第407回 NRIメディアフォーラム

汎用人工知能

チーフストラテジスト 長谷佳明

デジタルトラスト基盤事業本部
IT基盤技術戦略室

2026年3月24日

NRI

Envision the value,
Empower the change

汎用人工知能とは

汎用人工知能とは

■汎用人工知能（**AGI、Artificial General Intelligence**）とは、人間と**同等の知能**と**自律性**を発揮する人工知能。**知的作業のほとんどを代替**できると想定されている

■**高度な学習能力（メタ学習）と未知の課題**にも適応可能な**一般化能力のある推論力。人が与えた目的**から**計画を立案**し、**意思決定可能な自律性**を有するシステム

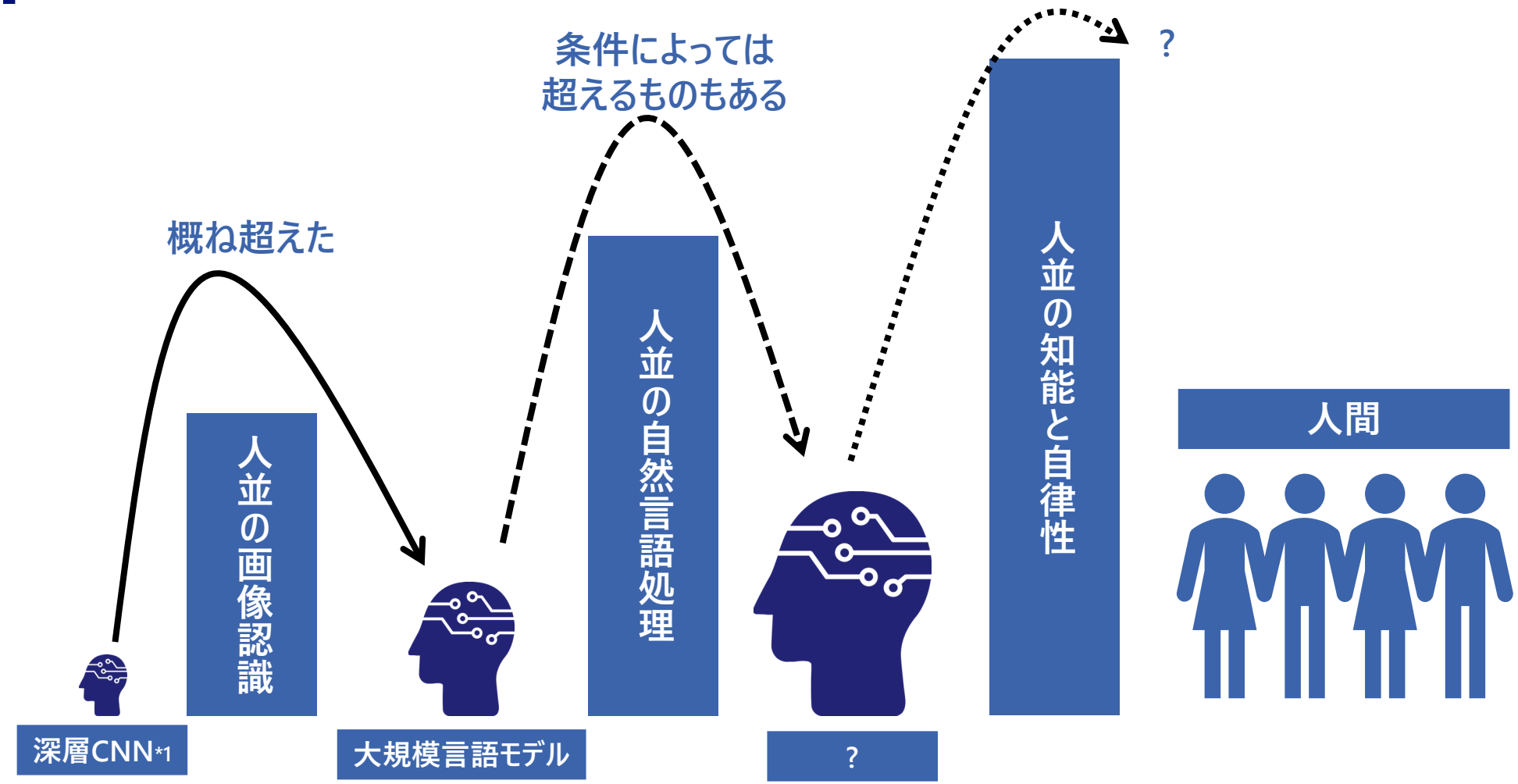
汎用人工知能

- ✓ **知能**
  - ✓ **学習**：知識やスキルを獲得する**高度な学習能力**
  - ✓ **推論**：**未知、既知**に関わらず対応可能な**推論能力**
- ✓ **自律性**
  - ✓ **行動計画**：目標達成のための**計画策定（階層的思考）**
  - ✓ **意思決定**：**行動の選択や評価**

✓ 汎用人工知能の実現は、機械の特性である**無限のスタミナ**、**膨大な記憶**から、人間を越える**超知能（Artificial Super Intelligence、ASI）の誕生も**意味し、**社会的インパクトが大きい**

汎用人工知能とは

①なぜ、今、汎用人工知能が注目されているのか？

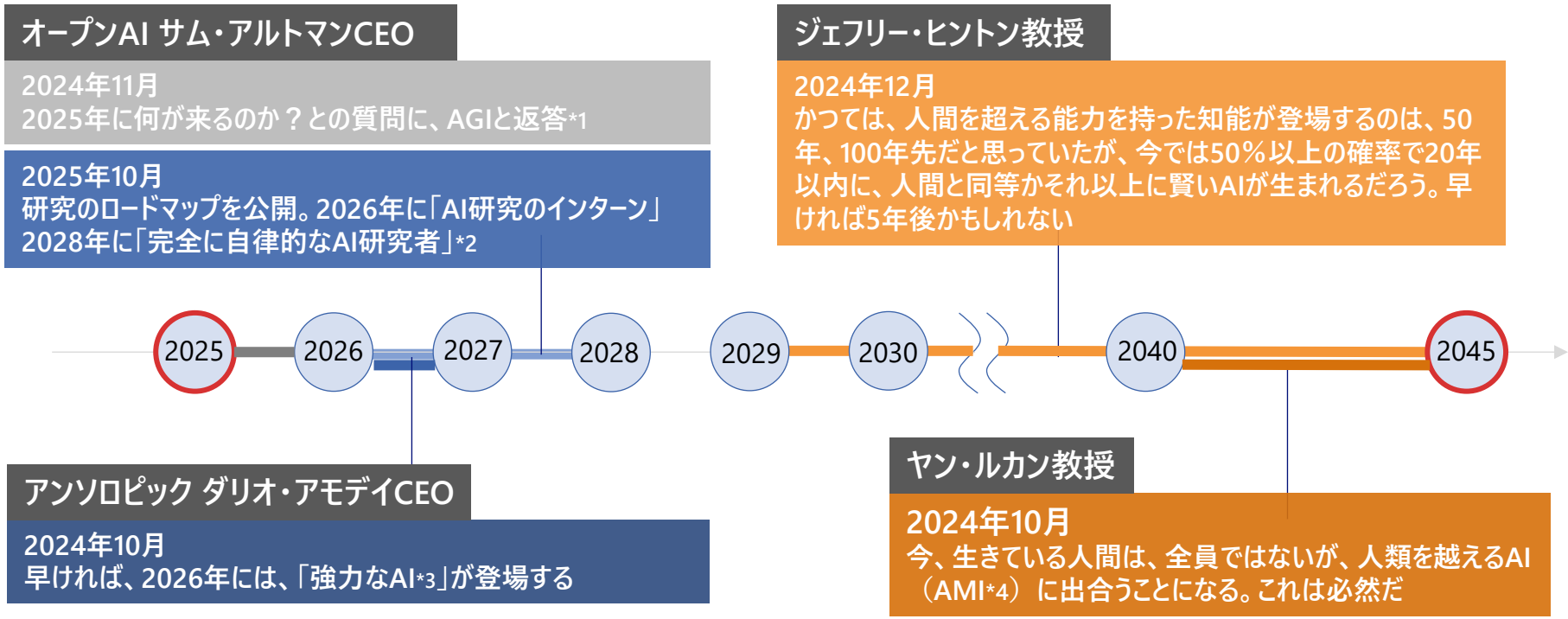


✓ **人並の画像処理に続き、人並の自然言語処理の実現**に道が開け、状況によっては人を上回るケースを確認（例：プログラミング）現在の技術の延長線上に**人並みの知能と自律性（＝汎用人工知能）**が実現するとの期待が高まっている

*1 CNN Convolutional Neural Network

②なぜ、今、汎用人工知能が注目されているのか？

■テック企業のCEOをはじめ、著名研究者から、**汎用人工知能に関する言及**が相次いでいる



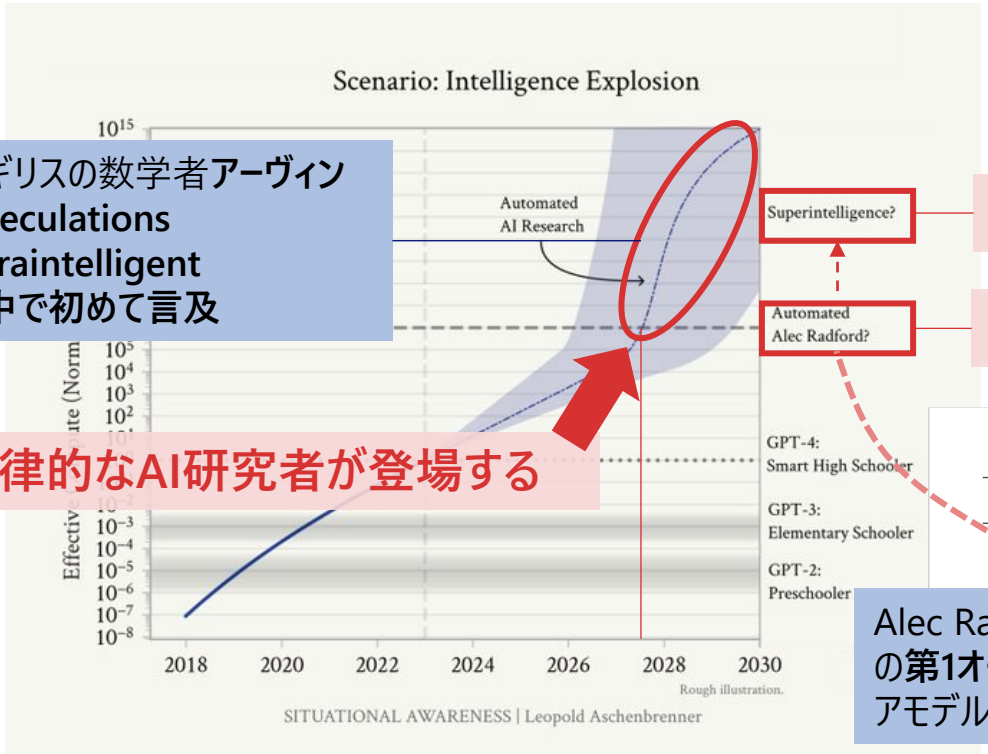
✓ 汎用人工知能は**SF（≒空想）**と思われてきたが、現実問題として真剣に捉えられ始めている
✓ ただし、実現時期には、**20年以上の差**があり、意図するものも**微妙に異なる**点に注意

*1 米国のベンチャーキャピタル Yコンビネーターとの動画「How To Build The Future: Sam Altman」の中での発言
*2 OpenAIの公式チャネル「Sam, Jakub, and Wojciech on the future of OpenAI with audience Q&A」の中での発言
*3 アモデイCEOのエッセイ「Machines of Loving Grace」、AGIという表現を好まず、あえて"Powerful-AI（強力なAI）"と表現
*4 コロンビア大学での講演「"How Could Machines Reach Human-Level Intelligence?"」の中での発言、Advanced Machine Intelligence（AMI）と表現

知能爆発 AIの急激な進化の始まりは、"人間"ではないだろう

■2024年6月、当時、オープンAIのスーパーアライメントチームのメンバーであった**レオポルド・アッシェンブレナー**は、**超知能**に関する考察を「**SITUATIONAL AWARENESS The Decade Ahead（SGIに対する状況認識 次の10年）**」にまとめ発表

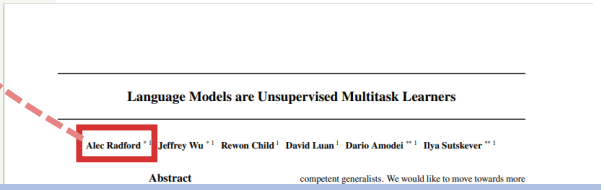
「**知能爆発**」については、イギリスの数学者**アーヴィング・ジョン・グッドが論文**「Speculations Concerning the First Ultraintelligent Machine」（1965年）の中で初めて言及



起こる事

引き金

2027年には、自律的なAI研究者が登場する


Language Models are Unsupervised Multitask Learners

Alec Radford Jeffrey Wu Rewon Child David Luan Dario Amodei Ilya Sutskever

Abstract

competent generalists. We would like to move towards more

Alec Radfordとは、GPT-1、GPT-2の論文の**第1オーサー**であり、オープンAIのフロンティアモデルの主要な研究者の名前

✓ "**Automated AI Researcher（自律的なAI研究者）**"が引き金となり、急激なAIの進化が起こり、汎用人工知能を越える**超知能（Artificial Super Intelligence）**が登場する

出所) https://situational-awareness.ai/wp-content/uploads/2024/06/situationalawareness.pdf
https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

汎用人工知能とは

そのアーキテクチャは、LLMに似るが、異なる

■2024年10月、アンソロピック ダリオ・アモデイCEOは、「Machines of Loving Grace（愛する恵みの機械）」と題したエッセイを公開


Dario Amodei

Machines o
Loving Grac

How AI Could Transform the World fo

October 2024

I think and talk a lot about the risks of p
the CEO of, Anthropic, does a lot of res
risks. Because of this, people sometimes
pessimist or "doomer" who thinks AI will be mostly bad or dangerous. I
don't think that at all. In fact, one of my main reasons for focusing on
a fundamentally positive future. I think that most people are
underestimating just
think most people are

主な主張

- **強力なAIのアーキテクチャは、LLMに似ているが異なる**
- **知能の観点では、ノーベル賞受賞者よりも賢い（未解決の問題を証明できる）**
- **インターフェースを持つ。話すだけでなく、コンピュータを操作したり、ビデオを観たり、動画を作り出したりする**
- **長期間のタスクを自律的に計画し、遂行する（長期プロジェクト）**
- **100万のAIのコピーが誕生し、それはデータセンターに住まう（「データセンターに天才の国」ができる？）**

早ければ、2026年には、「強力なAI」が登場する

I find AGI to be an imprecise term that has gathered a lot of sci-fi baggage and hype. I prefer "powerful AI" or "Expert-Level Science and Engineering" which get at what I mean without the hype.

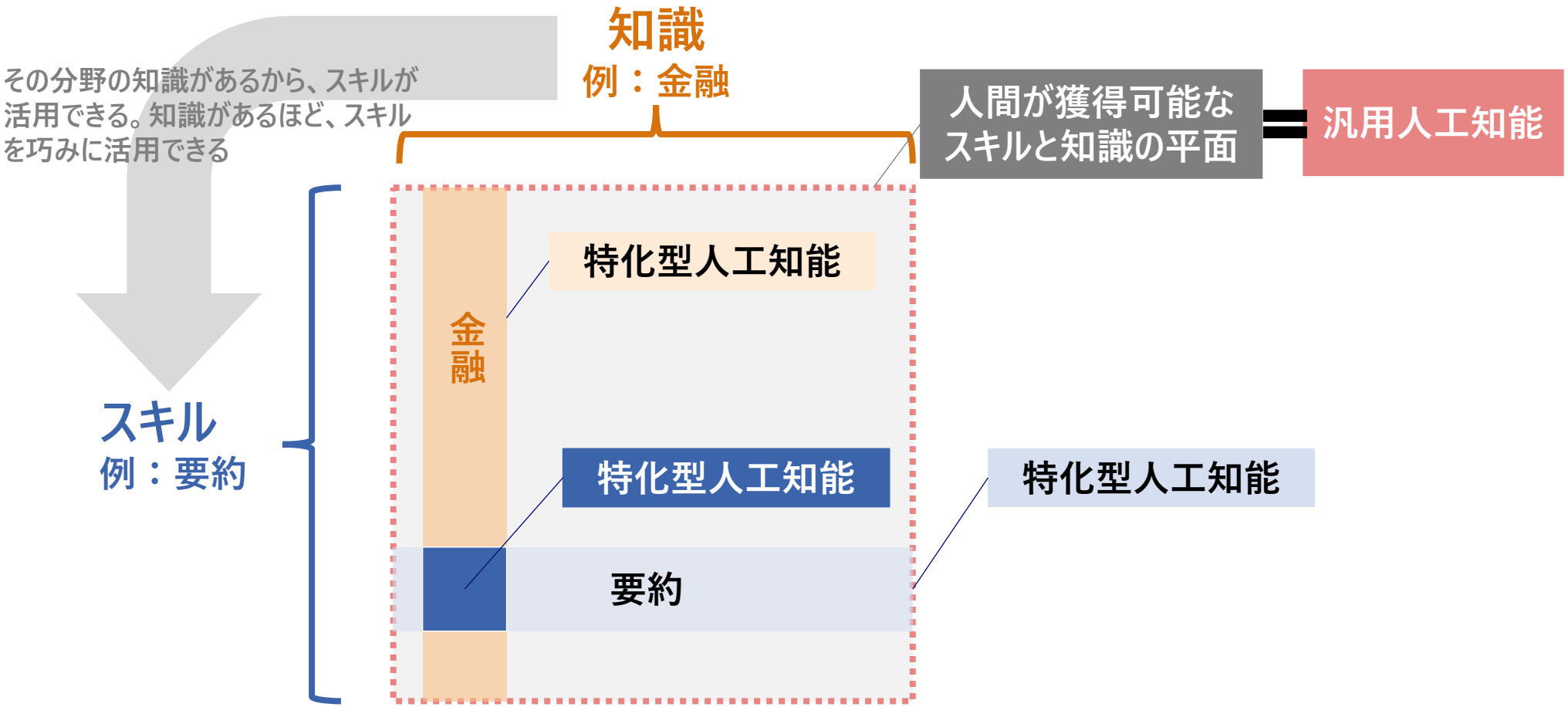
SFでは誇大広告を集めた不正確な用語のため、新たな用語として"強力なAI"を定義（AGIは使い古され、解釈が多すぎて、誤解を招くことを懸念）

✓「強力なAI」の実現に向け、**"欠けたパーツ"**を実装し、くみ上げるなど、高度なモデルに留まらない、**"システム"への進化を実践（⇒その成果が、Claude Code、CoWork）**

出所）https://darioamodei.com/essay/machines-of-loving-grace

汎用人工知能の特徴

特化型人工知能と汎用人工知能の違い（知識やスキルの範囲から生じる汎用性）



- ✓ 特化型人工知能は、スキルや知識を**特定の領域に絞り学習し**、実用性を引き出している
- ✓ 汎用人工知能は、領域に特化せず、**あらゆるスキルや知識を対象とする**
  - ⇒**知識×スキルの面積**が、AIの汎用性にも関係する
  - ⇒**タスク**を**自律的**に遂行するには、**スキルを"どう使うか"の知識がAIの中に必要である**

特化型人工知能と汎用人工知能の違い（スキルの活用に関する知識の外部化と内部化）

特化型人工知能

タスク：動物か判断する

知識：システム
画像認識を呼び出す

\+

スキル：画像認識
動物の写真を学習

エンジニアがスキルの
活用法に関する**知識**
を**システム**に変換する

スキルを活用する知識は外部化されている
スキルを活用する知識は外

汎用人工知能

タスク：動物か判断する

知識：動物とは
特徴、概念・・・
どのようなスキルで
解決できるか自ら
選択する

\+ **スキル：画像認識**
動物の写真を学習

\+ **スキル：動画認識**
動物の動きを学習

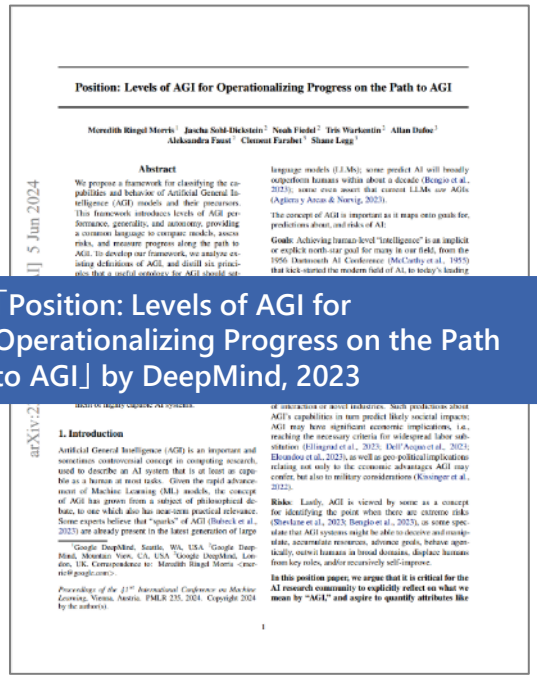
\+ **スキル：音声認識**
動物の発する音を学習

スキルを活用する知識は内部化されている
スキルを活用する知識もAIが持ち、判断する

✓ スキルを活用する主体が**外部（システム）**にあるか、**内部（AIが持つ）**か

特化型人工知能と汎用人工知能の違い（自律性のレベル）

■DeepMindは、2023年11月 AGIに向けた論文「Position: Levels of AGI for Operationalizing Progress on the Path to AGI」を公開


Position: Levels of AGI for Operationalizing Progress on the Path to AGI

「Position: Levels of AGI for Operationalizing Progress on the Path to AGI」by DeepMind, 2023

AGIに向けた自律性のレベル定義

	レベル	状態	
汎用人工知能	5	AIがすべてを自律的に遂行する （例：今は、ない）	完全自律
特化型人工知能	4	AIがタスクを主に実行し、人は判断やフィードバックのみ （例：タンパク質の構造予測のレベル）	半自律
	3	人とAIが協力して、対話的に目的を達成する （例：チェスAIのレベル）	
	2	人が指示し、AIがタスクを実行する （例：要約、コード生成のレベル）	
	1	人が全てのタスクを制御しタスクの一部をAIが自動化 （例：検索エンジン、文法チェックツールのレベル）	
それ以外	0	人が全て作業するシステム （例：テキストエディタのレベル）	手動/依存

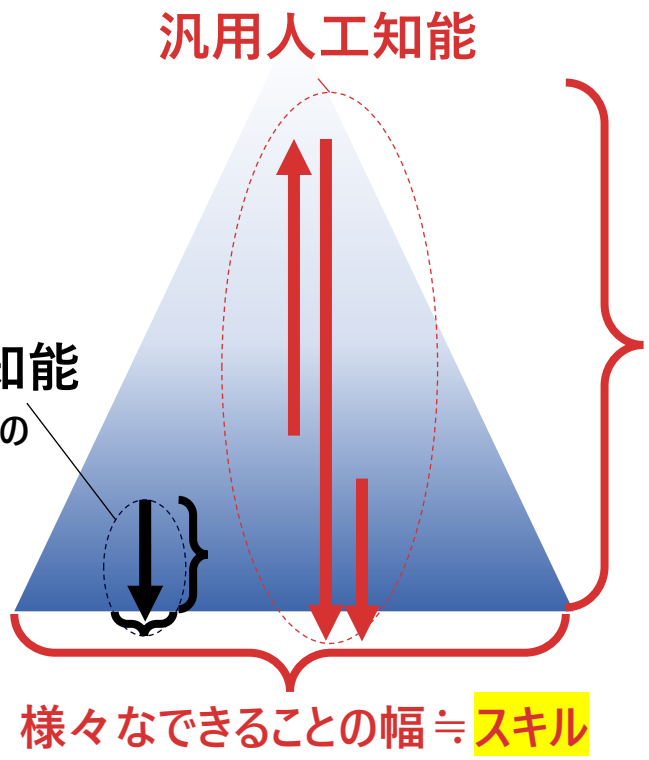
✓ 汎用人工知能と特化型人工知能の間には、レベル5とレベル4の間に”深い谷”がある

出所) https://arxiv.org/abs/2311.02462

特化型人工知能と汎用人工知能の違い（一般化能力とスキル、そして自律性）

抽象＝概念、目的

具体＝事実、行動



会議調整の成功とは、・・・

会議調整をお願いします

計画を作成（タスクへ）
参加者の人数把握、会議室の予約、・・・

- ✓ 特化型人工知能は**限られたスキルを持ち**、一般化能力も乏しく、**自律性も限定的**
- ✓ 汎用人工知能は**様々なスキルを遂行可能で**、優れた一般化能力を用いて依頼内容を**抽象化**、**取るべき行動のサブゴールを推論**。**計画に落とし込む**など、**高度な自律性をもつ**

汎用人工知能に関連する技術や論点

オルガノイドコンピューティングの勃興 “デジタル”から”モータル”へのシフトが、はじまるだろう

■トロント大学のジェフリー・ヒントン教授は、2024年2月、トロントの非営利人工知能研究機関Vector Instituteが開催するイベントで「Will Digital Intelligence Replace Biological Intelligence?」と題した講演を実施

主な主張

- **脳オルガノイドコンピューティングの勃興**
  デジタルから**モータル（死＝生物）**へ。あえて、デジタルの**不滅性を放棄**し、知識を継承できない代わりに、極めて小さい消費電力で、**安価に成長するハードウェア（脳細胞）**によりAIは実現すべきではないか
- **ただし、維持管理装置に課題**
  ピンの頭ほど（5万個のニューロンの集合体）で集約できるほど、**ニューロンの本体は小さい**が、それを維持、保管するための装置は、**一部屋を埋め尽くすほどの機材**になる場合がある
- **脳オルガノイド版の誤差伝搬法の開発も必要**
  誤差伝搬の実現が難しい。なぜなら、それぞれの細胞間のネットワークを**現在の技術水準では把握できない**し、制御するための手法が確立していないためである（人工ニューラルネットワークのように、順方向の計算に関する正確なモデルがモータルにはない）

✓ **ハードウェアが限界**に達している。“高価な”ハードウェアである**GPUに頼るAIは誤りだ**

トランスフォーマーでは先に進めない、実世界を学ぶには、別のアーキテクチャが必要だ

■ニューヨーク大学のヤン・ルカン教授は、2024年10月、コロンビア大学にて「"How Could Machines Reach Human-Level Intelligence?"」と題した講演を実施

主な主張

- **物理法則をはじめとした、常識への理解不足が問題だ**
  今のAIは、人間どころか、**動物以下**である。それは、世界がどのように機能しているのか、**法則を理解していないため**である
- **トランスフォーマーでは、先に行けない**
  トランスフォーマーは、文章を学習できても、**実世界の予測には不向きだ**。人間は、モノが倒れる際、詳細な状況を予測するのではなく、モノが落ちる、倒れる、など**抽象化した状況を予測**している
- **コグニティブアーキテクチャが必要である**
  センサーからの入力を受け付け、抽象空間で予測する**世界モデル**など、複数のニューラルネットワークを接続した**コグニティブアーキテクチャ**なくして、進化はない
- **JEPAが有望なモデルではないか**
  Joint Embedding Predictive Architecture （JEPA） （**非生成系モデル**）
- **AIは知のインフラになる**

✓ 現実世界の**抽象化**こそ最重要テーマ。言語モデルは抽象化したデータを扱っているに過ぎない

新たなスケーリング法則を探求しなければならない、"原点回帰"が必要ではないか

■オープンAIの共同設立者であったイリヤ・サツキバー氏は、2024年12月、AIに関するトップの国際学会であるNeurIPSで「Sequence to sequence learning with neural networks: what a decade」と題した講演を実施

主な主張

- **事前学習（だけで進化する）時代が終わる**
  GPUをはじめとしたインフラはなおも進化すると思われるが、データが枯渇する（"ネットは1つしかない"）ため、事前学習で性能が向上し続ける時代は**終焉を迎える**だろう
- **生物から学ぶ（"原点回帰"）**
  かつて、ニューラルネットワークが、**生物の脳の構造**から触発されたように、いま改めて、生物、具体的には**人間の特異性**から次を考える必要がある
- **人間の体に眠る、未知のスケーリング法則**
  生物の**体と脳のサイズ**には、**相関関係がある**が、人間を含む人類は、その関係から**外れている。**ここに何か理由があり、そこに次の進化のヒントがあるのではないか（⇒**身体性と脳**の特異な関係？）
- **世界モデルと意識**
  意識は世界モデルの一部であり、世界モデルは**機械に意識を芽生えさせるだろう**

✓ ニューラルネットワークは**生物に学んだ。**次は、**人間の特異性**に学ぶ時が来ている

汎用人工知能に関連する技術や論点

AI研究の権威の考える、汎用人工知能、超知能に対する考え方や論点

	ジェフリー・ヒントン	ヤン・ルカン	イリヤ・サツキバー
AGIや強力なAIは誕生するか	○	○	○
いつか	5年から20年*1	さほど遠くない未来*2	時期は言及せず
主な課題	**モデルとハードウェア** デジタルの限界、非効率性	**モデルとデータ** 実世界でうまく動作しないこと	**モデルとデータ** データ不足、学習の非効率性
解決策や重要技術	**脳オルガノイド** あえて、"モータル"へ 高価なGPUから 安価な成長する細胞へ	**真の世界モデル** Joint Embedding Predictive Architecture （JEPA） あえて抽象化し予測する	**新たなスケーリング則** ヒトの脳に生じているスケーリング測 より少ないデータから学ぶモデル
その他	• 脳オルガノイドコンピューティングには、**誤差伝搬にかわる技術**が必要。一方で数兆のモデルで、人間の100兆のモデルよりも桁違いに大量の知識を学習できるなら、誤差伝搬法の**情報圧縮効率は驚異的**だ • 目的の源泉は、生存欲求ではなく、**好奇心**である • AIに**意識（主観的体験）**は芽生えている。意識は、皆が考えるほど、特別で神秘的なものではない • 超知能は個人のように捉える風潮があるが、実際には、**コミュニティ**のようなものとなるだろう • **オープンモデルには強く反対する**	• 今のAIは、**世界の法則を理解していないため、**超知能からほど遠い。また、「**世界モデル（文章からではない）**」「**一般常識**」「**記憶、論理的思考、タスク分解能力や論理的思考**」が欠けている • 4歳児は視覚情報だけでも、現在のLLMの**約50倍**のデータ学習している • AIが組み込まれたモノのネットワークは、現在のインターネットを越え、データは、AIを通じて、集約、還元される（**知のインフラ**） • AIを**Object-Driven AI**とすれば目的にハードワイヤされ人間の指示が優先される • 遺伝子の**8Mbyte**に何があるか • **オープンモデルを強く支持**	• AIは、人間と比べ驚くべき程、賢い時もあれば、驚くべき程、賢くないときもある。それは、「**自律性**」「**論理的思考を含む推論能力**」「**理解力**」が欠けているから • 従来のAIは**システム1（認識）であり**、予測できるAIであった。しかし、今後は**システム2（論理）が高度化する。**その副作用として**予測不可能が高まる**（これは新たな課題である） • AIの幻覚に関しては、**AI自身が幻覚を認識し、自己修正する**ような機能が、論理的推論が高度化すれば可能になると思う • **意識は、世界モデルの一部。**世界モデルの完成は、AIに意識を持たせることになる

*1 読売新聞からのインタビュー「人間以上に賢いＡＩ、早ければ５年後」「ＡＩの脅威は単なるＳＦではない」…ノーベル物理学賞・ヒントン名誉教授が語る未来」（2024年12月4日）「かつては、人間を超える能力を持った知能が登場するのは、５０年、１００年先だと思っていたが、今では５０％以上の確率で２０年以内に、人間と同等かそれ以上に賢いＡＩが生まれるだろう。早ければ５年後かもしれない」
*2 今生きている人の一部は体験する

課題と展望

課題と展望

汎用人工知能の実現がもたらす社会的影響への対策なくして実現はありえない

■2025年1月、モントリオール大学のヨシュア・ベンジオ教授をはじめとした30カ国、94名からなる研究者らは、汎用人工知能を念頭にInternational AI Safety Report 2025を公開

想定される社会的影響

経済的視点：格差の拡大、外部不経済

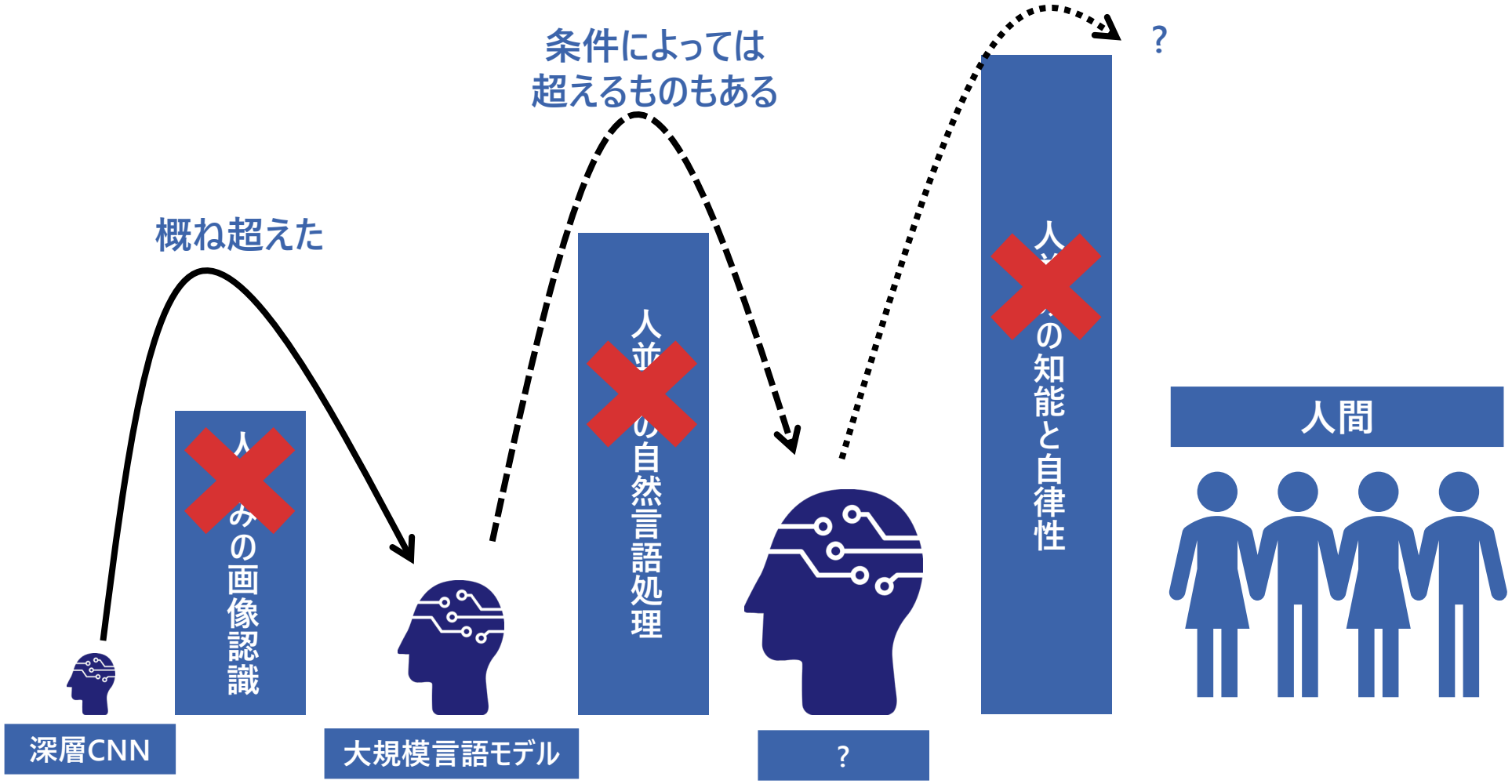
- **短期的には格差の拡大を招く**
  汎用人工知能の登場により、知的労働を中心に代替が急激に進み、**汎用人工知能を開発した企業**や、**AIを使いこなす企業**に**富が集中**する。なお悪いことに、その開発は**欧米**や**中国**に偏っている
- **外部不経済**
  汎用人工知能を開発した企業が**経済的利益を享受**する一方で、**暴走時の人類全体への影響**など負の側面への対応が**考慮されていないか、不十分である。**

倫理的視点：倫理的ジレンマ、倫理の主体の不在

- **人の価値観とAIの価値観との対立**
  AIが意思決定に関わるようになった際に、**価値観の衝突（倫理的ジレンマ）**が起こりえる。一方で、社会の多くの問題は、**公平性と他の目標とのトレードオフ**であり、尺度とは**相対的**であり、絶対的な倫理観はない。**社会全体の倫理観を学習させることは困難**である
- **倫理の主体の不在**
  誰の倫理観を学習させるかの方針、主体が現時点で**曖昧**である

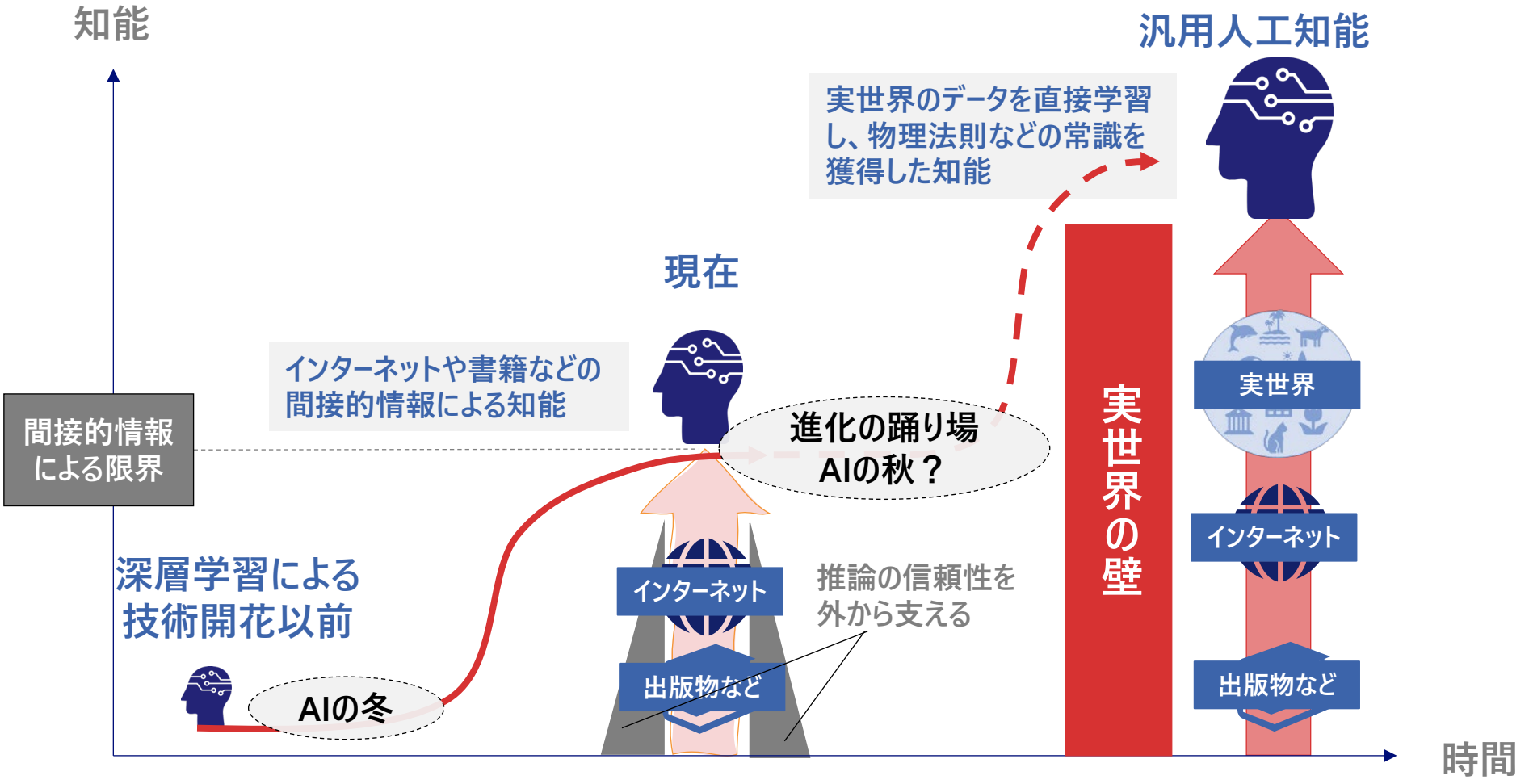
✓ 現在のAIは人類の統制下にあるが、将来に渡り安全に制御できる保証はない
✓ 暴走のリスクは現世代に留まらず、未来の世代にもリスクを背負わせることになる
✓ 2026年2月に公開された2026年版では、リスクに対する証拠がないことで、判断が遅れ、取返しがつかなくなる「証拠のジレンマ」に言及

現在のAIは、"本当はどのような状況"なのか？



- ✓ 人を越えるとも言われる画像認識でさえも、現状は"**特定の条件下**"などの**制約付き**
- ✓ **知識を活用する前提**となる**物理法則などの「常識」が欠けている**

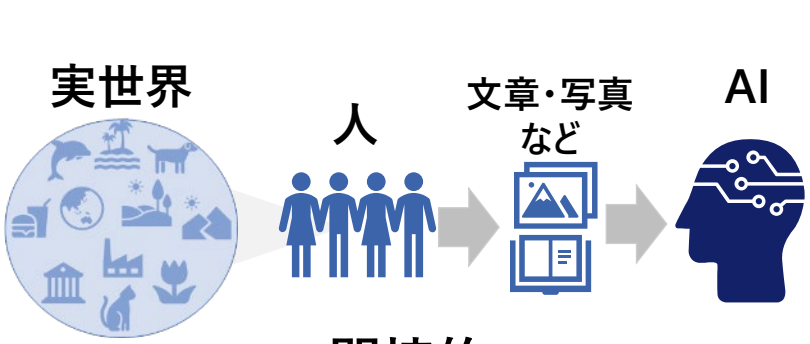
現在のAIは、自律性が開花しはじめているが、"知能"に関し、"進化の踊り場"にある



- ✓ 現在のAIは、インターネットなどの膨大な**間接的情報から知能が急激に高度化**
- ✓ ただし、間接的情報では、物理法則などの常識が獲得できず、**実世界の壁に直面**している

実世界を"抽象化"する機能をデジタルで"再発明"する （ポスト・トランスフォーマー）

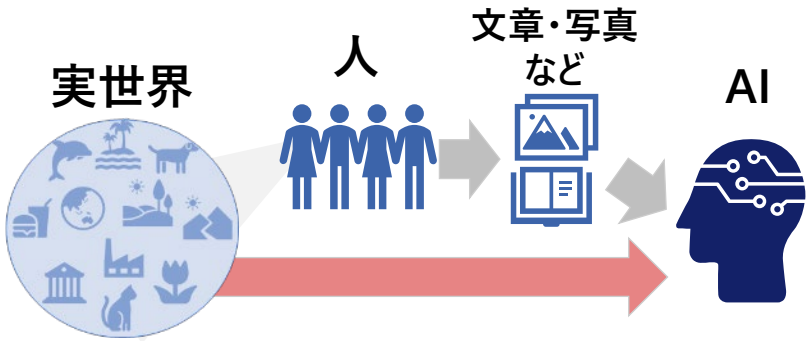
従来のAI開発
「人工物（人の"生成物"）」のみ
トランスフォーマーモデル



間接的

- ✓ 人がエンコーダーとなり、実世界を抽象化
- ✓ 人が抽象化した文章や画像を通じて学習
- ✓ 実世界のデータを直接取り扱うのが苦手
- ✓ AIのための実世界の単純化が必須（特化）

今後のAI開発
「人工物（人の"生成物"）」と「実世界」
ポスト・トランスフォーマーモデル



間接的＋直接的

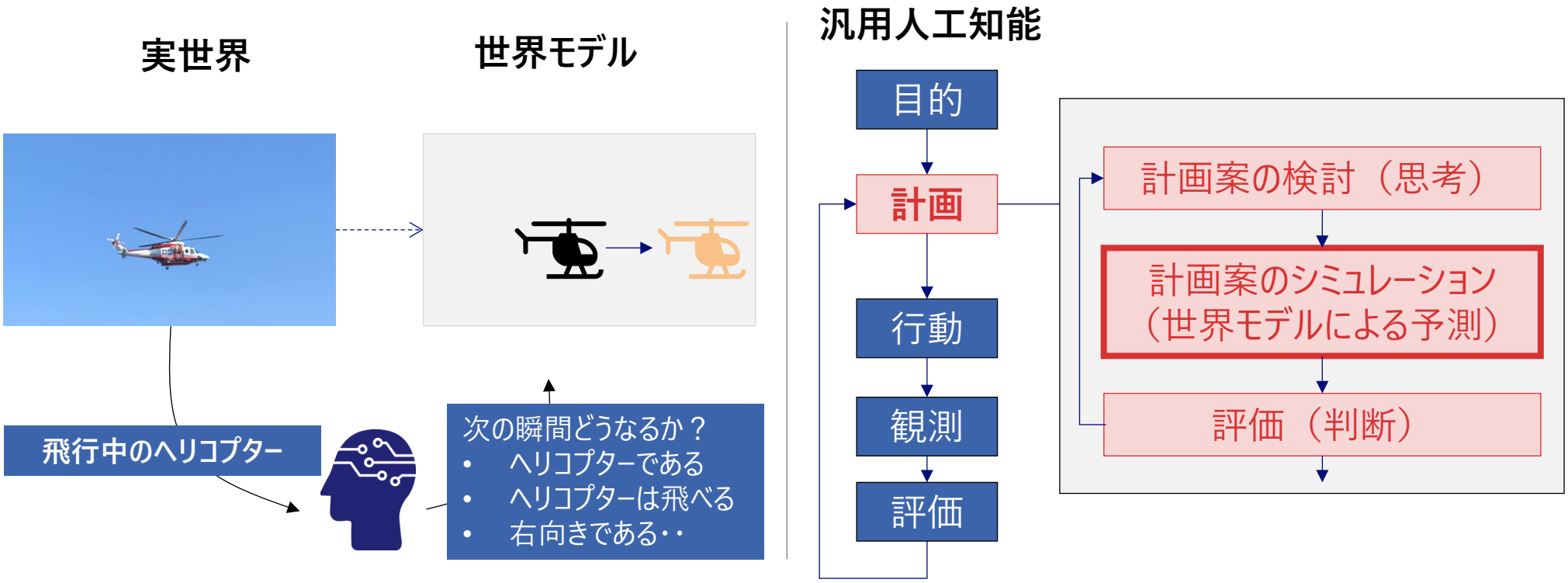
- ✓ 人とAIがエンコーダーとなり、実世界を抽象化
- ✓ 文章や画像、実世界の経験（データ）から学習
- ✓ 実世界のデータを直接取り扱うことが可能
- ✓ AIのための実世界の単純化が不要（汎用）

- ✓ これまでのAI開発は、人の作り出したデータに依存 （＝「抽象化は人任せ」）
- ✓ 実世界に適用できる汎用的なAIとなるためには、抽象化もAIが担わなければならない

AIの次の進化の鍵となる技術、世界モデルとは

■世界モデルとは、**実世界（外界）**の状況を抽象化するなどして、**再現した世界（内界）**

■AIに**想像力を与えるもの**といえ、**実世界の予測**のための**シミュレーション環境**として活用する



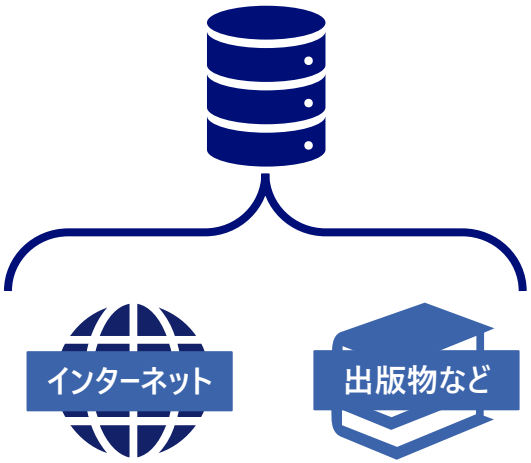
✓ 汎用人工知能は、世界モデルを活用し、**計画案を事前に評価。**実世界で直面する様々な課題に対して、考えられる中で**最善な行動**が可能となる

✓ 世界モデルにより、汎用人工知能の**計画能力が飛躍的に向上する**

世界モデルの獲得に必要なデータ量とは、どれほどのものになりうるのか？

■人並みの学習が必要とすると、現在の大規模言語モデルの50倍ものデータが必要になる

大規模言語モデル
10兆トークン（1トークン≒2バイト）
＝2.0×10の13乗バイト



50倍

満4歳児
1.6万時間×200万バイト×10
=1.1×10の15乗バイト

10回/秒（10Hz）

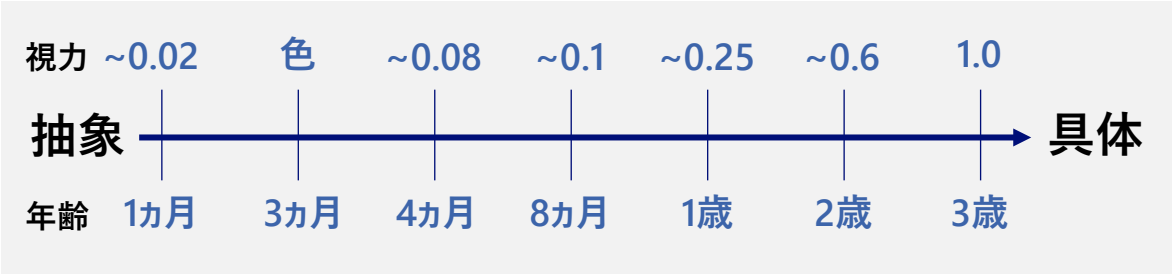
100万本×2

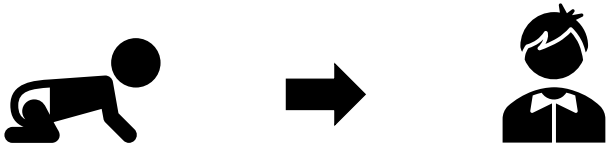
1.6万時間

✓ 大量のデータを学習できるモデルと、大規模な計算が可能な基盤（インフラ）が必要

世界モデルは、”人”を模倣し、”人の力”を借りることで、効果的に学習できうる

人の模倣
成長に伴う人の視力の推移





人の力を借りる
人の視線の先のデータ

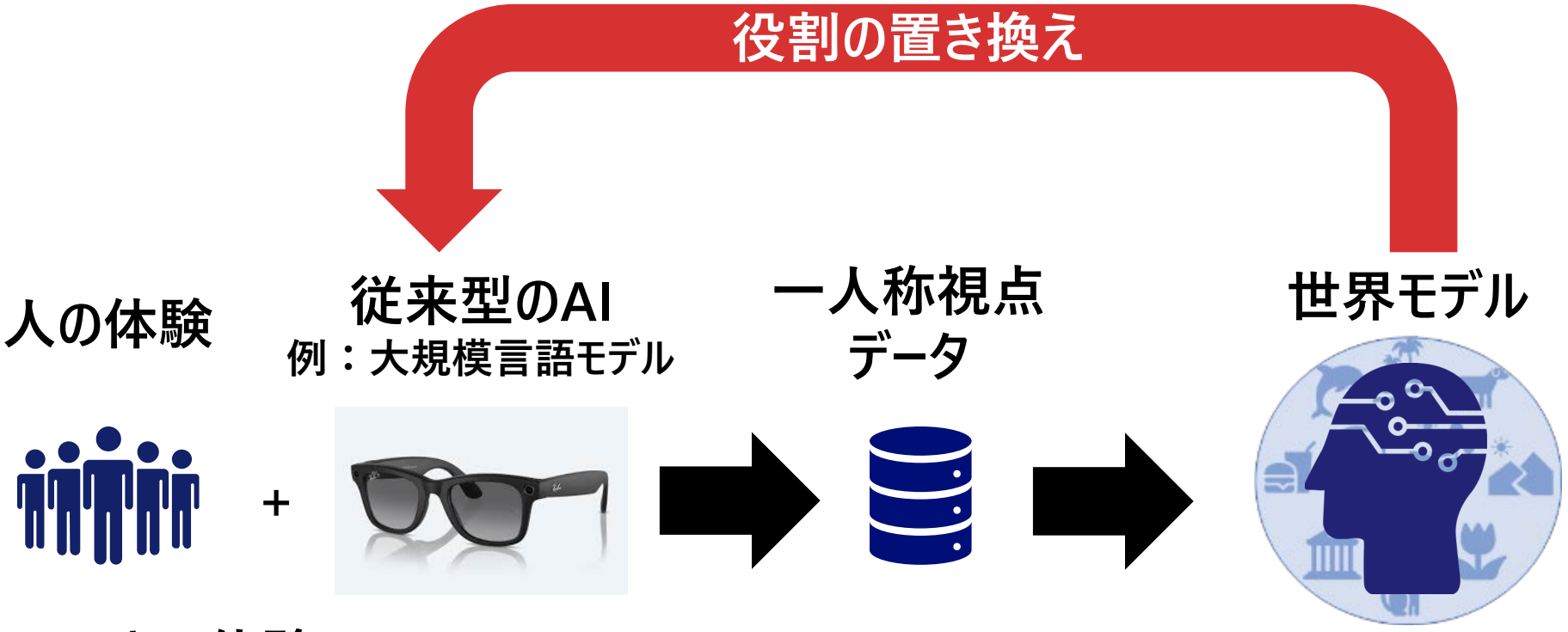
一人称視点
デジタルの仮想的な視覚

- ✓ 人の視覚の成長のメカニズムを模倣し、抽象性の高い状態を経て、具体性の高い状態を学習する
  ⇒ **世界モデルのカリキュラムラーニング**
- ✓ 何に注目すべきか人の視線を教師データとして学習する
  ⇒ **視線でラベル付けした動画データ**

- ✓ 人の**能力獲得メカニズムの模倣**、**人の視線を含むデータ**により、漫然と動画を学習するよりも、少ないデータによって、**実世界を抽象化する機能**を獲得できる可能性がある

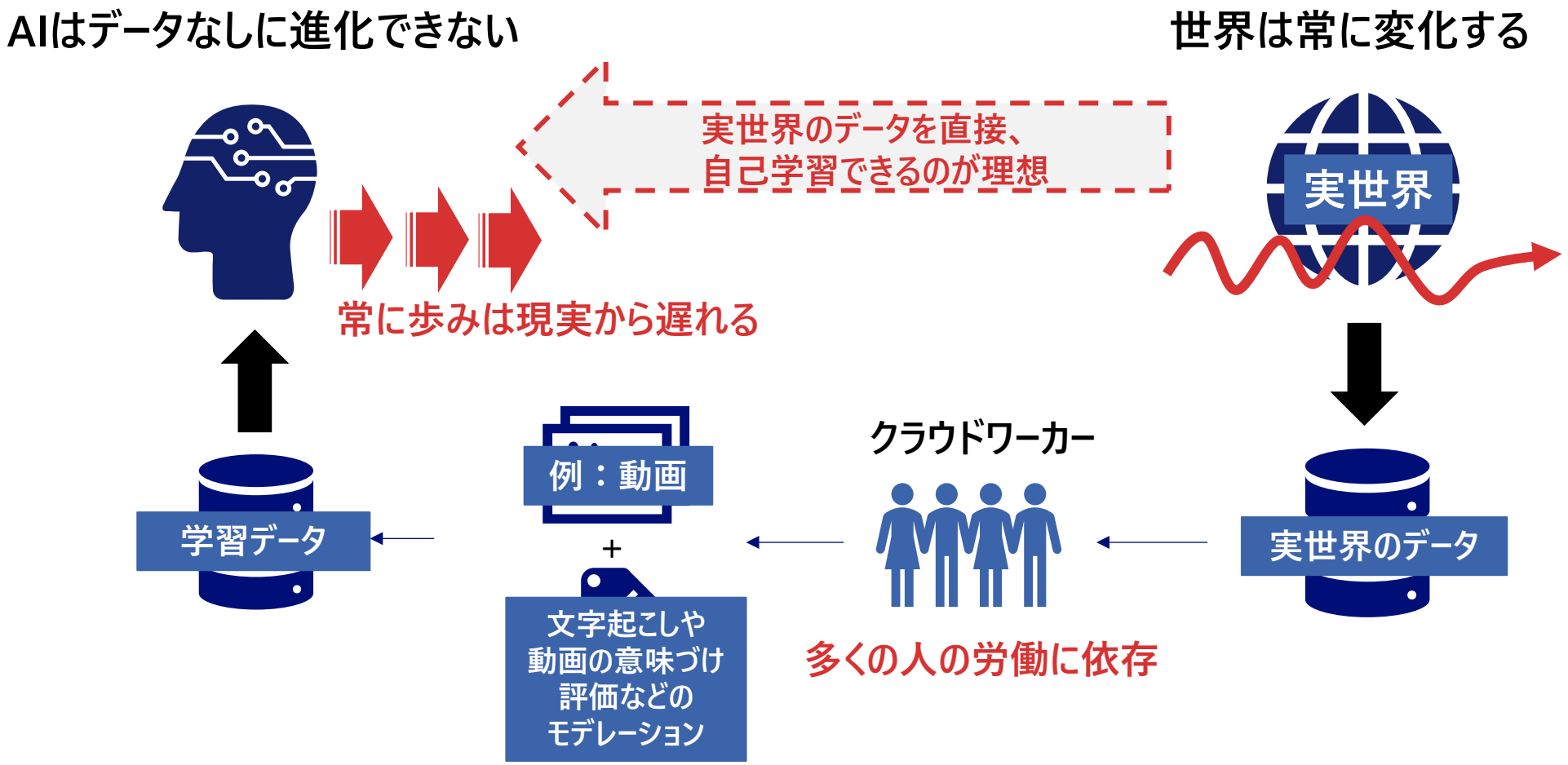
出所) https://www.meta.com/jp/ai-glasses/shop-all/

AIグラスは、特定の人のコンテキストに沿った一連の体験（＝データ）であり、世界モデルの学習には最適な資源となるだろう



- ✓ AIを介して、**新たな資源（＝データ）**が生み出され、**世界モデルが完成**する
- ✓ 一般常識を持つ世界モデルは、フィジカルの理解で既存モデルを圧倒し、大規模言語モデルを本来の役割である、**インタフェース（＝言葉）**と**データ（＝知識）**に**回帰させるのではないか（⇒AIのグレートリセット）**

自己学習 学習も人依存から、"脱却"しなければならない



- ✓ 自己学習とは、**AIが自らの経験から学び**、スキルや知識を獲得すること
- ✓ 新たな行動など、過去と**どれだけ異なる振舞いをしたか**を評価基準とし、**解を探索**する**新規性探索**のような手法もブレイクスルーになりうる（**人工生命研究との交差点**）

まとめ

まとめ

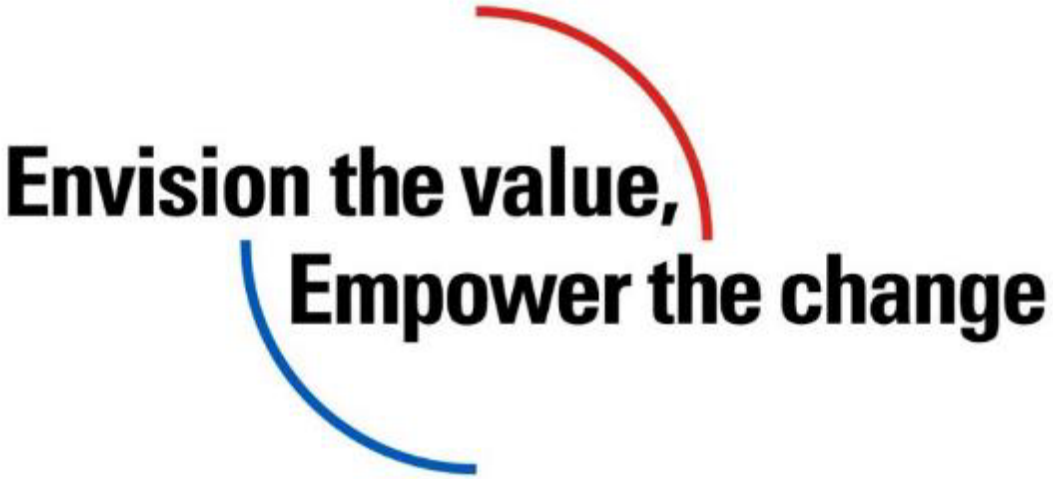
汎用人工知能の欠けた最重要ピースは世界モデル。その実現は"GPT"を超えた衝撃になる

■汎用人工知能は、**SF**でなはい
- **第一線の研究者**らも、**その可能性を真剣に**考えはじめている
- ただし、研究者により**定義に揺らぎ**があり、予測には、**5年から20年の幅**がある

■汎用人工知能は、**少なくとも3つの技術**が必要になる
- 人並の**抽象化**（「**ポスト・トランスフォーマー**」）
- 人並の**常識の獲得**（「**世界モデル**」）
- 人並の**学習**（「**自己学習**」「**新規性探索**」）

■その登場は、**ある日、突然ではない（社会が大混乱する状態にはならない）**
- **壁は何段**もあり汎用人工知能は**突如、登場せず**、技術革新の段階に応じ、徐々に**社会に変容**をもたらす。社会的な影響は結果的に大きいも、**一夜にして世界は変わらない**
- **ただし、自律的なAI研究者が引き金となり、進化が突如、加速する可能性がある**

■未来から**現在を見る**と何がいえるのか？
- 間接的情報（人工物）からの進化は**限界**にある。現在は**進化の踊り場（AIの秋）**にあり、**実世界の壁**を超えるには時間を要する（GPTでいえばブレイクスルーの兆しのあったGPT-2未満）一方で、近年の技術進化は目覚ましく、**AIを社会実装する好機でもある**


Envision the value,
Empower the change