

次世代AI研究開発の2つの潮流 ～米中2強時代に日本は？～

2020年9月28日～11月30日

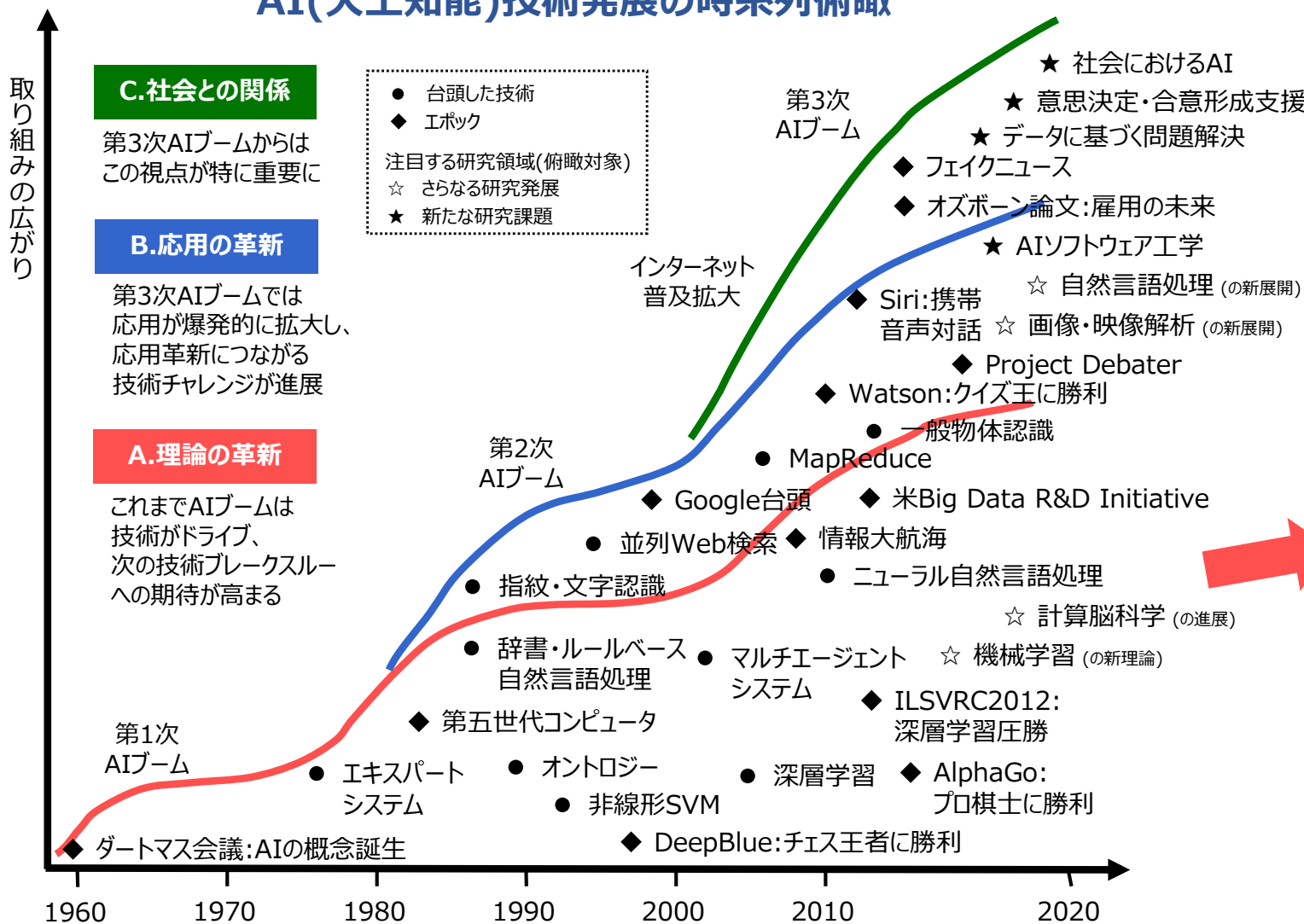
システム・情報科学技術ユニット 福島俊一



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

次世代AI研究開発の2つの潮流とは

AI(人工知能)技術発展の時系列俯瞰



潮流2

信頼されるAI

社会からの要請に応えるための研究開発

潮流1

第4世代AI

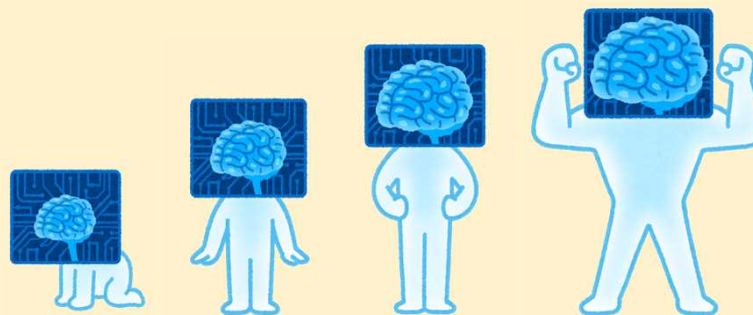
AI技術そのものの性能向上のための研究開発

[出典] 研究開発の俯瞰報告書 システム・情報科学技術分野(2019年)
<https://www.jst.go.jp/crds/report/report02/CRDS-FY2018-FR-02.html>

講演の構成

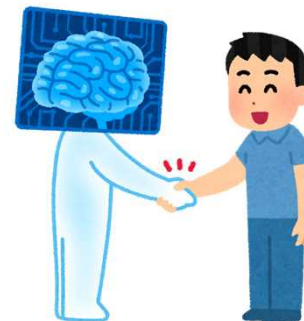
①潮流1: 第4世代AI

AI技術そのものの
性能向上のための研究開発



②潮流2: 信頼されるAI

社会からの要請に応えるための研究開発



③2つの潮流から見た 米中欧日の状況 米中2強時代に日本は？



第3次AIブームを深層学習(ディープラーニング)が牽引

- 様々な特定用途で 人間を上回る精度・性能 を次々達成
- 深層強化学習、自己教師あり学習、深層生成モデル等の技術発展も続く

画像認識(一般物体認識)で人間の精度を上回った

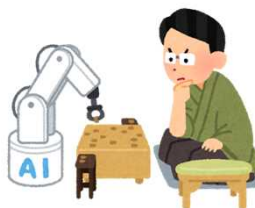
ILSVRCトップチームのエラー率の推移



[元情報] <http://www.image-net.org/challenges/LSVRC/>

ゲームでも人間に勝つシステムが続々と

- 囲碁ではGoogle DeepMindのAlphaGo (2015-2017)
- StarCraft IIではGoogle DeepMindのAlphaStar (2019)
- 6人制ポーカーではFacebookのPluribus (2019)



深層強化学習

膨大な回数の試行錯誤から
より良い行動決定方を学習

深層学習

多層ニューラルネット、
特徴設計まで自動化

文書読解テストでも人間の精度を上回った

SQuAD1.1	EM	F1
Dev set results without data aug		
BERT [10]	84.1	90.9
XLNet	88.95	94.52
Test set results on leaderboard,		
Human [27]	82.30	91.22
ATB	86.94	92.64
BERT* [10]	87.43	93.16
XLNet	89.90	95.08

<https://arxiv.org/pdf/1906.08237.pdf>

自己教師あり学習

データの一部分を隠して
その隠した部分の当て方を学習

識別モデルだけでなく生成モデルも著しく進展し、生成系応用が高品質化

ラフスケッチの写実画変換(GauGAN)



<https://youtu.be/MXWm6w4E5q0>

テキストからの
画像生成
(StackGAN)

This bird is white with some black on its head and wings, and has a long orange beak

This flower has overlapping pink pointed petals surrounding a ring of short yellow filaments

A big building with a parking lot in front of it

StackGAN
Stage-II
256x256
images



<https://arxiv.org/pdf/1612.03242.pdf>

深層生成モデル

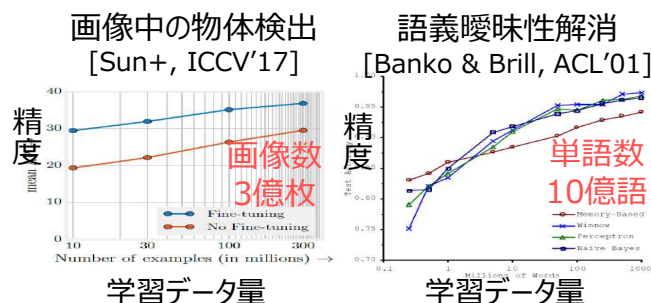
敵対的生成ネットワーク(GAN)や
変分自己符号化器(VAE)等、生成の仕方まで学習

現状の深層学習が抱える問題

① 学習に大量の教師データや計算資源が必要



- 膨大なデータ量の学習によって精度向上



- 最新アルゴリズムの実行には膨大な費用

自然言語解析の最高精度の学習アルゴリズム

- BERT (2018年) US\$ 6.9K/回
- XLNet (2019年) US\$ 61K/回

<https://syncedreview.com/2019/06/27/the-staggering-cost-of-training-sota-ai-models/>

② 学習範囲外の状況に弱く、実世界状況への臨機応変な対応ができない



- 学習範囲外ではどのような結果になるかわからない (学習データと傾向が変わらぬ環境で動くことが前提)
- 誤認識を誘発する攻撃が可能



<https://arxiv.org/pdf/1707.08945.pdf>

Adversarial Examplesに対する脆弱性 (深層学習が停止標識と認識できていたものを、人間は気につかない程度の小さな表示加工によって、速度制限標識と騙すことができてしまう)

③ パターン処理は強いが、意味理解・説明等の高次処理はできない



- 文書読解問題は意味を理解せずに傾向を当てているのが現状

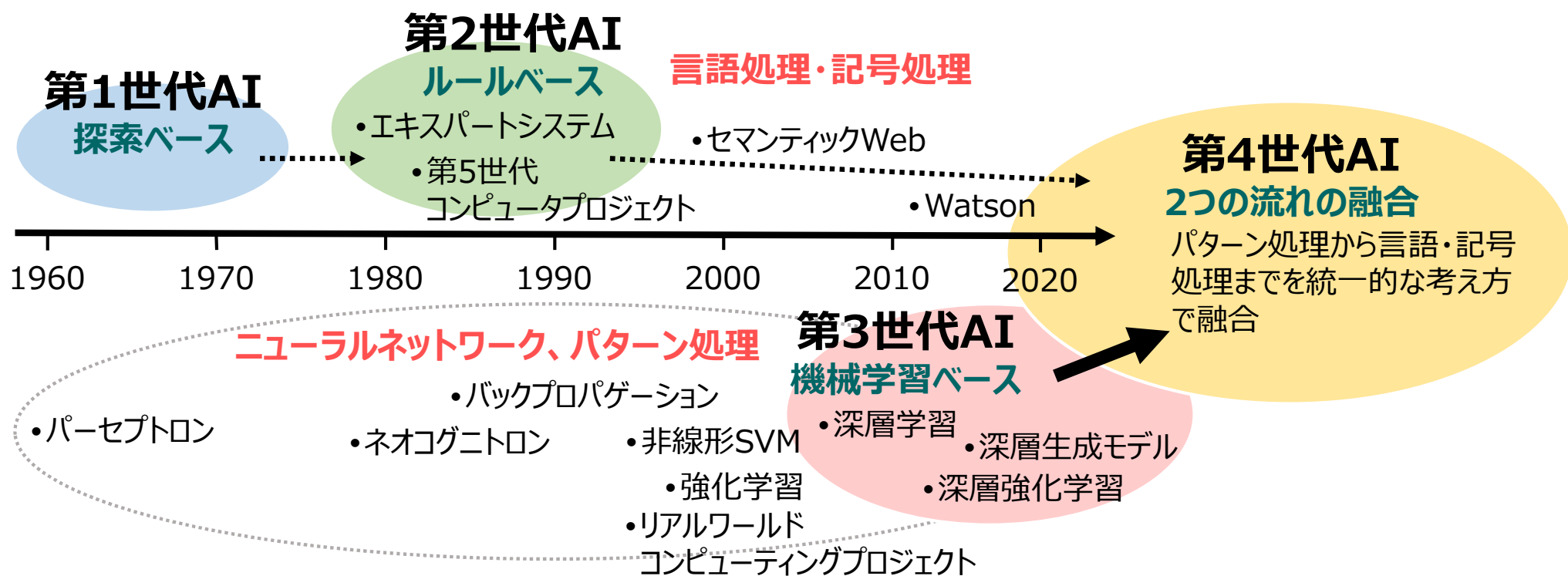


- 高い判定精度が得られたとしても、その判定理由は説明されない
- 社会的ルールに反する振る舞いを学習してしまう可能性がある



自動交渉エージェントに機械学習をさせた結果、「談合」を学習してしまった

潮流1: 第4世代AIへの進化



別々に発展してきた

● パターン処理と言語・記号処理の融合 により、第3世代AIの問題を克服

① 教師なしでも
知識獲得・成長
できるAIへ

② 実世界で
安全性・頑健性を
確保できるAIへ

③ コンテキストに応じて
適切な推論・
アクション可能なAIへ

● 人間の知能に関する基礎研究 に学ぶ

・脳科学。計算論的神経科学
・認知発達ロボティクス

- ・人間は大量の教師データなしで学習・成長し、学習したことを組み合わせて別な場面や状況にも応用できる
- ・人間の知能が持つ2つの側面のうち、現在の深層学習・強化学習がカバーするのは即応的知能のみ、熟考的知能の面が欠けている

第4世代AIに向けた注目動向

● NeurIPS 2019での基調講演

<https://youtu.be/T3sxeTgT4qc>

Yoshua Bengio (2018年チューリング賞を受賞した深層学習の父たちの一人)

“From System1 Deep Learning to System2 Deep Learning”

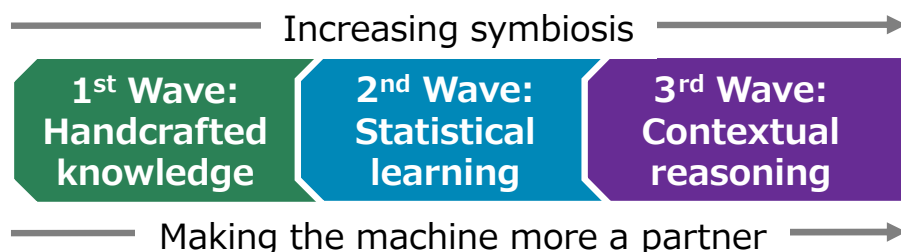
Daniel Kahneman「ファスト&スロー」	
System1: 即応的	System2: 熟考的
高速、無意識的、直感的、非言語的、習慣的	低速、意識的、論理的、言語的、計画・推論

現在の深層学習

将来の深層学習

● DARPAによるAI Next Campaign

DARPAが2018年9月に発表、次世代AI研究(AIの第3の波)に20億ドル以上の大型投資を実施



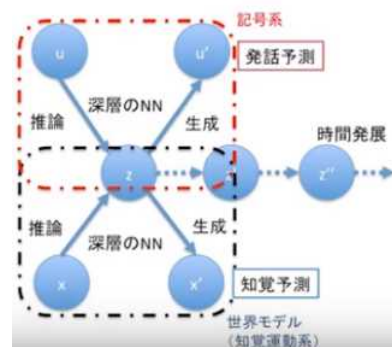
<https://www.darpa.mil/work-with-us/ai-next-campaign>

● 東大公開シンポジウム2018・2019「深層学習の先にあるもの－記号推論との融合を目指して－」

第1回: 2018年1月22日、第2回: 2019年3月5日

(講演者: 中島秀之、松尾豊、尾形哲也、麻生英樹、國吉康夫、谷口忠大、山川宏、他)

知能の2階建てモデル (東大 松尾豊教授)



- 下側は知覚・運動系(動物OS)、上側は言語系(BERT的な言語アプリ)を担う
- 下側は深層学習によって世界モデルをボトムアップに構築し、上側もそれを潜在構造として共有
- 下側は知覚予測、上側は発話予測を報酬系として持ち、2つの予測問題に対するマルチタスク学習が進む

[松尾2019]「意味理解と創造」(同シンポジウム2019での講演)

https://today.tv/contents-list/2018FY/beyond_deep_learning/01

● Neuroscience-Inspired AI

脳科学の研究成果が既にAIで活用された事例
深層学習、強化学習、アテンション、エピソード記憶、ワーキングメモリ、継続学習

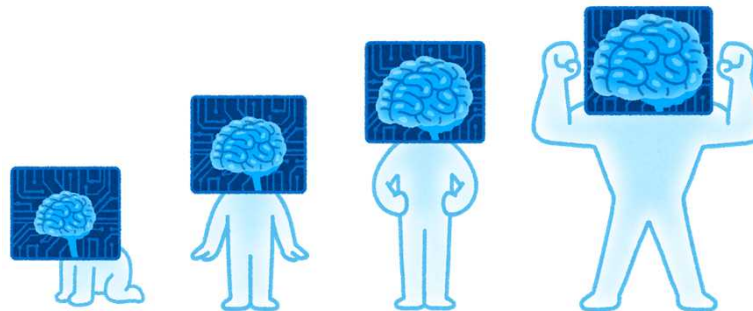
脳科学の研究成果・知見が今後AIに活用されると期待されるもの
物理的世界に関する直感的な理解、効率的な学習、転移学習、想像と計画、仮想的脳分析

- AlphaGoをはじめ最先端AI技術開発で注目されるGoogle DeepMind社の取り組み [https://www.cell.com/neuron/fulltext/S0896-6273\(17\)30509-3](https://www.cell.com/neuron/fulltext/S0896-6273(17)30509-3)
- 国内では新学術研究領域「人工知能と脳科学」(領域代表: 銅谷賢治) <http://www.brain-ai.jp/jp/>

講演の構成

①潮流1: 第4世代AI

AI技術そのものの
性能向上のための研究開発



②潮流2: 信頼されるAI

社会からの要請に応えるための研究開発



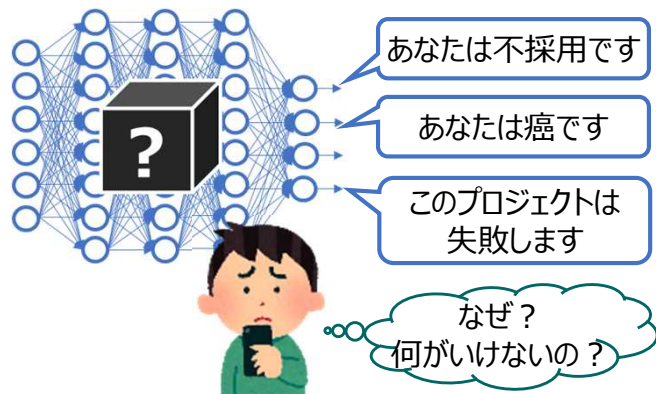
③2つの潮流から見た 米中欧日の状況 米中2強時代に日本は？



社会におけるAIの問題

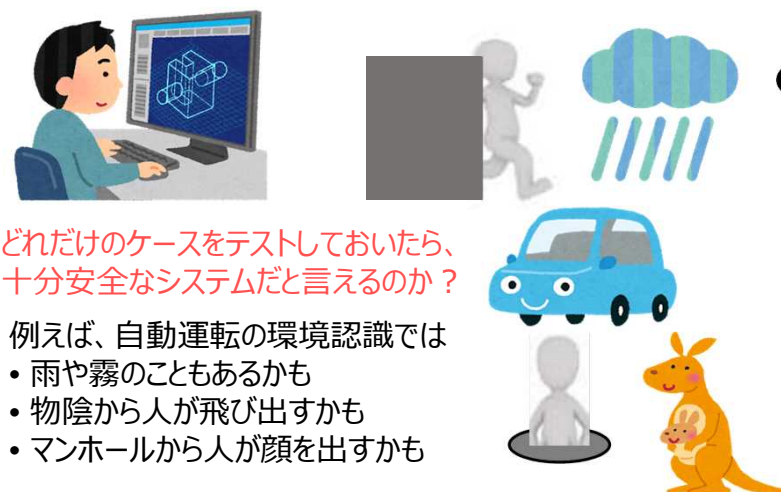
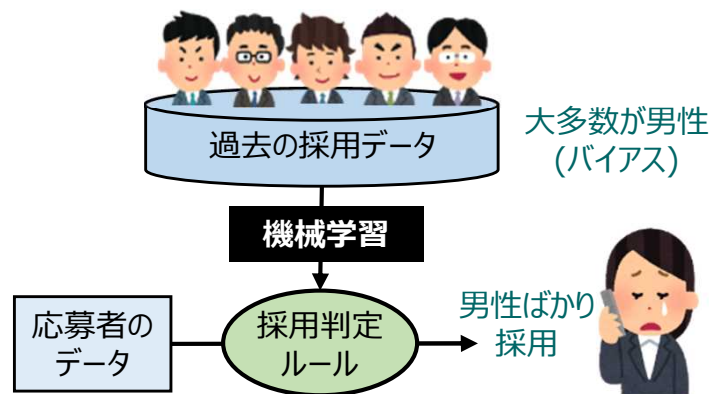
ブラックボックス問題

- 判定理由について、人間に理解可能な形で説明してくれない
- 事故発生時に、原因解明や責任判断ができない



バイアス問題

- 学習データが偏見を含んでいると、判定結果に偏見が反映される
- 学習データの分布の偏りが差別を生むこともある



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？

- 例えば、自動運転の環境認識では
- 雨や霧のこともあるかも
 - 物陰から人が飛び出すかも
 - マンホールから人が顔を出すかも

脆弱性問題

- 学習範囲外のデータに対して、どう振る舞うかは不明
- 誤認識を誘発する攻撃が可能 (Adversarial Examples)
- 悪意をもった追加学習で、不適切な振る舞いが引き起こされる



フェイク問題

- 本物と見紛うフェイク動画等の生成が可能になった
- フェイクニュースによる思考誘導・世論操作の危険性増大

- 自動生成された実在しない有名人風の顔画像
<https://youtu.be/XOxxPcy5Gr4>
- 政治家に思い通りの発言をさせるフェイク動画
<https://youtu.be/cQ54GDm1eL0>
- 別人の動作をコピーして振る舞わせたフェイク動画
<https://youtu.be/PCBTZh41Ris>

品質保証問題

- 仕様(正動作)が定義されないため、テストの成否が定まらない
- 精度100%は無理、間違いは不可避、動作保証ができない

AI社会原則・倫理指針

- 急速に社会に広がるAIに対して、社会からの要請を明文化する動きが、国・世界レベルで活発化、標準化活動も進められている

世界動向

- **日本「人間中心のAI社会原則」** 2019年3月
- 欧州委員会「信頼できるAIのための倫理指針」 2019年4月
(Ethics Guidelines for trustworthy AI)
- OECD「人工知能に関するOECD原則」 2019年5月
(OECD Principles on Artificial Intelligence) 42か国が署名
- G20サミット/貿易・デジタル経済大臣会合
「人間中心」を踏まえたAI原則 2019年6月
- 中国「次世代AIガバナンス原則－責任あるAIの発展」 2019年6月

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」をビジョンに掲げる

AI社会原則

- 人間中心の原則
- プライバシー確保の原則
- 公正競争確保の原則
- イノベーションの原則
- 教育・リテラシーの原則
- セキュリティ確保の原則
- 公平性・説明責任・透明性の原則

<https://www8.cao.go.jp/cstp/aigensoku.pdf>

学会および標準化活動

- 人工知能学会 倫理指針
- IEEE「倫理的に配慮されたデザイン」
(Ethically Aligned Design)
IEEE Standard Associationでの標準化活動も進行中
- ISO/IEC JTC1 SC42でAIに関わる標準化活動

産業界

- ソニーグループ AI倫理ガイドライン
- NECグループ AIと人権に関するポリシー
- 富士通グループ AIコミットメント
- NTTデータグループ AI指針
- Microsoft AI principles
- DeepMind: Building artificial intelligence: specification, robustness and assurance

だが、このような原則・指針を満たす「信頼されるAIシステム」はどう作ればよいのか？

潮流2: 信頼されるAIのための研究課題

AI応用
システム



システム開発のパラダイム転換が起きている

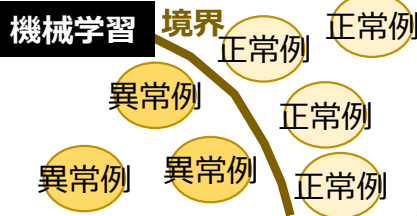
演繹的な作り方から帰納的な作り方へ転換

if 条件a then 処理A
else 処理B

条件a: 温度センサ > 90℃
(異常と判定する条件)

手続きや判定ルール
を明示的に書くことで
システムの動作が決
まる(プログラミング)

機械学習



データを例示すると、それを真似て
自動的に判定ルールができ、シス
テムの動作が決まる(機械学習)

品質保証問題

脆弱性問題

バイアス問題

ブラックボックス問題

フェイク問題

AIソフトウェア 工学

AI応用システムの安全
性・信頼性を確保する
開発方法論・技術体系

AI技術の進化

深層学習 深層生成モデル

深層強化学習 転移学習

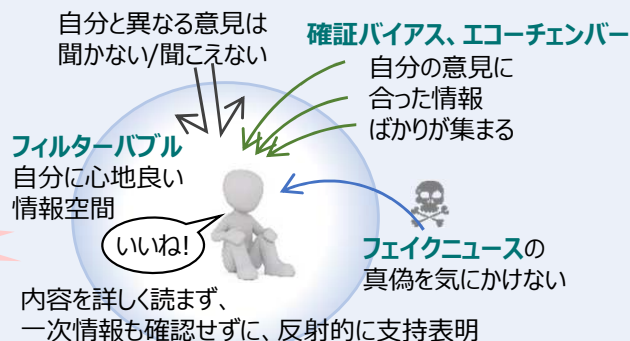
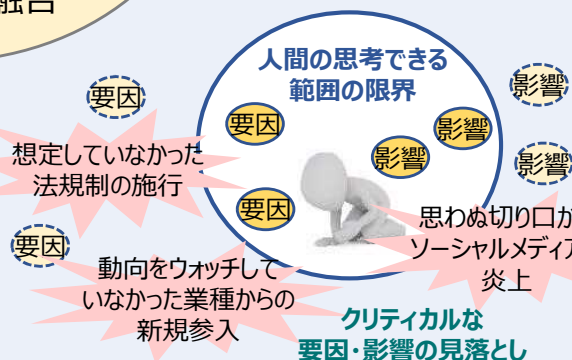
自己教師あり学習 ...

記号推論との融合

限定合理性問題
の深刻化

確証バイアス問題
の顕著化

AIと人間



主体的 意思決定の支援

フェイクに騙されず、熟
慮・熟議に基づく人間の
主体的意思決定を促進

人間の意思決定の主体性が揺らいでいる

信頼されるAIに向けた多面的な取り組み

説明可能AI

- ブラックボックス型を外側から調べて近似的説明を試みる手法
(LIME、Influence等、多数ツールがある)



(a) Original Image (b) Explaining Electric guitar (d) Explaining Labrador
<https://arxiv.org/pdf/1602.04938.pdf>

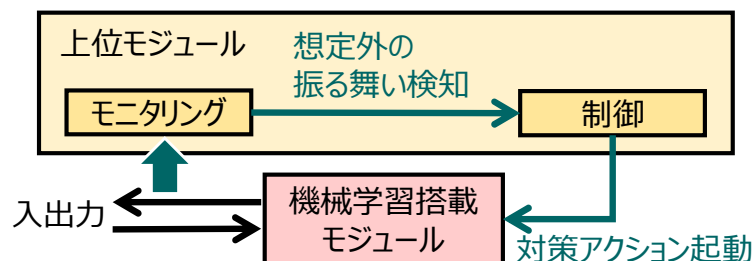
- もともと説明可能なホワイトボックス型の精度を高める手法
(異種混合学習等)

公平なAI

- 公平性配慮型機械学習: 公平性の指標(集団公平性・個人公平性等)を定義し、その指標と精度のトレードオフを調整

AIシステムのテスト

- 起こり得るケースの網羅は不可能だが、効率よく重要ケースをテスト
(メタモルフィックテスト、ニューロンカバレッジ等)
- 100%保証はできない機械学習型に、従来型を組み合わせシステム全体として安全性を確保
(Safety Envelope、Safe Learning等)



AI品質管理

- AI・機械学習の応用システムの品質の考え方やチェックリストの整備
(QA4AIコンソーシアム「AIプロダクト品質保証ガイドライン」、産総研「機械学習品質マネジメントガイドライン」等)

フェイク検知

- 発信元・伝播特性・表現等の分析や知識照合・相互比較によるフェイクニュース判定
- 映像の詳細解析によるフェイク動画判定

- ✓ 不自然なまばたきの仕方
- ✓ 不自然な頭部の動き、奇妙な目の色
- ✓ 画像から読み取れる人間の脈拍数
- ✓ 映像のピクセル強度のわずかな変化
- ✓ 照明や影などの物理的特性の不自然さ
- ✓ 日付・時刻・場所と天気との整合 等

<http://ainow.ai/2018/10/15/148590/>

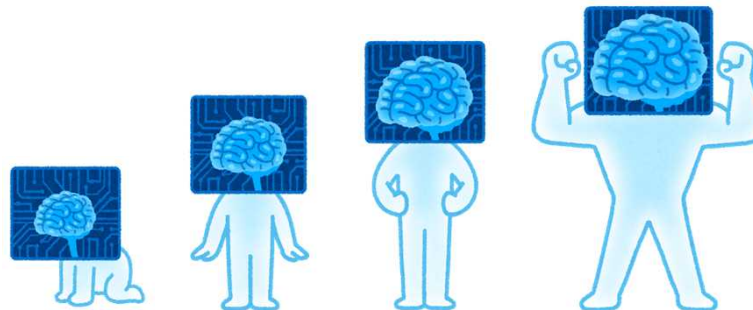
熟慮・熟議の促進

- 様々な立場の意見・言明を集めた集合知を構築・活用
(賛成・反対両面を見せる言論マップ、自動ファシリテーションによる大規模意見集約システムD-Agree等)

講演の構成

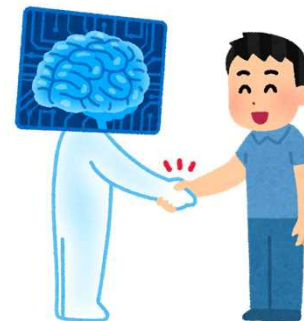
①潮流1: 第4世代AI

AI技術そのものの
性能向上のための研究開発



②潮流2: 信頼されるAI

社会からの要請に応えるための研究開発



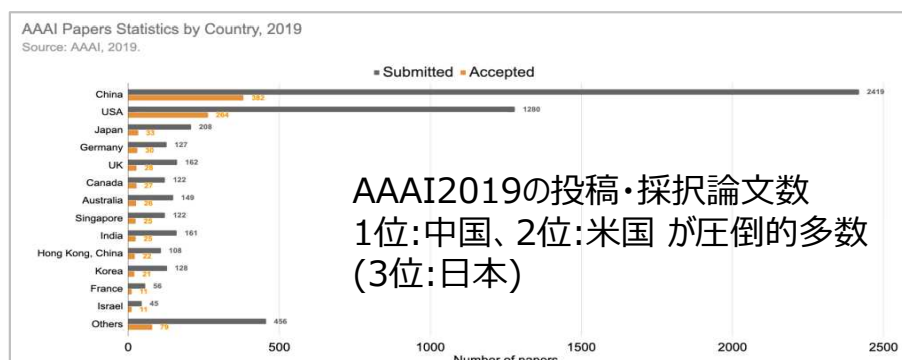
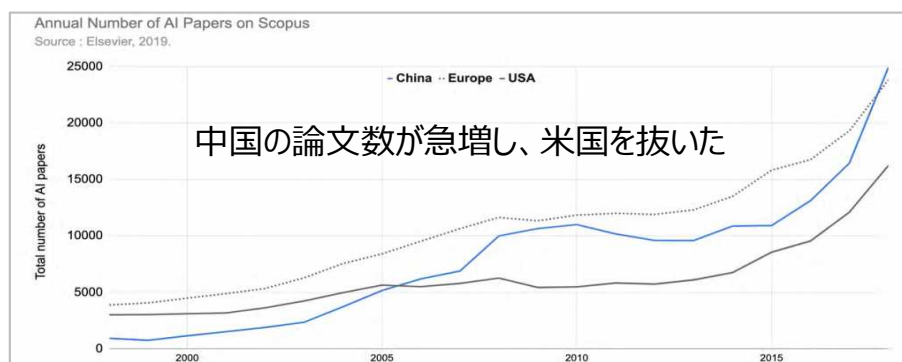
③2つの潮流から見た
米中欧日の状況
米中2強時代に日本は？



米中2強の状況、2つの潮流への取り組み方の違い

	潮流1: AI技術そのものの性能向上	潮流2: 社会からの要請に応える
米国	革新技术創出から産業化まで強みを保有・牽引	幅広い観点から研究の取り組みがあり、層が厚い
中国	技術の改良・実装速度で凌駕	国として原則は掲げたが、市場での実践は伴っていない
欧州	強い部分はあるが、米中ほど産業化の勢いが無い	理念・倫理ガイドラインを重視・施策化
日本	強い部分はあるが、米中に比べて層が薄い	信頼性・品質確保のための具体的取り組みでやや先行

AI分野の論文数



[出典] https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf

AI研究をリードする研究組織ランキング

Publication Index 2019 (NeurIPS, ICML)

順位	企業名	国	スコア
1	Google	米国	167.3
2	スタンフォード大学	米国	82.3
3	MIT	米国	69.8
4	カーネギーメロン大学	米国	67.7
5	UCバークレー	米国	54.0
6	マイクロソフト	米国	51.9

[出典] <https://medium.com/@chuvpilo/ai-research-rankings-2019-insights-from-neurips-and-icml-leading-ai-conferences-ee6953152c1a> (和訳) <https://ainow.ai/2020/01/13/182173/>

世界AIユニコーン企業ランキング

順位	企業名	国	評価額
1	Toutiao (Bytedance)	中国	\$75B
2	SenseTime	中国	\$7.5B
3	Argo AI	米国	\$7.25B
4	Automation Anywhere	米国	\$6.8
5	UiPath	米国	\$6.4
6	Face++ (Megvii)	中国	\$4B

[出典] <https://www.cbinsights.com/research-unicorn-companies>

AI関連の政府予算

(2018年)

- 米国: 5000億円
- 中国: 4500億円
- 日本: 770.4億円

AI関連の民間投資

(2018年)

- 米国: 7兆円超
- 中国: 6000億円超
- 日本: 6000億円超

[出典]

<https://www.sankeibiz.jp/macro/news/180226/mca1802260500001-n1.htm>

米中欧日の国際競争戦略

米国	- GAFAが<u>ビジネスと基礎研究の両面で圧倒的に優位</u> - GAFAやスタートアップによる民間の活発な技術開発の一方、<u>DARPAが国としての中長期的な研究投資(AI Next)</u>をシャープに打ち出している - 経済・国家安全保障のためのAIを強化、2019年2月「AIにおける米国のリーダーシップ維持」を大統領令で宣言(AIイニシアチブ)
中国	2017年「次世代人工知能発展計画(AI2030)」を掲げ、AIリード企業5社を選定、政府がAI産業を後押し <u>スタートアップの勢い</u>：技術実装スピード、質より量、BtoC中心にビッグデータ獲得がAI活用を加速 - 政府は<u>AIを活用した監視・管理社会</u>(社会信用システム、天網、金盾)の構築推進、他国と大きく異なるAI応用技術開発を推進
欧州	- 欧州委員会「信頼できるAIのための倫理ガイドライン」を策定(2019年4月) <u>データ保護戦略(GDPR)</u>を含めてAIに関わる国際ルール作りを通じて米中やGAFAに対抗 「AI for Europe」を掲げ、2018年12月「AI協調計画」を発表、各国のAI戦略に加えて、Horizon 2020/Europeによる国横断のAI研究推進
日本	「人間中心のAI社会原則」「AI戦略2019」を策定、<u>信頼される高品質なAI(Trusted Quality AI)</u>、信頼性のある自由なデータ流通に向けたDFFT(Data Free Flow with Trust)を打ち出し - 理研AIP、産総研AIRC、NICTが中核国研として国のAI研究を牽引 - 脳科学や<u>日本発の認知発達・記号創発ロボティクス</u>等、人間の知能に関する基礎研究に存在感

[出典] 戦略プロポーザル：第4世代AIの研究開発—深層学習と知識・記号推論の融合—

<https://www.jst.go.jp/crds/pdf/2019/SP/CRDS-FY2019-SP-08.pdf>

まとめ：米中2強時代に日本は？

①潮流1：第4世代AI

AI技術そのものの性能
向上のための研究開発



②潮流2：信頼されるAI

社会からの要請に
応えるための研究開発



③2つの潮流から 見た米中欧日の 状況



第4世代AIで先行を狙う× 高品質AIの強みを確立

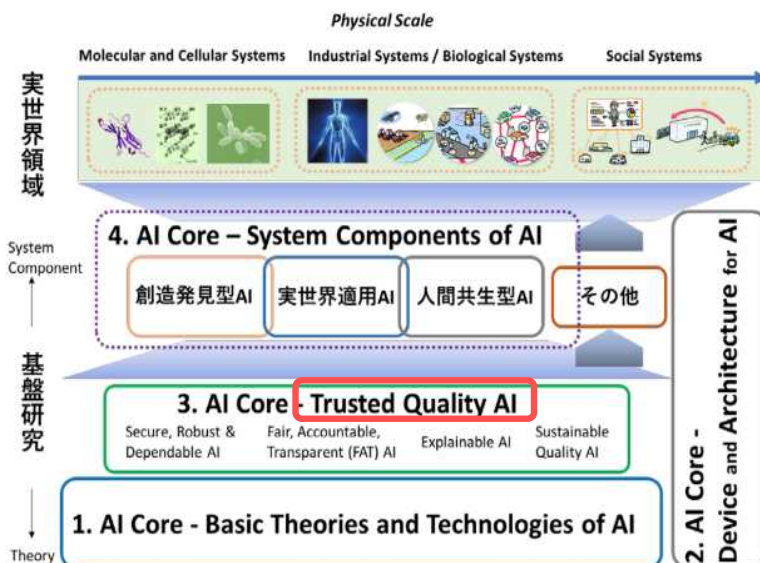
- 第4世代AIはどの国もスタートライン、日本にも先行チャンスあり
- 知能基礎研究の強みやロボットとの融合も活用



- 日本の産業界が従来から持っている品質・信頼性へのこだわりを活かす
- AI品質に関する具体的な取り組みで先行

日本政府 「AI戦略2019」

<https://www.kantei.go.jp/jp/singi/tougou-innovation/pdf/aisenryaku2019.pdf>



国のファンドによる研究開発プログラム

国立研究開発法人 科学技術振興機構(JST)

文科省戦略目標「信頼されるAI」を受けた戦略的創造研究推進事業

【CREST】信頼されるAIシステムを支える基盤技術

【さきがけ】信頼されるAIの基盤技術

国立研究開発法人 新エネルギー・産業技術総合開発機構(NEDO)

「人と共に進化する次世代人工知能に関する技術開発事業」

【研究開発項目1】人と共に進化するAIシステムの基盤技術開発

【研究開発項目2】実世界で信頼できるAIの評価・管理手法の確立

【研究開発項目3】容易に構築・導入できるAI技術の開発

より詳しい内容はCRDS報告書をご参照ください

- 研究開発の俯瞰報告書：システム・情報科学技術分野（2019年）
<https://www.jst.go.jp/crds/report/report02/CRDS-FY2018-FR-02.html>

潮流1に関して:

- 戦略プロポーザル：第4世代AIの研究開発—深層学習と知識・記号推論の融合—（2020年）
<https://www.jst.go.jp/crds/pdf/2019/SP/CRDS-FY2019-SP-08.pdf>
- 科学技術未来戦略ワークショップ報告書：深層学習と知識・記号推論の融合によるA I 基盤技術の発展（2020年）
<https://www.jst.go.jp/crds/pdf/2019/WR/CRDS-FY2019-WR-08.pdf>
- JSAI2020企画セッション報告書：次世代AI研究開発—さらなる進化に向けて—（2020年）[近日公開予定]

潮流2に関して:

- 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（2018年）
<https://www.jst.go.jp/crds/report/report01/CRDS-FY2018-SP-03.html>
- 科学技術未来戦略ワークショップ報告書：機械学習型システム開発へのパラダイム転換（2018年）
<https://www.jst.go.jp/crds/report/report05/CRDS-FY2017-WR-11.html>
- 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（2018年）
<https://www.jst.go.jp/crds/report/report01/CRDS-FY2017-SP-03.html>
- 科学技術未来戦略ワークショップ報告書：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（2017年）
<https://www.jst.go.jp/crds/report/report05/CRDS-FY2017-WR-05.html>
- 公開ワークショップ報告書：意思決定のための情報科学～情報氾濫・フェイク・分断に立ち向かうことは可能か～（2020年）
<https://www.jst.go.jp/crds/pdf/2019/WR/CRDS-FY2019-WR-02.pdf>

最後までご覧頂きありがとうございました。

■ CRDSの活動・成果・コラムはウェブサイトでご覧いただけます



<https://www.jst.go.jp/crds/>



各種報告書・コラム無料公開中

■ 最新情報はメールマガジン、Facebookでも配信中



<https://www.jst.go.jp/melmaga.html>



<https://fb.com/170314426446196/>

■ その他 お問い合わせはこちらまで



crds@jst.go.jp