CRDS 研究開発戦略センター
Center for Research and Development Strategy

国立研究開発法人科学技術振興機構

資料1-2
科学技術・学術審議会
基礎研究振興部会（第15回）
令和6年6月11日

文部科学省 科学技術・学術審議会 基礎研究振興部会 第15回

フィジカルAIシステムに関する基礎研究課題

2024年6月11日

科学技術振興機構(JST)
研究開発戦略センター(CRDS)
茂木 強
t2motegi@jst.go.jp

前回の振り返りと今回の報告の位置付け

前回の振り返り

現在のAIが抱える課題として、資源効率、実世界操作(身体性)、論理性、安全性、信頼性などが指摘されている

これらの課題克服によってもたらされる社会的価値として特に、AIが物理的な身体機能を獲得することで、デスクワークから現場作業まで幅広く労働市場に波及し、大きな社会・経済的インパクトをもたらし得ることが注目され、取り組みが急速に活発化している

これを実現するための研究開発課題として、
①次世代AIモデル
②AIと身体機能システムの融合
③人に安全なフィジカルAIシステム
が挙げられる

今回の報告で検討したこと

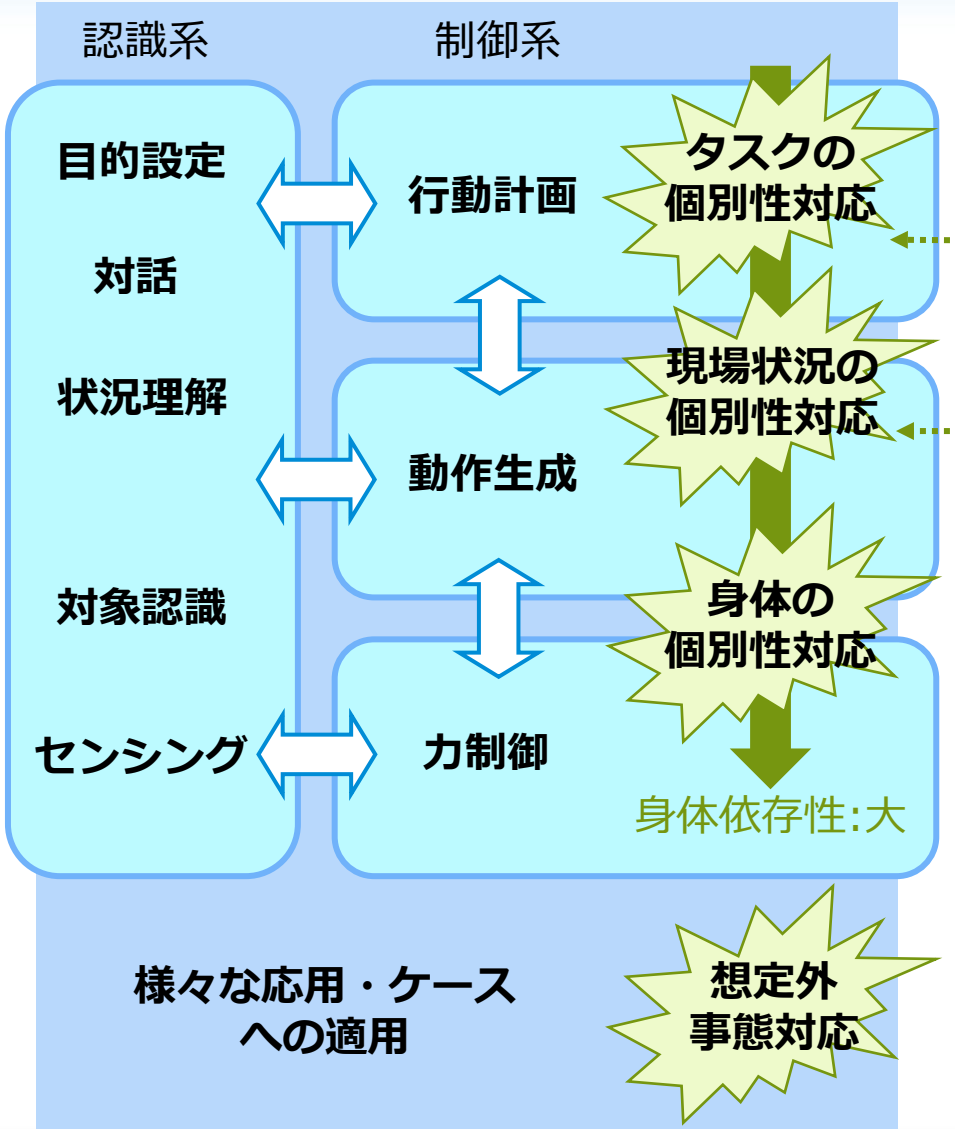
AIが物理的な身体機能を獲得して社会的価値を生み出すためには２つの発展の方向性が重要である

- より多様で複雑な実世界タスクの実行
- より多様で複雑な実世界環境への対応

発展の方向性から、達成すべきフィジカルAIシステムを3つのタイプに分け、取り組むべき研究開発課題を詳細化

- タイプP: 多様な実世界タスクをこなす
- タイプH: 実世界で人間と協働・共進化する
- タイプA: 多様な実世界環境で稼働する

AIが物理的な身体機能を獲得するための研究開発課題



第14回資料「基盤モデル後のAI研究開発動向」より再掲 (一部表現修正)

①次世代AIモデル

現在の基盤モデルの課題である資源効率・実世界操作(身体性)・論理性などを克服し、高い精度・効率と対話能力を持ち、タスクや現場状況や身体の個別性に柔軟に適応できるAIモデルを実現。また、学習から認知発達や進化への発展も探求。

②AIと身体機能システムの融合

①が非身体(AI)側からのアプローチであるのに対して、②は身体機能の側からのアプローチである。タスク/現場状況/身体の個別性まで把握し、ロボット側の個別対応(作り込み)を極力減らし、状況変化に対しても動的・適応的に動作できるようにする。さらに、所与の身体の制御に限らず、多様な身体への適応的制御や身体形状の適応まで可能性を探求。

③人に安全なフィジカルAIシステム

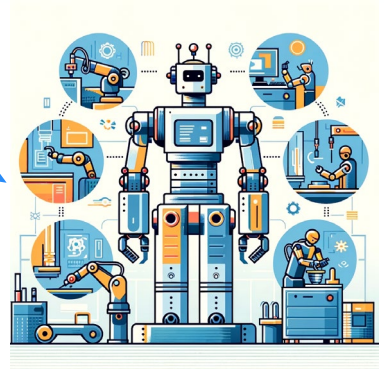
機械学習ベースのAIモデルは、原理的に100%の精度保証・動作保証はできない。現在は事前統制されたClosed環境を用意することで安全性を確保しているが、Open環境では想定外の状況が起こる。応用に合わせてClosedからOpenへ段階的に環境を拡張しつつ、想定外リスクを可能な限り回避・対策し、フィジカルAIシステムと人が安全に協働・共生できるようにする。

CRDS

フィジカルAIシステムの発展の方向性・タイプ

P: Performance
H: Humanoid
A: Adaptive

物理的な身体機能を獲得したAIシステム



タイプP.多様な実世界タスクをこなすフィジカルAIシステム

ある程度限定された環境のもとで、1台のシステムが様々なタスクを実行できるようにする。タスクに応じた様々な動作、繊細な動作、高速な動作を追求する。人間にはできないようなタスク・動作まで実現。

[想定する適用例] 産業用ロボットの知能化（セル生産、職人技の学習など）

タイプH.実世界で人間と協働・共進化するフィジカルAIシステム

人間のような知能・身体性・対話能力等を備え、人間の代わりに／人間と協働で様々なタスクを担えるようにする。人間自体を理解するためのサイエンス研究面も含み、能動的に学習し、発達・共進化するフィジカルAIシステムへと発展させる。

[想定する適用例] 発達・共進化（日本発の認知発達ロボティクス研究）

タイプA.多様な実世界環境で稼働するフィジカルAIシステム

タスクの種類は、ある程度限定されているが、環境や条件に多様性・複雑性・変動性・不完全性があっても、ロバストにタスクを遂行できるようにする。人間が立ち入れない環境でさえ稼働可能。

[想定する適用例] 屋外ロボットの知能化（農業、インフラ保守点検など）

より多様で複雑な実世界タスクの実行
（動作の種類拡大・繊細化・高速化）

タイプP

タイプH

タイプA

AI

人間に可能な程度の
実世界環境・タスク

より多様で複雑な実世界環境への対応
（場の広がり）

※上記画像は生成AIを用いて作成した

フィジカルAIシステムの発展研究がもたらす社会的価値

AIが物理的な身体機能を獲得すると、デスクワークから現場作業まで幅広く労働市場に波及、
大きな社会・経済的インパクトをもたらすことが期待できる

タイプ	現状の実用レベル	現状の研究レベル	最終的な目標レベル
P Performance	**タスク毎の作り込み** • 強く統制された現場環境 • 個々のタスクのプログラミング **実用例** • 自動車製造ラインの溶接ロボット • 半導体製造でのウェハ搬送ロボット	**タスク毎の事前学習** • 個別タスクの模倣学習と強化学習 • 熟練者技能の学習 **研究事例** • 米Covariant AI社 RFM-1 • 米Google社 RT-1/RT-2	**多様な実世界タスクをこなすシステム** • 1台でさまざまなタスクを実行 • 繊細な操作、高速な操作を追求 **事例：汎用型組立ロボット** • 車種やモデルごとに異なる部品の組み立てを1台のロボットで対応
H Humanoid	**単機能サービスロボット** • 限定環境での安全な定型作業 • 限定的話題や定型的対話(案内受付等) **実用例** • 飲食店における配膳ロボット • 病院における搬送ロボット	**対話的サービスロボット** • タスクに応じた行動計画推論 • 自由対話による指示理解・説明 **研究事例** • 米Figure AI社 Figure 01 • 米Google社 PaLM-SayCan	**実世界で人間と協働・共進化するシステム** • 人間同様の知能・身体・対話能力 • 能動的に学習し発達する **事例：サービスアシスタントロボット** • 家事やサービス業などで人間の作業を能動的に支援
A Adaptive	**専用機による作業** • 人間による機器の操作 • 限定環境での自動運転 **実用例** • 建設現場における重機ロボット • 農業における収穫ロボット	**限定環境で働くフィールドロボット** • 不整地対応ロコモーション • 移動とマニピュレーションの統合 **研究事例** • 米Agility Robotics社 Digit • スイスETH Zurich大 ANYmal C	**多様な実世界環境で稼働するシステム** • 多様・複雑・不完全性への頑強性 • 環境変化や想定外への適応 **事例：フィールドロボット** • 農業やインフラ点検作業など高負荷／危険な作業を人間に代わってこなす

フィジカルAIシステムの研究開発課題の全体観

フィジカルAIシステムが高い汎用性を実現するためには多くの課題がある。ここでは、①AIモデル、②身体機能との融合、③安全性の側面から、これらの課題と今後の研究方向性を示す。

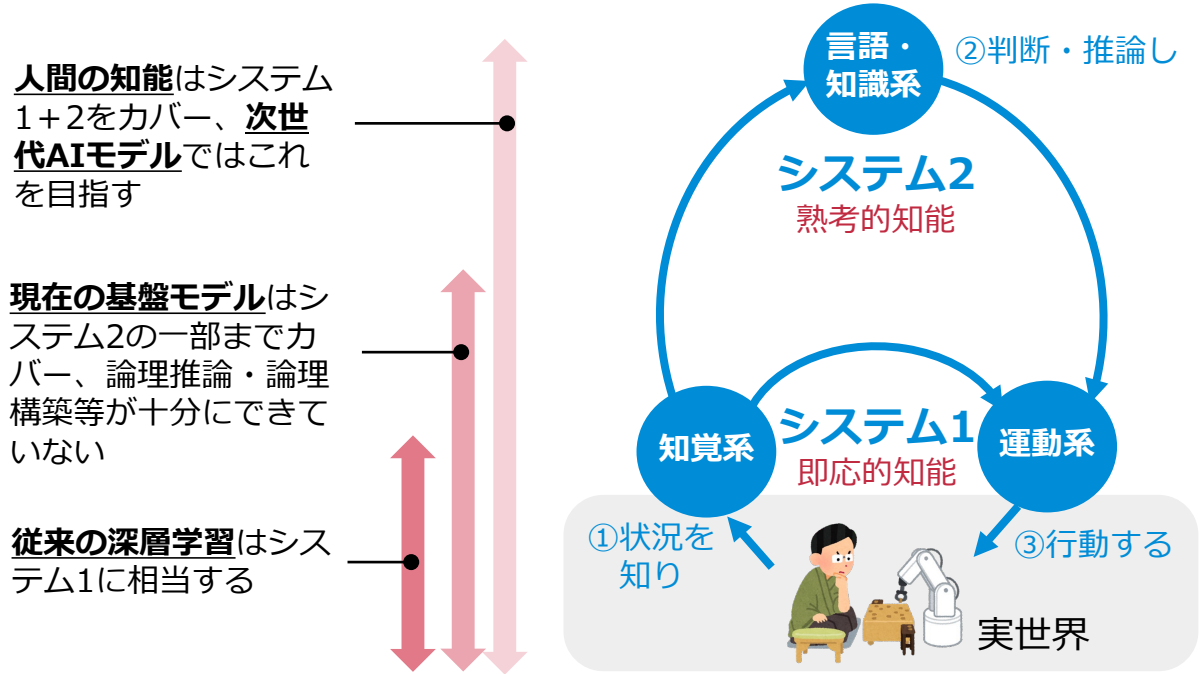
推進すべき領域 ▼

	現状の実用レベル	個々の現場ニーズに高度に応える技術開発研究レベル	高い汎用性をもたらす基礎研究
①次世代AIモデル	**限定定環境での認知・行動計画** • カメラによる環境認識 • 環境認識に基づく行動計画	**限定環境で働くフィールドロボット** • 多様なタスクの学習と自律実行 • 動的な環境認識による自律移動	**現状の課題を克服したAIモデル** • 高い精度・効率と対話能力 • タスクの現場状況や身体の個別性に柔軟に適応 • 能動的推論によりエッジで自律学習
②AIと身体機能システムの融合	**現場データからの学習** • 限定された環境での定型作業 • 個々のタスクのプログラミング	**タスク毎の事前学習** • 個別タスクの模倣学習と強化学習 • 熟練者技能の学習	**身体機能システムと融合するAIモデル** • タスク／現場状況／身体性の個別対応を極小化しオープン環境に適応 • リアルタイム動作
②'身体機能システム	**工場の生産ライン用のロボット** • 固定式の単腕または双腕ロボット • 車輪または二足で平面移動可能 • クライアントサーバ型の開発環境	**自律移動モバイルロボット** • 不整地移動対応ロコモーション • 精密かつ器用なマニピュレーション • 柔軟な身体の数理基盤と制御 • 高速低遅延通信	**自律的に動作可能な身体機能システム** • 環境認識に必要なセンシング機構 • エッジ対応推論チップ • 必要な時間動作可能な電源機構 • 他の個体やシステムとの通信機構
③人に安全なフィジカルAIシステム	**閉じた環境での特定タスクの実行** • 強く統制された現場環境 • 共存環境では安全な動作のみ実行	**限定環境での協働ロボット** • 自然言語による包括的な指示理解 • 人間の意図を把握した安全動作 • 想定内のリスクへに対応	**想定外に対応可能なフィジカルAIシステム** • 正確かつ詳細な環境認識 • 安全な行動計画可能な状況判断 • 最適な行動を即時で意思決定

重要な研究開発課題① 次世代AIモデル

現在の基盤モデルの課題である資源効率・実世界操作(身体性)・論理性などを克服し、高い精度・効率と対話能力を持ち、タスクや現場状況や身体の個別性に柔軟に適応できるAIモデルを実現。また、学習から認知発達や進化への発展も探求。

二重過程理論：経験に基づいた即応的な情報処理を担うシステム1と、抽象化されたモデル・知識を参照した熟考的な情報処理を担うシステム2から成るという知能のモデル。システム2で能動的・選択的に情報を取りに行けば学習データ量が減り、資源効率や論理推論能力を改善され得る。



予測符号化理論（能動的推論）：乳幼児からの成長のように、他者や環境との相互作用・実世界操作を通じて、自己・環境の認知、言語獲得、行動・推論等の認知機能を発達させていく過程を、予測誤差最小化原理(自由エネルギー原理)によって統一的に説明。大量の教師あり事前学習は不要になり、資源効率も改善され得る。

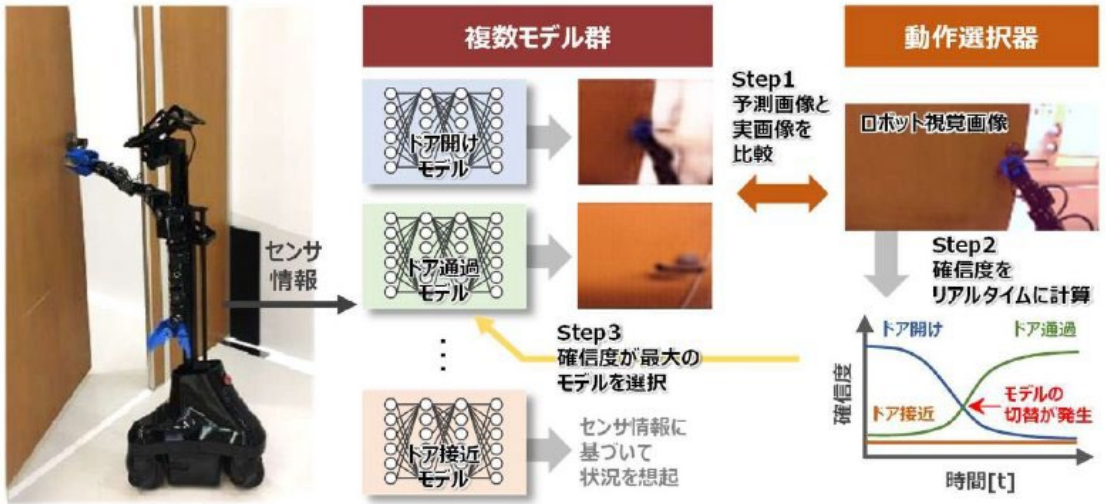


図 3 複数予測モデルのリアルタイム切替技術

1 つの予測モデルでは対応しきれない複雑な作業に対応可能

出典：深層予測学習型ロボット制御技術開発 https://www.waseda.jp/top/news/79811

重要な研究開発課題① 次世代AIモデル

実世界の物理モデルが組み込まれたモデルや、複数モデルの統合により、実世界における複雑なタスクが実行可能なモデルの実現することで、現在の基盤モデルの課題である実世界操作性と汎用性が向上する。

LLM▶VLAモデル▶身体化VLAモデル

マルチモーダルLLM

テキスト、画像、音声など複数のデータ形式を処理できる大規模言語モデル。OpenAIのGPT-4などが代表例。

VLAモデル（Vision-Language-Actionモデル）

視覚、言語、行動のデータを統合して、ロボットが複雑なタスクを実行するためのモデル。ロボットがカメラやセンサーを通じて環境を認識し、自然言語の指示を理解し、それに基づいて物理的な行動を行うことが可能。Google DeepMindのRT-2などが代表例。

Embodied VLAモデル（身体化VLAモデル）

通常のVLAモデルの拡張であり、ロボットの物理的な身体をより重視。ロボットの身体の形状や特性を考慮し、その体に最適化された行動を生成することを目指す。研究段階であり、実験環境の多くはサイバー空間に限られる。物理法則の厳格な模擬がポイント。

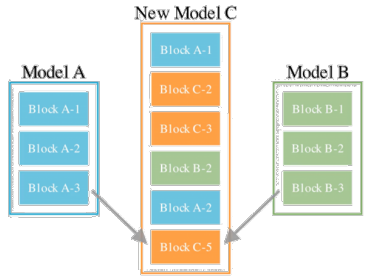
研究例：3D-VLA 3D-VLA (umass.edu)
[2403.09631] 3D-VLA: 3D視覚・言語・行動生成世界モデル (arxiv.org)

LEO（Embodied Generalist Agent）「
[2311.12871] An Embodied Generalist Agent in 3D World (arxiv.org)

複数のモデルの統合

サカナAI進化的モデルマージ

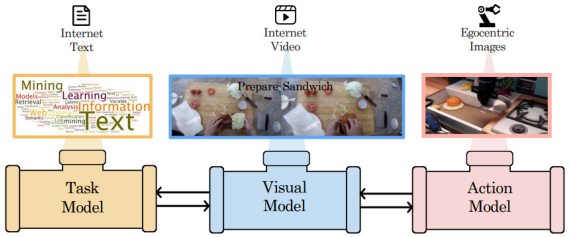
既存の複数のAIモデルを組み合わせて新しい高性能な基盤モデルを作成する技術



https://sakana.ai/evolutionary-model-merge-jp/

ロボットのための階層モデル

複数のAIモデルにより、複雑な計画を人間にわかりやすく実行



Compositional Foundation Models for Hierarchical Planning 2309.08587 (arxiv.org)

アプローチ	進化的モデルマージ	HiP（階層的計画）
概要	進化的アルゴリズムを用いて既存モデルを組み合わせて新モデルを生成	言語、視覚、行動を統合して階層的に計画・実行
主な目的	高性能な基盤モデルの自動生成	ロボットが複雑なタスクを遂行するための計画と実行
自動化	モデルの組み合わせを自動で最適化	反復的なフィードバックで計画と実行を調整
コスト効率	訓練不要で低コスト	高度な計画と視覚的検証が必要
応用分野	言語モデル、画像生成モデルなど広範囲	主にロボティクスに特化

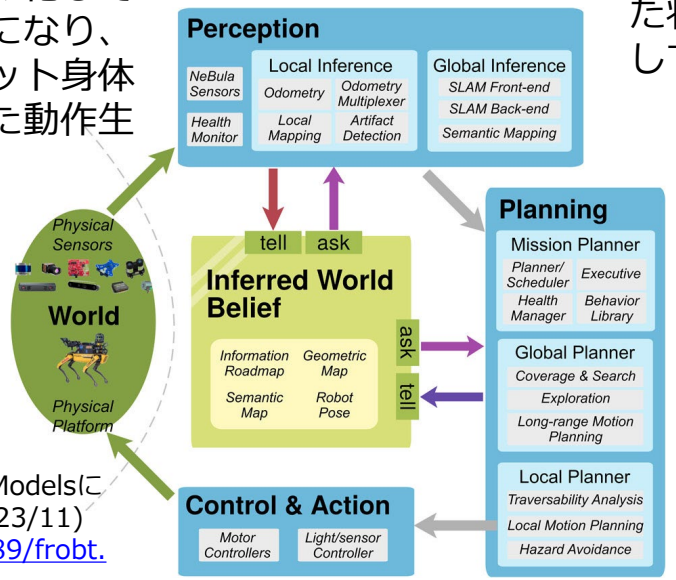
重要な研究開発課題② AIと身体機能システムの融合

①が非身体(AI)側からのアプローチであるのに対して、②は身体機能の側からのアプローチである。タスク/現場状況/身体の個別性まで把握し、ロボット側の個別対応(作り込み)を極力減らし、状況変化に対しても動的・適応的に動作できるようにする。さらに、所与の身体の制御に限らず、多様な身体への適応的制御や身体形状の適応まで可能性を探求。

世界モデルの構築と活用

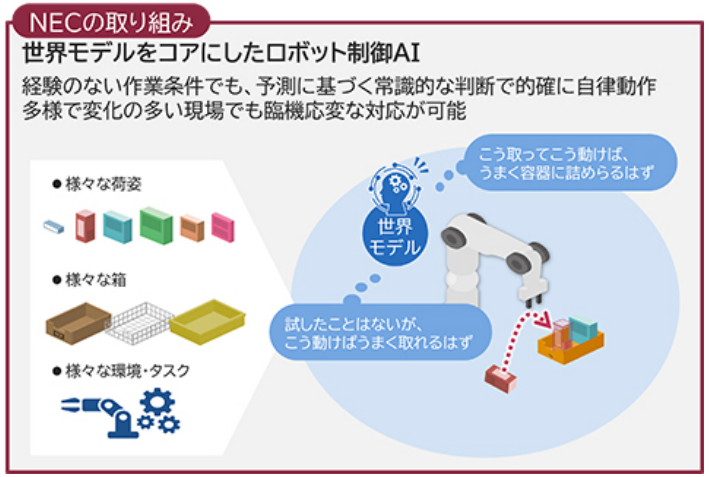
外界(世界)から得られる観測情報に基づき外界の構造を学習によって獲得するモデルは「世界モデル」(World Model)と呼ばれている
https://weblab.t.u-tokyo.ac.jp/20221130-1/

現場状況をモデル化して推論できるようになり、現場状況・ロボット身体の個別性に応じた動作生成が可能になる



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

世界モデルを用いたロボット制御の開発例:
世界モデルを用いた予測・シミュレーションによって、未学習の状況や部分的隠れが生じた状況に対しても推論して動作生成 (NEC)



[出典] NEC研究開発紹介Web https://jpn.nec.com/rd/technologies/202210/index.html (2023/3)
https://jpn.nec.com/rd/technologies/202316/index.html (2024/2)

大規模マルチモーダルデータ学習によるロボット基盤モデル構築

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)

RT-1 (2022/12)
ロボット実機13台×17か月
13万エピソードを学習
700以上のタスクに対応

▶

RT-2 (2023/7)
Web上のテキストと画像も学習、RT-1未学習の対象への対応可能に

▶

RT-X (2023/10)
世界33研究機関の参加による史上最大のオープンソースロボットデータセット構築

重要な研究開発課題② AIと身体機能システムの融合

AIと身体機能システムの融合にはエッジの知能化が不可欠であり、これによりフィジカルAIシステムが自律的に動作し、複雑なタスクを効率的に実行できるようになる。高性能なハードウェア、高度なソフトウェア、そして適切なアルゴリズムを統合的に組み合わせることで、高い社会受容性をもつフィジカルAIシステムの実現が可能になる。

エッジの知能化に求められる代表的な要件

ハードウェア

- **計算能力:** 専用のプロセッサ（GPU、AIアクセラレータ、推論チップなどによるデバイス上での高速処理
- **センサー:** LiDAR、深度カメラ、触覚センサーなど多様なセンサーの搭載
- **エネルギー効率:** バッテリー駆動による長時間稼働を実現する低消費電力設計

アルゴリズム

- **AIモデル**：ロボット基盤モデル、世界モデル
- **機械学習:** 強化学習、模倣学習、連合学習、転移学習など
- **ロボット制御:** SLAM、ナビゲーション、物体認識、行動計画、迅速な意思決定
- **リアルタイムでの処理:** センサーデータのリアルタイム処理

ソフトウェア

- **通信:** ローカルネットワークとクラウドの連携、リアルタイムなデータ送受信、マルチロボットシステムでの分散協調処理、システム内通信
- **ソフトウェアアーキテクチャ:** ROS 2などのロボット用ミドルウェアと連携
- **開発ツール:** ロボット向け開発フレームワーク、シミュレーションツール連携

その他

- **セキュリティ:** ロボットの動作とデータの保護、物理的セキュリティとサイバーセキュリティの両立

推論チップと軽量フレームワーク

■ NVIDIA が2024年3月の開発者向け会議でヒューマノイド ロボット向け Project GR00T 基盤モデルと Isaac Robotics プラットフォームおよび、エッジ用AIチップを紹介[1)]

① **Isaac sim:**ロボット向けのシミュレーションソフト。トレーニングシミュレータとして機能し、クラウドアクセスが可能になったことで、どこからでもシミュレーションにアクセスし、チームで協力してロボット開発を進めることができる。

② **NVIDIA Jetson Thor:** NVIDIA Orin™ SoC をベースにしたヒューマノイド ロボット用の新しいコンピューター。Jetson AGX Orin（**組込みシステム向け**チップ）の後継。

③ **Project GR00T 基盤モデル**：ロボットようのAI基盤モデル。少ないデタでロボットがトレーニングできる。

④ Disney Researchの**ヒューマノイドロボット GR-1:** Jetson Thor を搭載し、Isaac Sim プラットフォームで開発された。

Jetson Thor [1)]


GR-1 [1)]

1) NVIDIA GTC 2024 Keynote
https://www.youtube.com/live/Y2F8yisiS6E?si=20WncfYYHGBQiqPr

重要な研究開発課題②’ 身体機能システム

①、②が主にソフトウェア主体の課題であるが、本課題はハードウェア主体の課題である。フィジカルAIシステムがエッジで自律的に働くためには、サーバに頼らず実行可能なスタンドアローン動作、移動の柔軟性、把持の精密性や、省電力、リアルタイム性、柔軟な構造・動作など身体機能そのものの進化が必要。

模倣学習によるマニピュレーションの高度化

ハードウエアの開発と機械学習（特に模倣学習）を統合したアプローチが成果をあげつつある。

Shadow Dexterous Hand
出典：英国 Shadow Robot Company

1. 多指ハンドと触覚センシング:
- 多指ハンド ロボット
- ロボット 触覚センシング

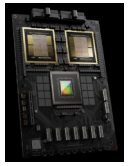
qb SoftHand Industry
出典：イタリアqbRobotics

2. ソフトロボティクスと材料科学:
- 柔軟な素材の活用:
- ソフトロボティクス ハンド

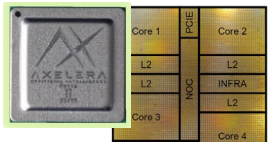
3. 模倣学習と強化学習の融合:
- 模倣学習: 熟練した職人の動作をモーションキャプチャリモート操作などで学習し、ロボットに模倣させることで、基本的な動作パターンを習得。
- 強化学習: 模倣学習で得られた基本動作を基に、ロボット自身が試行錯誤を繰り返しながら、より高度な技を習得。

エッジAIチップ

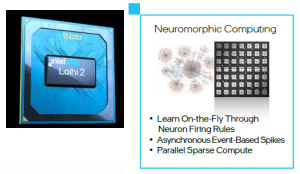
エッジ側での推論・学習用に、低消費電力化・高効率化を目指す積和演算の低ビット演算、メモリアクセスを抑制するためにメモリ内で演算するコンピュートインメモリ、スパイキングニューロンの模倣回路などの研究開発が活発化



NVIDIAの最新アクセラレータ（GB200 Grace Blackwell Superchip）（出典：https://nvidianews.nvidia.com/news/nvidia-blackwell-platform-arrives-to-power-a-new-era-of-computing ）



Axelera AI（オランダのスタートアップ）のAIエッジ用のチップ（Metis AIPU: 15TOPS/W ）（出典： https://www.axelera.ai/metis-aipu ISSCC2024 Session 11.3



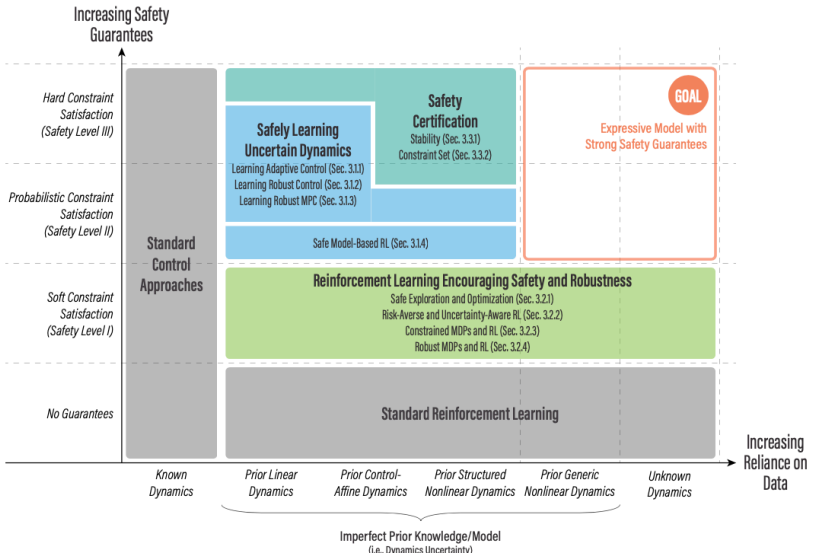
Loihi2ニューロモルフィックチップ（10^6 ニューロン、$1.2x10^8$シナプス）（出典：https://www.intel.com/content/www/us/en/research/neuromorphic-computing-loihi-2-technology-brief.html ）

重要な研究開発課題③ 人に安全なフィジカルAIシステム

機械学習ベースのAIモデルは、原理的に100%の精度保証・動作保証はできない。現在は事前統制されたClosed環境を用意することで安全性を確保しているが、Open環境では想定外の状況が起こる。応用に合わせてClosedからOpenへ段階的に環境を拡張しつつ、想定外リスクを可能な限り回避・対策し、フィジカルAIシステムと人が安全に協働・共生できるようにする。

安全な強化学習（Safe RL）

制御理論のモデルベースの予測能力と強化学習のデータ駆動の適応能力を組み合わせることで、動的な環境下でも安全に学習しながら制御を行う。



出典：Safe Learning in Robotics: From Learning-Based Control to Safe Reinforcement Learning (https://arxiv.org/abs/2108.06266)

安全から安心向けた人文学的研究

Perceived Safety（知覚的安全性）とは、ある環境や状況において、個人が主観的に感じる安全性の度合いをいう。

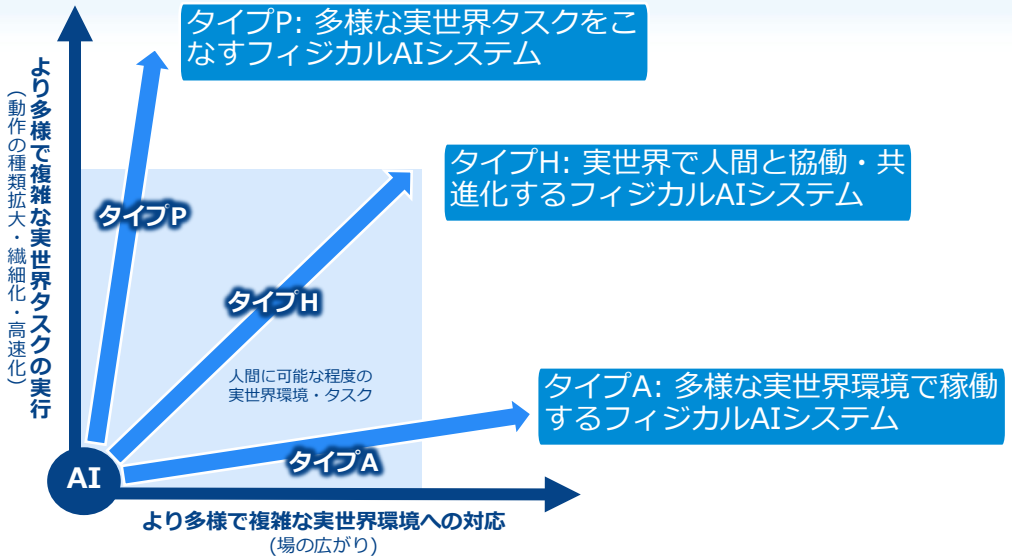
フィジカルAIシステム	心理的な安全性の向上のポイント
産業用マニピュレータ	ロボットの動きの予測可能性と一貫性が高いほど、ユーザーの心理的安全性が向上する。
移動ロボット	ロボットの速度と経路計画の安全性が重要。急激な動きや予期しない動きは安全性の知覚を低下させる。
移動マニピュレータ	ロボットの複雑な動きとタスク遂行能力が評価される。特に、人間との協働タスクにおいて予測可能な動きが重要。
ヒューマノイドロボット	人間に似た動作や表情が信頼感を増し、安全性の知覚を向上させるが、違和感がある場合は逆効果になることも。
ドローン	飛行高度や距離、速度が安全性の知覚に影響する。低速で安定した飛行が好まれる。
自動運転車	運転のスムーズさと予測可能性が重要。急停止や急加速は安全性の知覚を低下させる。

出典：Perceived safety in physical human–robot interaction—A survey
https://doi.org/10.1016/j.robot.2022.104047

まとめ

AIシステムの発展方向性・タイプ

- フィジカルAIシステムを「物理的な身体機能を獲得したAIシステム」として定義し、社会的価値を生み出すための二つの方向性を同定
  - より多様で複雑な実世界タスクの実行
  - より多様で複雑な実世界環境への対応
- フィジカルAIシステムの発展のタイプ



推進すべき研究領域

高い汎用性をもたらす基礎研究開発課題
①次世代AIモデル
②AIと身体機能システムの融合
③人に安全なフィジカルAIシステム