2.1.2 言語・知識系のAI技術

（1）研究開発領域の定義

知能を知覚・運動系と言語・知識系という2面で捉え、ここでは後者を俯瞰する（前者については前節2.1.1で俯瞰した）。研究開発領域としては、自然言語の解析・変換・生成などを行う自然言語処理（Natural Language Processing）、知識の抽出・構造化・活用を行う知識処理（Knowledge Processing）などが中心的に取り組まれてきた。

知覚系は実世界からの入力、運動系は実世界への出力として、知能の実世界接点の役割を担う。状況を知り、判断し、行動するという一連のプロセスは、知能において、知覚系と言語・知識系と運動系の連携によって熟考的に実行されることもあれば（ここでは熟考的ループと呼ぶ）、知覚系と運動系の間で即応的に実行されることもある（ここでは即応的ループと呼ぶ）[1]。近年、機械学習（Machine Learning）、特に深層学習（Deep Learning）が発展し、まずは知覚系（パターン認識）での活用が進み、次第に言語・知識系（自然言語処理、知識処理）や運動系（動作生成）にも広く活用されるようになった。知覚・運動系と言語・知識系の処理方式の共通性が高まり、それらを統一的に扱う枠組みも研究されるようになってきた。そこで、本節では、自然言語処理・知識処理そのものの研究開発の動向に加えて、知覚・運動系と言語・知識系を統合して熟考的ループを構成するための研究開発の動向についても取り上げる。

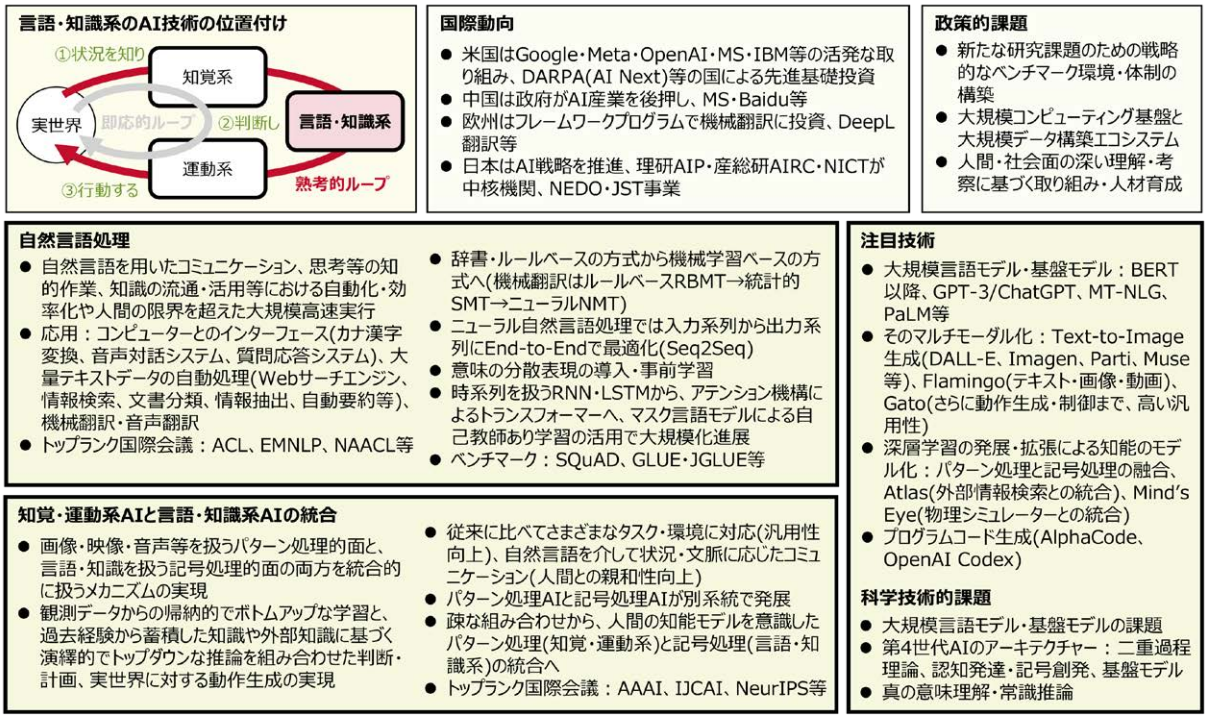

図2-1-4　　領域俯瞰：言語・知識系のAI技術

1　人間の思考は、直感的・無意識的・非言語的・習慣的な「速い思考」のシステム1と、論理的・系列的・意識的・言語的・推論計画的な「遅い思考」のシステム2とで構成されるという「二重過程理論」（Dual Process Theory）がある。社会心理学・認知心理学などの心理学分野で提案されていたが、ノーベル経済学賞を受賞したDaniel Kahnemanの著書「Thinking,Fast and Slow」[1)]でよく知られるようになった。本稿ではシステム1を「即応的ループ」、システム2を「熟考的ループ」と呼んでいる。

（2）キーワード

自然言語処理、知識処理、テキスト処理、機械学習、深層学習、分散表現、アテンション、トランスフォーマー、基盤モデル、二重過程理論、意味理解、文書読解、質問応答、機械翻訳、文章生成、知識獲得、プログラムコード生成、マルチモーダル、生成型AI、大規模言語モデル

（3）研究開発領域の概要

［本領域の意義］

自然言語は人間が日常の意思疎通のために用いる自然発生的な記号体系である[2]。人間にとって自然言語は、概念を表現する記号体系として、日常の意思疎通（コミュニケーション）だけでなく、思考の過程やその結果である知識の表現・保存にも用いられる。コンピューターによる自然言語処理と知識処理は、このような人間のコミュニケーション、思考などの知的作業、知識の流通・活用などを含むさまざまな場面に適用され得る。そして、その自動化・効率化や、人間の限界を超えた大規模高速実行を可能にする。その代表的な場面・システムのいくつかを以下に挙げる。

まず、コンピューターとのインターフェースに使われる自然言語処理として、カナ漢字変換入力システム、音声対話システム、質問応答システムなどが挙げられる。最近はスマートスピーカー（Amazon Echo、Google Homeなど）が家庭で使われ始めているが、自然言語で操作・指示できると、特別のコマンド入力・操作方法をあれこれ覚える必要がない。コンタクトセンターでの問い合わせ受付では、簡単な質問への対応の自動化によって、問い合わせ対応のスループット向上や質の安定が得られる。

また、大量のテキストデータの処理をコンピューターで行うことで、人間の負荷を軽減しようという自然言語処理システムがある。 Webサーチエンジンが代表例だが、膨大な情報の中から条件に合う情報を高速に見つけたり、整理したりするための情報検索・文書分類、その概要把握を助ける情報抽出・自動要約などに、自然言語処理が活用され、例えば、科学技術研究の加速につながっている。また、大量のWebテキストから、概念をノードで、概念の間の関係をリンクで表現した大規模なナレッジグラフ（知識グラフ、Knowledge Graph）を構築し、検索語の拡張、検索結果の品質向上、対話システムの話題拡大などに利用することも行われている。

さらに、機械翻訳・音声翻訳（自動通訳）も自然言語処理の代表的な応用である。母国語から他国語、あるいは、逆に他国語から母国語への翻訳・通訳は、他国語を話す人々とのコミュニケーションを支援するとともに、インターネットなどを介して世界に流通している膨大な量の他国語で書かれた情報を調査・分析する労力を大幅に軽減してくれる。

以上、自然言語処理・知識処理の意義や応用について述べたが、次に、本節で取り上げるもう一つのトピックである知覚・運動系と言語・知識系の統合（熟考的ループ）について、その意義や応用について述べる。この統合によって、人工知能（AI：Artificial Intelligence）やロボットを実現する上で、より総合的な知能の性質がカバーされる。すなわち、画像・映像・音声などを扱うパターン処理的な側面と言語・知識を扱う記号処理的な側面の両方を統合的に扱うメカニズムが実現される。また、知覚系を通して直接的に得られる外界の観測データからの帰納的でボトムアップな学習と、過去の経験を通して蓄積された知識や社会・他者と共有された外部知識に基づく演繹的でトップダウンな推論の、両方を組み合わせた判断や計画のメカニズムと、その結果を実世界に対する一連の動作として生成・実行するメカニズムが実現される。このようなメカニズムを備えたAI・ロボットは、従来に比べて、さまざまなタスクや環境により少ない学習で対応可能にな

2 自然言語に対して、プログラミング言語やマークアップ言語など、人工的に定義された言語がある。これらの人工的に定義された言語は解釈が一意に定まるように設計されているが、自然言語は文・句・単語などの意味や構造の解釈に曖昧性が生じ得る点、記号接地（記号と実世界における意味をどのようにして結び付けるか）や意図理解のように記号だけに閉じない問題が関わる点などが、その処理を難しいものにしている。

り（汎用性の向上）、自然言語を介して、実世界の状況・文脈に応じたコミュニケーションが可能になる。このことは、人間との親和性を向上させ、人間とAI・ロボットが協働する中で共に知識を創成し、共に成長する社会の実現につながると期待される。

［研究開発の動向］

❶ 自然言語の解析技術の発展（～2017年頃）

自然言語処理技術において共通的に必要とされる基礎技術はコンピューターによる自然言語解析であり、形態素解析（単語分割や品詞認定）、構文・係り受け解析、文脈・意味解析（語義の曖昧性解消や照応解析を含む）というステップで、より深い解析への取り組みが進められた。そのアプローチは、黎明期の1950年代から1990年代頃まで、人間が記述した辞書・文法を用いるルールベース方式が主流だった。しかし、大量のテキストデータが利用可能になったことや、機械学習技術が大幅に進化したことから、徐々に統計的な方式、機械学習を用いた方式に主流が移った。適用される機械学習技術は、ナイーブベイズに始まり、2010年頃にはSVM（Support Vector Machine）が主流となったが、2014年頃からはニューラルネットワークによる機械学習、特に深層学習が盛んに適用されるようになった[2), 3), 4)]。

このような技術発展は特に機械翻訳への取り組みによって牽引されてきた。機械翻訳方式は当初のルールベース機械翻訳（Rule-Based Machine Translation：RBMT）から、1990年代に大規模な対訳コーパス（元言語のテキストとターゲット言語のテキストを対にしたもの）と機械学習技術を用いた統計的機械翻訳（Statistical Machine Translation：SMT）に主流が移った。 SMTの精度改善に頭打ちが見えてきた2010年代に、ニューラル機械翻訳（Neural Machine Translation：NMT）[3), 5)] が考案され、顕著な精度改善がもたらされた。 SMTからNMTへの移行は、SMTで用いていた統計処理・機械学習のパートを単純に深層学習に置き換えたものではなく、機械翻訳のパラダイムを大きく転換させたものである。SMTでは、機械翻訳のプロセスを多段階に分け、各段階の処理モデルを統計的にチューニングして組み合わせていたのに対して、NMTでは、入力原文から翻訳結果の出力までを一つのニューラルネットワーク構造（Seq2Seqモデル）として扱い、End-to-Endの最適化を行う。

その際、自然言語の単語系列をニューラルネットワークで扱うため、意味の分散表現[3]が用いられる。これは単語・句・文・段落などの意味を固定長ベクトル（実際には数百次元程度）で表現したものである[6), 7)]。大量テキストにおける文脈類似性に基づき、ニューラルネットワークを用いて分散表現を高速に計算するWord2Vecが2013年に公開され、自然言語処理の基本的な手法として広く使われるようになった。それまで使われていたBag-of-Words形式（N次元のうちの1要素だけの値が「1」というOne-Hotベクトル）と異なり、分散表現はベクトル計算によって単語や文の意味の合成・分解や類似度計算が可能である。例えば、分散表現を用いると、「king」－「man」＋「woman」＝「queen」のような意味のベクトル計算が近似的に可能になる。従来の記号処理は厳密な論理演算をベースとした固いものだったが、分散表現を用いることで曖昧な条件を許した柔らかい演算が可能になった。

この分散表現とSeq2Seqモデルを用いてEnd-to-Endで最適化するアプローチは、

入力系列（End）→［エンコーダー］→分散表現→［デコーダー］→出力系列（End）

という流れになる。このような系列変換は、機械翻訳だけでなく、質問応答、対話、情報要約、画像・

3 分散表現（Distributed Representation）は、ニューラルネットワーク研究の分野では局所表現（Local Representation）に対する概念として考えられた。一方、自然言語処理研究においても、単語の意味を扱う方法論として分布仮説（Distributional Hypothesis）があり、これら両面が融合したものと考えられている[6)]。自然言語に限らず、なんらかの離散的な対象物の表現方法として、局所表現や分散表現を用いることができる。局所表現はone-hotベクトルのように一つないしは少数の要素で特徴を表現するのに対して、分散表現は多数の要素に特徴を分散させて表現する。また、埋め込み（Embeddings）という言い方も用いられる。たとえばDistributed Representation of WordsとWord Embeddingsは同義である。

映像に対する説明文生成など、自然言語処理のさまざまな応用に使われるようになった。

「2.1.1 知覚・運動系のAI技術」で述べたように、深層学習はまず画像認識・音声認識の分野に適用され、衝撃的な性能向上がもたらされた。自然言語処理の分野では、そのような性能向上はすぐにはもたらされなかったが、少し遅れて新たな技術発展が生み出され、自然言語処理においても著しい進展がもたらされた。その内容は❸で後述する。

❷ 大規模テキスト活用・知識活用の発展

テキスト検索は、コンピューターの処理性能が乏しかった時代、事前に人手で各テキストに付与したキーワードを索引に用いるしかなかったが、1990年代以降、コンピューターの性能向上、並列処理技術の発展、ストレージの大容量化などが進み、フルテキストサーチ（全文検索）方式に主流が移った。急激に大規模化したWebサーチエンジンが、その代表であるが、クエリのキーワードとWebのフルテキストの単純なマッチングでは高い検索精度が得られないことから、Webページ間の被リンク関係やアンカー文字列（リンク元テキスト）を考慮した検索結果のランキング法（ページランク）や、ユーザーの嗜好や目的に応じた適合ページの選別法など、さまざまな観点からWeb検索の精度を高める技術が開発された。また、大規模なWebを解析・検索するため、大規模自然言語テキストを解析・検索するための分散・並列処理、文字列の圧縮・索引処理などの技術が急速に発展した[8)]。 Webサーチエンジンは幅広い一般ユーザー向けのアプリケーションとして発展したが、インターネット上の多様な情報や企業内の大量文書から評判・意見、注目事象、傾向変化などを抽出し、企業経営、マーケティング、リンク管理などに活用するテキストマイニング・Webマイニングと呼ばれる技術・アプリケーションも開発が進んだ[8), 9)]。また、GoogleはWeb上の情報から大規模なナレッジグラフを構築し、ナレッジパネルとして検索結果とともに表示することを行っている。

さらにその発展として、大量のテキスト情報を知識源として用いる質問応答システムがある。代表的なシステムとしてIBMのWatson[10)] が挙げられる。 Watsonは、大量テキスト情報を知識源として自然言語で書かれた質問に回答する技術を中核とし、2011年に米国の人気クイズ番組「Jeopardy!」で人間のクイズ王に勝利するというグランドチャレンジに成功した。国内では、情報通信研究機構（NICT）が、大規模なWeb情報をもとに、自然言語による「なに？（いつ/どこ/だれ）」「なぜ？」「どうなる？」「それなに？」という4タイプの質問に回答するシステムWISDOM Xを開発・公開した。このような応用においても、近年はニューラルネットワークをベースとした方式に移行し、2022年11月末に公開されたChatGPTは大きな話題になっている（ChatGPTについては［新展開・技術トピックス］❶でも取り上げる）。

❸ ニューラルネット自然言語処理の最新動向（2017年～）

深層学習による画像認識には畳み込みニューラルネットワーク（Convolutional Neural Network：CNN）が主に用いられたが、自然言語処理には、当初、時系列情報を扱うのに適した回帰型ニューラルネットワーク（Recurrent Neural Network：RNN）やLSTM（Long Short-Term Memory）ネットワークが用いられた[6), 51)]。これを用いた系列変換（Seq2Seq）の際に、ニューラルネットワーク中のどこ部分（特定の単語など）に注目するかを動的に決定するアテンション（Attention）機構が考案され、出力系列生成の品質向上につながった。アテンションのアイデアは最初、機械翻訳に導入されたが、その後、自然言語処理全般で（さらには画像処理にも）用いられるようになった[5), 11)]。

このアテンション機構を最大限に生かした新しい深層学習モデルとして、2017年にトランスフォーマー（Transformer）がGoogleから発表された[12)]。トランスフォーマーは、RNNやCNNを使わずに、アテ

ンション機構[4]のみで構成した深層学習モデルである。RNNやCNNより計算量が抑えられ、訓練が容易で、並列処理もしやすく、複数の言語現象を効率よく扱えて、文章中の長距離の依存関係も考慮しやすいといった特長を持ち、機械翻訳や自然言語の意味理解タスクなどのベンチマークでも従来を上回る性能が示された。このことから、2018年以降、新たに提案されるモデルはトランスフォーマー一色となり、次に説明する自己教師あり学習による事前学習の手法と合わせて、自然言語処理のさまざまなベンチマークで最高スコアが更新されている。

ニューラルネット自然言語処理[51]で高い精度を達成するには、大量の訓練データが必要だが、さまざまなタスクのおのおのについて大量の訓練データを用意することは容易なことではない。そこで、まず、さまざまなタスクに共通的な汎用性の高いモデルを、大量のラベルなしデータで事前学習（Pre-Training）しておき、それをベースに個別のタスクごとに少量のラベル付きデータでの追加学習（Fine Tuning）を行うというアプローチが取られるようになった。この事前学習で作られたトランスフォーマー型の深層学習モデルは、2018年にGoogleから発表されたBERT[13]以降、自然言語処理においてスタンダードになった。ラベルなしの大量テキストデータで事前学習を行うため、BERTではMLM（Masked Language Model）が導入された[5]。MLMはもとのテキストに対して複数箇所をマスクし（隠し）、穴埋め問題のようにマスク箇所を当てるというタスクを、大量テキストデータで訓練するというものである。もとのテキストから穴埋め問題の答えは分かるので、このタスクは実質的に教師あり学習として訓練できる。これが自然言語処理における「自己教師あり学習」（Self-Supervised Learning）の代表的な成功例として定着した。

その後、BERTを改良・拡張したモデルが次々に考案され、GLUE（General Language Understanding Evaluation）、SQuAD（Stanford Question Answering Dataset）などの自然言語処理ベンチマークの最高スコアが次々に更新された。マスクする単語を動的に変更したRoBERTa[14]、軽量化したALBERT[15]、ナレッジグラフを組み込んだERNIE[16]、片方向や双方向の言語モデルを統合したUniLM[17]、マルチモーダルに拡張したViBERT[18]、VL-BERT[19]、UNITER[20]などがある。意味の分散表現のベクトル形式は、テキストデータだけでなく、画像・映像・音声などの異なるデータタイプの入力に対しても用いることが可能で、マルチモーダル処理を共通的なニューラルネットワーク構造で行うことが容易になった。

MLMによって大量テキストからの言語モデル学習が一気に進んだ。言語モデルの規模を表すパラメーター数は、BERTの場合、3.4億個であったが、2020年にOpenAIから発表されたGPT-3[21]では、事前学習に45TBのデータを用い、モデルのパラメーター数は1750億個となった[6]。さらに、2021年10月にMicrosoftとNVIDIAが発表したMT-NLG（Megatron-Turing Natural Language Generation）[22]のパラメーター数は5300億個、2022年4月にGoogleが発表したPaLM（Pathways Language Model）[23]のパラメーター数は5400億個に及んだ。これらは「大規模言語モデル」（Large Language Model：LLM）と呼ばれるが、高い汎用性を示すことから「基盤モデル」（Foundation Model）[24]とも呼ばれるようになった。詳しくは［新展開・技術トピックス］❶で述べる。

また、GPT-3においては、それまでのGPTと同様に後続の系列を予測する自己回帰型の自己教師あり学習が用いられ、タスクごとのファインチューニング学習をせずとも、最初に入力する系列にタスクの記述や

4 アテンション機構には大きく分けると、Self-AttentionとSource-Target-Attentionという2種類がある。アテンションを求める際に、Self-Attentionは対象文中の情報からウェートを計算し、Source-Target-Attentionは別文中の情報からウェートを計算する。トランスフォーマーでは、Self-Attention機構をマルチヘッドで動かすことで、複数の言語現象を並列に効率良く学習できるようにしている。

5 より詳細には、MLMとともに、NSP（Next Sentence Prediction）がBERTに導入された。 NSPは二つの文が連続する文かどうかを判定するタスクを学習するものである。

6 BERTやGPT-3などではパラメーター数の異なる複数のモデルがあるが、本稿中でのパラメーター数の比較は、それぞれの最も規模の大きいモデルをもとに記載している。

事例を含めること（プロンプト）で複数のタスクに対応することをIn-Context学習（Zero-Shot学習/One-Shot学習/Few-Shot学習）と呼び、言語モデルの汎用的な活用を開拓した。 In-Context学習では、大規模な基盤モデル自体のパラメーターを変更することなく、プロンプトの記述によって個別のタスクに適応させることができる。受け付けられるプロンプト長が拡大される傾向にある一方で、RAG（Retrieval-Augmented Generation、検索拡張生成）[55] と呼ばれる技術も活用されるようになった。 RAGは、プロンプトをもとに、関連する外部知識ベース（タスクに応じたドメイン知識・データベースなど）を検索し、その結果も考慮して応答を生成する仕組みである。これによって、応答の正確性を高めることが可能になるとともに、毎回プロンプトに大量の説明を盛り込まずとも、タスクやドメインに関する共通知識をまとめて与えることが可能になった。このように基盤モデル自体の確率モデルとしての精度・性能向上の補完するような機能を外付けで拡張する仕組みも発展している。

❹ 知覚・運動系AIと言語・知識系AIの統合に関わる動向

パターン処理を中心とした知覚・運動系のAI技術と、記号処理を中心とした言語・知識系のAI技術は、別系統で発展してきた。1950年代後半から1960年代にかけての第1次AIブームと、1980年代の第2次AIブームで扱われたのは、記号処理のAI研究であった。「2.1.1 知覚・運動系のAI技術」で述べたように、第1次AIブームと同時期に、ニューラルネットワークを用いたパターン認識の研究も活発に取り組まれていた。そして、深層学習を中心としたニューラルネットワーク型のAI技術の発展が、2000年代以降の第3次AIブームの主役となった。

このように別系統で発展してきたが、AIとしてパターン処理と記号処理の両面を扱う必要があることは古くからいわれていた。また、これまでも二つのタイプのAI技術を組み合わせたシステムは多く見られる。例えば、音声翻訳システムは、音声認識というパターン処理と、機械翻訳という記号処理をつなげたシステムである。また、統計的機械翻訳（SMT）は、記号処理をベースに構成されたシステム中のいくつかのパーツを、機械学習を用いてチューニングしたものである。物理モデルや事前知識モデルを用いたシミュレーションシステムに機械学習を組み合わせたり、機械学習の分類・判定結果を解釈するためにナレッジグラフを組み合わせたりといった取り組みも、二つのタイプのAI技術の組み合わせと見ることもできる。

これらに対して、深層学習の発展によって2018年頃から顕在化してきた取り組みは、人間の知能のモデルを意識したパターン処理と記号処理の統合に関する研究である。すなわち、人間の知覚・運動系と言語・知識系の関係や、それらが構成する即応的ループと熟考的ループの情報の流れと対応するような、あるいは、そこからインスパイアされたような、パターン処理と記号処理の統合モデルが検討されている[25]。

このような動きが顕在化したことを示したのが、NeurIPS 2019（The 33rd Conference on Neural Information Processing Systems）でのYoshua Bengioによる「From System 1 Deep Learning to System 2 Deep Learning」と題した招待講演[26] である。ここで、System 1とSystem 2は、2002年にノーベル経済学賞を受賞したDaniel Kahnemanの言う直観的な「速い思考」（システム1）と論理的な「遅い思考」（システム2）[1] のことで、本稿でいう即応的ループと熟考的ループに対応する。 Bengioは、現在の深層学習はシステム1に相当するが、システム2までカバーするような深層学習へ発展させるのが今後の方向性だと示唆した。さらに、Yoshua Bengio、Geoffrey Hinton、Yann LeCunの3人は、深層学習発展への貢献で2018年度ACM（Association for Computing Machinery）チューリング賞を受賞したが、AAAI 2020（The 34th AAAI Conference on Artificial Intelligence）における同賞記念イベントでは3人の講演に加えてKahnemanを交えたパネル討論も実施され、この方向性が論じられた。一方、日本国内では、これよりも早く、「深層学習の先にあるもの－記号推論との融合を目指して」と題した東京大学公開シンポジウムが2018年1月と2019年3月[27] に開催されている。

❺ 学会動向および国際動向

自然言語処理分野の最先端研究は、トップランク国際会議ACL（Annual Meeting of the Association for Computational Linguistics）、EMNLP（Conference on Empirical Methods in Natural Language Processing）、NAACL（North American Chapter of the Association for Computational Linguistics）などで活発に発表されている。 ACL 2020の国別採択論文数は、1位の米国が305件、2位の中国が185件、3位の英国が50件、4位のドイツが44件、日本は5位で24件であった。投稿論文数では中国が米国を上回っており、自然言語処理分野も米中2強になりつつある。

パターン処理と記号処理の統合（知覚・運動系と言語・知識処理の統合）については、AI全般のトップランク国際会議であるAAAI（Association for the Advancement of Artificial Intelligence）やIJCAI（International Joint Conferences on Artificial Intelligence）、あるいは、機械学習分野のNeurIPS（Neural Information Processing Systems）、コンピュータービジョン分野のICCV（International Conference on Computer Vision）やCVPR（Computer Vision and Pattern Recognition）、知能ロボット分野の IROS（International Conference on Intelligent Robots and Systems）などでの発表も目に付く。

米国はGoogle、Amazon、Meta、Apple、Microsoft、IBMなど、Big Tech企業を中心とする産業界による先進技術の研究開発・応用が活発である上に、技術政策として、国の安全保障目的も含めた自然言語処理の基礎研究への先行投資が際立っている。もともと米国国立標準技術研究所（NIST）による情報検索・質問応答技術、国防高等研究計画局（DARPA）による情報抽出技術や文章理解、高等研究開発局（ARDA）による質問応答技術の研究開発など、自然言語処理に関わる多くのプロジェクト（コンペティション型ワークショップを含む）が政府予算によって推進されてきた。さらに、DARPAは2018年に、AI研究に20億ドル以上の大型投資を実施するというAI Next Campaignを発表した。この発表では、AIの発展を、専門家の知識を抽出・活用するHandcrafted Knowledgeを「第1の波」、ビッグデータから知見を導く深層学習に代表されるStatistical Learningを「第2の波」とし、それに続く「第3の波」として、文脈を理解して推論するContextual Reasoningを挙げた。これによって、人間が把握・理解・行動する以上の速度でデータを生成・処理し、安全かつ高度に自律的な自動化システムを可能にしつつ、人間の意思決定を支援し、人間と機械の共生を促進することを狙っている。

中国政府は2017年7月に次世代AI発展計画を発表し、AI産業を強力に推進している。自然言語処理分野ではMicrosoftやBaidu（百度）が目に付く。Microsoftは2014年からチャットボットXiaoice（シャオアイス）を公開しており、ソーシャルネットやメッセージングのアプリケーションに導入され、世界中で6.6億人のユーザーがいるという。 BaiduはWebサーチエンジンで実績があるほか、前述のERNIE[16)] も開発している。

欧州は多数の国にまたがることから、欧州フレームワークプログラムの中で、機械翻訳を中心に自然言語処理に継続的に投資を行ってきている。産業界でも、ドイツに本社のあるDeepLの機械翻訳サービスが翻訳品質の高さで注目されている。

日本は現在「AI戦略」（2019年6月に発表、2021年・2022年にアップデートを加えている）を推進しており、文部科学省による理化学研究所革新知能統合研究センター（AIP）、経済産業省による産業技術総合研究所人工知能研究センター（AIRC）、総務省による情報通信研究機構（NICT）の三つを中核的なAI研究機関と位置付けている。自然言語処理については、NICTが機械翻訳・音声翻訳を中心に取り組んでおり、その実用化でも実績があるほか、AIPで基礎研究を推進している。パターン処理と記号処理を統合した次世代AI研究については、新エネルギー・産業技術総合開発機構（NEDO）による「次世代人工知能・ロボット中核技術開発」や「人と共に進化する次世代人工知能に関する技術開発」などの事業においてAIRCが中心的に取り組んでいるほか、文部科学省の2020年度戦略目標「信頼されるAI」とそれを受けた科学技術振興機構（JST）の戦略的創造研究推進事業（CRESTなど）でも基礎研究面が強化さ

れた。

［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・特許数ともますます増加の傾向にあり、その中で米国・中国が二強となっている。論文数シェア、Patent Asset Indexとも、中国が米国に迫っているが、米国が1位をキープしている。また、Top10%論文数では、1位の米国と2位の中国との間にまだ開きがある。欧州は総論文数・Top1%論文数で3位（Top10%論文数では中国とほぼ同程度）である。

「2.1.1 知覚・運動系のAI技術」領域で示したように、機械学習技術分野を先導する主要システムは米国から生まれており、その適用分野として言語分野が中心であることから、中国と比べて、本領域における米国の優位性が維持されている。そのため、論文数上位10位機関では、米国が6機関、中国が3機関と、米国が中国を上回っている。

（4）注目動向

［新展開・技術トピックス］

❶ 大規模言語モデル・基盤モデル

［研究開発の動向］❸で述べたように、2018年にGoogleがBERTを発表して以降、トランスフォーマー型の大規模言語モデル（Large Language Model：LLM）をMLMや次単語予測による自己教師あり学習で構築することが主流になった。さらに、2020年にOpenAIがGPT-3を発表し、学習に用いたデータ規模でBERTの約3000倍（45TB）、モデルのパラメーター数でBERTの500倍以上（1750億個）まで大規模化したことで、追加学習なしに少数の例示（Few-Shot学習/In-Context学習）だけで、さまざまな自然言語処理タスクに対応できることを示した。例えば、例示した文に続けてブログや小説を生成したり、簡単な機能説明文からプログラムコードやHTMLコードを生成したり、さまざまな質問文に対して回答を生成したりといった活用例が示された。さらに、2022年11月末に、GPT-3の改良版であるGPT-3.5に、人間のフィードバックを用いた強化学習RLHF（Reinforcement Learning from Human Feedback）を加え（InstructGPT[52]）、対話システムとしてファインチューニングされたChatGPTがWeb公開された。チャットという分かりやすいインターフェースを介して、さまざまな用途に自然言語で応えることが可能なことから、幅広く大きな話題になり、史上最速（当時）2カ月でアクティブユーザー数が1億人を突破した。

このような大量かつ多様なデータで訓練され、さまざまな下流タスクに適応できるモデルは「基盤モデル」（Foundation Model）[24]と呼ばれるようになった。計算リソース、学習（訓練）データ規模、モデル規模（モデルのパラメーター数）という3変数を大規模化するほど精度が高まるという「スケーリング則」（Scaling Law）[53]や、モデル規模が一定以上になると精度が急に高まるという「創発的能力」（Emergent Ability）[54]が観測されたことで、GPT-3以降はモデルの大規模化に拍車がかかった。［研究開発の動向］❸でも述べた通り、2021年10月にMicrosoftとNVIDIAが発表したMT-NLG（Megatron-Turing Natural Language Generation）はパラメーター数が5300億個、2022年4月にGoogleが発表したPaLM（Pathways Language Model）はパラメーター数が5400億個にも及んでいる。

さらに、OpenAIは、テキストだけでなく、画像と関連付いたテキストのペアを学習させたモデルCLIPを深層生成モデルと組み合わせて、テキストからの画像生成（Text-to-Image）も可能にした。2021年1月にDALL-E、2022年4月に改良版のDALL-E2が発表された。Googleからも、2022年5月にImagen、6月にParti、2023年1月にMuseが発表された（それぞれ異なる方式でText-to-Imageモデルを実現している）。画像生成AIの詳細は「2.1.1 知覚・運動系のAI技術」に記載している。

従来、機械学習ベースのAIは、タスクごとに学習させることが必要な目的特化型AIだったが、基盤モデルを用いたAIは、汎用性やマルチモーダル性が高まり、それ一つでさまざまなタスクに使えるようになりつつある。そのような発展として、Google DeepMindのFlamingo[28]やGato[29]が注目された（それぞれ

2022年4月と5月に発表された）。 Flamingoはテキスト・画像・動画を扱うことができ、Few-Shot学習で新しいタスクに適応できる（パラメーター数は800億個）。例えば、動画（画像の系列）とテキストのペアを例示して、その後に続くテキストを生成するというようなことが可能である。ここまで示してきたモデルが扱うタスクは、基本的にテキストや画像を生成・出力するものだったが、Gatoは一つのモデルで、テキストや画像の生成・出力から動作の生成・制御まで、さまざまなタスクを扱うことができる（パラメーター数は12億個）。その詳細は［注目すべき国内外のプロジェクト］❶で紹介する。

基盤モデル・生成AIの動向として、OpenAIのChatGPTは、2022年11月末に登場した時点ではGPT-3.5をベースとしていたが、2023年3月にはGPT-4が登場し、オプションとしてGPT-3.5とGPT-4のどちらを使うかが選択できるようになった。さらに、OpenAIに投資しているMicrosoftは、検索エンジンBingにGPT-4を組み合わせて、チャットとWeb検索を連携させた機能を提供した。 GPT-3.5やGPT-4は、それらのモデルの学習（訓練）に用いたデータ以降の出来事についての情報を持っていないが、検索エンジンと組み合わせることで、最新の情報にまで対応することが可能になった。一方、ChatGPTのような対話型生成AIは検索市場を脅かすと見られ、Googleはコードレッドを発して、対話型生成AIの開発に参入し、Bardを発表した。2023年2月に限定的な試験公開を行った後、5月に日本語にも対応したBardを一般公開した。 Bardのベースとなる大規模言語モデルは、当初LaMDAが用いられたが、一般公開時にはPaLM2（2022年4月に発表されたPaLMの性能をさらに高めたもの）に置き換えられた。その後、2023年12月にはPaLM2の後継モデルとしてGeminiが発表され、2024年2月にはGeminiモデルの大きなアップデートが行われ、そのタイミングでモデルだけでなく、対話型生成AIのサービス名もBardからGeminiに変更された。 Geminiモデルはマルチモーダルモデルとして強化されているのが特長であり、その規模や用途の異なる3種類のモデル（Gemini Ultra、Gemini Pro、Gemini Nano）が作られている。また、OpenAIも2024年5月に、最新モデルGPT-oを発表し、それを用いたChatGPT-4oを公開した。ChatGPT-4oでは、テキスト、音声、画像、映像がシームレスに扱われ、リアルタイム音声対話も可能になった。このような大規模言語モデル・基盤モデル、対話型生成AIまわりでは、新たな参入とバージョンアップが次々とあり、オープンソース版も公開され、動きが活発かつ急速である。

❷ 深層学習の発展・拡張による知能のモデル化

［研究開発の動向］❹で述べたように、パターン処理と記号処理を比較的疎な形で組み合わせたシステム化は以前から行われてきたが、深層学習が発展し、自然言語処理やナレッジグラフを用いた処理のような記号レベルの処理も深層学習によって再構成されるようになった。これにより、パターン処理と記号処理を統一的な考え方で統合する（共通の枠組み上に融合する）可能性が見えてきた。もともと深層学習や強化学習は、人間の知覚・運動系に類する学習モデルであり、その拡張として言語・知識系までカバーしようとするのは、自然な発展の方向である。その際、言語・知識系は二つの側面から位置付けられる。つまり、知覚・運動系からボトムアップに創発的に言語獲得・知識化が行われるという側面と、人間が他者から教えてもらったり本を読んだりするように外部の知識源から取り込むという側面がある。「2.1.7 計算脳科学」や「2.1.8 認知発達ロボティクス」の研究領域で人間の知能に関する研究が進展し、それに基づくニューロシンボリックAI（Neuro-Symbolic AI）、神経科学（脳科学）にインスパイアされたAI（Neuroscience-Inspired AI）、記号創発ロボティクス（Symbol Emergence in Robotics）といったコンセプトに基づく研究開発が進んでいる。

具体的な研究事例をいくつか挙げる。AI21 LabsのSenseBERT[30]は、深層学習ベースの言語モデル中に外部知識（ナレッジグラフ、意味ネットワーク）を組み込んだ。 Julassic-Xプロジェクトでは、言語モデルとデータベースなどの外部情報システムの結合を提唱している。MIT他によるNS-CL（Neuro-Symbolic Concept Learner）[31]は、画像や環境・空間の認識と質問応答という2系統を持ち、画像・空間系統は教師あり学習、質問応答系統は強化学習を用い、2系統のマルチタスクのカリキュラム学習を通して、視覚

的概念・単語・意味解析などを学習する。松尾豊（東京大学）の提案する知能の2階建てモデル[32]は、1階部分が深層学習をベースとした知覚・運動系で、その外界とのインタラクションを通して作られた世界モデルを介して、2階部分のトランスフォーマー的な言語・知識系が動くというものである。谷口忠大（立命館大学）らが提唱する記号創発ロボティクス[33]では、外界とのインタラクションを通して言語が創発的に形成されることを、確率的生成モデルをベースにモデル化することを試みている。

さらに、Metaから2022年8月に発表されたAtlas[34]は、大規模言語モデル（BERT派生のT5）に外部の情報を検索する機構（Contriever）を組み合わせて拡張した。Atlasの言語モデルの規模は110億パラメーターだが、64事例のFew-Shot学習により、質問応答タスクにおいて、5400億パラメーターのPaLMよりも高い精度を達成した。また、Googleから2022年10月に発表されたMind's Eye[35]は、大規模言語モデルに物理シミュレーター（DeepMindのMuJoCo）を組み合わせた。物理的推論を必要とする問題文[7]が与えられると、それは言語モデルによって物理シミュレーターを動かすためのコード（レンダリングのための情報を含む）に変換され、物理シミュレーションが実行される。そこで描画された物理世界での結果を用いて、問題文に対する答えを導出する。大規模言語モデルは基本的にテキストで訓練されるため、物理的推論が必要な問題に対しても、テキストを根拠として答えを出さざるを得ないという限界があった。Mind's Eyeは、この限界を克服する枠組みを示した[8]。

これらのシステムが示すように、今後、大規模言語モデルを人間とのインターフェースとしてさまざまな情報システムやAIシステムを結合・統合するようなアプリケーションの開発が急速に進みつつある。実際、大規模言語モデル・基盤モデルや生成AIのAPIが公開された後、それらでは足りない機能を外部モジュールとして用意して組み合わせる仕組みが多数提供されるようになった。［研究開発の動向］❸で述べたRAG（検索拡張生成）などもその一例である。さらに、与えられたゴールからそれを達成するための手順を計画し、大規模言語モデル・基盤モデルだけでなく、検索処理や蓄積処理などを含む複数の処理モジュールを組み合わせて実行することを自動的に行うAutoGPTやAgentGPTといったツールも開発されている。

なお、関連する話題として、AIとシミュレーションの融合、科学・数学へのAI活用、ゲームAIなどは、「2.1.6 AI・データ駆動型問題解決」で取り上げている。

［注目すべき国内外のプロジェクト］

❶ GATO[29]

前述したように、Gatoは一つのモデルで、テキストや画像の生成・出力から動作の生成・制御まで、さまざまなタスクを扱うことができる。具体的には、ビデオゲームをプレーしたり、チャットをしたり、文章を書いたり、画像にキャプションを付けたり、ロボットアームを制御してブロックを並べ替えたりすることができる。そのために604種類のタスクの実行例を学習した。そのようなさまざまな種類の実行例を学習するため、テキスト、画像、離散値、連続値などデータタイプの違いに応じたトークン化を行って、トランスフォーマーに入力している。現状で最大規模の基盤モデルと考えられるPaLMのモデル規模が5400億パラメーターであるのに対して、Gatoのモデル規模は12億パラメーターに抑えられている。これはロボットアームのリアルタイム制御のために、コンパクトなモデルにする必要があったためといわれている。個々のタスクについては必ずしもトップ性能とは言えないものがあるが、一つのモデルでこれだけ多種多様なタスクをそこそこの性能で実行できるというのは、汎用性という面で大きなインパクトのある成果である。開発元であるGoogle DeepMindは、GatoをGeneralist Agentと呼んでいる。

7 このような問題のベンチマークのためUTOPIAデータセットが構築された。

8 Mind's Eyeは物理世界での推論を扱うための拡張だが、関連して物理世界に対する動作計画・行動という面では、PalMSayCanやRT-1などの取り組みがある。これらについては「2.1.1 知覚・運動系のAI技術」のおける［注目すべき国内外のプロジェクト］❷❸で取り上げている。

❷ AlphaCodeとOpenAI Codex

Google DeepMindから2022年2月に発表されたAlphaCode[36]と、OpenAIから2021年8月に発表されたOpenAI Codex[37]は、トランスフォーマー型の言語モデルを用いてプログラムコードを生成する。

AlphaCodeは、競技プログラミングコンテストへの参加シミュレーションにおいて、コンテスト参加者の54%以内にランクされる成績となった。人間を上回るというような華々しい結果ではないが、競技プログラミングで人間並みの結果が出せるということを示した。モデルの学習は、まずGitHubで公開されているコードで事前学習が行われ、その後、競技プログラミングの小規模なデータセットで追加調整が行われた。AlphaCodeは、競技プログラミングの問題文が入力されたら、このモデルを用いて大量のC++とPythonのプログラムをいったん生成し、次いでそれらに対してフィルタリングやクラスタリングを行うことで、出力するプログラムコードを絞り込む。

OpenAI CodexはGPT-3をベースに、GitHubで公開されている大量のコードから学習しており、入力された文章に対して、Pythonをはじめ10以上のプログラミング言語でコードを出力可能である。 GitHubにおいてCopilotの機能として利用可能になっている。また、同じく OpenAI から発表された前述のChatGPTも、自然言語による指示からプログラムを生成したり、プログラムの中のバグを指摘したりすることが可能である。

（5）科学技術的課題

❶ 大規模言語モデル・基盤モデルの課題

［新展開・技術トピックス］❶に示した通り、トランスフォーマー型の深層学習モデルは大規模化とともに、汎用性・マルチモーダル性を高めている。さまざまなタスクに適用でき、あたかも人間の専門家が行ったかのような結果を出すことができつつある。特に2022年には、画像生成AI（Midjourney、Stable Diffusionなど）やChatGPTが、一般にも容易に利用できる形で公開されたため、機能・性能に関して大きな話題になった。しかし、その一方でさまざまな課題が指摘されつつある（「2.1.9 社会におけるAI」でもこの課題を詳しく取り上げる）。

まず、極めて大規模なモデルであることに起因して、次のような問題がある。

- 学習に極めて多大な計算リソースがかかる[9]。環境負荷もかかる。
- そのような多大な計算リソースをかけて、最先端の基盤モデルを作れるのは、世界中でもごく一部の企業（いわゆるBig Tech企業）に限られる[10]。
- 学習に使うデータが極めて大規模で、そのデータセットは公開されないため、他者による再現性評価がされない。
- 1回の学習に費用も時間もかかるため、頻繁に作り直すことは難しく、学習を実行した時期以降に起きた出来事に追従できない。

また、現状の機能においては、次のような懸念が出ている。

- 大量の学習データに基づき統計的な振る舞いをするのであって、（一見それらしく見えても）意味を理解しているわけではないし、物理法則・因果律・数学公式などの法則・知識にのっとっているわけではない。したがって、出力される結果の正しさは保証されず、むしろ、非常に自然な感じで間違った出力を生成するため、正しい結果と間違った結果を見分けにくくなっている（Hallucinationと呼ばれる）。

9 2018年発表のBERTの場合で、1回の学習に要する費用（クラウド利用料換算）は百万円前後という報告[38]があり、そこからのモデル規模拡大のペースは年10倍とも言われる。さまざまな基盤モデルの規模と訓練費用をプロットした1.1.2節の図1-1-7から、億円規模にも達していることが分かる。

10 オープンソース版も公開されるようになり、最大規模のモデルでなければ、Big Tech 企業以外でも作る動きが出てきている。しかし、大学の研究室レベルで作ることは難しい。

前述のMind's Eyeや、言語モデルを検索と組み合わせるperplexity.aiのような試みはあるが、それで十分に対応できているわけではない。

・膨大な学習データの中身は必ずしも十分に精査されてはおらず、倫理的な面、公平性の面、プライバシーの面などから問題のある出力をするリスクがある。このような問題に対して、ChatGPTでは、人間からのフィードバックを用いた強化学習によって調整する仕組み（RLHF）や、差別的・性的・暴力的などの要素を検出・フィルタリングする仕組み（Content Moderation）が組み込まれているが、いっそうの強化が必要であろう。

・上記のような不完全・不適切な面を持ちながらも、一見すると、まるで人間の専門家並みの結果を出すことから、その内容を人々が信じてしまいがちであることや、AIによるものだと気付くことが難しいことから、社会的な問題が起き得る。

上記最後の項目に関わるものとして、GPT-3を用いて掲示板に自動投稿されていた文章が、1週間、人間が書いたものでないと見破られなかったという事例や、ChatGPTは米国のMBA、法律、医療の試験に合格できるレベルにあるという報告がある。しかし、Metaが開発したGalacticaは、4800万件を超える科学技術文献から学習し、科学的知識に関する利用者からの質問に答えることができるシステムだったが、誤りや偏見を含む答えを出すことがあったため非難され、わずか3日間で公開停止となった。 ChatGPTを使って試験レポートや論文を書くことを禁じる動きが出てきている一方、それを見分けることができるのかという問題もある。Open AI はAI が生成した文章と人が生成した文章を見分けるツールも提供しているが、ChatGPTや画像生成AIを悪用したフェイク作成も深刻化しつつある問題である。

また、画像生成AIやプログラム生成AIに対して、学習にデータを使われたアーティストなどの側から強い反発が出されている。創作的な活動に関わる権利保護や、AI生成物に関わる著作権についての議論も必要になってくると思われる。

なお、スタンフォード大学の基盤モデル研究センター（Center for Research on Foundation Models：CRFM）は、基盤モデル周辺の倫理的・社会的な側面も含めて研究に取り組むとうたっている[24)]。

❷ 第4世代AIのアーキテクチャー

2.1節の冒頭（総論）で述べたように、機械学習ベースのAIを第3世代AIとしたとき、その次の世代、つまり第4世代AIへアーキテクチャーを発展させることが考えられている[25)]。第3世代AIは、a.学習に大量の教師データや計算資源が必要であること、b.学習範囲外の状況に弱く、実世界状況への臨機応変な対応ができないこと、c.パターン処理は強いが、意味理解・説明などの高次処理はできていないこと、といった問題を抱えており、第4世代AIではこれらの解決が望まれる。これらの問題の解決には、古くからAIの基本問題として掲げられている記号接地問題やフレーム問題も関わっており、これを再定義して段階的に解決していくようなアプローチが必要になると思われる。

第4世代AIに向けたアプローチとして、以下のような取り組みが注目される。

第1のアプローチとして、知覚・運動系AI（2.1.1節）と言語・知識系AI（2.1.2節）を融合しようという取り組みがある[25), 26), 27), 39)]。これは、人間の知能・思考は、即応的ループ（システム1）と熟考的ループ（システム2）とから成るという「二重過程理論」（Dual Process Theory）に相当する。社会心理学・認知心理学などの心理学分野で提案されていたが、ノーベル経済学賞を受賞したDaniel Kahnemanの著書「Thinking, Fast and Slow」[1)] でよく知られるようになった。脳・神経科学の面からも論じられている[40)]。従来の深層学習はシステム1に相当するものであり、深層学習研究でACMチューリング賞を受賞したYoshua BemgioはNeurIPS 2019で「From System 1 Deep Learning to System 2 Deep Learning」と題した招待講演[26)] を行った。また、松尾豊は「知能の2階建てアーキテクチャ」[32)] を提案している。

第2のアプローチは、認知発達ロボティクスの分野（2.1.8節）で研究されている認知発達・記号創発のモデルである。環境や他者とのインタラクションを通して、人間が他者との概念共有や言語の獲得をしていくプロセスに着目している。ここで近年注目されているのが、生物の脳の情報処理に関する「自由エネルギー原理」（Free-Energy Principle）[41), 42), 43), 44)] である。これは「生物は感覚入力の予測しにくさを最小化するように内部モデルおよび行動を最適化し続けている」という仮説である。ここでいう「予測のしにくさ」は、内部モデルに基づく知覚の予測と実際の知覚の間の予測誤差を意味し、変分自由エネルギーと呼ばれるコスト関数で表現される。さまざまな推論・学習、行動生成、認知発達過程を統一的に説明できる原理として注目されている。

他にも、知能を機能モジュールに細かく分けて組み合わせるアプローチや、人工ニューロンをベースとした全脳シミュレーションなどの取り組みもあり、「2.1.7 計算脳科学」や「2.1.8 認知発達ロボティクス」のような人間の知能の解明に関する知見から、新たなAIアーキテクチャーにつながるヒントが期待される。

その一方で、［新展開・技術トピックス］❶に示したように、トランスフォーマー型の深層学習モデルについても、極めて大規模なデータで訓練した基盤モデルは、人間と区別するのが難しいほどの生成能力を示すことが明らかになった。これはある意味、人間離れした方向へ進みつつあるわけだが、このような力任せのアプローチが第4世代AIにつながる可能性もあるかもしれない。

JST CRDSでは、以上の課題を踏まえつつ、より広いスコープで調査・検討を行い[56)]、2024年3月に戦略プロポーザル「次世代AIモデルの研究開発」[57)] を公開した。

❸ 真の意味理解・常識推論

さまざまな言語モデルが提案され、それらの性能評価には共通のベンチマークが用いられてきた。代表的なタスクとして、テキストを読んだ上で、その内容に関する質問に答える文書読解タスクがあり、そのベンチマークとしてよく使われているのがSQuADである。また、単一タスクでなく多数のタスクを評価できるベンチマークのセットとして、GLUEもよく使われている。 GLUEには含意関係判定、同義性判定、質問・回答解析、肯定的か否定的かの感情解析、文法チェックなどの約10種類のタスクが用意されている。2022年には日本語版のJGLUE[45)] も公開された。

言語モデルの改良により、これらのベンチマークでのスコアが人間の平均的スコアを上回ったという報告も出てきている。しかし、文章の意味を理解しなくても、統計的な傾向を捉えれば正解を当てられるような問題が多かったため、見かけ上、高いスコアが得られただけで、本当に意味を理解しないと当たらないような問題に対してはスコアが低くなることが指摘されている[46), 47)]。自然言語処理研究の黎明期からその難しさも含めて認識されている常識推論に向けて、含意関係認識やストーリー予測などのサブタスクの設定、敵対的サンプル（Adversarial Examples）なども取り入れたベンチマークの構築が求められる[48)]。その試みの一つとして、意味を理解しないと正答が難しい問題を多く含む文書読解タスクのベンチマークDROP（Discrete Reasoning Over the content of Paragraphs）[49)] が作られた。 SQuAD 2.0でのトップスコアが93%前後だが、DROPでは人間のスコアが96.4%であるのに対して、BERTのスコアが32.7%と、機械にとって難しいベンチマークとなっている（2022年末の時点でDROPのトップスコアは88%である）。

GPT-3などの基盤モデルにおいても、意味理解・常識推論は課題とされており、依然として難しく重要な課題である。

（6）その他の課題

❶ 新たな研究課題のための戦略的なベンチマーク環境・体制の構築

かつては形態素解析・構文解析などの自然言語処理のサブタスク別の精度評価が、技術改良における主たる目標になっていた。しかし、ニューラルネット自然言語処理では応用ごとのEnd-to-End最適化のアプローチが取られるため、サブタスクに注力することのウェイトが下がってきた。このような状況で、近年は、

GLUEやその日本語版の JGLUEのように問題ごとのベンチマークタスクを複数用意して、言語モデルの汎用性あるいは特定用途の有用性を評価するようになった。

同時に、AI分野では共通タスク・共通データセットでのコンペティション型ワークショップが盛んに実施されてきた。特に米国NISTが自然言語処理・情報検索領域でこれを推進してきたことは先駆的で、この分野の基礎研究の強化を大きく牽引した。このような活動の推進においては、タスク設定に関わる目利き人材がキーになる。取り組むことに大きな価値があり、しかも無理難題ではなく挑戦意欲をかき立てるようなタスク設定のさじ加減を適切にでき、コミュニティーをリードできるような人材が求められる。

このようなベンチマークやコンペティションのタスク設定・データセット構築は、米国がリードし、中心的に貢献してきたが、日本においても、データセット構築とコンペティション型ワークショップ運営に20年以上取り組んでいる国立情報学研究所のNTCIR（NII Testbeds and Community for Information access Research）プロジェクトなどの実績がある。今後、言語・知識系と知覚・運動系の統合AIや、マルチモーダルAI・実世界連結応用などの新しい研究課題に取り組んでいくにあたっては、新たなタスク設定・データセット構築が重要である。日本としての戦略的取り組みが期待される。

その一方で、科学技術的課題❶で述べたように、最先端の基盤モデルは大規模化し、それを作れるのは、世界中でもごく一部の企業（いわゆるBig Tech企業）に限られる状況であり、その大規模データセットは公開されず、他者による再現性評価がされない。そのため、上で述べたようなベンチマークやコンペティションによる取り組みに限界が生じている。

また、基盤モデルは、頻繁に作り直すことは難しく、学習を実行した時期以降に起きた出来事に追従できないという問題に対する取り組みとして、RealTime QA[50] というベンチマークタスクが実施されている。ここでは、最新のニュース記事を用いた質問が出題される。

❷ 経済安全保障上の課題、および、大規模コンピューティング基盤と大規模データ構築エコシステム

科学技術的課題で述べたように、最先端の基盤モデルの開発には極めて大量の計算リソースと大規模データを必要とする。最先端の基盤モデルはGoogleやOpenAIなどのBig Techのみが開発でき、日本はそのAPI（Application Programming Interface）を利用させてもらっているという状況である。今後、基盤モデルがさまざまなタスクの高度な自動化に使われていくならば、このBig Tech依存の状況は、「2.1.1 知覚・運動系のAI技術」でも述べた通り、経済安全保障上の懸念点にもなる。

計算リソース面を考えると、その実行環境はもはや大学の一研究室で確保できる規模ではなくなっている。大規模コンピューティング基盤の共同利用施設が不可欠であり、産業技術総合研究所のAI橋渡しクラウドABCI（AI Bridging Cloud Infrastructure）や理化学研究所のスーパーコンピューター「富岳」がこの役割を担っている。特にABCIでは、BERTの事前学習済みモデルを共同利用できるように公開している。このような共同利用施設・計算資源の継続的な強化・整備が極めて重要である。

データ規模の面については、Big Techと並ぶのは難しいし、今後果てしなく大規模化が続くというより、規模と質のバランスが重要になってくるであろう。日本として、何らかの付加価値の高いデータを構築・整備していくことを考えていくのがよい。そのためには、それを集中的に実行する組織を作るというよりも、さまざまな組織が何らかの形で貢献し合い、データの質や付加価値を高めていくような、データ構築エコシステムを考えていくべきかもしれない。

❸ 人間・社会面の深い理解・考察に基づく取り組み・人材育成

基盤モデルによって人間と区別できないような文章が生成されることが社会的懸念を生んでいるように、この研究分野から生み出される技術がもたらす社会的影響が増大している。また、今後の発展においては、人間の知能のメカニズムに学ぶという面が強くなりつつある。 AI全般の倫理的・法的・社会的課題（Ethical, Legal and Social Issues：ELSI）面については「2.1.9 社会におけるAI」にて論じるが、こ

の分野の技術の社会的影響や知能そのものに関する深い理解・考察とともに研究開発を進める必要があろう。人間の知能の情報処理メカニズムの理解という面では、「2.1.7 計算脳科学」や「2.1.8 認知発達ロボティクス」は日本が実績のある研究分野であり、その研究成果をこの分野の発展・強みに生かしていきたい。

1980年～2000年頃、ルールベース方式が主流だった時代は、カナ漢字変換、機械翻訳、サーチエンジンなどをターゲットとして、電気系大手企業の各社が自然言語処理の研究開発に力を入れていた。しかし、機械学習を用いる方式では、大量のテキストデータを使うことが研究開発に不可欠で、多数のユーザーを抱えるインターネットサービスを運営している企業において、自然言語処理への取り組みが活発になってきた。特にBig Techは保有するデータ量やそれを処理する計算機パワーが圧倒的で、データ収集や実験が行いやすいことは研究者に魅力的である。 AI分野の人材争奪戦が国際的に激しくなる中で、AI人材の教育・育成、幅広い人材の獲得や引き留めのための施策も重要である。

（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	○	↗	「AI戦略」を推進しており、理研AIP、産総研AIRC、NICTの三つを中核的なAI研究機関と位置付けている。自然言語処理については、NICTが機械翻訳・音声翻訳を中心に取り組んでおり、その実用化でも実績があるほか、AIPで基礎研究を推進している。パターン処理と記号処理を統合した次世代AI研究については、NEDO「次世代人工知能・ロボット中核技術開発」事業においてAIRCが中心的に取り組んでいるほか、文科省の2020年度戦略目標「信頼されるAI」とそれを受けたJST戦略的創造研究推進事業でも基礎研究面が強化された。言語処理学会に活気があり、ACL・EMNLPなどの国際的トップカンファレンスでも一定数の発表がなされている。言語資源や研究利用可能なコンテンツの整備が不足しているとともに、言語モデルの超大規模化に対して計算機環境面で劣勢である。
	応用研究・開発	○	→	NICTが音声翻訳、大規模情報分析、災害関連情報分析などで実用性の高い研究成果をリリースしてきた。 NEC、NTTドコモ、日本IBM、SoftBank、トヨタなどの民間企業でも、自然言語処理技術に基づく製品化・事業化が行われてきた。
米国	基礎研究	◎	→	Google、Meta、OpenAI、Microsoftなど、産業界による先進技術の研究開発・応用が活発である上に、技術政策として、国の安全保障目的も含めた自然言語処理の基礎研究への先行投資が際立っている。 ACL、EMNLPなどの有力な国際会議などでの発表の多くは米国発である。もともとNISTによる情報検索・質問応答、DARPAによる情報抽出・文章理解、ARDAによる質問応答など、自然言語処理技術に関わる多くのプロジェクトやコンペティションが政府予算によって推進されてきた。さらに2018年にDARPAはAI Next Campaignを発表した
	応用研究・開発	◎	↗	上述した基礎研究が大学教員の移籍などによって、比較的短期間にGoogle、Microsoft、IBM、Facebook、Amazonなどの応用研究・開発に回るサイクルが確立している。これらの企業の研究開発への投資も大きく、また、ベンチャーによる取り組みも活発である。
欧州	基礎研究	○	→	欧州は多数の国にまたがることから、欧州フレームワークプログラムの中で、機械翻訳を中心に自然言語処理に継続的に投資を行ってきているが、突出した研究は少ない。
	応用研究・開発	○	→	グローバル企業の研究所が存在し、一定のアクテビティーはあるが、米国主導による産業化の側面が強い。産業界では、DeepLの機械翻訳サービスが翻訳品質の高さで注目されている。

中国	基礎研究	◎	↗	北京大学・清華大学などの有力大学やMicrosoft Research Asia、Baiduなどの民間企業の研究所を中心に基礎研究が進められている。ACLなどのトップ国際会議でも論文採択数は米国に次いで中国が2位である。
	応用研究・開発	◎	↗	中国政府は2017年7月に次世代AI発展計画を発表し、AI産業を強力に推進している。自然言語処理分野ではMicrosoft やBaidu（百度）が目に付く。Microsoftは2014年からチャットボットXiaoice（シャオアイス）を公開しており、ソーシャルネットやメッセージングのアプリケーションに導入され、世界中で6.6億人のユーザーがいるという。 BaiduはWebサーチエンジンで実績があるほか、BERTを改良したERNIEも開発している。
韓国	基礎研究	△	→	KAIST、ETRI、KISTIなどの有力大学、国研を中心に基礎研究が進められている。ただ、ACLなどのトップカンファレンスでの韓国発の発表は減少している。
	応用研究・開発	○	↗	政府がNaver、サムソン、SK Telecom、HyundaiなどとArtificial Intelligence Research Instituteを設立予定。 Naverなどによるチャットボットや機械翻訳のサービスの開始が相次ぐ。

（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）トレンド　※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）Kahneman, D., Thinking, Fast and Slow（Farrar, Straus and Giroux, 2011）.（邦訳：村井章子訳,『ファスト＆スロー：あなたの意思はどのように決まるか?』, 早川書房, 2014年）

2）Christopher D. Manning, “Computational linguistics and deep learning”, Computational Linguistics Vol.41, No.4（2015）, pp.701-707. DOI: 10.1162/COLI_a_00239

3）Minh-Thang Luong, Hieu Pham and Christopher D. Manning, “Effective Approaches to Attention-based Neural Machine Translation”, Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing（EMNLP 2015; Lisbon, Portugal, September 17-21, 2015）, pp.1412-1421.

4）久保陽太郎,「ニューラルネットワークによる音声認識の進展」,『人工知能』（人工知能学会誌）31巻2号（2016年3月）, pp.180-188.

5）中澤敏明,「機械翻訳の新しいパラダイム：ニューラル機械翻訳の原理」,『情報管理』66巻5号（2017年8月）, pp.299-306. DOI: 10.1241/johokanri.60.299

6）坪井祐太・海野裕也・鈴木潤,『深層学習による自然言語処理』（講談社, 2017年）.

7）岡崎直観,「言語処理における分散表現学習のフロンティア」,『人工知能』（人工知能学会誌）31巻2号（2016年3月）, pp.189-201.

8）Anand Rajaraman and Jeffrey David Ullman, *Mining of Massive Datasets*（Cambridge University Press, 2012）.（邦訳：岩野和生・浦本直彦訳,『大規模データのマイニング』, 共立出版, 2014年）

9）大塚裕子・乾孝司・奥村学,『意見分析エンジン―計算言語学と社会学の接点』（コロナ社, 2007年）.

10）Special Issue: “This is Watson”, *IBM Journal of Research and Development* Vol. 56 issue 3-4 (May-June 2012).
11）西田京介・斉藤いつみ,「深層学習におけるアテンション技術の最新動向」,『電子情報通信学会誌』101巻6号（2018年）, pp. 591-596.
12）Ashish Vaswani, et al., “Attention Is All You Need”, *Proceedings of the 31st Conference on Neural Information Processing Systems*（NIPS 2017; Long Beach, CA, USA, December 4-9, 2017).
13）Jacob Devlin, et al., “BERT：Pre-training of Deep Bidirectional Transformers for Language Understanding”, arXiv: 1810.04805 (2018).
14）Yinhan Liu, et al., “RoBERTa：A Robustly Optimized BERT Pretraining Approach”, arXiv: 1907.11692 (2019).
15）Zhenzhong Lan, et al., “ALBERT：A Lite BERT for Self-supervised Learning of Language Representations”, arXiv: 1909.11942 (2019).
16）Yu Sun, et al., “ERNIE: Enhanced Representation through Knowledge Integration”, arXiv: 1904.09223 (2019).
17）Li Dong, et al., “Unified Language Model Pre-training for Natural Language Understanding and Generation”, arXiv：1905.03197 (2019).
18）Jiasen Lu, et al., “ViLBERT: Pretraining Task-Agnostic Visiolinguistic Representations for Vision-and-Language Tasks”, arXiv: 1908.02265 (2019).
19）Weijie Su, et al., “VL-BERT: Pre-training of Generic Visual-Linguistic Representations”, arXiv: 1908.08530 (2019).
20）Yen-Chun Chen, et al., “UNITER：UNiversal Image-TExt Representation Learning”, arXiv: 1909.11740 (2019).
21）Tom Brown, et al., “Language Models are Few-Shot Learners”, *Proceedings of the 34th Conference on Neural Information Processing Systems*（NeurIPS 2020; December 6-12, 2020).
22）Shaden Smith, et al., “Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, A Large-Scale Generative Language Model”, arXiv：2201.11990 (2022).
23）Aakanksha Chowdhery, et al., “PaLM：Scaling Language Modeling with Pathways”, arXiv: 2204.02311 (2022).
24）Rishi Bommasani, et al., “On the Opportunities and Risks of Foundation Models”, arXiv: 2108.07258 (2021).
25）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：第4世代AIの研究開発―深層学習と知識・記号推論の融合―」, CRDS-FY2019-SP-08（2020年3月）.
26）Yoshua Bengio, “From System 1 Deep Learning to System 2 Deep Learning”, *Invited Talk in the 33rd Conference on Neural Information Processing Systems*（NeurIPS 2019; Vancouver, Canada, December 8-14, 2019). https://slideslive.com/38922304/from-system-1-deep-learning-to-system-2-deep-learning (accessed 2023-02-01）
27）東大TV,「深層学習の先にあるもの－記号推論との融合を目指して（2）」(2019年3月5日）. https://todai.tv/contents-list/2018FY/beyond_deep_learning (accessed 2023-02-01）
28）Jean-Baptiste Alayrac, et al., “Flamingo: a Visual Language Model for Few-Shot Learning”, arXiv: 2204.14198 (2022).
29）Scott Reed, et al., “A Generalist Agent”, arXiv: 2205.06175 (2022).

30）Yoav Levine, et al., “SenseBERT: Driving Some Sense into BERT”, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*（ACL 2020; July 5-10, 2020）.
31）Jiayuan Mao, et al., “The Neuro-Symbolic Concept Learner: Interpreting Scenes, Words, and Sentences From Natural Supervision”, *Proceedings of the 7th International Conference on Learning Representations*（ICLR 2019; New Orleans, USA, May 6-9, 2019）.
32）松尾豊,「知能の2階建てアーキテクチャ」,『認知科学』(日本認知科学会誌）29巻1号（2022年3月）, pp. 36-46. DOI: 10.11225/cs.2021.062
33）谷口忠大,『心を知るための人工知能：認知科学としての記号創発ロボティクス』(共立出版, 2020年）.
34）Gautier Izacard, et al., “Atlas: Few-shot Learning with Retrieval Augmented Language Models”, arXiv: 2208.03299（2022）.
35）Ruibo Liu, et al., “Mind’s Eye: Grounded Language Model Reasoning through Simulation”, arXiv: 2210.05359（2022）.
36）Yujia Li, et al., “Competition-level code generation with AlphaCode”, *Science* Vol. 378, Issue 6624（December 8, 2022）, pp. 1092-1097. DOI: 10.1126/science.abq1158
37）Mark Chen, et al., “Evaluating Large Language Models Trained on Code”, arXiv: 2107.03374（2021）.
38）“The Staggering Cost of Training SOTA AI Models”, Synced（June 17, 2019）. https://syncedreview.com/2019/06/27/the-staggering-cost-of-training-sota-ai-models/（accessed 2023-02-01）
39）科学技術振興機構 研究開発戦略センター,「科学技術未来戦略ワークショップ報告書：深層学習と知識・記号推論の融合によるAI 基盤技術の発展（2020年1月30日開催）」, CRDS-FY2019-WR-08（2020年3月）.
40）Jeff Hawkins, *A Thousand Brains: A New Theory of Intelligence*（Basic Books, 2022）.（邦訳：大田直子訳,『脳は世界をどう見ているのか：知能の謎を解く「1000の脳」理論』, 早川書房, 2022年）
41）Karl J. Friston, James Kilner and Lee Harrison, “A free energy principle for the brain”, *Journal of Physiology-Paris* Vol. 100, Issues 1-3（July-September 2006）, pp. 70-87. DOI: 10.1016/j.jphysparis.2006.10.001
42）Karl J. Friston, “The free-energy principle: a unified brain theory?”, Nature Reviews Neuroscience Vol. 11, No. 2（January 2010）, pp. 127-38. DOI：10.1038/nrn2787
43）磯村拓哉,「自由エネルギー原理の解説：知覚・行動・他者の思考の推論」,『日本神経回路学会誌』25巻3号（2018年）, pp. 71-85. DOI：10.3902/jnns.25.71
44）乾敏郎・阪口豊,『脳の大統一理論：自由エネルギー原理とはなにか』(岩波書店, 2020年）.
45）栗原健太郎・河原大輔・柴田知秀,「JGLUE：日本語言語理解ベンチマーク」,『自然言語処理』(言語処理学会誌）29巻2号（2022年6月）, p. 711-717. DOI: 10.5715/jnlp.29.711
46）Robin Jia and Percy Liang, “Adversarial Examples for Evaluating Reading Comprehension Systems”, arXiv: 1707.07328（2017）.
47）Saku Sugawara, et al., “What Makes Reading Comprehension Questions Easier?”, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*（EMNLP 2018; Brussels, Belgium, October 31-November 4, 2018）, pp. 4208-4219.
48）井之上直也,「言語データからの知識獲得と言語処理への応用」,『人工知能』(人工知能学会誌）33巻3号（2018年5月）, pp. 337-344.
49）Dheeru Dua, et al., “DROP: A Reading Comprehension Benchmark Requiring Discrete

Reasoning Over Paragraphs”, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*（NAACL 2019; Minneapolis, Minnesota, USA, June 2-7, 2019）.
50）Jungo Kasai, et al., “RealTime QA: What’s the Answer Right Now?”, arXiv：2207.13332（2022）.
51）岡崎直観・他,『IT Text 自然言語処理の基礎』(オーム社, 2022年).
52）Long Ouyang, et al., “Training language models to follow instructions with human feedback”, arXiv: 2203.02155（2022）.
53）Jared Kaplan, et al., “Scaling Laws for Neural Language Models”, arXiv: 2001.08361（2020）.
54）Jason Wei, et al., “Emergent Abilities of Large Language Models”, arXiv: 2206.07682（2022）.
55）Yunfan Gao, et al., “Retrieval-Augmented Generation for Large Language Models: A Survey”, arXiv: 2312.10997（2023）. https://doi.org/10.48550/arXiv.2312.10997
56）科学技術振興機構 研究開発戦略センター,「科学技術未来戦略ワークショップ報告書：次世代AIモデルの研究開発 ～技術ブレークスルーとAI×哲学～（2023年11月23日・12月20日開催）」, CRDS-FY2023-WR-03（2024年2月）.
57）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：次世代AIモデルの研究開発」, CRDS-FY2023-SP-03（2024年3月）.