み、能動的検証の仕組みなど）も研究課題になり得る。

2.1

俯瞰区分と研究開発領域
人工知能・ビッグデータ

③ タンパク質言語モデル

深層学習による生物配列の解析技術の中でもタンパク質言語モデル (Protein Language Models) が急速に発達・普及している。言語モデルは、自然言語の文章を構成する各単語の特徴表現と文の生成確率関数を学習する統計モデルであるが、タンパク質言語モデルは各アミノ酸を「単語」とするタンパク質の一次配列という「文章」の特徴表現と文章の生成確率関数を学習する深層学習モデルである。これまでタンパク質言語モデルはword2vecや畳み込みニューラルネットワーク (CNN) などに基づくアーキテクチャーの提案がなされてきたが、近年のタンパク質言語モデルはTransformerベースが中心である。

タンパク質の表現学習はこれまでアミノ酸配列の構造を反映したベクトル (記述子) や多重配列アラインメント (Multiple Sequence Alignment: MSA) などが用いられてきたが、タンパク質言語モデルではその特徴表現をデータから獲得する。約2400万配列からなるタンパク質配列データベースUniRefから学習したタンパク質言語モデルUniRepは、その埋め込みベクトルにアミノ酸の物理化学的性質が反映され、分類や変異導入効果の予測などの下流タスクに利用可能なことが示されている⁴²⁾。この表現学習は、アミノ酸配列のマスクされた部分を周辺のアミノ酸残基の情報 (自然言語における「文脈」に相当) から推測させるタスクによる自己教師あり学習として行われ、アミノ酸配列中の残基の出現パターンの学習といえる。大規模なTransformerベースのタンパク質言語モデルとしてはESM (Evolutionary Scale Modeling)⁴³⁾、ProtTrans⁴⁴⁾、ProtT5⁴⁵⁾ などさまざまなものが提案され、獲得された特徴表現は種の分類や変異導入効果の予測などの下流タスクに応用可能である。

近年ではタンパク質言語モデルがMSAに似た情報を獲得していることを利用し、タンパク質の立体構造予測の高速化を狙う研究が盛んである。タンパク質立体構造予測ではAlphaFold2⁴⁶⁾ やRoseTTAFold⁴⁷⁾ など深層ニューラルネットを用いた高性能の予測モデルが既に標準的ツールとして利用されているが、いずれも入力としてMSAが必要であり、そこが処理時間上のボトルネックになっている。そこでMSAをタンパク質言語モデル (が保持するMSAに類する情報) で代替するという研究が生まれた。

タンパク質言語モデルを利用したタンパク質立体構造予測モデルの代表的な事例としてRGN2 (言語モデルはAmino BERT)⁴⁸⁾、HelixFold (DeBERTa)⁴⁹⁾、ESMFold (ESM2)⁵⁰⁾、EMBER3D (ProtT5)⁵¹⁾ が挙げられる。いずれのモデルもタンパク質言語モデルに基づく表現学習モジュールと構造予測モジュールに分かれた構造を持ち、タンパク質言語モデルから特徴ベクトルとアテンションマップが構造計算モジュールに受け渡される構成である。構造予測モジュールは多くの場合AlphaFold2で用いられたEvoformerというTransformerベースの深層学習モデルが元になっている。平均的な性能は単一入力配列 (MSAを使わない) の場合のAlphaFold2と同程度～高精度、実行速度では数桁高速になると報告されている⁵¹⁾。大規模タンパク質言語モデルESM2ではモデルサイズが15Bを超えると急激にPerplexityが改善し、言語モデルにとっての下流タスクである立体構造予測の精度も改善するなど基盤モデルの創発性と同様に、タンパク質言語モデルの大規模化にはメリットがあると示唆されている⁵⁰⁾。

[注目すべき国内外のプロジェクト]

① AlphaFold

AlphaFoldは、Google DeepMindが開発したタンパク質の立体構造予測を行うソフトウェアである^{34), 35)}。タンパク質構造予測の国際コンペティションCASP (Critical Assessment of protein Structure Prediction) に参加し、2018年のCASP13で最初のバージョンAlphaFold1で1位を獲得し、2020年のCASP14では改良されたAlphaFold2でさらに飛躍的なスコア向上を達成した。タンパク質構造予測は、タンパク質についてアミノ酸の配列が分かったとき、その立体構造を高精度に予測するという問題である。既に発見されたタンパク質は2億種類以上あるが、その働きを理解して、病気に対処したり新薬を開発したりするためには構造を知ることが重要である。X線結晶構造解析、低温電子顕微鏡、核磁気共鳴などの実験的手法では膨大な時間とコストがかかる。AlphaFoldが参加する前のCASP12では構造予

測スコアが40程度であったが、CASP14ではスコアが90前後まで向上した。AlphaFold2は、アテンション機構を持つトランスフォーマー型の深層学習（「2.1.2 言語・知識系のAI技術」を参照）を用い、既に明らかになっていた約17万種類のタンパク質の構造データを学習した。

その後、2021年7月にはソースコードが公開され、2022年7月には既知のタンパク質の配列約2億種類に対する構造予測が行われたことがDeepMindから発表された。AlphaFoldによる構造予測の限界として、変異や他タンパク質との相互作用による形状変化には対応していないことなどが指摘されているが、構造予測問題にほぼめどが付いたことによって、この分野の研究の進め方に大きな変革がもたらされた。

② Matlantis

Matlantisは、株式会社 Preferred Networks（PFN）とENEOS株式会社（ENEOS）が共同開発した、新物質開発・材料探索を高速化する汎用原子レベルシミュレーターである。両社の共同出資によって2021年6月に設立された新会社 Preferred Computational Chemistry（PFCC）からクラウドサービスとして契約ユーザー向けに提供されている。Matlantisでは、第一原理計算に基づく物理シミュレーションのために必要な密度汎関数理論（Density Functional Theory：DFT）計算に、従来は数時間から数週間を要していたのを秒・分レベルで実行できるまで高速化した³⁶⁾。物理シミュレーションと深層学習を統合しており、従来はデータセットが特定の原子構造・現象に絞って用意されていたが、より幅広く多様な元素・構造³を集めたデータセットを用意して汎用性を確保している。リチウムイオン電池のリチウム拡散、金属有機構造体の分子吸着挙動、金-銅合金の相転移現象などへの適用が紹介されている。秒・分レベルの高速化が実現されたことで、条件や観点を変えながらインタラクティブなシミュレーションが可能になり、研究開発の加速につながると期待される。

③ ロボティックバイオロジー

生命科学分野では、ロボットによる実験の自動化とAIによる条件探索や最適化を組み合わせた「ロボティックバイオロジー」の取り組みで、科学の自動化に向けた要素技術の開発が進められている。2022年6月には理化学研究所を中心とする研究チームが、汎用ヒト型ロボットLabDroid「まほろ」と新開発の最適化アルゴリズムを組み合わせた自動実験システムを構築し、iPS細胞（人工多能性幹細胞）から網膜色素上皮細胞（RPE細胞）への分化誘導効率を高める培養条件を、人間の介在なしに発見できることを実証した。この系にとどまらず、さまざまな生命科学実験において、これまで人間が試行錯誤をする必要のあった工程が、AIとロボットの導入により自律的に遂行可能になると期待される。

（5）科学技術的課題

① データ駆動型社会システムのための開発方法論と社会データ基盤の確立

AI・データ駆動型問題解決は、データ駆動型の社会システムの開発方法論でもある。社会システム科学として広く捉えられるほか、サイバーフィジカルシステムの開発方法論としてのReality 2.0³⁷⁾や、AI応用システムの開発方法論としてのAIソフトウェア工学³⁸⁾も関わりが深い。

また、国連で採択されたSDGs（Sustainable Development Goals：持続可能な開発目標）に掲げられているさまざまな社会課題に対しても、AI・データ駆動型問題解決は共通的に貢献する手段となる。ただし、貢献できる程度は社会課題の種類によって異なる。その差が生まれる要因として大きいのは、その社会課題の状況に関わる実世界ビッグデータを取得できるかという点と、状況把握・分析の結果を実世界

3 Matlantisは、2023年2月時点で72の元素をサポートしている。また、従来は既知の安定した原子構造データが主だったが、不安定な構造も含め、多様なパターンを集めている。

ヘフィードバックして状況改善に結び付ける制御手段が整っているかという点だと考える。そのため、そのような仕組みを強化した社会データ基盤の整備も重要課題である。

② AI・データ駆動型科学の方法論・技術群の研究開発

本節に示したようなデータ駆動型やAI駆動型と呼ばれる問題解決の方法論・技術群をいっそう強化していくことが求められる。[研究開発の動向] ④で科学研究の変革について、科学の四つのパラダイムを取り上げた。実験科学・理論科学・計算科学がなくなることはないが、データ・AI駆動科学が、科学研究・技術開発の国際競争力を左右するものになりつつある。そして、産業構造の変革や社会システムの変革にもそれが及ぶ。米国エネルギー省（Department of Energy：DOE）は2020年2月に「AI for Science」と題したレポート³⁹⁾を公開し、AI技術がさまざまな科学分野に波及し、その戦略的強化の必要性を示した。日本国内でも2020年4月に科学技術振興機構研究開発戦略センターから「デジタルトランスフォーメーションに伴う科学技術・イノベーションの変容」と題する同様のレポート⁴⁰⁾を公開している。第6期科学技術・イノベーション基本計画においても、データ駆動型研究の推進が掲げられている。マテリアル⁹⁾、バイオ⁴¹⁾、物理学³¹⁾などの科学分野を中心に、AI・機械学習技術の科学研究への活用が進んでいるが、[新展開・技術トピックス] ②で述べたように、人間の認知能力を超えた仮説生成・探索のための技術開発と、仮説評価・検証のハイスループット化をいっそう進めていくことが必要である。

③ 人材の再教育システムに関する研究開発

社会や産業の質的变化が起こってくる中で、「なくなる職業・仕事」に関する報告書⁴²⁾が話題になった。なくなる職業・仕事の一方で、新たな生まれる職業・仕事があることも指摘される。しかし、なくなる職業・仕事から新たな生まれる職業・仕事へ、必ずしも同じ人間が移行できるわけではない。社会や産業の質的变化に伴い、そこで働く人々に求められる能力・スキルも変化する。しかも、その変化が速いため、人間の能力・スキル獲得のスピードが追いつかないことが問題になる。社会制度（ベーシックインカムなど）や人材の再教育機会の整備を検討していく必要があるが、人材の再教育に関して、制度面の施策だけでなく、情報技術を活用した、よりの確で効率のよい再教育システムの研究開発も必要と考えられる。

(6) その他の課題

① 分野横断的な研究開発推進と人材育成・拠点構築

さまざまな技術分野・産業分野において、AI・データ駆動型問題解決のアプローチは必須のものと言える。国の政策としても、「AI戦略2019」において、「数理・データサイエンス・AI」の基礎は、デジタル社会の「読み・書き・そろばん」と位置付けられ、AI人材の育成、教育改革の推進が打ち出され、小中高・大学での教育、社会人リカレント教育などで取り組みが強化されている。文部科学省による大学変革の2023年度施策として、数理・データサイエンス・AI教育を全国レベルで推進するために、(1) 学部再編などによる特定成長分野（デジタル・グリーンなど）への転換等支援、(2) 高度情報専門人材の確保に向けた機能強化支援、が掲げられた。一方、大学においても、情報学部・データサイエンス学部を設置するところが急増している。

各分野でAI・データ駆動型アプローチを実践する上では、各分野のもともとの知識・技術と、AI・ビッグデータ技術の両方が分かる人材・組織の育成が重要である。AI・データ駆動型問題解決において、データ分析やAI・ビッグデータ技術は手段であり、問題解決・価値創造の側からの発想が重要である。分野横断的な技術者・研究者と、価値創造を牽引するリーダー人材と両面から人材育成を進めていくことが求められる。前者については、研究分野の異なる学会間の交流^{32), 33)}が盛んになりつつあり、同様の動きの広がりが期待される。

また、さまざまな技術を統合し、科学的発見のプロセスや問題解決のプロセスの全体を一気通貫で動か

2.1

俯瞰区分と研究開発領域 人工知能・ビッグデータ

するために拠点を構築することも有効と思われる。これは人材育成面にも寄与する。

② 制度設計・規制緩和

社会・産業などの質的变化に向けて、制度設計・規制緩和は必要になる。その際に、社会受容性に配慮した導入設計が重要になる。また、日本の社会適性という面だけでなく、グローバルな調和と競争環境も意識した設計が求められる。

③ 経済安全保障面の課題

AI・データ駆動型問題解決の枠組みは、幅広い産業・応用シーンが広がり、その重要性はますます高まるものと考えられるが、特にAI・データ駆動型科学の枠組み・基盤は、国の科学研究力・産業競争力を左右するものであり、経済安全保障面でも極めて重要なものになる。日本は、コンセプト提唱で先行したが、その実践では必ずしも先行優位を確保できておらず、その強化は重要課題である。

(7) 国際比較

国・地域	フェーズ	現状	トレンド	各国の状況、評価の際に参考にした根拠など
日本	基礎研究	○	↗	ノーベルチューリングチャレンジを提唱する中核的人材の存在や、AI戦略2019、第6期科学技術・イノベーション基本計画などで、AI人材育成やデータ駆動型研究推進が掲げられ、取り組みが強化されている。しかし、現時点では、人材育成やデジタル化の遅れが見られ、底上げが必要である。
	応用研究・開発	○	↗	実世界ビッグデータへの取り組みは強化されてきているが、社会・産業の変革に対する社会受容・制度設計などには課題がある。
米国	基礎研究	◎	↗	適切なグランドチャレンジの設定など、長期視点での変革につながる基礎研究への投資が国によって行われている（DARPA・DOEによる推進など）。
	応用研究・開発	◎	↗	Big Tech企業を中心とした産業界がAI・ビッグデータ技術の開発と社会・産業などの変革を牽引している。
欧州	基礎研究	○	↗	英国のアラン・チューリング研究所でThe Turing AI Scientist Grand Challengeプロジェクトがスタートした。
	応用研究・開発	○	→	インダストリー4.0など、ドイツでの取り組みが目される。
中国	基礎研究	○	→	国がAI・ビッグデータ研究開発に大型投資を行い、強力に推進している。深層学習を中心としたAI研究は米中2強となっており、中国の基礎研究は強化されている。
	応用研究・開発	◎	↗	国がAI・ビッグデータ技術を活用した監視・管理社会の構築を推進している。日本・欧米とは異なる文化・価値観だが、独自の社会変革を推進している。
韓国	基礎研究	△	→	特筆すべき点はない。
	応用研究・開発	○	→	特筆すべき点はないが、デジタル化は推進されている。

(註1) フェーズ

基礎研究：大学・国研などでの基礎研究の範囲

応用研究・開発：技術開発（プロトタイプの開発含む）の範囲

(註2) 現状 ※日本の現状を基準にした評価ではなく、CRDSの調査・見解による評価

◎：特に顕著な活動・成果が見えている

○：顕著な活動・成果が見えている

△：顕著な活動・成果が見えていない

×：特筆すべき活動・成果が見えていない

(註3) ティレンド ※ここ1～2年の研究開発水準の変化

↗：上昇傾向、→：現状維持、↘：下降傾向

参考文献

- 1) 喜連川優,「ビッグデータ」,『學士會会報』No. 918 (2016年), pp. 78-83.
- 2) 城田真琴,『ビッグデータの衝撃—巨大なデータが戦略を決める』(東洋経済新報社, 2012年) .
- 3) Yannick de Jong, “Levels of Data Analytics”, IThappens.nu (20 March 2019). <http://www.ithappens.nu/levels-of-data-analytics/> (accessed 2021-02-03)
- 4) 大向一輝,「オープンデータとLinked Open Data」,『情報処理』(情報処理学会誌) 54巻12号 (2013年12月), pp. 1204-1210.
- 5) 加藤文彦・他,「IMI共通語彙基盤」,『デジタルプラクティス』(情報処理学会デジタルプラクティス論文誌) 9巻1号 (2018年1月) .
- 6) 科学技術振興機構 研究開発戦略センター,「調査報告書: 計測横断チーム調査報告書 計測の俯瞰と新潮流」, CRDS-FY2018-RR-03 (2018年12月) .
- 7) 情報処理推進機構 AI白書編集委員会 (編),『AI白書2022』(KADOKAWA, 2022年) .
- 8) Tony Hey, Stewart Tansley and Kristin Tolle, *The Fourth Paradigm: Data-Intensive Scientific Discovery* (Microsoft Research, October 2009).
- 9) 岩崎悠真,『マテリアルズ・インフォマティクス: 材料開発のための機械学習超入門』(日刊工業新聞社, 2019年) .
- 10) 鷲尾隆,「計測指向情報処理技術と情報処理指向計測技術の共進化」,『Readout - HORIBA Technical Report』53号 (2019年10月), pp. 62-67.
- 11) Regina Nuzzo, “How scientists fool themselves - and how they can stop”, *Nature* Vol. 526, Issue 7572 (8 October 2015), pp. 182-186. DOI: 10.1038/526182a
- 12) 北野宏明,「人工知能がノーベル賞を獲る日,そして人類の未来—究極のグランドチャレンジがもたらすもの」,『人工知能』(人工知能学会誌) 31巻2号 (2016年3月), pp. 275-286.
- 13) Hiroaki Kitano, “Nobel Turing Challenge: creating the engine for scientific discovery”, *npj Systems Biology and Applications* Vol. 7, Article No. 29 (18 June 2021). DOI: 10.1038/s41540-021-00189-3
- 14) 丸山宏,「高次元科学への誘い」, CNET Japan ブログ (2019年5月1日) . https://japan.cnet.com/blog/maruyama/2019/05/01/entry_30022958/ (accessed 2021-02-03)
- 15) 高橋恒一,「AI 駆動科学とその社会と人間性への影響」,『ここまで 来ました: 右巻き左巻き・AI 駆動科学・がん医療の革新』(武田計測先端知財団編, 丸善プラネット, 2020年), pp. 41-77 (第2章) .
- 16) David Silver, et al., “Mastering the game of Go with deep neural networks and tree search”, *Nature* Vol. 529, No. 7587 (2016), pp. 484-489. DOI: 10.1038/nature16961
- 17) David Silver, et al., “Mastering the game of Go without human knowledge”, *Nature* Vol. 550, No. 7676 (2017) pp. 354-359. DOI 10.1038/nature24270
- 18) David Silver, et al., “Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm”, arXiv: 1712.01815 (2017).
- 19) Julian Schrittwieser, et al., “Mastering Atari, Go, chess and shogi by planning with a learned model”, *Nature* Vol. 588 (2020), pp. 604-609. DOI: 10.1038/s41586-020-03051-4
- 20) The AlphaStar team, “AlphaStar: Grandmaster level in StarCraft II using multi-agent reinforcement learning”, DeepMind blog (30 October 2019), <https://deepmind.com/blog/article/AlphaStar-Grandmaster-level-in-StarCraft-II-using-multi-agent-reinforcement-learning> (accessed 2023-02-01)
- 21) Noam Brown and Tuomas Sandholm, “Superhuman AI for heads-up no-limit poker: Libratus beats top professionals”, *Science* Vol. 359, Issue 6374 (26 Jan 2018), pp. 418-424. DOI:

2.1

俯瞰区分と研究開発領域
人工知能・ビッグデータ

10.1126/science.aao1733

- 22) Noam Brown and Tuomas Sandholm, “Superhuman AI for multiplayer poker”, *Science* Vol. 365, Issue 6456 (30 Aug 2019), pp. 885-890. DOI: 10.1126/science.aay2400
- 23) Anton Bakhtin, et al., “Human-level play in the game of Diplomacy by combining language models with strategic reasoning”, *Science* Vol. 378, Issue 6624 (22 Nov 2022), pp. 1067-1074. DOI: 10.1126/science.ade9097
- 24) János Kramár, et al., “Negotiation and honesty in artificial intelligence methods for the board game of Diplomacy”, *Nature Communications* Vol. 13, Article No. 7214 (6 December 2022). DOI: 10.1038/s41467-022-34473-5
- 25) Fawzi, Alhussein, et al., “Discovering faster matrix multiplication algorithms with reinforcement learning”, *Nature* Vol. 610, Article No. 7930 (5 October 2022), pp. 47-53. DOI: 10.1038/s41586-022-05172-4
- 26) Gal Raayoni, et al., “Generating conjectures on fundamental constants with the Ramanujan Machine”, *Nature* Vol. 590 (3 February 2021), pp. 67-73. DOI: 10.1038/s41586-021-03229-4
- 27) Aitor Lewkowycz, et al., “Solving Quantitative Reasoning Problems with Language Models”, arXiv: 2206.14858 (2022).
- 28) 井上克巳, 「人工知能による科学的発見」, 『電子情報通信学会誌』 98 巻 1 号 (2015 年), pp. 35-39.
- 29) 夏目徹・高橋恒一・神田元紀 (企画), 「特集: 新型コロナで変わる時代の実験自動化・遠隔化」, 『実験医学』 39 巻 1 号 (2021 年 1 月), pp. 1-52.
- 30) 科学技術振興機構 研究開発戦略センター, 「戦略プロポーザル: 人工知能と科学 ~ AI・データ駆動科学による発見と理解~」, CRDS-FY2021-SP-03 (2021 年 8 月) .
- 31) 田中章詞・富谷昭夫・橋本幸士, 『ディープラーニングと物理学: 原理がわかる、応用ができる』(講談社, 2019 年) .
- 32) 小林亮太・岡本洋・山川宏 (編), 「特集: 物理学と AI」, 『人工知能』(人工知能学会誌) 33 巻 4 号 (2018 年 7 月), pp. 391-448.
- 33) 「シリーズ: 人工知能と物理学」, 『日本物理学会誌』 74 巻 1 号~11 号 (2019 年) . <https://www.jps.or.jp/books/gakkaishi/seriesai.php> (accessed 2023-02-01)
- 34) John Jumper, et al., “Highly accurate protein structure prediction with AlphaFold”, *Nature* Vol. 596 (15 July 2021), pp. 583-589. DOI: 10.1038/s41586-021-03819-2
- 35) 森脇由隆, 「AlphaFold2 までのタンパク質立体構造予測の軌跡とこれから」, 『JSBi Bioinformatics Review』 3 巻 2 号 (2022 年 11 月), pp. 47-60. DOI: 10.11234/jsbibr.2022.3
- 36) So Takamoto, et al., “Towards universal neural network potential for material discovery applicable to arbitrary combination of 45 elements”, *Nature Communications* Vol. 13, Article No. 2991 (30 May 2022). DOI: 10.1038/s41467-022-30687-9
- 37) 科学技術振興機構 研究開発戦略センター, 「戦略プロポーザル: IoT が開く超スマート社会のデザイン - REALITY 2.0 -」, CRDS-FY2015-SP-02 (2016 年 3 月) .
- 38) 科学技術振興機構 研究開発戦略センター, 「戦略プロポーザル: AI 応用システムの安全性・信頼性を確保する新世代ソフトウェア工学の確立」, CRDS-FY2018-SP-03 (2018 年 12 月) .
- 39) Rick Stevens, et al., “AI for Science”, Report on the Department of Energy (DOE) Town Halls on Artificial Intelligence (AI) for Science (February 2020).
- 40) 科学技術振興機構 研究開発戦略センター, 「デジタルトランスフォーメーションに伴う科学技術・イノベーションの変容」(- The Beyond Disciplines Collection -), CRDS-FY2020-RR-01 (2020 年 4 月) .
- 41) 科学技術振興機構 研究開発戦略センター, 「AI × バイオ DX 時代のライフサイエンス・バイオメディカ

2.1

俯瞰区分と研究開発領域
人工知能・ビッグデータ

ル研究」(– The Beyond Disciplines Collection –) , CRDS-FY2020-RR-03 (2020年9月) .

- 42) Carl Benedikt Frey and Michael A. Osborne, “The Future of Employment: How Susceptible are Jobs to Computerisation?”, (September 17, 2013).
https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf
 (accessed 2023-02-01)
- 43) Daniel J. Mankowitz, et al., “Faster sorting algorithms discovered using deep reinforcement learning”, *Nature* Vol. 618 (7 June 2023), pp. 257-263. DOI: 10.1038/s41586-023-06004-9
- 44) Ethan C. Alley, et al., “Unified rational protein engineering with sequence-based deep representation learning”, *Nature Methods*, 16 (12), 1315-1322 (2019). DOI: 10.1038/s41592-019-0598-1
- 45) Alexander Rives, et al., “Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences”, *Proceedings of the National Academy of Sciences*, 118 (15), e2016239118 (2021). DOI : 10.1073/pnas.2016239118
- 46) Ahmed Elnaggar, et al., “Prottrans: Toward understanding the language of life through self-supervised learning”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44 (10), 7112-7127 (2022). DOI : 10.1109/TPAMI.2021.3095381
- 47) Colin Raffel, et al., “Exploring the limits of transfer learning with a unified text-to-text transformer”, *The Journal of Machine Learning Research*, 21(1), 140:5485-140:5551 (2020). DOI: 10.5555/3455716.3455856
- 48) John Jumper, et al., “Highly accurate protein structure prediction with AlphaFold”, *Nature* 596, 583-589 (2021). DOI: 10.1038/s41586-021-03819-2
- 49) Minkyung Baek, et al., “Accurate prediction of protein structures and interactions using a three-track neural network”, *Science*, 373 (6557), 871-876 (2021). DOI: 10.1126/science.abj8754
- 50) Ratul Chowdhury, et al., “Single-sequence protein structure prediction using a language model and deep learning”, *Nature Biotechnology* 40, 1617-1623 (2022). DOI: 10.1038/s41587-022-01432-w
- 51) Xiaomin Fang, et al., “HelixFold-Single: MSA-free Protein Structure Prediction by Using Protein Language Model as an Alternative”, arXiv: 2207.13921 (2023). DOI: 10.48550/arXiv.2207.13921
- 52) Zeming Lin, et al., “Evolutionary-scale prediction of atomic-level protein structure with a language model”, *Science* 379, 6637, 1123-1130 (2023). DOI: 10.1126/science.ade2574
- 53) Konstantin Weissenow, et al., “Ultra-fast protein structure prediction to capture effects of sequence variation in mutation movies”, bioRxiv (2022). DOI: 10.1101/2022.11.14.516473
- 54) Alex Davies, et al., “Advancing mathematics by guiding human intuition with AI”, *Nature* Vol. 600 (December 1, 2021), pp. 70-74. DOI : 10.1038/s41586-021-04086-x

2.1

俯瞰区分と研究開発領域
人工知能・ビッグデータ