RIKEN TRIP
Transformative Research Innovation Platform
of RIKEN platforms

# TRIP-AGIS:
# 理研のAI for Science Project

**A**dvanced **G**eneral **I**ntelligence for **S**cience of
**T**ransformative **R**esearch **I**nnovation **P**latform

泰地　真弘人 (Makoto Taiji)　　taiji@riken.jp

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

# 自己紹介：専用計算機・ASIC開発

RIKEN TRIP

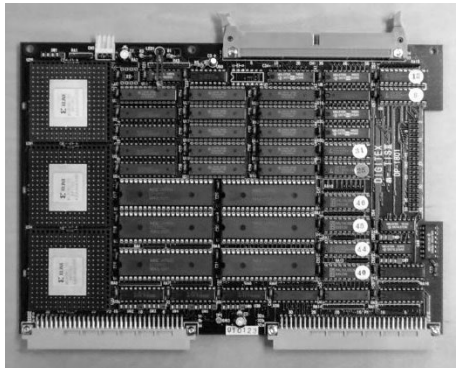
## ヘテロジニアス計算機開発の先駆

m-TIS II (1991)

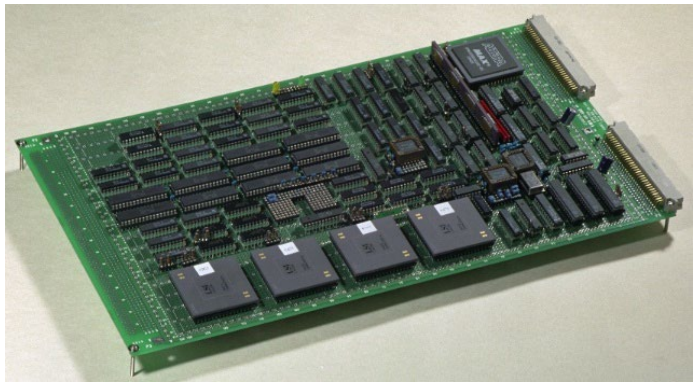
スピン系のモンテカルロ計算

最初期のFPGAを用いた再構成可能計算機

（東大理）



MD-GRAPE (1996)



最初のASICベースMD加速器

（東大教養）

GRAPE-4 (1995)

世界初のTFLOPS級計算機

(東大教養)

MDGRAPE-3 (2006)

世界初のPFLOPS級計算機

（理研）

MDGRAPE-4A (2019)

SoCベースのStrong Scaling Accelerator（理研）

# 複雑な自然・人工物の理解に向けて

RIKEN TRIP

- この世界は多様性に溢れ、また複雑である。基本的には、これらの複雑性は多自由度の非線形性に由来している。


*Inorganic*
*Eyjafjallajökull volcano*

*Organic*
*Flapjack octopus*


*Machine*
*Differential engine*

- 現代においては、人工物もまた非常に複雑である。例えば並列計算機は性能の予測が難しく、その最適化も自明でない。

*Videos from youtube*

# 非線形・複雑系の問題

RIKEN TRIP

非線形系を扱うための理論・計算的枠組は少ない

1. **厳密解の方法**

   ソリトン解や、近可積分系などの成果があるが、使える対象が限定的。

2. **繰り込み群の方法**

   相転移など、特異点があるケースで強力だが、限定的。

3. **シミュレーションの方法**

   基礎方程式・モデルがある場合には幅広く使える強力な手法。
   計算量の限界、計算可能なモデルがないときの問題がある。

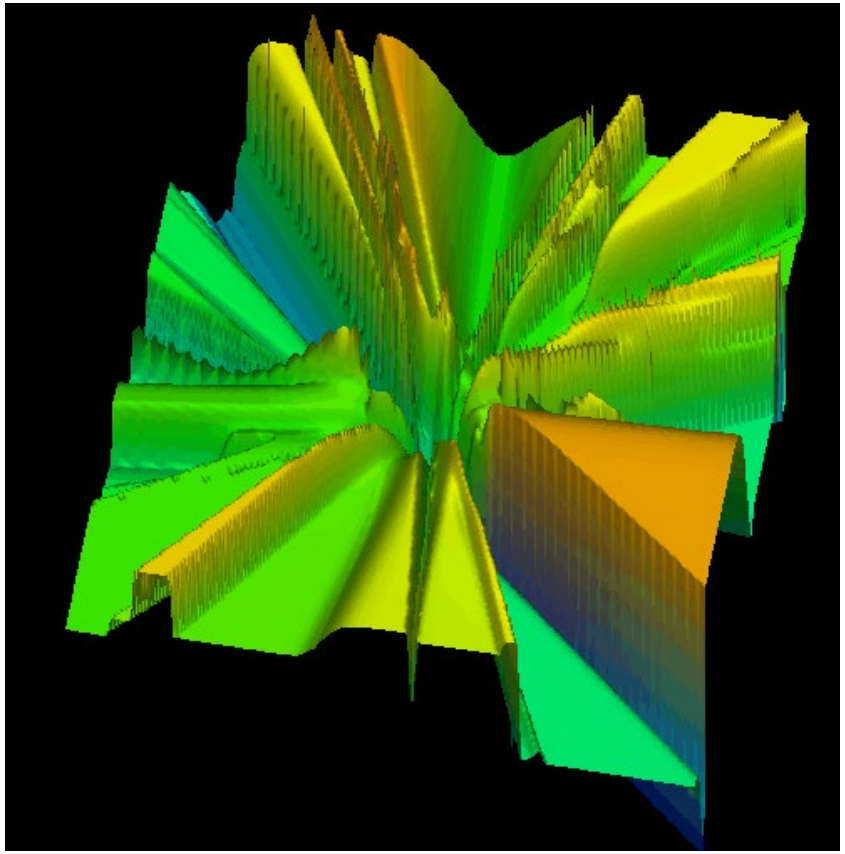
4. **機械学習の方法**

   近年の深層学習の発展により、非線形関数の汎用的な表現法として使えるようになってきた。データがあれば、モデルがなくても予測可能に。
   **これまでアプローチできなかった領域にアプローチできる可能性がある。**
   **（生命システムや社会システムなど）**

# 人工ニューラルネットワーク(ANN)の非線形性

RIKEN TRIP

- それ自身が強い非線形系で、複雑な大域的構造を持つ
- モデルの数理＋データの性質＋データの表現法が交絡
- 非線形の問題の「博物学性」：問題ごとの個別性の強さ



RNNの誤差のlandscape

Taiji, M. and Ikegami, T. “Dynamics of Internal Models in Game Players”, Phys. D (1999).

“Mathematical theory is not critical to the development of machine learning. *But scientific inquiry is.*”

L. Breiman, “Reflections after refereeing papers for NIPS”, in The Mathematics of Generalization (1994).

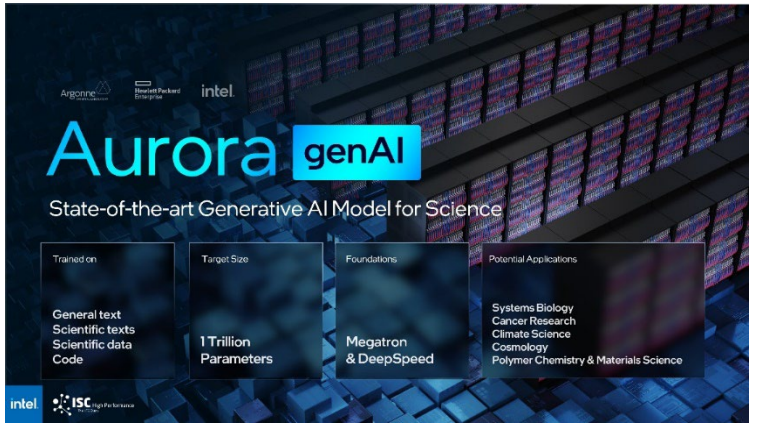
# AI･自動化による科学技術革命

RIKEN TRIP

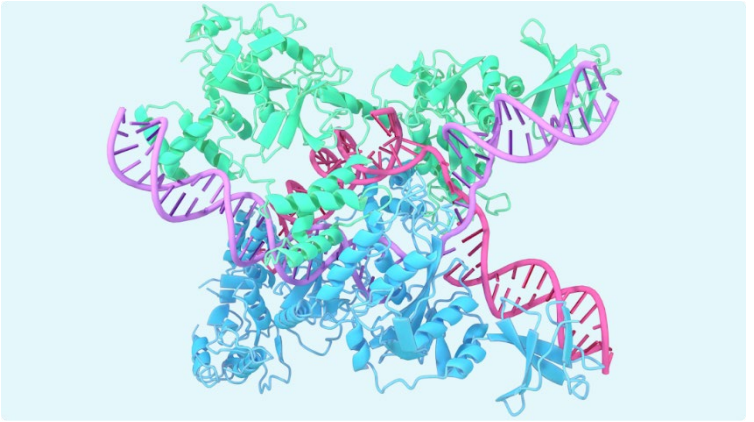
■ **第4次産業革命＝知的労働の自動化**
- ➢ 研究活動はその中核で、**価値の源泉**

■ 大規模言語モデルの汎用性と
科学研究における有用性
- ➢ LLMは色々な概念を理解しているように
ふるまい、適切なタスク選択が可能に
- ➢ 会話だけでなく、適切な道具や手法を
能動的に用いることが可能
  - ー数値的な分析手法・シミュレーション手法などの選択
  - ー実験の自動化

■ 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
大規模データから基盤モデルを作成する流れが加速中




Aurora GPT:
1兆パラメタの科学向け生成モデル
（計画中）
Argonne ＋ Intel



AlphaFold 3 by DeepMind

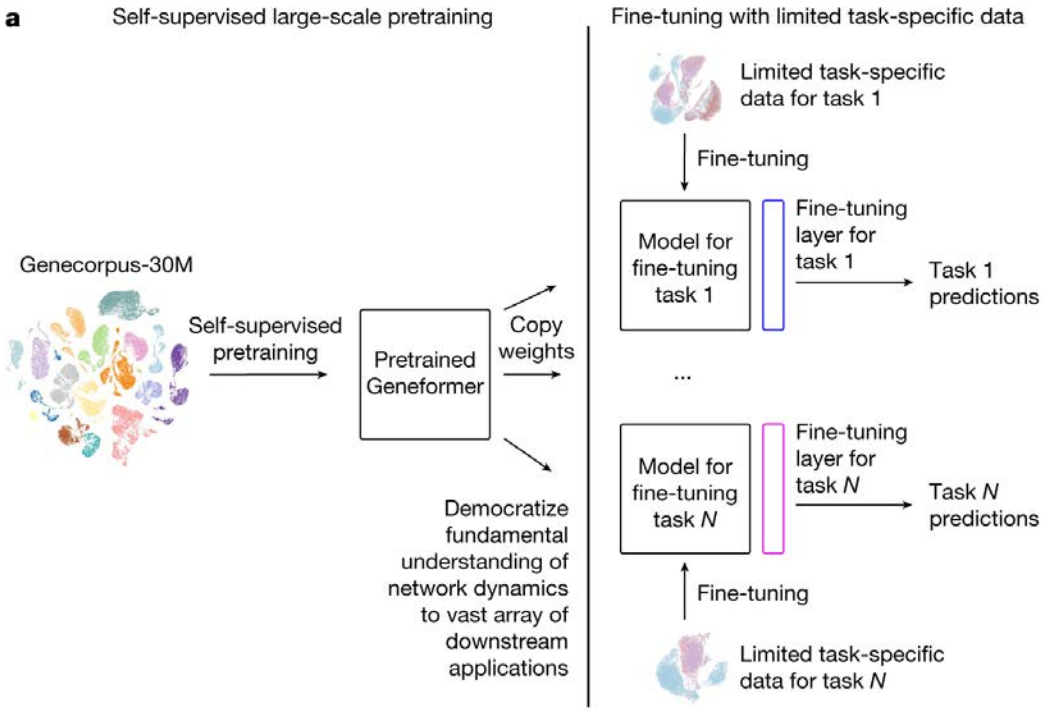
# 生命科学における基盤モデル・生成モデルの例

RIKEN TRIP

**生命分野は、様々なデータの蓄積を利用した基盤モデルの開発が進行中**

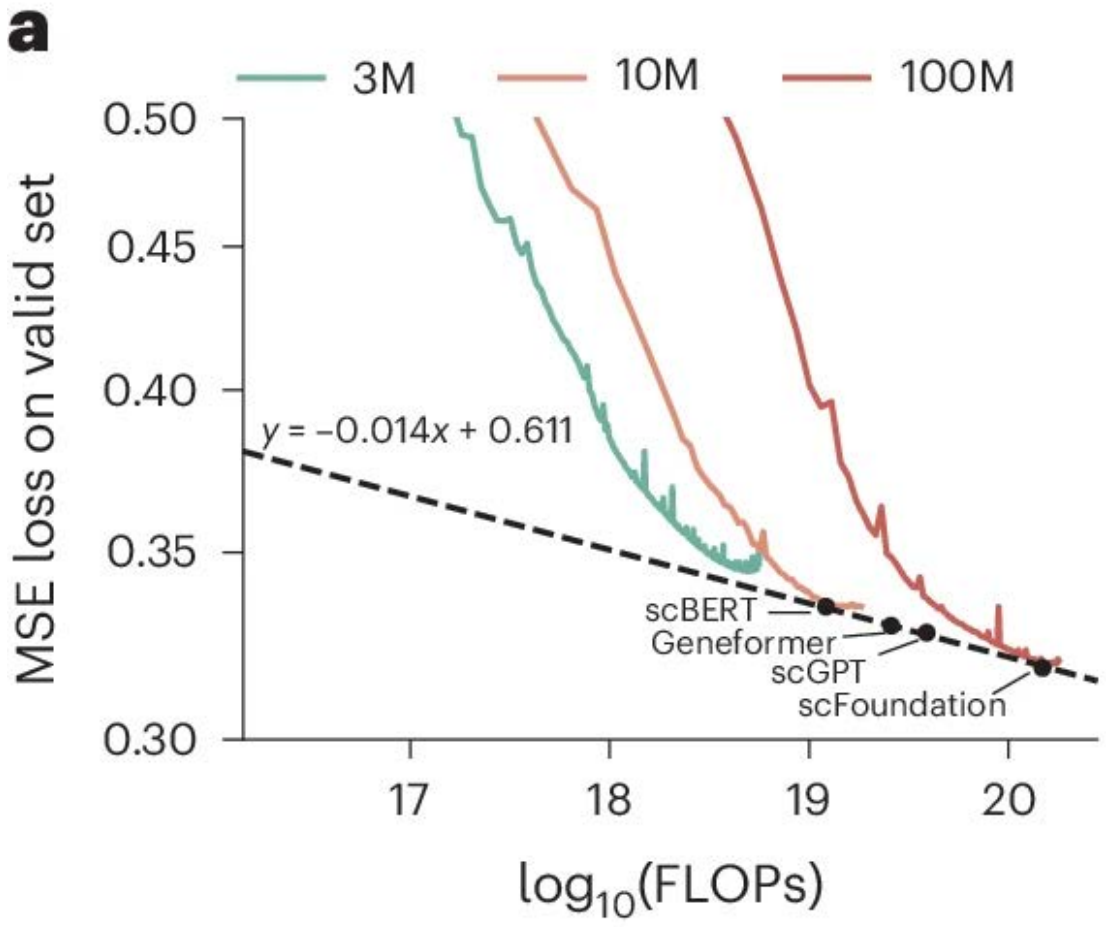
- 自然言語系：BERT/GPTをベースに生命科学向けに追加学習したモデル
  - ▷BioBERT/BioGPT：生命分野論文
  - ▷MedGPT/GatorTron：医療情報
- タンパク質
  - ▷ESM-2 (Meta, タンパク質の言語モデル)
  - ▷RF Diffusion (Baker Lab., タンパク質設計)
- ゲノム
  - ▷GenSLM 微生物・ウィルスの配列
  - ▷Geneformer / scGPT　一細胞遺伝子発現
- 化合物生成
  - ▷古くから多数の取り組み

Geneformer: 一細胞遺伝子発現の基盤モデル




Theodoris, C.V., Xiao, L., Chopra, A. *et al.* **Transfer learning enables predictions in network biology**. *Nature* **618**, 616–624 (2023). https://doi.org/10.1038/s41586-023-06139-9

# 遺伝子発現の基盤モデルのスケーリング則




LLM同様のスケーリング則が成立

～100M parametersと小規模な段階だが、まだまだ色々足りない

- 細胞種のVariation
- 遺伝的Variation
- 時系列・細胞間相互作用

Hao, M., Gong, J., Zeng, X. *et al.* Large-scale foundation model on single-cell transcriptomics. *Nat Methods* **21**, 1481–1491 (2024). https://doi.org/10.1038/s41592-024-02305-7

# タンパク質：ESM-2

RIKEN TRIP

- Evolutionary Scale Model-2, MetaAI

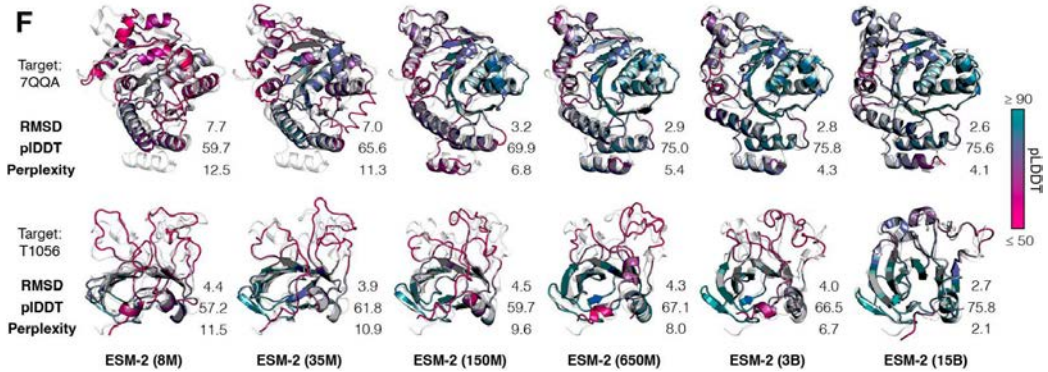
  Lin, Zeming, et al. "Evolutionary-scale prediction of atomic-level protein structure with a language model." *Science* **379**.6637 (2023): 1123-1130

- タンパク質のアミノ酸配列を学習させたモデル
  - ▷Data:UniRef 50(タンパク質のアミノ酸配列のDB)の配列-β+α 約60M
  - ▷Transformerで自己教師あり学習
  - ▷最大15Bパラメタ
- ESM-2＋Protein Data Bankの立体構造(おそらく$10^4$～$10^5$)で立体構造予測
  - ▷タンパク質内のアミノ酸のコンタクトを予測。単語間の関連に相当する。
  - ▷AlphaFold2と同等の性能を高速に実現
  - ▷最大モデルで最高の性能

# ESM-2で意外だったこと

RIKEN TRIP

- データ量は60Mとまあまあの量だが、実際のタンパク質の形の多様性は$10^5$程度。
- それでも、モデル規模を大きくするほど予測性能が上がる。




Lin, Zeming, et al. *Science* 379.6637 (2023): 1123-1130.

- 科学分野での基盤モデルの可能性、現状データでも思ったよりあるのではないか。
- 自然言語でも、実際に表現したいことの内容はそれほど多様なわけでもないが、その言語としての表現は多様。そうした組み合わせ的な自由度・確率的構造に対応するために大規模モデルが必要ということか。

# 科学研究基盤モデル開発プログラム　TRIP-AGIS

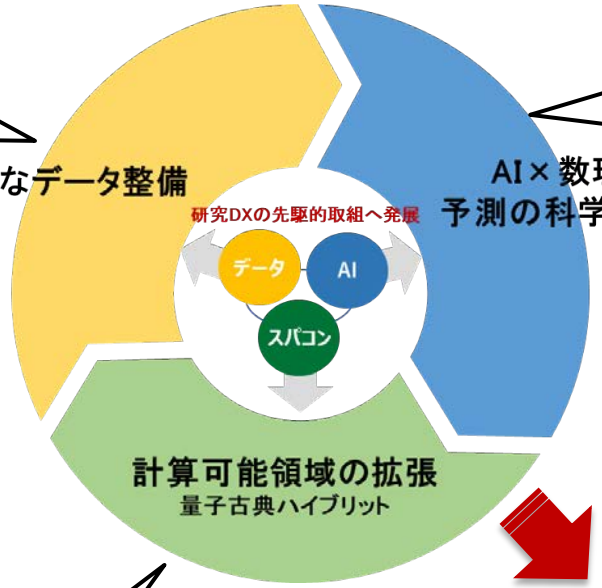
RIKEN TRIP

Advanced General Intelligence for Science Program (AGIS)

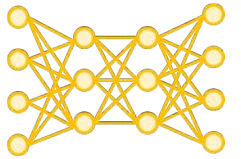
✓TRIP Second Stepとして、生成AIの技術も導入し、**科学研究向け生成AIモデル**を開発することで、**より一層の研究サイクルの加速を実現**
✓先端科学を社会インパクトへ導く活動を強化

**強みを有する計測技術や実験自動化を通じた良質且つ膨大なデータ生産**

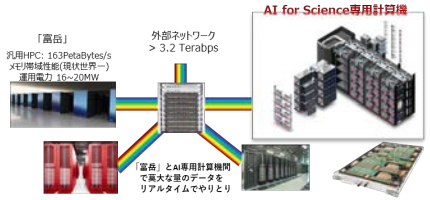





**科学研究向け生成AIモデルの開発・利用・共用（生命医科学・材料物性）**



**「富岳」とAI専用計算機を連携させ高度な学習・推論を実現**




**研究サイクルを加速し、基礎科学を起点とした先端科学を社会的インパクトに導く（Gx,包摂社会など）**

**生成AIによる科学技術研究加速と科学による生成AIの新たな学理と技術の開拓**

# TRIP-AGISの狙い

RIKEN TRIP

1. Scalingへの対応
   Data （量・質） / model / computation
2. 科学向け基盤モデルの開発と活用
   Multimodal Foundation models
3. 実験/シミュレーション・解析の自動化
   Active learning

**科学的成果の創出**
**Grand challenge課題の解決**

AI for Science / Science by AIの推進による

科学の領域の拡大

※なお我々は、AIには演繹的手法も含まれると考えています

# 科学分野でのマルチモーダル基盤モデルの重要性

RIKEN TRIP

## ■ 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

**超知性のような存在に進化可能**

# TRIP-AGISの全体像

RIKEN TRIP

## ① 科学研究向け生成AIモデル開発・共用の共通基盤技術

- 先進モデル：**多種多様なデータ（マルチモーダルなデータ）の学習・生成が可能**な**基盤技術開発**
- 良質なデータ：**学習に必要な大量のデータ創出**と、**モデルが生成した実験計画の自動実行**を両立できる**実験の自動化・高速化技術**

## ② 特定科学分野の科学研究向け生成AIモデルの開発・共用

### 生命・医科学分野

- 良質なデータ：薬物等を与えた際の細胞の経時的な変化、疾患が動物の行動や身体に与える変化　等
- 先進モデル：**ゲノム・細胞から生体全体までの現象を統合的に解釈**して予測できるモデルの実現

### 材料・物性科学分野

- 良質なデータ：材料の構造、物性、電子状態
  材料の作製方法　等
- 先進モデル：**無機・有機を問わず、物性、材料構造、作製方法等を統合的に解釈**してデータを生成できるモデルの実現

## ③ 革新的な計算基盤の開拓

- 計算資源：多様な種類の科学研究データの推論・学習・生成に最適化された、
  **科学研究向け生成AIモデルの開発・共用のための計算環境を整備・運用等**
  従来のGPUのみでは実現不可能な**演算処理の高度化・高速化を実現する新たな計算機原理の研究**

# 基盤モデルに多様な科学研究データを学習させる手法の開発

## ① 科学研究向け生成AIモデル開発・共用の共通基盤技術

◆ 科学研究では、言語に加えて、図表やグラフなどの数値情報のほか、様々な種類のデータや知識を使用。
このため、既存の基盤モデルも活用しながら、**多様なデータ・知識を学習・生成可能とするための技術開発**を実施。

◆ また、「②特定科学分野の科学研究向け生成AIモデルの開発・共用」で開発する**特定科学分野の科学研究向け生成AIモデルと、データベースや実験自動化ロボットを接続するための機能開発**を実施。

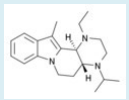
牛久祥孝
(Omron SinicX/
理研BDR)

**多様な科学研究データを学習・生成させる技術開発**

**データベース、論文等を用いて、追加学習、図表やグラフの読みとり・生成等を実施**

**既存の知識・データ（①で取得）**

・学術文献（図表・グラフ等の知識表現）
・科学研究データ（画像、測定データ）



$$\sigma_x = \lambda\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}\right)+2\mu\frac{\partial u}{\partial x}$$

**特定分野の科学研究データ（②で取得・創出）**

生命・医科学、材料・物性科学の様々なデータ

学習

既存の基盤モデル（言語）

科学研究向け生成AIモデルのベース（多様な科学研究データを生成・学習可能）

**科学研究向け生成AIモデル**

生成されたデータもシミュレーション等に活用

AIが生成した実験条件をもとに自動実験

科学研究データ、知識、実験条件等を生成

**接続のための機能開発**

**データベースや実験自動化ロボットなどと接続するためのインタフェースを開発・提供**

科学データベースから学習

自動実験の結果から学習

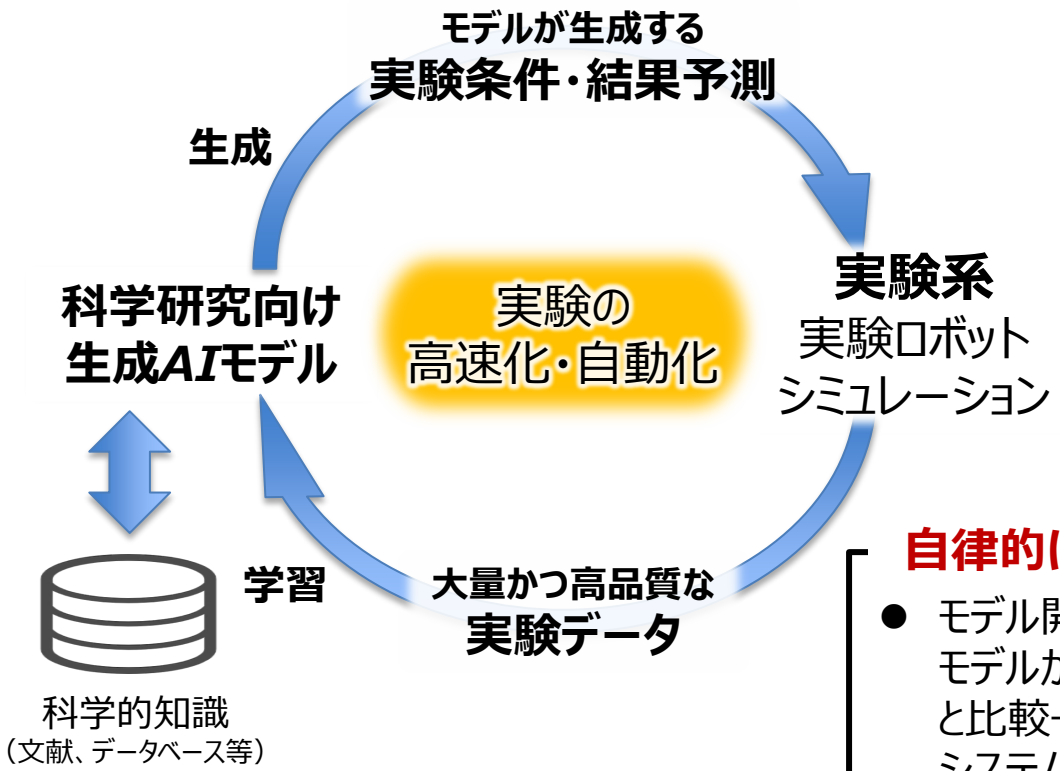
**生命・医科学や材料・物性以外の科学分野にも展開可能な基盤を提供。**

# 基盤モデルが必要とする膨大かつ良質なデータを自律的に取得する実験自動化技術の開発

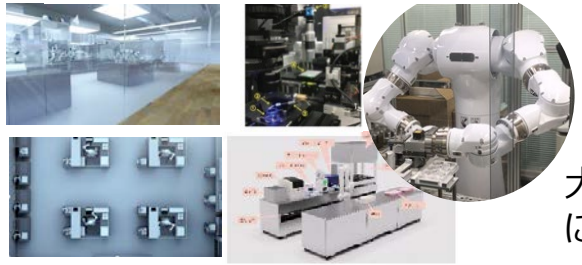
## ① 科学研究向け生成AIモデル開発・共用の共通基盤技術

高橋恒一
(理研BDR)

- ◆ 科学研究向け生成*AI*モデル開発には、従来とは次元の異なる大量の高品質な科学研究データが必要。このため、**実験の高速化・自動化に係る技術開発**を実施し、**厳密な実験条件の下で大量のデータを創出できる基盤を整備。**
- ◆ モデル自体に**自らがより賢くなるための実験条件を生成させる能動学習技術を開発。**学習の大幅な効率化を実現するとともに、将来的に**モデル自体が自律的に科学研究の一部を推進可能なシステムに発展することを見越す。**




### 自動実験系の開発



大規模自律ロボットラボ
による超並列実験

シミュレータによる大量の
定量データ生成

### 自律的に学習する方式の開発

- モデル開発中のある段階で、**モデルと自動実験系を結合。**
  モデルが実験条件・結果を予測→実験ロボット等による実験結果と比較→誤差をモデルが学習、というサイクルを自律的に繰り返すシステムを構築することで、**効率的なモデル高度化を実現。**

*Google*傘下の*DeepMind*開発の「*AlphaGo Zero*」の先例では、完全情報ゲームである囲碁で*AI*同士の対局による「自律的学習」を*3*日間続け、世界トップ棋士を破った「*AlphaGo*」に*100*戦*100*勝。
本*PJ*ではこれを**科学研究のような不完全情報・部分観測問題に拡張。従来より格段に高速で高精度な研究アウトプット・知的創造活動の拡充を目指す。**

# 分子レベル創薬モデル

## ②-1 特定科学分野の科学研究向け生成AIモデルの開発・共用（生命・医科学分野）

- ◆ 理研に蓄積されたデータや公共データベース、医療データ等の**マルチモーダルなデータを統合して体系的に連結させ、生成モデル化。**
- ◆ 並行して、分子レベルのモデルとして**化合物・タンパク質マルチモーダル生成AIモデル**を開発。
- ◆ これを知識グラフベースのモデルに融合し、*LLM*による自然言語インターフェイスを統合することで、**化合物ータンパク質ーオミックスー疾患を融合した分子レベル創薬モデルを開発。**

小島諒介
(京大/
理研R-CCS)

創薬、医療分野のマルチモーダルなデータ
（例：理研内データ、公共*DB*、医療データ等）

遺伝子、タンパク構造、イメージング、プロテオーム、
リピドーム、脳・神経活動、その他表現型　等

ToMMo　BIOBANK JAPAN　NBDC　PubMed　Life Intelligence Consortium

*病理画像*　*電子カルテ*

機械学習・*AI*の活用

**知識グラフ※の抽出・表現**　**非グラフデータの抽出**

**知識グラフの生成モデル化**
（非グラフデータも併せて学習）

※様々な知識を体系的に表した知識のネットワーク。
*AI*の実用的な基盤技術として活用される手法。

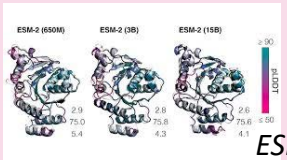
**化合物向け生成AIモデル**
- 低分子～中分子化合物の生成モデルを開発。
- モデル上で表現される化合物の構造が不連続等の課題克服を目指す。

*MoFlow/JTVAE etc.*

**タンパク質向け生成AIモデル**
- 既存のモデルに、タンパク質動態の取り込み、創薬向けのファインチューニング、核酸などへの拡張を行う。

*ESM-2/AlphaFold2 etc.*


化合物探索空間を効率よく拡大




*ESM-2*
タンパク質の特徴をよりよく把握
立体構造予測など

統合に向けた化合物表現方法の検討

**化合物・タンパク質マルチモーダル生成AIモデル**
- 結合可能なタンパク質と化合物の組み合わせを生成できるモデルを開発。
- 期待したい効果を狙った薬物の自動設計、化合物からのターゲットタンパク質（酵素）探索・同定を実現。

2つのモデルを結合

強化学習等によるファインチューニング
性能の向上、特定分野に向けた調整

**分子レベル創薬モデル**
- ✓ 新しい創薬リード化合物の全自動・半自動的な発見が実現。
- ✓ 疾患ターゲットを特定し、創薬研究に向けた利活用につなげる。

# 細胞レベル応答モデル

## ②-1 特定科学分野の科学研究向け生成AIモデルの開発・共用（生命・医科学分野）

◆ 現在より**2桁以上高い速度**で、薬物刺激などによる細胞内の摂動応答を**経時的・網羅的に計測することを可能とし、良質なデータセット※を取得。**

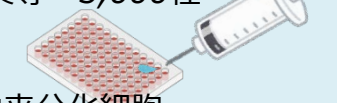
◆ 得られた細胞やRNA、タンパク質の時系列データ（画像・動画）と**ゲノムDNAデータ**を統合し、**細胞動態を予測するモデルを開発。**

※**データの量が多く、質が高く、幅が広いもの**を指し、これは既存技術よりもスループットが数桁高い計測技術と数倍高精度な計測により取得でき、**多様なタスクをこなせる生成AIモデル構築（＝日本の勝ち筋）**につながる。

二階堂愛
(医科歯科/理研BDR)

## 細胞応答マップの構築

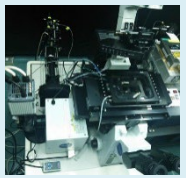
既存薬等～5,000種



iPS由来分化細胞・
オルガノイド　等　～100種

**理研の強み：高い計測技術**

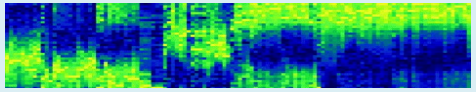
@TogoTV



**5,000種に及ぶ薬物等による細胞の応答を経時的・網羅的に計測し、100種の細胞のタンパク質等のオミックスデータを取得。**

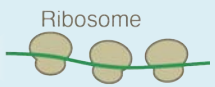
5,000種 × 100種 ＝ 50万の組合せの時系列データセットを構築

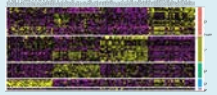
- 超大規模トランスクリプトーム（生体細胞内における遺伝子の発現状態を網羅的に把握）



スループット: 100倍
精度: 5倍

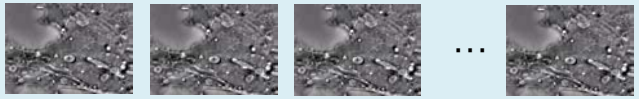
- タンパク質翻訳制御（細胞内のタンパク質量やRNA分解を網羅的に把握）



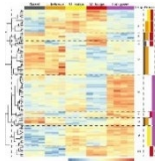
スループット: 100倍

- 高速ライブセルイメージング（細胞内の速い動きも生きたまま観察）



分解能: λ/3
速度: 10 ms/frame

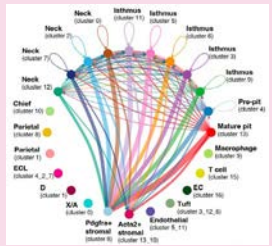
ゲノム配列



ATGCACGTCAGC
GACCGACGTAAT

## 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を開発
- ✓ これにより、薬物等による動的変化・遺伝子変異による差異予測が実現
- ✓ 創薬や再生医療・遺伝子治療の実現を通じて健康寿命の延伸やアンメット・メディカル・ニーズに対応



細胞応答マップのデータを学習したモデルを使うことで自在に細胞状態を制御可能

# 個体レベル行動・特性モデル

## ②-1 特定科学分野の科学研究向け生成AIモデルの開発・共用（生命・医科学分野）

桝屋啓志
(理研BRC)

- 健常モデル動物と疾患モデル動物（マウス・マーモセット）について、個体レベル及び集団レベルでの**日常的行動**※**データ（動画・音声等）を数か月単位で大規模に取得し、解析**。環境、薬剤投与等の各種外部条件も設定し、身体全体の理解を目指す。
- これら**行動データ**と、**ゲノムデータと身体構造データ**を組み合わせて、**行動から疾患の言い当てや各行動の意味の解釈、予測ができるモデル**を開発。

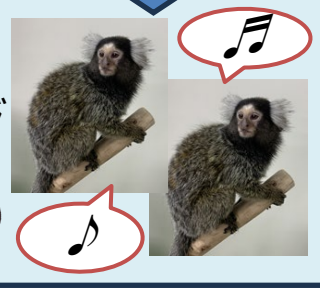
※「**行動**」は、時系列を追って計測すると、長期間の「文脈（日周、性周期、保育等）」に応じて短時間の「振る舞い」が変容する特徴を有する。
**自然言語に近い構造を持ち、身体の多くの不具合が反映される**ため、「行動」にフォーカスしてデータを収集する。

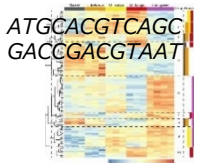
### 個体レベルの日常行動モニタリング
### 行動・音声データをAIツールを用いて解析

マウス　　マーモセット

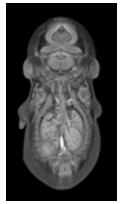

健常およびヒト疾患モデル　　ヒトに対応する多くの特徴

**行動および音声データ**






ゲノムデータ

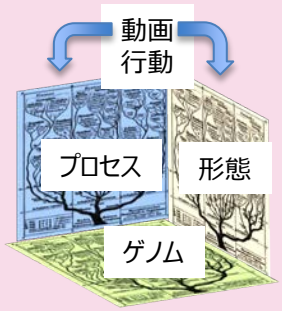


*CT*撮像

### 個体レベル行動・特性モデル

個体行動の多様性と複雑性を反映するモデルを開発。
精神・神経疾患のみならず、**多様な疾患及びその未病状態や前兆を外的観察から言い当てる。**

これにより、**個体レベルでの疾患や生命のプロセスの全体理解に貢献**するとともに、ヒトの行動モデルを補完する。
（ヒトは何故その行動に至るのか？を説明する）




マウスとマーモセットの特徴を生かしたデータ取得

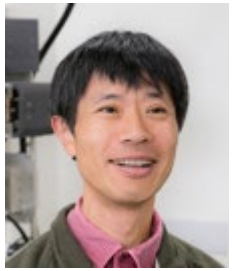
- マウスはヒトの疾患に対応した多くの遺伝子改変モデルが利用可能。表現型がゲノム情報に強く裏打ちされており、ヒトにおけるゲノム-疾患関連性のモデル。
- マーモセットは、霊長類特有の発達した高次脳機能を持ち、脳構造、行動、音声コミュニケーションによる社会行動等、表現型特徴が直接的にヒトに関連。

# 材料・物性科学分野の科学研究向けモデルを構築するための技術開発

## ②-2 特定科学分野の科学研究向け生成AIモデルの開発・共用（材料・物性科学分野）

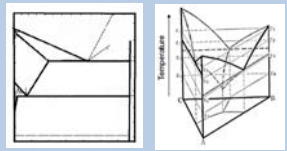
◆ 文献によって表記が異なる既存の物質データを整理・統合するための名寄せ用辞書を作製し、**材料・物性に関する実験・文献データを構造化**するとともに、**マテリアルデータに強みを有する研究機関とも連携**し、**良質なデータ取得を加速・蓄積。**

◆ 蓄積したデータを①の取組で開発する科学研究向け生成AIモデルのベースに追加学習させるとともに、**計算物理による構造最適化・物性予測を連携**させ、目指す材料機能を実現するための**物質構造やその作製方法を提案する生成AIモデル開発**を実施。

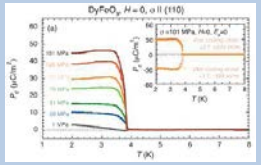


有馬孝尚
(理研CEMS)

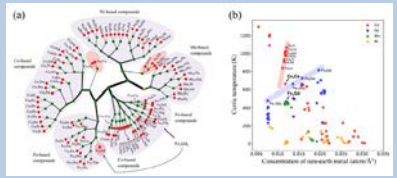
- 物質・材料作製方法に関する文献情報 （>$10^6$件を想定）
- 新規に取得する良質な実験データ



合成に関する情報



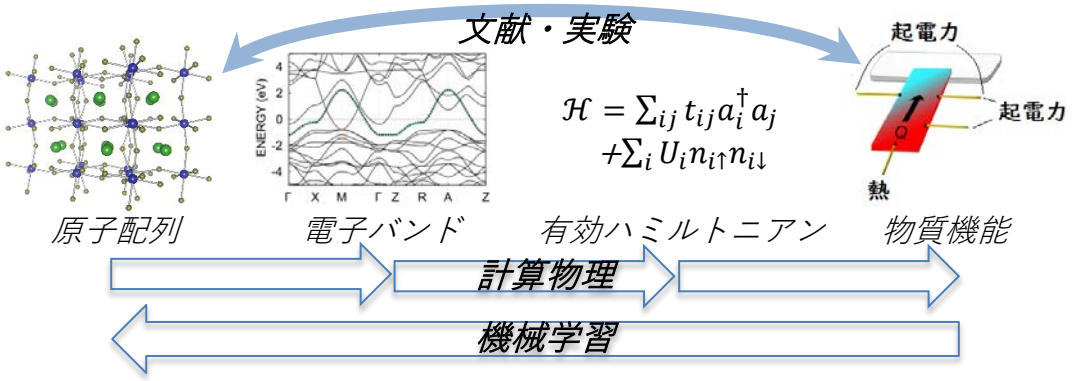
数値化されたグラフ情報



磁性体の特性に着目したクラスタリング情報

- 計算物理によるシミュレーションデータ

- 計算物理と機械学習の連携による物性予測




物質・材料研究機構のマテリアルデータベースを活用。

①の取組で開発する科学研究向け生成AIモデルのベース

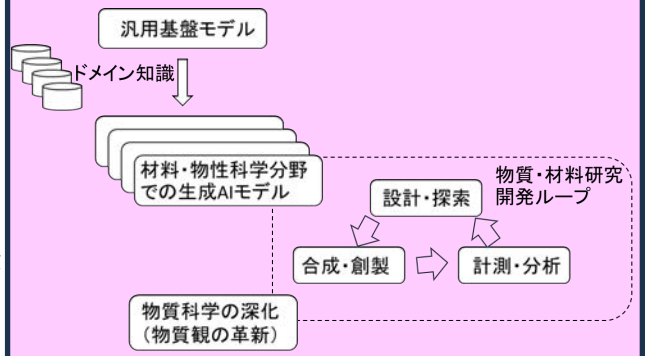
計算物理と連携（物理的な法則等に基づく正確な物性予測）

## 材料・物性科学分野指向 生成AIモデル開発

◆ 目指す材料機能を実現するための原子の3次元配列情報のAIモデルによる生成

◆ 提案された候補物質群の作製方法のAIモデルによる生成

→ **目指す機能を実現する材料開発を加速**



# 世界トップレベルの科学基盤モデルの開発・共用を可能とする計算環境と運用体制の構築

## ③ 革新的な計算基盤の開拓

松岡聡
(理研R-CCS)

- ◆ 多様なデータを扱う**科学研究向け生成*AI*モデルの開発・共用（学習・推論）に必要かつ最適な計算環境・体制**をスパコンで培った技術や既存の計算資源も活用し整備。
- ◆ 既存の計算資源と密接に連携し、科学研究向け生成*AI*モデルを効率的に開発・活用するために必要な**ソフトウェアを開発**。
- ◆ 計算資源を、①で開発する自動化技術（実験ロボット）等と連携させ、莫大な量かつ多様なデータをリアルタイムでやりとり可能とし、**高速で科学基盤モデルの開発に必要な試行錯誤（学習・推論）の繰り返しを実現**。

### ハード（計算機）の整備とソフト（ソフトウェア・アルゴリズム）開発の両面が不可欠

高度化　高速化　効率化

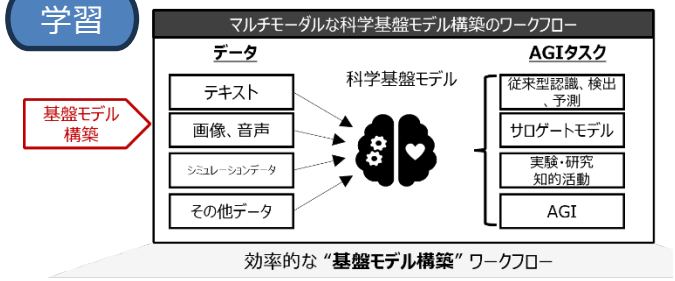
複雑な科学研究へのチャレンジに対応可能な高度な科学研究向け生成*AI*モデルの開発・共用（学習・推論）のため

- **大規模な推論と学習を並行して実施できる最適な計算環境と運用体制を構築・活用**

  既設の計算資源に密接に連携し、科学研究向け生成*AI*モデルに適した計算環境（＝密結合型科学研究向け生成*AI*モデル用計算環境）を経済的合理性を担保したうえで実現。科学研究データの生成やシミュレーションによる検証等と、多様な科学研究データの学習を並行して実施。

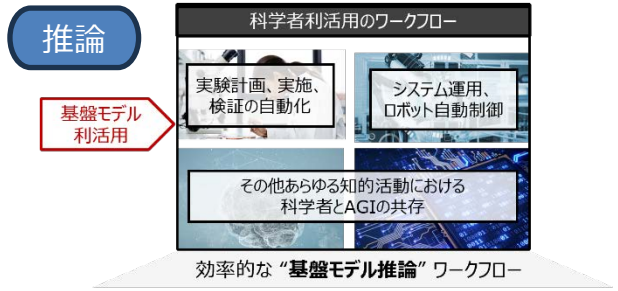
- 科学研究向け生成*AI*モデル開発環境の最適化に**必要な基盤的ソフトウェアを開発**
  - 計算資源の特性も踏まえた学習手法の高度化
  - 科学研究向け生成*AI*モデルの構築－管理－高速化のためのデータ管理の実現
  - 分野・目的毎のワークフローの最適化・効率化による処理の高速化　等

学習




科学研究向け生成*AI*モデル
に必要かつ最適な計算環境

推論



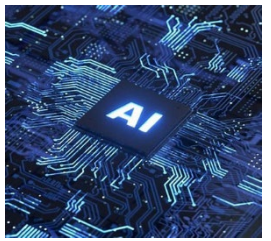
# 次世代のAI向けアーキテクチャの探索

RIKEN TRIP

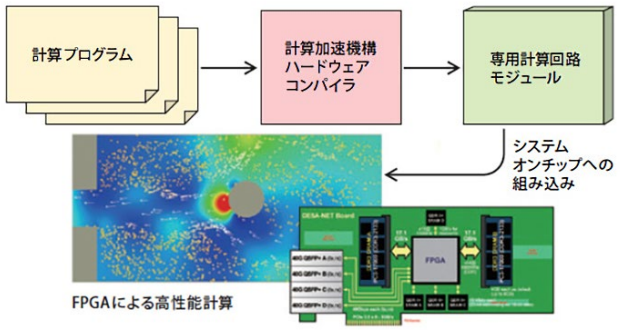
- AIの学習・推論のためのプロセッサアーキテクチャのコ・デザイン
- ２つの方向性で探索:
  1. エッジ向けを含む低消費電力プロセッサ

     Spiking NN etc.
  2. AIの信頼性向上

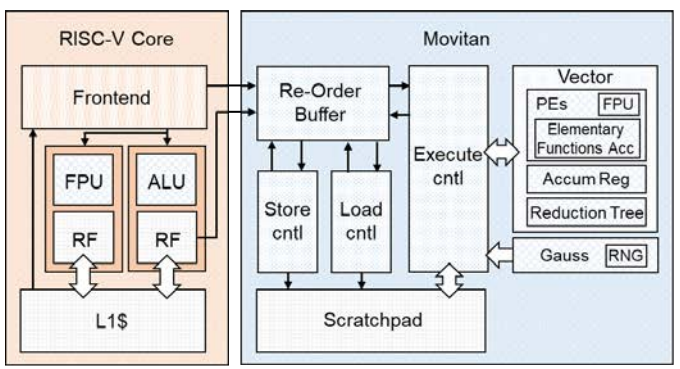
     Baysian Neural Network etc.

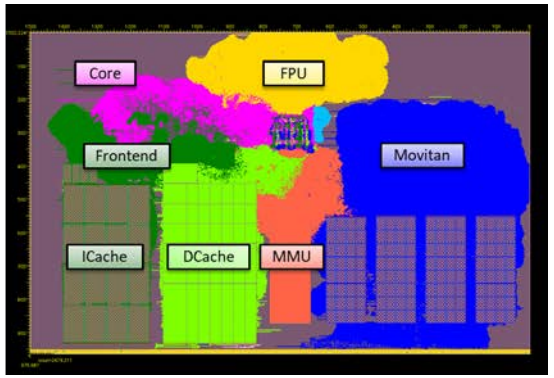



佐野健太郎
(理研R-CCS)




*Fast prototyping and evaluation of AI processors*



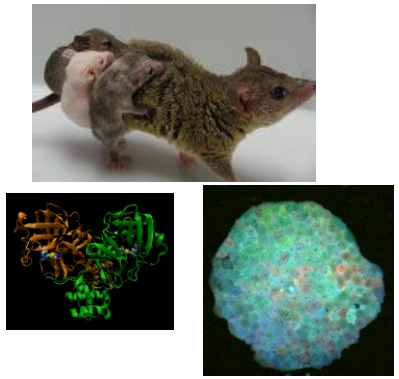

*Accelerators for training of Baysian NN*

# 多階層のマルチモーダル基盤モデル

RIKEN TRIP

- 究極の目標
- マルチモーダル基盤モデルは、科学の長年の課題
  ＝データ統合を実現するためのキーテクノロジ
  - ▷**複雑なシステムを、複雑なままモデル化して予測を可能に**
  - ▷特に生命システムは本来的に複雑性をもつ

**複雑システム**



***Digital Twin***

**マルチモーダル基盤モデル**

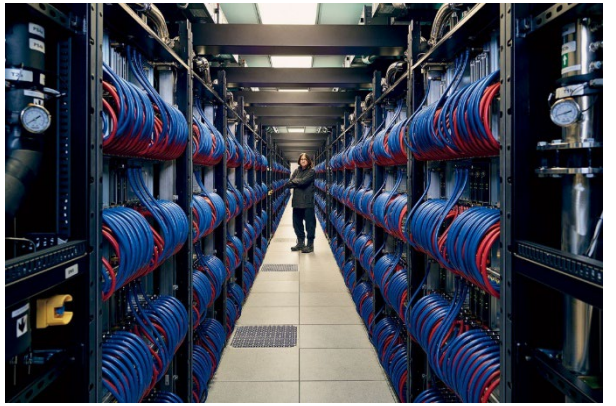
*Generated by DALL-E on Bing for “multimodal foundation model for life science”*

- 生命科学のみでなく、自然科学全般、ひいては社会科学までも統合できる可能性　＝　**学問の再統合**

# TRIP-AGISにおける国際協力

RIKEN TRIP

- 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”においても中心的な役割を果たしている。理研も発起人の一つ。



*Aurora supercomputer*
*https://www.chicagomag.com/*


*https://tpc.dev/*

# TRIP-AGISの今後の計画

RIKEN TRIP

1. FY2025: 最初のドメイン基盤モデルのリリース
2. ANLなど国際連携・国内連携の推進
3. 基盤モデル・研究自動化を活用した科学研究の推進
4. モダリティの追加
   - ▷ 生命科学
     - 医療データ（画像など）
     - オミックスデータの拡大
   - ▷ 材料・物性
     - ソフトマテリアル