2．俯瞰区分と研究開発領域

2.1　人工知能・ビッグデータ

人工知能（Artificial Intelligence：AI）技術は、人間の知的活動（認識、判断、計画、学習等）をコンピューターで実現するための技術群である。今日、AI 研究はさまざまな立場から取り組まれている。たとえば、人間の知能の本質・メカニズムをコンピューターで実現しようという立場（強い AI）や、人間の知能の働き・振る舞いに類する機能をコンピューターで実現しようとする立場（弱い AI）がある。また、人間の知能のさまざまな側面を広くカバーする汎用性の高いシステム（さまざまな状況で人間の知能のように動作するシステム）を目指す立場（汎用 AI）や、特定の機能や特定の状況下でのみ人間に近い（ときには精度で人間を上回る）振る舞いをするシステムを目指す立場（特化型 AI）がある。昨今、第 3 次 AI ブームと言われ、AI 技術がさまざまな実用的な応用に広がっているが、それらは現状、特化型 AI かつ弱い AI に相当する技術群である。

一方、ビッグデータ（Big Data）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・リアルタイム性等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。

これら AI 技術とビッグデータ（データそのもの、および、処理技術）は深く関係し合いながら発展している。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。

[AI・ビッグデータの俯瞰図（時系列）]

AI・ビッグデータの技術発展に関する俯瞰図（時系列）を図 2-1-1 に示す。この図では、横軸が年代、縦軸が取り組みの広がりをおおまかに表している。図中には、その時期に台頭した技術およびエポックをプロットした。AI は現在、3 回目のブームを迎えているが、これまでブームと連動して取り組みが広がってきた様子を示している。

第 1 次 AI ブーム（1950 年代後半から 1960 年代）では、AI に関わる基礎的な概念が提案され、AI が新しい学問分野として立ち上がったが、実用性という面ではまだトイシステムであった。第 2 次 AI ブーム（1980 年代）では、人手で辞書・ルールを構築・活用するアプローチが主流となり、エキスパートシステム、指紋・文字認識、辞書・ルールベース自然言語処理等（カナ漢字変換等）の実用化にも結び付いた。第 3 次 AI ブーム（2000 年代から現在）では、インターネットやコンピューティングパワーの拡大を背景として、ビッグデータ化と機械学習の進化がブームを牽引し、画像認識・音声認識、機械翻訳、囲碁・将棋等では人間に追いつき／上回る性能を示し、さまざまな AI 応用システム（認識・検索・対話システム等）が実用化され、社会に普及している。

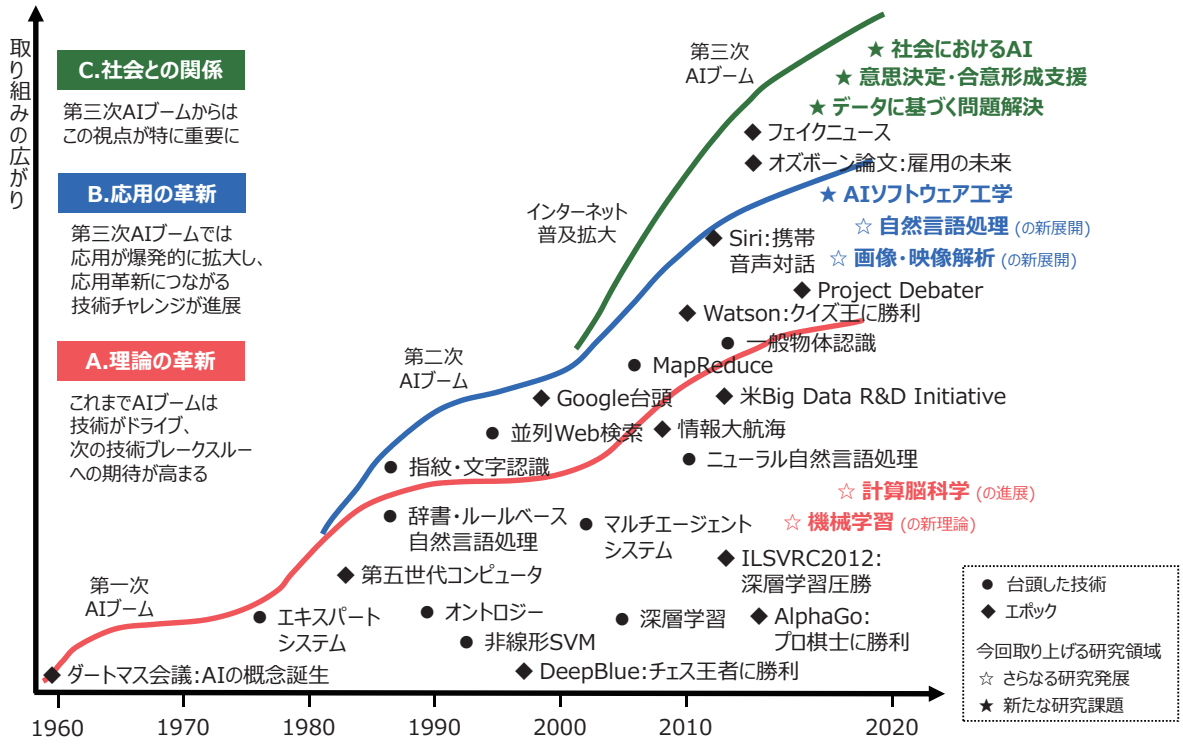


図2-1-1　人工知能・ビッグデータの俯瞰図（時系列）

このような技術発展を図 2-1-1 では 3 つの大きな流れでとらえている。

1 つめの流れは「A. 理論の革新」である（図中の赤ライン）。3 回の AI ブームはいずれも理論面の発展（知識表現・記号処理、辞書・ルールベース処理、機械学習・深層学習等）やコンピューティングパワーの増大等の技術進化によってドライブされた。

2 つめの流れは「B. 応用の革新」である（図中の青ライン）。第 2 次 AI ブーム以降は実用的な応用が生まれ始め、ビッグデータの高速並列処理・知識処理の実用化が進み、第 3 次 AI ブームでは、機械学習の応用分野が爆発的に拡大した。

3 つめの流れは「C. 社会との関係」である（図中の緑ライン）。これは第 3 次 AI ブームを迎えて、活発に議論されるようになった視点である。AI 技術のさまざまな応用が社会に広がったことに加えて、AI 技術の可能性が人間にとって恩恵だけでなく脅威や弊害ももたらし得るという懸念が強まったためである。

このようなこれまでの技術発展を踏まえて、図 2-1-1 にはさらに、今後の発展に関して、特に注目する技術テーマを 8 つ追加した（図の右端の赤字・青字・緑字の項目）。「機械学習」「計算脳科学」「画像・映像解析」「自然言語処理」という 4 つは、これまでも取り組まれてきた技術テーマであるが、理論や応用の革新という流れの中で、さらなる研究発展の勢いが感じられるものである。また、「AI ソフトウェア工学」「データに基づく問題解決」「意思決定・合意形成支援」「社会における AI」という 4 つは、AI の社会との関係や応用の革新といった流れの中で、新たな研究課題として立ち上がりつつあるものである。

[AI・ビッグデータの俯瞰図（構造）]

次に、これら８つの技術テーマを、AI・ビッグデータの技術スタックの中に位置付けた俯瞰図（構造）を、図 2-1-2 に示す。この図では、システムを構成する技術群を「処理基盤層」「処理コンポーネント層」「ソリューション層」に分けている。また、それと分けてもう一つ、「人間の理解、社会の受容」という層を設けている。これは、システムを設計する上で、その利用者となる人間や、それが組み入れられる社会についての理解・モデル化や規範・ガイドラインの議論も必要となるためである。図 2-1-2 には、今回注目する８つの技術テーマのほかにも、この技術スタック中に位置付けられる他の主要な技術テーマも配置している（図中で番号を振っていない薄い黄色の箱の項目）。

図2-1-2　人工知能・ビッグデータの俯瞰図（構造）

８つの技術テーマ（研究開発領域）について、その簡単な定義を以下に示す。それぞれの詳細は後続の節にまとめている。

①機械学習：データの背後に潜む規則性や特異性を発見することにより、人間と同程度あるいはそれ以上の学習能力をコンピューターで実現しようとする技術である。これにより、事象や対象物について、その観測データに基づく分類、予測、異常検知等が可能になる。

②画像・映像解析：カメラ等で撮影された画像や映像をコンピューターで解析して、その画像・映像の内容、つまり、そこに写っているものが何であるか（人や物体、文字等）、その位置や状態、あるいはシーン全体の状況を認識する技術である。

③自然言語処理：コンピューターによって自然言語を処理する技術である。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り、自然言語処理は、人間のコミュ

ニケーション、思考等の知的作業、知識の流通・活用等を支援する。

④ AI ソフトウェア工学：AI 応用システムを、その安全性・信頼性を確保しながら効率よく開発するための新世代のソフトウェア工学を意味する。従来の演繹型システム開発に加えて、機械学習を用いた帰納型システム開発にも対応した開発方法論・技術体系である。

⑤意思決定・合意形成支援：情報爆発等による可能性の見落としやフェイクニュース等による悪意・扇動意図を持った思考誘導操作の問題を、AI を含む情報技術によって軽減し、個人・集団が主体性・納得感を持って意思決定できるように支援する。

⑥データに基づく問題解決：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等による、問題解決手段の質的変化、産業構造・社会システム等の転換を生み出す研究開発領域である。

⑦計算脳科学：脳を情報処理システムとしてとらえて、脳の機能を調べる研究分野である。脳という情報処理システムについて、計算理論、表現・アルゴリズム、ハードウェアの 3 面からの理解を相互に深め、脳の情報処理機能を理解しようとする。

⑧社会における AI：AI 技術が社会に実装されていったときに起こり得る、社会・人間への影響や倫理的・法的・社会的課題（Ethical, Legal and Social Issues：ELSI）を見通し、あるべき姿や解決策の要件・目標を検討し、それを実現するための制度設計および技術開発を行う研究開発領域である。

［AI・ビッグデータの研究開発強化戦略］

以上のような技術発展や研究開発領域の俯瞰を踏まえて、AI・ビッグデータの研究開発強化には、特にソリューション層への戦略的な取り組みが重要だと、JST CRDS では考えている。以下、その背景や理由を述べる。

- 図 2-1-1 俯瞰図（時系列）に示したように、取り組みが理論から応用、さらに社会との関係に広がっており、社会から見た AI・ビッグデータ技術の価値という視点がますます重要になっている。また、社会に対する、より大きな価値や新しい価値の創造、あるいは、脅威・弊害の回避を狙うことが、新たな革新技術への要請（新たな価値を生み出すために求められる技術的要件の明示）につながる。すなわち、図 2-1-2 俯瞰図（構造）におけるソリューション層の戦略強化が、上位層（社会、人間）に対する価値創造と、下位層（処理コンポーネント層、処理基盤層）に対する技術的要請を生む。

- 図 2-1-2 俯瞰図（構造）における処理コンポーネント層や処理基盤層については、国の施策として、人工知能技術戦略会議や、理化学研究所・産業技術総合研究所・情報通信研究機構等での取り組みが強化されている。また、上位層の「社会における AI」についての検討も「人間中心の AI 社会原則検討会議」をはじめ、複数の取り組みが進められてきた。しかしながら、その間をつなぐソリューション層の戦略強化は必ずしも十分でなかった。これに対して、図 2-1-2 におけるソリューション層は、(1)「AI システム」、(2)「人間」、(3)「社会」という 3 視点に対応し、その強化の方向を示している。つまり、
  (1)「AI システム」の視点では、安全で信頼できる AI システムを作るにはどうすればよいかという「AI ソフトウェア工学」の強化、

(2)「人間」の視点では、AIシステムと共存することで人間の思考や意思決定がどのように変化し、それをどう支援すべきかという「意思決定・合意形成支援」の強化、
(3)「社会」の視点では、AIシステムが組み込まれることで社会・産業・学問等がどう変革されるかという「データに基づく問題解決」の強化
という方向性を示した。

・ AI・ビッグデータの研究開発は「理論×データ×ソフトウェア」という3つの掛け合わせが競争力を生む。また、AI・ビッグデータは社会との関係が深くなり、研究者には、情報科学技術だけでなく、人文・社会科学まで含めた幅広い視野が求められるようになってきた。ソリューション層を起点とした研究開発強化は、このような人材の育成にも貢献する。

以上のような考えから、JST CRDSでは2018年に、ソリューション層を起点とした研究開発強化の提言として、「AIソフトウェア工学」と「意思決定・合意形成支援」について戦略プロポーザルを作成した[1]。

[1]「AIソフトウェア工学」に関する戦略プロポーザル：「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月発行）
http://www.jst.go.jp/crds/report/report01/CRDS-FY2018-SP-03.html
「意思決定・合意形成支援」に関する戦略プロポーザル：「複雑社会における意思決定・合意形成を支える情報科学技術」（CRDS-FY2017-SP-03、2018年3月発行）
http://www.jst.go.jp/crds/report/report01/CRDS-FY2017-SP-03.html

２．１．１　機械学習

（１）研究開発領域の定義

機械学習（Machine Learning）は、データの背後に潜む規則性や特異性を発見することにより、人間と同程度あるいはそれ以上の学習能力をコンピューターで実現しようとする技術である。これにより、事象や対象物について、その観測データに基づく分類、予測、異常検知等が可能になる。

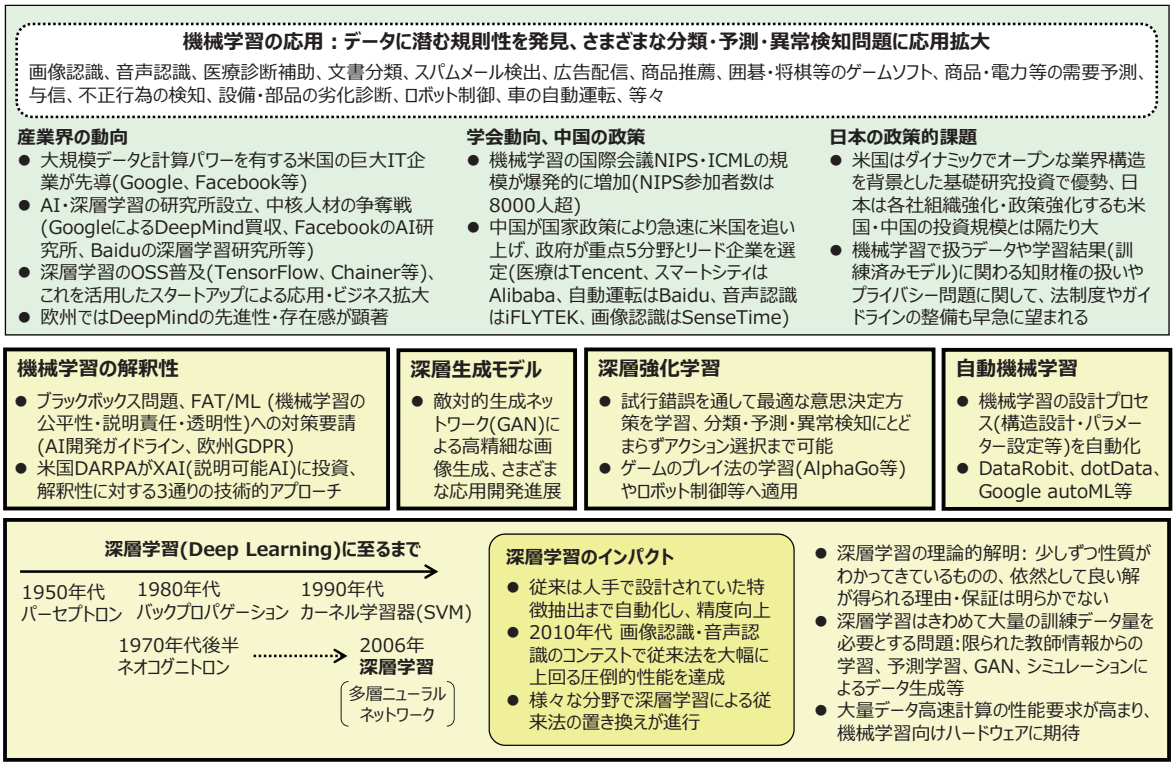


図2-1-3　領域俯瞰：機械学習

（２）キーワード

機械学習、ニューラルネットワーク、深層学習、深層生成モデル、深層強化学習、自動機械学習、ブラックボックス問題、訓練（学習）データ、訓練（学習）済みモデル

（３）研究開発領域の概要

［本領域の意義］

分類、予測、異常検知等の機能を実現したい目的に応じて、機械学習に投入するデータを用意することで、それらの機能が実現可能になる。ビッグデータの時代と言われる今日、さまざまな事象や対象物について大量の観測データが得られるようになり、機械学習技術は幅広い分野でさまざまな目的に利用されるようになった（例えば、画像認識、音声認識、医療診断補助、文書分類、スパムメール検出、広告配信、商品推薦、囲碁・将棋等のゲームソフト、商品・電力等の需要予測、与信、不正行為の検知、設備・部品の劣化診断、ロボット制御、車の自動運

転、等々）。

人工知能（AI）のモデル化は、トップダウンに人間が規則性を記述するアプローチと、ボトムアップに観測データから規則性を発見するアプローチの両面から取り組まれている。1980年代の第2次AIブームのときは前者が盛んに試みられたが、現在の第3次ブームを牽引しているのは後者である。ビッグデータの時代と機械学習技術の進化によって、人間には規則性を見出すのが難しいような現象のモデル化（規則性の記述）が可能になってきた。

［研究開発の動向］

① 技術発展の流れ

機械学習の基本的な処理構成は、訓練ステップ（学習ステップとも呼ばれる）と推論ステップに分かれる。訓練ステップは、訓練データ（学習データとも呼ばれる）を与えて、モデルを作るステップである。ここで作られたモデルは、訓練データの統計的傾向・規則性を表したものになる。推論ステップは、新たに入力されるデータを、訓練済みモデル（学習済みモデルとも呼ばれる）と照合することで、分類・予測・異常検知等の判定結果を出すステップである。訓練データに判定結果に関わるラベルが付与されているケースは教師あり学習、付与されていないケースは教師なし学習と呼ばれる（なお、教師あり学習・教師なし学習とは異なるタイプとして強化学習があるが、これについては後述する）。

機械学習の研究では、訓練ステップのアルゴリズム（学習アルゴリズム）、つまり、訓練データからそこに潜む統計的傾向・規則性をどのようにして見出すかが、一つの重要なポイントになる。モデルを訓練データにフィットさせ過ぎると、推論ステップで与えられるデータに対して必ずしも高い精度が得られないという問題（過学習と呼ばれる）も生じるため、汎化が適切に行われるような仕掛けが必要である。学習アルゴリズムの研究では、統計解析の手法とともに、人間の脳神経回路にヒントを得たニューラルネットワークを用いた手法が注目されるようになった。

このような研究は、古くは1950年代にパーセプトロン[1)]と呼ばれる単純なニューラルネットワークモデルが提案され、任意の線形分離関数を学習できることから1960年代に一大ブームを巻き起こした。しかし、単純なパーセプトロンでは排他的論理和（eXclusive OR: XOR）のような入り組んだ関数を学習できないことが明らかになり[2)]、1970年代には関連する研究は下火になった。

しかし、1980年代に入って、バックプロパゲーション（Backpropagation、誤差逆伝播法）[3)]と呼ばれる、階層構造を持つニューラルネットワークモデルの各ノードの重みを誤差逆伝播によって計算する勾配学習アルゴリズムが提案され、第2次ニューラルネットワークブームが到来した。ニューラルネットワークは、画像や音声の認識、ロボットの制御等、さまざまな問題に適用され、優れた性能を示した。しかし一方では、モデルの階層性に起因する非凸性（Non-convexity）のため、学習に非常に時間がかかり、一般に大域的な最適解を求めることができないという弱点が明らかになった。実際、複雑なニューラルネットワークをうまく学習するためには、学習パラメーターの初期値をうまくチューニングするためのノウハウや、学習を高速に実行するためのさまざまなヒューリスティックが必要であり、信頼性の高い学習システムを構築することは極めて困難であった。

そこで1990年代に入り、階層性を持たないサポートベクトルマシン（Support Vector

Machine：SVM）と呼ばれるカーネル学習器が提案された[4]。SVMの学習は凸最適化として定式化されるため、容易に大域的な最適解を求める事ができる。その後、大規模データに対する超高速学習アルゴリズム等の開発を経て、21世紀初頭にはカーネル法を中心とする機械学習の一大ブームが到来した。カーネル法は音声・画像・自然言語等の処理や、生命情報の解析等、さまざまな分野で優れた性能を発揮した。しかし、カーネル法の性能を十分に引き出すためには、特徴抽出に対応するカーネル関数をうまく設計する必要があり、さらなる学習性能の向上のために、特徴抽出も自動化したいというニーズが高まってきた。その後、カーネル関数の学習法が盛んに研究されるようになったが、未だ決定打に欠ける状況である。

カーネル法の全盛期の真っ只中であった2006年に、多層ニューラルネットワークの新しい学習アルゴリズムが発表された[5]。これは、教師なし学習手法によって事前学習した単層ニューラルネットワークを順々に積み上げていくことによって多層ニューラルネットワークを構築するという手法であり、特徴抽出を自動化する新たな可能性を切り開いた。また、カーネル法が全盛だった1980年代後半～1990年代前半にも、畳み込みニューラルネットワーク（Convolutional Neural Network：CNN）[6]と呼ばれる特殊な構造を持つ多層ニューラルネットワークが文字認識の分野で独立して研究されていた。これは、1970年代後半に提案されていたネオコグニトロン[7]と呼ばれる人間の脳の視覚野をモデル化したニューラルネットワークと本質的に同等である。カーネル法が一層の学習器であるのに対して、ニューラルネットワークは多層構造を持つ。そのことから、これらのニューラルネットワーク技術は、深層学習（Deep Learning）と呼ばれるようになった。

深層学習の技術は、2010年代に入ってから、良質かつ大規模な画像データセットであるImageNetを使った画像認識コンペティション（ImageNet Large Scale Visual Recognition Challenge: ILSVRC）等の画像認識や音声認識のコンテストで既存手法を大幅に上回る世界最高性能を達成するに至った。CNNは空間的構造の表現の取り扱いに適しており、その多層化による性能向上のメリットを最大限享受できる学習として、深層残差ネットワーク（Deep Residual Network）と呼ばれる迂回路をもつネットワーク構造が考案された。これは階層を深くしても効率よく学習が行えるという特徴があり、それまでの8倍程度の深さである152もの層を持つ深層残差ネットワークを用いたアルゴリズムが、2015年の画像認識コンテストで圧勝した。一方、時間的構造の表現の取り扱いに適した方式としては、再帰型ニューラルネットワーク（Recurrent Neural Network：RNN）が考案され、自然言語や時系列データのような連続性のあるデータの解析に用いられるようになった。

② コミュニティーの動向

こうした機械学習の基礎研究は、Neural Information Processing Systems Conference（NIPS、NeurIPS[1]）、International Conference on Machine Learning（ICML）等の国際会議で主に議論されている。これらの国際会議は規模が爆発的に大きくなっており、NIPSは、2013年が1,200人、2014年が2,400人、2015年が3,800人、2016年が5,000人、2017年・2018年[2]は8,000人規模の参加者があった。ICMLも参加者数は、2013年が900人、2014

[1] 2018年より略称が「NIPS」から「NeurIPS」に変更された。
[2] NeurIPS 2018は参加登録サイトがオープンしてからわずか11分で満席となった。2017年・2018年とも参加者数が同規模だが、勢いが頭打ちになったというわけではない。

年が1,200人、2015年が1,600人、2016年が3,000人、2017年が2,400人、2018年が5,000人と多数の参加者があり、第3次ニューラルネットワークブーム（同時に第3次AIブーム）の象徴となっている。

最先端の研究開発が進められている一方で、機械学習の基本的なアルゴリズムは、オープンソースソフトウェア（Open Source Software: OSS）として提供され、また、さまざまな商用プラットフォームにも実装され、容易に使い始められる市場環境になっている。OSSとして提供されている深層学習のフレームワークとしては、Torch・PyTorch（Facebook等）、TensorFlow（Google）、Chainer（Preferred Networks）等がよく知られている[8)]。また、企業がデータや問題をネット上で公開して、世界中の多数の人々に解かせる場（Kaggle等）も生まれ、さまざまなデータ分析問題で解法・精度が広く競われるようになってきた。このように道具立てとして、機械学習・深層学習のコモディティ化が進んでいる。

③ 主要国での取り組み

機械学習の研究開発・ビジネスは、米国が規模・質ともに世界をリードしている。AI分野の国家戦略・投資では、歴史的に米国国防高等研究計画局（Defense Advanced Research Projects Agency：DARPA）が中心的な役割を果たしてきたことに加え、GAFAと呼ばれる巨大IT企業（Google, Amazon, Facebook and Apple）が活発な取り組みを進めている。例えば、機械学習のトップレベル国際会議の一つICMLでは、2016年の採択論文のうち半数近くが米国発の論文であった。

その米国を中国が急激に追い上げており、Association for the Advancement of Artificial Intelligence（AAAI）、International Joint Conferences on Artificial Intelligence（IJCAI）といったトップレベル国際会議での採択論文数でも米国に迫りつつある。中国政府は2017年7月に次世代AI発展計画を発表し、2030年までに理論・技術・応用のすべての分野で世界トップ水準に引き上げ、中国のAI産業を170兆円に成長させるという目標を設定した。これに向けて、11月には次世代AI発展計画推進弁公室を設立し、政府主導で4つの重点AI分野を定め、医療分野はTencent（騰訊）、スマートシティーではAlibaba（阿里巴巴）、自動運転はBaidu（百度）、音声認識はiFLYTEK（科大訊飛）をリード企業として選定した。その後、5つめとして、画像認識でSenseTime（商湯科技）もあげられるようになった。

わが国では、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、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、総務省による情報通信研究機構脳情報通信融合研究センター／データ駆動知能システム研究センターが、政府投資に基づくAIの研究開発を推進している。政府は2016年4月に人工知能技術戦略会議を設立し、3省連携の司令塔機関の役割を持たせた。また、2018年9月に統合イノベーション戦略推進会議のもとで検討されたAI戦略案が発表された。

（4）注目動向

［新展開・技術トピックス］

① 機械学習の解釈性

深層学習を含むニューラルネットワーク系の機械学習はブラックボックス型である。学習結果はノードの重みに反映され、得られた規則性・モデルは人間が直接理解できる形では示され

ない。つまり、分類、予測、異常検知の結果について、理由を人間にわかりやすく説明することができない。

しかし、機械学習から得られた結果について、産業応用における分析担当者から顧客や意思決定者への説明、医療シーンにおける医師による医学的な所見との整合性確認、自動運転・ロボット制御等で事故が起きた場合の原因説明と責任の切り分け、公的機関での利用における公平性担保（人種・性別等によって不平等が生じない）といった要請がある。このような社会的要請から、透明性や説明責任がAI開発ガイドラインにおける重要な要件として論じられるようになってきた。国内では2016年から総務省のAIネットワーク社会推進会議において、AI利活用原則案・AI開発ガイドライン案が議論されている。国際的にも2014年から毎年「機械学習における公平性・説明責任・透明性」ワークショップFAT/ML（Fairness, Accountability, and Transparency in Machine Learning）が開催され、各国での議論も活発化している。欧州で2018年5月に施行された一般データ保護規制（General Data Protection Regulation：GDPR）には、AIの透明性を求める条文（GDPR第22条）が盛り込まれた。

そこで、この機械学習のブラックボックス問題に対する研究開発が活発になってきた[9]。米国DARPAは、説明可能なAI（Explainable Artificial Intelligence：XAI）への研究開発投資プログラムを開始した。このプログラムにおいて、機械学習の精度と解釈性を両立させることに向けたアプローチとして、次の3通りが取り組まれている。1つめはDeep Explanationである。深層学習について、入力の部分的な変化に対するモデル内の反応の変化を調べる等により、結果に特に影響を与えている要素・特徴を推定し、説明の根拠とする方法である。2つめはModel Inductionである。深層学習等のブラックボックス型の機械学習に対して、その入出力の振る舞いを、より解釈性の高い別モデルで近似する方法である。3つめはInterpretable Modelsである。ブラックボックスではない解釈性の高いモデルを用いる方法だが、決定木や重回帰分析のような従来のシンプルなモデルでは解釈性があっても高い精度が得られないため、解釈性がありながら精度も高い新しいモデルを作る。

Deep Explanationは、特に深層学習による画像認識応用において、認識結果が画像のどの部分やどんな特徴に着目して得られたのかを示す形で使われることが多い。また、Model Inductionでは、富士通が深層学習（Deep Tensor）の結果をナレッジグラフで説明する技術を発表し、Interpretable Modelsでは、NECが決定木的な場合分けと重回帰分析式を合わせて最適化する異種混合学習技術（因子化漸近ベイズ推論）を実用化している[10]。

② 深層生成モデル

機械学習でクラス分類を解くための手法には識別モデルと生成モデルがある。識別モデルはデータの属するクラスを同定するが、そのデータがどのように生成されたかは考えない。一方、生成モデルはデータがどのように生成されたか、その過程までモデル化する。深層学習に関して、前述のCNN等は識別モデルであるが、生成モデルにおいても著しい進展があった。特に、2014年にIan J. Goodfellow（当時モントリオール大学、現在Google）が発明した敵対的生成ネットワーク（Generative Adversarial Networks：GAN）[11]は、従来の生成モデルではできなかった高精細な画像を生成できることから、大いに注目され、画像生成を中心に精力的な技術開発と応用展開が進められている。本トピックについては「2.1.2 画像・映像解析」の節で詳しく述べる。

③ 深層強化学習

強化学習（Reinforcement Learning）[12] は、学習主体が、ある状態で、あるアクションを実行すると、ある報酬が得られるタイプの問題を扱う機械学習アルゴリズムである。より多くの報酬が得られるようにアクションを決定する意思決定方策を、アクション実行と報酬の受け取りを重ねながら学習していく。この強化学習では、ある状態で、あるアクションを実行することの良さを表す評価関数を求める必要があるが、この評価関数や方策を深層学習によって学習するのが深層強化学習である。これは試行錯誤を通してアクションを決定するものなので、ゲームのプレイや機器・装置の制御等への適用が広がっている。

Google DeepMind は深層強化学習を用いて、Atari ゲームのプレイ方法をほぼゼロから学習して人間よりも高いスコアを出したり、囲碁で世界トッププロに勝利したり（AlphaGo：詳細を後述）と注目を浴びた。また、Preferred Networks は CES（Consumer Electronics Show）2016 で「ぶつからない車」、CEATEC（Combined Exhibition of Advanced Technologies）2016 でドローン制御等のデモンストレーションを展示し、ロボット等の制御への適用・実用化を進めている。

④ 機械学習の設計自動化（自動機械学習）

OSS 化等、道具立てとして機械学習のコモディティ化が進んだが、その一方で、これらの道具を使いこなせる人材（いわゆるデータサイエンティスト）は依然として不足している。機械学習はその構造設計やパラメーター設定等にノウハウや試行錯誤が必要なことが多く、道具があっても使いこなすのは必ずしも容易ではない。また、大きな計算リソースを必要とする機械学習をクラウドで利用したいというニーズが高まっている。そこで、データと目的を与えれば結果が自動で出てくるような、機械学習の設計自動化、自動機械学習（Automated Machine Learning）の道具・フレームワークを目指した取り組みが活性化し、商用提供も始まった。これによって、データサイエンティスト不足が補われ、産業応用が広がることが期待される。

2012 年に米国ボストンで創業した DataRobot は、データサイエンティストの知識・ノウハウをソフトウェア化し、深層学習、決定木、ナイーブベイズ、SVM 等、100 以上のアルゴリズムから、問題やデータの性質に合う複数アルゴリズムを試し、精度の高い順にモデルを提示する機能を提供している。NEC は 2016 年にリレーショナルデータベースでの大規模データ予測分析プロセス全体を自動化する予測分析自動化技術（特徴量自動設計技術＋予測モデル自動設計技術）を発表し、2017 年にこの技術を製品提供する新会社 dotData を北米に設立した。銀行関連のデータ分析業務への適用事例では、従来 2 ～ 3 か月を要した作業期間が 1 日に短縮できたとのことである。

さらに Google はクラウド上で、Cloud Vision API や Natural Language API 等、訓練（学習）済みモデルによるサービスを提供してきたことに加えて、ユーザーが用意した訓練（学習）データを用いる自動機械学習サービス AutoML を発表した。2018 年 1 月に画像認識（Cloud AutoML Vision）、2018 年 7 月に自然言語処理（Cloud AutoML Natural Language）と機械翻訳（Cloud AutoML Translation）をリリースした。これらは、訓練済みモデルを別ドメインに転用して再学習する転移学習（Transfer Learning）技術と、対象ニューラルネットワークの構造やハイパーパラメーターを別のニューラルネットワークを用いた強化学習によって最

適化探索する NAS（Neural Architecture Search）技術によって実現されている [13]。

［注目すべき国内外のプロジェクト］

① DARPA の XAI プロジェクト [14]

前述の項「機械学習の解釈性」で触れたように米国 DARPA は、説明可能な AI（XAI）への研究開発投資プログラムを推進している（2017 年 5 月～ 2021 年 4 月の 4 年間）。この研究開発課題に対して、Data Analytics と Autonomy という 2 つの具体的なタスクを設定している。Data Analytics タスクは、マルチメディアデータを解析してターゲットを選択するに際して、選択の理由も示す（例えば、敵と判断した理由として銃の輪郭を強調表示する等）。Autonomy タスクは、ドローンやロボット等の自律システムにおいて、どういう状況でどういう理由で次の行動を決定したかを説明する。各タスクとも、説明可能なモデル（前述したように Deep Explanation、Model Induction、Interpretable Models という 3 通りのアプローチがある）だけでなく、説明のためのインターフェースも含んでいる。現在、11 チーム（UC Berkley、CMU、Xerox PARC 等）が採択されて、研究開発に取り組んでいる（1 チームあたりの予算規模は年間 $800K ～ $2M）。また、これら 2 つのタスクのほか、もう一つの研究開発課題として、説明の心理学的モデルが掲げられており、The Institute for Human & Machine Cognition（IHMC）が取り組んでいる。

② Google DeepMind の取り組み：AlphaGo から AlphaZero へ

ゲームを題材とした AI 技術開発は人間のチャンピオンに勝つことを、わかりやすいグランドチャレンジとして発展し、チェスでは IBM の Deep Blue が 1996 年に世界チャンピオンに初の 1 勝、翌年 1997 年には 2 勝 1 敗 3 引き分けで勝利した。将棋では 2010 年に「あから 2010」が女流王将に勝利し、その後、トッププロ棋士との直接対局は実現していないものの、対戦実績を重ねて結果から、2015 年にトッププロ棋士に追いついたとの判断がなされた。

完全情報ゲーム [3] として最も難度が高いとみられた囲碁は、さらに 10 年かかると言われていたが、Google DeepMind の AlphaGo（アルファ碁）が 2016 年～ 2017 年に世界トップランクプロに圧勝した。AlphaGo は、ゲーム AI 分野でよく使われているモンテカルロ木探索に、2 つの多層ニューラルネットワークを組み合わせた [15]。1 つは盤面を評価するための Value Network であり、もう 1 つは手を選択するための Policy Network である。これらは膨大なプロの棋譜からの教師あり学習と、膨大な回数の自己対戦による強化学習を通して訓練された。深層学習・深層強化学習によって、特定目的において人間の能力を超え得るというエポックメイキングな成果となった。

AlphaGo はその後、さらに進化し、2017 年 10 月に AlphaGo Zero というバージョンが発表された [16]。AlphaGo Zero では、Value Network と Policy Network を統合した 1 つの多層ニューラルネットワークを用い、盤面を評価し、次の手を選択する。その訓練にはもはや人間の棋譜は不要で、最初から自己対戦のみで訓練を重ねる。囲碁のルール以外の知識を与えず、自己対戦による強化学習のみを繰り返した結果、40 日後には従来の AlphaGo と 100 戦して

[3] 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。

89勝するまでになった。訓練にデータを必要とせず、自己対戦で成長するということは、ルールを変えるだけで（膨大なデータを用意する必要がなく）他の応用に適用でき、AlphaGo Zeroのアルゴリズムをさらに汎化したAlphaZero[17]は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。サイエンス系の探索問題への応用も検討されている。

なお、以上は完全情報ゲームとして定式化されるタイプの問題を扱っており、完全情報ゲームではないタイプの問題にはそのままでは適用できない。まだ限定的ではあるが、不完全情報ゲームへの取り組みも始まっている。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、AlphaStarというソフトウェアを開発し、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Libratusというソフトウェアが人間に勝利した。

（5）科学技術的課題

［新展開・技術トピックス］にあげた①～④についても、いっそうの研究開発が必要であるが、加えて、以下のような技術的課題への取り組みも重要である。

① 深層学習の理論的解明

カーネル法に代表される従来の凸最適化学習手法では、最適解の一意性が保証されることを用いて、さまざまな統計的理論解析が行われてきた[18]。一方、深層学習は非凸最適化学習手法であり、そもそも得られた解が何らかの誤差を最小にしている保証すらない。そのため、新たな理論解析手法の検討とそれによる深層学習の理論的解明が求められている。

これまでのところ、完全な解明には至ってはいないものの、深層学習の確率的勾配降下法は極小解にはまりにくいことに加えて、汎化性能が高いような幅が広い解を見つけられることや、ミニバッチ正規化やスキップ接続といったテクニックが学習を容易にし、適切なノイズを与えることで汎化性能をあげていること等がわかってきた。また、従来の機械学習の考え方ではパラメーター数が少ない方がモデルの複雑さは下がり、過学習しにくくなるのに対して、深層学習の場合は、モデルが大きければ大きいほど汎化性能は高いことも予想されている[19]。

② 訓練データ量の問題

深層学習は高い精度を得るためには、極めて大量の訓練データが必要である。しかし、通常、データ（しかも教師付き学習のためにはラベル付きデータ）を大量に用意することは容易ではない。一方、人間は学習のためにそれほど大量のデータを必要としていないと考えられており、より少ないデータ量での機械学習は重要な技術課題である。この課題に対してさまざまな取り組みが進められているが、以下に代表的なものをあげる。

・限られた教師情報からの学習

すべての訓練データにラベルを正しく付与することは多大なコストがかかる。そこで、一部のデータに付与されたラベルの情報や、分布の偏りに関する情報をもとに、精度よくモデルを推定する手法が研究されている。その一つが弱教師付き学習のアプローチである[20]。これは、完全な教師情報を含まない限られた情報から学習を行う枠組みであり、正例とラベルなしデータ、信頼度付きの正例、類似度データ対とラベルなしデータ、クラス事前確率の異なる2セットのラベルなしデータから二値分類器を学習できることが示されている。

・予測学習

現状の機械学習技術と人間の学習パターンを比較したとき、今後、教師なし学習に大きな技術発展の余地がある。特に将来予測の問題を扱う予測学習[21]は、常に時間遅れで実際の結果が得られるので、予測と実際の結果の差から学習することが可能である。しかし、対象物に関する知識や物理モデルの統合を含め、理論的な検討はまだ進んでいない。

・GAN

前述の深層生成モデル GAN は教師なし学習であり、ラベル付きのデータを用意する労力を省けるという意味で、訓練データ量の問題に対する一つのアプローチとなり得る。

・シミュレーションによるデータ生成

現実世界の観測データを訓練データに用いる場合、平常時のデータは集めやすいが、事故等の異常時のデータは十分に集めることが難しい。そのため、現実世界では起こりにくいケースも含めて、シミュレーションによって訓練データを生成・補完することで、訓練データのバリエーションや網羅性を高めることが試みられている。

③ 機械学習向けハードウェア

深層学習の性能をさらに向上させるべく、データ量は巨大化し、モデルも複雑化・深層化の傾向にあるため、学習時間は増加の一途をたどっている。そのため、深層学習の計算効率の改善はアルゴリズムだけでなくハードウェア面でも盛んに取り組まれている[22]。NVIDIA は、グラフィックス処理プロセッサ（Graphics Processing Unit：GPU）を応用し、深層学習を高速に行える GPU を開発している。Google は Tensor Processing Unit（TPU）と呼ばれる深層学習に適したプロセッサを開発し、GPU と比べて 10 倍の速さで学習が行えると発表した。これらの新しいハードウェアでは、従来のように浮動小数点計算を倍精度で行わず、単精度や半精度で行うことによって高速な学習を実現している。また、脳型の計算機構を持つ新しいプロセッサの研究も行われている。

（6）その他の課題

① ダイナミックでオープンな業界構造を背景とした基礎研究投資

機械学習を中心とした AI 技術の基礎研究に対して、各国とも国家としての戦略的な投資が盛んであるが、北米では GAFA による基礎研究への大型投資も目立つ。この傾向が生じているのは、GAFA が事業で大きな売上・収益を得て、その一部を将来事業のための先行投資に回した結果ともいえるが、そのような一企業の研究投資方針だけでなく、ダイナミックでオープンな業界構造がそれを促進している面もある。つまり、産学間の人材流動性や活発なベンチャー投資・企業買収等を含むダイナミックでオープンな業界構造が、基礎研究と事業の好循環や勢いのある事業成長を可能にしている。これが AI・機械学習分野の基礎研究投資において、日本が苦戦している要因にもなっている。

その参考のため、北米を中心としたダイナミックな人材の動きをあげる。深層学習は、現在のブームの火付け役であるトロント大学の Geoffrey Hinton、モントリオール大学の Yoshua Bengio、ニューヨーク大学の Yann LeCun、スタンフォード大学の Andrew Ng 等、北米の大学教授が基礎技術開発の中心的な役割を担っている。Hinton らはベンチャー企業 DNNResearch を起こし、それが 2013 年 3 月に Google に買収されると、Hinton も Google

の Distinguished Researcher に就任した。LeCun は 2013 年 12 月に Facebook に新設された AI 研究所の所長に就任した。Baidu は 2013 年 1 月に深層学習研究所を設立し、Ng が 2014 年 5 月に Chief Scientist に就任した。また、Google は 2014 年 1 月にロンドンの深層学習のベンチャー企業 DeepMind を買収した。さらに、企業間連携の動きとして、2015 年 12 月に SpaceX や Tesla Motors の CEO である Elon Musk らを中心に、AI 研究を行う非営利団体 OpenAI が設立された。2016 年 9 月には、Facebook、Amazon、Google、IBM、Microsoft が AI 研究での提携を発表した。

このような動きにおいて遅れている感は否めない日本だが、基礎研究とビジネスを連動させ、外部資金も調達して成長する若いベンチャー企業（CEO が 30 代）が生まれ、勢いを示しつつある。その代表が 2014 年創業の Preferred Networks である。深層学習を中心とした最先端の機械学習技術の研究開発をベースに、交通システム、製造業（ロボット等）、バイオ・ヘルスケアの 3 つを重点事業領域として、トヨタ自動車・ファナック等から計 100 億円を超える資金調達を得ている。また、NEC は 2018 年に、最先端の機械学習研究者が CEO に転じた予測分析自動化技術のベンチャー会社 dotData を米国シリコンバレーに設立した。カーブアウトのスキームを用いて外部資金も調達し、先端技術の育成とビジネス展開を進める。

② 人材獲得・人材育成の課題

勢いのある GAFA 等の企業が、機械学習を専門とする博士学生、ポスドク研究員、さらには大学教授も大量に囲い込もうと躍起になっており、人材獲得競争が熾烈になっている。近年は、金融系や自動車系の企業もこの競争に参入しており、人材不足が深刻な問題となりつつある。

機械学習は米国が強いが、中国やインドは、トップ人材を組織的に米国に送り、彼らが本国に戻って活躍するという流れを作っている。また、中国やインドはシリコンバレーの人材と技術を使いこなすために、拠点を構えている。

さらに、AI・機械学習はアルゴリズムを適切なソフトウェアとして実装してこそ威力を発揮する。日本は人材育成において、理論・アルゴリズムの基礎研究に加えて、ソフトウェア開発力においても強化施策が望まれる。

③ 法制度の整備

機械学習で扱うデータや学習結果（訓練済みモデル）等に関わる知的財産権の扱いやプライバシー問題は、産業応用における重要課題であり、法制度やガイドラインの整備が早急に望まれる。訓練済みモデルの再利用に関しては、単純にモデルをそのままコピーして使うケースだけでなく、追加学習して使うケース、複数の訓練済みモデルを組み合わせて使うアンサンブル（Ensemble）と呼ばれるケース、訓練済みモデルの振る舞い（どんな入力に対してどんな出力を出すか）の観測データを別のニューラルネットワークの学習に用いる蒸留（Distillation）と呼ばれるケース等も考えられ、知的財産権でどう扱うかは課題である [23]。

（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	○	↗	電子情報通信学会情報論的学習理論と機械学習（IBISML）研究会は 700 人を超え、人工知能学会は会員数 5000 人、全国大会参加者数 2600 人を超える等、AI・機械学習のコミュニティーは育っている。しかし、国際トップ会議（ICML、NIPS、AAAI 等）に採択される数は限られている。理化学研究所革新知能統合研究（AIP）センターが機械学習の基礎研究強化を打ち出し、JST CREST、ERATO も含めて上向き。
	応用研究・開発	○	↗	NEC、富士通、日立、パナソニック、NTT、Yahoo Japan!、楽天、リクルート等が AI 分野に積極的な技術開発投資を行っている。特に深層学習に関しては、Preferred Networks がロボット制御で Amazon Picking Challenge 2016 に初出場して 2 位（1 位と同スコア）、トヨタやファナック等の大手との提携による事業強化、ライフサイエンス分野への AI 適用等も進めている。
米国	基礎研究	◎	↗	大学、企業とも機械学習の研究を非常に盛んに行なっており、規模、質ともに世界をリードしている。例えば、機械学習のトップレベル国際会議の一つ ICML では、2016 年の採択論文のうち半数近くが米国発の論文であった。
	応用研究・開発	◎	↗	既に巨大 IT 企業となった Google、Facebook、Microsoft 等が AI・機械学習に積極的投資をしていることに加えて、Airbnb、Uber 等 AI 技術を活用したベンチャー企業が次々と誕生し、国際的に成功を収めている。例えば、Google の収益のほとんどを占める自動広告配信（年間売上 6 兆円）は機械学習技術によるものであり、戦略メッセージを Mobile First から AI First へと切り替えた。DeepMind をはじめ AI・深層学習のベンチャー買収も積極的に実行している。トヨタが AI 研究開発のために設立した TRI も 1000 億円規模の予算を投入すると発表された。
欧州	基礎研究	○	→	英国、ドイツ、フランス、スイス、イタリア、スペイン等の大学や研究機関にて機械学習の基礎研究が盛んに行われている。
	応用研究・開発	○	↗	ロンドンの Google DeepMind、ベルリンの Amazon Machine Learning 等、北米の企業の欧州支社が中心となり、応用研究開発を行っている。特に DeepMind が基礎・応用の両面で存在感を増している。ICML や NIPS 等での採択率もトップクラスである。
中国	基礎研究	○	↗	機械学習の主要な国際会議である ICML を 2014 年に北京でホストする等、当該分野の研究者人口が爆発的に増加している。北米とのコラボレーションも活発である。清華大学・MSRA（Microsoft Research Asia）等を中心に、機械学習の国際会議での中国からの採択数が伸びている（国別で米国に次ぐ）。
	応用研究・開発	◎	↗	Baidu、Huawei Noah's Ark Lab、MSRA、Horizon Robotics 等、企業による応用研究開発が活発に進められている。
韓国	基礎研究	△	→	ソウル大学、KAIST、POSTECH 等の主要大学にて関連の研究は行われているが、国際的に顕著なものは多くない。
	応用研究・開発	△	↗	韓国の大企業が共同で出資して設立した民間の研究所である知能情報技術研究院（AIRI）が 2016 年 10 月に始まり、応用研究のトレンドは上昇しつつある。また、Samsung 等も応用研究開発に力を入れている。

（8）参考文献

1）Frank Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain", *Psychological Review* Vol. 65, No. 6 (1958), pp. 386-408.

2）Marvin Minsky and Seymour A. Papert, *Perceptron: An Introduction to Computational Geometry* (MIT Press, Massachusetts, 1969).

3）David E. Rumelhart, Geoffrey E. Hinton and Ronald J. Williams, "Learning representations by back-propagating errors", *Nature* Vol. 323, No. 6088 (09 October 1986), pp. 533–536. DOI: 10.1038/323533a0

4）Bernhard E. Boser, Isabelle M. Guyon and Vladimir N. Vapnik, "A training algo-rithm for optimal margin classifiers", *Proceedings of the fifth annual workshop on Computational learning theory* (Pittsburgh, July 27-29, 1992), pp. 144-152. DOI: 10.1145/130385.130401

5）Geoffrey E. Hinton and Ruslan R. Salakhutdinov, "Reducing the dimensionality of data with neural networks", *Science* Vol. 313, Issue 5786 (28 July 2006), pp. 504-507.

DOI: 10.1126/science.1127647

6）Yann LeCun, Bernhard Boser, John S. Denker, Donnie Henderson, Richard E. Howard, Wayne Hubbard and Lawrence D. Jackel, “Backpropagation applied to handwritten zip code recognition”, *Neural Computation* Vol. 1, No. 4 (December 1989), pp. 541-551. DOI: 10.1162/neco.1989.1.4.541

7）Kunihiko Fukushima, “Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position”, *Biological Cybernetics* Vol. 36, Issue 4 (April 1980), pp. 193–202. DOI: 10.1007/BF00344251

8）岡野原大輔、「ディープラーニングフレームワーク、より簡単に、高速に、柔軟に」、『日経 Robotics』2016 年 6 月号（2016 年）、pp. 36-37。

9）Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Dino Pedreschi and Fosca Giannotti, “A Survey of Methods for Explaining Black Box Models”, arXiv:1802.01933 (2018).

10）高野敦、「もうブラックボックスじゃない、根拠を示して AI の用途拡大」、『日経エレクトロニクス』2018 年 9 月号（2018 年）、pp. 53-58。

11）Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville and Yoshua Bengio, “Generative Adversarial Networks”, arXiv:1406.2661 (2014).

12）岡野原大輔、「フィードバックから行動を獲得する強化学習」、『日経 Robotics』2015 年 11 月号（2015 年）、pp. 30-31。

13）森正和、「最新 AI 技術「NAS」（Neural Architecture Search）を紐解く ～ Google Cloud AutoML の裏側を見ていく～」、『Developers Summit FUKUOKA 2018』（福岡、2018 年 9 月 6 日）。

14）David Gunning, “Explainable Artificial Intelligence (XAI)”, DARPA Program Update (November 2017). https://www.darpa.mil/attachments/XAIProgramUpdate.pdf (accessed 2019-01-07)

15）David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Pan-neershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel and Demis Hassabis, “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

16）David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel and Demis Hassabis, “Mastering the game of Go without human knowledge”, *Nature* Vol. 550, No. 7676 (2017) pp. 354-359. DOI 10.1038/nature24270

17）David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Pan-neershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya

Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel and Demis Hassabis, “Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm”, arXiv:1712.01815 (2017).

18）Vladimir N. Vapnik, *Statistical Learning Theory* (Wiley, New York, 1998).

19）岡野原大輔、「ニューラルネットの逆襲から 5 年後」、Preferred Research Blog、2017 年 11 月 28 日。 https://research.preferred.jp/2017/11/deeplearning-5years-later/ (accessed 2019-01-07)

20）Takashi Ishida, Gang Niu and Masashi Sugiyama, “Binary classification from positive-confidence data”, *Proceedings of 32nd Conference on Neural Information Processing Systems* (Montréal, Canada, December 2-8, 2018).

21）岡野原大輔、「予測学習：Predictive Learning」、『日経 Robotics』2017 年 11 月号（2017 年）、pp. 36-37。

22）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：革新的コンピューティング ～計算ドメイン志向による基盤技術の創出～』、CRDS-FY2017-SP-02（2018 年 3 月）。

23）丸山宏・城戸隆、「機械学習工学へのいざない」、『人工知能』（人工知能学会誌）33 巻 2 号（2018 年 3 月）、pp. 124-131。

2.1.2　画像・映像解析

（1）研究開発領域の定義

カメラ等で撮影された画像や映像をコンピュータで解析して、その画像・映像の内容、つまり、そこに写っているものが何であるか（人や物体、文字等）、その位置や状態、あるいはシーン全体の状況を認識する技術である。

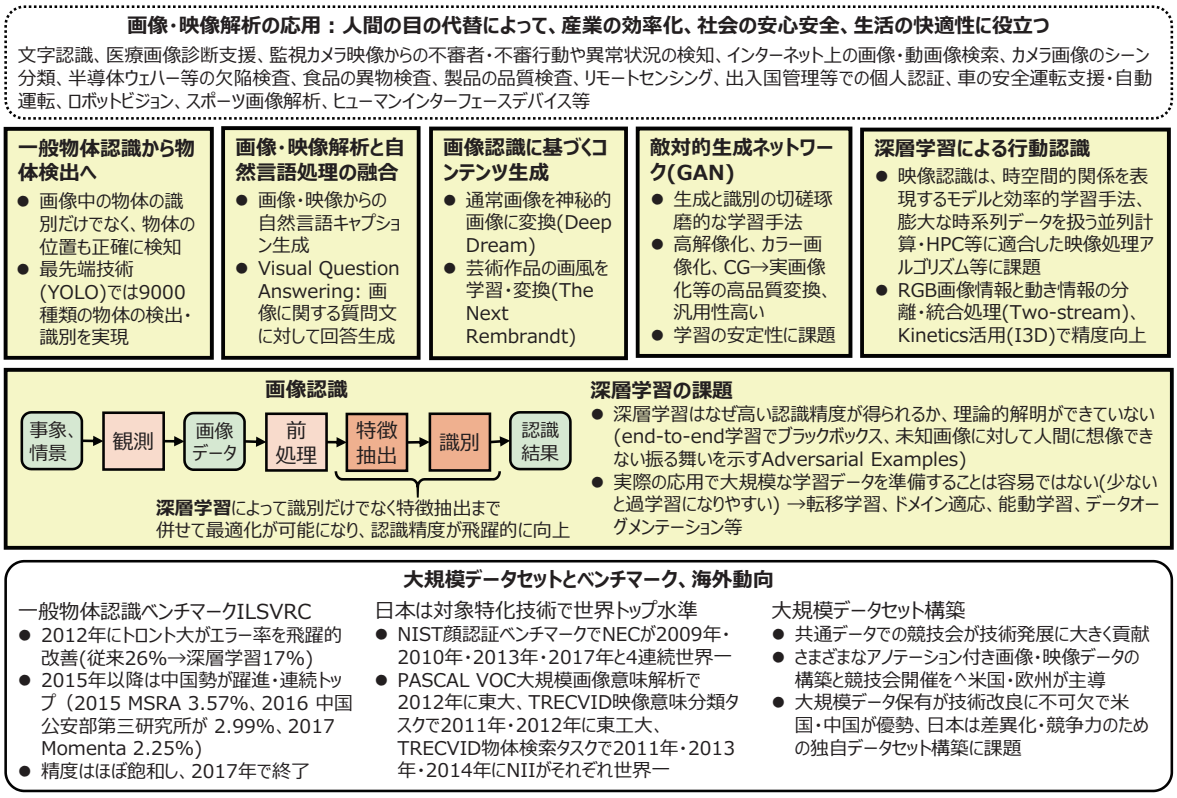


図2-1-4　領域俯瞰：画像・映像解析

（2）キーワード

画像認識、映像認識、パターン認識、メディア処理、一般物体認識、物体検出、顔認証、行動認識、深層学習、ニューラルネットワーク、敵対的生成ネットワーク、Adversarial Examples、大規模データセット

（3）研究開発領域の概要

[本領域の意義]

人間の目の代替となり、産業の効率化、社会の安全安心、生活の快適性に役立つ技術である。人間が行っている目視作業の自動化といった単なる省力化としての価値だけでなく、ヒューマンエラーを低減する判断・診断の支援や、人間では処理しきれない大量な画像・映像データの高速処理等、これまで得られなかった新たな価値を提供する。具体的な応用先としては、郵便区分機等での文字認識、マンモグラフィ等の医療画像診断支援、監視カメラ映像からの不審者・不審行動や異常状況の検知、インターネット上の画像・動画像検索、カメラ画像のシーン分類、

半導体ウェハーやフォトマスク等の欠陥検査、食品の異物検査、製品の品質検査、衛星画像等のリモートセンシング、出入国管理等での顔や指紋を用いた個人認証、自動車の安全運転支援や自動運転、ロボットビジョン、スポーツ画像解析、動作認識によるヒューマンインターフェースデバイス等があげられる。

[研究開発の動向]

① 画像・映像のパターン認識

画像・映像解析はパターン認識を画像・映像に適用するのが基本的な実現形になる。そのパターン認識処理は、観測、前処理、特徴抽出、識別から構成される[1)]。

観測は、実世界の事象を処理可能なデータに変換する処理である。画像・映像のデジタルデータは可視光カメラから得るのが一般的であるが、カメラから被写体までの距離情報を表す深度画像（デプス画像）、赤外線カメラ画像、マルチスペクトル画像等も用いられる。実世界は3次元立体であるが、カメラで撮影されるデータは2次元平面のため、被写体の姿勢変動や照明変動の影響で、被写体の見えが大きく変化する。また、隠れ（オクルージョン）によって情報の欠落も生じる。これらの変動を受けにくいセンシングをどう設計するかが、観測の課題である。文字認識やウェハー欠陥検査等は、被写体自身が2次元平面のため、姿勢変動や照明変動の影響を受けにくく、産業を効率化する技術として早期に実用化されている。一方、情景中の3次元オブジェクトは、姿勢や照明の変動を受けやすいため課題も多い。そのため、エラー発生にシビアな車の自動運転では、前方障害物のセンシングにレーザーレンジセンサー等の測域センサーも用いられている。

前処理は、以降の処理にかかる演算量を軽減するための処理であり、具体的にはデータの正規化やノイズの除去が行われる。画像からの対象領域の切り出しや白色化と呼ばれる信号の大きさの正規化、対象の姿勢の正規化、不明瞭な領域のS/N比（信号雑音比）を改善する鮮鋭化や霧等を除去するデヘイズ処理、あるいは画像の解像度を上げる超解像処理等も含まれる。これら画質を改善する処理は、視認性を高めることを目的に利用されることもある。

特徴抽出は、前処理後の画像・映像から識別に有効な特徴を抽出する処理である。主に局所フィルタを用いて、エッジやコーナー等の画像特徴が抽出されることが多い。処理の高速化、あるいは次元の呪いと呼ばれる認識精度の低下を抑えるために、得られた多数の特徴を少数個に低次元化する処理も含まれる。具体的には、識別に有効な特徴の組み合わせを選ぶ特徴選択や、識別に有効な特徴に作り直す線形あるいは非線形な特徴変換が行われる。

識別は、特徴抽出で得られた特徴をもとに、予め設定したクラス（あるいはカテゴリー）に分類する処理である。特徴抽出で得られた多数の特徴値を多次元特徴ベクトルとみなし、クラスごとに予め設定した代表ベクトルとの内積や距離に基づいて識別する方法や、特徴抽出で得られた特徴点の集合と、クラスごとに予め設定した特徴点の集合のマッチングに基づいて識別する方法等がある。クラスは目的に応じて人間が設定するものであり、たとえば人物と車両を識別する場合は、それぞれが1つのクラスとして設定される。一方、顔認証では、人物一人一人を識別する必要があるため、それぞれが1つのクラスとして設定される。

② 深層学習を中心とした最近の動向

パターン認識で高い精度を実現するには、観測から識別までの全体処理を最適設計すること

が重要である。観測と前処理は、専門家がこれまでの経験に基づき、目的に応じて設計している。識別は、計算機処理の発展とともに機械学習の技術開発が進み、テンプレートマッチングと呼ばれる単純な手法から機械学習で自動設計する手法に既に移行した。一方、特徴抽出は、従来から専門家が経験に基づき設計するのが一般的であったが、2010 年頃からの深層学習 [2] の進展に伴い、特徴抽出と識別を合わせて機械学習で自動設計できるようになってきた。

深層学習（Deep Learning）とは、深い層構造からなるニューラルネットモデル Deep Neural Network（DNN）の学習方法である。2010 年頃までは AutoEncoder や Restricted Boltzmann Machine といった教師なし学習に基づく手法が研究されていたが、現在では後述する CNN をベースとした教師あり学習に基づく手法が主流となっている。深層学習が登場した背景として、(1) インターネットや Social Networking Service（SNS）の普及によってインターネット上に大量の画像がアップロードされ、かつクローリング等で容易にダウンロードできるようになったこと、(2) クラウドソーシングによって安価に画像の正解付けができるようになったこと、(3) General-purpose computing on graphics processing units（GPGPU）を含む計算機性能の向上により、ネットワーク構造の決定に必要な多くの試行錯誤が可能になったこと、があげられる。

現在主流となっているモデルは、福島邦彦が 1979 年に発表したネオコグニトロン [3] がもとになっている。畳み込み層（S 層）とプーリング層（C 層）の組を複数積み重ねることで、パターンの局所変動に頑健になることが示されていた。それを誤差逆伝播学習法で最適化できるようにしたのが、1989 年に Yann LeCun が発表した畳み込みニューラルネットワーク Convolutional Neural Network（CNN）である。MNIST 等の手書き数字データベースで有効性が示されていたが、ネットワーク構造が深くなるほど伝播される誤差が小さくなり、学習が進まなくなる問題が指摘されていた。トロント大学教授の Geoffrey Hinton らは、誤差が深い層まで伝播するように活性化関数や正則化を工夫することでこの問題を解決し [4]、CNN をさらに深くしたニューラルネット構造を用いて、2012 年に ImageNet を使った画像認識コンペティション（ImageNet Large Scale Visual Recognition Challenge: ILSVRC）の一般物体認識のベンチマークテストでトップを獲得した。Hinton らは、従来手法がエラー率 26% だったのに対してエラー率 17% と、深層学習の適用によって一気に約 10% もの飛躍的な精度向上を達成したことが、画像認識・機械学習の研究者らに衝撃を与えた。ILSVRC を舞台として深層学習の改良が続いたが [5, 6, 7]、試行錯誤的な改良に終始しており、なぜ深層学習で高い認識精度が得られるかは、理論的にはまだ解明されていない。

深層学習の技術開発を先導しているのは、Google、Microsoft、Facebook 等の米国の巨大 IT 企業であり、膨大な画像と潤沢な計算リソースを保有していることが、その背景にある。また、深層学習がブームになった後、トロント大学の Hinton は 2013 年から Google 兼務、ニューヨーク大学教授の LeCun は 2013 年から Facebook 兼務になる等、人材の囲い込みが今も進んでいる。米国のコンピュータービジョンとパターン認識に関する著名な国際会議（International Conference on Computer Vision and Pattern Recognition: CVPR）で 2016 年のベストペーパーに選ばれた Deep Residual Network[5] を提案した Kaiming He は、2015 年の ILSVRC でトップを獲得した時は Microsoft Research Asia（MSRA）に所属していたが、直後に Facebook に移ったことが話題になった。

中国の Baidu（百度）は、2013 年にシリコンバレーに Institute of Deep Learning を設立し、

Google の AI プロジェクトの元責任者である Andrew Ng をトップに迎える等、深層学習に力を入れている。深層学習を使った自動運転技術の開発では、2017 年 4 月に、業界の垣根を越えた連携を促すオープンソース型のプラットフォーム「アポロ計画」を発表、わずか 1 年余りで自動車メーカーやサプライヤー、半導体メーカー等 100 社以上が集まった。Volvo とも提携し、2020 年以降に自動運転タクシーを実用化するとしている。2014 年 10 月、香港中文大学教授の湯暁鴎によるベンチャー企業 SenseTime（商湯）は、深層学習を活用した顔認識技術の研究と開発を手がけるが、2017 年 7 月にシリーズ B ラウンドで 4 億 1,000 万ドル、2018 年 4 月にはシリーズ C ラウンドで 6 億ドルを資金調達する等、大きな話題を呼んだ。映像監視や自動運転向け画像認識技術も手掛け、2017 年 12 月には本田技研工業と長期共同開発契約を締結している。これら中国企業が台頭する背景には、AI 技術に対する国策があり、中国政府は、2025 年までに AI 産業で世界をリードする目標を掲げている。前述の ILSVRC でも中国勢は世界トップの精度をたたき出しており、実用面で米国を脅かす存在になってきている。

日本では、産業の効率化として古くから文字認識の研究が進められ、郵便番号読み取り区分機を世界で初めて実用化したのを始め、マンモグラフィ等の医用画像スクリーニング、半導体製造の際のウェハーやフォトマスクの欠陥検査等が実用化されている。バイオメトリクス認証としては、虹彩認証、指紋認証、静脈認証、顔認証が実用化されている。静脈認証は世界で初めて日本国内の銀行で採用され、顔認証は 2001 年に発生した米国同時多発テロ以降のボーダーコントロール強化策として、世界各国に導入され始めている。日本の画像認識技術が世界トップ水準にあることは、国際的ベンチマークテストの成績でも示されている。米国国立標準技術研究所（National Institute of Standards and Technology: NIST）の静止画顔認証ベンチマークテストでは、NEC が 2009 年・2010 年・2013 年と 3 連続で世界第一位を獲得してきたが、2017 年に NIST が実施した動画顔認証ベンチマークテスト（Face in Video Evaluation: FIVE）でも圧倒的優位で一位を獲得、4 連続世界一を達成した。ImageNet データを用いた PASCAL Visual Object Classes（PASCAL VOC）における大規模画像意味解析コンペティションで東京大学教授の原田達也らのチームが 2012 年に世界第一位の認識精度を達成、TREC Video Retrieval Evaluation（TRECVID）の映像意味分類（Semantic indexing: SIN）タスクで東京工業大学教授の篠田浩一らのチームが 2011 年・2012 年に世界第一位の認識精度を達成、TRECVID の物体検索（Instance search: INS）タスクで国立情報学研究所教授の佐藤真一らのチームが 2011 年・2013 年・2014 年に世界第一位の検索精度を達成している。

（4）注目動向

［新展開・技術トピックス］

① 一般物体認識から物体検出へ

ILSVRC をベンチマークテストとして、一般文体認識の精度が飛躍的に改善されたが、さらに、画像に映っている物体の識別だけでなく、その位置も正確に検知する物体検出のための深層学習モデルが、近年数多く提案されている [8]。2014 年に、物体の候補領域を抽出する処理と CNN を統合した Regional CNN（R-CNN）[9] が提案されたのをきっかけに、2015 年に Fast R-CNN[10]、Faster R-CNN[11] と高速化が進んだ。2016 年には、画像をグリッドに区切った領域をもとに物体を抽出する Single Shot MultiBox Detector（SSD）[12]、You Only Look Once（YOLO）が提案され、さらなる高速化と高精度化が進んでいる。これまでは十数種類の物体

の検出・識別するモデルが多かったのに対し、2017 年に提案された YOLO 9000[13] は、9000 種類の物体の検出・識別を可能とし、コンピュータービジョン・画像認識のトップカンファレンス（IEEE/CVF International Conference on Computer Vision and Pattern Recognition: CVPR）2017 のベストペーパーに選出された。

② 画像・映像解析と自然言語処理の融合

注目分野として、画像や映像と自然言語処理の融合がある。画像や映像に対して、その内容を自然言語で記述したキャプション（たとえば「赤い服を着た女性が街中で電話をしている」等）を生成するタスクがその一例である [14]。このタスクを実施するためには、静止画像における ImageNet のような高品位のデータセットが必要となるが、画像キャプション生成では Microsoft Common Objects in Context（MSCOCO）がよく利用される。また、画像と自然言語処理を融合した新しい問題設定として Visual Question Answering（VQA）[15] がある。これは、計算機に画像と画像に関する質問を入力し、この質問に回答する知的システムを開発する課題である。

③ 画像認識に基づくコンテンツ生成

画像認識（画像データからクラスへの抽象化）の過程で、ニューラルネットワークの構造中に、画像に含まれるさまざまな意味的な要素が抽出されてくるが、それにグラフィクス系を融合し、意味的な要素を操作して新たな画像を生成するコンテンツ生成技術が進展している。2015 年に Google が DeepDream[16] と呼ばれるシステムを開発し、大きな話題となった。DeepDream は、通常の画像を夢に出てくるような見たこともない神秘的な画像に変換するシステムである。さらに Google は DeepStyle と呼ばれる入力画像の画風を変換するシステムを開発し、人工知能が製作する芸術作品としてメディア等で取り上げられている。Microsoft とデルフト工科大学は Rembrandt（レンブラント）の画風を学習し、3D プリンタを利用して新作を描く The Next Rembrandt プロジェクトを行っている。

④ 敵対的生成ネットワーク

深層学習によって、高精度な識別だけでなく、高精度な生成も可能になった。特に 2014 年に Goodfellow らが提案した Generative Adversarial Nets（GAN）[17] は、生成モデル G と識別モデル D から構成され、D は学習データと G が生成したデータを識別するように学習し、G は D が間違えるように学習する。G と D を切磋琢磨しながら学習させることで、精度のよいデータ生成が行える。2015 年には CNN を適用した Deep Convolutional GAN（DCGAN）[18] が提案され、精度のよいデータ生成が行えるようになったことで、近年ブームになっている。ラベル付きでデータを生成する Conditional GAN[19]、低解像画像を高解像化する Super Resolution GAN（SRGAN）[20]、モノクロ画像をカラー化したり、日中の画像を夜に変えたりする等、画像間の特徴を変換する pix2pix[21] 等、多数のモデルが提案されている。Apple も CG 画像から実画像に変換する SimGAN を発表し、CVPR 2017 のベストペーパーに選出された。Apple 初の論文として話題になった [22]。GAN は高品位なデータ生成モデルとして注目されているが、敵対的なコスト関数を最適化するため、学習の安定性に課題は残る。しかし、二つのモデルを敵対的に学習する枠組みは汎用性が高く、今後も多様な発展が期待される [23]。

⑤ 深層学習による行動認識

深層学習の登場により、人の行動認識等の動画像認識で、多くのモデルが提案されている。2014 年に提案された Spatio-temporal ConvNet は、単に 10 フレームを CNN へ入力するものであったが、2015 年に提案された Two-stream[24] は、見た目を表す RGB 画像を学習するネットワークと、動き情報であるオプティカルフローを学習するネットワークを組み合わせ、動きを直接的に扱うことで高い精度を実現した。各フレームを時間方向に並べ、3 次元の畳み込みとして学習するモデルも提案されたが、モデルのパラメーター数が多く、限られたデータでは精度がでなかった。2017 年に大規模な行動データ Kinetics の登場により、動画像を 3 次元として学習することが可能になり、I3D と呼ばれるモデルが高い精度を達成している [25]。

［注目すべき国内外のプロジェクト］

ILSVRC の終了と後継ベンチマークテスト

2012 年には深層学習が圧倒的な性能を披露し、一般物体認識の世界的ベンチマークテストとして画像認識技術の基本性能向上に大きく貢献してきた ILSVRC が、2017 年で終了となった。2017 年の結果は、上位 5 位までに正解が含まれないエラー率で、中国のベンチャー会社 Momenta が 2.25% で 1 位となった。2016 年は中国公安部第三研究所が 2.99% で 1 位であり、中国勢の躍進が目立つ。2015 年のトップが MSRA の 3.57% だったことを考えると、本ベンチマークテストに対する精度はほぼ飽和状態に達したとみられる。

ILSVRC は終了となったが、大規模な共通データセットを構築し、同じ土俵で適正に競争することが画像認識の技術発展に大きく貢献したことは明白であり、さまざまな課題設定のもとで同様の取り組みが展開されている。そのようなベンチマークテスト／コンペティションのプラットフォームとして活用されているのが Kaggle である。Kaggle は、企業等がベンチマークのためのデータセットを公開し、このベンチマークに対する精度を世界中からの参加者間で競う場である。Kaggle 運営会社は、2017 年 3 月に Google に買収された。

Google は 900 万枚超の画像データセット Open Images V4 を公開した。これは訓練用（学習用）データセットにおいて 14.6 百万個の物体位置矩形、600 種類の物体カテゴリーが与えられており、物体検出ベンチマークで世界最大規模のものである。さらに、ILSVRC 終了後の 2018 年夏に、その後継として、Open Images V4 を用いた Open Images Challenge というコンペティション（ベンチマークテスト）が Kaggle 上で実施された。その物体検出タスクは、世界中から 454 チームの参加があり、1 位は Baidu（スコア 58.657%）、僅差での 2 位は Preferred Networks（スコア 58.634%）という成績であった。

（５）科学技術的課題

［新展開・技術トピックス］であげたトピック・方向性は、さらなる技術開発が必要で、引き続き重点的に取り組むべき科学技術的課題である。以下では、それらを補完・補足する位置付けで、2 つの面から科学技術的課題を述べる。

① 深層学習の課題

画像認識で高い精度を達成し、現在最強力な手法と考えられる深層学習だが、いくつかの課題が指摘されている。（参考:深層学習とその課題については「2.1.1 機械学習」でも論じている）

まず、深層学習は、100 層以上にも及ぶ多層のニューラルネットを用い、パラメーター数が多いことから過学習になりやすい問題を内在している。過学習を軽減し、高い認識精度を得るには、大規模な学習データが必要である。しかしながら、実際の応用で大規模な学習データを準備することは容易ではない。学習データを十分に集められないケースには、(1) データは大量にあるが、正解ラベルがないため学習に使えない、(2) 正解ラベルが付いた類似のデータは大量にあるが、認識したいタスクとは異なるため使えない、(3) 正常データは大量にあるが、異常データは極めて少ない、(4) そもそもデータ自体が少ない、等がある。

(1) に対しては、いかに低コストで正解ラベルを付与するかが課題であり、クラウドソーシングを利用して人件費を削減するやり方の他、精度改善に有効な一部のデータだけに効率良く正解ラベルを付与する能動学習や、一部のラベルありデータを手掛かりにラベルなしデータも学習させる、半教師あり学習が求められる。(2) に対しては、大量の類似データで学習した DNN を初期値として、別タスクでファインチューニングする転移学習、類似データの分布を別タスクのデータ分布に合わせることで類似データを学習に流用するドメイン適応等が求められる。(3) に対しては、2 つのクラスのサンプル数のインバランスを考慮した学習や、異常データを正常データの外れ値とみなして検出する手法が求められる。(4) に対しては、物理シミュレーション等でデータを生成する方法や、パターン変動に関する先見知識を反映してデータを生成する、データオーグメンテーションと呼ばれる方法等が求められる。

また、深層学習はなぜ高い認識精度が得られるか、理論的解明ができていない。深層学習は大量の学習データに基づく End-to-End 学習、つまり、入力と望ましい出力の対からそれらの間の関係性だけを学習する方法で、それを実現するメカニズムはブラックボックスである。画像の意味分類が End-to-End でできるようになったが、これは画像の意味を認識したことになるのかという疑問が生じ、その解明も兼ねて、画像からのキャプション生成や質問応答の研究が進んだ。しかし、これらにおいても End-to-End の方式が高い精度を示し、結局、ブラックボックスであることは変わっていない。さらに、これと表裏一体の問題として、学習時とは異なる未知画像に対して、人間には想像できない振る舞いを示すという問題も明らかになっている。砂の嵐のような奇妙な画像なのにペンギン等と認識されてしまう画像、どう見てもパンダなのにテナガザルと断言されてしまう画像等が生成可能であること（Adversarial Examples）が指摘されている [26), 27)]。

② 映像認識の課題

静止画像認識の近年の進展に対して、映像認識はそれ固有の技術課題があり、技術強化が進められている。たとえば、実映像に対して時空間的な関係性を適切に表現するモデル構築、効率的な学習手法、膨大な時系列の映像データを扱うためのコンピューティング基盤（ストレージや計算パワーの大規模化、並列計算システムに適合した映像処理アルゴリズムの構築等）の検討が必要である。

そのような技術強化と並行して、具体的な映像認識アプリケーションの実用化も進められている。たとえば、国際体操連盟・日本体操協会との共創で富士通が開発した体操競技の採点支援システム [28)] があげられる。これは 3D レーザーセンサーで体操選手の動きを計測し、そのデータに機械学習技術を適用し、ひねりの回転数・角度等に基づいて体操の技と難度を自動判定するシステムである。2019 年 10 月の世界選手権で実用化され、2020 年の東京五輪でも一

部種目で導入される見込みである。また、日本のスタートアップ企業 Vaak は、映像認識技術を用いた万引き防止システムを実用化した。監視カメラ映像から人物を検出し、関節点の位置、歩幅、持ち物の有無、年齢性別等を認識し、その時系列変化から万引きらしい行動を判定する。3000 時間の映像で訓練したシステムを用いた横浜の実店舗での実証実験で、実際に万引き犯の検挙に貢献した。

映像の内容理解・検索の技術発展には NIST の TRECVID が大きく貢献してきた。また、ImageNet のような大規模で高品位なデータセットは映像理解においても重要であることから、さまざまな動画データセットの開発が進められている。たとえば、4800 クラス 800 万本 50 万時間の動画からなる YouTube-8M があげられる。この規模のデータセットになると大学レベルでの実現は困難であり、企業・政府等の力が必要となる。

（6）その他の課題

① 大規模データセット構築とベンチマークテスト

既に述べてきたように、この研究分野では、大規模な共通データセットを構築し、同じ土俵で適正に競争することが技術発展に大きく貢献してきた。一般物体認識ベンチマークテスト ILSVRC では、上位 5 位までに正解が含まれないエラー率が、2011 年に 25.77% だったのが 2017 年には 2.25% まで改善された。NIST の顔認証ベンチマークテストでは、他人受理率を 0.1% とした時の本人棄却率が、1993 年に 79% だったのが 2002 年に 20%、2010 年には 0.3% にまで改善された。なお、ILSVRC では 1000 クラスに対して 120 万枚を学習し、10 万枚で評価される。NIST の顔認証ベンチマークテストでは、160 万枚に及ぶ顔画像で評価されている。企業においては、大規模データセットは技術競争力の源泉として内部に蓄積してきたが、最近、Yahoo! Research の研究者らが、画像・映像合計 1 億にも及ぶデータセット YFCC 100M をさまざまな関連情報付きで公開した。オープイノベーションの流れと推測される。

日本国内でも、かつては電子技術総合研究所（現、産業技術総合研究所）が ETL9 と呼ばれる JIS 第一水準手書き漢字データセットを公開し、特に大学を中心に国内の手書き漢字認識に関する技術開発が大いに進展した経緯がある。しかし、クラス当たり 200 枚の画像であり、今となっては極めて小規模なものである。

分野全体の技術水準向上と分野活性化に共通データセットが大きく貢献してきた一方、技術開発競争では独自のデータセット構築が価値を持つ。NIST の顔認証ベンチマークテストで日本企業が上位成績を収めている背景には、企業各社が独自に構築した顔画像データセットがある。ILSVRC で米国ネット企業が好成績を収めているのも、大量画像データ収集において優位な立場にいることが要因であろう。したがって、この分野で日本の技術力を高めるためには、日本独自の大規模データセットを国が主導して構築し、大学・国研、企業も含めて適正な技術競争ができる土壌を造ることが最も有効である。特に、インターネット上では容易に得られない実世界の画像・映像データにフォーカスし、たとえば RGB-D のような深度情報も含めたデータセット構築が、今後の技術競争力を高める上で重要であろう。また、公共の場でのデータ収集や社会実証行う際の規制緩和も、検討すべき政策の一つである。

② 長期視点での研究育成、人材施策

昨今の画像・映像解析は深層学習一色だが、かつてのニューラルネットワークブームのあと2012年頃まで世界の第一線で活躍する画像認識研究者において深層学習の専門家はごく少数派だった。深層学習の成功は少数の研究者の長期間にわたる地道な努力が実を結んだものであって、流行に便乗することで成功したわけでは決してない。今後また全く新しい計算機アーキテクチャの出現により、深層学習とは異なる新規手法が生み出され、時代遅れとみなされた手法が再度注目される可能性も十分あり得る。そのために、すぐに役立つ技術だけではなく、10年後、20年後に役立つ可能性のある研究を支えていく制度が重要である。

一方で、深層学習を用いた実世界の画像・映像解析は、実社会のさまざまな課題解決に結びつき得る。複雑な実社会の課題分析から技術的・理論的にチャレンジングな研究ターゲットが見いだされ得る。このような課題解決では、画像・映像解析技術だけでなく、社会課題の分析や、音声・自然言語処理やロボティクスとの融合といったより広く複合的なアプローチのできる人材育成が求められる。このような境界領域研究の奨励や博士課程学生への支援制度、その後のキャリアパスの確保等も課題である。さらに、この分野ではグローバルな人材獲得競争が激化しているが、グローバル競争力の面で日本の給与体系は弱く、海外からの優秀な研究者・技術者の招聘による底上げが難しく、日本からの人材流出も懸念される。

③ 顔認識技術の社会的リスクへの対策

近年、顔認識技術がさまざまな応用に急速に広がっている。空港での入出国管理だけでなく、スマートフォンの認証にも使われ、SNSへの投稿写真中の人物タグ付けにも使われるようになった。中国では顔認識機能を有する犯罪者追跡システム「天網」が2000万台規模の監視カメラと連動して稼働している。

顔認識技術は以前からプライバシー保護の面からの懸念が指摘され、堅牢なセキュリティー確保や画像データを保存しない等の対策がとられてきた。しかし、従来は顔認識機能の利用が、そのような対策面で意識の高い大手企業に限られていたのが、裾野が拡大し、幅広い人・企業が簡単に利用できるような状況になりつつある。しかも、プライバシー保護の懸念だけでなく、特定の人種やマイノリティーの人々を差別してしまうリスク（学習データの質・量によっては、そういった人々の認識率が低く、場合によっては犯罪者と誤認識されやすい等）も指摘されている。さらに、顔の微妙な表情から感情追跡が可能になると、人の内面をのぞき込むような使われ方の懸念も生じる。

このような問題に対して、米国のシンクタンクであるAI Now Instituteは報告書[29]をまとめ、ガバナンスや監督・説明責任体制の整備が不十分なまま、顔認識技術が急速に広がりつつあることへの懸念と対策提言を示している。

（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	○	↗	NEC、オムロン、富士通等の企業で取り組まれているバイオメトリクス認証（顔・静脈等）を中心には世界的に競争力がある。大学や国研では一般画像の認識等の研究が進められている。ILSVRC、TRECVID では東京大学、東京工業大学、国立情報学研究所等による世界的にも競争力のある技術が開発されている。産総研、理研や多くの大学で AI 研究センターが立ち上がり、基礎研究を支援する体制が整いつつあるが、現状、国際会議での存在感は米国・中国ほど大きくない。
	応用研究・開発	◎	→	顔認証では NEC が NIST ベンチマークで 4 回続けてトップを獲得し、世界的にも大きな存在感を示している。オムロンの OKAO Vision は世界で 5 億ライセンスを出荷、世界トップレベルのシェアを持つ。静脈認証も日立や富士通が銀行向けに実用化し、国内大手電機メーカーの技術力は世界的に見ても高いものがある。
米国	基礎研究	◎	→	Stanford、UC Berkley、CMU 等の大学で基礎研究が盛んに行われている。大学のみならず巨大 IT 企業 Google、Facebook、Microsoft 等の研究所でも世界中の有能な研究者を集め、基礎研究を推進、この分野を主導。また、基礎研究に必要なデータセットの多くが米国の大学・巨大 IT 企業によって公開されている。研究すべきタスクの設定や研究コミュニティーへの情報発信等でも中心的な役割を果たしている。たとえばスタンフォード大学が一般物体認識向け大規模データベース ImageNet 構築し、深層学習のベンチマークテスト ILSVRC を主導。NIST が顔認証・指紋認証を初めとしてデータセット構築とベンチマークテストを継続的に実施。
	応用研究・開発	◎	↗	Google、Microsoft、Facebook 等の巨大 IT 企業では有能な技術者を全世界から集め、応用研究や開発が盛んに行われ、大学との連携も盛んに行われている。大学でも起業を目指した応用研究や開発も数多く実施されている。深層学習を利用した画像・映像処理のためのツール群を無償で公開し、画像・映像処理用のサービスをクラウド上で実施する等、他企業の応用研究のサポートも実施している。
欧州	基礎研究	◎	↗	オックスフォード大学、ETH、アムステルダム大学、INRIA、Max Planck 等に優秀な研究者が多数在籍、基礎研究力が高い。Google DeepMind、Facebook AI Research、Qualcomm 等の企業の欧州研究部門での基礎研究もインパクトのある成果をあげている。PASCAL VOC プロジェクト（2005 ～ 2012 年）が画像認識研究の底上げに大きく貢献した。ImageCLEF、MediaEval でデータセットを公開、行うべきタスク設定を行っている。
	応用研究・開発	○	↗	FP7 に続く Horizon 2020 が 2014 年に開始され産官学連携のプロジェクトや応用研究・開発が推進されている。上述の企業の研究部門は、応用研究でもインパクトある成果をあげている。DeepMind は、2014 年に Google に買収された。
中国	基礎研究	○	↗	MSRA が ILSVRC 2015 の多部門でトップ成績を出し、注目を浴びた。北京大学・清華大学・香港中文大学等はトップ会議におけるプレゼンスを高めている。欧米の有力研究者と連携で基礎研究の底上げを行っている例が多く見られる。欧米留学が多いことも研究レベルの底上げに寄与。
	応用研究・開発	◎	↗	MSRA、Baidu、アリババグループ等が応用研究・開発向けのソフトウェア公開・サービス提供を始めている（Baidu の PaddlePaddle を提供、アリババグループでは画像・映像処理も行える AI システム ET をクラウド上で実施）。SenseTime が多額の資金調達に成功、顔認証のみならず映像監視・自動運転の技術開発を進めている。
韓国	基礎研究	○	→	トップ会議において KAIST・ソウル大学校等のプレゼンスがある。画像・映像圧縮、コンピュータービジョン分野では顕著な成果がある。
	応用研究・開発	△	↗	サムスン等での応用研究はあるが、日本企業ほどのプレゼンス・影響力はない。韓国政府が人工知能を国家戦略プロジェクトにすえて、2026 年までに AI 専門企業を 1000 社、専門人材を 3600 人育成する計画がある。

（8）参考文献

1）佐藤敦、「安全安心な社会を支える画像認識技術」、『人工知能』（人工知能学会誌）29 巻 5 号（2014 年 9 月）、pp. 448-455。
2）麻生英樹・安田宗樹・前田新一・岡野原大輔・岡谷貴之・久保陽太郎・ボレガラダヌシカ（人工知能学会監修、神嶌敏弘編）、『深層学習―Deep Learning―』（近代科学社、2015 年）。
3）福島邦彦、「位置ずれに影響されないパターン認識機構の神経回路モデル―ネオコグニトロン―」、『電子情報通信学会論文誌 A』J62-A 巻 10 号（1979 年）、pp. 658-665。

4）Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, *Proceedings of the 26th Annual Conference on Neural Information Processing Systems* (NIPS 2012; Lake Tahoe, Nevada, December 3-6, 2012), pp. 1097-1105.
5）Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun, “Deep Residual Learning for Image Recognition”, *Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition* (CVPR 2016; Las Vegas, Nevada, June 26-July 1, 2016), pp. 770-778.
6）Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato and Lior Wolf, “DeepFace: Closing the Gap to Human-Level Performance in Face Verification,” *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition* (CVPR 2014; Co-lumbus, Ohio, June 23-28, 2014), pp. 1701-1708.
7）Florian Schroff, Dmitry Kalenichenko and James Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering”, *Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition* (CVPR 2015; Boston, Massachusetts, June 7-12, 2015), pp. 815-823.
8）原田達也、『画像認識』（講談社、2017）。
9）Ross Girshick, Jeff Donahue, Trevor Darrell and Jitendra Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation”, *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition* (CVPR 2014; Columbus, Ohio, June 24-27, 2014), pp. 580-687.
10）Ross Girshick, “Fast R-CNN”, *Proceedings of the 2015 IEEE International Conference on Computer Vision* (ICCV 2015; Santiago, Chile, December 13-16, 2015), pp. 1440-1448.
11）Shaoqing Ren, Kaiming He, Ross Girshick and Jian Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, *Proceedings of 29th Conference on Neural Information Processing Systems* (NIPS 2015; Montréal, Canada, December 7-12, 2015) pp. 91-99.
12）Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu and Alexander C. Berg, “SSD: Single Shot MultiBox Detector”, *Proceedings of the 14th European Conference on Computer Vision* (ECCV 2015; Amsterdam, The Netherlands, October 11-14, 2016), pp. 21-37.
13）Joseph Redmon and Ali Farhadi, “YOLO9000: Better, Faster, Stronger”, *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition* (CVPR 2017; Honolulu, Hawaii, July 21-26, 2017), pp. 7263-7271.
14）Andrej Karpathy and Li Fei-Fei, “Deep Visual-Semantic Alignments for Generating Image Descriptions”, *Proceedings of 28th IEEE Conference on Computer Vision and Pattern Recognition* (CVPR 2015; Boston, Massachusetts, June 7-12, 2015), pp. 3128-3137.
15）Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C. Lawrence Zitnick, and Devi Parikh, “VQA: Visual Question Answering”, *Proceedings of the 2015 IEEE International Conference on Computer Vision* (ICCV 2015; Santiago, Chile,

December 7-13, 2015), pp. 2425-2433.

16）Alexander Mordvintsev, Christopher Olah and Mike Tyka, “Inceptionism: Going Deeper into Neural Networks”, Google AI Blog (17 June 2015). https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html (accessed 2019-01-07)

17）Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville and Yoshua Bengio, “Generative Adversarial Nets”, *Proceedings of 28th Conference on Neural Information Processing Systems* (NIPS 2014; Montréal, Canada, December 8-13, 2014), pp. 2672-2680.

18）Alec Radford, Luke Metz, and Soumith Chintala, “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks”, arXiv:1511.06434 (2015).

19）Mehdi Mirza and Simon Osindero, “Conditional Generative Adversarial Nets”, arXiv:1411.1784 (2014).

20）Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi, “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network”, arXiv:1609.04802 (2016).

21）Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros, “Im-age-to-Image Translation with Conditional Adversarial Networks”, arXiv:1611.07004 (2016).

22）Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Josh Susskind, Wenda Wang, and Russ Webb, “Learning from Simulated and Unsupervised Images through Adversarial Training”, *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition* (CVPR 2017; Honolulu, Hawaii, July 21-26, 2017), pp. 2242-2251.

23）篠崎隆志、「GAN―敵対的生成ネットワーク―の発展」、『人工知能』（人工知能学会誌）33巻2号（2018年3月）、pp. 181-188。

24）Karen Simonyan and Andrew Zisseman, “Two-Stream Convolutional Networks for Action Recognition in Videos”, *Proceedings of 28th Conference on Neural Information Processing Systems* (NIPS 2014; Montréal, Canada, December 8-13, 2014), pp. 568-576.

25）Joao Carreira, and Andrew Zisserman, “Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset”, *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition* (CVPR 2017; Honolulu, Hawaii, July 21-26, 2017), pp. 4724-4733.

26）Anh Nguyen, Jason Yosinski and Jeff Clune, “Deep Neural Networks are Easily Fooled”, *Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition* (CVPR 2015; Boston, Massachusetts, June 7-12, 2015), pp. 427-436.

27）Ian Goodfellow, Jon Shlens and Christian Szegedy, “Explaining and Har-nessing Adversarial Examples”, *Proceedings of the 3rd International Conference on Learning Representations* (ICLR 2015; San Diego, California, May 7-9, 2015).

28）藤原英則・伊藤健一、「ICTによる体操競技の採点支援と3Dセンシング技術の目指す世界」、『FUJITSU』69巻2号（2018年3月）、pp. 70-76。

29）Meredith Whittaker, Kate Crawford, Roel Dobbe, Genevieve Fried, Elizabeth Kaziunas, Varoon Mathur, Sarah M. West, Rashida Richardson, Jason Schultz and Oscar Schwartz, “AI Now Report 2018”, AI Now Institute (December 2018). https://ainowinstitute.org/AI_Now_2018_Report.pdf (accessed 2019-01-07).

２．１．３　自然言語処理

（１）研究開発領域の定義

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。自然言語処理はコンピューターによって自然言語を処理する技術を意味する。

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等、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・単語等の意味や構造の解釈に曖昧性が生じ得る点、シンボルグラウンディング（記号と実世界における意味をどのようにして結びつけるか）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等が、その処理を難しいものにしている。

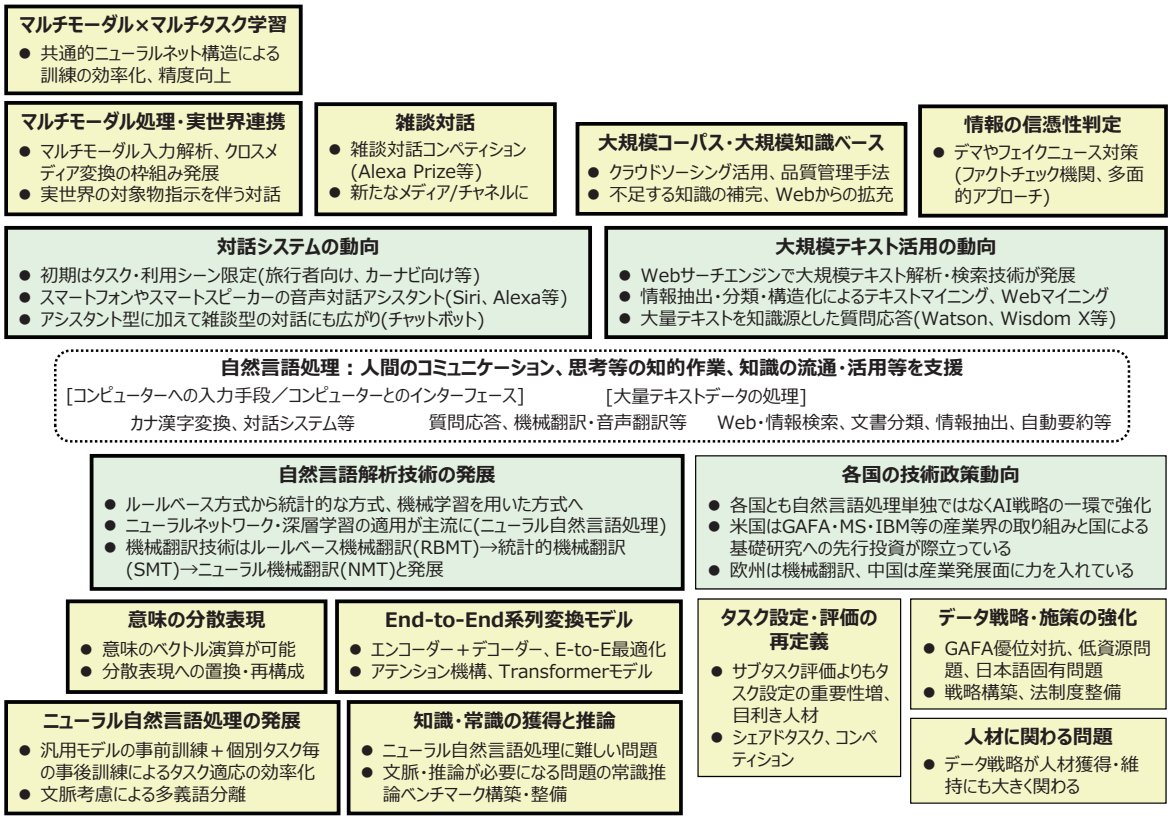


図2-1-5　領域俯瞰：自然言語処理

（２）キーワード

自然言語処理、形態素解析、構文解析、意味解析、機械学習、深層学習、分散表現、質問応答、機械翻訳、対話システム、情報抽出、知識獲得、常識推論

（３）研究開発領域の概要

[本領域の意義]

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は、このような人間のコミュニケーション、思考等の知的作業、知識の流通・活用等を含むさまざまな場面に適用され得る。その代表的な場面・システムのいくつかを以下にあげる。

まず、コンピューターへのテキストの入力手段であるカナ漢字変換システムは、既に多くの人々に使われている自然言語処理システムの一つである。カナ列やアルファベット列を、単語や句といった単位に区切って言葉として解釈することがコンピューターで実現されている。さらに、コンピューターへのテキスト入力だけでなく、コンピューターとのインターフェース全般に自然言語を使う対話システムや質問応答システムがあげられる。具体的には、コンタクトセンターへの問い合わせや、家庭で使われ始めたスマートスピーカー（Amazon Echo、Google Home 等）に対する操作・指示命令を自然言語で行えるといったものである。人間がコンピューターやそのアプリケーションサービスを使う際に、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く、普段使っている言葉で使えるというのが大きなメリットである。コンタクトセンターへの問い合わせのケースでは、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というメリットもある。

また、大量かつ多様なテキストがデジタル化され、インターネット等を介して発信・共有される今日、人間が読むことのできる量をはるかに超えたテキストデータがまわりに溢れている。このような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。既に日常的に利用されるようになった Web サーチエンジンがその代表である。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・自動要約等も自然言語処理のターゲットである。また、上述の質問応答システムは、自然言語を用いたインターフェースであるとともに、大量テキストから容易に情報・知識を得るための手段でもある。

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ターゲットとして長年取り組まれている。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、インターネット等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。

［研究開発の動向］

① 自然言語の解析技術の発展

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、近年は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、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 [1), 2), 3)]。

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）[2), 4)]が考案され、顕著な精度改善を生んでいる。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1つのモデル（ニューラルネットワーク構造）として扱い、End-to-Endの最適化を行う。今日、このEnd-to-End最適化のアプローチが、機械翻訳だけでなく、質問応答、情報要約、画像・映像に対する説明文生成等、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[5)]。

このように、自然言語処理の分野でも、ニューラルネットワーク・深層学習の適用が進んでいるが、画像認識・音声認識の分野ほど、一気に華々しい性能向上が得られたというわけではない。SMTからNMTへ移ったようなパラダイムの転換、処理プロセスの再構成が行われ、その後、使いこなし、チューニングが進んだ結果、この1～2年で徐々に従来方式の性能を上回るようになってきたという状況である。

② 対話システムの動向

対話的インターフェースは、まず、タスクや利用シーンを限定した形で実用化が進んだ。このような限定によって、扱うべき語彙・文型・文脈を現実的な規模に抑えることができる。たとえば、コンタクトセンターでの問い合わせインターフェース、車載端末におけるカーナビゲーションの音声インターフェース、旅行会話の音声翻訳等があげられる。Nuance、IBM、NEC等から法人向けソリューションが提供されたほか、情報通信研究機構（National Institute of Information and Communications Technology：NICT）で開発された個人旅行者向け多言語音声翻訳アプリVoiceTraは、スマートフォンアプリとして、民間企業や、スポーツイベントでの案内、病院診察での外国人対応等に使われ始めており、2020年の東京オリンピックに向けた総務省のグローバルコミュニケーション計画では外国人観光客対応の活用も計画されている。

この間、スマートフォンが普及したことで、そのフロントエンドのインターフェース（音声対話アシスタント）としても、対話システムが使われるようになった。2011年にリリースされたAppleのSiriや2012年にリリースされたNTTドコモの「しゃべってコンシェル」がその代表である。また、2014年には、MicrosoftがWindowsのフロントエンドとして、対話インターフェースを持つCortanaをリリースした。同じ2014年にはAmazon Echo（アシスタントソフトウェアはAlexa）が発売され、家庭向けの音声対話アシスタント内臓スピーカー（スマートスピーカーやAIスピーカーとも呼ばれる）という新しい製品形態が生まれた。同種の製品として、その後、2016年にGoogle Home（アシスタントソフトウェアはGoogle Assistant）、2017年にLINE Wave（アシスタントソフトウェアはClova）等が続いた。さらに、アシスタント型の対話よりも雑談型の対話が中心の「チャットボット」（Chatbot、会話（Chat）とロボット（Bot）を組み合わせた造語）と呼ばれるソフトウェアも利用が広がりつつある。

たとえば、Facebook Messenger や LINE 等のメッセージアプリにおいて、ユーザーの会話の相手の一人としてチャットボットが振る舞うといった使われ方があり、企業にとって新たな顧客接点としても注目されている。

以上は対話システムのアプリケーション形態の広がりを中心に述べたが、技術面は機械翻訳のケースと同じような発展を経ている。Siri の初期のモデルや「しゃべってコンシェル」等は人間が記載したルールと従来型の機械学習の組み合わせであったが、近年はニューラルネットワーク・深層学習の適用が進んでいる。Microsoft の「りんな」等には深層学習による発話選択が組み込まれている。さらに、予め準備された発話を選択するのではなく、発話内容自体も深層学習・強化学習によって獲得する枠組みも研究されている。一方で、学習機能の使い方によっては不適切な対話をしかねないという、ある種の脆弱性が判明し[6]、発話内容を適切にコントロールできるようにすることが課題である。

③ 大規模テキスト活用の動向

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等が進み、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。Web 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や、ユーザーの嗜好や目的に応じた適合ページの選別法等、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、圧縮・索引処理等の技術が急速に発展した[7]。

また、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を解析し、企業活動・意思決定等に活用するテキストマイニング（テキストアナリティクスともいわれる）や Web マイニングと呼ばれるアプリケーション（および技術）も開発が進んだ[7, 8]。商品やサービスに関する評判・意見をインターネット上のサイトから集めて、肯定的か否定的かを自動判定し、マーケティングや風評・リスク対策に活用するためのシステム、論文情報・求人情報等の特定情報のみを選択的に収集・分類・構造化し、傾向把握や注目情報への迅速なアクセスを可能にするシステム等がその例である。

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[9] があげられる。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、NICT が開発・公開した WISDOM X は大規模な 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。また、WISDOM X の技術を応用した対災害 SNS 情報分析システム DISAANA、災害状況要約システム D-SUMM も開発された。災害時には、多様な災害関連テキスト（SNS、ニュース、報告書等）がバースト的に流通するが、厳しい時間制約・人資源制約の下、それら

を精査・統合して適切なアクションを決定することは非常に難しい。DISAANA は 2016 年の熊本地震の際に政府で活用され、自然言語処理技術が災害時の迅速な意思決定にも活用可能であることを示した。

④ 技術政策動向

昨今、各国とも自然言語処理単独の技術政策ではなく、人工知能（AI）分野の技術政策の中で自然言語処理技術の強化を図っている。近年、日本の技術政策においても、自然言語処理を前面に打ち出した研究投資プログラムは進められていない。ただ、AI 技術政策の一環における自然言語処理の位置付けは、米中欧の政策それぞれに特徴が見られる。

米国は GAFA（Google, Amazon, Facebook and Apple）や Microsoft、IBM 等、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等、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。2014 年からは DARPA の Big Mechanism プロジェクトが 4,500 万ドル規模の予算、3 年半の予定で推進されており、技術論文から因果関係の情報を抽出し、そのモデル化を行う技術が研究されている。

欧州は多数の国にまたがることから機械翻訳技術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。EU の第 7 次フレームワークプログラム FP7（2007 年～ 2013 年）では、言語解析ツールの相互運用や機械翻訳のプロジェクトが Work Programme として推進されていた。その後継となる Horizon 2020 プロジェクト（2014 年～ 2020 年）でも、機械翻訳、対話、テキスト分析、文生成といったテーマが選ばれている。

中国は国が AI による産業活性化を推進している。中国工業情報化部（工信部）が 2017 年 12 月に「新世代の人工知能（AI）産業発展を促進するための 3 年行動計画（2018 年～ 2020 年）」を発表した。その中で、育成すべき AI 応用製品として、自動車、ロボット、無人機械、医療機器（映像補助診断システム）、画像映像による身分識別システム、自動音声応答装置、翻訳システム、インテリジェント家具という 8 分野を指定した。自動音声応答装置や翻訳システムは自然言語処理技術に直接関わる製品分野である。

（4）注目動向

[新展開・技術トピックス]

① 意味の分散表現

自然言語処理における意味の分散表現[1]は単語・句・文・段落等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[10), 11), 12), 13)]。大量テキストにおける文脈類似性に

[1] 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[13)]。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 1 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。

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形式もベクトル表現だが、各単語に1次元ずつ割り当てるOne-Hotベクトル（N次元ベクトルの要素N個のうち1個だけが値「1」で他の要素はすべて「0」という形式）であった。Bag-of-Words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。たとえば、分散表現を用いると、「king」－「man」＋「woman」＝「queen」のようなベクトル計算が可能である。

この分散表現の導入によって、自然言語処理は急速に再構成が進んだ。つまり、自然言語処理のプロセスを分解して、単語や意味の要素表現を分散表現に置き換える動きが進んだ。分散表現はニューラルネットへの入力形式としても適しており、深層学習を自然言語処理に適用する際の基本的手法になった。さらに、シソーラスや辞書にも分散表現の学習が取り入れられ、意味選択や質問応答等のタスクも分散表現の演算として扱われるようになりつつある。従来の記号処理は厳密な論理演算をベースとした硬いものだったが、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ということができる。

② End-to-End系列変換モデル

前述のように、ニューラル機械翻訳NMTでは、入力原文から翻訳結果の出力までを1つのニューラルネットワーク構造として、End-to-Endの最適化を行っている。このアプローチは、入力系列から出力系列への変換という意味でSeq2Seq（Sequence to Sequence Model）[14]と呼ばれ、機械翻訳以外の系列変換型の応用（自動要約、画像のキャプション生成等）にも広く用いられるようになった。系列変換のためのニューラルネットワーク構造は、エンコーダーとデコーダーを、分散表現（実数ベクトル形式）を中間に置いて連結した形をとり[5]、テキスト入力に対するエンコーダー処理では分散表現への変換が行われる。つまり、

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）

という構成をとる。

デコーダーには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）を用いた生成処理が行われるが、デコーダー処理においてRNN中のどこ部分（特定の単語等）に注目するかを決定するアテンション機構が組み込まれたことが、生成結果の精度向上につながった。このアテンション機構は強力で、さまざまな形で発展・活用されている。RNNとアテンションの組み合わせは、機械翻訳だけでなく音声認識・対話等にも使われ、CNNとアテンションの組み合わせは画像のキャプション生成等にも使われる。さらに、RNNやCNNを使わずにアテンション機構のみで従来よりも高い精度が達成可能なことを示したTransformerモデル（自己アテンション機構を使用）[15]が注目されている。

③ マルチモーダル×マルチタスク学習

一つの試みとして、テキストに対する意味の分散表現は実数ベクトル形式だが、画像・映像・音声等の異なるタイプの入力に対しても同様のベクトル形式に変換することで、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考えられる。さらに、それらの入力に対して行うタスクも共通構造をベースに組み上げることが考えられる。実際に、画像内の対象物の認識、キャプションの生成、音声の認識、4組の言語間での翻訳、文法と構文の解析

といった異なる複数のタスクを、1つのニューラルネットワーク構造で表現して訓練した結果、各タスクを別々のニューラルネットワーク構造で訓練したケースよりも高い精度が得られたことが報告された[16]。共通構造を用いることで、あるタスクでの訓練結果が別のタスクの精度向上にも効果があったということから、深層学習がマルチモーダルな応用場面に広がるだけでなく、効率的な訓練手法にもつながると期待できる。

④ 大規模コーパス・大規模知識ベース

ビッグデータの時代と言われるように、自然言語処理で用いるコーパスや知識ベースも大規模化が進んできたが、そのための管理技術も発展している。

まず、その構築手法として、不特定多数の作業者に振り分けて作業を実施するクラウドソーシング[17]が活用されるようになった。クラウドソーシングには、1つ1つのタスクで見ると非常に安価に実施できる一方、作業者による品質のばらつきが避けられないという問題がある。この品質問題に対して、複数の作業者の結果を統計処理するとか、作業者を評価・選別するとか、管理手法が整備されてきている。また、利用経験を重ねたことで、クラウドソーシングの適用が適するケース、あまり適さないケースも見極められるようになってきた。

次に、質問応答・推論等を支える知識ベースの大規模化と有効活用も進んでいるが、その知識ベースの効率的構築技術や品質管理技術も発展している。知識の関係を補完する技術や、知識の足りない部分をWeb・テキストから拡充する技術等も整備されてきている。

［注目すべき国内外のプロジェクト］

雑談対話コンペティション

目的に着目した対話システムの類型として、タスク指向対話か非タスク指向対話（雑談対話）かという分け方がある。これまで対話システムの開発・実用化の中心は、観光案内、交通案内、チケット予約、スケジュール予約、レストラン検索といったタスク指向対話であったが、近年、雑談対話を行える対話システムの研究開発が活性化している。

それを示す事例として雑談対話コンペティションの開催があげられる。特に知られているのはAmazonが開催するAlexa Prize[18]である。その最終目標は、5段階評価で4.0以上の評価が得られる魅力的な会話を、人間と20分間以上続けることとされている。米国の大学の研究チームが対象で、優勝チームは第1回（2017年）がワシントン大学、第2回（2018年）がカリフォルニア大学デイビス校であった。しかし、いずれも平均対話時間は10分前後、5段階評価で3.1前後と、最終目標にはまだ遠い。他にも、機械学習のトップランク国際会議Annual Conference on Neural Information Processing Systems（NIPS、NeurIPS）でも2017年からThe Conversational Intelligence Challengeが開催されている。国内でも、2015年から対話破綻検出チャレンジが開催されており、2018年には対話システムライブコンペティションが開催された。

ビジネス上の狙いを考えると、Amazonは家庭向けにスマートスピーカーAmazon Echo（Alexa）をリリースしているが、魅力的な雑談対話を長く続けられるようになれば、家庭におけるテレビやネットアクセス等の時間を奪い取り、新たなメディアあるいはチャネルになり得る。雑談対話を通してユーザーの嗜好を獲得し、より的確で自然なレコメンデーションが可能になる。さまざまなタスク指向対話の前段としての位置付けることも可能である。

（５）科学技術的課題

① ニューラル自然言語処理の発展

ニューラル自然言語処理で高い精度を達成するには、大量の学習データが必要だが、さまざまなタスクのそれぞれについて大量の学習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、大量のデータで事前訓練（Pre-training）しておき、それをベースにして個別のタスク毎の事後訓練（Post-training）を行うというアプローチが検討されるようになった。事後学習においては、個別のタスク毎に必要な学習データはそれほど大量でなくてもよい。このような事前訓練の手法として特に注目されたのは、2018 年に提案された OpenAI GPT[19]、ELMo[20]、BERT[21] である。OpenAI GPT は left-to-right 単方向の Transformer モデル、BERT は双方向の Transformer モデル、ELMo は left-to-rigt と right-to-left の連結モデルである。これらのモデルでは、事前訓練のアプローチに、文脈を考慮する仕組みも取り入れられている。従来、Word2Vec では多義語の意味を分離できなかった。たとえば「bank」という単語は「銀行」と「土手」という異なる 2 つの意味を持つが、分散表現において、これら 2 つの意味を分離できないという問題である。この問題に対して、ELMo や BERT は双方向の文脈を考慮した計算モデルを用いることで多義語を分離することを可能にしている。ELMo、さらに BERT は、複数のベンチマークで従来を上回る精度が得られたと報告されている。

また、End-to-End 系列変換モデルでは、入力系列と出力系列の間で言語要素の過不足が生じ得るという問題がある。つまり、Seq2Seq で実数ベクトルを介して出力系列を生成するため、入力系列になかった単語が出力系列に現れたり、その逆で入力系列にあった単語が出力系列で欠落してしまったりすることがある。また、現状、文単位の変換を行っているため、同じ単語が別の文では異なる表現で出力される等、文章全体での可読性の面でも課題がある。

分散表現と End-to-End 系列変換を用いた新しい枠組みで、自然言語処理のさまざまなタスクが再構成されつつあるが、さらなる技術発展が期待される。

② マルチモーダル処理・実世界連結

ニューラル自然言語処理によってマルチモーダル入力解析やクロスメディア変換を行う基本的な枠組みが得られた。たとえば、映像と音声から成る動画の内容解析や、画像や映像に対する自然言語のキャプション生成等が可能になってきた。今後さらなる性能向上やさまざまな応用開発が期待される。

また、対話システムにおける実世界連結もロボット応用等で重要性が増している。その具体的な成果例として、Preferred Networks が CEATEC JAPAN 2018（2018 年 10 月）でデモンストレーション展示を行った「全自動お片付けロボットシステム」が注目された。このシステムでは、部屋の中に散らかったさまざまな物について、片付けロボットに対して「それは赤い箱に片づけて」というような物体移動の命令を言語指示で行うことができる[22]。このような物体移動の命令は、自然言語（指示代名詞も含む）と実世界の物体を連結する仕組みを用いることで処理できる。対話システムの利用シーン拡大には、このような実世界連結の技術の発展が求められる。

③ 知識・常識の獲得と推論

自然言語処理の枠組みとしてニューラル自然言語処理が広がりつつあるが、従来のルールベースや統計ベースの方式でそれなりの精度が得られていた問題の解法を、ニューラル方式に置き換えて精度を上げたというのがこれまでの状況である。文脈理解や常識推論等、従来方式で難しかった問題はニューラル方式を使おうと依然として難しい。また、ニューラル自然言語処理によって、本来は記号レベルの処理である自然言語処理が End-to-End ブラックボックス化している。こういったことから、ニューラル自然言語処理とは異なるアプローチも組み合わせていくことが必要なことは間違いない。特に知識・常識の獲得やそれを用いた推論の問題でその必要性が高い。

常識推論は自然言語処理研究の黎明期から取り組まれてきたが、近年は問題がいくつかの具体的な形で定義され（含意関係認識、ストーリー予測、Adversarial Examples 等）、そのベンチマークが作られるようになった [23]。

含意関係認識（Recognizing Textual Entailment;. RTE）問題は、常識推論の古典的なベンチマークとみなせる。2 つの文 T と H が与えられたとき、T が H を含意するか（T が真のとき、H も真といえるか）を判定するのが RTE 問題である。2006 年から Pascal RTE Challenge コンペティションも開催されるようになった。当初のベンチマーク用データセットは数百事例程度の規模であったが、近年は Stanford Natural Language Inference（SNLI、画像の説明文ドメイン、約 57 万事例）、Multi-NLI（さまざまなドメイン、約 43 万事例）等、クラウドソーシングを活用した大規模データセットが構築されるようになった。

ストーリー予測は、連続する数文が与えられ、その原因（前提）や結果（エンディング）に該当する文を、別に与えられた数文の中から選択する問題である。10 万ストーリー規模のデータセットが構築され、やはりコンペティションも実施されている。

その後、単に規模が大きければよいわけではなく（統計的な傾向を捉えれば結果が出せるようなものではなく）、本当に文脈を捉えたり、推論したりすることが必要になる問題セットを作るべきだという認識が高まった。そこで、Adversarial Examples の考え方を取り入れたベンチマーク構築が検討されるようになった。このようなコンピューターの言語理解能力を評価するタスクは Machine Reading Comprehension（MRC）と呼ばれ [24]、自然言語処理のトップランクの国際会議でワークショップが開催される等、常識推論のベンチマークとして取り組みが活性化している。

④ 情報の信憑性判定

Web や SNS の情報を扱う上で常に問題となるのは情報の信憑性である。2011 年の東日本大震災ではインターネット上に大量のデマが出回った。2016 年の米国の大統領選挙ではフェイクニュースが結果に影響を与えたとも言われる。このような課題を受けて、ファクトチェック機関が立ち上がったり、Fake News Challenge コンペティションが開催されたりと、対策の検討が進められているが、技術のみで解決できる課題ではなく、多面的なアプローチが必要である [25]。

（6）その他の課題

① タスク設定・評価の再定義（コンペティションの設計・推進）

従来は形態素解析・構文解析等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、ニューラル自然言語処理では応用毎の End-to-End 最適化のアプローチがとられるため、サブタスクに注力することのウェイトが下がってきた。このような状況で技術開発を進めるにあたって、適切なタスクを設定し、評価の指標や条件を適切に定めることの重要性が高まっている。

これに関連して、AI 分野ではシェアドタスク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いる。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匙加減を適切にでき、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。

また、雑談対話のように人間を相手にするシステムの評価は難しい。タスク指向対話では目的達成までの対話回数が少ない方が効率的でベターだが、雑談対話では長く続けられた方が良い。ただし、単なる長さだけでなく、魅力的な／楽しい対話という主観的な質の評価を扱わなくてはならない。前述の Alexa Prize 等のコンペティションでもそのような面での検討が進むと期待される。

② データ戦略・施策の強化

深層学習で高い性能を得るためには大量の学習データが必要である。日々大量のデータを集める GAFA が優位な中、欧州では一般データ保護規則（GDPR）を成立させて対抗している。日本は、大量データの収集・整備・共有の仕組み作りをどのように進めるのか、法制度の整備も含めて、まだ十分な戦略が構築できていない。データの収集から学習の実行まで、高い効率・生産性で回す深層学習ファクトリーのような姿を目指した取り組みも必要である。フェアユースの導入も含めた著作権法等の法制度の整備も急務である。

その一方で、必ずしも大量の学習データを集めることができないタイプの問題も、現実には多い。GAFA のようなインターネットサービスからデータを集められるケースを除けば、現実世界のビジネスの現場で、最初から大量の学習データが得られるようなケースはあまりない。また、英語は流通しているデジタルデータの量が圧倒的に多いのに対して、そもそも集め得るデータ量が格段に少ない国の言語（低資源言語と呼ばれる）もある。このような学習データを十分大量に集めることが困難なタイプの問題に対する技術・体制・取り組み方等も考えていく必要がある。

また、テキストデータの場合、国際的な競争力のため英語や多言語対応が必要である一方、日本語固有の問題へも戦略的に取り組んでおくことも忘れてはいけない。

③ 人材に関わる問題

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、サーチエンジン等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GAFA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。つまり、人材の獲得・維持にデータ戦略が大きく関わっている。

（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	○	→	言語処理学会に活気があり、ACL・EMNLP等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、自然言語処理単独の基礎研究のための政府プロジェクトがなく、AI全般の研究振興のなかで相対的に埋没する危険がある。
	応用研究・開発	○	↗	NICTが音声翻訳システムVoiceTra、大規模情報分析システムWISDOM Xや災害関連情報分析システムDISAANA、D-SUMMを開発・公開（一部は商用化）。NEC、NTTドコモ、日本IBM、Softbank、トヨタ等の民間企業でも、含意認識技術、Watson技術（日本語版）の他、対話系の「しゃべってコンシェル」、Pepper、Kirobo mini等の実用化が進んでいる。
米国	基礎研究	◎	→	スタンフォード大学、CMU、MIT、ペンシルバニア大学、ワシントン大学等の有力大学のみならず、Google等の有力企業はそれぞれ強力な研究チームを持っており、基礎研究を遂行している。ACL、EMNLP等の有力な国際会議等での発表の多くは米国発である。
	応用研究・開発	◎	↗	上述した基礎研究が大学教員の移籍等によって、比較的短期間にGoogle、Microsoft、IBM、Facebook、Amazon等の応用研究・開発に回るサイクルが確立している。Amazonの226億ドルを筆頭にGoogle、Microsoft、Apple等100億ドル規模の研究開発費を投入している企業が多数存在する（2018年時点）[26]。自然言語処理・音声対話等のビジネスもさまざまな応用の開発・展開で先行している[27]。
欧州	基礎研究	○	→	EUの第7次フレームワークプログラム（FP7）が2013年で終了、現在は後継プロジェクトとしてHorizon 2020（2014～2020）が約800億ユーロの予算規模で進行中。多国語を公式言語として使用する環境のため、機械翻訳への投資が伝統的に大きいが、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はあるが、米国主導による産業化の側面が強い。一方で欧州言語間の機械翻訳の精度は、人間による翻訳プロセスにおける下訳として使えるレベルにある。
中国	基礎研究	◎	↗	北京大学・清華大学等の有力大学やMicrosoft Research Asia、Baidu等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、WWW等のトップコンファレンスでも中国発の論文が多数発表されている。中国工業情報化部が2017年12月に発表した「新世代の人工知能（AI）産業発展を促進するための3年行動計画（2018年～2020年）」において、育成すべきAI応用製品8分野中に自動音声応答装置と翻訳システムがあげられた。
	応用研究・開発	◎	↗	Microsoftのチャットボット「シャオアイス」、検索エンジン最大手であるBaidu、チャットサービスのテンセント、EC分野のアリババ等、盛んな研究開発投資と、サービスを通して得られる巨大データに背景にしたAI分野全般の実用化能力は脅威である。
韓国	基礎研究	△	→	KAIST、ETRI、KISTI等の有力大学、国研を中心に基礎研究が進められている。ただ、ACL等のトップコ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等とArtificial Intelligence Research Instituteを設立予定。Naver等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。

（8）参考文献

1）Christopher D. Manning, “Computational linguistics and deep learning”, *Computational Linguistics* Vol.41, No.4 (2015), pp.701-707. DOI: 10.1162/COLI_a_00239
2）Minh-Thang Luong, Hieu Pham and Christopher D. Manning, “Effective Ap-proaches 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.
3）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。
4）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
5）情報処理推進機構 AI 白書編集委員会（編）、「自然言語を中心とする記号処理」、『AI 白書 2017』（KADOKAWA、2017 年）、pp.64-78（1.4 節）。
6）Mark Molloy, “Microsoft ‘deeply sorry’ after AI becomes ‘Hitler-loving sex robot’”, The Telegraph, 26 March 2016. https://www.telegraph.co.uk/technology/2016/03/26/microsoft-deeply-sorry-after-ai-becomes-hitler-loving-sex-robot/ (accessed 2019-01-25)
7）Anand Rajaraman and Jeffrey David Ullman, *Mining of Massive Datasets* (Cam-bridge University Press, 2012). （和訳：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）
8）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）。
9）“Special Issue: This is Watson”, *IBM Journal of Research and Development* Vol. 56 issue 3-4 (May-June 2012).
10）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。
11）Maximilian Nickel, Kevin Murphy, Volker Tresp and Evgeniy Gabrilovich, “A Review of Relational Machine Learning for Knowledge Graphs”, *Proceedings of IEEE* Vol. 104, No. 1 (2016), pp.11-33.
12）Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado and Jeffrey Dean, “Dis-tributed Representations of Words and Phrases and Their Compositionality”, *Proceedings of 27th Annual Conference on Neural Information Processing Systems* (NIPS 2013; Nevada, United States, December 5-8, 2013), pp.3111-3119.
13）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、機械学習プロフェッショナルシリーズ、2017 年）。
14）Ilya Sutskever, Oriol Vinyals and Quoc V. Le, “Sequence to Sequence Learning with Neural Networks”, arXiv:1409.3215 (2014).
15）Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser and Illia Polosukhin, “Attention Is All You Need”, arXiv:1706.03762 (2017).
16）Lukasz Kaiser, Aidan N. Gomez, Noam Shazeer, Ashish Vaswani, Niki Parmar, Llion

Jones and Jakob Uszkoreit, “One Model To Learn Them All”, arXiv:1706.05137 (2017).

17）鹿島久嗣・小山聡・馬場雪乃、『ヒューマンコンピュテーションとクラウドソーシング』（講談社、機械学習プロフェッショナルシリーズ、2016 年）。

18）Ashwin Ram, Rohit Prasad, Chandra Khatri, Anu Venkatesh, Raefer Gabriel, Qing Liu, Jeff Nunn, Behnam Hedayatnia, Ming Cheng, Ashish Nagar, Eric King, Kate Bland, Amanda Wartick, Yi Pan, Han Song, Sk Jayadevan, Gene Hwang and Art Pettigrue, “Conversational AI: The Science Behind the Alexa Prize”, arXiv:1801.03604 (2018).

19）Alec Radford, Karthik Narasimhan, Tim Salimans and Ilya Sutskever, “Improving language understanding with unsupervised learning”, OpenAI Technical report, OpenAI Blog (11 June 2018). https://blog.openai.com/language-unsupervised/ (accessed 2019-01-25)

20）Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee and Luke Zettlemoyer, “Deep contextualized word representations”, arXiv:1802.05365 (2018).

21）Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, arXiv:1810.04805 (2018).

22）羽鳥潤・菊池悠太・小林颯介・高橋城志・坪井祐太・海野裕也・Wilson Ko・Jethro TanKo、「実世界におけるインタラクティブな物体指示」、『言語処理学会第 24 回年次大会発表論文集』（2018 年 3 月）、pp.943-946。

23）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。

24）Saku Sugawara, Kentaro Inui, Satoshi Sekine and Akiko Aizawa, “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.

25）笹原和俊、『フェイクニュースを科学する：拡散するデマ、陰謀論、プロパガンダのしくみ』（化学同人、DOJIN 選書、2018 年）。

26）Barry Jaruzelski, Robert Chwalik and Brad Goehle, “WHAT THE TOP INNOVATORS GET RIGHT”, strategy+business (PwC Strategy&), 30 October 2018. https://www.strategy-business.com/feature/What-the-Top-Innovators-Get-Right?gko=e7cf9 (accessed 2019-01-25)

27）八山幸司、「米国における自然言語処理技術と人工知能のコミュニケーションをめぐる動向」、JETRO ニューヨークだより、2016 年 11 月。 https://www.ipa.go.jp/files/000055592.pdf (accessed 2019-01-25)

2.1.4　AIソフトウェア工学

（1）研究開発領域の定義

AIソフトウェア工学は、AI（Artificial Intelligence：人工知能）応用システムを、その安全性・信頼性を確保しながら効率よく開発するための新世代のソフトウェア工学を指す[1)]。

従来型のシステム開発においては、安全性・信頼性を確保し、効率よくシステム開発を行うための技術体系・方法論がソフトウェア工学の中で整備されてきた。ここでいう従来型とは、プログラム（手続き）を書くという演繹型のシステム開発方法を意味する。これに対して、AI応用システムの開発では、データを例示することによる、機械学習を用いた帰納型の開発方法が用いられる。AIソフトウェア工学は、従来の演繹型システム開発のためのソフトウェア工学から、AI応用システム向けの帰納型システム開発にも対応したソフトウェア工学へ拡張した技術体系・方法論である。

なお、AIソフトウェア工学とほぼ等しい用語として、国内では「機械学習工学」[2), 3)]、海外では「Software 2.0」[4), 5)]がよく用いられている。

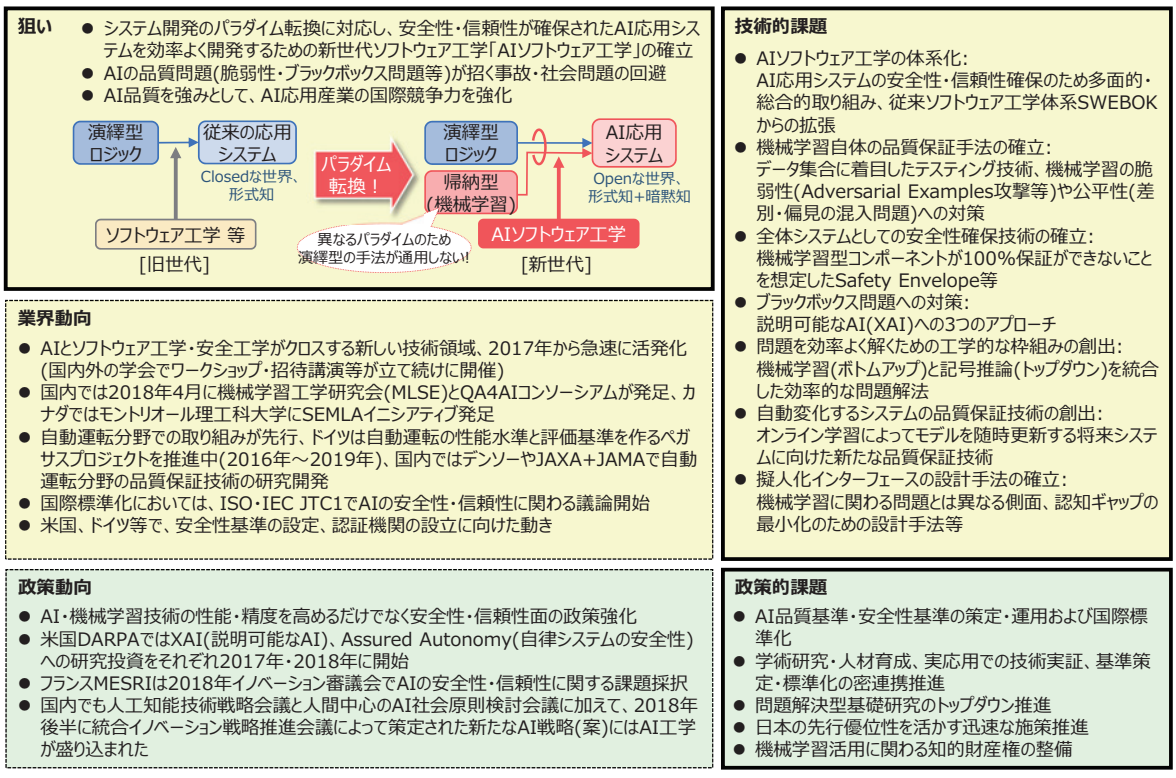


図2-1-6　領域俯瞰：AIソフトウェア工学

（2）キーワード

機械学習工学、ソフトウェア2.0、AI応用システム、ソフトウェア開発、ソフトウェア工学、安全工学、安全性、信頼性、公平性、解釈性、ソフトウェアテスティング、ソフトウェア品質、品質基準、訓練済みモデル、訓練データ、データ品質、データ検査、学習データ工学

（3）研究開発領域の概要

［本領域の意義］

現在のAIブームを牽引しているのは、深層学習（Deep Learning）をはじめとする機械学習技術の進化である。機械学習技術はさまざまな製品・システムに組み込まれ、実社会での応用・実用化が急速に広がっている。

しかし、機械学習による帰納型のシステム開発方法は、従来の開発スタイルとは大きく異なる。そのため、従来ソフトウェア工学として構築・整備されてきた技術・方法論（V字モデル等）は必ずしも適さず、システムの要件定義や動作保証・品質保証にも新しい考え方が必要になる。システム開発のパラダイム転換が起きているのである[1), 2), 3), 4), 5)]。

このパラダイム転換によって、システム開発に必要な人材スキルや方法論が刷新され、この変化に追従できないと、ソフトウェア産業やシステムインタグレーターは競争力を失いかねない。また、動作保証・品質保証等の考え方が整備されないまま、機械学習技術を組み込んだシステムが急激に社会に入っていくと、そこで発生した問題や事故が社会問題化する懸念もある。具体的な例として、画像認識等の誤認識を誘発するAdversarial Examples攻撃[6)]や、機械学習が訓練データ（学習データ）に潜んでいた差別・偏見を反映してしまう問題[7)]等が指摘されている（これらの詳細は後述）。

こういった問題・懸念は、AI応用システムの安全性・信頼性が本質的に低いということを意味するわけではない。パラダイム転換に対して、システム開発のための技術体系・方法論がまだ整備されていないためである。AIソフトウェア工学を確立していくこと[1]が、このような問題・懸念への対策になる。

また、「2.1.1 機械学習」「2.1.2 画像・映像解析」「2.1.3 自然言語処理」で述べてきたように、機械学習技術を用いることで、さまざまな応用において人間の判断を上回る精度の分類・予測・異常検知等が可能になった。これは、人間が形式知化できていないような規則性が機械学習技術によって獲得可能になり、コンピューターによってシステム化・自動化できる機能が広がってきたということである。AIソフトウェア工学は、このような新たな価値を生み出すAI応用システムの安全性・信頼性を確保するとともに、その効率のよい開発を可能にする。

［研究開発の動向］

① 学術界・産業界の動向

AIソフトウェア工学は、AIとソフトウェア工学（Software Engineering）や安全工学（Safety Engineering）がクロスする新しい技術領域である。「AIソフトウェア工学」という名称は2018年12月に発行したJST戦略プロポーザル[1)]で用いた。「AIソフトウェア工学」に相当する「機械学習工学」や「Software 2.0」という名称が研究コミュニティー内で認知されるようになったのも2017年からである[2), 4)]。

それ以前にもシステム開発のパラダイム転換への対応が必要だという指摘・議論はあったが、国内では2017年初頭から学会・業界イベント（2017年2月の情報処理学会ソフトウェアジャパン2017、同年8月の情報処理学会ソフトウェア工学シンポジウムSES2017、他多数）で

[1] ここでは、開発者の視点で書いたが、システムの安全性・信頼性の確保に向けては、システムの利用者や開発依頼者がAI応用システムの安全性・信頼性に関する考え方や特性を理解し、どのように受け入れていくか、という側面も考えていく必要がある。

基調講演や企画セッション等が立て続けに開催され、一気にホットトピック化した[1)]。

さらに2018年4月に、日本ソフトウェア科学会に機械学習工学研究会MLSE（Machine Learning Systems Engineering）が発足した[8)]。MLSEは、機械学習応用システムの開発・運用にまつわる生産性や品質の向上を追求する研究者とエンジニアが、互いの研究やプラクティスを共有し合う場として、研究発表会、ワークショップ、勉強会等、さまざまな活動を展開している。機械学習を用いたシステムの要件定義から設計・開発・運用まで、プロセス管理や開発環境・ツール、テスト・品質保証の手法、プロジェクトマネジメントや組織論も含めて、機械学習を用いたシステム開発全般について幅広いスコープで活動している。

同じ2018年4月には、AIプロダクト品質保証コンソーシアムQA4AI（Consortium of Quality Assurance for Artificial-Intelligence-based products and services）も発足した[8)]。産学27の発起人・団体で設立され、AIプロダクトの品質保証に関する調査・体系化、適用支援・応用、研究開発を推進するとともに、AIプロダクトの品質に対する適切な理解を啓発する活動を行っている。

一方、海外でも2017年後半から、機械学習を従来型プログラミングに対する新しいパラダイムと捉える動きが見られた。新しいパラダイムは「Software 2.0」（Teslaによる）[4)]や「Machine Teaching」（Microsoftによる）[9)]と呼ばれ、国際学会（2018年6月のISCA 2018、2018年12月のNeurIPS 2018等）で基調講演[5)]も行われるようになった。カナダのモントリオール理工科大学にSEMLAイニシアティブ（The Software Engineering for Machine Learning Applications initiative）が発足し、2018年6月にキックオフシンポジウムが開催された。

以上に示したように、国内外とも2017年から2018年にかけて急速に関心が高まり、議論や取り組みが活性化してきた。この背景には、AI・機械学習技術の応用がさまざまな分野に急速に広がり、品質保証クリティカルな応用分野にも適用されるようになってきたことがある。安全性・信頼性に関する要求レベルは応用毎に異なる。応用システムが持つ3つの性質「ミスの深刻性」「AI寄与度」「環境統制困難性」[2]に着目すると、一般に、ミスの深刻性が高く、AI寄与度が高く、環境統制困難性の高い応用ほど、品質保証がクリティカルになる[1)]。機械学習の応用として、商品のレコメンド機能や、文字認識による郵便物の自動読取区分システム等は10年以上前から実用化されているが、これらはミスの深刻性や環境統制困難性が比較的低い応用である。これに対して、昨今注目される自動運転や医療診断といった応用分野は、ミスの深刻性や環境統制困難性が高い（ミスが人命に関わり、多様な環境条件で使われる）。そのため、事前（および運用時）の品質保証が極めて重要なものになっている。

産業界の中でも特に問題意識が高く、検討が先行しているのが自動車業界である。自動運転の実現に向けてAI・機械学習技術の役割が増しており、上でも述べたように品質保証クリティカルな応用分野として具体的な検討が進められている。海外で特に注目されているのは、ドイツ経済エネルギー省（Bundesministerium für Wirtschaft und Energie：BMWi）が主導す

[2]「ミスの深刻性」は、AI・機械学習が誤った判定結果を出したときに生じる問題がどれくらい深刻であるかを意味する。人命に関わるような場合は深刻性が高い。「AI寄与度」は、問題解決のために実行されるアクションの決定にAI・機械学習がどれくらい大きく寄与するかを意味する。AI・機械学習の出力（判定結果）がそのまま反映される場合は寄与度が高く、AI・機械学習の出力（判定結果）を参考にして人間が最終的に判断する場合は寄与度が低い。「環境統制困難性」は、AI・機械学習を実行する際の環境条件をコントロールすることの難しさを意味する。環境条件を列挙することが難しく想定外のことがいろいろ起こり得る場合は困難性が高く、環境条件を統制することが容易であれば困難性が低い。

るペガサス（Pegasus）プロジェクト[10]（2016年1月～2019年6月）である。ペガサスプロジェクトには、自動車関連主要企業が参加し、自動走行システムの期待性能水準と評価基準を明確化するべく取り組んでいる（詳細は後述）。国内では、宇宙航空研究開発機構（JAXA）と日本自動車工業会（JAMA）が連携し、AIによる自律化システムのソフトウェア品質確保の考え方をまとめようとしている[8]。また、デンソーは、自動運転を含むAI搭載システムの品質保証体系（仕組み）の構築を進めるとともに、(1) アプリケーション、(2) サブシステム、(3) コンポーネント（モデル／データ）の3階層で品質技術を開発している[11]。

その一方で、AI・機械学習の研究者は、機械学習の精度・性能を高める競争が激しく、総じて開発法自体への関心は低かった。しかし、社会におけるAIについての議論が活発に行われるようになり、機械学習の脆弱性（Adversarial Examples[6]等）、ブラックボックス問題[12), 13)]への対処等が検討され始めた。たとえば、2014年から毎年開催されているFAT/ML（Fairness, Accountability, and Transparency in Machine Learning）ワークショップ等で、機械学習における公平性・説明責任・透明性等が議論されている（詳細は後述）。これらは開発法そのものの研究ではないが、開発法を設計する上での指針になる。

② 基準策定・標準化の動向

上で述べたような問題意識の急速な高まりと連動して、AI品質関連の標準化活動や安全基準策定に関わる動きが生まれている。国際標準化機構ISO（International Organization for Standardization）と国際電気標準会議IEC（International Electrotechnical Commission）の第一合同技術委員会JTC1において、SC7がソフトウェア工学、SC42が人工知能を扱っている（JTC：Joint Technical Committee、SC：Subcommittee）。これらのSCでAIの品質や安全性・信頼性に関わる議論が始まっている。特にSC7下のWG6がソフトウェア品質、WG26がソフトウェアテスティングを扱っており、さらに自律システムの倫理等を扱うWGも新設される予定である（WG：Working Group）。また、ISO TC159で人間工学、その下のSC4でヒューマンコンピューターインタラクションを扱っており、AIを使う際の利用者視点でのガイドラインを作ろうとしている。IEEE Standard AssociationでもGlobal Initiative for Ethical Considerations in AI and Autonomic Systemsが議論されている。

他にも、米国ベンチャーEdge Case Researchと、米国の認証機関Underwriters Laboratoriesが共同で自律システムの安全性基準を策定しようとしている。Google DeepMindは自社のAI開発ガイドライン[14]を、仕様、頑健性、保証という3面からまとめていると発表した。

ドイツでは、ドイツ人工知能研究センター（The German Research Center for Artificial Intelligence：DFKI）とドイツの認証機関TÜV SÜDが共同で、TÜV for Artificial Intelligence策定の活動を進めている。日本国内では、ソフトウェア品質知識体系SQuBOK（Software Quality Body of Knowledge）策定部会が、2020年に発行予定の次期バージョンV3において、AI応用システムの品質も取り上げる予定である[15), 16)]。

③ 科学技術政策の動向

AIに関する科学技術政策は、いま各国が国としての戦略を掲げ、重点投資を進めている。その中で、AI・機械学習技術の性能・精度を高める技術開発競争が強く意識されてきたが、徐々

に安全性・信頼性の面にも目が向けられるようになってきた。

わが国では、総理指示を受けたAI研究の体制として、2016年に「人工知能技術戦略会議」とその下での総務省・文部科学省・経済産業省の3省連携による推進体制が構築され、AI技術の研究開発が強化されている。一方、内閣府に「人間中心のAI社会原則検討会議」が設けられ、AIの社会的・経済的・倫理的・法的な課題（Ethical, Legal and Social Issues：ELSI）を含む社会から見たAIへの要請が検討されている[3]。ここで次に「AI社会原則が示す姿（満たすことが望ましい要件）を実現するために、AI技術群を用いて、どのようにAI応用システムを作ればよいか」という問いが生じるが、それに答えるのがAIソフトウェア工学である。2018年6月の閣議決定を受けて新たに「統合イノベーション戦略推進会議」が設置され、その下で有識者検討に基づく「AI戦略（案）」が9月末に提言された。この「AI戦略（案）」ではAIソフトウェア工学を含むAI工学が取り組むべき課題として取り上げられている。

米国では、国防高等研究計画局（Defense Advanced Research Projects Agency：DARPA）が推進するさまざまなAI関連プログラムの中で、安全性・信頼性に関わる研究開発も進められている。特に知られているのが2017年に始まったXAI（Explainable AI：説明可能なAI）プロジェクト[17]である。また、2018年に始まったAssured Autonomyプロジェクト[18]においても、自律システムの安全性確保が検討されている（これらのプロジェクトの詳細は後述）。

フランスでは、高等教育・研究・イノベーション省（Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation：MESRI）が、2018年9月にイノベーション審議会によって採択されたAIに関する2つの主要課題を発表した。その主要課題の1つが「AIを活用するシステムの安全確保、認証、信頼性確保の方法」である。

（4）注目動向

［新展開・技術トピックス］

① AIソフトウェア工学の体系化への取り組み

システムの安全性・信頼性の確保は一つの技術で解決するものではなく、全体像を押さえつつバランスよく、技術開発を総合的に取り組んでいく必要がある。このような認識のもと、個々の技術開発と並行して、AIソフトウェア工学のスコープや技術体系が整理・検討されている[1)]。AIソフトウェア工学のスコープ・技術体系を考えるにあたって、従来のソフトウェア工学の体系をもとにすることができる。具体的には、IEEE Computer Societyによってまとめられたソフトウェア工学の知識体系SWEBOK（Software Engineering Body of Knowledge）が参考になる。SWEBOK V3.0では、ソフトウェア要求、ソフトウェア設計、ソフトウェア構築、ソフトウェアテスティング、ソフトウェア保守、ソフトウェア構成管理、ソフトウェアエンジニアリングマネージメント、ソフトウェアエンジニアリングプロセス、ソフトウェアエンジニアリングモデルおよび方法、ソフトウェア品質、ソフトウェアエンジニアリング専門技術者実践規律、ソフトウェアエンジニアリング経済学、計算基礎、数学基礎、エンジニアリング基礎

[3] 社会から見たAIへの要請については、「人間中心のAI社会原則検討会議」に先立ち、内閣府の「人工知能と人間社会に関する懇談会」、総務省の「AIネットワーク社会推進会議」、経済産業省の「AI・データ契約ガイドライン検討会」等で検討されてきた。特に「AIネットワーク社会推進会議」では、透明性、利用者支援、制御可能性、セキュリティー確保、安全保護、プライバシー保護、倫理、アカウンタビリティという8つの原則を掲げた「AI開発原則（案）」や「AI利活用原則（案）」等を示してきた。内閣府は、2018年4月に「人間中心のAI社会原則検討会議」を発足させ、先行の検討結果を踏まえつつ、同年度末までに「人間中心のAI社会原則（案）」をまとめる予定である。

があげられている。AI ソフトウェア工学もこれらを扱うとともに、それぞれの項目において、機械学習型のシステム開発で新たに発生する課題に対応するための拡張が必要である。

また、従来のソフトウェア工学との対比で語られることが多いが、AI・機械学習の応用システム開発は、ソフトウェアだけに閉じず、ハードウェアやデータ管理も含めたシステムが対象スコープになる。狭い意味でのソフトウェア工学に限らず、安全工学やシステム工学も検討範囲に含まれる。具体的な手法として、機能間の関係性を踏まえ、制御系が環境と相互作用することで起きうる事故モデルを使った安全性分析・ハザード解析手法として知られる STAMP/STPA[19), 20)]（System Theoretic Accident Model and Processes / System Theoretic Process. Analysis）や FRAM（Functional Resonance Analysis Method）等を、AI・機械学習の応用システムにも拡張適用する検討が進められている[8)]。

② 機械学習のテスティング技術

機械学習を用いて作られたシステムは、どれだけテストすれば十分なのか、テストの方法や品質指標がまだ確立されていない。従来型の簡単なテストのイメージは、「ボタン A を押したら光る」「ボタン B を押したら音が鳴る」というような動作ロジックに沿って、すべてのケースと正しい結果を事前にリストアップすることができ、その通りの結果が得られるかを確認すればよいというものである。従来型でも、一定以上の規模や複雑さを持つシステムになると、すべてのケースはリストアップできず、テストが難しくなるが、機械学習型の場合は、動作ロジックの記述ではなくデータ例示によって動作を定義するので、そもそもすべてのケースをリストアップするための手がかりがない。たとえば、自動運転における環境認識では、「雨や霧のこともあるかもしれない」「物陰から人が飛び出すかもしれない」等、実世界で車が遭遇し得る環境の可能性をどう数え上げ、どれだけのケースをテストしておいたら十分安全なシステムだと言えるのか、という問題は機械学習型においていっそう難しい。

このことから、事前に想定していなかったケースに対するシステムの振る舞いが保証できず、脆弱性が生じる。この脆弱性を突く攻撃が Adversarial Examples 攻撃[6)]である。たとえば、機械学習を用いた画像認識システムがそれまで正しく認識できていた画像に対して、人間には気にならない程度の小さな加工（ごく小さなノイズ等）を加えて、それまでと全く異なる誤認識結果を出させるというものである。道路標識を対象とした実験で、停止標識を速度制限標識と誤認識させた（停止しなければ事故を招く）と報告されている。

機械学習のテスティング技術は、AI ソフトウェア工学分野で特に活発に取り組まれている研究テーマの一つである。機械学習は訓練データによってシステムの動作が定まるが、起こり得るすべてのケースを訓練データやテストデータとして事前にカバーすることはできない。そのような前提のもと、テストデータのカバレッジやパターン量を適切かつ効率よく増やすためのさまざまな手法が開発されている[23), 27)]。具体的には、ニューロンカバレッジ、メタモルフィックテスティング、サーチベースドテスティング、データセット多様性等のアイデアが知られている。ニューロンカバレッジは、深層学習等ニューラルネットワーク系の機械学習において、ニューラルネットワーク内の活性化範囲を調べ、それを拡げるようなテストパターンを生成しようという考え方である。メタモルフィックテスティングは、入力を変えると出力はこう変わるはずという関係を検証し、既存テストケースから多数のテストケースを生成する手法である。サーチベースドテスティングは、メタヒューリスティックを用いて、欲しいテストケー

スを表すスコアを最大化するようテストケースを生成する手法である。

③ 機械学習における公平性・解釈性・透明性（FAT/ML）

AI・機械学習技術がさまざまな応用に使われ、社会に広がるにつれて、社会における AI という視点からの議論が活発に行われるようになった。各国の政府機関・学会・企業等が AI あるいは AI 応用システムが満たすべき要件をまとめ、公開する動きも進んだ[28]。研究開発という側面では、FAT/ML（Fairness, Accountability, and Transparency in Machine Learning）というテーマで、技術的な定式化や対策検討が行われている。

公平性（Fairness）に関する問題[29), 30), 31)]は、主に訓練（学習）データに潜んでいた差別・偏見をシステムが学習してしまうというものである。たとえば、罪を犯した人間の再犯率を予測するために、人間による判断結果を機械学習にかけたとする。もしその判断を行った人間が特定人種に偏見を持っていたならば、機械学習による判断結果もその傾向を反映してしまう。また、個々の訓練データに偏見を持った判断が含まれているか否かだけでなく、訓練データの分布の偏りや量が差別を生むケース（特定人種の誤認識が多い、採用判定を性別で差別した等）も指摘されている。

説明責任（Accountability）あるいは解釈性に関する問題は、機械学習で獲得されたモデルやルールが人間には理解が難しい、また、機械学習による判定について理由を説明できないというブラックボックス問題[12)]である。すべての機械学習アルゴリズムがこれに該当するわけではないが、近年応用が拡大している深層学習はこれに該当する。顧客や利用者への説明が求められるケースや、事故・エラーが発生したときの原因解明・切り分けのために解釈性が求められるケース等、対策が求められる。

透明性（Transparency）に関する問題は、説明責任・公平性と要件的には重なるが、関連するトピックとして、欧州で 2018 年 5 月に施行された一般データ保護規制（General Data Protection Regulation：GDPR）に AI の透明性を求める条文（GDPR 第 22 条）が盛り込まれたことから、規制遵守という面での対応が求められる。

以上のような FAT/ML の問題に対して、AI・機械学習の主要な国際会議（NIPS、ICML、KDD）とも併設される形で 2014 年から FAT/ML ワークショップが毎年開催されており、2018 年からは ACM FAT* が開催されるようになった。技術的には、前述の XAI や、FADM（Fairness-aware Data Mining）[30), 31)]といった研究分野が立ち上がっている。XAI については、DARPA のプロジェクトの概要を後述するので、ここでは省略する。FADM については、不公平さの検出と不公平の防止という 2 つのタスクを中心に取り組まれている。訓練に用いるデータ集合に差別・偏見が潜んでいないかという視点だけでなく、訓練時にセンシティブな属性をどう扱うかという視点も必要になる。人種・性別等の属性を除外して機械学習を実行したとしても、他の属性に人種・性別等と相関の高いものがあれば、不公平な結果を出してしまうかもしれない。また、通常、機械学習の精度と公平性・解釈性の確保はトレードオフ関係が生じるので、応用・目的に応じた設計が必要である。

［注目すべき国内外のプロジェクト］

① 機械学習工学研究会 MLSE

前述の通り、2018 年 4 月に、日本ソフトウェア科学会に機械学習工学研究会 MLSE が発足した。活動の趣旨やスコープについては前述したので、ここでは繰り返さないが、発足後の具体的な活動として、2018 年 5 月にキックオフシンポジウム、7 月にワークショップ、12 月に国際ワークショップ iMLSE（1st International Workshop on Machine Learning Systems Engineering）が開催された。6 月・8 月・9 月には関連国内学会の全国大会で企画セッションも開催された。情報処理学会誌 2019 年 1 月号（60 巻 1 号）では「機械学習工学」特集が組まれ、取り組みを俯瞰する 6 本の解説論文 [21), 22), 23), 24), 25), 26)] が掲載された。他に論文読み会や勉強会といった場も運営されている。上記キックオフシンポジウムは 500 名以上、産業界からの参加者比率も高く、業界の問題意識・関心が顕在化された。国際ワークショップ iMLSE では、前述の SEMLA イニシアティブのオーガナイザーメンバーである Giuliano Antoniol（モントリオール理工科大学教授）と Foutse Khomh（同大学准教授）を招き、SEMLA イニシアティブとの国際連携も図られた。AI ソフトウェア工学への取り組みが国際的にも活性化しつつある中で、日本は具体的な活動でやや先行しているが、MLSE の牽引によるところが大きい。

② ペガサスプロジェクト

ドイツ経済エネルギー省（BMWi）が主導するペガサスプロジェクト [10)] は、自動走行システムの期待性能水準（自動運転車両はどれくらいの性能が求められるか？）と評価基準（求められる性能を満たすことをどのようにして確認するか？）を明確化するための産学官共同プロジェクトである。BMW、Daimler、Volkswagen、Audi 等の自動車メーカー、Robert Bosch GmbH（ボッシュ）等のサプライヤー、ドイツ航空宇宙センターやダルムシュタット工科大学等の大学・研究機関、第三者認証機関 TÜV SÜD（テュフズード）等、計 17 団体が参加している。従来の V 字モデルに代わる自動運転の安全性評価フレームワークについて、自動走行車の機能・要件定義、走行環境・走行状況のデータ化、それらを組み合わせたテストシナリオ作成、テストシナリオに基づくシミュレーション・実走行テストとリスク許容性を踏まえた安全性評価という 4 項目が検討されている。プロジェクト期間は 2016 年 1 月～ 2019 年 6 月で、終了が近づいているが、活動は継続し、評価手法の国際標準化を目指すとのことである。

③ 米国 DARPA における関連プロジェクト（XAI、Assured Autonomy）

米国 DARPA の AI 関連プログラムの中で、安全性・信頼性に関わり、特によく知られているのが XAI プロジェクト [17)]（2017 年 5 月～ 2021 年 4 月の 4 年間）である。このプロジェクトは、人間の意思決定を支援するパートナーとしての AI を、人間の兵士が理解・信頼し、管理することを目指している。つまり、技術的には、機械学習の精度と解釈性の両立が目標であり、具体的なタスクとして Data Analytics と Autonomy という 2 つを設定している。Data Analytics タスクは、マルチメディアデータを解析してターゲットを選択するに際して、選択の理由も示す（たとえば、敵と判断した理由として銃の輪郭を強調表示する等）。Autonomy タスクは、ドローンやロボット等の自律システムにおいて、どういう状況でどういう理由で次の行動を決定したかを説明する。

これらのタスクにおいて、機械学習の精度と解釈性を両立させるアプローチは 3 通りが検討

されている。1つめはDeep Explanationである。深層学習について、入力の部分的な変化に対するモデル内の反応の変化を調べる等により、結果に特に影響を与えている要素・特徴を推定し、説明の根拠とする方法である。2つめはModel Inductionである。深層学習等のブラックボックス型の機械学習に対して、その入出力の振る舞いを、より解釈性の高い別モデルで近似する方法である。3つめはInterpretable Modelsである。ブラックボックスではない解釈性の高いモデルを用いる方法だが、決定木や重回帰分析のような従来のシンプルなモデルでは解釈性があっても高い精度が得られないため、解釈性がありながら精度も高い新しいモデルを作る。また、説明可能なモデルだけでなく、説明のためのインターフェースや説明の心理学的モデルの研究にも取り組まれている。

もう一つ、Assured Autonomyプロジェクト[18]（2018年5月～2022年4月の4年間）では、自動運転車やドローン等の自律システムの安全性確保が検討されている。そのために、実行時の入出力をモニタリングし、想定された振る舞いから外れたときに、それを検知して動作を止める等の対策アクション（事前に設計しておいたもの）を起動するような枠組みが研究開発されている。

（5）科学技術的課題

システムの安全性・信頼性の確保は、1つの技術によって達成し得るものではなく、多面的・総合的な取り組みが不可欠である。全体像を押さえつつ、見通しよく、バランスよく、取り組むべきである。そのため、個々の技術開発と並行してAIソフトウェア工学の体系化を進めることが重要である。これについては、前述の通り、ソフトウェア工学の知識体系SWEBOKをベースに体系を描くと見通しがよい。

ここで、AIソフトウェア工学に対する多面的・総合的な取り組みは、（A）実践から得られる知見を体系化していくアプローチや（B）理論・原理に関わる学術研究のアプローチが可能である。これらのうちアプローチ（A）は産業界を中心に自ずと進んでいくに違いない。その一方で、システム開発のパラダイム転換に対応するには、より根本的な取り組みとして、アプローチ（B）が不可欠である。このアプローチ（B）の学術的な基礎研究として重要な技術課題として、以下の4点があげられる[1]。

① 機械学習自体の品質保証手法の確立

データ集合に着目した機械学習の品質（精度、公平性等）の理論やテスティング手法が重要である。機械学習の精度や想定外リスクを左右するのは訓練データ集合なので、テストや訓練に用いるデータ集合の選び方やデータ集合の質の評価等が検討されている。また、機械学習をシステムに組み込み、運用する過程において、区別して評価・管理すべき品質の種類・性質（プロダクト品質、サービス品質、プラットフォーム品質）[32]も検討されている。研究開発が活性化しつつあるが、理論的整備や公平性確保、裏付けのある品質基準策定等、まだ多くの研究課題が残されている。

② 全体システムとしての安全性確保技術の確立

従来型と機械学習型の混在システムの全体としての安全性評価法やリカバリー処理設計法が必要である。機械学習型コンポーネントが100%保証はできないことを想定した全体システム

設計（Safe Learning、Safety Envelope）[33]、機械学習型の入力・出力をモニタリングして例外処理・リカバリー処理を起動するシステム構成法[18]が検討されている。また、安全性分析・ハザード解析手法（STAMP/STPA、FRAM）[19), 20)]に機械学習の特性も取り入れた拡張も今後の研究課題である。

③ ブラックボックス問題への対策

前述したように、機械学習応用に不安を与えている要因として、機械学習が判定したとき、その理由を人間が理解できないというブラックボックス問題がある。（すべての機械学習方式がブラックボックスというわけではないが）システムにブラックボックスな部分があると、設計時や利用時に、設計者あるいは利用者がシステムの動作を予測することが難しく、システムの安全性を確保する上で妨げになる。また、もしシステムが事故を起こしてしまった場合、その原因理解や責任の切り分けが難しいということにもなる。そこで、システム動作の解釈性を向上させ、事故発生時の原因理解や設計時・利用時の動作予測の容易にするため、XAI（説明可能なAI）と呼ばれる研究テーマへの取り組みが進められている[13), 17)]。

④ 問題を効率よく解く工学的な枠組みの創出

機械学習（ボトムアップ）と記号推論（トップダウン）を統合した効率的な問題解法を追求する。深層学習が広がったことで帰納型・ボトムアップ型の問題解決が主流になっているが、本来、演繹型・トップダウン型の記号推論も重要である。それらのあるべき組み合わせや新しい問題解決の枠組みを探究することは、AI応用システムのアーキテクチャ設計という面で重要である。新たな問題解法、効率的な処理構成として、さまざまな可能性が考えられる[34), 35)]。

以上の①～④は既に取り組みが始まっている重要な技術課題をあげたが、他にも今後取り組みが必要になると思われる技術課題を以下に2つあげる。

⑤ 自動変化するシステムの品質保証技術の創出

機械学習は、訓練データを与えてモデルを生成する訓練フェーズ（学習フェーズとも呼ばれる）と、その訓練済みモデルを用いて、新たに入力されたデータを判定する判定フェーズ（推論フェーズとも呼ばれる）を持つ。上で述べてきた品質保証に関わる手法は基本的に、訓練フェーズのバッチ的実行を想定している。すなわち、初期の訓練であれ、追加の訓練であれ、訓練フェーズを実行したら、判定フェーズに入る前に、必ず訓練済みモデルを評価・テストがされなければならない。しかし、機械学習の使い方として、オンライン学習によって、モデルを随時更新しながら、判定にも使っていく形があり得る。このような形の場合、システムの品質保証ははるかに難しく、新たな技術チャレンジが必要である。

⑥ 擬人化インターフェースの設計手法の確立

ここまで、AI応用システム開発における問題を、機械学習に起因するものにフォーカスして論じたが、機械学習以外にも問題になり得る要因が考えられる。その一つは擬人化インターフェースである。2次元（画面表示）にせよ3次元（ロボット形状）にせよ、人間の形状・表情・対話を模したインターフェース（擬人化インターフェース）を持つAI応用システムが提

供されつつある。擬人化インターフェースの利点は、そのシステムと相対する利用者にとって、システムがどのような応答をするかのモデルを仮定しやすいことである。しかし、それは逆に、利用者が思い込みをしやすい面があり、利用者が仮定したモデルと、実際のシステムの応答モデルとの間のギャップが、想定外の状況を生む可能性を持つ。そのような認知ギャップを最小化するような設計手法が求められる。

（6）その他の課題

① AI 品質基準・安全性基準の策定・運用および国際標準化

AI 品質を日本の強みにして、AI 適用産業の国際競争力を強化するチャンスである。そのため、国として適切な品質基準・安全性基準を策定し、その認証を行う機関を設立・運用（評価のための適切なデータセットの構築・管理も含む）、さらには国際標準化も推進すべきである。粗悪な AI 商品・サービスの流通は事故を多発させ、AI 応用システムに対する消費者の不安・懸念を招き、厳しい規制や国内 AI 関連産業の委縮につながる。そのような事態を回避し、市場の健全化と産業の活性化を促進するため、十分な根拠を持った適切な基準策定と、AI 応用システムの社会受容性を確保するための施策が求められる。

品質基準・安全性基準の策定・標準化は、自動車分野での取り組みが進みつつある。この分野で先行的な取り組みを推進し、さらに他分野への展開を進めたい。

② 学術研究・人材育成、実応用での技術実証、基準策定・標準化の密連携推進

AI ソフトウェア工学の研究開発は、(1) 学術研究＆人材育成、(2) 実応用での技術実証、(3) 基準策定・標準化、という 3 つの活動を密連携させて推進することが不可欠である。そのための司令塔の役割を持つ部門も必要である。産業界の AI 応用システムの開発現場に導入され、効果を生むものでなくてはならない一方で、パラダイム転換に対応できる新たな原理・理論を学術的な基礎研究から生み出していく必要がある。

③ 問題解決型基礎研究のトップダウン推進

システムの安全性・信頼性の確保には、さまざまな技術を用いて、多面的・総合的に取り組む必要がある。問題解決の視点から技術体系を構築していくことが必要になる。基礎研究を強化・推進するためのファンディングの考え方として、シーズプッシュ型（リニアモデル）は馴染まない。問題解決型でさまざまな解法を試み、その中から筋のよいものを見いだし、必要ならば、それらを組み合わせて、より優れた解法を創出していくようなアプローチが適する。このような問題解決型の基礎研究をトップダウンに強化・推進するファンディングを考える必要がある。

④ 日本の先行優位性を活かす迅速な施策推進

AI ソフトウェア工学は、学問分野としてまだ黎明期だが、AI に関わる他分野と同様に、動きがはやいことは間違いない。しかも、現状、国内での取り組みは海外にやや先行した面もあり、決して後追い状況ではなく、世界的にリードできるチャンスがある。また、品質基準・安全性基準等を打ち出して、標準化活動を進めるにあたっても先行優位性はキープしたい。したがって、日本の先行優位性を活かす迅速な施策推進が強く望まれる。

⑤ 機械学習活用に関わる知的財産権の整備

機械学習に用いるデータや解析結果に関わる知的財産権に加えて、機械学習固有の問題として訓練（学習）済みモデルの知的財産権の問題がある。訓練済みモデルの再利用のパターンは、(1) Copy：そのまま複製して使う、(2) Fine Tuning：ある訓練済みモデルにさらにデータを与えて追加訓練したものを行う、(3) Ensemble：複数の訓練済みモデルの出力を束ねて（平均・多数決等）使う、(4) Distillation：ある訓練済みモデルの振る舞い（どんな入力を与えたときにどんな出力が得られるか）を訓練データとして作ったモデルを使う、という 4 通りがある [3]。このようなパターンを含めて、訓練済みモデルの知的財産権をどのように保護すべきか、法整備が必要である。

（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	○	↗	JST や NEDO のプログラムにおいて、AI 応用システムの安全性や品質保証に取り組む研究課題が採択されている。具体的には、JST ERATO「蓮尾メタ数理システムデザインプロジェクト」（2016 ～ 2022、NII 等）、NEDO（AI・ロボット中核技術）「機械学習 AI の品質保証に関する研究開発」（2018 ～ 2022、産総研・NII）、JST MIRAI（サイバー世界とフィジカル世界を結ぶモデリングと AI）「高信頼な機械学習応用システムによる価値創造」（2018 採択、NII）が進められている。現状、少数の先進的研究者によって牽引されていて、研究者層がまだ薄い。国の AI 戦略に AI 工学が掲げられ、今後、政策強化が期待される。
	応用研究・開発	○	↗	2018 年 4 月に機械学習工学研究会 MLSE と QA4AI コンソーシアムが発足し、産業界からの多数の参画もあり、活発に活動が進められている。JAXA と JAMA の共同検討やデンソーでの取り組み等、自動車分野での安全性保証に関する技術開発も進められている。
米国	基礎研究	○	↗	DARPA が 2017 年から XAI プロジェクト、2018 年から Assured Autonomy プロジェクトをスタートさせており、基礎研究への比較的大型の政府投資がなされている。
	応用研究・開発	○	↗	GAFA は AI 応用システム開発に関する実践的な手法や知見を保有している。Edge Case Research と Underwriters Laboratories が共同で安全性基準策定・認証機関への検討を進めている。
欧州	基礎研究	○	↗	ドイツの産官学連携によるペガサスプロジェクトが、自動運転分野の安全性評価の基準や評価手法の開発で国際的に先行している。
	応用研究・開発	○	↗	イギリスの DeepMind が自社の AI 開発ガイドラインをまとめ、公開している。ドイツでは DFKI と TÜV SÜD が共同で AI に関する第三者認証の検討を始めた。
中国	基礎研究	△	→	2018 年 11 月に深圳で開催された第 17 回全国ソフトウェア応用大会において「AI に基づくソフトウェア工学」セッションが開催された。
	応用研究・開発	△	→	欧米企業に比べると、安全性・信頼性に関する中国企業での取り組みは現状あまり見えていない。
韓国	基礎研究	×	→	現状、特段の活動が見られない。
	応用研究・開発	×	→	現状、特段の活動が見られない。

（8）参考文献

1）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）。
2）丸山宏、「機械学習工学に向けて」、『日本ソフトウェア科学会第 34 回大会講演論文集』（2017 年 9 月）。
3）丸山宏・城戸隆、「機械学習工学へのいざない」、『人工知能』（人工知能学会誌）33 巻 2 号（2018 年 3 月）、pp. 124-131。
4）Andrej Karpathy, “Software 2.0”, Medium (2017.11.12). https://medium.com/@

karpathy/software-2-0-a64152b37c35 (accessed 2019-01-18)
5）Kunle Olukotun, “Designing Computer Systems for Software 2.0”, Invited Talk (December 6, 2018) in *the 32nd Conference on Neural Information Processing Systems* (NeurIPS 2018; Montréal, Canada, December 3-8, 2018).
6）Kevin Eykholt, Ivan Evtimov, Earlence Fernandes, Bo Li, Amir Rahmati, Chaowei Xiao, Atul Prakash, Tadayoshi Kohno and Dawn Song, “Robust Physical-World Attacks on Deep Learning Models”, arXiv:1707.08945 (2017).
7）Kate Crawford, “The Trouble with Bias”, Invited Talk (December 5, 2017) in *the 31st Conference on Neural Information Processing Systems* (NIPS 2017; Long Beach, Calfornia, December 4-9, 2017).
8）進藤智則、「深層学習や機械学習の品質をどう担保するか？新しいソフト開発手法と位置付け「工学体系」構築へ」、『日経ロボティクス』2018年6月号（2018年）、pp. 3-10。
9）Patrice Y. Simard, Saleema Amershi, David M. Chickering, Alicia Edelman Pelton, Soroush Ghorashi, Christopher Meek, Gonzalo Ramos, Jina Suh, Johan Verwey, Mo Wang and John Wernsing, “Machine Teaching: A New Paradigm for Building Machine Learning Systems”, arXiv:1707.06742 (2017).
10）Ulrich Eberle, Olaf Schädler, Sven Hallerbach, Matthias Büker , Sebastian Gerwinn, Tino Brade and Christian Amersbach, “Automated Driving & Development Processes － A Brief Look Beyond”, *Pegasus Symposium* 2017 (Aachen, Germany, November 9, 2017).
11）桑島洋・安岡宏俊・中江俊博、「セーフティクリティカルな機械学習システムの品質保証の課題」、『第1回機械学習工学ワークショップ』（MLSE2018；2018年7月）。
12）Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Dino Pedreschi and Fosca Giannotti, “A Survey Of Methods For Explaining Black Box Models”, arXiv:1802.01933 (2018).
13）高野敦、「もうブラックボックスじゃない、根拠を示してAIの用途拡大」、『日経エレクトロニクス』2018年9月号（2018年）、pp. 53-58。
14）Pedro A. Ortega, Vishal Maini and the DeepMind safety team, “Building safe artificial intelligence: specification, robustness, and assurance”, Medium (2018.9.28.) https://medium.com/@deepmindsafetyresearch/building-safe-artificial-intelligence-52f5f75058f1 (accessed 2019-01-18)
15）鷲崎弘宜、「アジャイル品質保証の知識体系 － SQuBOK 2020予定より－」、『enPiT-Proスマートエスイーセミナー』（2018年9月7日）。
16）SQuBOK策定部会（編）、「SQuBOK Review 2018」（2018年）。https://www.juse.or.jp/sqip/squbok/file/squbok_rev2018_20181016.pdf (accessed 2019-01-18)
17）David Gunning, “Explainable Artificial Intelligence (XAI)”, DARPA Program Update, November 2017. https://www.darpa.mil/attachments/XAIProgramUpdate.pdf (accessed 2019-01-18)
18）Sandeep Neema, “Assured Autonomy”, DARPA Program, August 2017. https://www.darpa.mil/attachments/AssuredAutonomyProposersDay_Program%20Brief.pdf

(accessed 2019-01-18)
19）情報処理推進機構 技術本部ソフトウェア高信頼化センター、『はじめての STAMP/STPA ～システム思考に基づく新しい安全性解析手法～』（情報処理推進機構、2016 年 3 月）。https://www.ipa.go.jp/files/000051829.pdf (accessed 2019-01-18)
20）情報処理推進機構 技術本部ソフトウェア高信頼化センター、『はじめての STAMP/STPA（活用編）～システム思考で考えるこれからの安全～』（情報処理推進機構、2016 年 3 月）。https://www.ipa.go.jp/files/000065199.pdf (accessed 2019-01-18)
21）丸山宏、「機械学習工学の狙いと展開」、『情報処理』（情報処理学会誌）60 巻 1 号（2019 年 1 月、特集：機械学習工学）、pp. 12-16。
22）今井健男・太田満久、「機械学習応用システムのテストと検証」、『情報処理』（情報処理学会誌）60 巻 1 号（2019 年 1 月、特集：機械学習工学）、pp. 17-24。
23）石川冬樹・徳本晋、「機械学習応用システムの開発・運用環境」、『情報処理』（情報処理学会誌）60 巻 1 号（2019 年 1 月、特集：機械学習工学）、pp. 25-33。
24）吉岡信和、「機械学習応用システムのセキュリティとプライバシ」、『情報処理』（情報処理学会誌）60 巻 1 号（2019 年 1 月、特集：機械学習工学）、pp. 34-39。
25）五十嵐健夫、「機械学習のためのヒューマンインタフェース」、『情報処理』（情報処理学会誌）60 巻 1 号（2019 年 1 月、特集：機械学習工学）、pp. 40-47。
26）本橋洋介、「機械学習応用システムのプロジェクト管理と組織」、『情報処理』（情報処理学会誌）60 巻 1 号（2019 年 1 月、特集：機械学習工学）、pp. 48-55。
27）中島震、「データセット多様性のソフトウェア・テスティング」、『日本ソフトウェア科学会第 34 回大会講演論文集』（2017 年 9 月）。
28）情報処理推進機構 AI 白書編集委員会（編）、『AI 白書 2019 ～企業を変える AI 世界と日本の選択～』（KADOKAWA、2018 年）。
29）MIT テクノロジーレビュー編集部（翻訳・編集）、『AI and bias：人工知能は公平か？』（MIT テクノロジーレビュー Special Issue Vol. 7；2018 年 5 月 31 日発行、角川アスキー総合研究所）。
30）福地一斗、「公平性に配慮した学習とその理論的課題」、『第 21 回情報論的学習理論ワークショップ』（IBIS2018；2018 年 11 月 6 日）。
31）神 嶌 敏 弘、「Fairness-Aware Data Mining」、2018 年。http://www.kamishima.net/archive/fadm.pdf (accessed 2019-01-18)
32）中島震、「機械学習ソフトウェアの品質：製品、サービス、プラットフォーム」、『電子情報通信学会 知能ソフトウェア工学研究会』（KBSE；2018 年 11 月 9 日）。
33）蓮尾一郎、「統計的機械学習と演繹的形式推論：システムの信頼性と説明可能性へのアプローチ」、『日本数学会 2018 年度秋季総合分科会 数学連携ワークショップ』（2018 年 9 月 24 日）。http://group-mmm.org/~ichiro/talks/20180924okayama.pdf (accessed 2019-01-18)
34）David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel and Demis Hassabis, “Mastering the game of Go with deep neural networks and tree

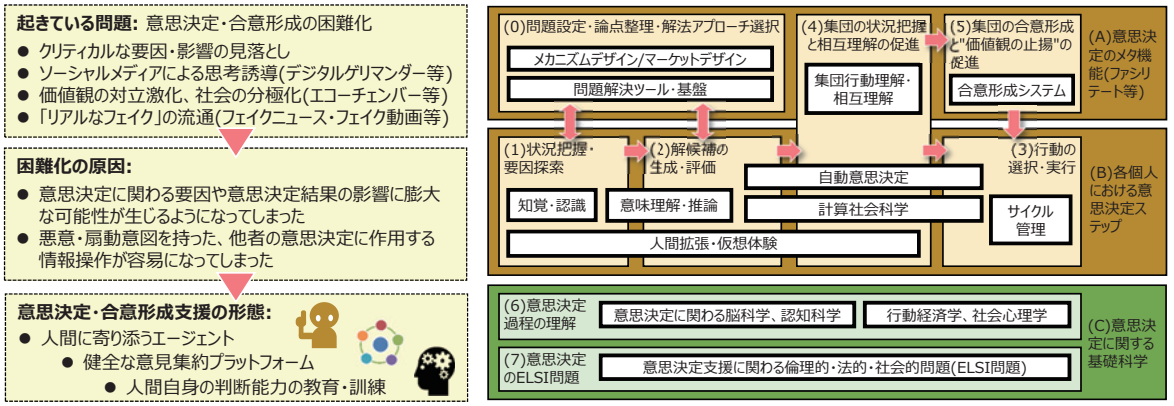
search”, *Nature* Vol. 529, (2016) pp. 484-489. DOI:10.1038/nature16961
35）我妻広明、「データ駆動型人工知能と論理知識型人工知能の融合による解釈可能な自動運転システムに関する研究」、『NEDO「次世代人工知能・ロボット中核技術開発」（人工知能分野）中間成果発表会』（2017 年 3 月 29 日）。https://www.airc.aist.go.jp/info_details/docs/170329/1730-Wagatsuma.pdf (accessed 2019-01-18)

2．1．5　意思決定・合意形成支援

（1）研究開発領域の定義

「意思決定」は、個人や集団がある目標を達成するために、考えられる複数の選択肢の中から一つを選択する行為である。その選択では個人の価値観が拠りどころとなるが、集団の意思決定では、必ずしも関係者（メンバーやステークホルダー）全員の価値観が一致するとは限らない。関係者内で選択肢に関する意見が分かれたとき、その一致を図るプロセスが「合意形成」である。

情報爆発等による可能性の見落としやフェイクニュース等による悪意・扇動意図を持った思考誘導操作といった問題が顕在化し、意思決定ミスを起こしてしまうリスクが高まっている。このような問題・リスクを人工知能（AI）技術等の情報技術によって軽減し、個人・集団が主体性・納得感を持って意思決定できるように支援することを目指す研究開発領域である[1)]。



研究開発課題（番号は右上図との対応）	これまでの取り組み	今後の課題・方向性
膨大な可能性の探索・評価(1)(2)	因果関係推論、マルチエージェントシミュレーション等	因果関係知識獲得、常識推論等
自動意思決定(2)(3)	強化学習、予測型意思決定最適化等	多様な価値観や不完全情報での意思決定等
多様な価値観の把握・可視化(4)	gIBIS、言論マップ生成、VR・ゲーミングによる追体験等	議論マイニング、ディベートAI、法学AI等
交渉・合意形成・集合知支援(5)	討論型世論調査、自動交渉エージェント、インセンティブメカニズム等	高度な自動交渉、価値観の相互作用モデル等
フェイク対策(0)(1)	ソーシャルネット分析、メカニズムデザイン等	高度なファクトチェック支援、フェイク検出等
意思決定に関する基礎科学(6)(7)	脳の意思決定メカニズム、行動経済学、認知バイアス等	ELSI・社会受容性、利用者から見た透明性等

関連動向
- AI・ビッグデータ技術による分析ビジネスは自動意思決定・意思決定支援を提供する処方的分析へ(Gartner)
- 意思決定問題は、古くから主に人文・社会科学の分野で取り組まれてきたが、今日、情報科学技術の関わりが増大
- AI(機械学習・最適化)が扱うのは評価関数が定まり解が1つ求まる問題、多様な価値観が混在する問題は新たな技術チャレンジ
- 欧米では、AIへの大型投資に加えて、学際的な意思決定研究が進んでいる(大規模意見集約、討論型世論調査、市民科学等)
- 米国では政治・ジャーナリズム分野でもAI技術との融合的な取り組みや議論が進展

図2-1-7　領域俯瞰：意思決定・合意形成支援

（2）キーワード

意思決定、合意形成、意見集約、フェイクニュース、フェイク動画、デジタルゲリマンダー、議論マイニング、マルチエージェント、自動交渉、メカニズムデザイン、計算社会科学、行動経済学、処方的分析

（3）研究開発領域の概要

[本領域の意義]

我々は日々さまざまな場面で意思決定を行っている。クリティカルな場面での意思決定ミスは個人や集団の状況を悪化させ、その存続・生存さえも危うくする。例えば、企業の経営における意思決定ミスは、企業の業績悪化・競争力低下を招き、国の政策決定・制度設計における意思決定ミスは、国の経済停滞や国民の生活悪化にもつながる。また、個人の意思決定における判断スキル・熟慮の不足は、その個人の生活におけるさまざまなリスクを誘発するだけでなく、世論形成・投票等における集団浅慮という形で、社会の方向性さえも左右する。

情報技術が発展し、社会に浸透した今日、情報の拡散スピードが速く、膨大な情報があふれ、影響を及ぼし合う範囲が思わぬところまで広がっている。そのような意思決定の行為自体の難しさが増していることに加えて、意思決定の際の拠りどころとなる価値観の多様化[2), 3)]によって、合意形成の難しさも増している。さらには、価値観の対立から悪意・扇動意図を持った思考誘導の情報操作（フェイクニュース、フェイク動画等）まで行われるという問題も顕在化し[4), 5)]、社会問題化している。

このような意思決定の困難化（意思決定ミスを起こすリスクの増大）という状況に対して、AI 技術等の情報技術の活用は、問題のすべてを解決できるわけではなくとも、リスクを軽減し、状況を改善する手段になり得る。技術発展とともに急速に変化し、複雑化する社会環境にあっても、個人・集団が主体性・納得感を持って意思決定できるように、情報技術を用いて支援することが、上で述べたようなさまざまな意思決定場面で効果を生むと期待される。

また、集団での意思決定は、単に異なる意見の間の調整・交渉ではなく、多様な視点・考えからの集合知が期待でき、情報技術はその活性化・活用促進にも効果がある。

[研究開発の動向]

① 意思決定問題への取り組み

個人・集団の意思決定問題は古くから検討されてきた問題である。意思決定に関する先駆的な研究としては、1978 年にノーベル経済学賞を受賞した Herbert A. Simon の取り組み[6), 7)]がよく知られている。Simon は意思決定プロセスを、(1) 情報（Intelligence）活動、(2) 設計（Design）活動、(3) 選択（Choice）活動というステップで構成されるとした。(1) で意思決定に必要な情報を収集し、(2) で考えられる選択肢をあげ、(3) で選択肢を評価し、どれを選択するか決定する。これらのステップにおいて、必要な情報をすべて集めることができ、可能性のあるすべての選択肢をあげることができ、各選択肢を選んだときに起こり得るすべての可能性を列挙して評価することができるならば、合理的に最良の選択が可能になる。しかし、現実にはそのようなすべての可能性を考えて意思決定することはできず、人間が合理的な意思決定をしようとしても限界がある。この Simon が導入した「限定合理性」（Bounded Rationality）という概念は、意思決定に関する研究発展の基礎となった。Simon は、経営の本質は意思決定だと考え、限定合理性を克服するための組織論も展開した。

そのように人間の判断・行動が必ずしも合理的になり得ず、心理・感情にも左右されるものであることを踏まえて、行動経済学が発展し、その中では意思決定に関わる興味深い知見が示されている。特に有名なのは、Simon の後、行動経済学の分野でノーベル経済学賞を受賞した二人、Daniel Kahneman（2002 年受賞）と Richard H. Thaler（2017 年受賞）の研究であ

る。Kahneman は、直観的な「速い思考」のシステム 1 と論理的な「遅い思考」のシステム 2 というモデル[8]を提唱し、Thaler は、軽く押してやることで行動を促す「ナッジ」（nudge）という考え方[9]を提唱した。

また、脳科学分野における脳の意思決定メカニズムの研究も進んでいる（詳細は「2.1.7 計算脳科学」を参照）。ドーパミン神経細胞の報酬予測誤差仮説等が見いだされ、モデルフリーシステムによる潜在的な意思決定と、モデルベースシステムによる顕在的な意思決定が協調および競合しつつ、人間の意思決定が動作していることがわかってきた[10]。モデルフリーシステムは、事象と報酬との関係を直接経験に基づき確率的に結び付ける。モデルベースシステムは、事象と報酬との関係を内部モデルとして構築し、直接経験していないケースについても予測を可能にする。このような 2 通りのシステムは Kahneman のモデル（システム 1・システム 2）とも整合しており、意思決定が合理性だけによるものではないことの裏付けにもなる。

このような人文・社会科学分野や脳科学分野における意思決定に関する研究が、主に人間の側から掘り下げられてきた一方で、近年の情報技術の発展、Web やソーシャルメディアの普及は、意思決定を行う人間の環境を大きく変化させた。その結果、意思決定問題は新たな様相を呈するようになり、以前とは異なる困難さが生じている。今日、意思決定問題は情報技術との関わりが大きなものになっている。

② 意思決定問題の新たな様相・困難さ

本研究開発領域で取り組む意思決定・合意形成支援は、上で述べたような新たな様相・困難さに対処するために、情報技術を活用する。具体的な技術内容を説明する前に、新たに生じている困難さを示す事象（問題）として顕著なものを 4 つあげる。

1 つめは、クリティカルな要因・影響の見落としの問題である。たとえば、グローバル化したビジネス競争環境において、世界のあらゆる地域、思ってもいなかった業種から新たな競合が生まれ、想定していなかった法規制やソーシャルメディアで思わぬ切り口からの炎上も起こり得る。膨大な情報があふれ、社会がボーダーレス化した今日、意思決定に関連しそうな要因や意思決定結果の影響に膨大な可能性が生じ、人間の頭でそのあらゆる可能性をあらかじめ考えるのは極めて難しい。Simon のいう限定合理性が極度に進み、問題として深刻化している状況である。

2 つめは、ソーシャルメディアによる思考誘導の問題である。Web やソーシャルメディアを用いた情報発信・交流が広がり、それが人々の意思決定や世論形成に与える影響は無視できないものになっている[11), 12)]。2016 年の米国大統領選挙はその顕著な事例であり[4), 5)]、SNS（Social Networking Service）等のソーシャルメディアを用いた政治操作は「デジタルゲリマンダー」と呼ばれ[13)]、フェイクニュースが社会問題化した。SNS では、価値観が自分に近い相手としかつながらず、自分の価値観に沿った情報しか見ない、いわゆる「フィルターバブル」状態[14)]に陥りやすいことも、SNS が思考誘導の道具になりやすい原因になっている。

3 つめは、価値観の対立激化、社会の分極化の問題である。集団の合意形成に難航し、対立が激化する傾向が強まっている。価値観の対立は古くから起こってきた事象だが、社会のボーダーレス化に伴う関係者範囲の広がりや、SNS での同調圧力やエコーチェンバー現象による意見同質集団の形成強化が、対立を強め、社会の分極化（Polarization）や政治的分断と言われる事態も引き起こされている[15), 16)]。

4つめは、まるで本物のようなフェイク動画・画像の流通の問題である。前述のフェイクニュースは言葉（SNSテキスト等）で伝達されるものが主であったが、深層学習を用いた敵対的生成ネットワーク（Generative Adversarial Network：GAN）技術（GAN技術自体については「2.1.2 画像・映像解析」を参照）によって、まるで本物のように見えるフェイク動画やフェイク画像が簡単に作れてしまうようになった（Deepfakes、OpenFaceSwap等）。特にフェイク動画は本物だと信じ込まれやすく、政治家や有名人の架空の発言・行為等を作るためにこれが悪用され、社会に流通すると、何が真実で何かフェイクか、真偽判断を見誤るリスクが増大し、さまざまな混乱が生まれると危惧される[15), 17)]。

以上の問題に見られるように、(1) 意思決定に関わる要因や意思決定結果の影響に、膨大な可能性が生じるようになってしまったこと、(2) 悪意・扇動意図を持った、他者の意思決定に作用する情報操作が容易になってしまったことが、意思決定の困難化の原因として顕著である。

③ 意思決定・合意形成支援のための技術群

図2-1-7の右上部に、個人・集団の意思決定プロセス（合意形成を含む）に対応させて、関連する技術群を示した。Simonの3ステップに相当する (B) 各個人における意思決定ステップを中心に、(A) 意思決定のメタ機能と (C) 意思決定に関する基礎科学を上下に配置した3層構造で技術群を整理した。以下、これらを5つの技術群に分けて、取り組みの現状と今後の方向性について述べる。

膨大な可能性の探索・評価

上述の原因への対策としてまず求められるのは、意思決定に関わる要因や意思決定結果の影響における膨大な可能性を探索し、それらの組み合わせの中から目的に合うものを評価して絞り込む技術である。自然言語処理による因果関係推論[18)]やマルチエージェントシミュレーションによる影響予測[19)]等の研究開発が進められており、具体的なシステムとして「なぜ？」「どうなる？」等の因果関係に関する質問応答を扱うことができる情報通信研究機構（NICT：National Institute of Information and Communications Technology）のWISDOM X[20)]が知られている。しかし、さまざまな分野・文脈で推論が行えるようにするには、常識を含め推論に必要な知識の獲得や、推論が成立する前提条件の精緻化等、取り組まなくてはならない技術課題がまだ多く残されている。

自動意思決定・自動交渉

価値観に対応する評価関数・効用関数が定められるケースは、AI技術の一分野として自動意思決定や自動交渉の研究開発が進められている。機械学習・最適化技術を活用した自動意思決定は、強化学習[21)]や予測型意思決定最適化[22)]等が開発されている（詳細は「2.1.1 機械学習」を参照）。さらに、異なる価値観（効用関数）を持ったエージェント間の自動交渉の研究があり、2010年から毎年、国際自動交渉エージェント競技会ANAC（Automated Negotiating Agents Competition）が開催されている[23)]。しかし、自動交渉は異なる価値観をもつ者の間の勝負という面があり、集団の合意形成を目的とするならば、ファシリテーターの役割[24)]や、相手への共感による価値観の変化といった面も取り込んでいく必要があり、自動ファシリテーション技術やフィールド実験にも取り組まれている[25)]。

多様な価値観の把握・可視化

多様な価値観が混在する状況下での意思決定・合意形成に向けては、その状況や価値観の違いを可視化する技術が有効である。賛成・反対の各立場から意見と根拠を対比する言論マップ生成[26]、主張・事実等への言明とその間の関係（根拠・支持、反論・批判等）を推定する議論マイニング（Argumentation Mining）[27]、議題に対して賛成・反対の立場でディベートを展開するシステム（IBMの「Project Debater」、日立の「ディベートAI」[28]等）が研究開発されている。これらは自然言語処理技術を用いた手法だが、集団の相互理解促進のためにはVR（Virtual Reality）技術やゲーミング手法を用いて相手の立場を追体験させるアプローチも効果がある。

フェイク対策

ソーシャルメディア上での情報伝播の傾向や、そこで起きている炎上、フェイクニュース、エコーチェンバー、二極化等の現象を把握・分析すること[4), 11), 12), 16), 17), 29)]は、フェイク対策のための基礎的研究となる。フェイクニュースへの対抗としては、発信された情報が客観的事実に基づくものなのかを調査し、その情報の正確さを評価・公表するファクトチェックという取り組みが立ち上がっている[4), 30)]。機械学習・自然言語処理等を活用してフェイクニュースを検出する競技会Fake News Challengeも2016年から始まった。ただし、フェイクの検出技術とフェイクの作成法は往々にしていたちごっこになるため、技術開発だけでなく、メディアリテラシーの教育・訓練や法律・ルールの整備による対策も進めることが必要である[4), 17)]。不適切な状況をなるべく回避するようなルール設計という面では、メカニズムデザインの理論[31)]も役立つ。

意思決定に関する基礎科学

情報技術によって人間の意思決定を支援するにあたって、そもそも人間の意思決定とはどういうものか、どうあるべきかを理解しておくことは重要である。既に言及した通り、行動経済学や脳科学の分野で意思決定プロセスのモデルやメカニズムが研究されてきた。また、社会心理学等の分野で研究されている確証バイアスを含む認知バイアスも意思決定に大きく関わる。加えて、人間の意思決定・合意形成を支援する機能がELSI（Ethical, Legal and Social Issues：倫理的・法的・社会的課題）の視点から適切であるかについても常に考えておかねばならない。

（4）注目動向

[新展開・技術トピックス]

① データ分析の発展段階（処方的分析へ）

米国の調査・アドバイサリー企業であるGartnerは、データ分析の発展を記述的分析（Descriptive：何が起きたか）、診断的分析（Diagnostic：なぜそれが起きたか）、予測的分析（Predictive：これから何が起きるか）、処方的分析（Prescriptive：何をすべきか）という4段階で自動化が進むとし、4段階目の処方的分析は「意思決定支援」と「自動意思決定」という2通りがあるとしている[32)]。この段階が進むほど、データ分析の顧客価値が高く、ビジネス上の競争も処方的分析へと進みつつある。

「自動意思決定」はデータ分析の結果に基づき、何をすべきかというアクションまで自動決定するものであり、「意思決定支援」はアクションの候補を人間に提示し、どんなアクションを実行するかは最終的に人間が決定するものである。一見すると、意思決定支援よりも自動意思決定の方が、より発展したものであるかのように思えるが、現状、意思決定問題の性質が異なると考えるのが適切である。すなわち、コスト、精度、速度、売上等のような明確な指標（いわば価値観に相当）が定められ、それを評価関数として合理的に解が一つ定められる意思決定問題は「自動意思決定」が可能になる。それに対して、さまざまな価値観が混在している状況下、あるいは、価値観が不確かな状況下での意思決定問題は、最終決定に人間が関わる「意思決定支援」の形が基本になる。「自動意思決定」タイプの問題は、機械学習・最適化技術等を用いた実用化が盛んに進められているが、「意思決定支援」タイプの問題にはそれとは異なる新たな手法の確立が必要である。

② 大規模意見集約システム

集団の意見集約・合意形成のために情報技術を活用するシステムは、古くはグループウェアや CSCW（Computer Supported Cooperative Work）の研究分野での取り組みが見られる。たとえば gIBIS という意思決定支援ツール[33]がよく知られている。gIBIS は、ワークショップにおける対話のファシリテーションの技法の一つとして、Issue（課題、論点）をベースに木構造の表現でまとめる IBIS（Issue Based Information Systems）法をグラフィカルに実現したシステムである。複雑な問題の本質を可視化・共有するために、このようなツールが有効なことが認識された。

一方、政治学の分野では、討論型世論調査（Deliberative Poll）[34]が、熟議に基づく民主主義を実現するための方法論として有効だと認識されるようになった。この方法論では、あるテーマについて回答を得る前に、回答者にグループ討論をしてもらうようにする。それにより、テーマに関する理解・熟慮が進み、回答の意見分布に明確な差が生じる。

米国マサチューセッツ工科大学（MIT）に 2006 年に設立された MIT Center for Collective Intelligence（集合知研究センター：CCI）では、学際的な取り組みを含め、集合知の収集・活用を推進する研究を進めている。インターネットを使った大規模な議論を、その論理構造の可視化（議論マップ）によって支援するシステム Deliberatorium[35]や、地球温暖化問題を取り上げて、解決プランを協議するシステム The ClimateCoLab 等のプロジェクトを進めている。さらに、CCI のトップである Thomas W. Malone は、2018 年の著書[36]で、人間の集合知に AI との協働を含めた Superminds の方向性を示した。

さらに、名古屋工業大学では大規模合意形成支援システム COLLAGREE[37]の研究開発を進めている。議論構造の可視化に加えて、エージェント技術によるファシリテーター機能の開発により、より柔軟で創造的な合意形成支援を目指している。名古屋・浜松等の自治体での社会実験にも取り組んでいる。

ここで述べたような集団の意思決定の支援では、価値観の異なる人々の間で意見を調整して合意点を見出すというよりも、むしろ、多様な考え・視点を持つ人々のコラボレーションに基づく集合知を通して、より良い解を見出すことに重点が置かれている。そのような集合知をより高めるための AI・情報技術による支援機能が検討されている。

［注目すべき国内外のプロジェクト］

① IBM の Project Debater

Project Debater は IBM が開発したディベート AI システムで、人間とのディベート（討論）をライブで行うことができる。IBM のイスラエルの研究所で 2011 年から研究開発がスタートし、2018 年 6 月に米国サンフランシスコで開催されたイベント Watson West にて、イスラエルの 2016 年度ディベートチャンピオンとライブ対戦し、「政府支援の宇宙探査を実施すべきか否か」という議題で Project Debater が勝利した。ライブ対戦の進行は、対戦する両者が議題の肯定派と否定派に分かれ、まず 4 分間ずつ主張を述べ、次に 4 分間ずつ相手の主張に対する反論を述べ、最後に 2 分間ずつまとめを述べるという形で行われ、その勝敗は聴衆の支持数で決まる。

Project Debater の技術詳細はまだ論文として公開されていないが、データ駆動型のリアルタイム音声対話生成、長い話し言葉中から重要な概念や主張を識別する聞き取り能力、独自のナレッジグラフによる人間のジレンマのモデル化等の技術を新たに開発・適用したということである。ディベートの議題は直前に与えられ、その場で相手の主張も踏まえつつ、自分の主張を組み立てる。ニュース記事や学術論文を 3 億件の収集・構造化しておいて用いており、2011 年に米国のクイズ番組 Jeopardy! で人間のチャンピオンに勝利した IBM Watson[38] の自然言語処理や、研究が活発化している議論マイニング等の技術が応用されているものと考えられる。

なお、類似する国内での研究開発事例として、日立のディベート AI[28] があるが、より応用をフォーカスし、論理構築・推論を深める研究として法学 AI への取り組み [39), 40)] がある。

② 産業競争力懇談会（COCN）「人工知能間の交渉・協調・連携」、

2016 年度・2017 年度にかけて、NEC をリーダー企業として 20 前後の企業・組織が共同で調査・検討した報告書 [41)] が、COCN から出された。それぞれの行動基準（価値観、効用関数）をもった AI システムが混在する社会を想定したときに必要になる、AI システム間の交渉・協調・連携の仕組み作りを提言したものである。製造バリューチェーン、交通・人流、電力・水、自動運転車・移動体という 4 つの産業分野を中心としたユースケースの検討、課題の分析、共通の仕組み（交渉プロトコル等）、国への提言等がまとめられた。

アカデミアを中心に取り組まれてきている自動交渉エージェント技術を踏まえつつ、その具体的な産業応用に向けた検討が、産業界を中心に進められた。

（５）科学技術的課題

まず、［研究開発の動向］の項にあげたような意思決定・合意形成支援に関わるさまざまな要素技術の研究開発は、それぞれさらなる研究開発が必要である。それに加えて、それらの要素技術を活用・統合して、個人・集団の意思決定・合意形成を支援する機能・サービスを、具体的なシステムとして実現する取り組みが求められる。たとえば、以下に示すような支援形態①②③が考えられる。ただし、これらは一例であり、どのような機能・サービスが支援形態として有効かということ自体が研究課題である。

① 人間に寄り添うエージェント

意思決定の困難化の原因(1)として、意思決定に関わる要因や意思決定結果の影響に膨大な可能性が生じるようになってしまったことをあげた。この問題に対して、膨大な可能性を探索することはコンピューターが得意なタスクであることから、例えば、人間（個人や集団）に寄り添うエージェント型の支援形態が考えられる。近年、スマートフォンの音声対話インターフェースや家庭に置かれたスマートスピーカー等、エージェントインターフェースのアプリケーションが広がりつつある。その将来的な拡張機能として、個人向けエージェントならば、個人の意思決定場面で膨大な可能性の中から適切な選択肢をアドバイスしてくれたり、集団向けエージェントならば、議論・交渉の状況をモニタリングしていて、有望な合意点や新たな観点を提示してくれたりといった支援機能が考えられる。

② 健全な意見集約プラットフォーム

意思決定の困難化の原因(2)としてあげたのは、悪意・扇動意図を持った、他者の意思決定に作用する情報操作が容易になってしまったことである。情報技術による対策には限界があり、多面的に取り組む必要がある問題だが、Web・SNS等の情報システムをプラットフォームとして用いた意見集約（意見の発信・収集・投票等）のケースは、情報技術による対策が考えられる。健全な意見集約を行えるように、「悪意・扇動意図を持った、他者の意思決定に作用する情報操作」を検出・防止する機能やメカニズムが組み込まれたプラットフォームであってほしい。今後期待される対策機能としては、たとえば、フェイクニュースやデマ等を検出・排除する機能、ニュース等の一次情報や意見の根拠を追跡・確認する機能、声の大きい意見だけでなく公平に意見を集める機能、ある意見について異なる立場の意見も併記・比較する機能、投票や意見集約プロセスにおいて不正直な申告や裏工作の効果がなくなる（正直申告が最良となる）メカニズム等が考えられる。

③ 人間自身の判断能力の教育・訓練

意思決定の困難化の2つの原因に対する直接的な支援形態は上記①②の通りだが、もう一つ考えておくべき支援形態として、人間自身の判断力を高めるような教育・訓練がある。なぜならば、上に述べたような支援機能によって意思決定・合意形成が楽なものになると、人間は支援機能に頼り、自分自身で考えなくなっていき、人間の判断能力が低下してしまうと懸念されるからである。また、フィルターバブルに陥りやすいか否かには個人差があるものの、誰しもある程度の確証バイアスを持つことは避けられない。そこで、意思決定・合意形成支援の一環で、人間自身の判断能力の教育・訓練という側面も考えておきたい。その実現形態としては、判断能力の教育・訓練用ツールのような直接的な実装だけでなく、①や②に示したような支援機能を利用する過程で、人間自身が深く考えるように促したり、より多面的な検討を促したりするといった組み合わせ型の実装も考えられる。

（6）その他の課題

① ELSIおよび社会受容性に配慮した研究開発

人間の意思決定・合意形成を支援する機能が、倫理的・法的・社会的な視点（ELSI面）から適切であるかを常に考えておかねばならない。たとえば、人間の支援機能を意図したものが、

思考誘導や検閲（表現の自由の制限）と受け止められる可能性もある。

また、支援機能の考え方や仕組みについて利用者から見た透明性を確保し、社会受容性に配慮した技術開発が求められる。そのためには、実社会の具体的な問題に適用して社会からのフィードバックを受けるプロセスを、短いサイクルで回していくのがよい。技術・機能の効果や筋の良し悪しを早い段階で見極め、ELSI 面の懸念点や良い効果が得られないケース等への対策を早期に実施する。

そのような社会への試験的な適用の内容やプロセスは、原則公開で進めるようにするのがよい。どのような取り組みが進められているか、どのような条件でどのような結果が得られているのかを公開して、さまざまな視点から議論し、問題点の指摘やその解決のためのアイデアを得ることが、リスクの早期回避や知見の共有につながる。

② 分野横断の研究開発体制・推進施策

本研究開発領域は、AI 技術等の情報技術だけでなく、計算社会科学、脳科学、認知科学、心理学、経済学、政治学、社会学、法学、倫理学等が重なる学際的な領域であり、分野横断の研究開発体制・推進施策が必要である。そのためには、初期段階から分野横断で研究者を共通の問題意識・ビジョンのもとに束ねる研究開発マネジメントが望ましい。現状、本研究開発領域の個々の技術課題・要素技術に関わる研究者は多いものの、研究者それぞれの取り組みは全体の問題意識に対してまだ断片的なものにとどまっている感が強く、分野横断の連携・統合による骨太化が求められる。

また、情報技術分野と人文・社会科学分野の連携の仕方は、具体的な問題に対する問題の定式化の段階から共同で取り組むことが望ましい。情報技術の側で扱いやすい形の問題にしてしまうとか、人文・社会科学の側から結果に対して駄目出しするとかではなく、具体的な問題に対する定式化において意識合わせをすべきと考える。つまり、実社会への適用において発生するさまざまな制約事項を、アルゴリズム・原理のレベルで扱うのか、運用上の制約（法規制等）の形で扱うのかによって、技術的なアプローチは変わってくる。基礎研究のスタイルとしても、シーズプッシュ型ではなく、問題解決型の基礎研究として推進するのが適する。

（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	○	↗	マルチエージェントシステムの分野で、オークション・マッチングの理論研究やインセンティブメカニズムの研究が多い。
	応用研究・開発	○	↗	大規模合意形成支援システム等で先端的な取り組みや、AI 間の交渉・協調・連携に関する COCN の取り組みがあるが、まだ大きな流れにはなっていない。
米国	基礎研究	◎	↗	MIT CCI の Deliberatorium や The ClimateCoLab、Stanford University の討論型世論調査をはじめ、学際的な基礎研究が根付いている。AI・マルチエージェントシステムの分野で、メカニズムデザイン、オークション・マッチングの理論研究が広く行われている。
	応用研究・開発	◎	↗	上記基礎研究がそのまま応用研究やベンチャーによる産業化につながる傾向が強い。国および企業による AI 分野への大型投資が行われている（Facebook の自動交渉エージェント等）。
欧州	基礎研究	◎	↗	Imperial College London、Oxford University、Delft University of Technology 等、自動交渉の基礎研究が強く、論理的なアプローチによる自動交渉の研究もおこなわれている。
	応用研究・開発	◎	↗	市民からの意見集約や合意形成のためのシステム・応用に盛んに取り組まれている。自動交渉の応用ソフトウェア（電力売買等）への取り組みも見られる。

中国	基礎研究	○	↗	メカニズムデザインや自動交渉の基礎理論は Hong Kong Baptist University で活動が盛んだが、他は見当たらない。
	応用研究・開発	△	→	顕著な活動は見当たらない。
韓国	基礎研究	△	→	顕著な活動は見当たらない。
	応用研究・開発	△	→	顕著な活動は見当たらない。

（8）参考文献

1）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：複雑社会における意思決定・合意形成を支える情報科学技術』、CRDS-FY2017-SP-03（2018 年 12 月）。
2）Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon and Iyad Rahwan, “The Moral Machine experiment”, *Nature* Vol. 563 (24 October 2018), pp. 59-64. DOI: 10.1038/s41586-018-0637-6
3）亀田達也、『モラルの起源－実験社会科学からの問い』（岩波書店、2017 年）。
4）笹原和俊、『フェイクニュースを科学する―拡散するデマ、陰謀論、プロパガンダのしくみ―』（化学同人、2018 年）。
5）湯淺墾道、「米大統領選におけるソーシャルメディア干渉疑惑」、『情報処理』（情報処理学会誌）58 巻 12 号（2017 年 12 月）、pp. 1066-1067。
6）Herbert A. Simon, *Administrative behavior: a study of decision-making processes in administrative organization* (Macmillan, 1947).（邦訳：二村敏子・桑田耕太郎・高尾義明・西脇暢子・高柳美香訳、『新版 経営行動：経営組織における意思決定過程の研究』、ダイヤモンド社、2009）。
7）Herbert A. Simon, *The Sciences of the Artificial* (The MIT Press, 1969).（邦訳：稲葉元吉・吉原英樹訳、『システムの科学 第 3 版』、パーソナルメディア、1999 年）
8）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）
9）Richard H. Thaler and Cass R. Sunstein, *Nudge: Improving Decisions About Health, Wealth, and Happiness* (Yale University Press, 2008).（邦訳：遠藤真美訳、『実践 行動経済学』、日経 BP 社、2009 年）
10）坂上雅道・山本愛実、「意思決定の脳メカニズム―顕在的判断と潜在的判断―」、『科学哲学』（日本科学哲学会誌）42-2 号（2009 年）。
11）遠藤薫、『ソーシャルメディアと〈世論〉形成』（東京電機大学出版局、2016 年）。
12）田中辰雄・山口真一、『ネット炎上の研究』（勁草書房、2016 年）。
13）金子格・須川賢洋（編）、「小特集 ディジタルゲリマンダとは何か―選挙区割政策からフェイクニュースまで」、『情報処理』（情報処理学会誌）58 巻 12 号（2017 年 12 月）、pp. 1068-1088。
14）Eli Pariser, *The Filter Bubble: What the Internet is Hiding from You* (Elyse Cheney Literary Associates, 2011)。（邦訳：井口耕二訳、「閉じこもるインターネット」、早川書房、2012 年）
15）「Politics and Technology テクノロジーは民主主義の敵か？」、『MIT テクノロジーレビュー Special Issue』Vol. 11（2018 年）。
16）田中辰雄・浜屋敏、「ネットは社会を分断するのかーパネルデータからの考察ー」、『富士通総研 経済研究所 研究レポート』No. 462（2018 年）。

17）「特集:フェイクニュース」、『DIAMOND ハーバード・ビジネス・レビュー』2019 年 1 月号、pp. 16-82。
18）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。
19）服部宏充・栗原聡、「エージェント研究におけるシミュレーション」、『人工知能』（人工知能学会誌）28 巻 3 号（2013 年 5 月）、pp. 412-417。
20）水野淳太・田仲正弘・大竹清敬・呉鍾勲・Julien Kloetzer・橋本力・鳥澤健太郎、「大規模情報分析システム WISDOM X, DISAANA, D-SUMM」、『言語処理学会第 23 回年次大会発表論文集』（2017 年）、pp. 1077-1080。
21）牧野貴樹・澁谷長史・白川真一（編著）、他 19 名共著、『これからの強化学習』（森北出版、2016 年）。
22）藤巻遼平・村岡優輔・伊藤伸志・矢部顕大、「予測から意思決定へ ～予測型意思決定最適化～」、『NEC 技報』69 巻 1 号（2016 年）、pp. 64-67。
23）藤田桂英・森顕之・伊藤孝行、「ANAC: Automated Negotiating Agents Competition（国際自動交渉エージェント競技会）」、『人工知能』(人工知能学会誌) 31 巻 2 号 (2016 年 3 月)、pp. 237-247。
24）桑子敏雄、『社会的合意形成のプロジェクトマネジメント』（コロナ社、2016 年）。
25）伊藤孝行・藤田桂英・松尾徳朗・福田直樹、「エージェント技術に基づく大規模合意形成支援システムの創成　―自動ファシリテーションエージェントの実現に向けて―」、『人工知能』（人工知能学会誌）32 巻 5 号（2017 年 9 月）、pp. 739-747。
26）水野淳太・Eric Nichols・渡邉陽太郎・村上浩司・松吉俊・大木環美・乾健太郎・松本裕治、「言論マップ生成技術の現状と課題」、『言語処理学会第 17 回年次大会発表論文集』（2011 年）、pp. 49-52。
27）岡崎直観、「自然言語処理による議論マイニング」、『人工知能学会全国大会（第 32 回）』(2018 年）1D2-OS-28a-a（OS-28 招待講演）。 https://www.slideshare.net/naoakiokazaki/ss-100603788
28）柳井孝介・小林義行・佐藤美沙・柳瀬利彦・三好利昇・丹羽芳樹・池田 尚司、「AI の基礎研究: ディベート人工知能」、『日立評論』98 巻 4 号（2016 年 4 月）、pp. 61-64。
29）Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cam-eron Marlow, Jaime E. Settle and James H. Fowler, “A 61-million-person experiment in social influence and political mobilization”, *Nature* Vol. 489 (12 September 2012), pp. 295–298. DOI: 10.1038/nature11421
30）Tsubasa Tagami, Hiroki Ouchi, Hiroki Asano, Kazuaki Hanawa, Kaori Uchiyama, Kaito Suzuki, Kentaro Inui, Atsushi Komiya, Atsuo Fujimura, Hitofumi Yanai, Ryo Yamashita and Akinori Machino, “Suspicious News Detection Using Micro Blog Text”, *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation* (PACLIC 32, Hong Kong, Dec 1-3, 2018).
31）横尾真、「メカニズムデザインと最適化」、『論理と推論の理論・実装・応用に関する合同セミナー』招待講演（2013 年）。 http://www.edu.kobe-u.ac.jp/istc-tamlab/cspsat/pdf-open/cspsat2_seminar06_yokoo.pdf (accessed 2019-02-15)

32）John Hagerty, “2017 Planning Guide for Data and Analytics”, *Gartner Technical Professional Advice* (13 October 2016). https://www.gartner.com/binaries/content/assets/events/keywords/catalyst/catus8/2017_planning_guide_for_data_analytics.pdf (accessed 2019-01-31)
33）Jeff Conklin and Michael L. Begeman, “gIBIS: a hypertext tool for exploratory policy discussion”, *Proceedings of the 1988 ACM conference on Computer-supported cooperative work* (CSCW '88: Portland, USA, September 26-28, 1988), pp. 140-152. DOI: 10.1145/62266.62278
34）James S. Fishkin, *When the People Speak: Deliberative Democracy and Public Consultation* (Oxford University Press, 2011).（邦訳：岩木貴子訳、曽根泰教監修、『人々の声が響き合うとき：熟議空間と民主主義』、早川書房、2011 年）
35）Mark Klein, “Enabling Large-Scale Deliberation Using Attention-Mediation Metrics”, *Computer Supported Cooperative Work* Vol. 21, No. 4-5 (2012), pp. 449-473. DOI: 10.2139/ssrn.1837707
36）Thomas W. Malone, *Superminds: The Surprising Power of People and Computers Thinking Together* (Little, Brown and Company, 2018).
37）伊美裕麻・伊藤孝行・伊藤孝紀・秀島栄三、「オンラインファシリテーション支援機構に基づく大規模意見集約システム COLLAGREE －名古屋市次期総合計画のための市民議論に向けた社会実装」、『情報処理学会論文誌』56 巻 10 号（2015 年）、pp. 1996-2010。
38）“Special Issue: This is Watson”, *IBM Journal of Research and Development* Vol. 56 issue 3-4 (May-June 2012).
39）佐藤健、「論理に基づく人工知能の法学への応用」、『コンピュータソフトウェア』（日本ソフトウェア科学会誌）27 巻 3 号（2010 年 7 月）、pp. 36-44。DOI: 10.11309/jssst.27.3_36
40）日本学術会議公開シンポジウム「AI による法学へのアプローチ」（2019 年 1 月 24 日）。http://research.nii.ac.jp/~ksatoh/ai-law-symposium/ (accessed 2019-01-31)
41）産業競争力懇談会（COCN）、「人工知能間の交渉・協調・連携」、『産業競争力懇談会 2017 年度プロジェクト最終報告』（2018 年 2 月 21 日）。

２．１．６　データに基づく問題解決

（１）研究開発領域の定義

人工知能（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等による、問題解決手段の質的変化、産業構造・社会システム等の転換を生み出す研究開発領域である。

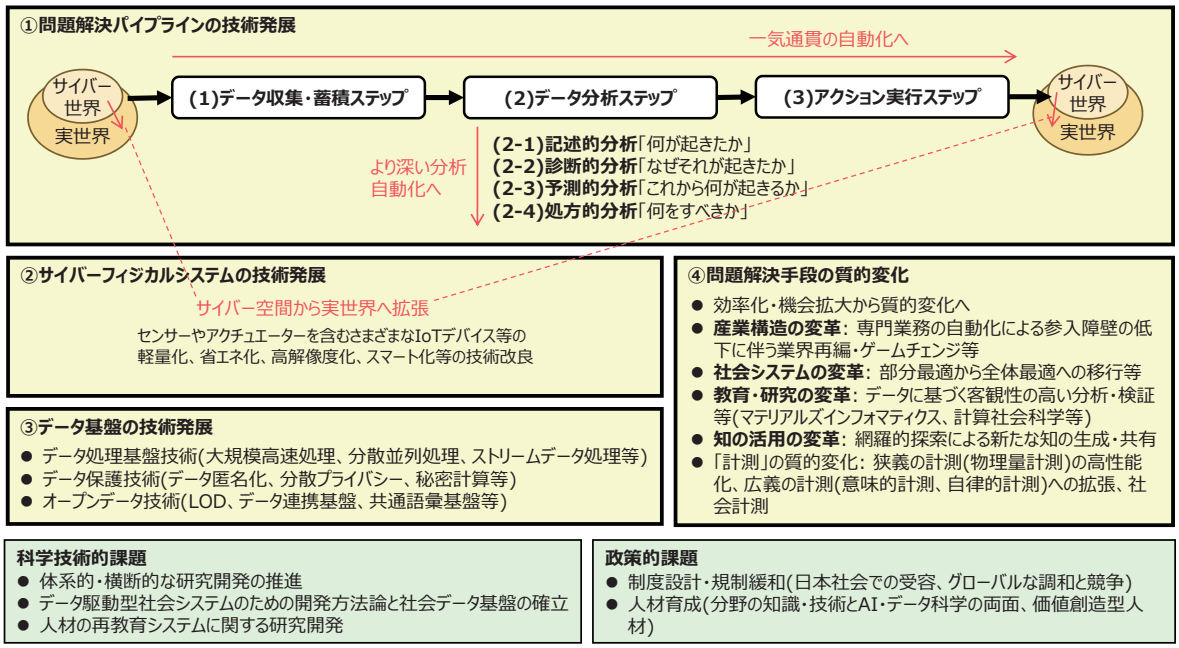


図2-1-8　領域俯瞰：データに基づく問題解決

（２）キーワード

ビッグデータ、データ駆動型社会、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、計測

（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・リアルタイム性等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等のサーチエンジンにおける検索連動型広告や、Amazon等のショップ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上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。たとえば、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[3)]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、エネルギー等、さまざまな分野で新たな価値創出が進められている。

［研究開発の動向］

「データに基づく問題解決」の基本的な枠組みに着目して、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、という3面から述べる。次に、「データに基づく問題解決」がもたらす4つの変革（産業の変革、社会の変革、教育・研究の変革、知の活用の変革）の動向について、特に問題解決手段の質的変化がもたらされる可能性について触れる。

① 問題解決パイプラインの技術発展

「データに基づく問題解決」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・アドバイザリー企業である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[4)]。(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、たとえば (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ればよい）。

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（そのための具体的な技術内容は「2.1.1 機械学習」等を参照）

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、高解像度化、スマート化等の技術改良が進められている（その具体的な技術内容は「2.4.6 IoT アーキテクチャ」を参照）。

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データ基盤にも取り組まれてきた。ここでいうデータ基盤は、データ処理基盤技術、データ保護技術、オープンデータ技術を含む。

データ処理基盤技術は、大規模なデータを高速に処理するための技術群である。ますます大規模なデータを、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等が発展している（その具体的な技術内容は「2.4.4 データ処理基盤」を参照）。

データ保護技術は、分析対象となるデータの保護のための技術群で、暗号化等のセキュリティー技術に加えて、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型データ分析技術が取り組まれている。データ匿名化、分散プライバシー、秘密計算等の技術がある（それらの詳細は「2.1.8 社会における AI」を参照）。

一方、オープンデータ（Open Data）は、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る。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（メタデータ等）の記述形式等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）[5] 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、3 年以内に整備、5 年以内に本格稼働の計画で、共通語彙基盤（Infrastructure for Multilayer Interoperability : IMI）の構築等が進められている[6),7)]。IMI、NIEM、SEMIC の間の国際的な相互運用性も検討されている。

④ 問題解決手段の質的変化

AI・ビッグデータ技術を活用した「データに基づく問題解決」の枠組みによって、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等が可能になり、問題解決手段の質的変化が起きる。これが「1.1.1 社会の要請、ビジョン」にあげた 4 つの変革（産業構造の変革、社会システムの変革、教育・研究の変革、知の活用の変革）につながる。

AI・ビッグデータ技術を活用した「データに基づく問題解決」の枠組みは、既に多くの業種・分野に広がっている。ここでそれらを列挙することは省くが（それらの詳しい内容は「AI 白書 2019」[8)] 等を参照）、さまざまな種類、多数の事例が生まれている。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・社会システム等の転換のような大きな質的変化も生まれる。効率化（コスト削減等）と機会拡大（売上拡大等）は従来の土俵の上での競争だが、この質的変化は土俵を変える（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、4 つの変革のそれぞれについて、質的変化の可能性に着目する。

1 つめの産業構造の変革は、人工知能技術戦略会議がまとめた産業化ロードマップ[9)] 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 3 つのフェーズでとらえている。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 は概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変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。その結果、業界再編・ゲームチェンジが起こり得る。

2 つめの社会システムの変革は、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、質的変化が生まれると考えられる。このような社会システムデザインに関わる技術群や研究開発課題については「2.3 社会システム科学」で取り上げている。なお、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・交通等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。

3つめの教育・研究の変革では、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り、質的変化が起きる。マテリアルズインフォマティクス（Materials Informatics：MI）[10), 11)]、計算社会科学（Computational Social Science）[12)] のアプローチ等がその一例である。

4つめの知の活用の変革では、コンピューターを用いた網羅的な可能性の探索から、それまで人間が気づいていなかった新たな知が生まれ得る。たとえば、囲碁の世界において、コンピューターソフトウェアであるAlphaGo[13)]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かもしれない[14)]。

（4）注目動向

［新展開・技術トピックス］

① マテリアルズインフォマティクス（MI）の展開[10), 11)]

MIは、材料科学とAI・データ科学を融合し、新材料や代替材料等の材料開発を大幅に効率化するデータ駆動型材料開発の取り組みである。従来の材料開発は、勘と経験で新たな材料を合成し、その材料の特性を調べることによって進められてきた。これに対してMIでは、材料に関するデータベース構築、シミュレーション、機械学習や統計解析（スパースモデリング等）を用いて、必要な特性を実現し得る材料を探索する（従来と逆問題）。

2011年に米国でMaterials Genome Initiativeが発足したことから注目され、欧州でも関連コンソーシアムが立ち上がる等、世界的な動きが起こっている。日本でも、2015年に物質・材料研究機構（National Institute for Materials Science：NIMS）を中心とする推進体制が構築された。JSTイノベーションハブ構築事業として、NIMSを中心とした情報統合型物質・材料開発イニシアティブMI2I（Materials research by Information Integration Initiative、2015年～）が立ち上がり、内閣府戦略的イノベーション創造プログラムSIPで「革新的構造材料」（2014年～2018年）が実施された。

MIを活用した具体的な成果としては、マサチューセッツ工科大学（MIT）とサムスン電子によるLiイオン電池の固体電解質材料の発見[15)]を契機とし、日本でも京都大学とシャープの産学協同研究による2次電池材料の開発にMIが活用された[16)]。NECが熱電材料分野でスピン流熱電変換素子向け材料として、それまでの数百倍の性能（出力密度）を持つ新材料をMIを用いて開発した[17)]のをはじめ、日立・富士通等を含め、ここ数年で産業界での取り組みが急速に活性化している。

②「計測」の質的変化

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、現在の状況を把握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、計測は、AI・ビッグデータ技術と結びついて、その概念を広げている。いわば「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・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・重量等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・データ分析処理等の情報処理（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、(1) 物理量計測の高性能化、(2)「意味的計測」、(3)「自律的計測」が可能になってきた[18]。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 (1)(2)(3) について簡単に説明する（詳細は調査報告書[18]にまとめた）。

(1) 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。たとえば、カメラ画像の超解像（画像処理によって解像度を高める）等がある。

(2)「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・顔認証等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。

(3)「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。たとえば、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。

また、上記 (1)(2)(3) は計測の高次化であるが、物理量計測を出発点としている。このような物理量計測に加えて、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[12]も、近年、取り組みが活発になっている。

［注目すべき国内外のプロジェクト］

AI による科学的発見というグランドチャレンジ

北野宏明（システム・バイオロジー研究機構、ソニーコンピュータサイエンス研究所、他）が、2016 年に AI による新たなグランドチャレンジを立ち上げた[14]。そのチャレンジは「2050 年までに、ノーベル賞かそれ以上の科学的発見を行う人工知能を開発する」ことを目標とするものであり、彼自身は特に生命科学分野をターゲットとして取り組んでいる。仮説として「科学的発見は、仮説空間の網羅的探索と検証によってなされる」と考えており、上述の MI と通じるところがあるが、ノーベル賞級の科学的発見を行うために道具として AI を使うという立場ではなく、科学的発見の本質を解明し、人間の知識の飛躍的拡大をサポートする AI 技術の実現によって、人類が直面する問題の解決につなげることを狙った取り組みである。

（５）科学技術的課題

① 体系的・横断的な研究開発の推進

AI・ビッグデータ技術の適用はさまざまな分野に広がっているが、前述した 4 つの変革に向けた体系的な研究開発、特に質的変化とそれから派生する問題への対応は必ずしも十分でない。また、社会、産業、教育・研究、知の活用と広い分野に及び、分野横断かつ多様な立場か

らの研究開発が不可欠である。

② 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

「データに基づく問題解決」は、データ駆動型の社会システムの開発方法論でもある。「2.3 社会システム科学」において関連する動向をまとめているほか、サイバーフィジカルシステムの開発方法論としての Reality 2.0[19] 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[20] もかかわりが深い。

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、「データに基づく問題解決」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へ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。

③ 人材の再教育システムに関する研究開発

社会や産業の質的変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[21] 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。社会や産業の質的変化に伴い、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、その変化がはやいため、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等）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、より的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 制度設計・規制緩和

社会・産業等の質的変化に向けて、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。

② 人材育成

データ科学のアプローチや AI・ビッグデータ技術がさまざまな技術分野・産業分野に取り入れられるようになり、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わかる人材・組織の育成が重要になっている。また、「データに基づく問題解決」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。価値創造型人材の育成も重要である。

（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	○	→	各分野で取り組み強化されているが、人材・資金や体系的な取り組みの面ではまだ十分でない。
	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・制度設計等には課題がある。
米国	基礎研究	◎	→	適切なグランドチャレンジの設定等、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。
	応用研究・開発	◎	↗	GAFA 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・産業等の変革を牽引している。
欧州	基礎研究	○	→	
	応用研究・開発	○	→	
中国	基礎研究	○	→	
	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	△	→	
	応用研究・開発	○	→	

（8）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）内　閣　府、「Society 5.0」https://www8.cao.go.jp/cstp/society5_0/index.html (accessed 2019-02-01)
4）John Hagerty, “2017 Planning Guide for Data and Analytics”, Gartner Technical Pro-fessional Advice (13 October 2016). https://www.gartner.com/binaries/content/assets/events/keywords/catalyst/catus8/2017_planning_guide_for_data_analytics.pdf (accessed 2019-02-01)
5）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。
6）内閣官房 IT 総合戦略室、「分野間データ連携基盤の整備」、平成 30 年度第 1 回情報共有基盤推進委員会（2018 年 8 月 3 日）資料 1　https://imi.go.jp/contents/2018/08/t20180810_1.pdf (accessed 2019-02-01)
7）加藤文彦・武田英明・田代秀一・平本健二・松澤有三、「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）。
8）情報処理推進機構 AI 白書編集委員会（編）、『AI 白書 2019 －企業を変える AI、世界と日本の選択』（KADOKAWA、2018 年）。
9）人工知能技術戦略会議、「人工知能技術戦略」（2017 年 3 月 31 日）https://www.nedo.go.jp/content/100862413.pdf (accessed 2019-02-01)
10）加藤幸一郎、『材料開発の新潮流 ～材料科学とデータ科学の融合：Materials Informatics ～』、みずほ情報総研レポート Vol. 15（2018 年）。
11）「特集：待ったなし！ AI で材料開発」、『日経エレクトロニクス』2018 年 10 月号。
12）笹原和俊、「私のブックマーク：計算社会科学（Computational Social Science）」、『人工知能』（人工知能学会誌）30 巻 6 号（2015 年 11 月）、pp. 856-859。

13）David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel and Demis Hassabis, “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

14）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。

15）Yifei Mo, Shyue Ping Ong and Gerbrand Ceder, “First Principles Study of the $Li_{10}GeP_2S_{12}$ Lithium Super Ionic Conductor Material,” *Chemistry of Materials* Vol. 24, No. 1 (2012), pp. 15-17. DOI: 10.1021/cm203303y

16）Motoaki Nishijima, Takuya Ootani, Yuichi Kamimura, Toshitsugu Sueki, Shogo Esaki, Shunsuke Murai, Koji Fujita, Katsuhisa Tanaka, Koji Ohira, Yukinori Koyama and Isao Tanaka, “Accelerated discovery of cathode materials with prolonged cycle life for lithium-ion battery,” *Nature Communications* Vol. 5, Article No. 4553 (2014), pp. 1-7. DOI: 10.1038/ncomms5553.

17）石田真彦・岩崎悠真・澤田亮人・桐原明宏・白根昌之、「スピン流熱電変換 〜インフォマティクスを活用した材料開発と適用領域〜」、『NEC 技報』71 巻 1 号（2018 年）、pp. 92-96. https://jpn.nec.com/techrep/journal/g18/n01/pdf/180119.pdf (accessed 2019-02-18)

18）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）。

19）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）。

20）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）。

21）Carl Benedikt Frey and Michael A. Osborne, “The Future of Employment: How Sus-ceptible are Jobs to Computerisation?”, (September 17, 2013). https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf (accessed 2019-01-18)

2.1.7　計算脳科学

（1）研究開発領域の定義

脳を情報処理システムとしてとらえて、脳の機能を調べる研究分野である。計算論的神経科学（Computational Neuroscience）とも称される。視覚の計算理論等で知られる David Marr は、情報処理システムを理解するにあたって、(A) 計算理論、(B) 表現とアルゴリズム、(C) ハードウェアという 3 つの水準を併存させた理解が重要であると述べているが [1]、脳という情報処理システムについて、(A) の明確化を行うことで、(A)(B)(C) の 3 つのレベルの理解を相互に深め、脳の情報処理の機能を理解しようとするのが計算脳科学である。

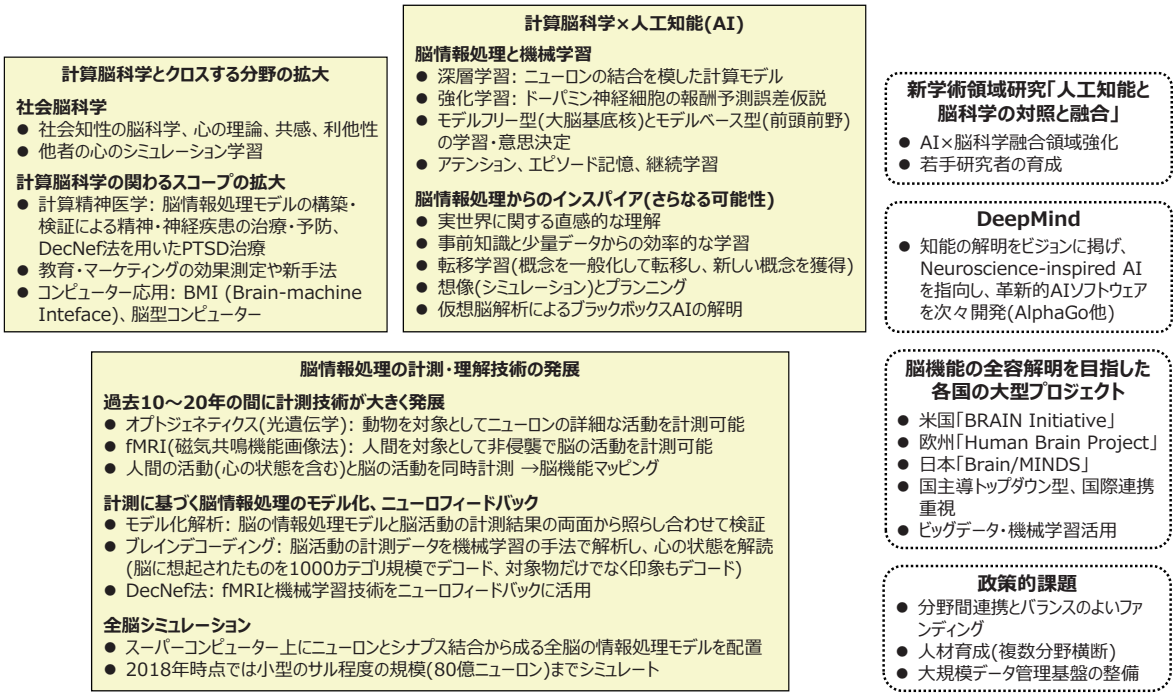


図2-1-9　領域俯瞰：計算脳科学

（2）キーワード

計算脳科学、計算論的神経科学、脳情報処理、脳活動計測、ブレインデコーディング、モデル化解析、ニューロフィードバック、深層学習、強化学習、社会知性、社会脳科学、計算精神医学、全脳シミュレーション

（3）研究開発領域の概要

[本領域の意義]

計算脳科学は人工知能（AI）の研究発展にさまざまな形で貢献し得る。たとえば、AI を創るために、人間の脳で行われている情報処理のメカニズムを知ることは、より高度な機能や高い処理性能を実現する方式のヒントになる。また別な面では、AI（あるいはその要素技術を組み込んだシステム）と人間がインタラクションをする際に、人間（特にその脳）の応答パターンを知ることは、より良いインタラクションを設計・評価することにつながる。

現在のAIブームを牽引している深層学習（Deep Learning）[2]は、脳を構成するニューロン（Neuron：神経細胞）の結合を模した計算モデルをベースとしている。深層学習は、計算脳科学の成果に基づき、画像認識・音声認識等のパターン認識の機能において、さまざまな条件下で、既に人間を上回る認識精度を達成するようになった。これらの成果は素晴らしいものであるが、同時にこれは脳の物体認識や知覚機能の実現に対応しているにすぎない。脳は、他にも運動・認知・言語・感情・意識等の優れた情報処理機能を実現しており、AI研究が脳研究から得られることはまだまだ多い。たとえば、深層学習は大量の学習データを必要とするのに対して、人間は比較的少量のデータからでも学習できている。また、深層学習は大きな計算パワー（消費電力）を必要とするのに対して、人間の脳の消費電力は約20ワット（薄暗い電球程度）である。これらは、計算脳科学がAIの研究発展に大きく貢献してきたこと、および、これからもさらに貢献し得ることを示す一例である。

［研究開発の動向］

① 脳情報処理の計測・理解技術の発展

過去10～20年の間に、脳の機能・活動を知るための計測技術は大きく発展した。

その一つは、動物を対象として、ニューロンの詳細な活動を調べることができるオプトジェネティクス（Optogenetics：光遺伝学）[3]である。これは、光によって活性化されるタンパク分子を遺伝学的手法で特定の細胞に発現させ、その機能を光で操作する技術である。従来の電気刺激を用いる手法や薬理学的手法では難しかったレベルの高い分解能が得られ、マイクロ～ミリ秒オーダーのタイムスケールで特定の神経活動のみを制御できるようになった。たとえば、マウスを使った実験結果によると、記憶をスイッチしたり[4]、誤りの記憶を形成したり[5]といった操作が行える。Nature Method誌が科学全分野の中から選ぶ「Method of the Year 2010」に選定されたことが、この技術が画期的であったことを示している。

もう一つは、人間を対象として、非侵襲で脳の活動を調べることができるfMRI（Functional Magnetic Resonance Imaging, : 磁気共鳴機能画像法）[6]である。fMRIは、神経活動に伴う血管中の血液の流れ（血流量）や酸素代謝の変化を、磁気共鳴画像装置（MRI）を用いて計測・可視化する技術である。人間の脳の活動を頭皮の外から測定する方法として、従来は脳波測定法や陽電子を用いるPET（ポジトロン断層映像法）があったが、これらに比べてfMRIは空間分解能が高く、PETの課題である被爆の心配もない。大きな病院に普及している臨床用の通常のMRI装置を活用できるという経済性もあり、fMRIは1990年代初頭に考案された後、急速に普及し、人間の高次の脳機能を調べるために活用されるようになった。

fMRIによって、人間の行動（心の状態を含む）と脳の活動の同時計測が可能になり、どのような行動や心の状態のときに、脳のどの部分が深く関わっているのか（脳機能マッピング）が調べられるようになった。さらに、マッピングだけでなく、脳の情報処理のモデル化が考えられるようになった。モデル化の主なアプローチとして、モデル化解析（Model-based Analysis）、ブレインデコーディング（Brain Decoding）等がある。

モデル化解析[7]では、脳の情報処理モデルを行動と脳の活動の両面から検証する。そのため、まず複数考えられる仮説について、それぞれの処理モデルがどれだけ行動データを説明できるかを調べる。さらに、この行動データへのフィッティングを通してモデルの自由パラメーターを推定する。その結果から脳のどの部分での活動かを導出し、脳の活動データと照らし合わせ

て検証する。

ブレインデコーディングは、fMRI 等によって計測された人間の脳の活動データを、機械学習の手法を用いて解析することで、人間の心の状態を解読しようとする技術である。当初 2005 年頃は、fMRI の計測データのパターンと、少数のカテゴリーとの間の対応関係を学習するものであった [8]。その後、深層学習や分散表現（Word2Vec）等の機械学習の新たな手法も取り込み、脳に想起されたものを、1,000 を超える数のカテゴリーと対応付ける一般物体デコーデイング [9] や、対象物（名詞）やその動作（動詞）だけでなく、それらの印象（形容詞）のデコード [10] にも迫りつつある。

② 脳情報処理と機械学習

機械学習を中心とする AI 技術の発展は、上記①の発展を通して明らかになってきた脳情報処理の (A) 計算理論や (B) 表現とアルゴリズムと、結びつきが強いものになっている。前述した通り、深層学習は脳を構成するニューロンの結合を模した計算モデルをベースとしている。さらに、強化学習、アテンション、エピソード記憶、作業記憶、継続学習等についても、脳情報処理の知見・発見との結びつきが強いことが知られている [11]。

脳情報処理への関心が触媒となっている機械学習の手法は大変多い。深層学習はもとより、その源流であるニューラルネットワーク・誤差逆伝搬法・自己組織化マップ・表現学習・独立成分解析・強化学習そして情報幾何などがある。これらの研究の発展において、脳科学と機械学習の両方で、日本の研究は大いに貢献してきている。

たとえば、強化学習（Reinforcement Learning）は、ドーパミン神経細胞の報酬予測誤差仮説によって、AI 研究における強化学習と脳の強化学習とが強く結びついている [12), 13)]。AI 研究における強化学習は、学習主体が、ある状態で、あるアクションを実行すると、ある報酬が得られるタイプの問題を扱い、より多くの報酬が得られるようにアクションを決定する意思決定方策を、アクション実行と報酬の受け取りを重ねながら学習していく機械学習アルゴリズムである。一方、中脳にあるドーパミン神経細胞は、報酬予測誤差（実際に得た報酬量と予測された報酬量との誤差）に基づいてドーパミンを放出し、これが大脳基底核に運ばれることで、脳における強化学習の学習信号として働くということがわかってきた。また、脳における学習・意思決定のプロセスにはモデルフリー型とモデルベース型があり、モデルフリー型では、刺激と反応の関係性を報酬の程度・確率に直結した形で学習し、モデルベース型では、刺激や反応の間の関係性を状態遷移等の内部モデルとして学習する。モデルフリー型は上述の大脳基底核、モデルベース型は大脳新皮質、特に前頭前野が重要な役割を果たしていると見られている。

このように、脳情報処理における科学的発見が AI 的手法の理論的な裏付けになるとともに、脳情報処理の知見を取り込むことが AI 技術の発展につながり得るという事例が、機械学習を中心に積み上げられつつある。

③ 社会脳科学

人間は社会の中で他者との関わりを持ちながら考え、行動している。このような社会行動の根幹には、人々が互いの心や振る舞いを推断するときに働かせる社会知性（Socio-intelligence）がある。この他者の行動を予測し、その予測を踏まえた意思決定をする脳機能は、しばしば「心の理論」（Theory of Mind）と呼ばれる。そこには、他者の気持ち・感情を感じ取る能力である「共

感」（Empathy）や、自分の利益のみにとらわれず他者の利益を図るように行動する性向である「利他性」（Altruism）も関わる。この社会知性の脳科学（社会脳科学）がこの 15 年ほどで著しい発展を見せている。

この計算理論は、②で触れた脳の強化学習の計算理論をベースに発展させたものが考えられており、①で述べた fMRI による計測とモデル化解析の手法を用いて、脳計算モデルの検証が行われている[7), 12)]。この脳計算モデルでは、自己の行動選択を報酬予測誤差信号に基づいて学習することに加えて、同様のプロセスが他者の心の中でも行われているというシミュレーションを自己の心の中で行って学習する。この他者の心のシミュレーション学習は、シミュレーションにおける他者報酬予測誤差信号だけでなく、他者の観察から得られる他者の行動予測と実際の行動との差を示す他者行動予測誤差信号も用いたハイブリッドな構成で行われていることが明らかになってきた。

（4）注目動向

［新展開・技術トピックス］

① Decoded Neurofeedback

前述のブレインデコーディングを発展させ、fMRI と機械学習技術をニューロフィードバックに活用した DecNef（Decoded Neurofeedback）法[14)]が国際電気通信基礎技術研究所（ATR）によって開発された。ニューロフィードバックは、特定の脳領域の活動を被験者にフィードバックし、被験者自身による脳活動の操作を促すことによって、その領域に対応した認知機能の増進や補綴を誘導するアプローチである。この DecNef 法は、fMRI と機械学習によるブレインデコーディング結果を被験者にリアルタイムにフィードバックすることで、従来に比べて細かい脳領域の操作を可能にした。特定の刺激に対して感覚が鋭くなる知覚学習の因果関係も示され、つらい記憶を思い出すことなく消すことの可能な、心的外傷後ストレス障害（PTSD）の新しい治療法につながる可能性も見いだされた[15)]。

② 計算精神医学

計算脳科学による脳情報処理モデルの構築・検証は、精神・神経疾患の解明や治療・予防にも貢献すると期待されるようになってきた。上述の DecNef 法を用いた PTSD 治療の可能性もその一例である。また、前述のモデル化解析を用いて、精神・神経疾患の患者（社会的不適応が認められる意思決定を伴う）と健常者の意思決定の間で、どの脳計算ステップや内的変数の扱いに違いが表れるのかを解明するアプローチも検討されている[7)]。現代社会におけるさまざまなストレス要因や高齢化社会で増加する認知症への対策も含め、精神・神経疾患の解明や治療・予防に向けて、計算脳科学と精神医学を融合した計算精神医学は重要性を増している。

③ 全脳シミュレーション

前述のような脳情報処理の計測技術の発展と脳計算モデルの理解の深まりによって、スーパーコンピューターを用いた全脳シミュレーションへの取り組みが進められるようになった。ニューロンやシナプス結合等で構成される全脳の情報処理モデルをスーパーコンピューター上に配置し、その振る舞いのシミュレーションを行い、その実行結果と、実際に全脳の活動を計測した結果とを比較することで、脳のより深く正確な理解が可能になる。さらに、パーキンソン

病・てんかん・うつ病を含む多くの脳疾患は、複数の脳領域が直接的・間接的に影響し合っていると言われており、そのような脳疾患の解明には、全脳シミュレーションのアプローチが有効と考えられている。

具体的な成果として、2013 年に日本とドイツの共同研究チーム（理化学研究所、ユーリッヒ研究所、沖縄科学技術大学院大学）によって、「京」コンピューターと NEST シミュレータを用いた大脳皮質神経回路シミュレーション[16]で、17.3 億個のニューロンと 10.4 兆個のシナプスのシミュレーション実行が確認された。2018 年には電気通信大学のプロジェクトにおいて、JAMSTEC の暁光システム（PEZY-SC：2048 コア、10000 チップ、28.19 PETA FLOPS）を用い、80 億の神経細胞からなる小脳モデルのリアルタイムシミュレーション[17]が実現され、視機性眼球反応に対応する神経活動の再現が確認された。これらのニューロン規模は小型のサル程度（マーモセット：約 6 億個、ヨザル：約 14 億個、マカクザル：約 63 億個）に相当する。人間は約 860 億個と言われており、「京」の 100 倍の性能を持つ次世代機「ポスト京」で人間の全脳シミュレーションを目指すプロジェクト（ポスト京 萌芽的課題 4「思考を実現する神経回路機構の解明と人工知能への応用」、2016 年 8 月～ 2020 年 3 月、沖縄科学技術大学院大学・京都大学・理化学研究所・電気通信大学・東京大学）が進められている。

[注目すべき国内外のプロジェクト]

① 脳機能の全容解明を目指した各国の大規模プロジェクト

2013 年 ～ 2014 年 に、 米 国 で は The Brain Research through Advancing Innovative Neurotechnologies（BRAIN）Initiative、欧州では Human Brain Project（HBP）、日本では「革新的技術による脳機能ネットワークの全容解明プロジェクト」（Brain Mapping by Integrated Neurotechnologies for Disease Studies：Brain/MINDS、革新脳）という脳科学研究の大型プロジェクトが相次いで立ち上がった。BRAIN Initiative はアポロ計画やヒトゲノム計画に匹敵する巨大科学プロジェクトとして構想されたと言われるが、いずれも脳機能の全容解明に向けて、国主導のトップダウン型で、国際連携にも重点を置いたプロジェクト推進が必要という共通的な認識がある。一方、米国の BRAIN Initiative は技術開発、欧州の HBP は計算論に基づいた脳のモデル化、日本の Brain /MINDS は霊長類モデルを活用したマップ作成等、各国の取り組みの特色も出されている。前述のように、オプトジェネティクスや fMRI 等の革新的な計測技術や、ビッグデータ解析・機械学習技術の進化が、脳機能の可視化の可能性を飛躍的に高めたことが、脳機能の全容解明を目指す方向性につながっており、これらのプロジェクトの中でも、脳情報処理の理論やデータ解析といった計算脳科学の側面は重きが置かれている。

② 新学術領域研究「人工知能と脳科学の対照と融合」

国内においては、上記①にあげた Brain/MINDS プロジェクト（革新脳）のほか、脳科学研究戦略推進プログラム（脳プロ）、戦略的国際脳科学研究推進プロジェクト（国際脳）等も推進され、脳科学研究の強化が図られている。そういった中、特に計算脳科学にフォーカスしたものとして、新学術領域研究「人工知能と脳科学の対照と融合」（研究期間：2016 年 6 月 30 日～ 2021 年 3 月 31 日）がある。「予測と知覚」、「運動と行動」、「認知と社会性」という 3 つの研究課題を掲げ、人工知能と脳科学の主要研究者を集め、融合研究を推進している。さらに、サマースクールやハッカソン等によって、融合領域の若手研究者の育成にも力を入れている。

③ DeepMind

DeepMind は Demis Hassabis らが創業した英国企業で、2010 年に創業され、2014 年に Google に買収された。革新的な機械学習技術を組み込んだソフトウェアを次々に開発し、2015 年にコンピューターゲーム（Atari 2600）のプレイ方法を自力で学習して高い成績を達成することを示し、2016 年・2017 年には囲碁ソフトウェア AlphaGo が世界トップランク棋士に圧勝して大きな話題となった。2018 年にはタンパク質の折り畳みを解析するソフトウェア AlphaFold が、アミノ酸の配列からタンパク質の構造を予測する CASP（Critical Assessment of Structure Prediction）競技会において、他を圧倒する世界トップ精度を達成した。

DeepMind は AI・機械学習のスタートアップとして注目されているが、「知能の解明」を企業ビジョンとして掲げており、Demis Hassabis 自身は脳科学研究での高い実績も有する（海馬とエピソード記憶に関する研究成果で Science 誌による 2007 年 10 大ブレイクスルーの一つに選ばれた）。2017 年には DeepMind のメンバーと「Neuroscience-Inspired Artificial Intelligence」と題した論文 [11] を Neuron 誌に発表し、脳科学を重視した AI への取り組み姿勢を示した。AlphaGo 等で深層学習・強化学習の先進的活用事例が知られているが、DeepMind は計算脳科学の有効性・可能性を示しているとも言える。

（5）科学技術的課題

① 脳情報処理の計測・理解技術のさらなる革新と脳の多階層な構造・機能の解明

前述のように、オプトジェネティクス、脳波測定法、PET、fMRI 等、脳の活動を計測する技術が発展し、低侵襲・非侵襲化、分解能向上が図られてきた。これにビッグデータ解析・機械学習技術を組み合わせて、ブレインデコーディング、モデル化解析、ニューロフィードバック・DecNef 法等、脳情報処理をより深く理解する手段も生み出されてきた。

このような計測・理解技術に基づく脳の構造・機能の解明は、個々のニューロンや脳内各部の神経回路といったミクロなレベルから、脳全体の活動をとらえるマクロなレベルまで、さまざまな階層で進められてきた。それら多階層の成果を統合し、脳情報処理を総合的に解明していく取り組みが今後いっそう重要になっていく。そのために、多階層でビッグデータを蓄積していくことや、前述した全脳シミュレーションのためのコンピューティング基盤の研究開発も重要である。

② AI 研究課題に対する計算脳科学からの貢献

前述したように、脳情報処理の知見・発見にインスパイアされた AI 研究トピックとして、深層学習、強化学習、アテンション、エピソード記憶、作業記憶、継続学習等があげられる。今後さらに脳情報処理の知見・発見が AI 研究の発展に貢献していくと期待され、たとえば Demis Hassabis ら [11] は具体的に以下のような AI 研究課題をあげている。

- 実世界に関する直感的な理解：現状の AI は画像・映像から個々の物体に分解して認識し、それらの関係を組み立てようとする。それに対して人間は、幼児の段階から実世界における空間・数・物理感等を直感的に把握できる。
- 効率的な学習：現状の AI は大量の学習データを必要とする。それに対して人間は、事前知識と少量のデータから新しい概念を獲得することができる。

- 転移学習（Transfer Learning）：人間は概念を一般化し、転移して、新しい概念を獲得することができる（たとえば、ある乗り物の運転をできる人は、類似する乗り物の運転もできる）が、現状の AI ではまだ難しい。
- 想像とプランニング：人間は将来を想像・シミュレーションして、状況変化があっても柔軟に行動を選択できる。それに対して現状の AI は、過去の傾向の延長線上での行動に縛られている。
- 仮想脳解析（Virtual Brain Analytics）：計算脳科学が脳情報処理を解明しようとしているように、同様の手法・考え方を用いて、ブラックボックス化している AI システムを解明する試みが考えられる。

③ 計算脳科学とクロスする分野の拡大

脳の活動はさまざまな人間の活動の根幹であるから、計算脳科学の研究成果・知見はさまざまな分野に波及し得る。前述した社会脳科学では、他者との関係、社会知性を扱っており、個々の人間の意思決定のメカニズムから社会活動のメカニズムにまで、計算脳科学の関わるスコープは広がる。精神・神経疾患の解明や治療・予防にも貢献する計算精神医学の動向については既に触れたが、病気の治療法、さらには教育・マーケティング等の効果測定や新手法に貢献している。コンピューター応用においても、新しいインターフェースとしての BMI（Brain-machine Interface）、新しい計算メカニズムとしての脳型コンピューター等が検討されている。

（6）その他の課題

① 分野間連携とバランスのよいファンディング

脳の情報処理メカニズムはいまだ未知の部分が多く、その解明には長期的な基礎研究の継続が不可欠である。その一方で、コンピューターに実装され、さまざまな応用・ビジネスへと展開が進んでいる深層学習・強化学習技術は、脳の情報処理メカニズムとの関係が深い。また、脳の機能や情報処理メカニズムの理解には、認知科学・心理学等も関係が深く、ELSI（Ethical, Legal and Social Issues：倫理的法的社会的課題）の面も考慮する必要がある。このような幅広い視点からの議論や分野間連携を促進するような研究プロジェクト体制も効果的である。長期的な基礎研究への継続投資を進めつつ、このような分野間連携の活動へもバランスよく研究投資していくことが重要である。

② 人材育成

上記①で述べたように、計算脳科学の研究には、複数分野横断の幅広い視野・知見を持った人材が必要であるが、現状はそのような人材が非常に少ない。研究プロジェクトにおいて、複数分野の研究者を 1 つの拠点で共同・交流させるような体制を作ることが望ましい。さらに、AI や計算機科学そして計算脳科学と脳科学を同時に学べるような、新たな大学院研究科・学部創設も検討すべきである。また、医学系の学生はもともと数学の素養が高いので、プログラミングや統計・数理・データ解析等、コンピューター科学を学び、活用する機会を継続的に設けることは有効と考えられる。

③ 大規模データ管理基盤の整備

脳活動の計測技術の進化や、脳科学研究の大型プロジェクトの実施を背景として、脳活動に関わる大規模データが取得・蓄積されるようになってきた。データ解析が研究発展への貢献も高まってきており、大規模データの保管・共有・効率的解析のための基盤整備が、今後の研究加速のために求められる。

（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	◎	→	fMRI 法、DecNef 法、京による全脳シミュレーション等、脳情報処理を計測・理解するための基本的手法の創出を主導してきた。国として脳科学の基礎研究プロジェクトを多階層で推進し、Brain/MINDS 等、国際的にも認知されている。
	応用研究・開発	○	→	米国と比べると、民間財団・ベンチャー企業での取り組みが相対的に弱い。
米国	基礎研究	◎	→	BRAIN Initiative をはじめ大型研究投資がなされており、分子細胞レベルからシステムレベルまで脳科学に関する層の厚い研究開発が進められている。
	応用研究・開発	◎	→	民間財団・ベンチャー企業での取り組みが活発で、基礎研究から応用への展開が円滑に進められる。大規模なデータベースやツール類の整備が進んでいる。
欧州	基礎研究	◎	→	Human Brain Project（HBP）で欧州連携の大型投資が進められている。英国 DeepMind が、脳科学に基づく先進的 AI 技術開発に取り組んでいる。
	応用研究・開発	◎	→	HBP では脳科学と情報科学の融合分野を強化しており、計算脳科学のコンピューティング基盤の整備も進んでいる。
中国	基礎研究	○	↗	第 13 次 5 カ年計画（2016 年～ 2020 年）で特に成長が見込まれる 5 分野の 1 つとして脳科学があげられ、15 年計画の（2016 年～ 2030 年）「Brain Science and Brain-inspired Intelligence」プロジェクトが立ち上げられた。上海の復旦大学が十数校および中国科学院（CAS）と脳科学協同イノベーションセンターを設立した。
	応用研究・開発	○	↗	中国は AI 分野の研究開発・ビジネスで米国と二強になりつつあり、脳科学を AI と連携させて強化する方針が打ち出されている。
韓国	基礎研究	○	→	韓国科学技術研究院（KIST）に機能的コネクトミクスセンターが設立された。さらに、Korean Brain Initiative が 10 年計画（2018 ～ 2027 年）でスタート、さまざまな階層での脳マップの作製や AI 関連研究等を推進。
	応用研究・開発	○	→	

（8）参考文献

1）David Marr, Vision: *A Computational Investigation into the Human Representation and Processing of Visual Information* (W. H. Freeman and Company, 1982).

2）麻生英樹・安田宗樹・前田新一・岡野原大輔・岡谷貴之・久保陽太郎・ボレガラダヌシカ（人工知能学会監修、神嶌敏弘編）、『深層学習—Deep Learning—』（近代科学社、2015 年）。

3）Karl Deisseroth, “Control the Brain with Light”, *Scientific American*, November 2010, pp. 49-55.

4）Roger L. Redondo, Joshua Kim, Autumn L. Arons, Steve Ramirez, Xu Liu, and Susumu Tonegawa, “Bidirectional reversal of the valence associated with the hippocampal memory engram”, *Nature* Vol. 513 (18 September 2014), pp. 426-430. DOI: 10.1038/nature13725

5）Steve Ramirez, Xu Liu, Pei-Ann Lin, Junghyup Suh, Michele Pignatelli, Roger L. Redondo, Thomas J. Ryan, and Susumu Tonegawa, “Creating a false memory in the hip-pocampus”, *Science* Vol. 341, Issue 6144 (26 July 2013), pp. 387-391. DOI: 10.1126/science.1239073

6）Seiji Ogawa, Tso-Ming Lee, Alan R. Kay and David W. Tank, “Brain magnetic resonance imaging with contrast dependent on blood oxygenation”, *Proceedings of the National Academy of Sciences of the United States of America* Vol. 87, No. 24 (December 1990), pp. 9868-9872.
7）中原裕之・鈴木真介、「意思決定と脳理論：人間総合科学と計算論的精神医学への展開」、『Brain and Nerve』65 巻 8 号（2013 年 8 月）、pp. 973-982。
8）Yukiyasu Kamitani and Frank Tong, “Decoding the visual and subjective contents of the human brain”, *Nature Neuroscience* Vol. 8 (24 April 2005), pp. 679-685. DOI: 10.1038/nn1444
9）Tomoyasu Horikawa and Yukiyasu Kamitani, “Generic decoding of seen and imagined objects using hierarchical visual features”, *Nature Communications* Vol. 8, Article number 15037 (22 May 2017). DOI: 10.1038/ncomms15037
10）Satoshi Nishida and Shinji Nishimoto, “Decoding naturalistic experiences from human brain activity via distributed representations of words”, *NeuroImage* Vol. 180, Part A (15 October 2018), pp. 232-242. DOI: 10.1016/j.neuroimage.2017.08.017
11）Demis Hassabis, Dharshan Kumaran, Christopher Summerfield and Matthew Botvinick, “Neuroscience-Inspired Artificial Intelligence”, *Neuron* Vol. 95, Issue 2 (19 July 2017), pp. 245-258. DOI: 10.1016/j.neuron.2017.06.011
12）中原裕之、「社会知性を実現する脳計算システムの解明：人工知能の実現に向けて」、『人工知能』（人工知能学会誌）32 巻 6 号（2017 年 11 月）、pp. 863-872。
13）田中慎吾・坂上雅道、「推移的推論の脳メカニズム―汎用人工知能の計算理論構築を目指して―」、『人工知能』（人工知能学会誌）32 巻 6 号（2017 年 11 月）、pp. 845-850。
14）Kazuhisa Shibata, Takeo Watanabe, Yuka Sasaki and Mitsuo Kawato, “Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation”, *Science* Vol. 334, Issue 6061 (09 Dec 2011), pp. 1413-1415. DOI: 10.1126/science.1212003
15）Ai Koizumi, Kaoru Amano, Aurelio Cortese, Kazuhisa Shibata, Wako Yoshida, Ben Seymour, Mitsuo Kawato and Hakwan Lau, “Fear reduction without fear: Reinforcement of neural activity bypasses conscious exposure”, *Nature Human Behaviour* Vol. 1, Article Bumber 0006 (21 November 2016). DOI: 10.1038/s41562-016-0006
16）Susanne Kunkel, Maximilian Schmidt, Jochen M. Eppler, Hans E. Plesser, Gen Masumoto, Jun Igarashi, Shin Ishii, Tomoki Fukai, Abigail Morrison, Markus Diesmann and Moritz Helias, “Spiking network simulation code for petascale computers”, *Frontiers in Neuroinformatics* 8:78 (10 October 2014). DOI: 10.3389/fninf.2014.00078
17）Tadashi Yamazaki and Wataru Furusho, “Realtime simulation of cerebellum”, *International Symposium on New Horizons of Computational Science with Heterogeneous Many-Core Processors* (Riken Wako campus, Japan, February 27-28, 2018).

２．１．８　社会におけるＡＩ

（１）研究開発領域の定義

人工知能（AI）技術が社会に実装されていったときに起こり得る、社会・人間への影響や倫理的・法的・社会的課題（Ethical, Legal and Social Issues：ELSI）を見通し、あるべき姿や解決策の要件・目標を検討し、それを実現するための制度設計および技術開発を行う研究開発領域である。

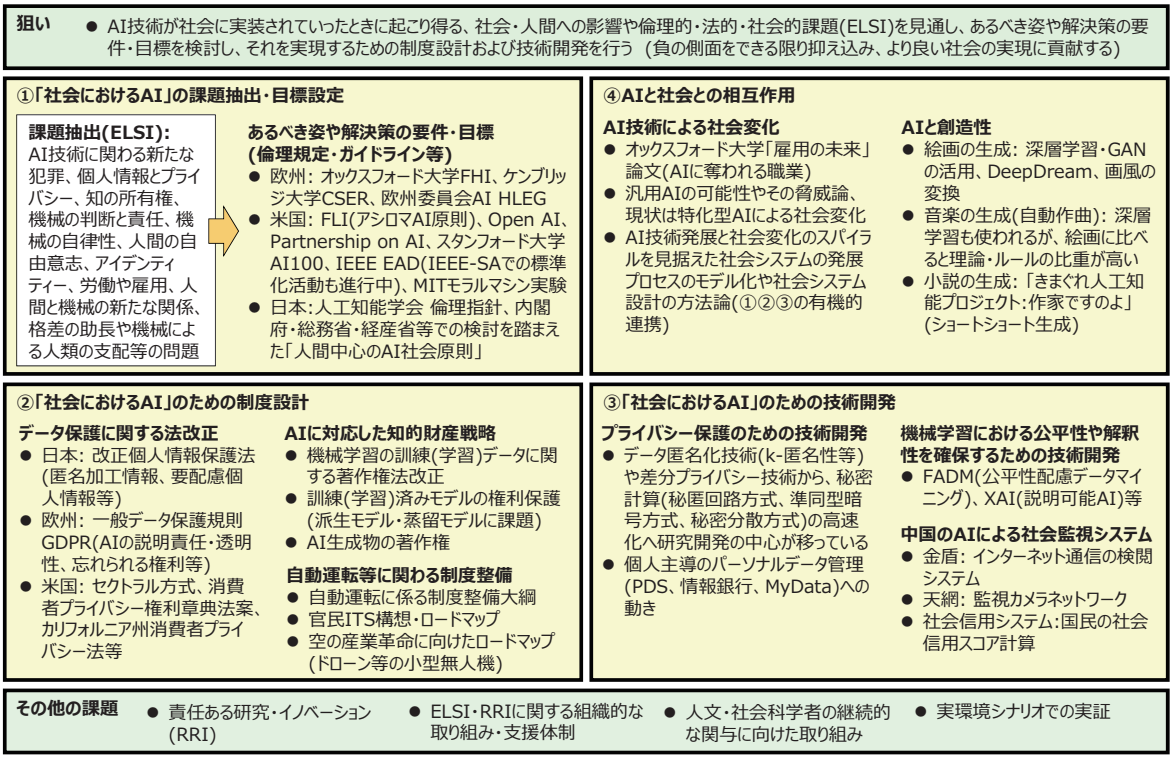


図2-1-11　領域俯瞰：社会におけるAI

（２）キーワード

ELSI、RRI、FAT、倫理、公平性、説明責任、透明性、創造性、法制度、プライバシー、知的財産権、ビジネスモデル

（３）研究開発領域の概要

［本領域の意義］

AI 技術は、人間の知的作業がコンピューターによって代行される可能性を広げることから、社会における人間の役割を変え、人間の働き方やモチベーションにも影響を及ぼし、社会の仕組み・あり方も変貌させる可能性を持っている。これによって、便利で効率的な社会を築くことができ、人間は快適な生活を過ごせると期待される一方で、AI が職業を奪うとか、プロファイリング（個人の性格・特徴を分析する技術）によってプライバシーを侵害されるとか、負の側面に対するさまざまな不安・懸念が指摘されている。

本研究領域の取り組みは、それら起こり得る影響・課題を事前に把握し、その対策を制度と技術の両面から実現することによって、負の側面をできる限り抑え込み、より良い社会を実現するために貢献する。

［研究開発の動向］

本領域の取り組みを、①「社会における AI」の課題抽出・目標設定、②「社会における AI」のための制度設計、③「社会における AI」のための技術開発、④ AI と社会との相互作用、という 4 つに分け、その概要と動向を述べる。

①「社会における AI」の課題抽出・目標設定

①は AI 技術が社会に実装されていったときに起こり得る、社会・人間への影響や倫理的・法的・社会的課題（ELSI）を抽出し、あるべき姿や解決策の要件・目標を定める活動である。

まず ELSI として考えられる問題を表 2-1-1 に例示した。JST CRDS では 2013 年から 2014 年にかけて、人文・社会科学と情報科学の双方から有識者を広く集めたワークショップを複数回 [1), 2), 3)] 開催しており、その中で議論された ELSI に関する論点をまとめたものである。AI 技術に直接関わるものだけでなく、情報技術全般に共通するものや、科学技術一般の社会受容そ

表2-1-1　倫理的・法的・社会的課題(ELSI)と想定される問題

ELSI	想定される問題	実際に発生した主な事例
新たな犯罪	・ハッキングされた機械やシステムによる詐欺や盗聴・盗用 ・ユーザーの個人情報を用いた詐欺、恐喝（オレオレ詐欺など） ・フェイクニュース、デマによる実質的な被害	・Microsoft の Twitter 上のチャットボット Tay が、人種差別的発言や暴力的発言をし始め、デビュー後数時間で停止（2016 年、米国）
個人情報とプライバシー	・個人に関わる情報の利活用に関するさまざまな問題 ・EU の一般データ保護規則への抵触 ・機微情報の扱い、非言語情報の多くはプライバシー情報 ・看視と監視（監視カメラや通行履歴など） ・平時や非常時（災害、犯罪、国防など）の使い分け	・JR 東日本が Suica の乗降履歴を日立に販売し、利用者やマスコミが大反発（2014 年、日本） ・Cambridge Analytica が Facebook から不適切な方法で約 5000 万人のデータを政治操作等に悪用疑惑（2016 年、英国・米国）
知の所有権	・知の断片化の促進と専門知の経済基盤の流動化 ・二次情報、三次情報の所有権、知のエコシステム、AI の著作物	・AI が描いた絵画が著名オークションで 4900 万円で落札（2018 年、米国）
機械の判断と責任	・機械の判断の正当性・妥当性の担保 ・機械の法的電子人格 ・結果に対する責任の帰着、責任のトップは人間だけか	・Amazon が AI 採用打ち切り、女性差別の欠陥露呈で（2018 年、米国）
機械の自律性	・自律的に動作する機械およびそれらの競合・協調動作をいかに制御するか ・判断・推論の信頼性の保証	・NY 証券取引所でフラッシュクラッシュ（2017、米国）
人間の自由意志	・個人や集団の合意形成や意思決定の制御・誘導 ・選挙や国民投票への介入	・ソーシャルメディア等を用いた政治操作（デジタルゲリマンダー）が米国大統領選挙の結果にも影響を与えたと言われる(2016 年、米国)
アイデンティティー	・サイバー上の情報やプロファイリングによる虚像とそれに基づく不公平や差別の助長 ・情報の標準化による個性の喪失	・ロシアのコンピュータプログラム「ユージーン」がチューリングテストに合格（2014 年、英国）
労働や雇用	・機械により奪われる雇用問題 ・実践知の蓄積による名人・職人の雇用喪失	・ゴールドマン銀行の株式デスクに生身のトレーダー 3 人、昔は 500 人（2018 年、米国）
人間と機械の新たな関係	・技術の進歩に人間はどう対応すべきか（抑制か活用か） ・社会システムの技術への過剰依存、サイバーテロ ・ポストヒューマニティー	・ジョージア工科大学の AI 教育助手 Jill Watson に、学生は気づかず（2016 年、米国）
格差の助長や機械による人類の支配	・情報リテラシーの有無によるデバイドや格差の拡大 ・スーパーヒューマニティーの出現と無用者階級	・時給 20 ドル以下の仕事の 83% で AI 優勢、時給 40 ドル以上の仕事で 4% と予想（2016 年、米国）

のものに関わるものも含んでおり、やや広めにあげている。加えて、表 2-1-1 には、実際に発生して報じられた主な事例も併せて載せた。また、参考として、論文[4]では AI と倫理に関わる問題を、AI の持つリスク、AI に関わる人間のリスク、失業などの社会的インパクト、法律や社会の在り方に関する問題という 4 つに分けて論じている。

次に、このような問題を議論し、あるべき姿や解決策の要件・目標を定めようとする取り組みが国内外で推進されている[5),6),7)]。以下にその代表的なものをあげる。

欧州では、比較的早い時期から、特に英国の大学・研究機関を中心に取り組まれてきた。まず、2005 年にオックスフォード大学の哲学科の下部組織として Future of Humanity Institute（FHI）が設立された。FHI は、技術変化によってもたらされる倫理的ジレンマやリスク（AI だけでなく気候変動や経済破綻等も含む）に対して、長期的にどう選択・対処していくべきか、学際的な研究を進めている。また、ケンブリッジ大学では、2012 年に人文・社会科学部局の下部組織として Cambridge Center for Existential Risk（CSER）が設立された。CSER では、AI、バイオ、ナノ等の先端技術のリスクに対する哲学的・倫理的な研究が行われており、産官学ワークショップ等を実施している。最近では、欧州委員会の AI HLEG（High-Level Expert Group on Artificial Intelligence）による Draft Ethics Guidelines for Trustworthy AI が 2018 年 12 月に公開され、パブリックコメントを受けた後、2019 年 3 月に最終版が出される予定とされている。

米国では、産業界主導で取り組みが始まり、それを追うように学術界での取り組みが立ち上がった。産業界主導の活動としては The Future of Life Institute（FLI）、OpenAI、Partnership on AI がよく知られている。FLI は、2014 年 3 月に Jaan Tallinn（Skype 共同創始者）らによって設立され、Elon Musk から 1000 万ドルの寄付も得て、AI 等の新技術を人類がうまくコントロールできるようにする研究を支援し、オープンレターも発信している。FLI はさらに、2017 年 1 月に 5 日間にわたるアシロマ[1]での会議の結果として、AI の研究課題、倫理と価値観、長期的な課題を含む 23 項目のガイドライン「アシロマ AI 原則」（Asilomar AI Principles）を公表し、多くの署名賛同を得ている[2]。OpenAI は、2015 年 12 月に Elon Musk を含む複数の投資家からの寄付金 10 億ドルをもとに設立された非営利研究機関で、安全な AI の恩恵をオープンソース（透明性の高い形で）で広く配分することを目指している。Partnership on AI は、2016 年 9 月に Amazon、Google、Facebook、IBM、Microsoft の 5 社が、AI のベストプラクティス構築、社会的影響の議論を目的として設立した非営利団体で、その後、Apple、Intel、Sony 等を含め、参加企業・団体数を大幅に増やしている。

一方、米国の学術界での取り組みとしては、スタンフォード大学の One Hundred Year Study on Artificial Intelligence（AI 100）、IEEE（The Institute of Electrical and Electronics Engineers、米国電気電子学会）の自律インテリジェントシステムの倫理に関する IEEE グローバルイニシアティブ（The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems）がよく知られている。AI 100 は、AAAI（Association for the Advancement of Artificial Intelligence、米国人工知能学会）の会長を 2007 年～ 2009 年に努めた Eric Horvitz（現 Microsoft）らが中心となって 2014 年 12 月に開設し、AI が法・経済・

[1] 米国カリフォルニア州のアシロマは、遺伝子組み換えに関するガイドラインが議論されたアシロマ会議が、1975 年に開催された場所である。このアシロマ会議は、科学者自らが研究の自由を束縛してまで自らの社会的責任を表明したもので、科学史に残る象徴的な場所で再び AI に関して同様の議論がなされた。

[2] 2018 年末時点の公開情報として、AI・ロボット工学研究者 1197 名、その他 2320 名がこの原則に署名したとのことである。

社会等にもたらす100年にわたる長期的な影響を学際的に調査・議論している。IEEEグローバルイニシアティブは、2016年4月に活動を開始し、「倫理的に配慮されたデザイン（Ethically Aligned Design）：自律インテリジェントシステムで人間の福祉を優先するためのビジョン」と題されたレポートを作成しており、2016年12月に初版（EADv1）、2017年12月に第2版（EADv2）を公開した（詳細は後述）。これはIEEEでの標準化も目指している。

日本では、学会・政府主導のガイドライン策定が推進されている。学会では2014年に人工知能学会が倫理委員会を立ち上げ、同委員会での議論や公開討論を経て、2017年2月に「人工知能学会 倫理指針」を公開した。9項目から成り、研究者倫理に焦点が置かれているが、最後の第9条「人工知能への倫理遵守の要請」はAI自体が倫理的であるべきということを掲げたのが特徴である。また、政府主導の活動としては、内閣府の「人工知能と人間社会に関する懇談会」、総務省の「AIネットワーク社会推進会議」、経済産業省の「AI・データ契約ガイドライン検討会」等が進められてきたが、それらを踏まえた活動として、内閣府の「人間中心のAI社会原則検討会議」が2018年5月に始まり、同年12月に「人間中心のAI社会原則（案）」が公開された。国内外から広く意見を募った上で2019年3月に策定という予定とされている。人間中心のAI社会原則（案）は、人間の尊厳が尊重される社会（Dignity）、多様な背景を持つ人々が多様な幸せを追求できる社会（Diversity & Inclusion）、持続性ある社会（Sustainability）という3つの価値を基本理念とし、「AI-Readyな社会」をビジョンに掲げ、人間中心の原則、教育・リテラシーの原則、プライバシー確保の原則、セキュリティー確保の原則、公正競争確保の原則、公平性・説明責任・透明性の原則、イノベーションの原則という7つをAI社会原則としてあげている。

②「社会におけるAI」のための制度設計

①で導出した要件・目標の実現に向けて、制度設計面での取り組みが進められている。具体的には、プライバシー・個人情報保護等のデータ保護に関する法改正、AIに対応した知的財産戦略、自動運転に関わる制度整備等が議論されている。ここでは、特にデータ保護に関する法改正の動向について述べる[8),9)]。他2点については「(4) 注目動向」で取り上げる。

日本におけるデータ保護の法制度としては、まず2003年に制定された、個人情報保護法を含む個人情報保護関連5法がある。その後、2013年からIT総合戦略本部の「パーソナルデータに関する検討会」が中心になって、個人情報保護法の改正に向けた議論が行われ、2015年に改正個人情報保護法が成立し、2017年5月に施行された。この改正では、事業者間での轌転流通が認められる匿名加工情報の新設、要配慮個人情報の導入、高い独立性を持つ個人情報保護委員会の設立等、データを取得した個人の同意なしでデータを流通・利用活するための新しい枠組みが創設された。

欧州（EU）では、1995年に制定されたデータ保護指令（Data Protection Directive: 95/46/EC）が存在したが、2012年からインターネット、デジタル化といった技術進化やグローバル環境変化を踏まえた全面的な見直しが進められた結果、パーソナルデータの取り扱い（Processing）と移転（Transfer）に関わる規則（Regulation）を定めた一般データ保護規則（General Data Protection Regulation：GDPR）が2016年4月に成立し、2018年5月から適用が開始された[3]。先のデータ保護指令は、各国に一定の法律の制定を義務付けているが、

[3] GDPRに関連するニュースとして、2019年1月、欧州委員会から日本は十分性認定（Adequacy Decision）を受けた。GDPRでは、EU域内の個人に関するデータをEU域外へ移転することを原則として認めていないが、十分なデータ保護政策がとられている国であれば、十分性認定を受けることで、移転が認められる。

指令（Directive）が各国に直接適用されるわけではない。それに対してGDPRは、各国に直接適用されるため、運用面での位置付けが大きく異なる。規則の内容の面では、特にAIとの関係が深いものとして、プロファイリングに基づく自動意思決定に対する説明責任・透明性を要求していること（GDPR第22条）があげられる。そのほかにも「削除権」（「忘れられる権利」、同第17条）、「開示請求権」（同第15条）、「データポータビリティー権」（同第20条）、罰則の強化等にも特徴がある。

米国では、公的部門については1974年のプライバシー法が存在するものの、民間部門については包括法がなく、自主規制を基本としている。規制対象を限定して個別領域毎に個別法が制定されるセクトラル方式がとられている。ただし、GAFA（Google, Amazon, Facebook and Apple）をはじめとするインターネット関連事業者に膨大な利用者データが集まる状況に際して、2012年に「ネットワーク化された世界における消費者データプライバシー」という政策大綱が公表され、その中で、事業者によるインターネット上の追尾・追跡（トラッキング）を消費者が拒否（オプトアウト）できる「Do Not Track（DNT）」という概念を明確化して基本方針とした「消費者プライバシー権利章典」が提案された。さらに、その後、2015年に、この権利章典をもとにした「消費者プライバシー権利章典法案」が公開された。また、州法レベルでのさまざまなプライバシー保護法が制定されているが、特にカリフォルニア州は先進的な法律を制定してきた。たとえば、2002年に同州が最初に制定したセキュリティー侵害通知法（California Security Breach Notification Act：情報漏洩が発生した場合に事後的な通知・報告を義務付け）は、その後、ほぼ全州が同様の法律を制定した。さらに、カリフォルニア州は2018年6月に新たに消費者プライバシー法（California Consumer Privacy Act）を制定した（2020年1月施行）。この新法では、消費者が収集・販売された情報について開示・消去等を請求できるようになる。

③「社会における AI」のための技術開発

①で導出した要件・目標の実現に向けて、技術開発面での取り組みが進められている。特に活発に取り組まれている技術課題としては、機械学習における公平性や解釈性を確保するための技術開発や、プライバシー保護のための技術開発があげられる。前者については、FADM（Fairness-aware Data Mining）[10), 11)]やXAI（Explainable AI）[12), 13), 14)]等の技術開発が進められているが、「2.1.1 機械学習」「2.1.4 AIソフトウェア工学」の節でも取り上げているので、本節では詳細を省略する。ここでは特に後者についての技術開発状況を述べる。

なお、社会におけるAIに関して、①②の面だけでなく③の面、つまり、どのような情報技術によって対処できるかについても論じられる国際会議としては、2014年から毎年開催されているFAT/ML（Workshop on Fairness, Accountability, and Transparency in Machine Learning）や、2018年から始まったACM FAT*（ACM Conference on Fairness, Accountability, and Transparency）　やAIE’S（AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society）がある。

プライバシー保護のために取り組まれている代表的な技術として、(1) 解析対象データにおけるプライバシー保護のためのデータ匿名化技術、(2) データベース問い合わせにおけるプライバシー保護のための差分プライバシー技術、(3) 計算過程におけるデータ内容の漏洩防止のための秘密計算（秘匿計算と呼ばれることもある）技術等がある。

1つめのデータ匿名化技術は、データとデータ主体（あるいは所有者）との間の相関を取り除く技術である。パーソナルデータの収集において、姓名等の識別子を削除しただけでは、上記の相関は完全には取り除けず、他の属性情報・履歴情報を束ねて見ることで個人が特定され得るリスクがある。このようなリスクを定式化し、低減するための考え方としてk-匿名性[15]がよく知られている。具体的には、表形式データについて、パーソナルデータの属性値の組み合わせが同じであるデータが、パーソナルデータ集合中にk個以上存在している状態が、k-匿名性が成立した状態である。データの正確性は犠牲になるが、パーソナルデータを改変することで、k-匿名性を成立させ、個人特定を困難にする。その後、k-匿名性を基礎概念として、匿名化対象を表形式データからグラフや時系列データに拡張する研究や、k-匿名性モデルにおいて十分にプライバシーを保護できない状況下におけるより強力な匿名性定義の研究等が進められてきた（l-多様性、t-近似性等）。改正個人情報保護法で新設された匿名加工情報の実装において実務上重要な技術である。

2つめの差分プライバシー技術[16]は、データベース問い合わせとその応答に関わる情報漏洩を理論的に評価し、その対応策を与える技術である。一般に、統計的クエリの応答値開示は個人情報の開示とは見なされにくいが、母集団数が少ない場合は、統計値から個別情報がある程度推測できてしまうケースが起こり得る。そこで、差分プライバシーは、個人に関するデータを格納したデータベースに対する統計的クエリの応答値から、そのデータベースに、ある個人が含まれているか否かを判定できることをリスクと捉え、このリスクを低減するために、統計値を雑音によって摂動させて開示するアプローチを取る。個人がデータを提供するときに個人の下で雑音を加算するローカル差分プライバシーのアイデアも注目されている。差分プライバシーは、その安全性が攻撃者の背景知識によらないため、k-匿名性等に比べて、より強力な安全性を保証できる。ただし、データベース中に個人に関わるデータが含まれるかわからなくするために、大きな雑音が必要となることがある。

3つめの秘密計算[17]とは、複数の者が互いに相手と共有することができない秘密のデータを所持しているときに、その秘密データを互いに他者に開示することなく、秘密データを入力に取る関数を評価し、その出力をいずれかの者あるいは第三者のみに報告する分散計算の総称である。その実現方式は、大きく分けて、秘匿回路を用いた方式、準同型暗号を用いた方式、秘密分散を用いた方式という3種類が知られている。

秘匿回路方式は、1982年に秘密計算の概念が提案されたときに示されたアイデアで、二者間で一方が暗号化された秘匿回路を構築し、両者がおのおののデータを暗号化した上で回路に入力すると、回路による演算処理の結果が他方に渡されるというものである。論理回路で記述可能な任意の関数を秘密計算として実装することが可能で汎用性が高いが、計算時間・通信量が大きい（通信回数は少なくできる）。

準同型暗号とは、暗号化したまま復号することなく（秘密鍵の知識なしに）、加法・乗法等の演算が可能な暗号系である。加法または乗法のいずれかに対応した準同型性暗号系は以前から知られていたが、2009年に加法と乗法の両方で同時に準同型性を満たす暗号「完全準同型暗号」が発表されたのがブレイクスルーとなり、研究が進展した。秘匿回路方式と比較すると、算術演算をベースとしているため、数値データ解析や行列演算をベースとするアルゴリズムに適している（一方、秘匿回路方式は文字列操作等の離散的な処理の方が向く）。

これらに対して、秘密分散は暗号系を用いない秘密計算手法である。秘密情報を複数の情報

の断片（シェアと呼ぶ）に分散させる。単独のシェアからは秘密情報を復元することはできないが、複数のシェアを集めることで秘密情報を復元することができる。秘密情報を複数の者にランダムシェアとして分散させ、ランダムシェア同士の演算によって秘密計算を実現する。実現可能な関数の種類は限定されるが、暗号処理を含まないので一般に計算効率がよい。

いずれの方式も当初は膨大な計算時間を要したが、近年、高速化改良が進み、実用性が高まっている。秘密計算ライブラリーが公開されるようになり、また、いくつかの企業で秘密計算を用いたサービスが実施されている（エストニアの Cybernetica による Sharemind、デンマークの Partisia、イスラエルの Unbound 等）。

④ AI と社会との相互作用

AI 技術は社会を変え、変わった社会がさらに発展あるいは安定化するために、また新たな AI 技術を要求する、というような連鎖（AI 技術→社会変化→ AI 技術→社会変化→…）が起こってくる。そのような社会変化や AI 技術の発展から、人間のあり方や思考の仕方も影響を受ける。連鎖のスピードや方向に、必ずしもすべての人々が追従できるわけではない。連鎖がどのような方向へ進むかを迅速に予測・把握し、連鎖の進行をうまくコントロールするための対策を的確に講じていくことが望ましい。

その際に、現状の AI 技術でどのようなことまでできるのか、さらに、今後どのようなことまでできる見込みなのか、といった見極めが求められる。これに関連した話題として、英国オックスフォード大学から 2013 年に発行された「雇用の未来（The Future of Employment）」と題する論文[18]が、「今後 10~20 年程度で、米国の総雇用者の約 47% の仕事が自動化されるリスクが高い」という予想を示したことから、AI 技術による職業や雇用機会の変化が盛んに論じられるようになった。また、さまざまな状況下でさまざまな目的に対応できる汎用 AI（Artificial General Intelligence : AGI）[19]の可能性や、それに対する脅威論も盛んに話題に取り上げられている[20), 21)]。その詳細をここでは論じないが、現状の AI 技術は汎用 AI からははるかに遠く、いま急速に社会に広がっている AI 応用システムは特定目的のみに対応した特化型 AI によるものである。ただし、特化型 AI であっても、社会や人間に大きな影響を与え得るものであり、社会を変えつつある。

また、AI 技術でどこまでできるかという面で、最近よく話題に上がるようになったものとして、AI による芸術作品（絵画、音楽）や文学作品等の生成、あるいは、AI の創造性という論点がある。まず AI 技術による絵画の生成については、深層学習の適用によって大きく進展した。2015 年に Google が開発した DeepDream[22]は、深層学習のニューラルネットワーク構造中の特徴情報を操作することで、通常の画像を、夢に出てくるような見たこともない神秘的な画像に変換する。さらに、ゴッホやムンク等の絵画作品から画風を学習し、入力画像をゴッホ風やムンク風の画風に変換するシステムも開発された[23]。Microsoft とデルフト工科大学は Rembrandt（レンブラント）の画風を学習し、3D プリンタを利用して新作を描く The Next Rembrandt プロジェクトの成果を公開し[24]、話題を集めた。深層学習を用いた敵対的生成ネットワーク（Generative Adversarial Networks : GAN）[25]による画像生成手法が発展し、きわめて自然で高精細な画像を生成できるようになった。2018 年 10 月には、14 世紀から 20 世紀に描かれた肖像画 1.5 万点のデータをもとに、GAN によって自動生成された肖像画が、米国ニューヨークの有名オークション「クリスティーズ」で、約 4900 万円という高値で、初めて

のAI作品として落札された。

AI技術による音楽の生成（作曲）についても、同様に深層学習の適用が進みつつあり、バッハ風やビートルズ風といった新曲を作ることが可能になった[26), 27)]。しかし、コンピューターを活用した作曲では、古くから音楽理論やルールをベースとした手法の蓄積があり、絵画のようにGANの活用は進んでいないようである。

AI技術による文学作品の生成については、2012年にスタートした「きまぐれ人工知能プロジェクト：作家ですのよ」がよく知られている。星新一のショートショート小説作品をコンピューターで解析して生成した作品を星新一賞に応募する試みを行っている。2016年の星新一賞の一次審査を通過した（受賞はならず）。ただし、登場人物の設定や話の筋、文章の部品に相当するものは人間が用意しておき、AI技術はそれらを用いた文章生成を行うものである。

人間の持つ創作欲求（美意識や自己表現等）そのものをAIが持つことは当分難しいとしても、表現の種や見本となるものがあるときに、そのスタイルをまねたり、新たな連想を生み出したりといったステップにはAI技術も活用できるようになりつつある。これは、人間の創作活動を支援する道具として、AI技術の活用が進む可能性を示唆している。

（4）注目動向

［新展開・技術トピックス］

① 個人主導のパーソナルデータ管理への動き

これまでパーソナルデータは、企業が運営するサービスの利用者に関するデータとして、各企業のもとに集められ、管理される形態が大半であった。企業が集中管理しているパーソナルデータが漏洩する事故がたびたび発生し、運営会社のセキュリティー管理が必ずしも万全ではないという不安も生じている。このような企業主導のパーソナルデータ管理ではなく、個人主導のパーソナルデータ管理へ移行しようという動きがある[8)]。自分のパーソナルデータがどこまでどの企業に開示されているのかを、自分自身で把握し、コントロールしたい（すべき）という考えである。そのために法律面では、自己情報コントロール権（開示請求権、削除権・訂正権、データポータビリティー権等が含まれる）の確保が考えられている。一方、技術・システム面では、自分の管理下にパーソナルデータを集約・管理する仕組みPDS（Personal Data Store）が設計されている。また、PDSとして自分の手元に置く代わりに、情報銀行（情報信託銀行の略称）に管理を委託するビジネスモデルも考えられている。内閣官房IT総合戦略室は、2016年9月からデータ流通環境整備検討会を開始し、その「AI・IoT時代におけるデータ活用ワーキンググループ」において、PDS、情報銀行、データ取引市場等の仕組みを集中的に検討し、2017年3月に中間とりまとめ[28)]を公表した。海外においても、米国のエネルギー分野のグリーンボタンや医療分野のブルーボタン、英国のmidataイニシアティブによるエネルギー・モバイル・金融・小売データ分野での取り組み等、個人主体のパーソナルデータ管理に向かう動きがみられる。

② AIに対応した知的財産戦略の動向

国内では、内閣の知的財産戦略本部での議論の中で、AIに対応した知的財産戦略・法改正等が議論されている。2017年3月に関連する報告書[29)]が公表された。

まず、機械学習の訓練（学習）データに関わる著作権法が改正され、機械学習のために、よりデータを利活用しやすくなった。もともと日本の著作権法は、コンピューターによる情報解析を目的とした複製等を許容する権利制限規定を有しており（旧 47 条の 7）、営利目的であっても、第三者の著作物が含まれていても、一定限度で著作権者の許諾なく著作物を利用することが可能とされている。さらに、2019 年 1 月 1 日施行の改正著作権法では、旧法では制限がかかっていたと解釈されるいくつかのケースに関して制限が緩和された（新 30 条の 4 第 2 号）。すなわち、旧法では訓練データを作成する主体と機械学習を実行する主体が同一であることを前提としていたのに対して、新法ではその前提が排除された。すなわち、作成した訓練データセットを他者に提供することも許容され、その利活用がいっそう促進される。

次に、訓練（学習）済みモデルの権利保護の課題について述べる。訓練済みモデルの再利用の仕方は、単純にモデルをそのままコピーして使うケース（複製）だけでなく、追加学習して使うケース（派生）、複数個の訓練済みモデルを組み合わせて使うケース（アンサンブル：Ensemble）、訓練済みモデルの振る舞い（どんな入力に対してどんな出力を出すか）の観測データを別のニューラルネットワークに学習させて新たなモデルを作るケース（蒸留：Distillation）が知られている [30]。このうちアンサンブルで使うモデルは複数個だが、1 つ 1 つは複製モデルに相当するので、権利保護を考えるべきモデルの種類は複製モデル・派生モデル・蒸留モデルの 3 種類である。このうち、派生モデルと蒸留モデルは、もとの訓練済みモデルとの関係性の立証が難しいことが権利保護上の課題である。契約・特許権・著作権等でどこまで保護できるか、新しい権利による保護が必要か、営業秘密として不正競争防止法で保護できるケースはどのようなケースか、といった検討がなされている [29]。

また、AI 生成物の著作権については、次のような解釈がされる。人間が AI 技術を道具として利用した創作物は、その人間に創作意図と創作的寄与があれば、その創作物は著作物であると認められる。一方、人間に創作的寄与が認められないケースは、AI 創作物とされ、現行の著作権法では著作物と認められない。AI はパラメーターを少しずつ変えながら休むことなく膨大なバリエーションの生成物を出力することが可能なので、それらに著作権を与えたら、大きな弊害を生む。ただし、ここでいう創作的寄与というのがどの程度のものならば該当するのかは、今後の課題として残っている。加えて、AI 創作物を人間による創作物だと偽られる懸念や、機械学習を用いた場合に生成物が訓練データと類似してしまう問題等も課題である。

③ 自動運転に関わる制度整備の状況

自動運転の研究開発、実用化に向けたトライアルが急速に進みつつあり [4]、その推進のための制度整備が、内閣をはじめ国土交通省・経済産業省・警察庁等の各省庁で検討されている。特に全体戦略としては、高度情報通信ネットワーク社会推進戦略本部（IT 総合戦略本部）官民データ活用推進戦略会議から公表された「自動運転に係る制度整備大綱」（2018 年 4 月）[31] と「官民 ITS 構想・ロードマップ」（毎年更新、同年 6 月に 2018 年版）[32] に示されている。

最新 2018 年版のロードマップによると、自家用車については、2020 年までに自動運転レ

[4] 自動運転への取り組みの現状は「SIP-adus Workshop 2018」に詳しい。
http://en.sip-adus.go.jp/evt/workshop2018/
[5] 自動運転レベルの定義は、レベル 0:運転自動化なし、レベル 1:運転支援、レベル 2:部分運転自動化、レベル 3:条件付き運転自動化、レベル 4：高度運転自動化、レベル 5：完全運転自動化とされている [31]。

ベル[5]2、2020年を目途に自動運転レベル3、2025年を目途に高速道路での自動運転レベル4の市場化が可能となる環境整備が掲げられている。物流サービスは、高速道路でのトラック走行について、2021年までに自動運転レベル2以上での後続車有人隊列走行、2022年以降にその無人隊列走行、2025年以降に自動運転レベル4が掲げられている。移動サービスは、2020年までに自動運転レベル4による限定地域での無人自動運転移動サービス、2022年以降に自動運転レベル2以上による高速道路でのバスの自動運転を目指すとされている。

これに向けた制度整備のための重点課題として、(1) 安全性の一体的な確保、(2) 自動運転車の安全確保の考え方（道路運送車両法等）、(3) 交通ルールの在り方（道路交通法等）、(4) 責任関係（自動車損害賠償保障法、民法、製造物責任法、自動車運転死傷処罰法等）、(5) 運送事業に関する法制度との関係等があげられている[31)]。

[注目すべき国内外のプロジェクト]

① IEEE「倫理的に配慮されたデザイン（Ethically Aligned Design）」

自律インテリジェントシステムを実現する仕組みの確立に向けた公の議論（過度な期待や恐怖の払拭を含む）、標準化や認証プログラムの策定、各国あるいは国際的な政策立案を促すような指針を与えることを目的とし、2016年12月に初版（EADv1）が公表され、そこからのフィードバックを踏まえて、2017年12月に第2版（EADv2）が公表された。(1) 倫理的に配慮されたデザインのための一般原則、(2) 自律インテリジェントシステムへの価値観の組み込み、(3) 倫理的な研究や設計のための方法論やガイド、(4) 汎用人工知能や人工超知能の安全性や便益、(5) パーソナルデータとアクセス制御、(6) 自律型兵器システムの再構築、(7) 経済・人道的課題、(8) 法律、(9) 感情（アフェクティブ）コンピューティング、(10) 政策、(11) 情報通信技術における伝統的倫理観、(12) 複合現実、(13) ウェルビーイング、という13項目から成る。(9) ～ (13) がEADv2で新たに追加された。

EADはデザインという言葉を使っている通り、倫理そのものではなく、設計論・設計思想、それをどのように技術に落とし込めるかといった論点が整理されていることが特徴である[33)]。また、このような議論を全世界から、AI研究者だけでなく、人文・社会科学研究者、政策関係者、企業関係者等、多様な250人以上の参加者を得て進めている。自律型兵器システムのような問題にも踏み込んでいる。さらに、IEEE標準化にもつなげようとしている点で、国際的な影響力が大きい。この標準化プロジェクトは、既にIEEE-SA（Standard Association：標準規格）に承認されている。EAD報告書執筆陣を中心に、IEEE-SA P7000シリーズとして標準化を目指す14のワーキンググループでドラフトを作成中である[34)]。

② MITメディアラボのモラルマシン実験

モラルマシンは自動運転車版のトロッコ問題（倫理的ジレンマ）に関する思考実験である。つまり、自動運転車にブレーキ故障等が生じ、事故で犠牲者が出ることが避けられない状況を示し、その中で一部の人だけ免れるとしたとき誰が優先されるべきかを、さまざまな人々に問い、回答を集めた。一人につき13通りのシナリオを示し、どちらを優先するか2択の質問に回答してもらうアンケートサイトを2016年6月に公開した。233の国・地域、230万人からのべ4000万件の回答が得られ、そのうち回答者が100人を超す130の国・地域のデータを分析した結果が2018年10月に発表された[35)]。

その結果によると、たとえば、個人主義的な文化を持つ地域では、より多くの人数を救うことが優先される傾向や、一人当たりの GDP が低く、法の規律も低い地域では、法遵守違反に寛容だという傾向や、経済格差の大きい地域では、社会的な地位の高さが優先される傾向等が見られ、地域の文化的背景との相関、各地域の特徴、地域間の類似性等が示された。ここで示された世界規模の倫理観の多様性に対して、倫理ルール作りをどのように進めていくべきか、あるいは、ある倫理観にしたがって作られた製品・サービスの各地域での受容性をどう考えていくか等、より検討を深めていくことが望まれる。

③ 中国の AI による社会監視システム

中国では、従来の「金盾」（Great Firewall）システムに加えて、「天網」（Sky Net）システムと「社会信用システム」の構築を進めており、政府による社会や国民の監視・管理が、AI 技術を用いて強化されている。

「金盾」はインターネット通信の検閲システムであり、Web 検索エンジンの検索語、電子メールやインスタントメッセンジャーの通信内容、Web サイトや SNS（Social Networking Service）のコンテンツ等に対して検閲・遮断が行われる。2003 年頃から稼働し、その後も段階的に強化されている。

「天網」は監視カメラネットワークで、2012 年に北京市に本格的に導入され、2015 年には中国内の都市エリアが 100% カバーされた。2017 年時点で監視カメラ 1.7 億台の規模になり、さらに 3 年で 4 億台の設置が目指されている。顔認証技術が組み込まれており、人混みの中から指名手配犯を見つけ、逮捕できたという実績もあげている。深圳市では、交差点に設置された監視カメラから信号無視等の違反者を見つけ、警告する試みも実施された。

さらに中国政府による「社会信用システム」の構築が 2014 年～ 2020 年の計画で進められている。所得・社会的ステータス等の政府が保有するデータに加えて、インターネットや現実社会での行動履歴も含めて評価し、各国民の社会信用スコアを計算するという計画である。一方、民間の電子マネー運営企業・電子決済運営企業においては既に、利用者のプロファイルや行動履歴から独自に信用スコアを算出し、そのスコアに応じて利用者に優遇や制限を与える（公共交通機関の割引・制限、病院診察やビザ取得手続きでの優遇等）ことが行われている。特に Alibaba（阿里巴巴）が展開する信用スコア「芝麻信用」は中国内で利用者が 5 億人を超えるという電子決済サービス Alipay と連動しており、大きな存在感を示している [6]。

政府によるさまざまな行動の監視や信用スコアに基づく賞罰によって、品行方正に振る舞う人々が増え、犯罪・違反の抑制や迅速な逮捕にもつながるという効果が得られているという。中国の「金盾」「天網」「社会信用システム」そのものは、表現の自由やプライバシーを重んじる欧米・日本には適合しないシステムであるが、AI が組み込まれた社会の一形態として非常に興味深い。

[6] 日本国内でも 2017 年・2018 年から信用スコアに類するサービスが立ち上がりつつある。たとえば、みずほ銀行とソフトバンクが設立した J.Score では「AI スコア」を提供している。ただし、中国の「芝麻信用」とは性格が異なり、氏名・住所の情報を与えずに匿名で始めることができ、アンケートに答える形で、その個人の信用力や可能性をスコア化するサービスである。

（５）科学技術的課題

①「社会における AI」の課題抽出・目標設定に関わる研究開発課題

前述の IEEE EAD は、2017 年 12 月に公開された EADv2 に対する意見のフィードバックを受けて、2019 年に最終版となる EADv3 が公開される見込みである。それと並行して、IEEE-SA P7000 シリーズとしての国際標準化も進行中である。また、前述のような各国での検討に基づき、2017 年からは、OECD（経済協力開発機構）や G7 情報通信・産業大臣会合のような国際的な議論の場においても、AI に関わるガイドラインが議題として取り上げられるようになった。なお、日本の「人間中心の AI 社会原則（案）」も 2019 年 3 月策定後、OECD・G7 にて発信される予定である。

IEEE EAD 等と比較すると、日本のガイドラインはやや一般的・現実的な視点でまとめられている印象があり、長期的な AI の可能性や軍事・デュアルユース問題には踏み込んでいない。その背景として、IEEE EAD は非常に多様な人々が多数議論に参加していることがあげられる。国際的な原則・ガイドラインの策定のためのプロセスや議論の仕方等、議論の場のデザインにも目を向けていく必要がある [33]。

②「社会における AI」のための制度設計に関わる研究開発課題

研究開発の動向や注目動向のパートで取り上げたように、(1) プライバシー保護を含むデータ保護に関わる制度設計（PDS や情報銀行への対応も含む）、(2) AI に対応した知的財産戦略（訓練済みモデルや AI 生成物の扱い等）、(3) 自動運転に関わる制度整備に関しては、日本としての戦略・制度改革方針に沿って手が打たれつつあるが、まだ課題も残されており、引き続き検討・施策推進が求められる。

それら (1)(2)(3) 以外で、制度整備が急がれるものとして、ドローン等の小型無人機に関わる制度整備があげられる。これに関しては、小型無人機に係る環境整備に向けた官民協議会から 2018 年 6 月に「空の産業革命に向けたロードマップ 2018」[36] が公表された。このロードマップでは、小型無人機の飛行レベルについて、レベル 1：目視内での操縦飛行、レベル 2：目視内での自動・自律飛行、レベル 3：無人地帯での目視外飛行（補助者の配置なし）、レベル 4：有人地帯（第三者上空）での目視外飛行（補助者の配置なし）と定めている。そして、2018 年頃からレベル 3、2020 年代前半からレベル 4 の利活用を本格化させるとしている。そのための制度整備や実証環境整備として、空の産業革命に向けた総合的な検討、目視外・第三者上空飛行等の要件に関する検討、機体の性能評価基準の策定、無人航空機 JIS の策定、操縦・運航管理に係る人材等の育成、航空機・無人航空機相互間の安全確保と調和、ドローン情報基盤システム、福島ロボットテストフィールドの整備・活用、国家戦略特区制度による「規制のサンドボックス」制度の創設、電波利用の環境整備があげられている。

なお、国の制度設計においては、他国の動きもウォッチし、国際的な方向性との整合性・連動性も考慮していくことも必要である。

また、AI・ロボット技術の今後の進展を見据えて、その法的な課題が「ロボット法」という視点で議論されている [6), 37)]。将来に向けて重要な視点である。

③「社会における AI」のための技術開発に関わる研究開発課題

プライバシー保護技術については、匿名化技術や差分プライバシー技術から秘密計算の高速

化へ研究開発の中心が移ってきている。しかし、準同型暗号を用いた秘密計算は、暗号化されていても個人情報に該当し、データ主体の同意なしに処理することはできないという見解がある。また、改正個人情報保護法で新設された匿名加工情報も、実施に際しての具体的な仕組みが詰められておらず、活用は広がっていない。②の制度設計と③の技術開発が連動するように、取り組みを推進する必要がある。

また、機械学習の解釈性の問題に関しては、人間とAIの共存において重要なのは、解釈性よりも、むしろトラストではないかという見方もあり、問題設定をより高い視点から検討していくことも求められる。

自動運転に関して、②制度設計の面で論じたが、インタラクション／コミュニケーション技術の面の課題もある。特に、完全な自動運転に移行する中間的な段階で発生する問題として、レベル3で機械から人間にいきなり制御主体が切り替わる状況で人間はどう対応できるのか、人間が運転する車と自動運転車の混在する状況でどう相手を認識し、コミュニケーションするか、といった問題を考えていく必要がある。

④ AIと社会との相互作用に関わる研究開発課題

AI技術発展と社会変化のスパイラルを見据え、そのようなスパイラルを組み入れた社会システムの発展プロセスのモデル化や社会システム設計の方法論の研究開発にも取り組んでいく必要がある。その中で、上記の①「社会におけるAI」の課題抽出・目標設定、②「社会におけるAI」のための制度設計、③「社会におけるAI」のための技術開発を有機的に連携させていくことが重要である。また、職業の変化や人材育成・教育への取り組みも、その中で描いていくことが必要である。「AI社会論」[38]として取り上げられている問題も関係が深い。

（6）その他の課題

① 責任ある研究・イノベーション（Responsible Research and InnovatŠn：RRI）

「社会におけるAI」へのRRIの考え方が不可欠である。RRIとは、新しい技術の創出・展開を進めるにあたって、生み出される成果が倫理的・法的・社会的に受容可能で、社会的価値・持続可能性等の面でも好ましいものであることを担保しながら取り組むことである。RRIは2000年代前半から欧米で議論されており、欧州のHorizon 2020においても政策的課題として重視されている。RRIを成り立たせる要件として、予見的であること（Anticipatory）、応答的であること（Responsive）、熟議的であること（Deliberative）、自己反省的であること（Reflective）があげられている[33), 39)]。

② ELSI・RRIに関する組織的な取り組み・支援体制

研究者のELSI・RRIに関する普及啓発（後期教養教育として幅広い研究者に対する継続的な教育実践を含む）や相談（事前検証と事後対応の両面）を支援して、ELSI問題を適切に解決する仕組みが必要である。個々の研究者に自覚を持たせることは重要だが、ELSIガイドライン等を設定して研究者本人に任せるというだけでなく、組織的な支援体制作りが求められる。つまり、知的財産センターが研究者の知的財産の創造・保護・活用を多面的に支援・促進するのと同じように、ELSIについても継続的に支援する組織体制を作り、ノウハウを組織的に蓄積することが効果的である。

③ 人文・社会科学者の継続的な関与に向けた取り組み

ELSIについては、倫理学者、哲学者、法学者、憲法学者、社会学者、政治学者、経済学者などの人文社会科学者の研究者が自身の研究課題として積極的に関与することが望ましい。欧州では人文・社会科学者が、Horizon 2020プロジェクトにおいて、人文・社会科学に対する賢明な投資が有益であると宣言した（ビルニウス宣言、2013年リトアニア）。プロジェクトの終了と共に活動やコミュニティーが終息することがないよう、継続的な取組みが必要である。

④ 実環境シナリオでの実証

Paolo Dario（イタリア聖アンナ大学院大学）は、EU/FP7 RoboLawプロジェクトにおいてロボットを市民が暮らす環境の中で働かせて試験した。さまざまな問題に遭遇し、対応に多大な努力を要したが、その結果、法規制の欠如、政治的手段、想定外の障害などが明らかになったという。AI関連アプリケーションサービスにおいても特区などの制度を活用して実際の適用に向けた課題の抽出と必要な対処の試行を効率的に実施すべきである。

（7）国際比較

①「社会における AI」の課題抽出・目標設定、②「社会における AI」のための制度設計、③「社会における AI」のための技術開発、④ AI と社会との相互作用 という 4 つの活動のうち、①②を基礎研究、③④を応用研究・開発として扱い、下表にまとめる。

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	◎	→	人工知能学会、内閣府、総務省、経済産業省等による学会・政府主導のガイドライン策定が推進されてきた。内閣府「人間中心の AI 社会原則（案）」がまとまり、国際的に発信される予定。データ保護では、事業者間での轢転流通が認められる匿名加工情報の新設等を盛り込んだ改正個人情報保護法を施行。
	応用研究・開発	○	↗	NTT・NEC 等で秘密計算の研究開発・実用化が推進されている。PDS・情報銀行への取り組みも進んでいる。
米国	基礎研究	◎	→	産業界主導で取り組み（FLI、OpenAI、Partnership on AI 等）が始まり、それを追うように学術界での取り組み（AI 100、IEEE EAD 等）が立ち上がった。EAD は IEEE 標準化を並行して進めており、国際的な影響力が大きい。
	応用研究・開発	◎	→	差分プライバシー、完全準同型暗号等、多くの理論的アイデアは米国の大学・企業の研究者から提案され、実装への取り組みも活発である。機械学習の解釈性に関する DARPA XAI、創造性に関する Google DeepDream 等、大学・企業の研究者層が厚く活発である。
欧州	基礎研究	◎	→	オックスフォード大学の FHI、ケンブリッジ大学の CSER 等、英国の大学・研究機関を中心に、比較的早い時期から取り組まれている。データ保護では、AI 処理の説明責任・透明性の要求や忘れられる権利等を盛り込んだ一般データ保護規則 GDPR を施行。
	応用研究・開発	◎	↗	プライバシー保護の基礎研究、匿名化技術を中心に著名な研究者がいる。ドイツを中心に、自動運転の制度設計への取り組みで先行している。
中国	基礎研究	△	→	2017 年 6 月に個人情報保護も含む中国インターネット安全法（中華人民共和国網絡安全法）が制定された。また、北京大学の ROBOLAW.ASIA というサイトでは、ロボットと AI の法律・倫理問題のさまざまな論考に取り組んでいる。
	応用研究・開発	△	↗	倫理・プライバシー保護面の取り組みは弱いが、AI 技術開発とその社会実装への取り組みは急成長している。中国独自の AI 監視・管理社会のための技術開発・システム化も注目される。
韓国	基礎研究	△	→	ロボットと人間との関係について定めたロボット倫理憲章の草案が 2007 年に産業資源部によって発表された。
	応用研究・開発	△	→	特に目立った活動は見られない。

（8）参考文献

1）科学技術振興機構 研究開発戦略センター、『ワークショップ報告書：知のコンピューティング ―人と機械が共創する社会を目指して― Wisdom Computing Summit 2013』、CRDS-FY2013-WR-05（2013 年 10 月）。
2）科学技術振興機構 研究開発戦略センター、『ワークショップ報告書:科学技術未来戦略ワークショップ「知のコンピューティングと ELSI/SSH」報告書』、CRDS-FY2014-WR-09（2014 年 10 月）。
3）科学技術振興機構、「激論！先端 ICT の光と影」、サイエンスアゴラ 2014 ワークショップ（2014 年 11 月）。
4）松尾豊・西田豊明・堀浩一・武田英明・長谷敏司・塩野誠・服部宏充・江間有沙・長倉克枝、「人工知能と倫理」、『人工知能』（人工知能学会誌）31 巻 5 号（2016 年 9 月）、pp. 635-641。
5）江間有沙、「「人工知能と未来」プロジェクトから見る現在の課題」、『人工知能学会第 29 回全国大会』2I5-OS-17b-1（2015 年）。DOI: 10.11517/pjsai.JSAI2015.0_2I5OS17b1
6）弥永真生・宍戸常寿（編）、『ロボット・AI と法』（有斐閣、2018 年）。
7）総務省情報通信政策研究所、「AI ネットワーク社会推進会議報告書 2018 － AI の利活用の促進及び AI ネットワーク化の健全な進展に向けて－」（2018 年 7 月 17 日）。
8）中川裕志・高崎晴夫・石井夏生利・森亮二・佐古和恵・神嶌敏弘・高橋克巳、「特集：パーソナルデータの利活用における技術および各国法制度の動向」、『情報処理』（情報処理学会誌）55 巻 12 号（2014 年 11 月）、pp. 1333-1380。
9）総務省、『情報通信白書（平成 30 年版）』（2018 年）。
10）福地一斗、「公平性に配慮した学習とその理論的課題」、『第 21 回情報論的学習理論ワークショップ』（IBIS2018；2018 年 11 月 6 日）。
11）神嶌敏弘、「Fairness-Aware Data Mining」、2018 年。http://www.kamishima.net/archive/fadm.pdf (accessed 2019-01-18)
12）Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Dino Pedreschi and Fosca Giannotti, “A Survey Of Methods For Explaining Black Box Models”, arXiv:1802.01933 (2018).
13）David Gunning, “Explainable Artificial Intelligence (XAI)”, DARPA Program Update, November 2017. https://www.darpa.mil/attachments/XAIProgramUpdate.pdf (accessed 2019-01-18)
14）高野敦、「もうブラックボックスじゃない、根拠を示して AI の用途拡大」、『日経エレクトロニクス』2018 年 9 月号（2018 年）、pp. 53-58。
15）Latanya Sweeney, “k-Anonymity: A Model for Protecting Privacy”, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* Vol. 10, Issue 5 (October 2002), pp. 557-570. DOI: 10.1142/S0218488502001648
16）Cynthia Dwork, Frank McSherry, Kobbi Nissim and Adam Smith, “Calibrating Noise to Sensitivity in Private Data Analysis”, *Proceedings of Third Theory of Cryptography Conference* (TCC 2006; New York, USA, March 4-7, 2006), pp. 265-284. DOI: 10.1007/11681878_14
17）竹之内隆夫・高橋克巳・菊池浩明（編）、「特集：安全なデータ活用を実現する秘密計算技

術」、『情報処理』（情報処理学会誌）59 巻 10 号（2018 年 10 月）、pp. 872-915。
18）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 2019-01-18)
19）Sam S. Adams, Itamar Arel, Joscha Bach, Robert Coop, Rod Furlan, Ben Goertzel, J. Storrs Hall, Alexei Samsonovich, Matthias Scheutz, Matthew Schlesinger, Stuart C. Shapiro and John F. Sowa, “Mapping the Landscape of Human-Level Artificial General Intelligence”, AI Magazine (Spring 2012), pp. 25-42. （和訳：篠田孝祐・市瀬龍太郎・ジェプカラファウ・寺尾敦・船越孝太郎・松島裕康・山川宏、「人間レベルの汎用人工知能の実現に向けた展望」、『人工知能』（人工知能学会誌）29 巻 3 号、2014 年 5 月、pp. 241-257）
20）Nick Bostrom, Superintelligence: Paths, Dangers, Strategies (Oxford University Press Keith Mansfield, 2014).（邦訳：倉骨彰訳、『スーパーインテリジェンス：超絶 AI と人類の命運』、日本経済新聞出版社、2017 年）。
21）中川裕志、「AGI の群雄割拠」、『人工知能』（人工知能学会誌）33 巻 3 号（2018 年 5 月）、pp. 351-356。
22）Alexander Mordvintsev, Christopher Olah and Mike Tyka, “Inception-ism: Going Deeper into Neural Networks”, Google AI Blog (17 June 2015). https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html (accessed 2019-01-18)
23）Leon A. Gatys, Alexander S. Ecker and Matthias Bethge, “A Neural Algorithm of Artistic Style”, arXiv:1508.06576 (2105).
24）“The Next Rembrandt”, https://youtu.be/IuygOYZ1Ngo (accessed 2019-01-18)
25）Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville and Yoshua Bengio, “Generative Adversarial Networks”, arXiv:1406.2661 (2014).
26）Jean-Pierre Briot, Gaëtan Hadjeres and François Pachet, “Deep Learning Tech-niques for Music Generation - A Survey”, arXiv:1709.01620 (2017).
27）徳井直生、「Deep Learning を用いた音楽生成手法のまとめ［サーベイ］」、medium (Nov 21, 2017)。
28）IT 総合戦略本部 データ流通環境整備検討会「AI、IoT 時代におけるデータ活用ワーキンググループ　中間とりまとめ」（2017 年 3 月）。
29）知的財産戦略本部 検証・評価・企画委員会 新たな情報財検討委員会、「新たな情報財検討委員会報告書 ーデータ・人工知能（AI）の利活用促進による産業競争力強化の基盤となる知財システムの構築に向けてー」（2017 年 3 月）。
30）丸山宏・城戸隆、「機械学習工学へのいざない」、『人工知能』（人工知能学会誌）33 巻 2 号（2018 年 3 月）、pp. 124-131。
31）高度情報通信ネットワーク社会推進戦略本部・官民データ活用推進戦略会議、「自動運転に係る制度整備大綱」（2018 年 4 月 17 日）。
32）高度情報通信ネットワーク社会推進戦略本部・官民データ活用推進戦略会議、「官民 ITS

構想・ロードマップ 2018」（2018 年 6 月 15 日）。
33）江間有沙、「倫理的に調和した場の設計：責任ある研究・イノベーション実践例として」、『人工知能』（人工知能学会誌）32 巻 5 号（2017 年 9 月）、pp. 694-700。
34）江間有沙・長倉克枝、「「倫理的に調和した設計」の論点整理 ―異分野・異業種によるワークショップからの示唆―」、『情報法例研究』第 4 号（2018 年 11 月）、p. 3-14。
35）Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon and Iyad Rahwan, “The Moral Machine experiment”, *Nature* Vol. 563 (2018), pp. 59-64. DOI: 10.1038/s41586-018-0637-6
36）小型無人機に係る環境整備に向けた官民協議会、「空の産業革命に向けたロードマップ 2018：小型無人機の安全な利活用のための技術開発と環境整備」（2018 年 6 月 15 日）。
37）Ugo Pagallo, The Laws of Robots: Crimes, Contracts, and Torts (Springer, 2013).（邦訳：新保史生・松尾剛行・工藤郁子・赤坂亮太訳、『ロボット法』、勁草書房、2018 年）。
38）高橋恒一・井上智洋（編）、「特集：AI 社会論」、『人工知能』（人工知能学会誌）32 巻 5 号（2017 年 9 月）、pp. 640-700。
39）平川秀幸、「責任ある研究・イノベーションの考え方と国内外の動向」、文部科学省 安全・安心科学技術及び社会連携委員会（第 7 回）資料 4-3（2015 年 4 月 14 日）。