



RIKEN TRIP

Transformative Research Innovation Platform
of RIKEN platforms

資料 4

科学技術・学術審議会
基礎研究振興部会（第18回）
令和7年7月7日

科学研究基盤モデル開発プログラム（AGIS）と 国内外の取組 — 科学向けAIによる研究の革新 —

泰地 真弘人 (Makoto Taiji)

理化学研究所

科学研究基盤モデル開発プログラム ディレクター

生命機能科学研究センター 副センター長

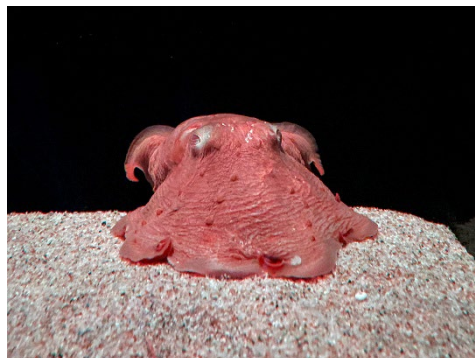
理事長補佐

なぜ科学技術でAIが大事なのか？

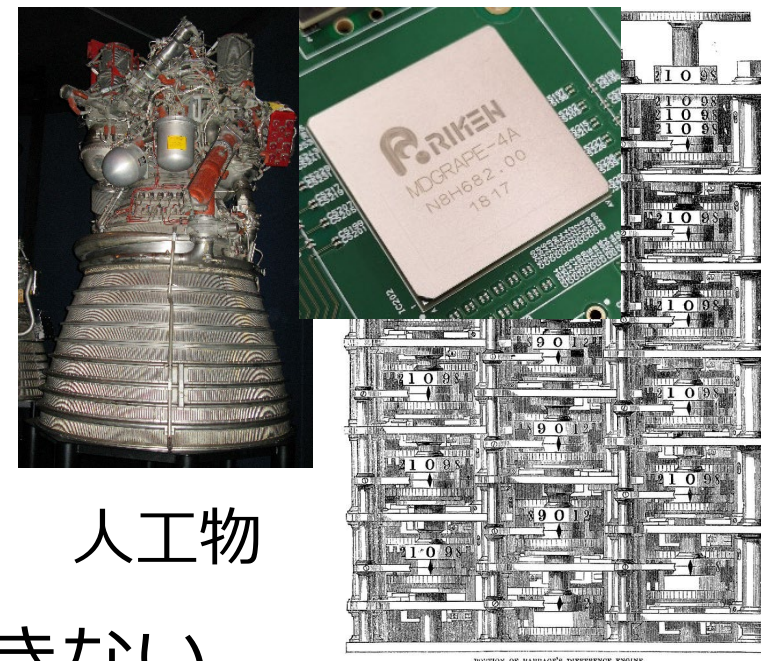
■ 科学・工学の対象が複雑になってきた



自然



生命



人工物

- 複雑なものはなかなか紙と鉛筆で理解できない。
コンピュータ上で、シミュレーションかAIを使うことがどうしても必要。
- **近年のAIの発展で、AIの適用範囲が急速に拡大**

■ 第4次産業革命＝知的労働の自動化

研究活動はその中核で、**イノベーションの源泉**

■ 大規模言語モデルの汎用性と 科学研究における有用性

AIエージェントによる研究支援・自動研究

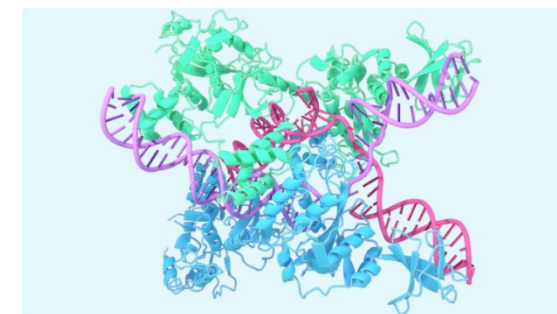
■ 科学の各ドメイン固有の基盤モデルの重要性

既にタンパク質モデル、ゲノムモデルなどで成果

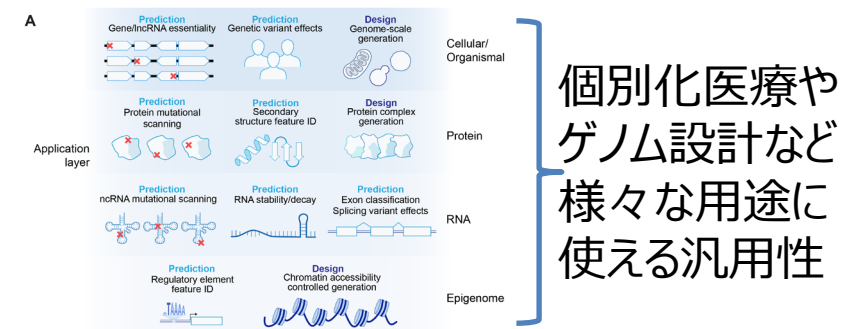
**大規模データから作成した基盤モデルは汎用性が高く、
様々な研究開発に適用可能**



Google/DeepMind AI co-scientist



AlphaFold 3 (DeepMind)



タンパク質のアミノ酸配列からの立体構造予測を可能に

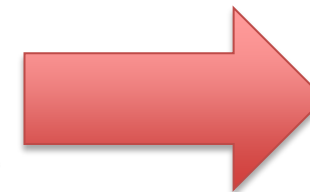
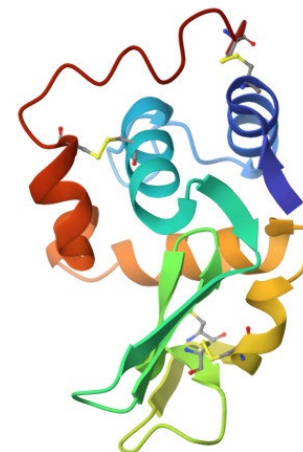
タンパク質の
アミノ酸配列

MKALIVLGLVLLSVTVQGKVFERC
ELARTLKRLGMDGYRGISLANWM
CLAK **次世代シーケンサで
簡単に読める** GDRS
TDYC PGAV
NACH ACAC
RVVRDPQGIRAWVAWRNRCQNR
DVRQYVQGCGV

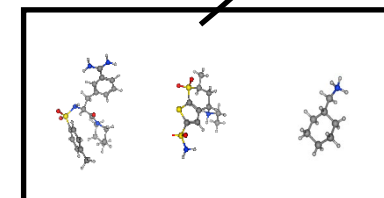
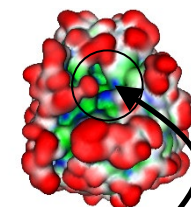


AlphaFold

タンパク質の立体構造



**実験で決めるには多大な
コスト・時間が必要**



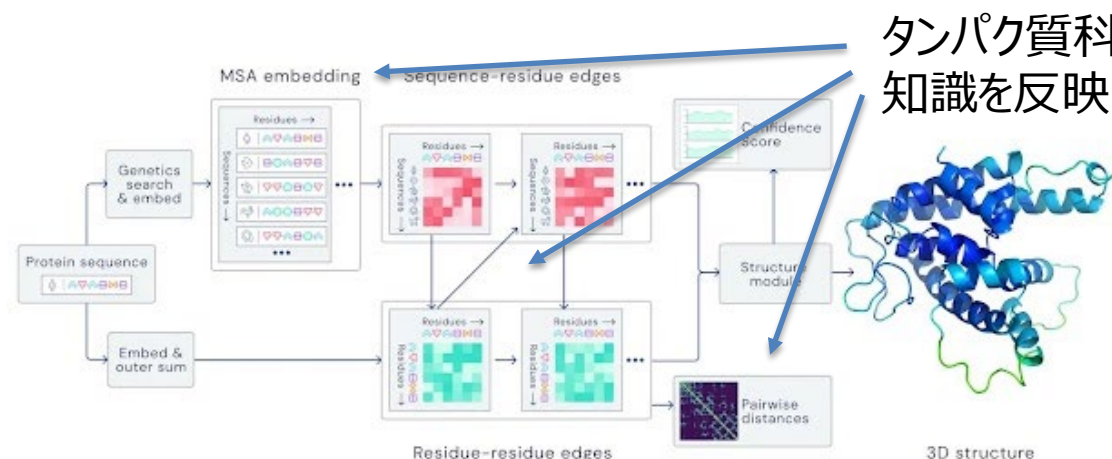
薬の設計

60年にわたって多くの科学者が取り組んできた問題を解決。
創薬を含む産業応用への展開が即座に始まると共に、
科学技術におけるAIの重要性を示した。

AlphaFoldはなぜ成功したか

AlphaFoldは、単純にデータをAIに学習させるだけでできたわけではない

2024ノーベル化学賞



AlphaFold2の構成
<https://deepmind.google/discover/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology/>



Demis Hassabis : AI



John Jumper: タンパク質科学

最先端のAI
科学的な知識
大規模計算（スパコン）

これらが密接に連携する体制が重要

**理研TRIPでもAI+科学+計算の密接な連携を推進し、
科学のグランドチャレンジ問題に挑戦**

Google/DeepMind	AlphaFoldを筆頭に、科学向けのAIモデル開発、AIエージェント開発では最先端を走っている。GNoME(物性)、WeatherNext(気象)、AlphaEvolve(アルゴリズム開発)など多数。
OpenAI	大規模言語モデルをベースに開発。生命科学向けのGPT-4b microなど。
Microsoft	MatterGen(物性)・Aurora(気象)などの開発を活発に推進。
Anthropic	プログラミング自動化向けのエージェントを提供中（Claude Code）。
Meta	タンパク質言語モデルESMの開発などを推進。Chan Zuckerberg Initiativeでは仮想細胞モデルを開発中。
Nvidia	AI for ScienceをGPUの有効な活用対象として重視しており、AIモデルのチューニング、利用容易化も推進。Evo2(ゲノム言語モデル)への貢献なども行っている。

Preferred Networks	物性分野の機械学習ポテンシャルMatlantisを開発。
Sakana AI	計算機科学分野のAI科学者を開発。

科学研究基盤モデル開発プログラム AGIS

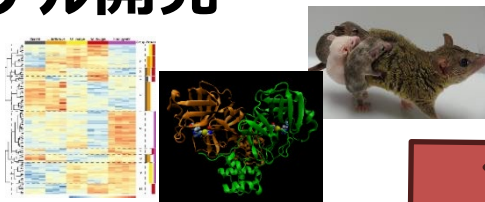
Advanced General Intelligence for Science Program (AGIS)

- ✓TRIP構想の枠組の中で、生成AIの技術も導入し、**科学研究向け生成AIモデルを開発することで、より一層の研究サイクルの加速を実現**するため、**2024年4月に開始**
- ✓理研の総合力を活用し、**先端科学を社会インパクトへ導く活動を強化**

科学向けAI開発 ターゲット領域 →理研の強みを生かせる2領域を選定

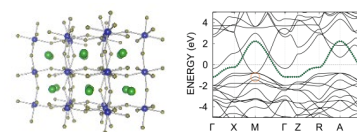
生命・医科学基盤モデル開発

分子から個体・基礎から医療まで、多様な階層のモデル開発



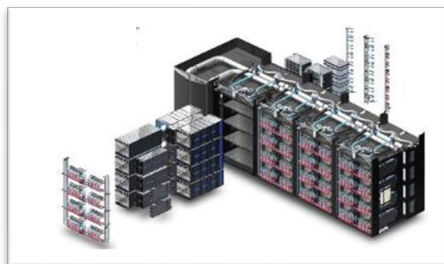
材料・物性基盤モデル開発

固体物性・高分子、加工技術等のモデル開発



計算基盤

AI専用計算機による高度な学習・推論

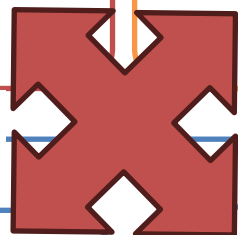


共通基盤

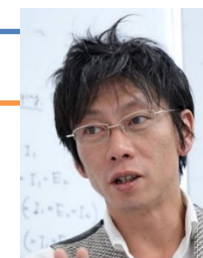
実験自動化・研究自動化
データ学習手法開発



密接に連携



泰地真弘人
(理研AGIS)



吉田亮
(理研AGIS/統数研)



松岡聡
(理研R-CCS)



高橋恒一
(理研BDR)

※現在9センターが参画

(生命機能科学研究センター、バイオリソース研究センター、脳神経科学研究センター、生命医科学研究センター、創発物性科学研究センター、光量子工学研究センター、計算科学研究センター、革新知能統合研究センター、開拓研究所)

1. Scalingへの対応

データ（量・質）/モデル/計算の規模拡大

2. 科学向け基盤モデルの開発と活用

マルチモダリティを重視

3. 実験/シミュレーション・解析の自動化



科学的成果の
創出

Grand challenge
課題の解決

理研の取り組みは、現場の科学者を中心に、データ取得まで含めて行うAI for Scienceへの組織的な取り組みとして先駆的なもの

【参考】米国アルゴンヌ国立研究所でAI for Scienceの組織的枠組の検討を開始したのは2024年6月から

<https://www.newswise.com/doescience/argonne-to-support-new-ai-for-science-projects-as-part-of-the-national-ai-research-resource-pilot>

■ 大規模言語モデル

(Large Language Model)

人間のこれまでの言語活動
からの学習



**人間の知性を大きく超えることは
難しい**

かなりかしこい人間止まり



■ マルチモーダル基盤モデル

(Multimodal Foundation Model)

自然言語

+ 数値・画像情報

+ 高速計算・情報処理



人間が本来苦手な数値情報も加えることで
これまでにない付加価値を創出可能

例：細胞の遺伝子状態 + 画像



マルチモーダル技術は、
カルテ + 医療画像 + 日常活動データの
総合的判断などにも適用可能



例：細胞応答の基盤モデル

データ取得：細胞応答マップの構築

理研の強みである

高い計測技術を活かし、

細胞の刺激に対する動的応答のマルチモーダル

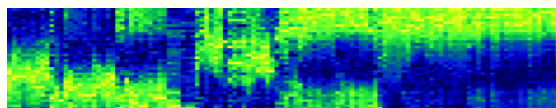
データを時間軸に沿って大量に取得

細胞の種類 × 刺激の種類 × 時間 : 250万点以上

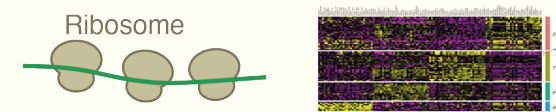


@TogoTV

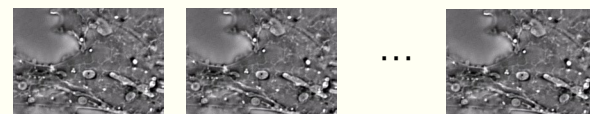
- 遺伝子の動き



- タンパク質への翻訳



- 細胞の動画



他に類のない
マルチモーダル
データ

モデル開発： 細胞レベル 応答モデル

外部刺激に対する細胞
動態が予測できるモデル

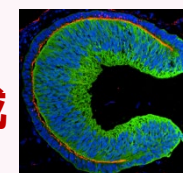
基盤モデルとして
様々な応用が可能

個別化医療

薬の副作用
予測



iPS細胞から
ミニ臓器の作成



二階堂愛
(科学大/AGIS)



岡田康志
(東大/BDR)



岩崎信太郎
(CPR)

AI for Scienceにおいても、大規模言語モデル向け同等以上の計算能力が必要

例：ゲノム言語モデルEvo2のトレーニングでは、ABCI3.0数か月分を消費



松岡聡
(理研R-CCS)

理研計算科学研究センター（R-CCS）でのハードウェア・ソフトウェアの開発・整備

■ AI for Scienceのための専用スパコンの導入

▷ **最新GPUを用いたAIスパコンを2025年度に導入**

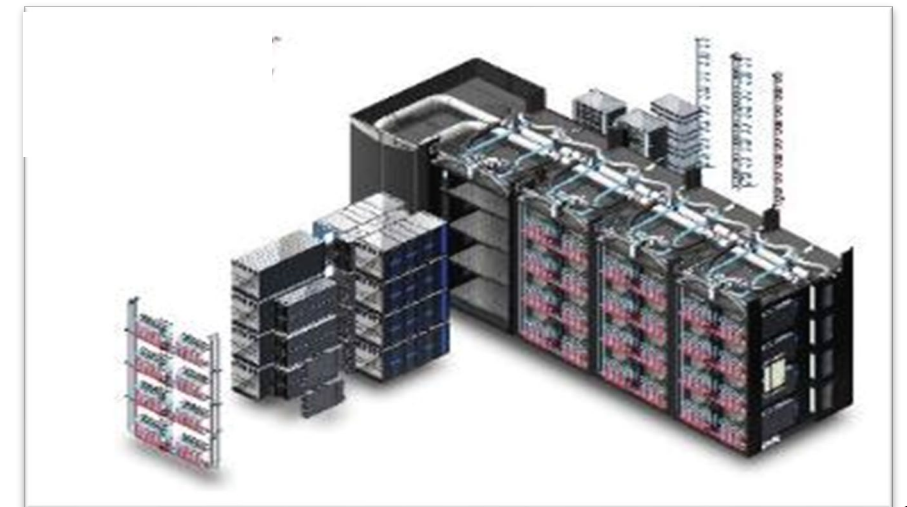
ABCI3.0の2/3程度の性能となる見込み

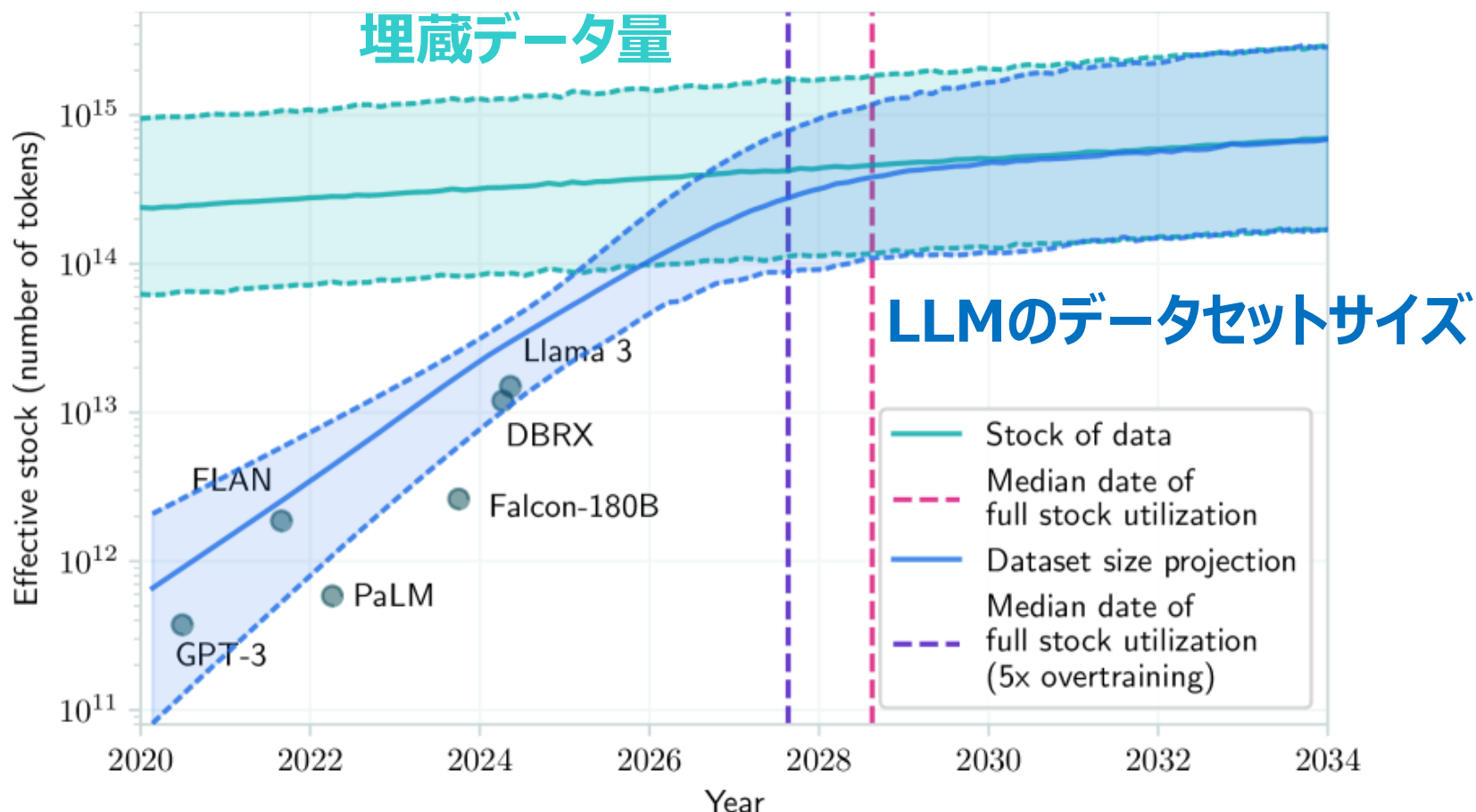
▷ 低精度演算を用いれば、**AI for Scienceで重要な推論計算では ABCI3.0を上回る性能**を実現可能

▷ パイロット版を2024年度に導入し、 基盤モデルの開発・利用に必要な環境を検証中

▷ 国内のAI for Science研究に幅広く活用

■ 富岳でのシミュレーションとAIの融合





P. Villalobos et al., "Will we run out of data?
Limits of LLM scaling based on human-generated
data", *arXiv:2211.04325v2*

データはAIモデルにとっての
「石油」

言語モデルにおいては
もうすぐデータがなくなる

科学データは無尽蔵かつ
足りない！

**科学データの採掘の加速
が重要**

**AIが必要とする（AIを賢
くする）データ取得戦略
が重要**

■ AIとロボットで24時間365日実験を進める未来型ラボ

→ 科学研究の自動化



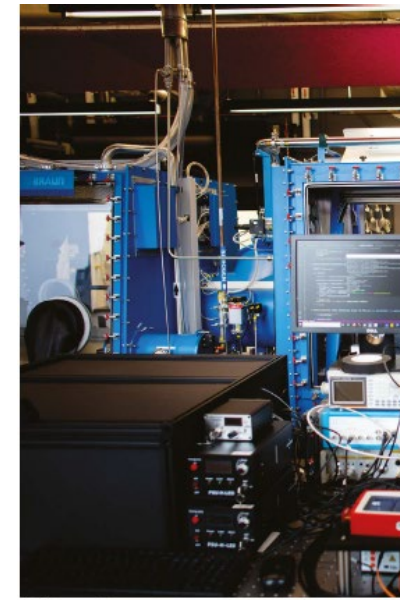
高橋恒一
(理研BDR)



自動実験系（ロボットラボ）の開発



Nature誌
2025年の注目技術7選
1. 自律駆動ラボ
4. 生命科学基盤モデル



ロボットラボは
世界的にも注目
理研では10年前から
先駆的に取り組んできた

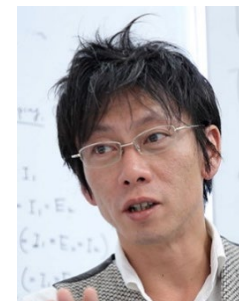
Self-driving labs, such as this one at the Acceleration Consortium in Toronto, Canada, use algorithms and robots to advance materials science.

SEVEN TECHNOLOGIES
TO WATCH IN 2025

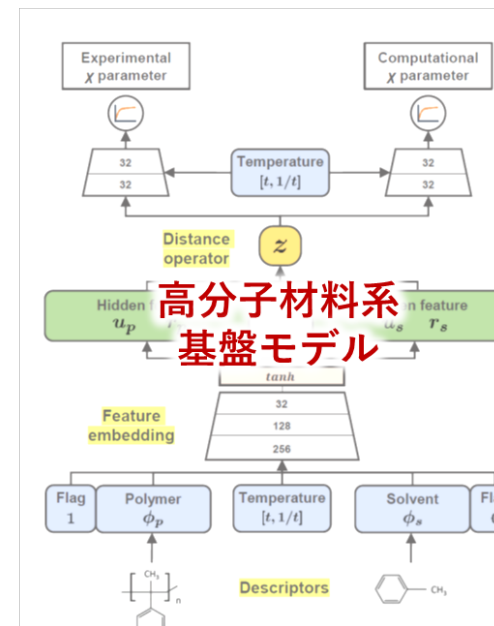
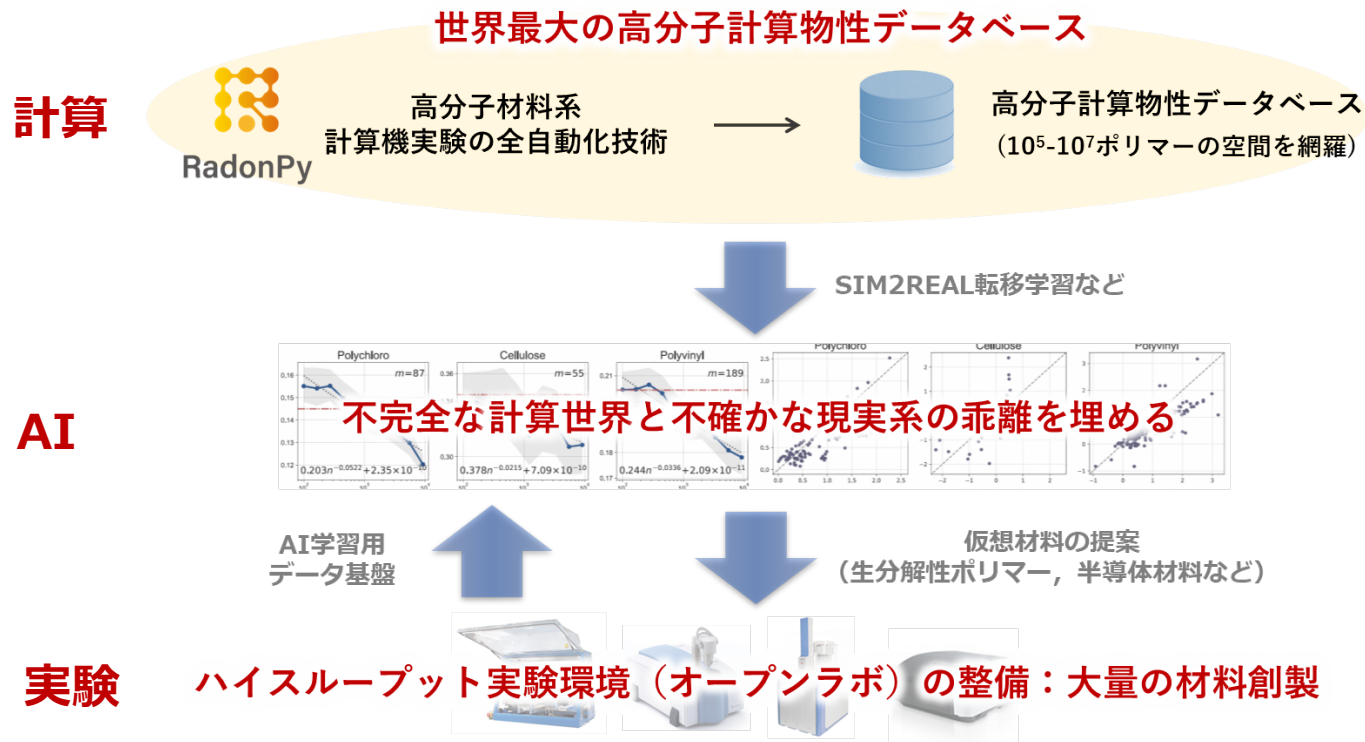
■例：高分子の基盤モデル開発（統数研・理研）

▷ 富岳等を用いた**大規模シミュレーションでデータを生成**

▷ シミュレーション＋実験データからAIモデルを開発



吉田亮
(AGIS/統数研)



企業を含めた
コンソーシアムを形成

気象分野などでも、シミュレーションデータを活用したモデル開発が進んでいる。防災など実験が難しい分野でも重要。

■ AIエージェントは、様々なAIや機器をつなぐ「司令塔AI」

▶ AIエージェントの開発により、AIが自律的に処理を行えるようになってきた。

- プログラミング (Claude codeなど)
- 研究 (AI scientist (Sakana AI), AI co-scientist (Google Deepmind)など)

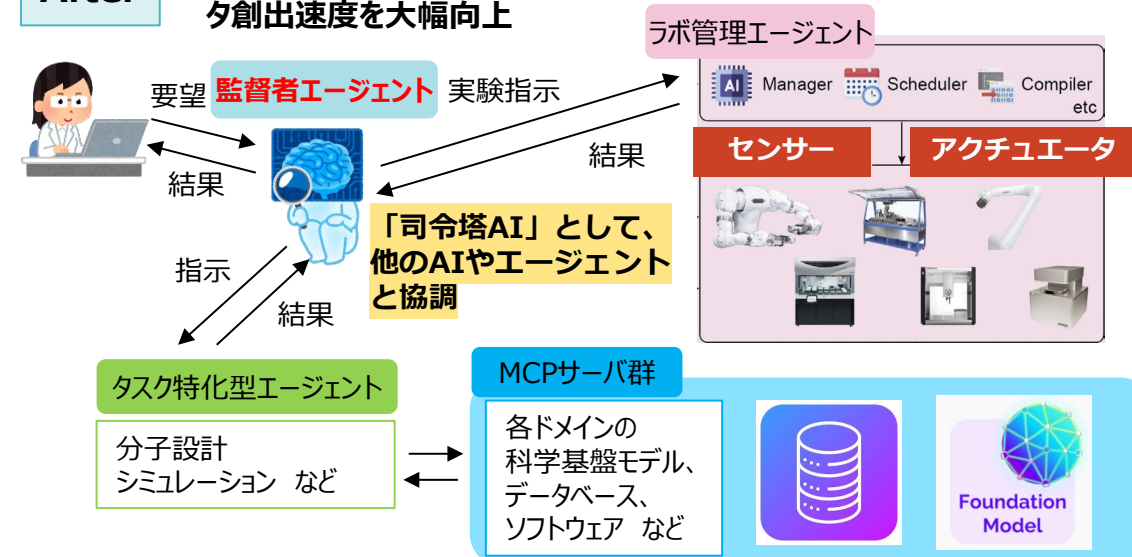
▶ 実験自動化技術の更なる活用に向けて、より省人化・高速化を追求するには、なお相当の人手が必要となっている**実験管理の自律化を追求することが必要**。

▶ AIによるシミュレーション利用についても自律化レベルを向上させ、**実験データの信頼性確保・データ創出速度の向上**を目指す。

Before 実験の計画管理・実行管理については、自動化は限定的



After AIエージェントによる実験管理自動化+シミュレーション等の利用でデータ創出速度を大幅向上



- Argonne National Laboratory (US) とのMoU締結(2024/4/5)
 - ▷ ANLはDOEのAI for Scienceプロジェクトの中心機関
 - ▷ 1兆パラメタ規模の科学向け基盤モデル開発に向けた国際協力“Trillion Parameter Consortium”（理研も発起人）のとりまとめ役。



- その他、海外企業との連携を議論中
 - ▷ AI for Scienceは、Big Techから見てもアカデミアへの期待が大きい分野となっている。

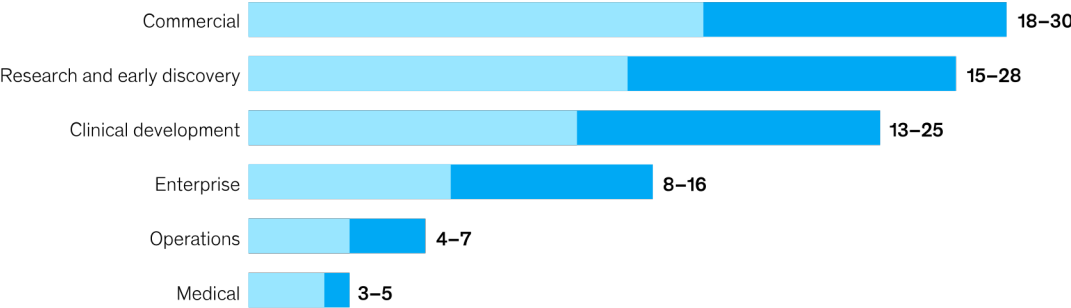
- 理研AGISプロジェクトは、科学向けAIの開発を、実験・シミュレーション自動化技術、マルチモーダルなAI向けデータ取得と融合させた先駆的な試み
- **科学向けAIの活用は、今後ほぼ全ての科学者にとって必須**
 - ▷ AI for Scienceの利用支援・エージェント開発
 - ▷ 国内への展開：ライフ分野で全国版のAI for Scienceプロジェクトを検討中（文科省ライフ課）
 - ▷ 科学の領域の広大さを考えると、Big Techを含めた国際連携は必須。その上でも、**他にない価値のあるデータ・モデル、科学者のコミットメント**が重要
- 産業開拓
 - ▷ AI for Scienceに取り組む企業との連携・ベンチャー化などの支援
- 新しい分野の開拓
 - ▷ これまでコンピュータで扱えなかった分野も、AIでDX化可能。
 - ▷ 例えば社会科学なども、AIエージェントでの研究が拡大中。

補足資料

生成AIは製薬・医療製品業界で 年間8.6兆円から14兆円の経済価値へ

Generative AI is expected to produce \$60 billion to \$110 billion in annual value across the pharmaceutical industry value chain.

Expected value annually, \$ billion

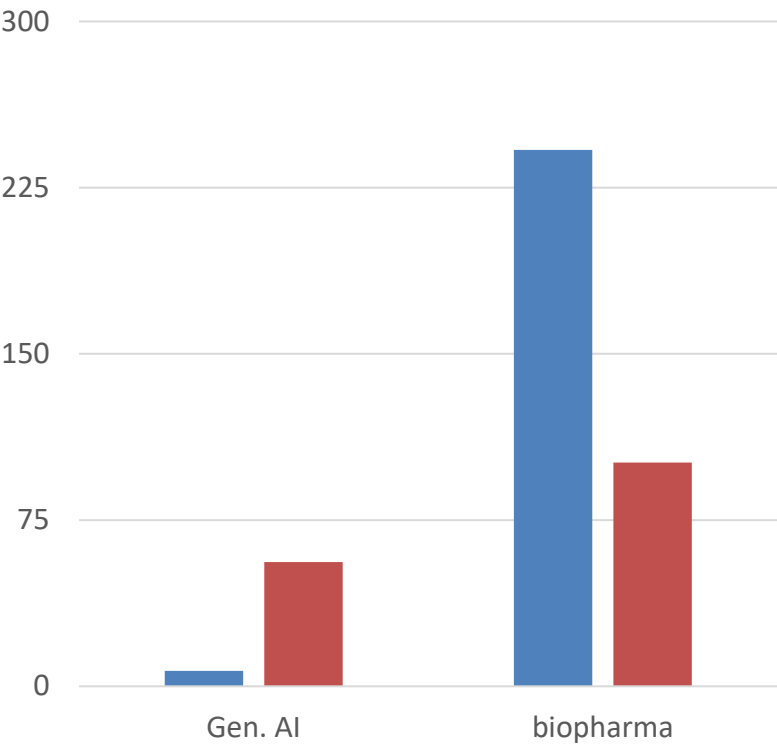


Source: McKinsey analysis

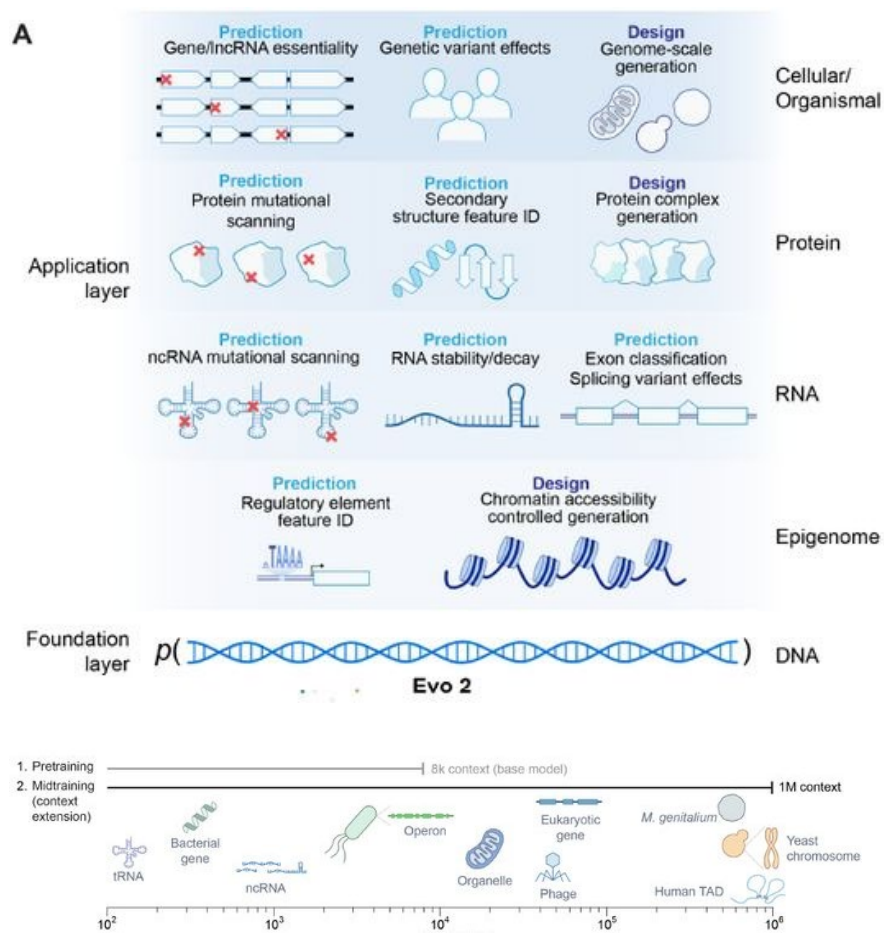
McKinsey & Company

McKinsey Global Institute (MGI)
Generative AI in the pharmaceutical industry: Moving from hype to reality. 2024

Gen AI. スタートアップ企業は8倍に バイオ医薬スタートアップ企業は1/2に



Silicon Valley Bank (SVB)
Healthcare Industry Trends 2024 Annual Report



■世界最大のゲノムDNA言語モデル

- ▷ Multi-hybrid StripedHyena 2 architecture
- ▷ 9.3兆塩基、40Bパラメータ、100万トークンのコンテキストウィンドウ

■ABC13.0の4カ月分以上の計算資源を投入

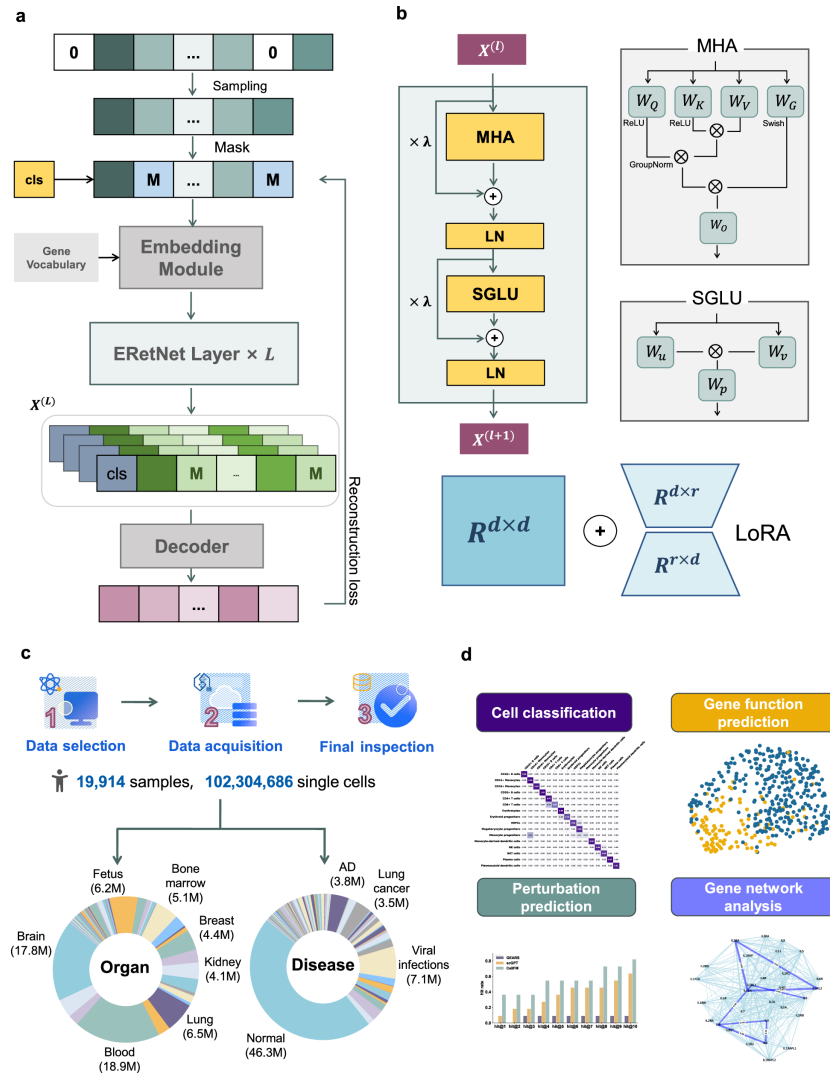
■米国Arc Institute, Stanford U., Nvidia等が開発（論文著者52名）

■応用タスク

- ▷ 変異の病原性予測
- ▷ がん遺伝子の変異分類
- ▷ ゲノム構造アノテーション
- ▷ 新規ゲノム生成

■未達成タスク

- ▷ 遺伝子発現変化などダイナミクスの予想
- ▷ 個人ゲノム配列生成など



Zeng Y. et al. Nature Comm. 2025

- 最大の遺伝子発現の基盤モデル
- ERetNet
- 1億細胞、8億パラメータ
- HUAWEIと中山大学 (中国)
 - ▷HUAWEI MindSpore AIフレームワーク
 - ▷HUAWEI Ascend 910 NPU x 4 (Atlas800)
 - Ascend 910は米国の対中輸出規制下で中国で開発されているGPU (NPU)
- GPUの開発からモデル改良、学習までを中国のみで可能であることを証明した