

JSAI2019 企画セッション KS-10

AIマップタスクフォースの活動 —AI初学者・異分野研究者のためのAI研究の俯瞰—

JST CRDS「研究開発の俯瞰報告書」から 人工知能分野の俯瞰図ご紹介

2019年6月6日

福島 俊一



国立研究開発法人 科学技術振興機構 研究開発戦略センター
Center for Research and Development Strategy Japan Science and Technology Agency

JST CRDS「研究開発の俯瞰報告書」

crds 俯瞰報告書

検索

<http://www.jst.go.jp/crds/report/report02/index.html>

CRDS 国立研究開発法人科学技術振興機構 研究開発戦略センター

JSTトップ > 研究開発戦略センター (CRDS) > 報告書等 > 研究開発の俯瞰報告書

研究開発の俯瞰報告書 2019年

国内外の社会や科学技術イノベーションと関連政策の動向を俯瞰し、今後のあるべき方向性を展望

研究開発の俯瞰報告書 システム・情報科学技術分野 (2019年)

「研究開発の俯瞰報告書」は、国内外の社会や科学技術イノベーションの動向及びそれらに関する政策動向を把握・俯瞰・分析し、今後のあるべき方向性を展望するものです。主要な研究開発領域ごとに国内外の動向や主要国間の国際比較を記載しているほか、主要国の科学技術政策立案体制、科学技術基本政策、研究開発投資戦略等についても記載しています。

2019年度

- 研究開発の俯瞰報告書 環境・エネルギー分野 (2019年)
- 研究開発の俯瞰報告書 システム・情報科学技術分野 (2019年)
- 研究開発の俯瞰報告書 ナノテクノロジー・材料分野 (2019年)
- 研究開発の俯瞰報告書 ライフサイエンス・臨床医学分野 (2019年)
- 研究開発の俯瞰報告書 主要国の研究開発戦略 (2019年)
- 研究開発の俯瞰報告書 日本の科学技術イノベーション政策の変遷 ~科学技術基本法の制定から現在まで~

システム・情報科学技術分野 (2019年)

目次

研究開発の俯瞰報告書 システム・情報科学技術分野 (2019年)

研究開発の俯瞰報告書 概要 システム・情報科学技術分野 (2019年)

エグゼクティブサマリー [PDF:1.2MB]

目次 [PDF:0.7MB]

はじめに [PDF:1.0MB]

1. 研究対象分野の全体像 [PDF:3.3MB]

- 1.1 分野の範囲と構造
- 1.2 分野の研究開発を取り巻く現状
- 1.3 今後の展望・方向性

2. 俯瞰区分と研究開発領域 [PDF:6.1MB]

2.1 人工知能・ビッグデータ [PDF:1.9MB]

機械学習、画像・映像解析、自然言語処理、AI ソフトウェア

2.2 ロボティクス [PDF:3.8MB]

認知発達ロボティクス、生活支援ロボット、医療ロボット、ディクス、生物規範型ロボティクス、産業用ロボット、研

2.3 社会システム科学 [PDF:1.5MB]

計算社会科学、社会インフラマネジメント、サイバーフィ

2.4 コンピューティングアーキテクチャ [PDF:1.7MB]

プロセッサアーキテクチャ、量子コンピューターサイエンス、IoT アーキテクチャ、ブロックチェーン

JST CRDSの役割と俯瞰報告の位置付け

- JSTの主要事業は国の科学技術政策に基づくアカデミアへの戦略的ファンディング
- CRDSはそのためのターゲティングや研究開発戦略を提言
- 提言に先立ち、各分野の有識者の協力を得て、分野俯瞰を隔年発行



研究開発戦略の立案

～未来を共創する研究開発戦略を立てる～



- ✓ 俯瞰報告
- ✓ 戦略プロポーザル
- ✓ ワークショップ など

研究開発戦略センター ほか

- システム・情報科学技術ユニット
- 環境・エネルギーユニット
- ナノテクノロジー・材料ユニット
- ライフサイエンス・臨床医学ユニット
- 科学技術イノベーション政策ユニット
- 海外動向ユニット

科学技術イノベーションのための戦略プログラムの運営

～知を創造し、経済・社会的価値へ転換する～

戦略的創造研究推進事業

- CREST (国の戦略目標へチーム型研究)
- さきがけ (総括助言下での個人独立研究)
- ERATO (卓越リーダー下の独創的研究)
- ACCEL (成果の実用化・権利化推進)
- ACT-C (先導的物質返還領域)
- ALCA (先端的低炭素化技術開発)
- RISTEX (社会技術研究開発) ほか

未来社会創造事業 MIRAI

革新的研究開発推進プログラム ImPACT

ほか

科学技術イノベーションのための基盤整備

～社会との対話を推進し、人材を育成する～

データベースやツールの提供

(J-GLOBAL、J-STAGE、NBDC、researchmapほか)

次世代人材の育成

(スーパーサイエンスハイスクール、科学の甲子園、JREC-INほか)

情報発信・コミュニケーション

(日本科学未来館、サイエンスアゴラ、サイエンスポータルほか)

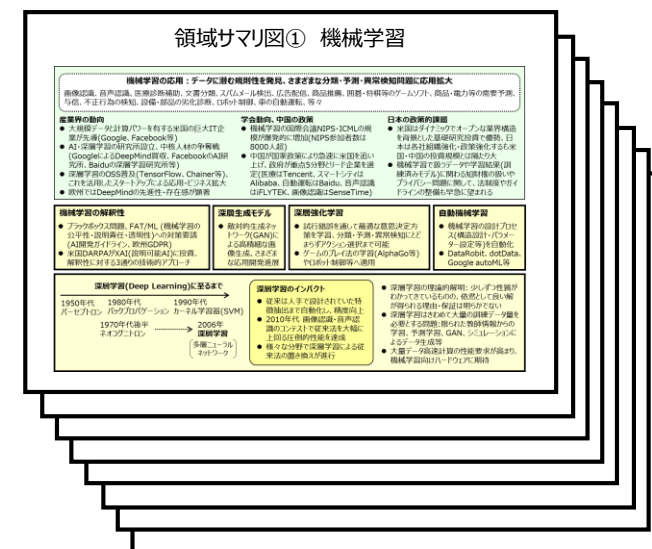
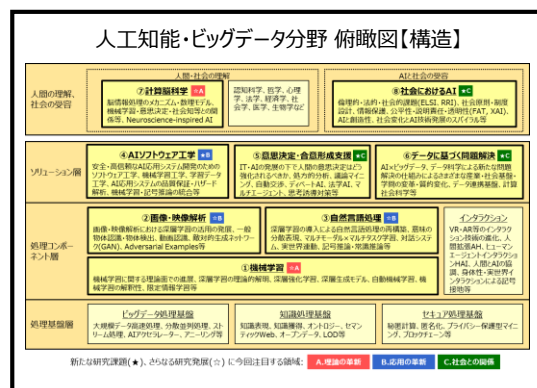
各研究開発領域の説明

領域サマリ図

(記載内容)

- (1)研究開発領域の定義
- (2)キーワード
- (3)研究開発領域の概要
 - [本領域の意義]
 - [研究開発の動向]
- (4)注目動向
 - [新展開・技術トピックス]
 - [注目すべき国内外のプロジェクト]
- (5)科学技術的課題
- (6)その他の課題
- (7)国際比較
- (8)参考文献

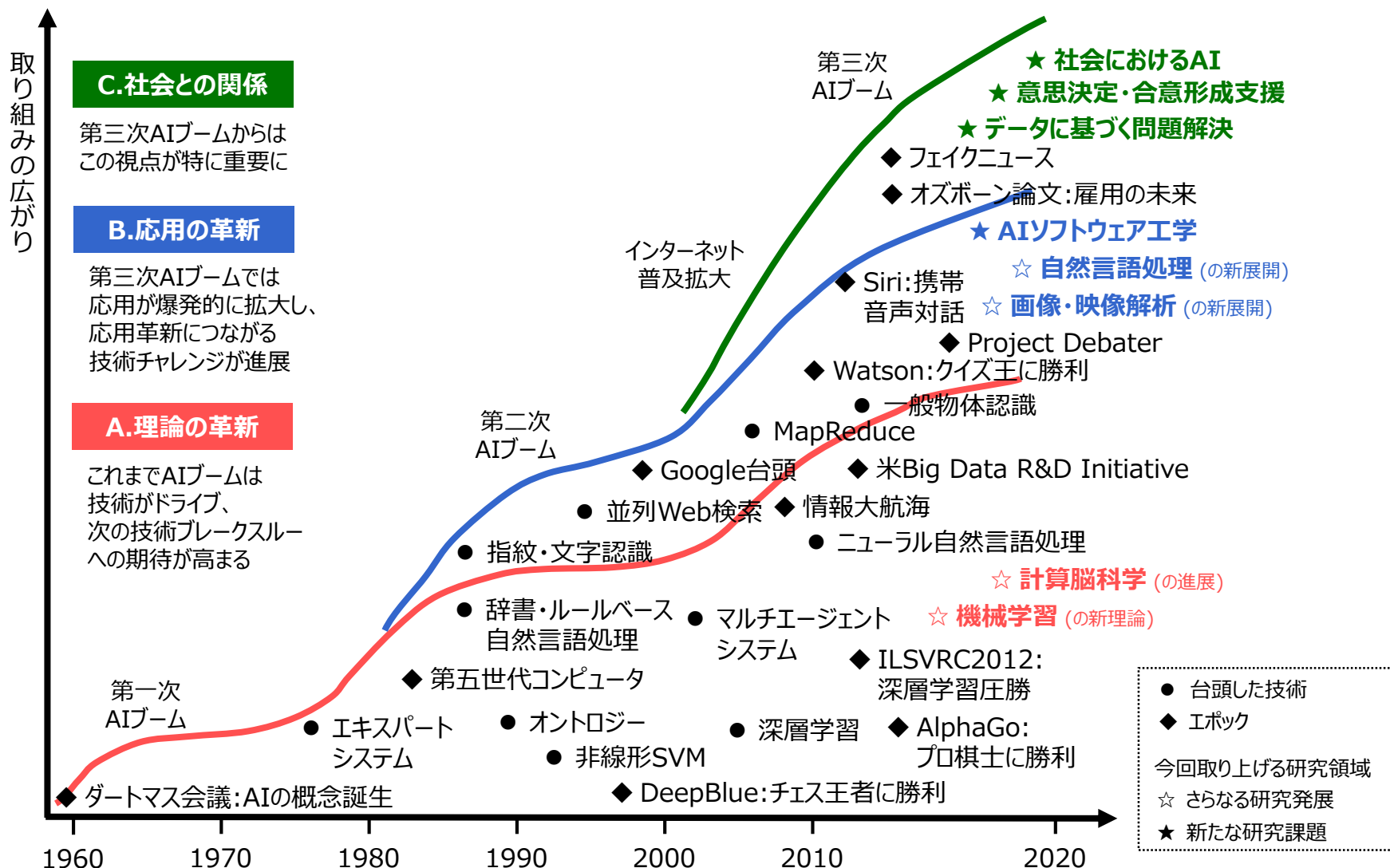
國際比較表



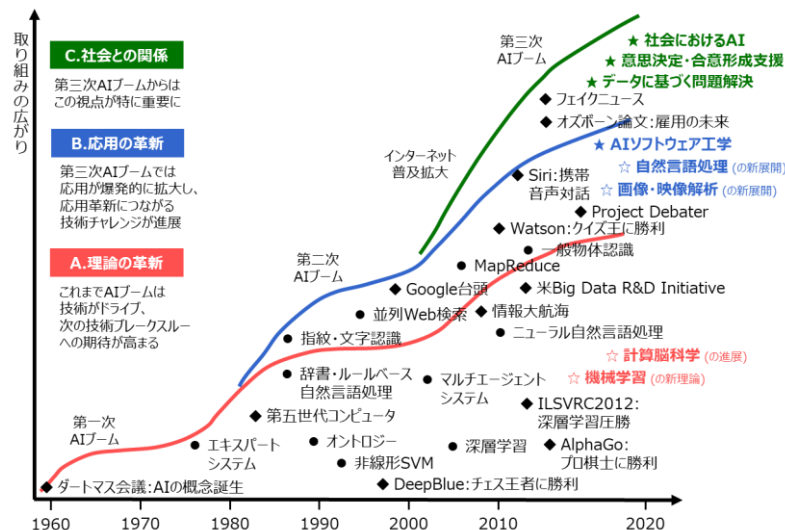
国	フェーズ	機械学習	画像・映像解析	自然言語処理	AIソフトウェア工学	意思決定・合意形成支援	データに基づく問題解決	計算脳科学	社会におけるAI
日本	基礎研究	○↗	○↗	○→	○↗	○→	○→	○→	○→
	応用研究・開発	○↗	○→	○↗	○↗	○↗	○↗	○→	○↗
米国	基礎研究	○↗	○→	○→	○↗	○↗	○→	○→	○→
	応用研究・開発	○↗	○↗	○→	○↗	○↗	○→	○→	○→
欧州	基礎研究	○→	○↗	○→	○↗	○↗	○→	○→	○→
	応用研究・開発	○↗	○↗	○→	○↗	○↗	○→	○→	○→
中国	基礎研究	○↗	○↗	○↗	△→	○↗	○→	○↗	△→
	応用研究・開発	○↗	○↗	○↗	△→	△→	○↗	○↗	△→
韓国	基礎研究	△→	○→	△→	×→	△→	△→	○→	△→
	応用研究・開発	△↗	△↗	○↗	×→	△→	○→	○→	△→

- 注目する8つの研究開発領域
- ① 機械学習
 - ② 画像・映像解析
 - ③ 自然言語処理
 - ④ AIソフトウェア工学
 - ⑤ 意思決定・合意形成支援
 - ⑥ データに基づく問題解決
 - ⑦ 計算脳科学
 - ⑧ 社会におけるAI

人工知能・ビッグデータ分野 俯瞰図【時系列】



注目する8つの研究開発領域



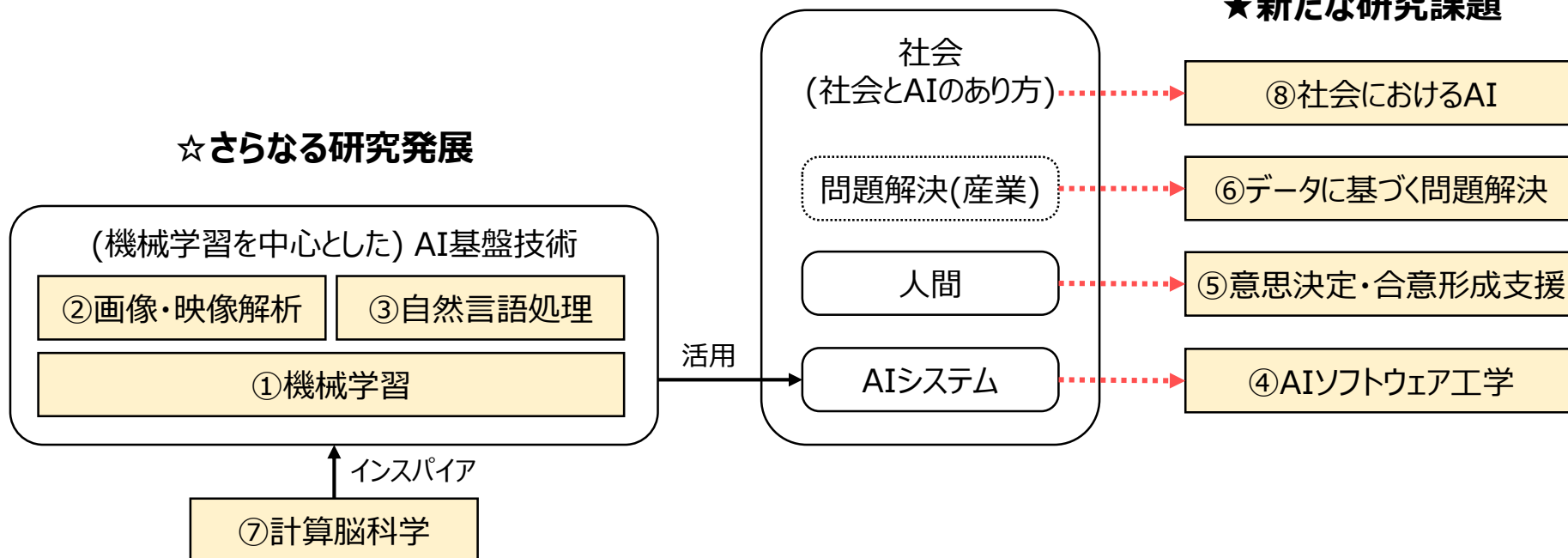
★新たな研究課題

AIと社会との関係が広がることで生まれる研究課題に注目

☆さらなる研究発展

機械学習を中心としたAI基盤技術のさらなる発展に注目

★新たな研究課題



人工知能・ビッグデータ分野 俯瞰図【構造】



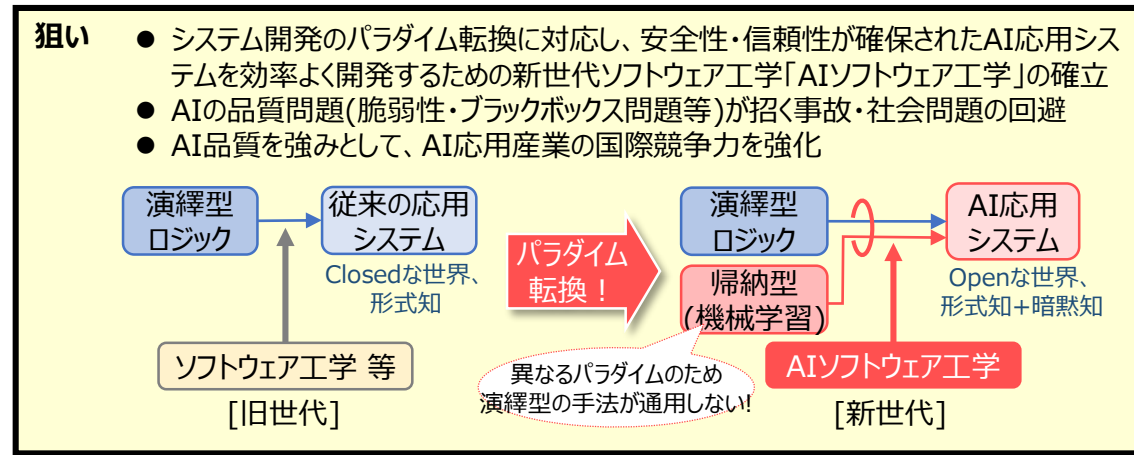
新たな研究課題(★)、さらなる研究発展(☆)に今回注目する領域:

A.理論の革新

B.応用の革新

C.社会との関係

領域サマリ図④ AIソフトウェア工学



- 業界動向**
- AIとソフトウェア工学・安全工学がクロスする新しい技術領域、2017年から急速に活発化(国内外の学会でワークショップ・招待講演等が立て続けに開催)
 - 国内では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、カナダではモントリオール理工科大学にSEMLAイニシアティブ発足
 - 自動運転分野での取り組みが先行、ドイツは自動運転の性能水準と評価基準を作るペガサスプロジェクトを推進中(2016年～2019年)、国内ではデンソーやJAXA+JAMAで自動運転分野の品質保証技術の研究開発
 - 国際標準化においては、ISO・IEC JTC1でAIの安全性・信頼性に関わる議論開始
 - 米国、ドイツ等で、安全性基準の設定、認証機関の設立に向けた動き

- 政策動向**
- AI・機械学習技術の性能・精度を高めるだけでなく安全性・信頼性面の政策強化
 - 米国DARPAではXAI(説明可能なAI)、Assured Autonomy(自律システムの安全性)への研究投資をそれぞれ2017年・2018年に開始
 - フランスMESRIは2018年イノベーション審議会でAIの安全性・信頼性に関する課題採択
 - 国内でも人工知能技術戦略会議と人間中心のAI社会原則検討会議に加えて、2018年後半に統合イノベーション戦略推進会議によって策定された新たなAI戦略(案)にはAI工学が盛り込まれた

- 技術的課題**
- AIソフトウェア工学の体系化:
AI応用システムの安全性・信頼性確保のため多面的・総合的取り組み、従来ソフトウェア工学体系SWEBOKからの拡張
 - 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テスト技術、機械学習の脆弱性(Adversarial Examples攻撃等)や公平性(差別・偏見の混入問題)への対策
 - 全体システムとしての安全性確保技術の確立:
機械学習型コンポーネントが100%保証ができないことを想定したSafety Envelope等
 - ブラックボックス問題への対策:
説明可能なAI(XAI)への3つのアプローチ
 - 問題を効率よく解くための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
 - 自動変化するシステムの品質保証技術の創出:
オンライン学習によってモデルを随時更新する将来システムに向けた新たな品質保証技術
 - 擬人化インターフェースの設計手法の確立:
機械学習に関わる問題とは異なる側面、認知ギャップの最小化のための設計手法等

- 政策的課題**
- AI品質基準・安全性基準の策定・運用および国際標準化
 - 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
 - 問題解決型基礎研究のトップダウン推進
 - 日本の先行優位性を活かす迅速な施策推進
 - 機械学習活用に関わる知的財産権の整備

国際比較表

国	フェーズ	機械学習	画像・映像解析	自然言語処理	AIソフトウェア工学	意思決定・合意形成支援	データに基づく問題解決	計算脳科学	社会におけるAI
日本	基礎研究	○↗	○↗	○→	○↗	○↗	○→	◎→	◎→
	応用研究・開発	○↗	◎→	○↗	○↗	○↗	○↗	○→	○↗
米国	基礎研究	◎↗	◎→	◎→	○↗	◎↗	◎→	◎→	◎→
	応用研究・開発	◎↗	◎↗	◎↗	○↗	◎↗	◎↗	◎→	◎→
欧州	基礎研究	○→	◎↗	○→	○↗	◎↗	○→	◎→	◎→
	応用研究・開発	○↗	○↗	○→	○↗	◎↗	○→	◎→	◎↗
中国	基礎研究	○↗	○↗	◎↗	△→	○↗	○→	○↗	△→
	応用研究・開発	◎↗	◎↗	◎↗	△→	△→	◎↗	○↗	△↗
韓国	基礎研究	△→	○→	△→	×→	△→	△→	○→	△→
	応用研究・開発	△↗	△↗	○↗	×→	△→	○→	○→	△→

まとめ

- 国の研究開発戦略・科学技術政策立案への提言のための前段としてのマップ(俯瞰図)
- 国の基礎研究投資という視点で注目する研究開発領域を整理
 - ・ 必ずしもAI分野のすべての研究トピックをカバーしているわけではない
- 注目動向や国際比較を踏まえ、戦略構築シナリオを組み合わせて、提言につなげる
 - ・ AI×社会のための研究開発、次期AI基盤技術の研究開発
 - ・ 国の「人間中心のAI社会原則」「AI戦略2019」と連動

戦略構築の基本的シナリオ

1)強い技術を核とした骨太化

既に保有している、あるいは、育ちつつある強い技術を足掛かりとして、技術の国際競争力を骨太化するシナリオ

2)強い産業の発展・革新の推進

既に保有している、あるいは、育ちつつある強い産業を足掛かりとして、国際競争力のある技術群を育てるシナリオ

3)社会課題の先行解決

課題先進国として、先端技術の社会受容性で先行できることを活かして、国際競争力を構築するシナリオ

4)社会基盤を支える根幹技術の確保

社会基盤を支える根幹技術は、国として保有・強化しなくてはならないという考えで取り組むべき技術開発シナリオ

より詳しい紹介を予定しています

連続セミナー 2019



データ駆動で新時代を切り拓く

第6回： 2019年12月9日(月)
IT分野の研究開発動向を俯瞰する

領域サマリ図① 機械学習

機械学習の応用：データに潜む規則性を発見、さまざまな分類・予測・異常検知問題に応用拡大

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

産業界の動向

- 大規模データと計算パワーを有する米国の巨大IT企業が先導(Google、Facebook等)
- AI・深層学習の研究所設立、中核人材の争奪戦(GoogleによるDeepMind買収、FacebookのAI研究所、Baiduの深層学習研究所等)
- 深層学習のOSS普及(TensorFlow、Chainer等)、これを活用したスタートアップによる応用・ビジネス拡大
- 欧州ではDeepMindの先進性・存在感が顕著

学会動向、中国の政策

- 機械学習の国際会議NIPS・ICMLの規模が爆発的に増加(NIPS参加者数は8000人超)
- 中国が国家政策により急速に米国を追い上げ、政府が重点5分野とリード企業を選定(医療はTencent、スマートシティはAlibaba、自動運転はBaidu、音声認識はiFLYTEK、画像認識はSenseTime)

日本の政策的課題

- 米国は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で優勢、日本は各社組織強化・政策強化するも米国・中国の投資規模とは隔たり大
- 機械学習で扱うデータや学習結果(訓練済みモデル)に関わる知財権の扱いやプライバシー問題に関して、法制度やガイドラインの整備も早急に望まれる

機械学習の解釈性

- ブラックボックス問題、FAT/ML(機械学習の公平性・説明責任・透明性)への対策要請(AI開発ガイドライン、欧州GDPR)
- 米国DARPAがXAI(説明可能AI)に投資、解釈性に対する3通りの技術的アプローチ

深層生成モデル

- 敵対的生成ネットワーク(GAN)による高精細な画像生成、さまざまな応用開発進展

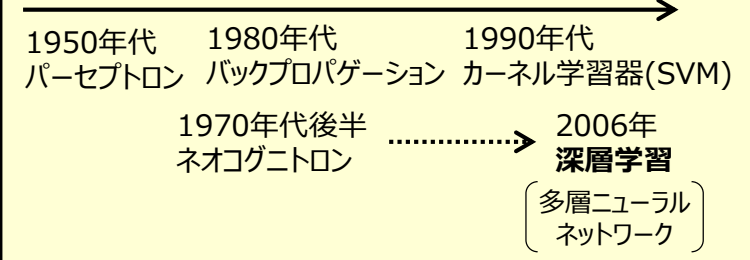
深層強化学習

- 試行錯誤を通して最適な意思決定方針を学習、分類・予測・異常検知にとどまらずアクション選択まで可能
- ゲームのプレイ法の学習(AlphaGo等)やロボット制御等へ適用

自動機械学習

- 機械学習の設計プロセス(構造設計・パラメータ設定等)を自動化
- DataRobot、dotData、Google autoML等

深層学習(Deep Learning)に至るまで



深層学習のインパクト

- 従来は人手で設計されていた特徴抽出まで自動化し、精度向上
- 2010年代 画像認識・音声認識のコンテストで従来法を大幅に上回る圧倒的性能を達成
- 様々な分野で深層学習による従来法の置き換えが進行

- 深層学習の理論的解明: 少しずつ性質がわかってきているものの、依然として良い解が得られる理由・保証は明らかでない
- 深層学習はきわめて大量の訓練データ量を必要とする問題: 限られた教師情報からの学習、予測学習、GAN、シミュレーションによるデータ生成等
- 大量データ高速計算の性能要求が高まり、機械学習向けハードウェアに期待

領域サマリ図② 画像・映像解析

画像・映像解析の応用：人間の目の代替によって、産業の効率化、社会の安心安全、生活の快適性に役立つ

文字認識、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・動画検索、カメラ画像のシーン分類、半導体ウェハー等の欠陥検査、食品の異物検査、製品の品質検査、リモートセンシング、出入国管理等での個人認証、車の安全運転支援・自動運転、ロボットビジョン、スポーツ画像解析、ヒューマンインターフェースデバイス等

一般物体認識から物体検出へ ● 画像中の物体の識別だけでなく、物体の位置も正確に検知 ● 最先端技術(YOLO)では9000種類の物体の検出・識別を実現	画像・映像解析と自然言語処理の融合 ● 画像・映像からの自然言語キャプション生成 ● Visual Question Answering: 画像に関する質問文に対して回答生成	画像認識に基づくコンテンツ生成 ● 通常画像を神秘的画像に変換(Deep Dream) ● 芸術作品の画風を学習・変換(The Next Rembrandt)	敵対的生成ネットワーク(GAN) ● 生成と識別の切磋琢磨的な学習手法 ● 高解像化、カラー画像化、CG→実画像化等の高品質変換、汎用性高い ● 学習の安定性に課題	深層学習による行動認識 ● 映像認識は、時空間的關係を表現するモデルと効率的学習手法、膨大な時系列データを扱う並列計算・HPC等に適合した映像処理アルゴリズム等に課題 ● RGB画像情報と動き情報の分離・統合処理(Two-stream)、Kinetics活用(I3D)で精度向上
---	--	--	---	--

画像認識

深層学習によって識別だけでなく特徴抽出まで併せて最適化が可能になり、認識精度が飛躍的に向上

深層学習の課題

- 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(end-to-end学習でブラックボックス、未知画像に対して人間に想像できない振る舞いを示すAdversarial Examples)
- 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(少ないと過学習になりやすい) → 転移学習、ドメイン適応、能動学習、データオーグメンテーション等

大規模データセットとベンチマーク、海外動向

一般物体認識ベンチマークILSVRC ● 2012年にトロント大がエラー率を飛躍的改善(従来26%→深層学習17%) ● 2015年以降は中国勢が躍進・連続トップ(2015 MSRA 3.57%、2016 中国公安部第三研究所が 2.99%、2017 Momenta 2.25%) ● 精度はほぼ飽和し、2017年で終了	日本は対象特化技術で世界トップ水準 ● NIST顔認証ベンチマークでNECが2009年・2010年・2013年・2017年と4連続世界一 ● PASCAL VOC大規模画像意味解析で2012年に東大、TRECVID映像意味分類タスクで2011年・2012年に東工大、TRECVID物体検索タスクで2011年・2013年・2014年にNIIがそれぞれ世界一	大規模データセット構築 ● 共通データでの競技会が技術発展に大きく貢献 ● さまざまなアノテーション付き画像・映像データの構築と競技会開催を米国・欧州が主導 ● 大規模データ保有が技術改良に不可欠で米国・中国が優勢、日本は差異化・競争力のための独自データセット構築に課題
---	--	--

領域サマリ図③ 自然言語処理

マルチモーダル×マルチタスク学習

- 共通のニューラルネット構造による訓練の効率化、精度向上

マルチモーダル処理・実世界連携

- マルチモーダル入力解析、クロスメディア変換の枠組み発展
- 実世界の対象物指示を伴う対話

雑談対話

- 雑談対話コンペティション (Alexa Prize等)
- 新たなメディア/チャンネルに

大規模コーパス・大規模知識ベース

- クラウドソーシング活用、品質管理手法
- 不足する知識の補完、Webからの拡充

情報の信憑性判定

- デマやフェイクニュース対策 (ファクトチェック機関、多面的アプローチ)

対話システムの動向

- 初期はタスク・利用シーン限定(旅行者向け、カーナビ向け等)
- スマートフォンやスマートスピーカーの音声対話アシスタント(Siri、Alexa等)
- アシスタント型に加えて雑談型の対話にも広がり(チャットボット)

大規模テキスト活用の動向

- Webサーチエンジンで大規模テキスト解析・検索技術が発展
- 情報抽出・分類・構造化によるテキストマイニング、Webマイニング
- 大量テキストを知識源とした質問応答(Watson、Wisdom X等)

自然言語処理：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

[コンピューターへの入力手段/コンピューターとのインターフェース]

[大量テキストデータの処理]

カナ漢字変換、対話システム等

質問応答、機械翻訳・音声翻訳等

Web・情報検索、文書分類、情報抽出、自動要約等

自然言語解析技術の発展

- ルールベース方式から統計的な方式、機械学習を用いた方式へ
- ニューラルネットワーク・深層学習の適用が主流に(ニューラル自然言語処理)
- 機械翻訳技術はルールベース機械翻訳(RBMT)→統計的機械翻訳(SMT)→ニューラル機械翻訳(NMT)と発展

各国の技術政策動向

- 各国とも自然言語処理単独ではなくAI戦略の一環で強化
- 米国はGAFA・MS・IBM等の産業界の取り組みと国による基礎研究への先行投資が際立っている
- 欧州は機械翻訳、中国は産業発展面に力を入れている

意味の分散表現

- 意味のベクトル演算が可能
- 分散表現への置換・再構成

End-to-End系列変換モデル

- エンコーダー＋デコーダー、E-to-E最適化
- アテンション機構、Transformerモデル

タスク設定・評価の再定義

- サブタスク評価よりもタスク設定の重要性増、目利き人材
- シェアドタスク、コンペティション

データ戦略・施策の強化

- GAFA優位対抗、低資源問題、日本語固有問題
- 戦略構築、法制度整備

ニューラル自然言語処理の発展

- 汎用モデルの事前訓練＋個別タスク毎の事後訓練によるタスク適応の効率化
- 文脈考慮による多義語分離

知識・常識の獲得と推論

- ニューラル自然言語処理に難しい問題
- 文脈・推論が必要になる問題の常識推論ベンチマーク構築・整備

人材に関わる問題

- データ戦略が人材獲得・維持にも大きく関わる

領域サマリ図⑤ 意思決定・合意形成支援

起きている問題: 意思決定・合意形成の困難化

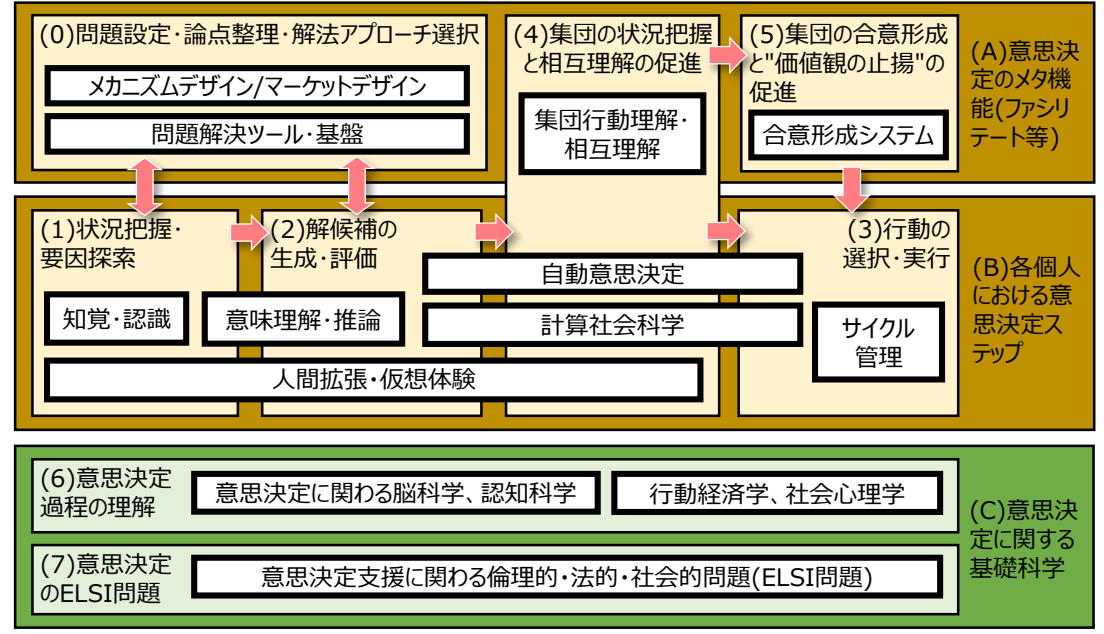
- クリティカルな要因・影響の見落とし
- ソーシャルメディアによる思考誘導(デジタルグリマンダー等)
- 価値観の対立激化、社会の分極化(エコーチェンバー等)
- 「リアルなフェイク」の流通(フェイクニュース・フェイク動画等)

困難化の原因:

- 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
- 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

意思決定・合意形成支援の形態:

- 人間に寄り添うエージェント
 - 健全な意見集約プラットフォーム
 - 人間自身の判断能力の教育・訓練

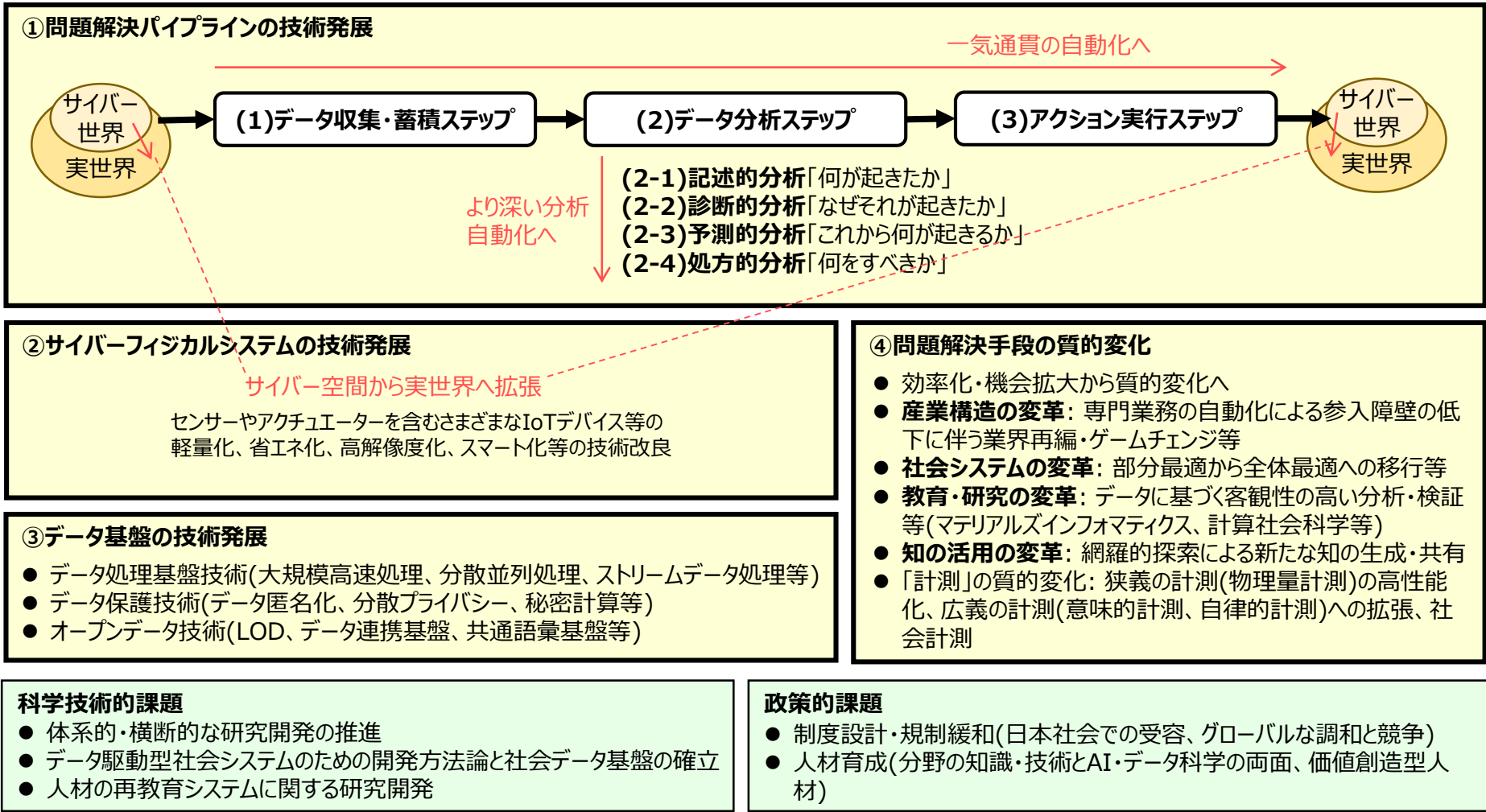


研究開発課題 (番号は右上図との対応)	これまでの取り組み	今後の課題・方向性
膨大な可能性の探索・評価 ⁽¹⁾⁽²⁾	因果関係推論、マルチエージェントシミュレーション等	因果関係知識獲得、常識推論等
自動意思決定 ⁽²⁾⁽³⁾	強化学習、予測型意思決定最適化等	多様な価値観や不完全情報での意思決定等
多様な価値観の把握・可視化 ⁽⁴⁾	gIBIS、言論マップ生成、VR・ゲーミングによる追体験等	議論マイニング、ディベートAI、法学AI等
交渉・合意形成・集合知支援 ⁽⁵⁾	討論型世論調査、自動交渉エージェント、インセンティブメカニズム等	高度な自動交渉、価値観の相互作用モデル等
フェイク対策 ⁽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

領域サマリ図⑥ データに基づく問題解決



領域サマリ図⑦ 計算脳科学

計算脳科学とクロスする分野の拡大

社会脳科学

- 社会知性の脳科学、心の理論、共感、利他性
- 他者の心のシミュレーション学習

計算脳科学の関わるスコープの拡大

- 計算精神医学：脳情報処理モデルの構築・検証による精神・神経疾患の治療・予防、DecNef法を用いたPTSD治療
- 教育・マーケティングの効果測定や新手法
- コンピューター応用：BMI (Brain-machine Interface)、脳型コンピューター

計算脳科学×人工知能(AI)

脳情報処理と機械学習

- 深層学習：ニューロンの結合を模した計算モデル
- 強化学習：ドーパミン神経細胞の報酬予測誤差仮説
- モデルフリー型(大脳基底核)とモデルベース型(前頭前野)の学習・意思決定
- アテンション、エピソード記憶、継続学習

脳情報処理からのインスパイア(さらなる可能性)

- 実世界に関する直感的な理解
- 事前知識と少量データからの効率的な学習
- 転移学習(概念を一般化して転移し、新しい概念を獲得)
- 想像(シミュレーション)とプランニング
- 仮想脳解析によるブラックボックスAIの解明

新学術領域研究「人工知能と脳科学の対照と融合」

- AI×脳科学融合領域強化
- 若手研究者の育成

DeepMind

- 知能の解明をビジョンに掲げ、Neuroscience-inspired AIを指向し、革新的AIソフトウェアを次々開発(AlphaGo他)

脳情報処理の計測・理解技術の発展

過去10～20年の間に計測技術が大きく発展

- オプトジェネティクス(光遺伝学)：動物を対象としてニューロンの詳細な活動を計測可能
- fMRI(磁気共鳴機能画像法)：人間を対象として非侵襲で脳の活動を計測可能
- 人間の活動(心の状態を含む)と脳の活動を同時計測 → 脳機能マッピング

計測に基づく脳情報処理のモデル化、ニューロフィードバック

- モデル化解析：脳の情報処理モデルと脳活動の計測結果の両面から照らし合わせて検証
- ブレインデコーディング：脳活動の計測データを機械学習の手法で解析し、心の状態を解読(脳に想起されたものを1000カテゴリ規模でデコード、対象物だけでなく印象もデコード)
- DecNef法：fMRIと機械学習技術をニューロフィードバックに活用

全脳シミュレーション

- スーパーコンピューター上にニューロンとシナプス結合から成る全脳の情報処理モデルを配置
- 2018年時点では小型のサル程度の規模(80億ニューロン)までシミュレート

脳機能の全容解明を目指した各国の大型プロジェクト

- 米国「BRAIN Initiative」
- 欧州「Human Brain Project」
- 日本「Brain/MINDS」
- 国主導トップダウン型、国際連携重視
- ビッグデータ・機械学習活用

政策的課題

- 分野間連携とバランスのよいファンディング
- 人材育成(複数分野横断)
- 大規模データ管理基盤の整備

領域サマリ図⑧ 社会におけるAI

狙い ● 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 (負の側面をできる限り抑え込み、より良い社会の実現に貢献する)

①「社会におけるAI」の課題抽出・目標設定

課題抽出(ELSI):
AI技術に関わる新たな犯罪、個人情報とプライバシー、知の所有権、機械の判断と責任、機械の自律性、人間の自由意志、アイデンティティ、労働や雇用、人間と機械の新たな関係、格差の助長や機械による人類の支配等の問題

あるべき姿や解決策の要件・目標 (倫理規定・ガイドライン等)

- 欧州: オックスフォード大学FHI、ケンブリッジ大学CSER、欧州委員会AI HLEG
- 米国: FLI(アシロマAI原則)、Open AI、Partnership on AI、スタンフォード大学AI100、IEEE EAD(IEEE-SAでの標準化活動も進行中)、MITモラルマシン実験
- 日本: 人工知能学会 倫理指針、内閣府・総務省・経産省等での検討を踏まえた「人間中心のAI社会原則」

④AIと社会との相互作用

AI技術による社会変化

- オックスフォード大学「雇用の未来」論文(AIに奪われる職業)
- 汎用AIの可能性やその脅威論、現状は特化型AIによる社会変化
- AI技術発展と社会変化のスパイラルを見据え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(①②③の有機的連携)

AIと創造性

- 絵画の生成: 深層学習・GANの活用、DeepDream、画風の変換
- 音楽の生成(自動作曲): 深層学習も使われるが、絵画に比べると理論・ルールの比重が高い
- 小説の生成: 「きまぐれ人工知能プロジェクト:作家ですよ」(ショートショート生成)

②「社会におけるAI」のための制度設計

データ保護に関する法改正

- 日本: 改正個人情報保護法(匿名加工情報、要配慮個人情報等)
- 欧州: 一般データ保護規則GDPR(AIの説明責任・透明性、忘れられる権利等)
- 米国: セクトラル方式、消費者プライバシー権利章典法案、カリフォルニア州消費者プライバシー法等

AIに対応した知的財産戦略

- 機械学習の訓練(学習)データに関する著作権法改正
- 訓練(学習)済みモデルの権利保護(派生モデル・蒸留モデルに課題)
- AI生成物の著作権

自動運転等に関わる制度整備

- 自動運転に係る制度整備大綱
- 官民ITS構想・ロードマップ
- 空の産業革命に向けたロードマップ(ドローン等の小型無人機)

③「社会におけるAI」のための技術開発

プライバシー保護のための技術開発

- データ匿名化技術(k-匿名性等)や差分プライバシー技術から、秘密計算(秘匿回路方式、準同型暗号方式、秘密分散方式)の高速化へ研究開発の中心が移っている
- 個人主導のパーソナルデータ管理(PDS、情報銀行、MyData)への動き

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

- FADM(公平性配慮データマイニング)、XAI(説明可能AI)等

中国のAIによる社会監視システム

- 金盾: インターネット通信の検閲システム
- 天網: 監視カメラネットワーク
- 社会信用システム: 国民の社会信用スコア計算

その他の課題 ● 責任ある研究・イノベーション(RRI) ● ELSI・RRIに関する組織的な取り組み・支援体制 ● 人文・社会科学者の継続的な関与に向けた取り組み ● 実環境シナリオでの実証