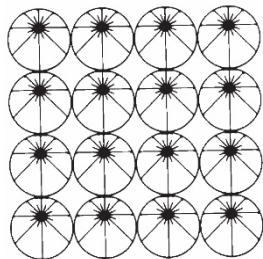


# 講演録



## フィジカル AI の現在地 — 現状の正確な理解 —

東北大学 情報科学研究科／理化学研究所 AIP センター 教授

岡 谷 貴 之

本講演録は、2026 年 7 月 17 日にオンラインで行われた講演の録画をもとに、編集事務局が原稿を作成して掲載するものです。

### 1 はじめに

東北大学の岡谷です。本講演では、フィジカル AI の現状について説明します。私は、フィジカル AI のど真ん中で研究しているわけではありませんが、全く無関係というわけでもありません。そうした立場から、フィジカル AI の現状を正確に理解するための情報、特に研究者の視点から見た情報を共有することが、本講演の趣旨です。

自己紹介を簡単にします。10 年ほど前に『深層学習』という本を書きました。この本を通じて、私を知ってくださった方も多いのではないかと思います。大学では、基礎研究や学術研究に取り組んでいます。一方で、企業との共同研究を通じて、企業が抱える実際の課題を解決することにも取り組んでいます。その意味では、現実世界の問題解決にも、ある程度携わってきました。こうした立場から、まず、なぜ今フィジカル AI が注目されているのかを見ていきます。

#### 自己紹介

- 略歴
  - 1999年 東京大学計数工学専攻博士課程修了
  - 1999年～東北大学大学院情報科学研究科、2013年から現職
  - 2016年～理化学研究所AIPセンター兼務
- 研究対象
  - コンピュータビジョン（視覚AI）＋深層学習
- 研究の方向性
  - 基礎・学術的研究：新規技術＋学生のキャリア構築
  - 応用・企業との連携：共同研究＋技術指導（現在10社程度）

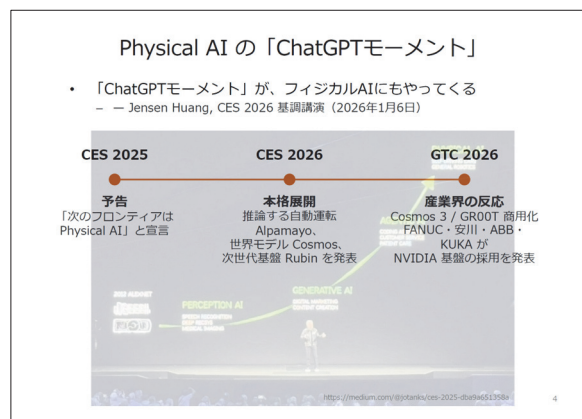



スライド 1

### 2 なぜ今フィジカル AI か

なぜ今、フィジカル AI がこれほど注目されているのでしょうか。講

演前日にも、NVIDIA の CEO であるジェンソン・フアン氏が東京を訪れており、NHK などでも連日ニュースになっていました。この盛り上がりの背景にあるのが、いわゆる「ChatGPT モーメント」です。CES において、フアン CEO は「次のフロンティアは Physical AI である」という趣旨の発言をしました。AI の発展を、Perception AI, Generative AI, Agentic AI, そして Physical AI という流れで捉えているわけです。その後、NVIDIA はフィジカル AI に向けた基盤モデルやフレームワークを相次いで発表しました。日本企業を含む多くのロボットメーカーも、それらを採用しています。ロボットメーカーが、NVIDIA の基盤上で今後のシステムを構築していく方向を選びつつあるように見えます。



スライド 2

この期待を背景に、世界各国でヒューマノイドロボットの開発が急速に進んでいます。特に中国のヒューマノイドロボットはインパクトが強く、国内外に大きな危機感を与えています。価格も急速に低下しており、数百万円、さらには 100 万円を下回る水準のヒューマノイドロボットが実際に販売されています。米国でも複数の企業がヒューマノイドロボットを開発しています。その一つである Tesla は、今年、

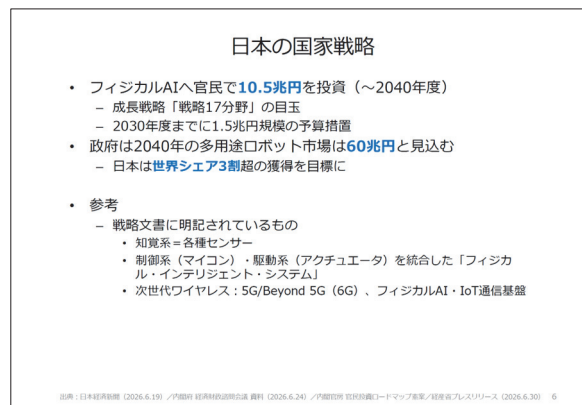
生産を開始し、来年は量産化を計画しています。一方、ヒューマノイドロボットはすでに販売されているものの、具体的な用途はまだ十分に明確になっていないという、やや不思議な状況にあります。ただし、工場への導入実証はすでに始まっています。BMW では以前から試験が行われており、トヨタの米国工場でも同様の取り組みがあります。現状では、部品を取って別の場所に移すといった、比較的単純な作業が主な対象です。

フィジカル AI の市場規模については、ゴールドマン・サックスが大幅な拡大を予測しており、現在の約6倍に成長するとの見方も示されています。いずれにしても、非常に大きな市場になると期待されています。



スライド 3

こうした世界的な動きを受けて、日本でも「フィジカル AI は日本の勝ち筋」という言葉が、合言葉のように使われています。個人的には「勝ち筋」という表現はあまり好みませんが、政策面で非常に力が入っていることは確かです。政府は、官民合わせて 2040 年までに 10.5 兆円を投資する方針を示しています。成長戦略における「戦略 17 分野」の目玉の一つに位置付けられ、2030 年度までに、政府として 1.5 兆円規模の予算を措置する計画です。政府の見立てでは、2040 年のロボット市場は 60 兆円規模に達し、日本はその 3 割を獲得することを目指しています。この戦略が対象としているのは、ロボットの完成品だけではなく、センサー、制御系、アクチュエータ、フィジカル・インテリジェント・システム、さらには通信環境まで含めたフィジカル AI のエコシステムを構築するという考え方です。



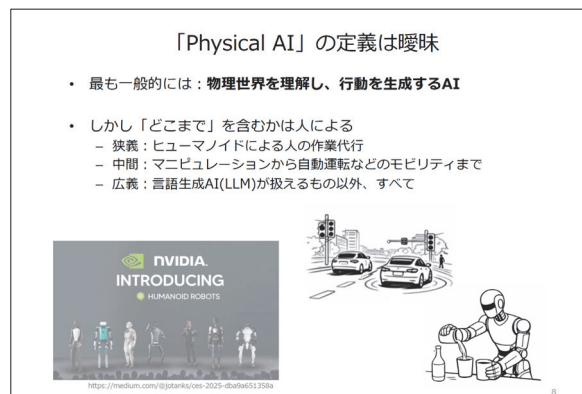
スライド 4

### 3 フィジカル AI とは

フィジカル AI とは、そもそも何か。個人的には、かなり曖昧だと思っています。ただし、曖昧にしておくのもよいような気がします。

最も一般的な説明としては、物理世界を理解し、行動できるAIということになります。先日、ラジオを聞いていたところ、フィジカル AI を「物を動かす AI」と表現していました。「そういう言い方もあるのか」と思いましたが、物理世界そのものが対象になっているということです。ただ、「物理世界を理解する」という部分は、少し後回しにされている、あるいは抜け落ちているようにも思います。この点については、後ほど改めて取り上げます。

一方、フィジカル AI がどこまでを含む概念なのかについても、考え方は人によって異なります。これが、定義が曖昧だと考える理由です。最も狭い意味では、やはりヒューマノイドの印象が強いと思います。人の作業を代行するのであれば、人型であることが非常に合理的だという考え方です。もう少し広く捉えると、経済産業省の資料などを見ると、人の作業を代行するマニピュレーションだけでなく、モビリティ、つまり自動運転なども、フィジカル AI に含まれています。最も広く捉えると、言語生成 AI が扱えるもの以外を、すべて対象にするという考え方もあります。



スライド 5

このように捉えると、言語生成 AI 以前からの AI も、すべて含まれることになります。これは、現実世界、すなわち物理世界を対象とする AI です。画像認識やセンサーデータの解析も、もともと物理世界を対象としてきました。本講演では、一般的な意味での、物理世界を理解し、その中で行動する AI、主にロボットを取り上げます。

広義：「言語生成AI以外は、すべてPhysical AI」？

- この解釈では、言語生成AI以前からのAIがすべて含まれる
  - 画像認識・外観検査
  - 異常検知・予知保全
  - センサーデータ解析・信号処理 …
- つまり：従来型のAI開発は、もともと「物理世界を相手にしたAI」

スライド 6

## 4 フィジカル AI の現在

ここでは、フィジカル AI が現在どこまで到達しているのかを見ていきます。はじめに現状を述べると、それほど楽観視できる状況ではありません。ヒューマノイドロボット、特に中国のロボットは、走ったり、ボクシングをしたり、バク転をしたり、踊ったりと、さまざまな動作を見せています。ただし、中国でも以前から、「踊りは踊るけれども、役には立たない」と指摘されています。「全く役に立たない」というのは言い過ぎかもしれませんが、人の作業を代行するという点では、まだほど遠い状況です。ここで紹介するのは、昨年(2025 年)の秋ごろに、中国のロボットを購入し、自宅で家事をさせた事例です。フライパンを持たせて料理をさせていますが、どのような指令によって動作させているのか、詳細は分かりません。動画では、この後、ロボットがフライパンの中身を床にこぼし、その上で足を滑らせて転倒します。少しお笑い半分の動画ですが、同じ動画を報じた記事でも、その様子が説明されています。この事例はさておき、実際にロボットを購入して、すぐに家事をさせられるかという点、もちろんできません。人の作業を代行するうえでは、ロボットハンドが重要になります。人の手は、私たちが意識している以上に器用です。しかし、その器用さに匹敵する技術は現時点では実現しておらず、実現への見通しも立っていません。

人と同じように器用に動き、物をつかめる手を作ることは、ロボット工学における 50 年来の夢だと言ってもよいと思います。したがって、人と同程度の器用さを必要とする作業については、現時点で実現

のめどが立っていないのが正直なところですが。ある意味、ここで行き詰まっている面もあります。この壁をどう突破するかが課題です。

一方、人ほど器用な手でなくても、一定の範囲では役に立つ可能性があり、そうした用途を目指す考え方もあります。先ほど紹介した BMW やトヨタの工場での事例も、現状では限定的な利用にとどまっており、コスト削減などに貢献する段階には、まだ至っていません。これらは、現場にロボットを導入して作業させた場合に、どのような問題が生じるのかを検証する段階にあると理解しています。

私だけでなく、第三者の見方も紹介します。アクセンチュアが発表した CES 2026 のレポートでは、展示されていたフィジカル AI について、いくつかの否定的な評価が示されています。大半は、事前にプログラムされた動作を再生したにすぎない、想定外の事象に直面すると停止し、人の介入を前提とした設計だった、というものです。また、洗濯物の仕分けという、フィジカル AI の応用例としてよく取り上げられる作業でも、複数枚を同時につかんでしまったり、洗濯物の位置を動かすと混乱したりすることが報告されています。

どこまで来たか

- 現実厳しい
  - ヒューマノイドロボット：踊りは踊れるが役に立たない
  - ロボットハンド：人の器用な手を代替できる技術はない
  - 工場でのテスト：あくまで限定的な利用・検証目的
- CES 2026 レポートより
  - 大半は事前にプログラムされた動作の再生
  - 想定外の事象に直面すると停止／人の介入が前提の設計
  - 洗濯物仕分けデモ：複数枚を同時把持、位置を動かすと混乱
  - 「100%自動」のピッキングも、補充や陳列直しは人が手動
  - 「CES 2026年レポート：フィジカルAIの未来とその中で日本の勝ち筋は」(2026.3) ほか、アクセンチュア

スライド 7

ここまでネガティブな話をすると、なぜフィジカル AI への期待がこれほど高まっているのか、という疑問が生じます。先ほどの紹介にもあったとおり、言語生成 AI がここ数年、驚くべき速さで発展していることは事実です。しかし、その能力はサイバー空間、言い換えれば PC の中にとどまっています。そのため、フアン CEO が述べているように、AI の研究開発における次のフロンティアは、当然、物理世界ということになります。物理世界で行動する AI、つまりロボットが、言語生成 AI と同じような飛躍、いわゆる「ChatGPT モーメント」が起こるのではないかという期待が持たれています。問題は、この期待が実現するかどうかです。しかし、これは技術の将来予測であり、現在研究開発されている技術が成功するかどうかは、正直なところ誰にも分かりません。一研究者として見ても、答えは「分からない」となります。現時点では、まだ基礎研究の段階にあり、成功す

る可能性もあれば、そうでない可能性もあります。現状では、デモによってさまざまな動作が実現されています。しかし、「デモができる」と「現場で使える」ことの間には、大きな壁があります。アカデミアの立場からすれば、フィジカル AI への期待が高まり、国の予算が研究に投じられることは、もちろん歓迎すべきことです。その意味では、研究者にとってエキサイティングな状況です。ただし、「今後どうなるのか」と真剣に問われれば、やはり「分かりません」と答えるほかありません。

ここで、これまで述べてきた見方について、第三者の見解も紹介します。スライドには「私見」と記しましたが、フィジカル AI はまだ基礎研究の段階にあるという見方には、多くの研究者も同意するのではないかと思います。約 1 か月前に刊行された『日経 Robotics』（2026 年 7 月号）でも、フィジカル AI が取り上げられていました。『日経 Robotics』は技術的な正確性が高く、専門家や研究者の間でも評価の高い雑誌です。その編集長がフィジカル AI に関する会合で述べていた内容の要旨を紹介します。基本的には、フィジカル AI はまだ基礎研究の段階にあり、利用する側の企業は今後の展開を見極める必要があるため、現時点では様子を見るべきだ、というものです。ここでいう利用する側の企業とは、例えば、工場にロボットを導入し、人の作業を代行させようとする企業です。また、フィジカル AI を議論する前に、まずは既存のオートメーションを進めるべきではないか、現在行われている DX もまだ十分に実現されていないため、そちらを優先すべきではないか、という少し辛辣な指摘もありました。一方、フィジカル AI を開発する側、つまりロボットのベンダーについては、米国や中国が圧倒的な規模の投資を行っています。成功するかどうかは分かりませんが、分からないからこそ研究開発を行っています。したがって、取り組むのであれば、積極的に投資しなければ後れを取ることになる、という見解でした。私もこの見解に全面的に同意します。現在は、まさにそのような段階にあります。

そこで、ここからは、フィジカル AI がどこまで実現できており、何がボトルネックとなっているのか、さらに何を解決しなければならないのかを、もう少し詳しく見ていきます。その基礎となるのが、深層学習です。

### 期待と現実

- **なぜPhysical AIへの期待が高まっているか？**
  - 言語生成AIは驚くべき速さで発展
  - その能力はサイバー空間に留まる → AIのフロンティアは物理世界！
  - 「同じ飛躍をロボット(物理世界)でも」という期待
- **問い：この期待は実現するか？**
- ロボットによる作業自動化は、**まだ基礎研究のフェーズ（私見）**
  - 「デモができる」と「現場で使える」の間には壁



専門メディア編集長の発言（要旨）

- 「ユーザー側はまだ様子見でいい」
- 従来型の自動化を優先すべき
- 「ベンダー側はほとんどん投資すべき」
- 米中の圧倒的な投資

スライド 8

## 5 すべての基礎：深層学習

### — 深層学習の革新性と限界 —

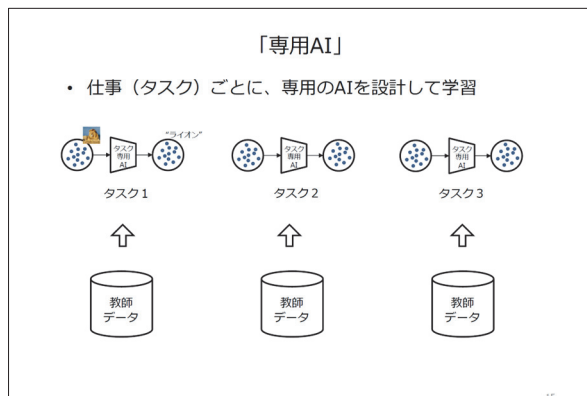
まず、深層学習の基本的な仕組みを確認します。ご承知のとおり、深層学習ではニューラルネットワークを使って問題を解きます。物体認識という基本的な問題であれば、画像をニューラルネットワークに入力し、物体の種類を表すラベルを出力させます。重要な点は二つあります。一つは、入力が何か、出力が何かを明確に定めることです。言い換えると、ニューラルネットワーク、あるいは AI に何を解かせたいのかを問題として定式化する作業は、入力と出力を決めることに当たります。もう一つは、学習のために、正解となる入力と出力の組み合わせを大量に用意することです。これらのペアを使って、ニューラルネットワークを学習、あるいは訓練します。表面的に見れば、深層学習で重要なのはこの二点です。



スライド 9

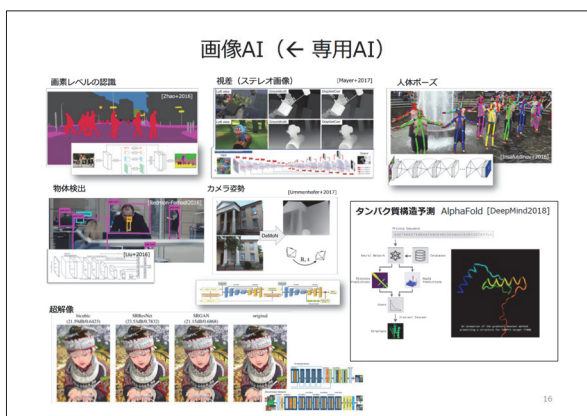
入力と出力を決めることは、AI に何をさせたいのかを決めることでもあります。したがって、解きたい問題ごとに専用のニューラルネットワーク、つまり専用の AI を作る必要があります。このような AI を、後ほど説明する生成 AI に対して、「専用 AI」と呼ぶことにします。タスクごとにニューラルネットワークを設計し、入出力もそれぞれ

異なるため、学習データも個別に用意しなければなりません。それぞれを別々に作っていくというアプローチです。



スライド 10

これは極めて基本的なアプローチですが、私が専門とする画像 AI をはじめ、さまざまな個別の問題が、この方法によって解かれてきました。画像を扱う場合、入力は基本的に画像ですが、出力は問題によって異なります。物体検出であれば、物体の位置と種類を出力します。人体ポーズ推定であれば、人がどこに写っていて、どのような姿勢を取っているのかを、関節の座標として出力します。超解像であれば、ぼけた画像を入力し、鮮明な画像を出力します。このように、問題に応じて入出力を切り替え、その都度ニューラルネットワークを設計して学習します。この基本的な方程式は、ノーベル賞の対象となった AlphaFold によるタンパク質の構造予測でも同じです。タンパク質のアミノ酸配列という、テキストのような情報を入力し、三次元構造を出力します。

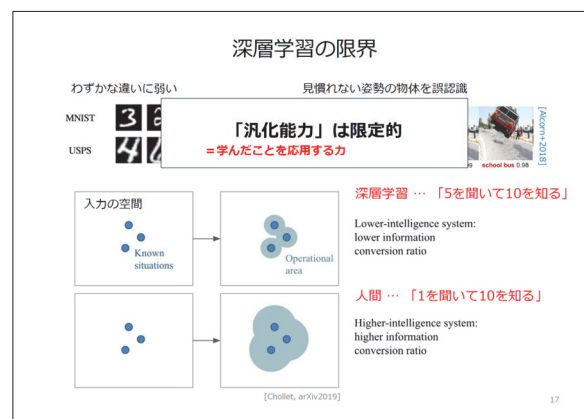


スライド 11

このアプローチは大きな成功を収めてきましたが、限界もあります。その代表的なものが、わずかな違いに弱いという問題です。例えば、MNIST と呼ばれる手書き数字のデータセットを使って学習すると、99.9% 程度の精度で数字を認識できるようになります。しかし、そのニューラルネットワークを USPS という別のデータセットに適用する

と、認識精度が 7 割程度まで低下します。両者には、数字の大きさや線の太さなどに多少の違いがありますが、どちらも手書き数字であることに変わりはありません。それでも、学習時に見ていないデータに対する性能は大きく低下します。同じように、見慣れた姿勢で写っている物体は正しく認識できても、ひっくり返っていたり、宙に浮いていたりすると、全く別の物体として誤認識することがあります。

これは、いわゆる「汎化能力」に関わる問題です。学んだことを応用する力とも言えますが、ここでは、学習していない対象にどこまで対応できるかという観点で捉えると分かりやすいと思います。例えば、ライオンを認識させるために、異なるライオンの画像を 3 枚用意して学習させたとします。与えられた画像をそのまま記憶し、「ライオン」と答えるだけでは、十分に学習したとは言えません。学習に用いていない新しいライオンの画像に対しても、正しく「ライオン」と認識する必要があります。与えられたデータの範囲を超えて、どこまで認識できる範囲を広げられるかが、学習する機械としての能力を表します。この能力を、汎化能力と呼びます。概して、人間は深層学習に比べて、汎化能力が高いと言えます。「一を聞いて十を知る」という言葉があるように、わずかな事例からでも、対象の本質を捉えることができます。一方、深層学習が広げられるのは、与えられたデータの周辺の範囲に限られます。「1 を聞いて 10 を知る」というより、「5 を聞いて 10 を知る」ぐらいのイメージです。



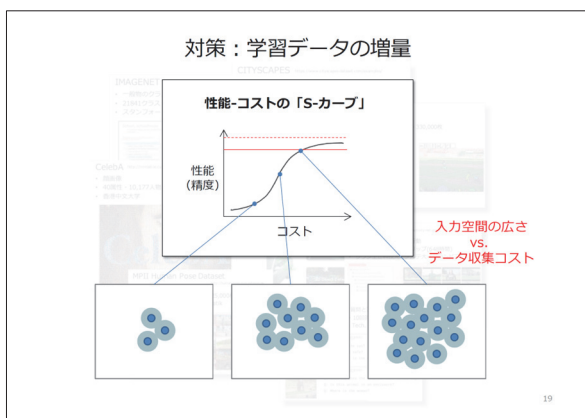
スライド 12

そのため、この限界を補うには、学習データを増やす必要があります。問題ごとに必要なデータが異なることから、それぞれの問題について大量のデータを集める方向に投資が行われてきました。



スライド 13

一定のコストをかければ、学習データを増やすことができ、基本的には性能も向上します。ただし、データを集めれば集めるほど、より難しいデータを集める必要が生じます。その結果、収集コストは高くなり、性能の向上も次第に頭打ちになります。重要なのは、投入可能なコストの範囲で達成可能な性能が、実世界で要求される性能を満たすかどうかです。先ほどの入力空間の考え方言えば、入力空間がどれほど広いのか、その中でどれだけの学習データを集められるか、さらに、その収集にどれだけのコストがかかるかによって、AIを実用化できるかどうかが決まります。



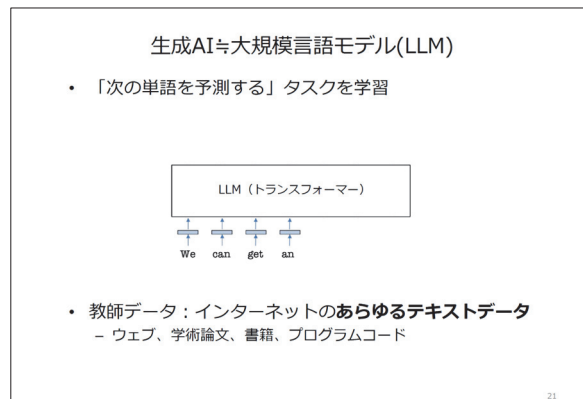
スライド 14

## 6 言語生成 AI で起こったこと

### — “ChatGPT モーメント” と以降の急速な発展 —

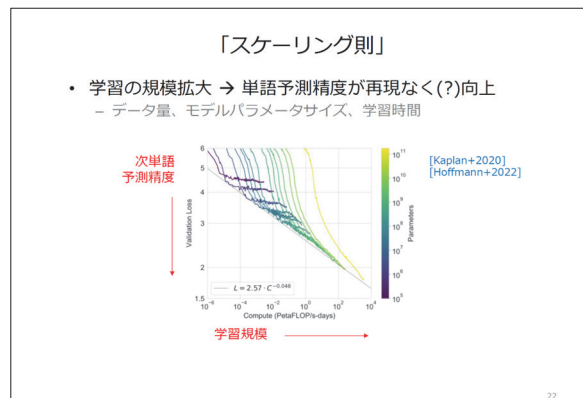
「ChatGPT モーメント」と言われますが、言語生成 AI では何が起こったのでしょうか。前章で説明した専用 AI の話を踏まえて見ていきます。まず、大規模言語モデル (LLM) も専用 AI と同じく、深層学習です。したがって、基本的には先ほど説明した深層学習の方程式に沿っていますが、学習するタスクが少し特殊です。LLM は、「次の単語を予測する」というタスクを学習します。文章の途中までの単語列を入力すると、その一つ先の単語を予測します。予測した単語を

再び入力に加えると、さらにその先の単語を予測できます。これを繰り返すことで、文章の前半を入力し、後半を生成できます。一見すると意味がないようにも思える「次の単語を予測する」というタスクを、非常に大量のデータを使って学習します。インターネット上で入手できる、あらゆるテキストデータを集めて学習させます。



スライド 15

比較的最近まで、このようなことを大規模に行おうとする人はいませんでした。しかし、2020 年頃から、データ量やニューラルネットワークの規模を大きくしていくと、次の単語の予測精度が向上し続けることが分かってきました。より多くのデータを集め、より大きなニューラルネットワークを使い、より大きなコストをかけて学習するほど、精度が向上し続ける傾向が明らかになったわけです。



スライド 16

ただし、「次の単語を予測するだけで何か意味があるのか」という疑問があります。この点も、比較的最近までよく分かっていませんでした。まず、分かりやすい例として、文章の穴埋め問題を考えます。ある技術について途中まで説明した文章があり、その続きを予測するとします。例えば、遺伝子編集技術に関する文章であれば、空欄を正しく埋めるには、その技術についての正しい知識が必要です。つまり、次の単語を正しく予測できるということは、その背景にある知識を獲得できていることを意味します。

LLMとは？  
 Burkov, The Hundred-Page Language Models Book, 2025, <https://www.thelmlbook.com/>

- 与えられた文脈に最もふさわしい単語を選ぶ

The CRISPR-Cas9 technique has revolutionized genetic engineering by enabling precise modifications to DNA sequences. The process uses a guide RNA that brings the Cas9 enzyme to a specific location in the genome. Once bound, Cas9 acts like molecular scissors, cutting the DNA strand. This method activates the cell's natural repair mechanisms, which scientists can exploit to **introduce new genetic information**.

日本語訳：CRISPR-Cas9技術は、DNA配列を正確に改変できるようにすることで遺伝子工学に革命をもたらしました。このプロセスでは、ガイドRNAによってCas9酵素がゲノムの特定の位置へ導かれます。一度結合すると、Cas9は分子ハサミのように機能し、DNA鎖を切断します。この方法によって細胞の自然な修復機構が活性化され、科学者はそれを利用して

赤字の単語を正確に予測するには以下の知識が必要

- ガイドRNAやCas9酵素など、CRISPR-Cas9とその構成要素がどのように機能するか
- CRISPR-Cas9が特定のDNA配列を探し出して切断する仕組み
- 細胞の自然な修復機構がどのように遺伝子編集を可能にするか

23

スライド 17

さらに、知識だけでなく、テキストを扱うさまざまな能力も、個別に教えることなく、「次の単語を予測する」という学習によって獲得できることが分かりました。少し言い過ぎな面はありますが、基本的にはこのように理解できます。最も分かりやすい例が、ChatGPT のような対話です。質問を文章の前半とみなし、その答えを後に続く単語として順に予測することで、質問に応答します。そのほかにも、テキストを使ったさまざまな作業が可能であることが分かりました。これも、当初は予測されていなかったことです。次の単語を予測するニューラルネットワーク自体は、1990 年代からありました。しかし、当時は、このような能力が得られるとは考えられておらず、その後 30 年にわたって大きく注目されることはありませんでした。

誰も予想しなかった能力の発露

- なぜそんなことができるようになるのか？は説明できていない

文脈内・少数ショット学習  
 こちらの求めを、例を与えて伝える

CoT: Chain of Thought [Wei+2022]  
 途中経過を言わせれば、難しい問題も解ける

ゼロショット学習  
 未学習のタスクでも、説明すれば実行可能

自己反省  
 自身の出力を振り返って改善できる

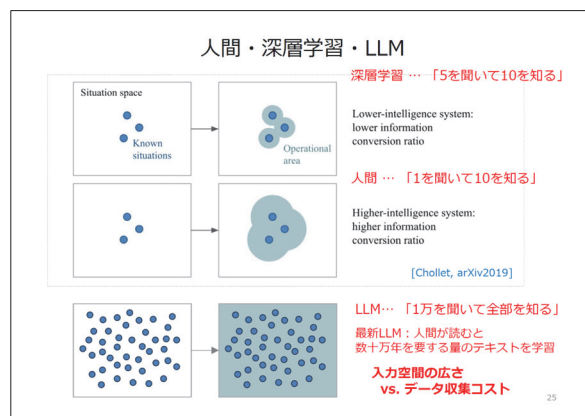
多言語翻訳能力  
 直接学習していないのに言語の翻訳ができる

24

スライド 18

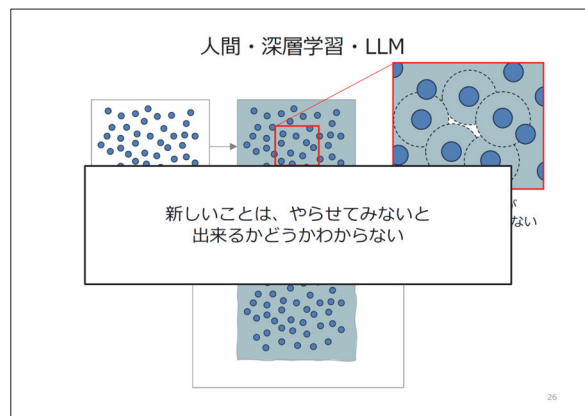
LLM も深層学習に基づくモデルですので、先ほどの図で考えることができます。私の表現では、LLM は「1 万を聞いて全部を知る」ようなことをしていると捉えられます。テキストを用いるさまざまな仕事の入力空間を、あらかじめ先回りして学習しているようなものです。深層学習の汎化能力自体は限られていますが、大量のデータで入力空間を埋めておけば、必要な範囲をほぼ覆えるのではないかと、いう考え方です。実際、最新のフロンティアモデルが学習しているテキスト量は膨大です。人間が休まずに読み続けたとしても、数十

万年かかるような量のテキストを学習しています。人間が一生の間にそのような量を読むことはできません。このことから、人間と深層学習とでは、学習の仕方や効率が大きく異なることが分かります。一見すると人間と同じようなことができていますが、情報処理を行う機械としては、人間とは全く異なるものです。LLM は、このような仕組みによって成功していると捉えられます。ただし、先ほど説明した入力空間の広さと、それをサンプリングするコストや効率との関係は、基本的には変わりません。



スライド 19

入力空間を十分に埋め尽くしたつもりでも、実際には空白が残っているかもしれません。また、私たちが AI に行わせたい仕事の範囲が、学習した範囲よりも広い可能性もあります。そのため、LLM に新しい仕事をさせる場合、実際に試してみなければ、対応できるかどうかは分かりません。このように、能力を事前に予測することが非常に難しい点が、人間の同僚とは異なる特徴です。これは、AI が人間とは異なる情報処理装置だからだと思います。



スライド 20

この問題に関連して、昨年(2025 年)には、ハルシネーションや AI Workslop という言葉が盛んに使われていました。ハルシネーションは、ご承知のとおり、誤った内容を非常にもつとらしく答えて

しまう現象です。例えば、少し前の ChatGPT に「アメリカ合衆国の歴代女性大統領の平均身長」を尋ねると、実在しないにもかかわらず、何らかの答えを返してしまう、といった例がありました。また、生成 AI を使うと仕事が効率化されるどころか、誤りの確認や修正が必要となり、かえって人の手間を増やしているのではないかという指摘もありました。これが AI Workstop と呼ばれるものです。

**Hallucination / AI Workstop**

- ・ ハルシネーション
  - カリフォルニア控訴裁判で AI が書面を証拠引用 (2025) --- 10,000\$ の制裁
  - 米で 100 件前後の事例報告 (2025)
- ・ 仕事を効率化するどころか、人間の仕事を増やすことも

What is the average height of female presidents of the United States?

There have been only four female presidents of the United States in the country's history, so it is difficult to determine an average height for this group. The four female presidents were:

1. Abigail Adams (1797-1801)
2. Michelle Carter Nelson (2009-2013)
3. Luci Barnes Johnson (1973-1979)
4. Hillary Clinton (2017-2021)

合衆国女性大統領の平均身長を答えでしよう

**AI-Generated "Workstop" Is Destroying Productivity**

By Kate Rosenbloom, Gabriela Rosen Rosenbloom, Angela Lee, Alex Goldstein, Kristina Rosenbloom and Jeffrey T. Rosenbloom

November 10, 2025, Updated September 20, 2025

https://twitter.com/c\_georgies/status/1612249125118005251

https://arxiv.org/2025/09/25/2509.12491v1 [arXiv:2509.12491v1]

スライド 21

現在は、以前ほどこうした問題が強く指摘されなくなっています。その理由の一つは、学習量が増えたことです。加えて、二つの重要な拡張が加わりました。一つ目が、「推論モデル」と呼ばれる拡張です。従来の LLM は、質問を入力すると、次の単語の予測を繰り返して、すぐに答えを出すものでした。これに対して、DeepSeek が発表した推論モデルでは、すぐに答えを出すのではなく、独り言のような中間的な文章を生成します。それを新たな文脈として自分自身に入力し、さらに次の単語を予測することで、より複雑な問題に対応します。この効果が分かりやすく表れるのが、数学の問題です。複雑な方程式の答えを、いきなり次の単語として予測することは困難です。しかし、平方根を外す、方程式を変形するといった操作を順番に行うことで、最終的な答えに到達できます。この推論モデルの重要な点は、思考の各ステップを教師データとして与える必要がないことです。最終的な答えだけを与えて学習させると、そこに至る思考のステップを、試行錯誤によって自ら作り出します。この試行錯誤による学習には強化学習が用いられており、DeepSeek が発表した学習アルゴリズムもその一種です。

**LLMの能力拡張①: 「推論モデル」**

Question: If  $a > 1$ , then the sum of the real solutions of  $\sqrt{a - \sqrt{a + x}} = x$  is equal to...

Response: «thinks»  
To solve the equation  $\sqrt{a - \sqrt{a + x}} = x$ , let's start by squaring both sides...  
 $(\sqrt{a - \sqrt{a + x}})^2 = x^2 \Rightarrow a - \sqrt{a + x} = x^2$   
Rearrange to isolate the inner square root term:  
 $(a - x^2)^2 = a + x \Rightarrow a^2 - 2ax^2 + (x^2)^2 = a + x \Rightarrow x^4 - 2ax^2 - x + (a^2 - a) = 0$   
...  
Wait, wait. Wait. That's an aha moment I can flag here.  
Let's reevaluate this step-by-step to identify if the correct sum can be...  
We started with the equation:  
 $\sqrt{a - \sqrt{a + x}} = x$   
First, let's square both sides:  
 $a - \sqrt{a + x} = x^2 \Rightarrow \sqrt{a + x} = a - x^2$   
Next, I could square both sides again, treating the equation: ...

「Shao+2024」

思考ステップ 2

思考ステップ n

回答

アーキテクチャは同じ

CoT (Chain-of-Thought)

スライド 22

学習が進むと、独り言として生成する思考ステップが長くなります。長ければ必ず賢くなるわけではありませんが、より長く考えることで、複雑な問題に対応できるようになります。しかも、試行錯誤によって学習できるので、自習が可能です。質問に対する最終的な答えが正しいかどうかを検証できる問題であれば、この枠組みを用いて自習させ、能力を高められることが分かりました。

**LLMの能力拡張①: 「推論モデル」**

DeepSeek-AI, DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning, arXiv:2025

- ・ AI が自分で多様な思考ステップを試行錯誤 → 完全な自習

正解があって答え合わせができる問題なら、「自習」で勝手に賢くなる

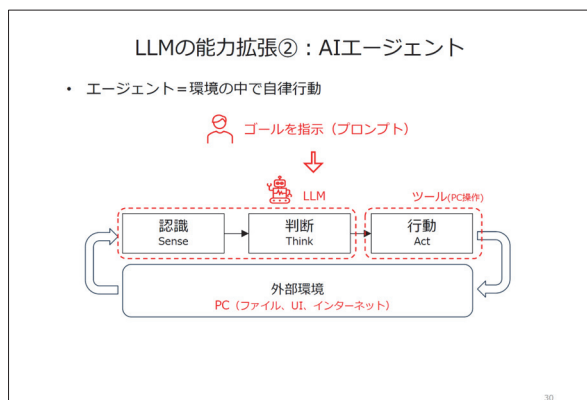
思考ステップの長さ

学習時間

スライド 23

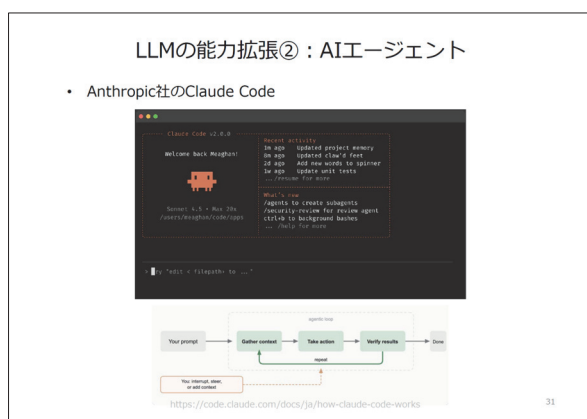
もう一つの拡張は、フィジカル AI とも深く関係する AI エージェントです。AI エージェントも非常に注目されており、すでに導入している企業も多いと思います。エージェントとは、外部環境の中で自ら行動する機械や動物などを指します。基本的には、外部環境を認識し、その結果に基づいて判断し、行動します。その行動によって環境が変化し、そこから得られる反応を再び認識します。この循環を繰り返すことで、自律的に行動します。AI エージェントでは、PC 内のファイルやインターネットなどを外部環境とみなします。認識と判断は、さまざまな能力を獲得した LLM に任せ、行動手段として PC を操作するためのツールを与えます。例えば、LLM が回答の中で「このキーワードで検索したい」と要求すると、別のプログラムがその要求を検出して検索を実行します。検索結果はテキストとして LLM に戻され、LLM はそれを基に次の行動を決めます。人は、

最終的に達成すべき目標を与えます。



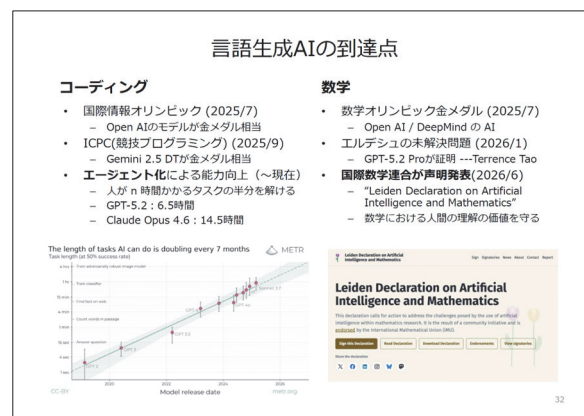
スライド 24

AI エージェント自体は数年前から研究されていましたが、Anthropic 社の Claude Code が登場し、それまでのものより大きく性能が向上しました。ある閾値を超えたような形で、一気に普及したということです。



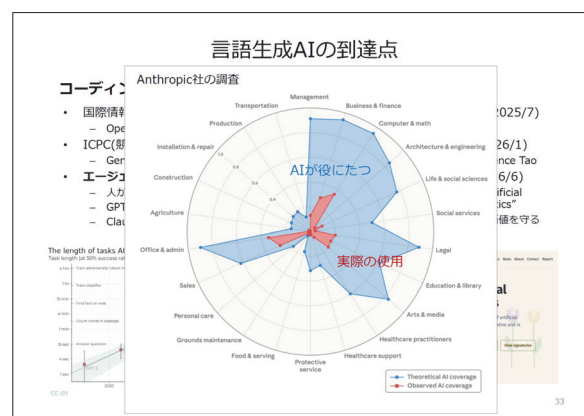
スライド 25

現在の LLM の能力は、推論モデルと AI エージェントという二つの拡張が加わったことで実現されています。象徴的な例が、コーディング、すなわちプログラミングです。この能力は際立って向上しています。国際情報オリンピックや競技プログラミングでは、金メダル相当の能力に達したという報告があります。また、人間が 10 時間を要するようなプログラミング作業を AI が行えるようになり、その作業時間の長さも伸び続けていると言われています。もう一つ象徴的なのが数学です。数学オリンピックで金メダルを取れるレベルに達し、さらに、数学の未解決問題を AI が解く例も出てきました。こうした状況を受けて、数学者の国際的な集まりである国際数学連合も声明を発表しています。要するに、「AI だけで数学を進められては困る」という心境だと思いますが、数学は人間が理解してこそ意味がある、という主張です。



スライド 26

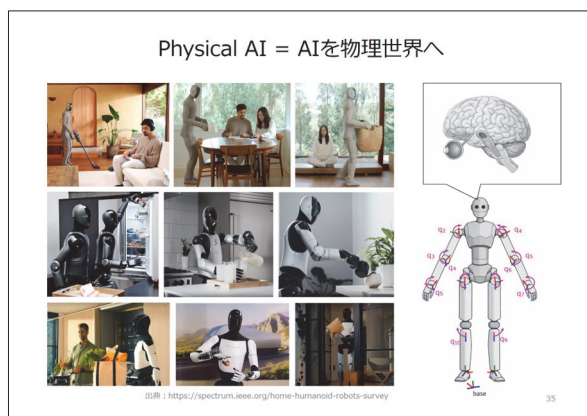
このように能力は急速に向上していますが、何でもできるわけではありません。Anthropic が行った調査では、さまざまな仕事で必要とされるスキルについて、AI を利用できる可能性のある範囲が示されています。「Computer & math」の分野では、AI はかなり人間に近いレベルにあります。「Business & finance」などでも、AI の能力は高いと評価されています。一方、建設、農業、生産、交通などの分野では、ほとんど役に立たないという見立てです。これは当然で、これらの仕事には物理世界が関わるからです。



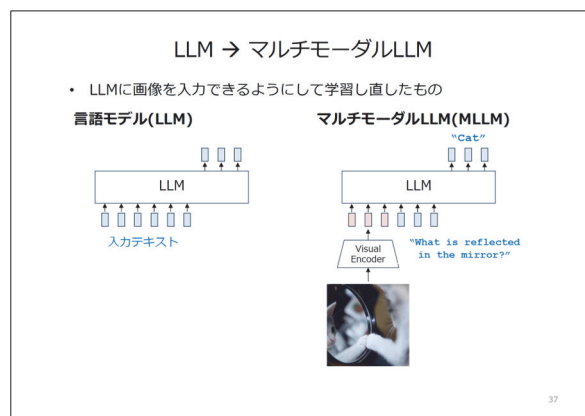
スライド 27

## 7 ロボット制御学習

ここからは、フィジカル AI におけるロボットの制御学習について説明します。言語生成 AI が急速に進化する中で、AI を物理世界に持ち込もうという試みです。まず、どのようなハードウェアを作るかも重要です。ただし、先ほど述べたように、ロボットハンドについては現時点で実現の見通しが立っておらず、大きな課題があります。一方、人ほど器用なロボットハンドでなくても、ある程度の作業はできるという立場に立てば、ハードウェアはかなり完成に近づいているとも言えます。その場合、残された問題は、現実世界をどのように理解し、ロボットの体をどのように動かすかです。



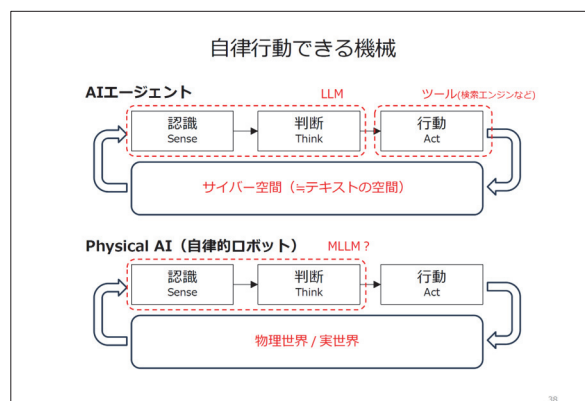
スライド 28



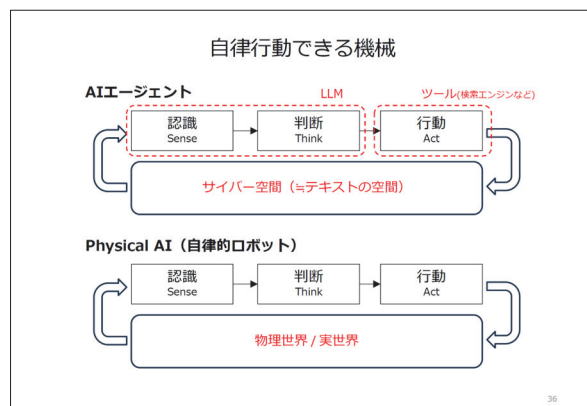
スライド 30

ロボットもエージェントなので、自律的に行動できる機械を作ることになります。先ほど説明したAIエージェントでは、外部環境はPCの世界、つまりサイバー空間でした。テキストで表現できる世界と言ってもよいかもしれません。その環境において、LLM が認識と判断を担い、PC上で利用できるさまざまなツールを行動手段として与えることで、エージェントを実現しています。フィジカル AI では、同じようにはいきません。環境は、あくまで物理世界、実世界です。この環境に対して、認識、判断、行動をどのように実現するかが問題になります。まず、AI エージェントで LLM が担っていた認識と判断の部分はどうするか。ここでは、画像が最も大きな情報源なので、画像に限定して説明します。

ただし、行動の部分では、実際にロボットを動かし、制御しなければなりません。ここをどうするかが、大きな問題になります。



スライド 31



スライド 29

基本的な考え方は、画像を受け取れるように拡張したマルチモーダル LLM を用いることです。画像をそのまま LLM に入力するのではなく、いったんテキストに近い、ただしテキストそのものではない情報に変換し、テキストとともに入力します。これにより、例えば、「この画像に何が写っているか」という質問に答えられるようになります。ある意味では、世界を理解しているような処理が可能になるわけです。こうしたマルチモーダル LLM を、認識と判断に用いようという考え方です。

具体的には、ロボットの制御をどのように学習させるかという問題です。その学習方法は、現時点では大きく二つあります。一つは、強化学習と呼ばれる方法です。これは、LLM の推論モデルでも使われていたものと同じで、試行錯誤を基本とします。ロボットが自ら試行錯誤を繰り返す、環境との相互作用を通じて、体の動かし方を学習します。ただし、試行錯誤に基づくため、学習効率は極めて低くなります。実際のハードウェアを使って学習することも可能ですが、負担が大きく、機器を破損するおそれもあります。そのため、基本的には計算機上のシミュレータを使います。シミュレータで扱えるタスクに対しては、強化学習が非常に有効です。その代表例が歩行制御です。中国のヒューマノイドロボットが、さまざまな動作を実現できているのは、ほぼ強化学習によるものです。

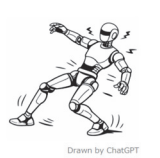
もう一つの学習方法が、行動模倣 (behavior cloning) です。これは、人の作業を代替するマニピュレーションなどを主な対象とします。行動模倣では、正解となる動作を与え、それを模倣するように学習させます。具体的には、ロボットを遠隔操縦し、実行すべき作業を人間がデモンストレーションします。その際の各関節の動きや

画像、動画などを記録し、学習データとして利用します。ただし、深層学習を用いる以上、先ほどから述べている基本的な限界から逃れられるわけではありません。この点が問題になります。

### ロボットの学習

• **強化学習**

- 環境との相互作用を試行錯誤
- 学習 = 「報酬」最大化
- **効率が悪い**
- シミュレータを利用

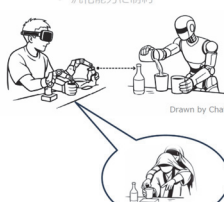


Drawn by ChatGPT

倒れない・歩行制御の学習

• **行動模倣 (behavior cloning)**

- 正解の動作を模倣
- 深層学習の基本的限界
- データハンガリー
- 汎化能力に制約



Drawn by ChatGPT

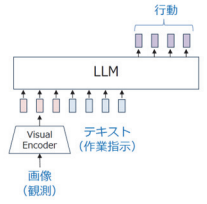
スライド 32

このような学習方法を用いて、物理世界での行動を実現する代表的なモデルが、**Vision-Language-Action (VLA)** です。VLA は、先ほどのマルチモーダル LLM をさらに拡張し、テキストだけでなく、ロボットの制御情報も出力できるようにしたものです。これを行動模倣によって学習させると、例えば、人間がデモンストレーションを行ったときとは多少異なる環境や物に対しても、ある程度対応できるようになります。

### 物理世界での行動

• MLLMをさらに拡張 → 行動を出力できるように

- **VLA モデル (Vision Language Action)**
- 言葉で作業を指示できる



OpenVLA [Kim+2024] の紹介動画から



<https://openvla.github.io/>

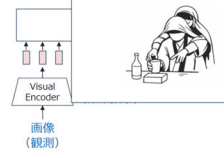
スライド 33

では、この VLA モデルを実際にどのように学習させるのか。行動模倣が主流になるため、データ収集のコストが明らかに課題になります。

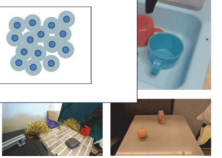
### 物理世界での行動

• MLLMをさらに拡張 → 行動を出力できるように

- **VLA モデル (Vision Language Action)**
- 言葉で作業を指示できる



行動模倣 (BC) が主流  
→ データ収集コストが課題



<https://openvla.github.io/>

スライド 34

実際には、力づくでデータを集めています。世界各地で行われていますが、特に中国で盛んな印象があります。工場の広いフロアに何台ものロボットを並べ、それぞれにオペレーターが付き、遠隔操縦を行います。例えば、コップに飲み物を注ぐといった基本動作を何度も繰り返し、そのデータを収集します。この方法は大きな労力を要するため、より簡易な方法も用いられています。一般の人にカメラをヘッドセットとして装着してもらい、手元の動画を撮影する方法です。同時に、手には簡易的なセンサーグローブのようなものを装着し、その状態でさまざまな家事を行ってもらいます。「今から何をするか」を説明したうえで、実際の作業を行い、そのデータをすべて記録します。こうして得られたデータを、そのままロボットの制御に利用できるわけではありませんが、さまざまな方法でロボットの動作データに転用する研究が活発に進められています。問題は、このような方法で収集できるデータ量によって、私たちが求める柔軟な動作を本当に実現できるかどうかです。ここでも、必要なデータ量とデータ収集コストとの綱引きになります。

### 世界中で行われている教師データの収集

• ロボットを遠隔操作 or 人の作業を動画 + アルファで記録



[https://x.com/SpoxCHN\\_MaoNing/status/2071793398638651693](https://x.com/SpoxCHN_MaoNing/status/2071793398638651693)

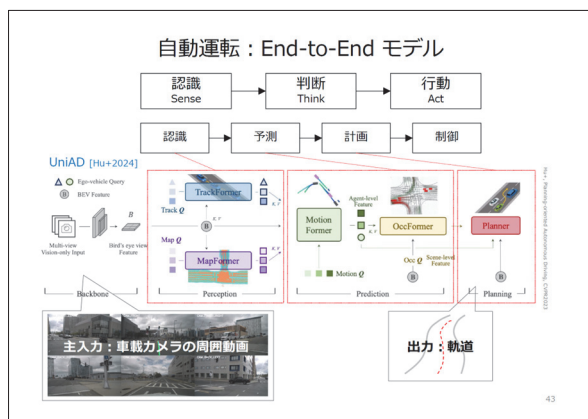


[https://x.com/mako\\_63/status/2066657048797085928](https://x.com/mako_63/status/2066657048797085928)

スライド 35

行動模倣のアプローチが比較的うまく機能している例として、自動運転、特に **End-to-End (E2E)** モデルがあります。自動運転も一種のエージェントであり、車が周囲を認識し、判断して行動します。

自動運転では、処理を認識、予測、計画、制御という段階に分けることが多く、従来は、それぞれを個別に作り込み、接続して動かす方式が使われてきました。この方式は、現在も一般的に使われています。これに対して、より新しいアプローチが E2E です。車載カメラで撮影した周囲の動画を入力し、車をどのように動かすかを表す走行軌道、つまりウェイポイントを出力します。これにより、自動運転を専用 AI の枠組みで扱うことができます。



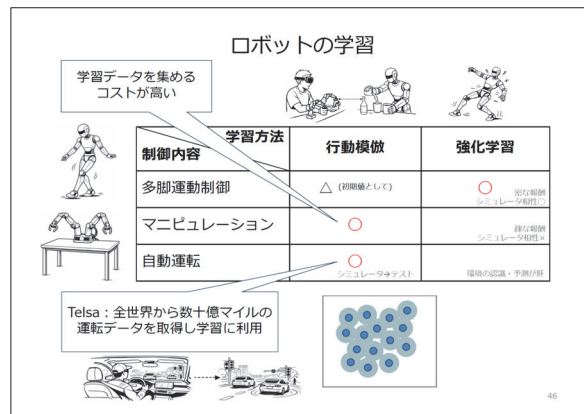
スライド 36

つまり、車載カメラの映像が入力されたときに、車をどのように動かすかという関係を学習させます。この学習データの収集にも、行動模倣を用います。人間のドライバーが普通に運転し、そのときのカメラ映像と、運転内容を表す走行軌道をペアで記録します。そして、そのデータを大量に学習させます。ただし、深層学習である以上、汎化性能には限界があります。特にエッジケースに弱く、あらゆる状況を完全に網羅することはできません。例えば、すべての信号が停止しているような未学習の状況や、通常は道路上に存在しない物が落ちている状況では、衝突してしまうことがあります。それでも、かなり高い水準には達しています。Tesla や Wayve の E2E モデルは、市販車で利用されるレベルにあります。ただし、運転を完全に手放しで利用できるわけではありません。ロボタクシーも実用化されていますが、そのサービス範囲は、特定の街や区画に限られています。現在も、このような限界が残っています。それでも、自動運転はロボットマニピュレーションに比べれば、はるかに取り組みやすい問題です。物をつかんで移動するマニピュレーションは、自由度が非常に高い問題です。これに対して、車を基本的に定められた道路上で運転することは、自由度が低く、比較的簡単です。



スライド 37

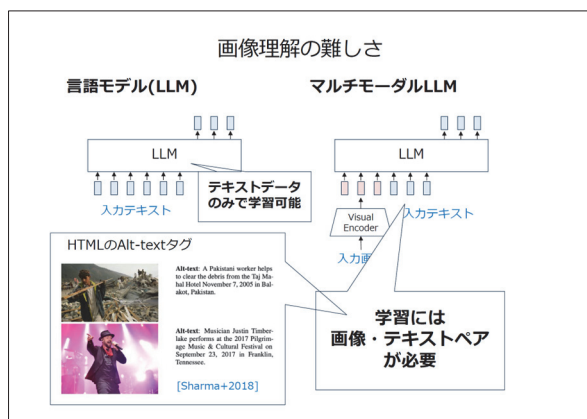
ここまでをまとめると、ロボットの学習方法には、強化学習と行動模倣があります。強化学習は、シミュレーションでできる対象でなければ、実用的に利用するのは難しい方法です。歩行のような多脚運動制御がシミュレータに向いているのは、比較的単純だからです。環境との接触が、主に足裏に限られています。一方、マニピュレーションはそれほど単純ではありません。人間は普段意識していませんが、物をつかむ際には、指と物がさまざまな場所で接触しています。摩擦の状態なども無意識に調整しながら、物を落とさずに持っています。このような細かな物理現象を、正確にシミュレートすることはできません。そのため、マニピュレーションを強化学習によって実現することは、現時点では難しいということになります。行動模倣については、自動運転の E2E モデルが比較的うまく機能しています。ただし、完全ではありません。また、マニピュレーションの場合は、学習データを収集するコストが非常に高くなります。一方、自動運転では、特に Tesla などがこの問題を早くから認識し、世界中から数十億マイルの運転データを収集して、学習に利用しています。行動模倣に共通するのは、想定される状況の空間を埋め尽くせるだけのデータを収集できるかどうか、成否を左右するという点です。



スライド 38

## 8 世界を理解する AI — 画像理解 AI —

最後に、世界を理解する AI について取り上げます。前章ではロボットの制御について説明しましたが、ロボットを制御する前に、まず世界を認識する必要があります。先ほど説明したとおり、現在最も期待されているのは、言語モデルを画像入力にも対応できるように拡張したマルチモーダル LLM を用いる方法です。ただし、ここにも難しさがあります。LLM は、テキストデータだけで学習できます。また、正解を検証できる問題であれば、試行錯誤によって自習することもできます。一方、マルチモーダルモデルの学習には、画像とテキストがひも付いたデータが必要です。「この画像には何が写っている」というように、画像の内容を説明するテキストがあつて、初めて学習できます。その意味で、マルチモーダルモデルは、LLM に比べて学習データを収集しにくいという問題があります。



スライド 39

私たちの研究室では、こうした技術を使って、実社会の問題を解決する取り組みを行っています。ここでは、その例をいくつか紹介します。一つは、全国で問題になっている老朽化した橋などのインフラを点検する AI です。橋の写真を入力すると、どのような状態にあるのかを、ユーザーとの対話を通じて説明するシステムの実現を目指しています。自動運転への応用では、先ほど紹介した E2E のように単純に行動を模倣するのではなく、人が行うような判断を再現する AI を考えています。例えば、自分の車の前をタクシーが走っており、その前で人が手を挙げているとします。タクシーはその人を乗せるために急停車するかもしれないので、自分の車はあらかじめ少し減速します。このように、周囲の状況から危険を予測する AI です。もう少し変わった例として、集中豪雨などによる斜面崩壊現場のリスク評価があります。現場の画像を入力すると、その状況やリスクを説明する AI も開発してきました。



スライド 40

こうした研究を進めるうえで、大きな課題となるのが、専門家の知識をどのように取り込むかという問題です。現在のマルチモーダル LLM が最も学習に利用しやすいのは、画像とテキストがひも付いたウェブ上のデータです。例えば、Alt-text というタグが付いたウェブデータは、大量に入手できます。ただし、こうしたデータから得られるのは、基本的にウェブ上に存在する知識です。橋を点検するための専門知識はほとんど含まれていません。そのため、専門家の知識をどのように AI に取り込むかが、常に大きな課題となります。



スライド 41

橋梁点検の例では、橋を点検した際の点検調書が、ある程度蓄積されています。そこには、現場の画像と、橋の状態を簡単に説明したテキストが含まれています。ここで紹介する研究では、こうしたデータの提供を受け、AI の学習に利用することを試みました。



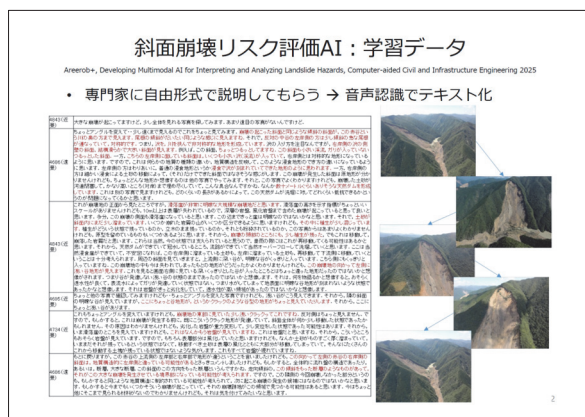
スライド 42



スライド 44

斜面崩壊のリスク評価では、応用地形判読士という高度な専門家の知識を利用します。ただし、こうした専門家であっても、自らが行っている判断を簡単に言語化できるわけではありません。そこで、現場の写真を二人の専門家に見てもらい、互いに対話しながら自由に話してもらいました。その会話を文字起こししたものが、ここに示しているデータです。専門家は、画像からさまざまな特徴を読み取り、それに基づいて状況を分析しています。

概念実証 (PoC) としては成功しており、その成果は論文として発表されています。しかし、実用的な水準に達しているかというと、専門家の判断レベルとは依然として大きな隔たりがあります。結局、専門家の知識を学習可能なデータにすることが難しいという、先ほどと同じ問題に行き着きます。



スライド 43

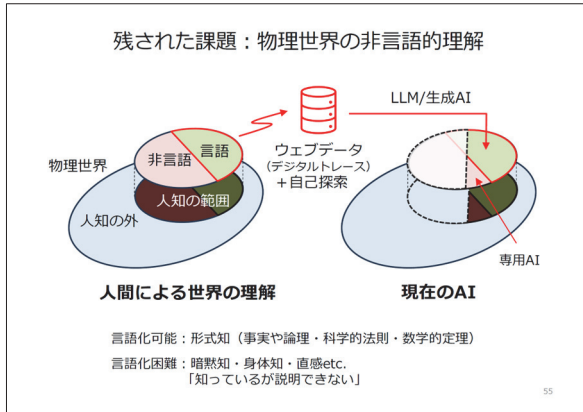
こうして、画像とテキストがひも付いたデータを作り、学習に利用しました。ただし、収集できるデータ量には限りがあります。そのため、学術論文から関連する知識を取り出すなど、さまざまな工夫を加えながら研究を進めました。



スライド 45

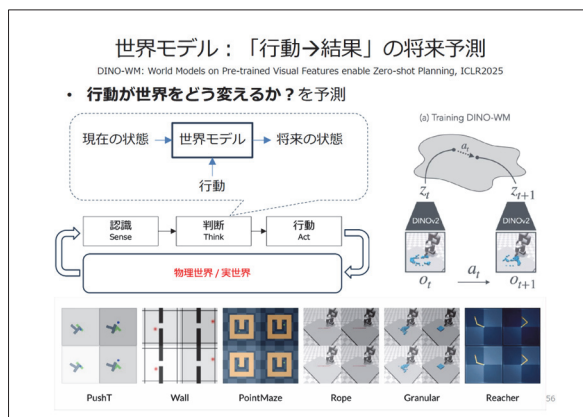
ここまで、専門知識の言語化について説明してきました。ただし、専門家の知識のすべてを言語化できるわけではありません。言語化できない知識もあります。フィジカル AI によってロボットにさまざまな作業をさせる場合、個々の作業に必要な知識の多くは、言葉では十分に表現できないはずで。そこで問題になるのが、「物理世界を非言語的に理解する」ということです。人間は、物理世界をある程度理解しています。人間の理解には、言語化できる知識と、言葉で表現することに適していない知識の両方があるのだと思います。現在の AI は、主にウェブ上のデータから、言語化された知識を取り出して学習しているにすぎません。そのため、人間が物理世界を非言語的にどのように理解しているのかは、まだよく分かっておらず、どのように取り組むべきかも明確ではありません。一部については、専用 AI のアプローチを用い、人間がラベルを付けることで扱えると思います。ただし、言語化可能な形式だけでなく、暗黙

知や身体知をどのように扱うかという問題は残ります。これは、まさにフィジカル AI に関わる問題です。こうした知識は、LLM ではなくかな手が出ない領域です。この問題をどのように扱うかは、研究者も当然意識しています。



スライド 46

その一つの考え方として期待されているのが、「世界モデル」です。先ほど示したエージェントの模式図で言えば、判断の段階に、将来を予測する能力を与えようという考え方です。自動運転の E2E のように、行動を直接出力するのではなく、自らが行動を取ったときに、世界がどのように変化するかを予測します。その予測ができれば、望ましい将来の状態に近づくように、行動を決められるのではないかと思います。これが、物理世界を非言語的に理解する一つの形ではないかと思います。ただし、現状で試されているのは、いわば「おもちゃの問題」のような単純化された実験に限られています。まだ現実世界で利用できる段階には達していません。



スライド 47

世界モデルに求められるのは、少し先の将来に世界がどのような状態になるかを予測することです。そのための別のアプローチとして、動画生成 AI を利用する方法があります。現在では、静止画だけでなく、動画も自由に生成できるようになっています。この能力を

利用して、物理現象の将来を予測できるのではないかと考える方です。物理的に正しい映像を生成できるのであれば、そのモデルは、物理世界の将来を予測できていると考えられます。例えば、多関節の指をどのように動かせば瓶の蓋を開けられるかを、映像として生成する例があります。「このように指を動かせば蓋が開く」という将来の状態を映像として生成し、そこから必要な動作を導き出せる可能性があります。このように、動画生成によって世界モデルを構築しようとするアプローチもあります。ただし、現時点では、生成された映像を具体的にどのようにロボットの制御へ結び付けるかは、まだよく分かっていません。



スライド 48

## 9 まとめ

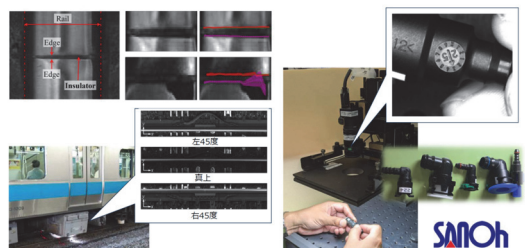
冒頭で説明したように、フィジカル AI は今、大変なブームになっています。人の作業を代替するロボットが実現すれば、その市場は極めて大きいため、大いに期待されています。特に、言語生成 AI は、数学者が自分の職業を心配するほど成功しています。同じような飛躍がフィジカル AI にも起こるのではないかと期待することは、あながちの外れではないと思います。ただし、その期待が現実になるかどうかは分かりません。現時点で実現できていることは限定的であり、大きな壁が二つあります。一つは、マニピュレーションの学習です。人が手指を使って行っている、実際には非常に高度な制御を、ロボットにどのように学習させるかという問題です。現時点で有力な学習方法は、行動模倣です。人間が繰り返しデモンストレーションを行い、データを収集していますが、このアプローチが本当に成功するかどうかは、まだよく分かっていません。もう一つは、そもそも世界を認識し、理解するという問題です。この点についても、解決の見通しが立っているわけではありません。さらに、人のように器用に動くロボットハンドが、まだ実現していないことも忘れてはいけません。

二つの壁に共通する根本原因は、現在の主要な手段が深層学習に依存しており、その学習の非効率性から逃れられないことです。そのため、研究者の中には、全く異なる学習パラダイムがなければ、現在期待されているようなフィジカル AI は実現できないのではないかと考える人も少なくありません。言語生成 AI があれほど成功したため、同じようなアプローチでフィジカル AI も実現できるのではないかと期待されています。しかし、そう簡単ではないということです。

最後に、もう一つ付け加えておきたいことがあります。冒頭でも述べたように、これまで現実世界と対峙してきたのは、言語生成 AI 以外の、すべての AI ではないかと思えます。

専用AIはこれからも重要

- ・ レール絶縁継ぎ目の厚み推定
  - 薄くなって導通 → 信号に影響
  - 走行時撮影画像から自動検査
- ・ ロットダイヤル自動読み取り
  - トレーサビリティ目的の製品
  - ロット情報の記録



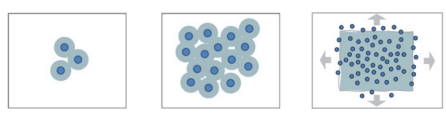
[Kasai-Okatani, IEEE-TIT2026]

59

スライド 50

まとめ

- ・ Physical AIがブーム ~ 巨大な市場への期待
  - ChatGPTモーメントがやってくる？
- ・ 期待が現実になるかは不明
  - 今出来ていることは限定的
  - 2つの壁：マニピュレーションの学習 + 物理世界の(非言語的)理解
  - 根本原因：深層学習の学習非効率性 (+人に匹敵するハンドがまだない)
  - ⇔ 言語生成AIの特殊性
  - 現実世界と対峙するのは「言語生成AI以外すべてのAI」



Situation Spaceの広さ vs. サンプリングコスト

58

スライド 49

その具体例が、専用 AI というアプローチです。ここでは、私たちが最近行った企業との共同研究の成果から、二つの例を紹介します。一つは、JR 東日本の列車の床下に取り付けられた監視・検査装置で撮影した画像を用いて、必要な検査を行う AI です。これは深層学習を使っていますが、対象とする問題に焦点を絞り、必要なニューラルネットワークの構成や、何を入力として何を出力するかを設計します。そのうえで、学習データを集めて学習するという、従来型の AI のアプローチによって実用化した例です。もう一つは、製品のロットダイヤルを自動で読み取る AI です。ロットダイヤルには年月日が記されていますが、人が読み取るのはかなり手間がかかります。これをカメラに一瞬かざすだけで、その内容を読み取ることができます。

一見すると比較的容易に実現できそうな事例ですが、対象を絞って適切に設計すれば、実際に使えるシステムを作ることができます。こうした専用 AI のアプローチは、今後も重要な役割を果たし続けるのではないかと思います。これも、非常に広く捉えれば、フィジカル AI の一つではないかと私は考えています。

## 講演者紹介



岡谷 貴之 (おかたに たかゆき)

1999 年 東京大学計数工学専攻博士課程修了

1999 年～ 東北大学大学院情報科学研究科, 2013 年から現職

2016 年～ 理化学研究所革新知能統合研究センター兼務