

全脳アーキテクチャシンポジウム2024

# 基盤モデル・生成AIの限界とAGIの可能性

2024年9月18日

福島 俊一 JST CRDS

[toshikazu.fukushima@jst.go.jp](mailto:toshikazu.fukushima@jst.go.jp)



# 基盤モデル・生成AIの限界とAGIの可能性

## 1. 講演の背景・立場

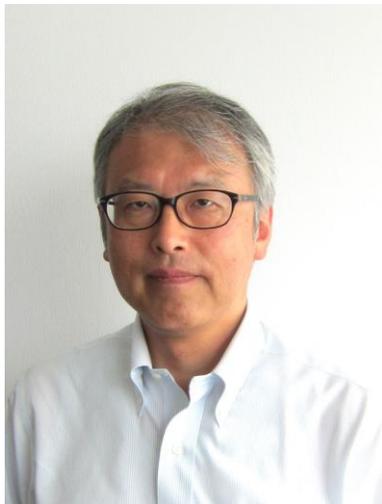
2. 戦略プロポーザル「次世代AIモデルの研究開発」の骨子

3. 基盤モデル・生成AIの限界

4. 限界克服に向けた次世代AIモデルの研究動向とAGI

# 自己紹介

## 福島 俊一 Toshikazu Fukushima



1982年東京大学理学部物理学科卒業、NEC入社。以来、中央研究所にて自然言語処理・サーチエンジン等の研究開発・事業化、人工知能(AI)・ビッグデータ研究開発戦略を担当。工学博士。情報処理学会フェロー。

2016年4月から科学技術振興機構(JST)研究開発戦略センター(CRDS)フェロー。 AIを中心とした情報科学技術分野の俯瞰的調査・戦略提言に従事。 2019年からNEDO技術委員を兼任、AI・ロボット関連NEDOプロジェクトの審査・推進にも参画。

2011～2013年東京大学大学院情報理工学研究科客員教授。1992年情報処理学会論文賞、1997年情報処理学会坂井記念特別賞、2003年オーム技術賞ほかを受賞。2015～2017年人工知能学会理事、2018～2020年人工知能学会監事。

趣味は料理・蕎麦打ち・パン作り、絵を描くことやアート展巡りなど。

# JST CRDS 報告書 ～AI・トラスト関連～

## 分野俯瞰

- **人工知能研究の新潮流2 ～基盤モデル・生成AIのインパクト～** (2023年)
- 人工知能研究の新潮流 ～日本の勝ち筋～ (2021年)
- 俯瞰ワークショップ報告書：エージェント技術 (2022年)
- 俯瞰ワークショップ報告書：ヒューマンインタフェース研究動向 (2023年)
- 研究開発の俯瞰報告書：システム・情報科学技術分野 (2023年)
- プレプリントサーバーarXivを利用したAI分野の研究動向俯瞰調査 (2024年)

## 戦略提言(1)次世代AIモデル

- **戦略プロポーザル：次世代AIモデルの研究開発** (2024年)
- 科学技術未来戦略ワークショップ報告書：次世代AIモデルの研究開発 ～技術ブレークスルーとAI×哲学～ (2024年)
- 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)
- 科学技術未来戦略ワークショップ報告書：深層学習と知識・記号推論の融合によるAI基盤技術の発展 (2020年)
- 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)

## 戦略提言(2)AIソフトウェア工学

- **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)
- 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)

## 戦略提言(3)意思決定・合意形成支援

- **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)
- 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)
- 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)



## CRDS報告書 情報・システム分野

全文PDFダウンロード可能

## 戦略提言(4) デジタル社会のトラスト

- **戦略プロポーザル：デジタル社会における新たなトラスト形成** (2022年)
- 俯瞰セミナー&ワークショップ報告書：トラスト研究の潮流 ～人文・社会科学から人工知能、医療まで～ (2022年)
- 科学技術未来戦略ワークショップ報告書：トラスト研究戦略 ～デジタル社会における新たなトラスト形成～ (2022年)
- 公開シンポジウム報告書「デジタル社会における新たなトラスト形成 ～総合知による取り組みへ～」 (2023年)

## 戦略提言(5) AI駆動科学

- **戦略プロポーザル：人工知能と科学～AI・データ駆動科学による発見と理解～** (2021年)
- 俯瞰セミナーシリーズ報告書：機械学習と科学 (2021年)
- 科学技術未来戦略ワークショップ報告書：人工知能と科学 (2021年)
- 計測横断チーム調査報告書 計測の俯瞰と新潮流 (2018年)

# 活用いただいた実績やアウトリーチ事例 (の一部)

## 戦略提言(1)(2)(3)

- ▶ **文科省** 2020年度**戦略目標**「信頼されるAI」
- ▶ **JST CREST**「信頼されるAIシステム」2020年度～
- ▶ **JSTさきがけ**「信頼されるAI」2020年度～

## 戦略提言(4)

- ▶ **JST RISTEX**「デジタル ソーシャル トラスト」2023年度～

## 戦略提言(5)

- ▶ **文科省** 2024年度**戦略目標**「自律駆動による研究革新」
- ▶ **JSTさきがけ**「研究開発プロセス革新」2024年度～
- ▶ **文科省**「AI for Science」(TRIP-AGIS) 2024年度～

## 人工知能研究の新潮流2

- ▶ **文科省**「科学技術・イノベーション白書」2024年版

## 戦略提言(2)

- ▶ **経団連**「AI活用戦略」2019年



## 人工知能学会全国大会での企画セッション等

### 戦略提言(1)

- ▶ JSAI2024「次世代AIモデルの研究開発へ」
- ▶ JSAI2022「AI・シミュレーション融合研究の展望と戦略」
- ▶ JSAI2020「次世代AI研究開発」

### 戦略提言(2)

- ▶ JSAI2019「機械学習における説明可能性・公平性・安全性への工学的取り組み」

### 戦略提言(3)

- ▶ JSAI2019～2021(OS)「複雑化社会における意思決定・合意形成のためのAI技術」

### 戦略提言(4)

- ▶ JSAI2023「人工知能とトラスト」

### 戦略提言(5)+

- ▶ JSAI2021「量子×AI」「AIと数学」

# 本講演に対応するCRDS報告書



## 人工知能研究の新潮流2 ~基盤モデル・生成AIのインパクト~

CRDS-FY2023-PR-02 (2023年7月)



## 次世代AIモデルの研究開発 ~技術ブレークスルーとAI×哲学~

CRDS-FY2023-WR-03 (2024年2月)



## 戦略プロポーザル：次世代AIモデルの研究開発

CRDS-FY2023-SP-03 (2024年3月)



# 専門家ヒアリング

- 戦略プロポーザル等の作成では、文献調査だけでなく、専門家へのヒアリングや意見交換を重視
- 年間100件超の学会・講演会等の聴講、50名超の専門家インタビュー、ワークショップでの議論等
- 「次世代AIモデルの研究開発」の戦略プロポーザル作成では、2タイプのワークショップを実施

## ワークショップ 第1回: 技術ブレークスルー編

2023年11月23日(木) 13:00~17:00

**話題提供:** 黒橋禎夫(NII)、東中竜一郎(名古屋大)、牛久祥孝(OSX)、山川宏(WBAI)、尾形哲也(早大)、松尾豊(東大)

**ディスカッション:** 大森久美子(京都哲学研究所/NTT)、高橋恒一(理研)、谷口忠大(立命館大)、銅谷賢治(OIST)、村上祐子(立教大)

**論点:** 現在のAIの本質的問題と次世代AIモデルへの技術的アプローチを議論

- ① 現在の基盤モデル・生成AIの本質的な問題点は何か？
- ② その問題点を解決する次世代AIモデルへの重要なアプローチは何か？ その世界的な動向や日本のポジションは？
- ③ 米中二強で劣勢の日本が取り得る戦略・方策は？

## ワークショップ 第2回: AI×哲学編

2023年12月20日(水) 9:00~12:30

**話題提供:** 三宅陽一郎(スクエニ)、鈴木貴之(東大)、谷口忠大(立命館大)、田口茂(北大)、大森久美子(京都哲学研究所/NTT)、高木俊一(京都哲学研究所/京大)、村上祐子(立教大)、高橋恒一(理研)

**ディスカッション:** 牛久祥孝(OSX)、尾形哲也(早大)、黒橋禎夫(NII)、銅谷賢治(OIST)、山川宏(WBAI)

**論点:** AIの知能としての根本的課題やAIと人間・社会との関係のあるべき姿を議論

- ① 人工知能学会「AI哲学マップ」の「哲学からAIへの15の批判」を起点としてを深掘り(a.現在のAIの知能としての根本的課題、次世代AIモデルに求める要件、b. AIと人間・社会の相互作用・スパイラル、AIと人間・社会との関係のあるべき姿)
- ② このような議論を継続し、AI研究に反映していくための学際的研究の進め方・課題

# 基盤モデル・生成AIの限界とAGIの可能性

1. 講演の背景・立場

2. 戦略プロポーザル「次世代AIモデルの研究開発」の骨子

3. 基盤モデル・生成AIの限界

4. 限界克服に向けた次世代AIモデルの研究動向とAGI

# 基盤モデル・生成AIに関する課題の全体観と戦略提言の狙い

- **現状認識**：日本国内では基盤モデル・生成AIの後追い開発や応用開発への取り組みが活発化しており、国際的には喫緊の問題への対策・ルール整備なども進められている状況
- **戦略提言の狙い**：活発化している基盤モデル・生成AIの後追い開発や応用開発にとどまらず、その先の次世代AIモデルを創出する基礎研究の戦略強化を狙う



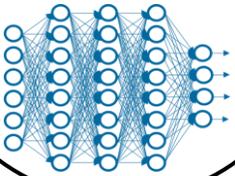
既に活発な取り組みが、国際競争の中で進んでおり、走りながら迅速に手を打っていくべき課題

基礎研究として重点的に取り組むべき課題

# 次世代AIモデルの創出に向けた研究開発課題

なぜあれほど賢く見える振る舞いを示すのか、理由が明らかになっていない

現在の  
基盤モデル  
・生成AI



資源効率、実世界操作(身体性)、論理性・正確性、安全性・信頼性などの限界が指摘されており、その克服が望まれる

**研究開発課題①**  
次世代AIモデルの基本原則

社会課題の解決、より良い情報社会の実現に向けて、次世代AIによるプロセス革新が期待される

次世代  
AIモデル

人・AI  
共生社会

AIの急速な発展に伴い、社会において発生する様々なリスクを予見し対処する方法論・技術の確立が求められる

**研究開発課題②**  
AIリスクへの対処技術

**研究開発課題③**  
AI駆動型プロセス革新

人とAIの関係や社会の在り方として、どのような姿を目指すかを考えてAIの研究開発を進めるべき

**研究開発課題④**  
人・AI共生社会の在り方

# AI分野の研究開発形態の変化を踏まえた推進方策

## ビッグサイエンス化

基礎研究に必要な計算資源・データの超大規模化、その運営は研究者だけでなく多様なスタッフを擁した大規模プロマネが必要

## ハイスピード化・ハイインパクト化

社会・生活に広く影響を与え得る技術、その影響範囲の予測困難な技術が次々に(数週間単位で)生まれている

## 非オープン化

ビッグテック企業が最先端の研究成果・知見を保有し、その内容が公開されない(公共財化されない)傾向が強まっている

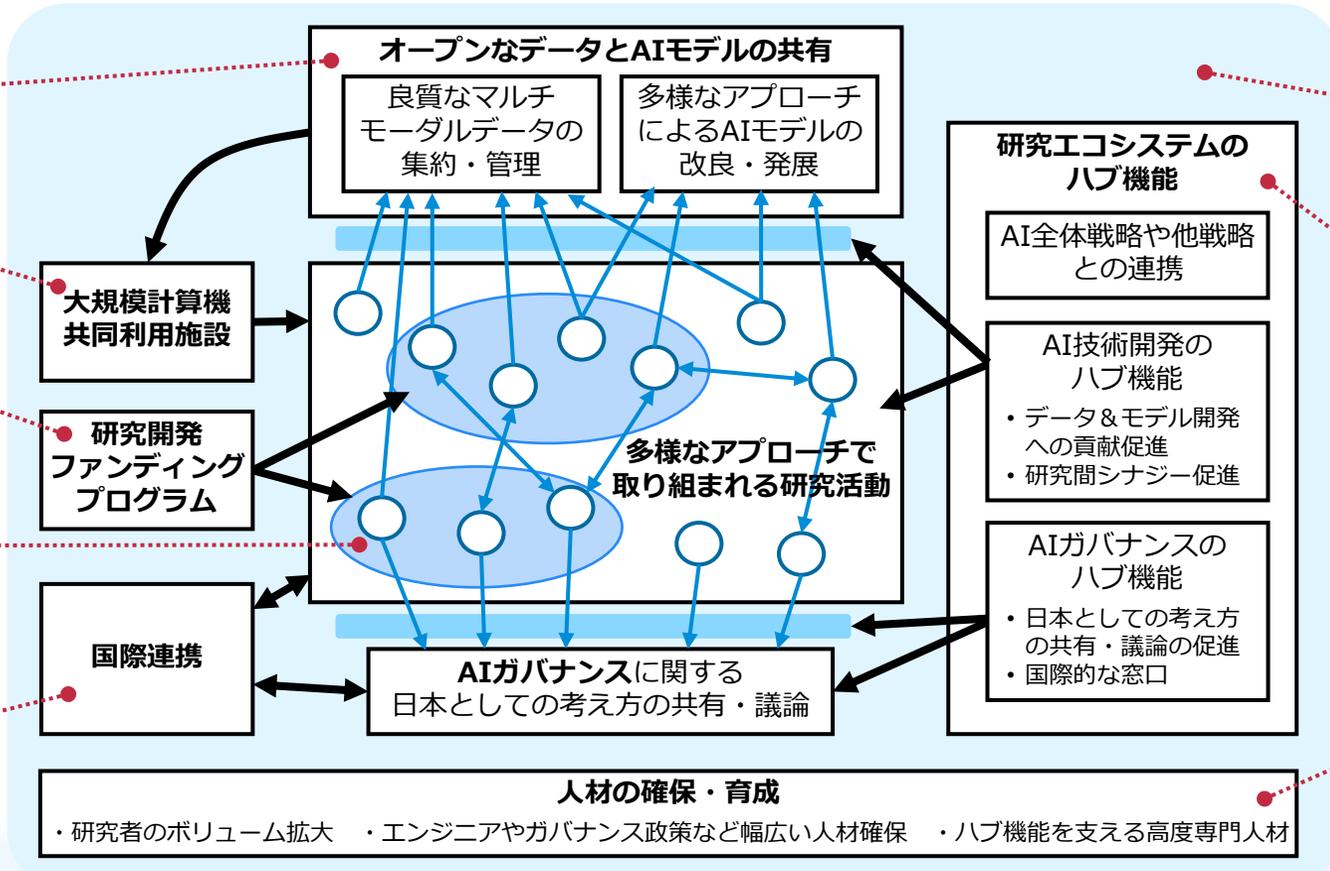
c. AIモデルとマルチモーダルデータの集約・共有・評価・管理体制の整備

b. 大規模計算機の共同利用施設の継続的な運用・強化

g. 研究エコシステムを生かした柔軟でアジャイルなプログラム運営

e. 技術系研究者のみならず人文・社会系研究者の主体的参画を促進するプログラム設計

d. 基礎研究とルールメイキングにおけるオープンな国際連携とその支援体制構築



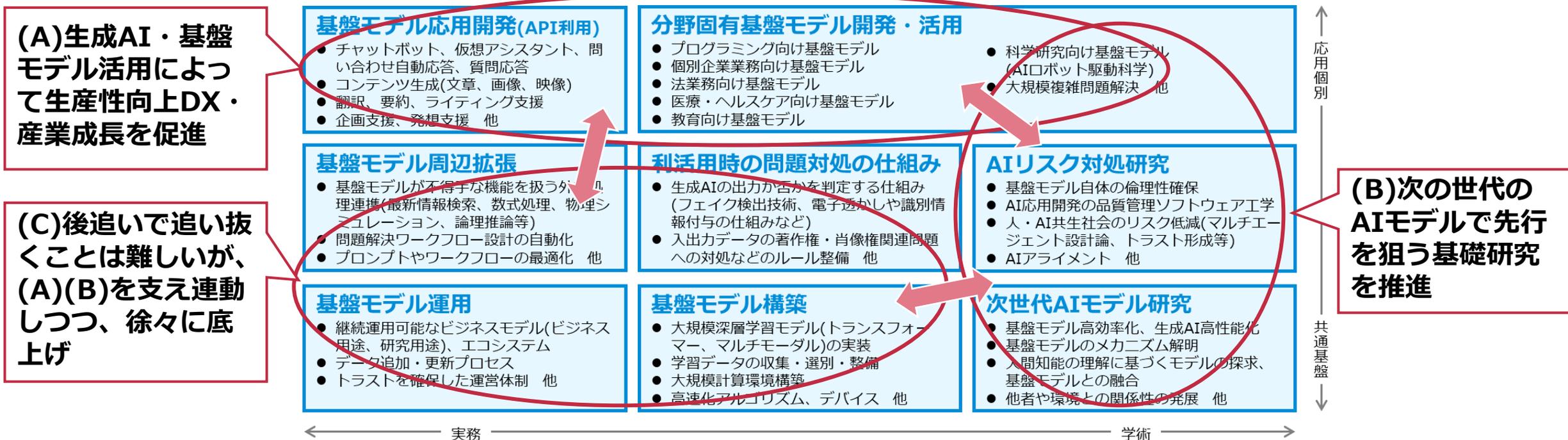
a. さまざまな研究機関・組織が結集し協力し合う研究エコシステムの形成

f. 研究エコシステムのハブ機能を担う組織の設置

h. 研究エコシステムを支える人材の確保・育成

# ビッグテック企業先行の状況下 後追いの日本の戦略は？

- 次世代AIでの先行を狙うものの、そのための基礎研究(B)だけで勝てるというものではなく、下図(A)(B)(C)を連動した推進が不可欠
- 米中の巨額投資に国家投資だけで並ぶのは無理、産業成長(A)とともに回していけるかがカギ
- そういった中で「AI品質・トラスト」「人・AI共生社会」との親和性は日本の強みになり得るのではないかと



# 基盤モデル・生成AIの限界とAGIの可能性

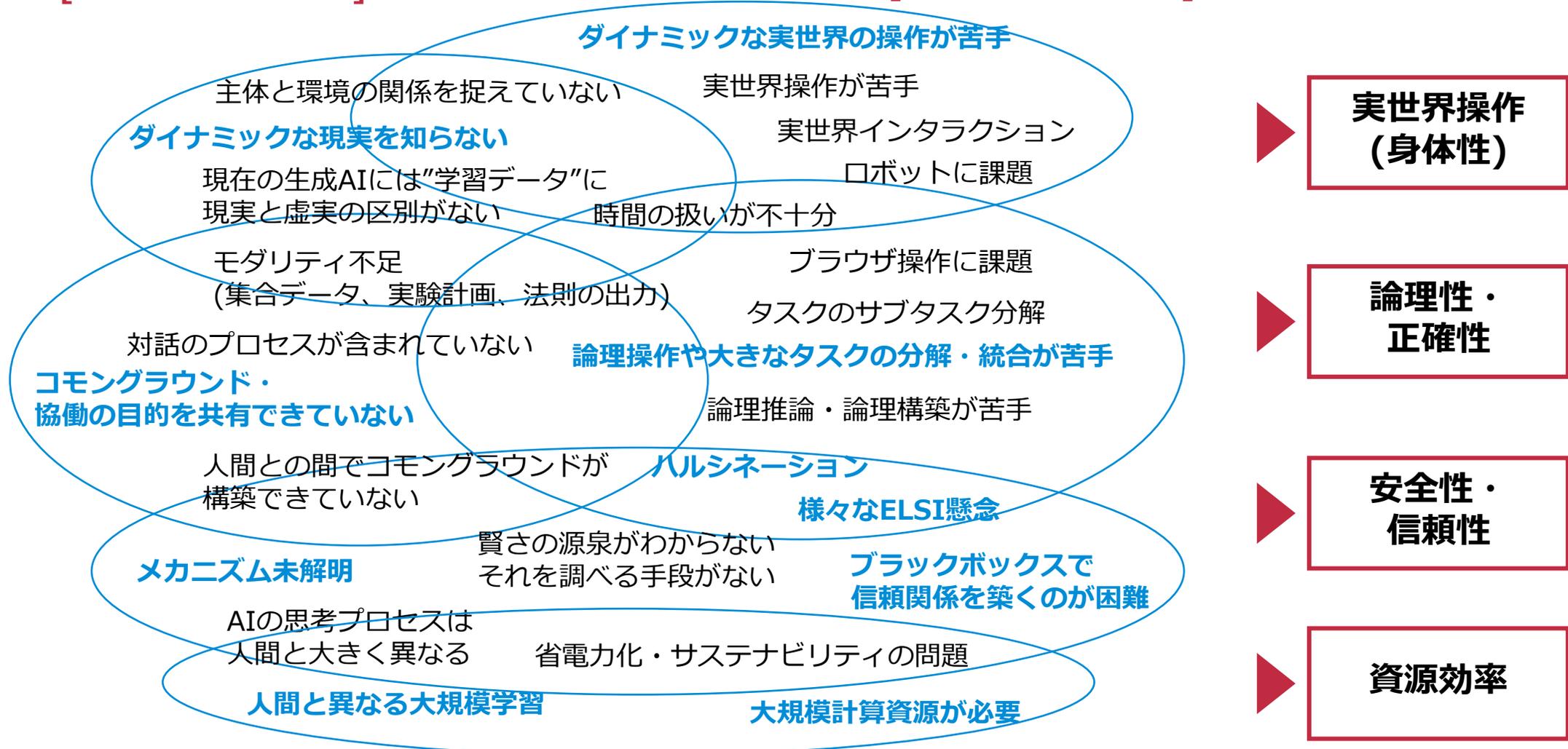
1. 講演の背景・立場
2. 戦略プロポーザル「次世代AIモデルの研究開発」の骨子
3. 基盤モデル・生成AIの限界
4. 限界克服に向けた次世代AIモデルの研究動向とAGI

# 現在の生成AI・基盤モデルの限界

ワークショップでの議論から

## [原因視点の問題認識]

## [結果視点の問題認識]



# 現在の生成AI・基盤モデルの限界

- 現在のAIモデルは、専門的な知識・能力を備えたような応答を返すが、基本原理は確率モデルによる予測 ▶ **以下のような点で問題**

## 資源効率

### 極めて大規模な計算資源/データ/電力が必要

- 最先端の基盤モデルは、1回の学習実行に百億円超の計算費用がかかる [AI Index Report 2024]
- 人間の脳と消費電力は20ワット程度

## 論理性・正確性

### 厳密な論理操作が苦手、ハルシネーション

- 確率モデルによる予測であって、厳密に論理立てて推論や問題分解をしているのではない

9/12リリースされたOpenAI o1では論理推論の仕組みが導入され、苦手でもなくなりつつあるが

# まとめ

## 実世界操作(身体性)

### 実世界状況に適応した操作・行動が苦手

- 仮想世界での学習であり、実世界状況の動的変化や個別性には十分適応できていない

## 信頼性・安全性

### 人間にとってブラックボックスに思え、動作・精度の100%保証はされない

- 振る舞いが確率的で不安定性に見え、期待通り・想定通りの動作や結果が得られるとは限らないため、信じ切ることは難しい、安全性に不安が残る

## 自発性

目的や価値基準は外部から与えられ、自ら生み出すことができない【この扱いは要議論】

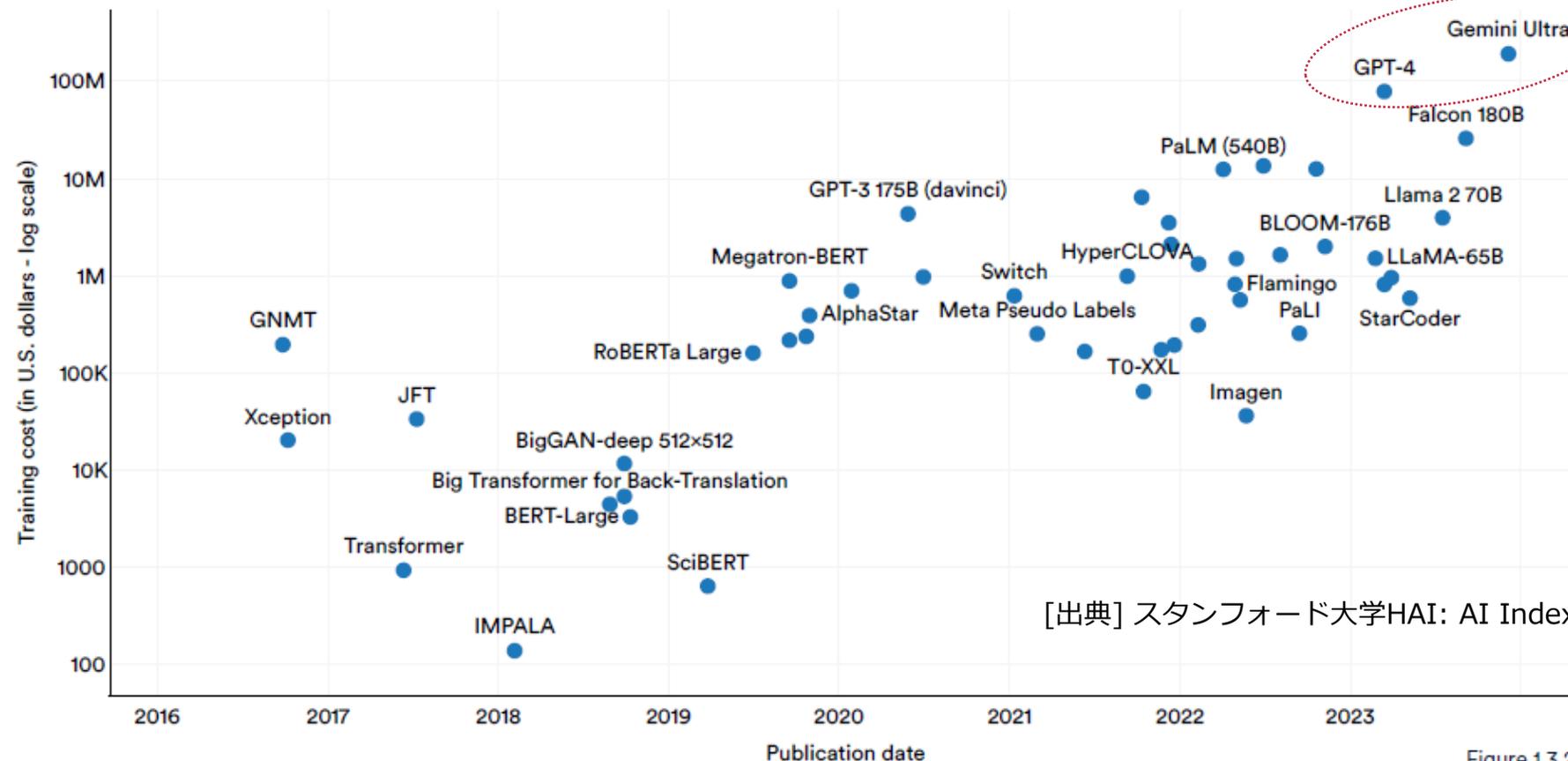
# 資源効率の問題：基盤モデルの訓練コスト

- 最先端の基盤モデルは、1回の学習実行に**百億円超の計算費用**がかかる

OpenAI GPT-4: 推定7800万ドル相当  
Google Gemini Ultra: 1億9100万ドル

Estimated training cost of select AI models, 2016–23

Source: Epoch, 2023 | Chart: 2024 AI Index report

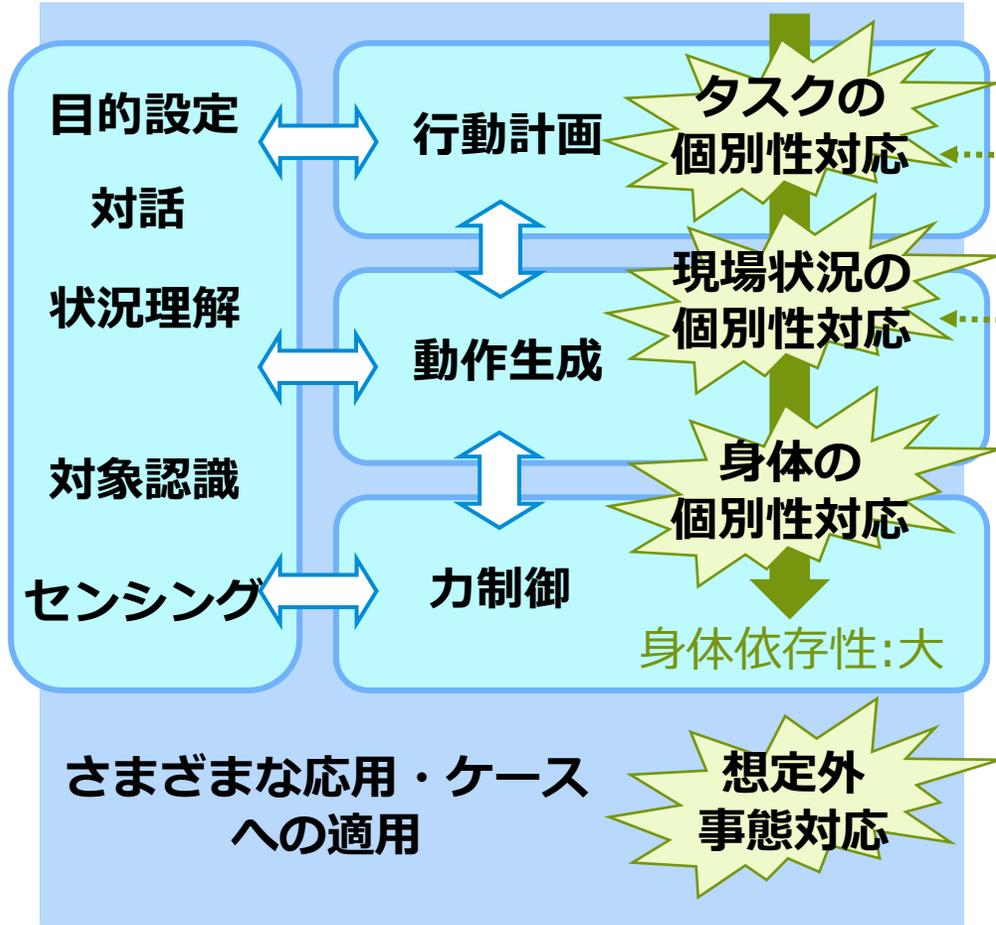


[出典] スタンフォード大学HAI: AI Index Report 2024

Figure 1.3.22

# 実世界操作(身体性)の問題

■ 基盤モデルによる行動計画は、実世界状況の動的変化や個別性には十分適応できていない



現状

基盤モデル

タスクごとに、ある程度汎化された手順の生成

[例] 自動販売機まで移動 → 缶飲料を選択  
→ 硬貨を投入 → 缶飲料を取り出す

強化学習・  
模倣学習

個別の状況ごとと個別の身体ごとに個別学習

[例] 缶飲料の選択動作、硬貨の投入動作、  
缶飲料の取り出し動作など

学習データ量を増大化すれば、  
無限の個別性や想定外事態に対応し得るのだろうか

# 安全性・信頼性の問題：生成AIによって深刻化するAIリスク

- 権利侵害・情報漏洩、犯罪・攻撃への悪用、社会・文化への影響まで **広範にAIリスクが拡大**
- 生成AI出力か人間応答かを見分けるのが困難な状況になり、 **人間・社会に内在するリスクを増幅**

## 生成AIの出力から生じる問題

- ウソや架空の出来事をあたかも事実であるかのように語る(ハルシネーション) 
- 差別・偏見、偏った価値観が応答中に表れる(社会的バイアス) 
- 学習データやプロンプトから **個人情報・機密情報が漏洩** 
- 学習データや生成データの **著作権・肖像権の問題** 
- クリエーターや俳優・声優などの **反発、創作市場・文化への影響** 

- AIによる **労働者の置き換え・失業** 
- 学習過程における **低賃金労働者搾取** 
- 大量の電力・水の消費による **環境インパクト** 

## 生成AIの信頼問題

- 個人情報・機密情報は **学習に使わない** 
- **信頼できる良質な生成AI** 
- **正確性・安全性・倫理等を確保**するようにモデルを調整 
- **粗悪な生成AI** 
- **偽りの生成AI** 
- **邪悪な生成AI** 
- 犯罪向けの **ワームGPT**
- 特定主義・思想の **プロパガンダ**
- 利用者の **個人情報抜き取り**
- **汚染されたモデル、仕込まれたバックドア**
- 個人情報・機密情報の除外や **品質確保**をしていない
- **偽ってもバレないだろう**

## 社会の在り方・文化への影響

- 超知能などが **予期せぬ挙動・事態**を引き起こす懸念 
- **教育の在り方**への影響、AI依存による **思考力低下**の懸念 
- フェイク拡散による世論誘導・選挙干渉、対立激化による **民主主義の質的低下** 
- 証拠の信憑性の低下による **犯罪捜査・司法のゆらぎ** 

## 生成AIの悪用問題

- **フェイク動画やフェイクニュース**を生成、SNSで拡散して世論を誘導・干渉、ハラスメント・攻撃 
- なりすましや詐欺メール等を生成し、 **犯罪利用** 
- 人々を思い通りに **誘導・洗脳**、思考停止させ、依存させる 
- 武器や毒薬の作り方などの **悪知恵**を聞き出し(脱獄/Jailbreak)※   
 ※生成AIに対するPrompt Injection攻撃の一種 

# 基盤モデル・生成AIの限界とAGIの可能性

1. 講演の背景・立場
2. 戦略プロポーザル「次世代AIモデルの研究開発」の骨子
3. 基盤モデル・生成AIの限界
4. 限界克服に向けた次世代AIモデルの研究動向とAGI

# 動向① 現在の基盤モデルを出発点とした改良・発展

- 基盤モデルがなぜあれほど賢い振る舞いをするのか不明、そのメカニズムの数理的説明は、より高性能・高効率のAIモデルを生み出す基礎として重要 ▶LLM-jpプロジェクト
- 基盤モデルのさらなる大規模化・マルチモーダル化、複数モデルの統合・マージ、外部処理連携などによる精度・機能・汎用性の向上が進展、その一方で軽量高速化・カスタム化

## In-context Learning プロンプトの工夫

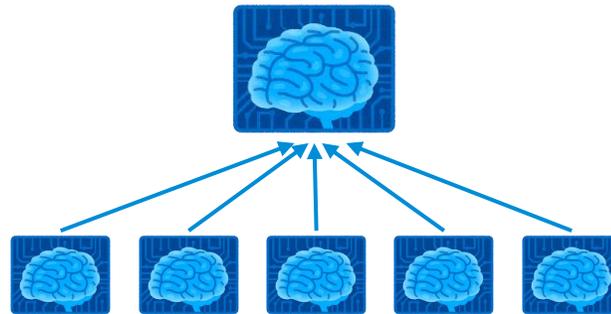
- なりきりプロンプト
- Step by Stepプロンプト (Chain-of-Thought)
- 深呼吸プロンプト 等

## マルチモーダルモデル

- LLM (Large Language Model)
- LMM (Large Multimodal Model)
- VLA (Vision Language Action Model) 等

## モデルの統合・マージ

- Mixture of Experts
- Model Merge



異なる得意・特性を持った  
エキスパートモデル群

## 外部連携・カスタム化

- 検索拡張生成RAG (Retrieval-Augmented Generation)
- 生成AIエージェント (タスク実行に必要な手順を計画して自動実行: AgentGPT、AutoGPT、Devin/Devika等)
- カスタムGPT (ChatGPTをもとにMyGPTs)

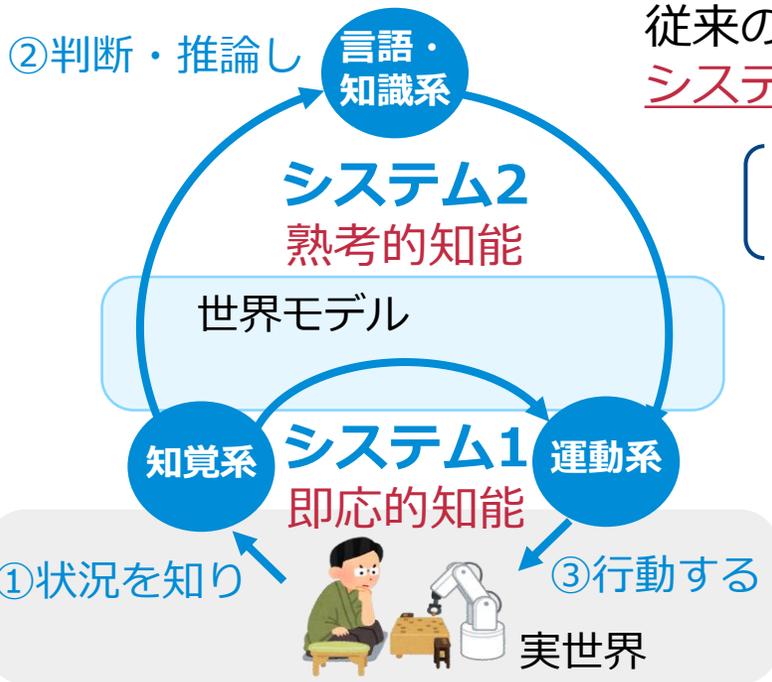
## 論理推論

- OpenAI o1 [動向①+②]

# 動向② 人間の知能からヒントを得た新原理研究

- 人間知能は未解明だが、様々な知見が得られつつあり、現AIの限界克服につながり得る
- 現状の帰納型(ボトムアップ)AIは膨大な事前学習が必要だが、能動的(トップダウン)に選択獲得・推論する仕組みによって、資源効率や論理性・正確性を改善し得る
- 予測誤差からモデルを修正する仕組みは、大量の教師あり事前学習を必要としない

## 二重過程理論

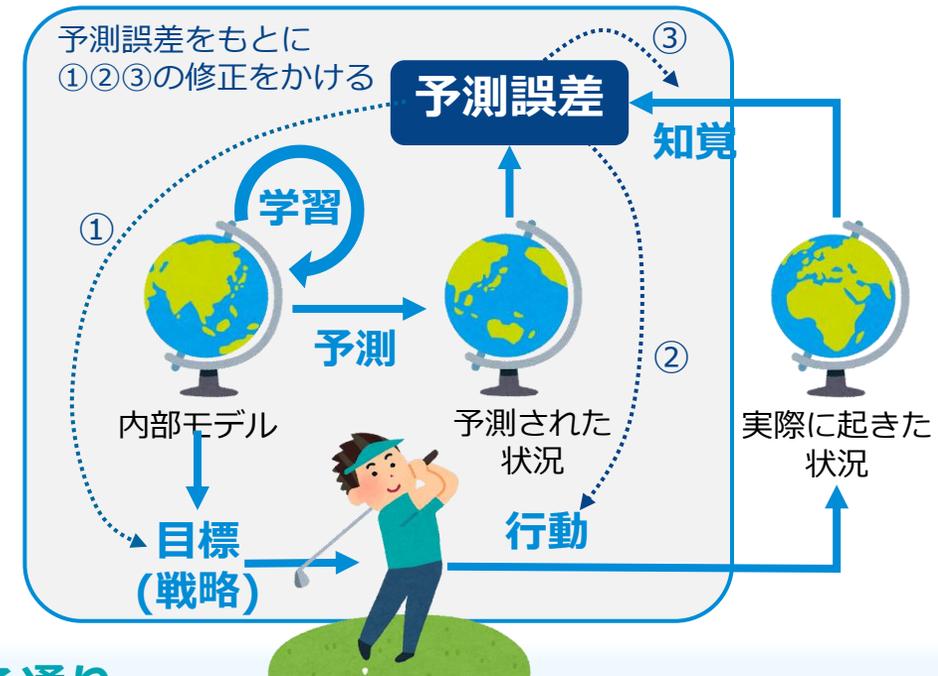


従来の深層学習はシステム1が中心で、システム2への拡張が必要

[ OpenAI o1はシステム2を  
実現したように見える ]

乳幼児からの認知発達の原理を、他者や環境とのインタラクションにおける予測誤差の最小化によって統一的に説明

## 予測符号化理論 [動向②+③] (予測誤差最小化/自由エネルギー原理)



▶WBAの取り組みは既に本日前半で説明されている通り

# 動向③ 他者や環境との関係性の中で発展する知能の研究

- AI単体での高度化だけでなく、外部との関係性の中で発展すると捉えた研究
- 現状、必ずしもまとまった大きな流れにはなっていないが、多数のAIと人間が協働・共存するマルチエージェント社会において考えておくべきAIモデルの観点がいくつか指摘されつつある

## コモングラウンド



コミュニケーションを取る上で欠かせない、相手との共通理解や会話のバックグラウンド



現在のChatGPTと利用者との間には コモングラウンド がないのでは

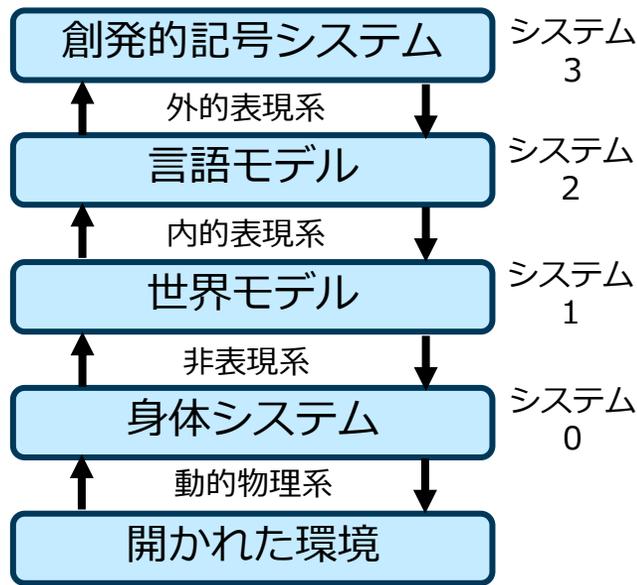


共同作業による目的達成が可能に

西田豊明「AI時代のコモングラウンド」  
[https://www.tasc.or.jp/educate/monthly/article\\_2/pdf/article\\_2\\_1911.pdf](https://www.tasc.or.jp/educate/monthly/article_2/pdf/article_2_1911.pdf)

## 集合的予測符号化

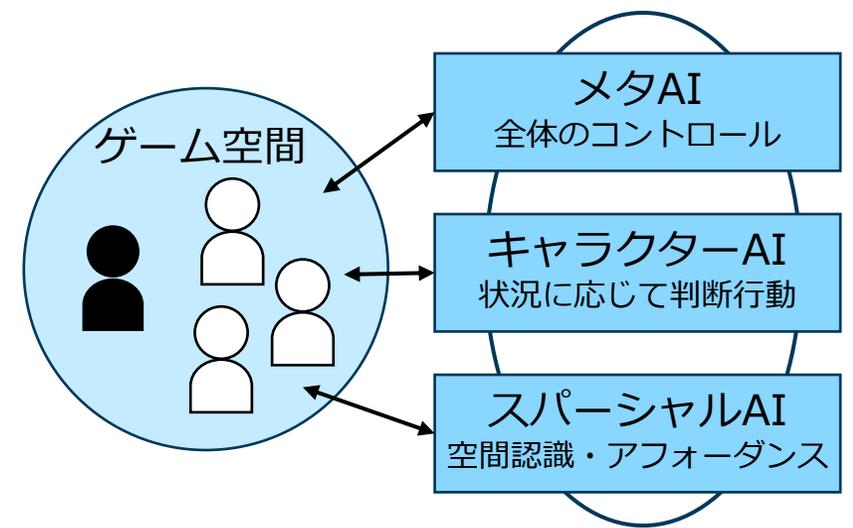
集団の中で 記号(言語系) が 創発 されるモデル(仮説)



前述のワークショップでの谷口忠大氏の講演をもとに作成

## 複数種AI連携モデル

個々のキャラクターの知能化だけでなく、オブジェクトや環境にも知能を持たせるアプローチ



三宅陽一郎氏の「MCS-AI動的連携モデル」の資料をもとに作成

# 動向①+③ 実世界インタラクションへの拡張

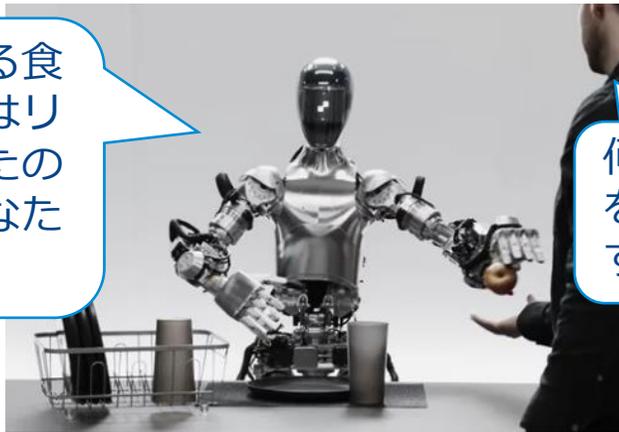
- 実世界操作のためには、身体性に基づく環境とのインタラクションが不可欠であり、AIとロボットの融合を通して研究が活発化

## ロボット基盤モデル

ロボットに基盤モデルを組み合わせることで、自然言語による指示・対話や実世界操作を可能にする技術開発が活発化している

### ChatGPT搭載人型ロボットFigure 01

テーブルにある食べられるものはリンゴだけだったので、それをあなたに渡しました



何か食べ物をもらえますか？

対話 認識 記憶  
計画 推論 説明  
動作制御

<https://www.youtube.com/watch?v=Sq1QZB5baNw>

(OpenAI+Figure, 2024/3)

## 世界モデル

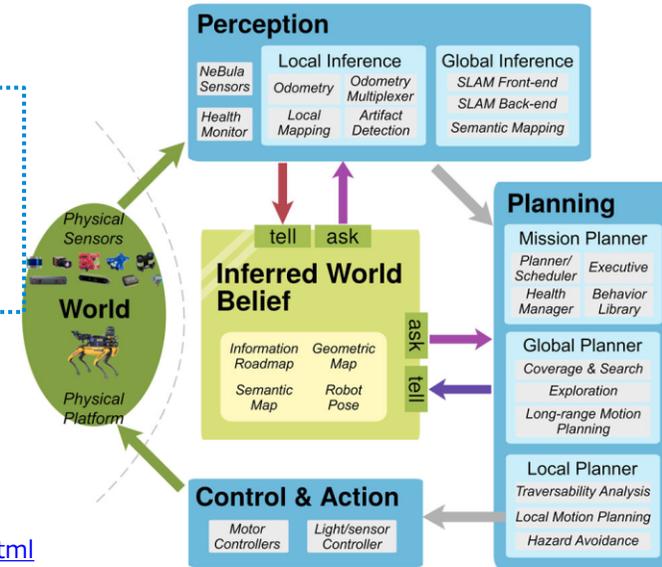
外界(世界)から得られる観測情報に基づき外界の構造を学習によって獲得するモデル

<https://weblab.t.u-tokyo.ac.jp/20221130-1/>

現場の状況に応じた動作生成のために不可欠

- 動作に対する物品の動きを予測
- 重なりあって隠れた物体の形状を予測

<https://jpn.nec.com/rd/technologies/202316/index.html>



[出典] Robotic World Modelsレビュー論文(2023/11)  
<https://doi.org/10.3389/frobt.2023.1253049>

## ロボット基盤モデルの構築 RT-1▶RT-2▶RT-X

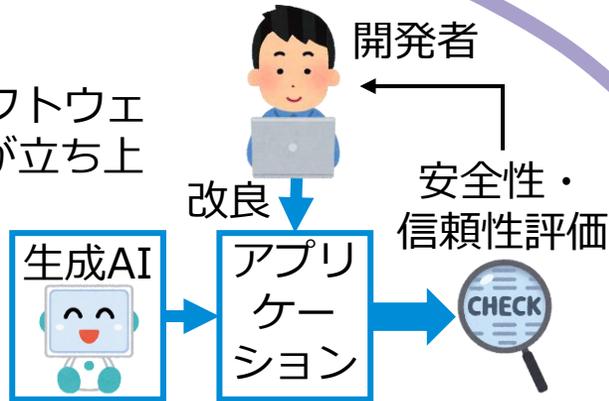
より多様でより大量のマルチモーダルデータに拡大して学習することで、基盤モデルを用いたロボットの行動計画・動作生成を、より多様なケースに適用可能に (Google)

# 生成AIリスク対策

6つのケースに整理

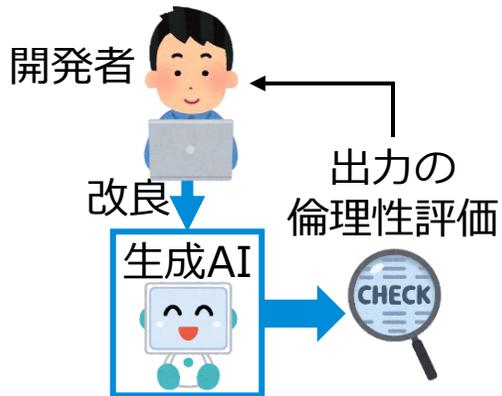
## ② 生成AIアプリ開発での品質管理

機械学習工学(AIソフトウェア工学)の研究分野が立ち上がり精力的取り組み



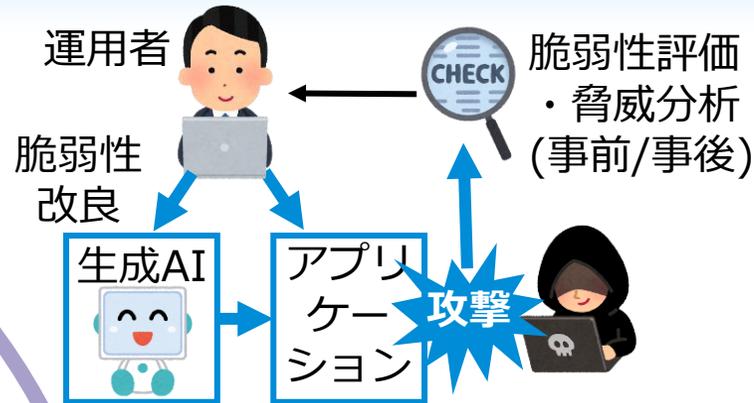
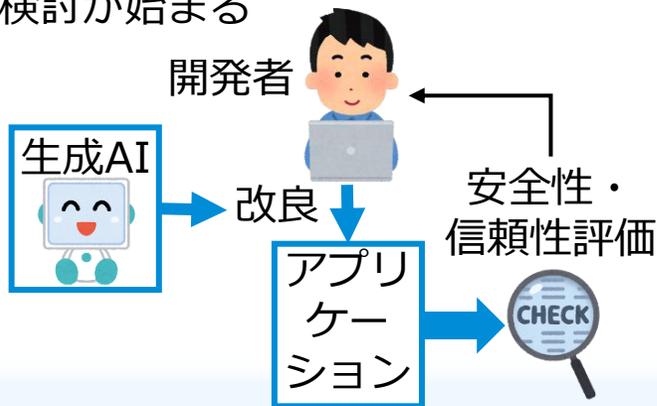
## ① 生成AI出力の倫理性確保

RLHF・DPO等、生成AIベンダーで取り組み活発



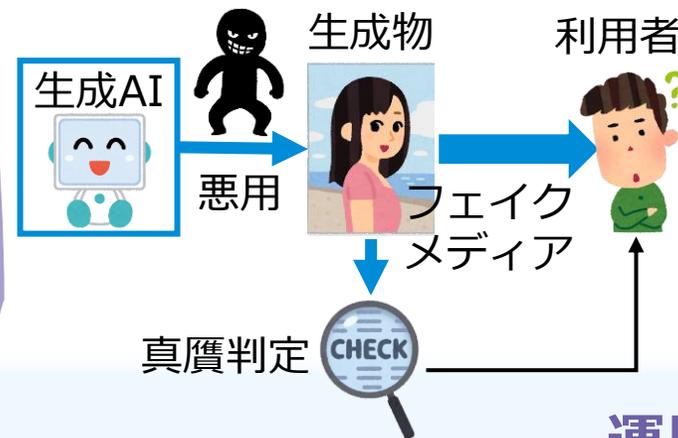
## ③ 生成AIを用いて開発したアプリの品質管理

新たな取り組み課題として検討が始まる



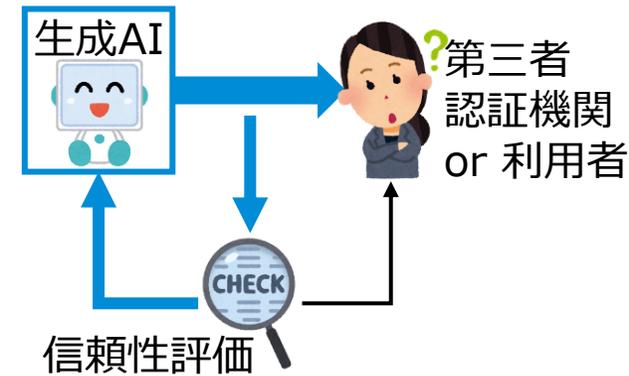
## ④ AI特有の脆弱性攻撃への対処

誤作動させる攻撃やモデル/データを搾取する攻撃など、対策が検討されている



## ⑥ 邪悪・粗悪なAIを見分ける手段

第三者認証制度・機関設置の動き(技術的裏付けは課題)



## ⑤ フェイクに騙されないように

フェイク検出技術や出所証明・追跡技術の開発が活発化

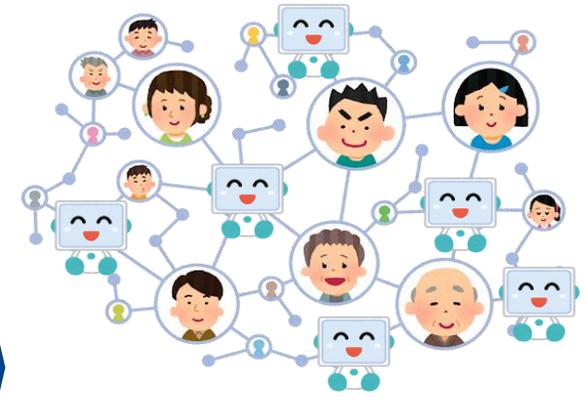
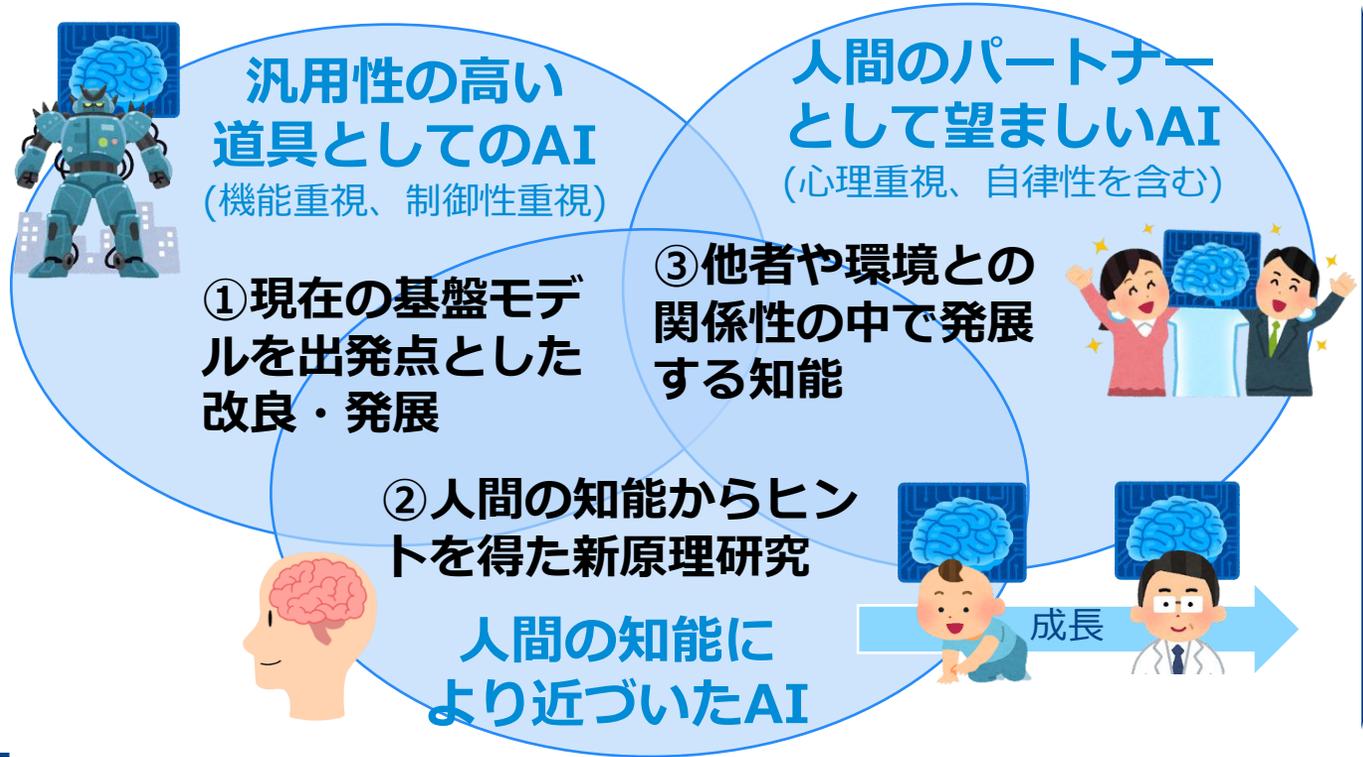
開発時

運用・利用時

# 次世代AIモデルの方向性・目指す姿

- AIが極めて高い能力・性能を持つようになったことで、その能力・性能をさらに高める研究開発だけでなく、どのような社会、人とAIのどのような関係を目指すべきか、「人・AI共生社会」の在り方を考えていくことが必要に
- リスク対策も各ケースの対策積み上げだけでなく、共生社会の在り方から制約を考える取り組みも

- 資源効率
- 実世界操作 (身体性)
- 論理性・正確性
- 信頼性・安全性



## 人・AI共生社会

- 多様な人々と多様なAIが共生・協働して
- 様々な社会課題を解決
  - 社会システムを全体最適化
  - 社会の持続性・安全性を確保

これらの限界克服

これらの取り組み・方向性は徐々に融合

# AGIの可能性についての私見

- 生成AIは道具として高い汎用性を持ったAIといえるかもしれない
- AGIには自律性が想定され、特に自発性(主目的・価値観の生成)の必要性をどう考えるか？
- 下表の他課題は現在の取り組みの発展・融合によって、ある程度改善されそうだが、自発性は？

現在のAIの課題	①現在の基盤モデルを出発点とした改良・発展	②人間の知能からヒントを得た新原理研究	③他者や環境との関係性の中で発展する知能
資源効率	×：大規模事前学習に依存する限り難しい	○	△：分散化によって改善されそう
実世界操作(身体性)	△：基盤モデルだけでなく②③との組み合わせか	○	○
論理性・正確性	△：基盤モデル単体では難しく外側に組み合わせ	○：人間並みには可能	?
信頼性・安全性	△：実用上の対策が積み上げられていく	信頼性は○、安全性は△	○
自発性	現在の取り組み①②③の延長ではたどり着かなそうにも思える： 自発性が生まれ得るのか？ 最高次の主目的・価値観(生きる動機?)のみ与えるのか？		

この表の内容はかなり主観的・感覚的なもので、異なる解釈・異論があると思うので、たたき台としての提示です

# 基盤モデル・生成AIの限界とAGIの可能性

1. 講演の背景・立場
2. 戦略プロポーザル「次世代AIモデルの研究開発」の骨子
3. 基盤モデル・生成AIの限界
4. 限界克服に向けた次世代AIモデルの研究動向とAGI

ご清聴  
ありがとうございました

