

# 生成AIの研究開発について

---

2024年1月  
文部科学省  
基礎・基盤研究課

# 生成AIの開発力強化と人材育成の推進

令和6年度予算額（案） 24億円  
令和5年度補正予算額 377億円  
※運営費交付金中の推計額を含む



国民が生成AIに対して感じるリスクの声に応えるとともに、わが国の科学技術の競争力を強化するため

- ① アカデミアを中心としたオープンな生成AIモデル研究開発を通じた**透明性・信頼性の確保によるリスクの軽減** [AI for Society]
  - ② 開発された基盤モデルを活用した、**科学研究向け基盤モデルの開発及び多様な科学分野での利活用** [AI for Science]
  - ③ 若手研究者・博士後期課程学生に対する**人材育成** [Cross AI Talent Development]
- を推進し、生成AIモデルの基礎的な研究開発力を国内で醸成する。

## AI For Society

### 生成AIモデルの透明性・信頼性の確保に向けた研究開発拠点形成

令和6年度予算額（案）： 7億円（新規）  
令和5年度補正予算額： 42億円

- ✓ 創発性の観測される規模の基盤モデルに関する研究開発を通じ、一連の構築プロセスの検証を実施するとともに、知見・ノウハウを共有。
- ✓ 例えば、学習コーパスが明らかなモデルに対し、コーパス検索機能を活用して入出力関係を分析するなど原理解明等の研究開発を実施。



## AI For Science

### 科学研究向け基盤モデルの開発・共用 (TRIP-AGIS)

令和6年度予算額（案）： 17億円（新規）  
※理化学研究所運営費交付金中の推計額  
令和5年度補正予算額： 122億円

- ✓ 特定科学分野（ドメイン）に強い他の研究機関と連携し、理研を中核として、基盤モデルを活用した科学研究向け基盤モデル（科学基盤モデル）を開発。
- ✓ 開発した科学基盤モデルを他のドメインにも展開しつつ、その利用については広く開放し、日本の科学研究を世界に先駆けて革新。



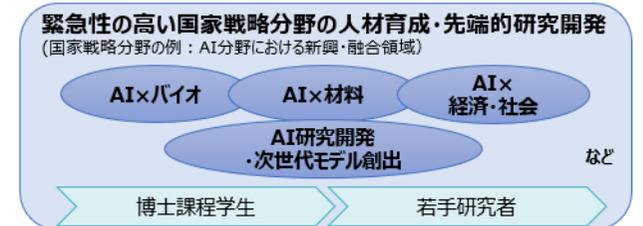
知見・人材・ノウハウ等の共有・活用

## Cross AI Talent Development

### 国家戦略分野の若手研究者及び博士後期課程学生の育成 (次世代AI人材育成プログラム)

令和5年度補正予算額： 213億円  
※JST創発的研究推進基金の積み増し

- ✓ 緊急性の高い国家戦略分野として、AI分野における新興・融合領域を設定し、当該分野の人材育成を推進。
- ✓ 研究者の流動性及び人材獲得力を高めるため、人件費を上乗せ支援。



民間主導の大規模基盤モデル構築に資する環境整備を推進

# 科学研究向け基盤モデルの開発・共用

～ Artificial General Intelligence for Science of Transformative Research Innovation Platform (TRIP-AGIS) ～

令和6年度予算額（案） 17億円（新規）  
※運営費交付金中の推計額  
令和5年度補正予算額 122億円



- **特定科学分野（ドメイン）に強みを有する研究機関と連携体制を構築し、基盤モデルを活用して、科学研究データを追加学習（マルチモーダル化）等することで、ドメイン指向の科学研究向け基盤モデル（科学基盤モデル）を開発**
- **開発した科学基盤モデルの利用を産学に広く開放することで、多様な分野における科学研究の革新（科学研究サイクルの飛躍的加速、科学研究の探索空間の拡大）をねらう**

AIに関する暫定的な論点整理  
（令和5年5月26日、AI戦略会議）

【AI開発力】

- AIの研究成果がAI以外の分野の研究開発の加速に寄与することもほぼ確実である。
- 生成AIによって世界の変革がもたらされようとしている中、可及的速やかに生成AIに関する基盤的な研究力・開発力を国内に醸成することが重要である。
- 世界からトップ人材が集まり切磋琢磨できる研究・人材育成環境の構築や産学官の基盤開発力の強化を進めていくことが期待される。

## 良質なデータ

- トレーニングやファインチューニング、インストラクションなどに必要なデータを良質な形で整備
- データを蓄積する関係研究機関と連携
- 特定科学分野：まずは、  
生命・医科学分野（例：薬剤候補の探索や細胞の刺激応答予測、疾患への適応予測）  
材料・物性科学分野（例：材料機能を実現する物質構造やその作製方法の提案）など

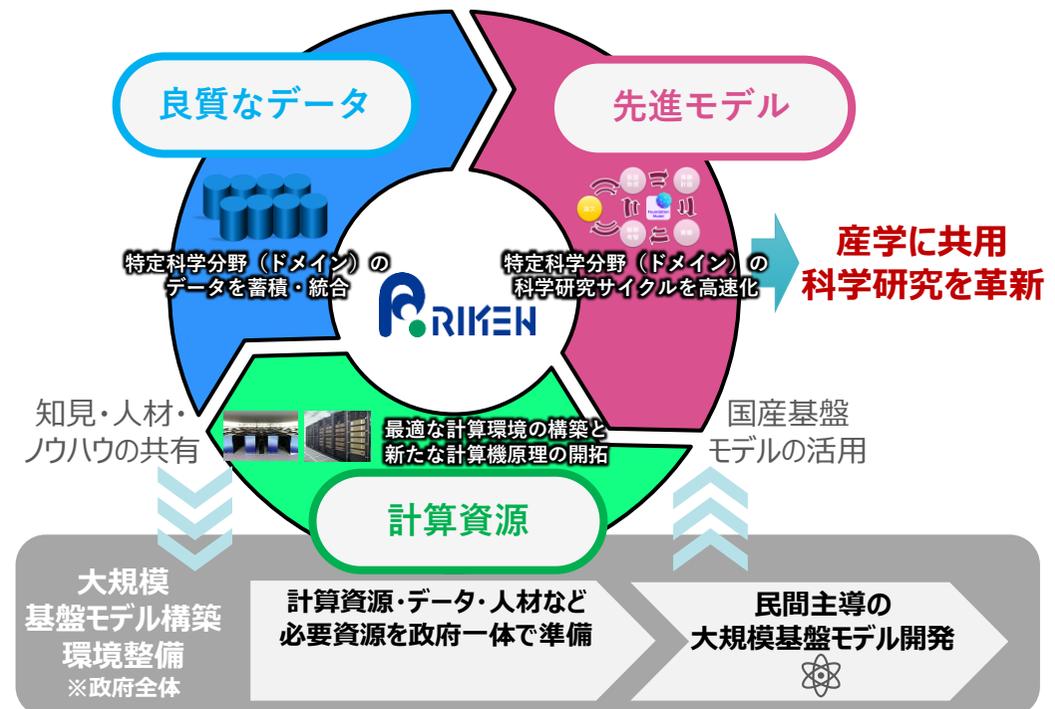
## 先進モデル

- 基盤モデルを活用し、特定科学分野（ドメイン）指向の科学基盤モデルを開発・運用・共用
- 並行して、マルチモーダルデータを読み込・学習・生成するために必要な研究開発

## 計算資源

- スパコン「富岳」の大規模言語モデル分散並列学習手法の開発（実施中）、成果の活用
- 試行錯誤を繰り返して、小規模モデルから徐々に大規模化し、大規模計算時は政府全体として整備する計算資源を活用
- 並行して、「高速」、「セキュア」、「エコ」を実現する革新的な計算資源の研究開発

## “科学基盤モデル”による研究革新



※科学基盤モデル： 基盤モデル（一般文章・画像等）に科学研究データ（科学論文、実験データ、シミュレーションデータ等）を追加学習、推論等させ、科学研究向けに調整した基盤モデルのこと

## ① 科学研究向け基盤モデル開発・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基盤モデルの開発・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、**科学研究向け基盤モデルの開発・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**

# ① 科学研究向け基盤モデル開発・共用の共通基盤技術

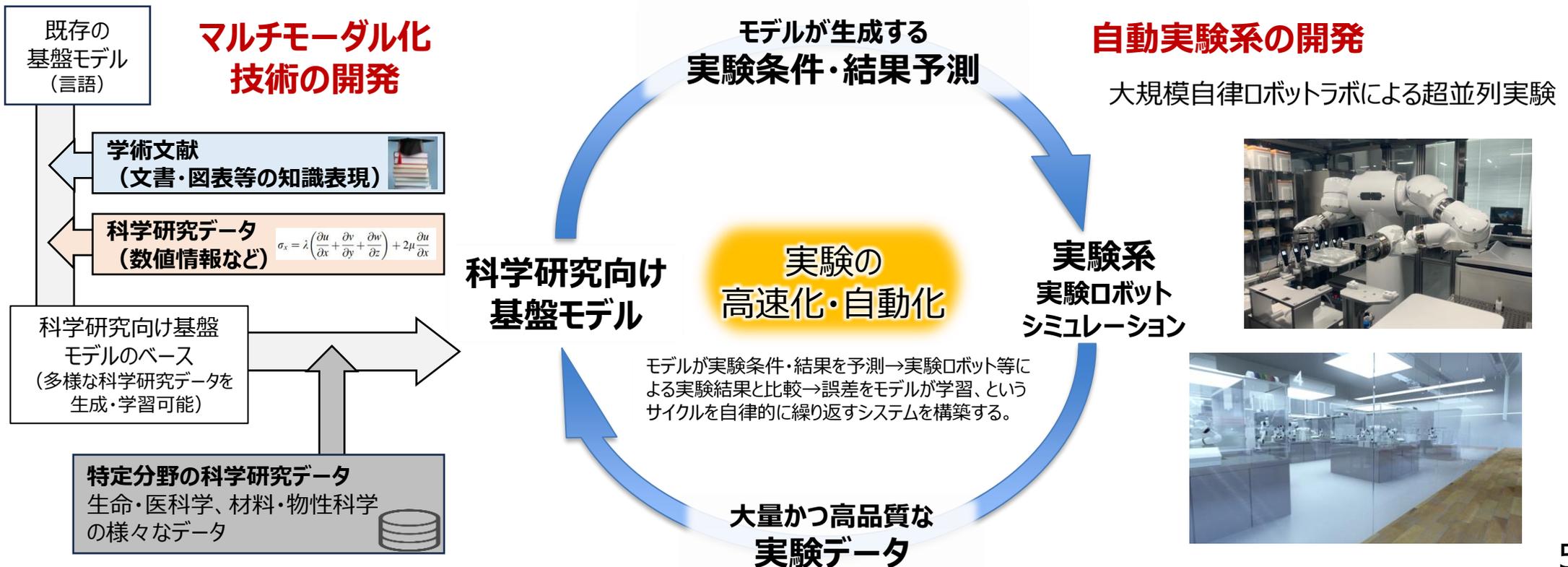
## 目標

- ◆ 多様な科学研究データの学習と生成を可能にする**高度な「マルチモーダル化技術」**を確立。
- ◆ **高品質のデータを大量に創出**して学習を行い、科学研究に活用可能な**高い信頼性**を実現。
- ◆ **自律的・能動的に学習するシステム基盤を開発・提供**し、**科学研究の加速**に貢献。

## 新たな取組

基盤モデルに多様な科学研究データを学習・生成させる手法の開発

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



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

### 目標

- ◆ 遺伝子変異や薬物接種によって**体内で起こる変化や疾患への影響**を予測。
- ◆ 疾患の要因を網羅的かつ効率的に探索し、**創薬プロセスの加速化・医療技術の高度化**を実現。

### 新たな取組

疾患ターゲットの特定と  
新たな創薬リード化合物の発見

薬物等への応答に対する  
細胞動態を予測

動物の行動から疾患を予測

理研内外の既存データ  
化合物・タンパク質データ

薬物刺激等に対する  
細胞内応答データ

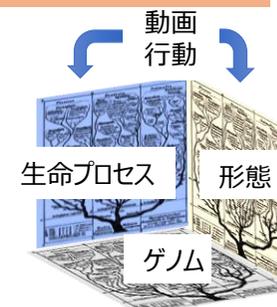
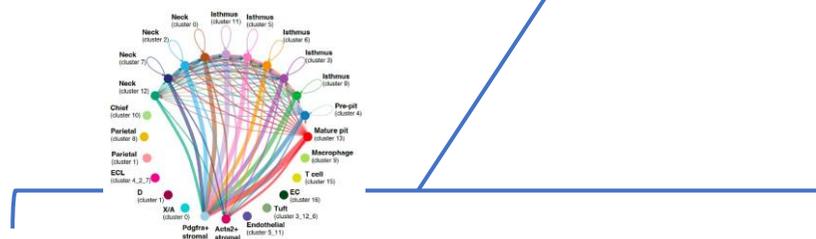
疾患モデル動物等の  
行動・特性データ

分子、細胞、個体の3つの階層に分けてモデルを構築し、統合

分子レベル創薬モデル

細胞レベル応答モデル

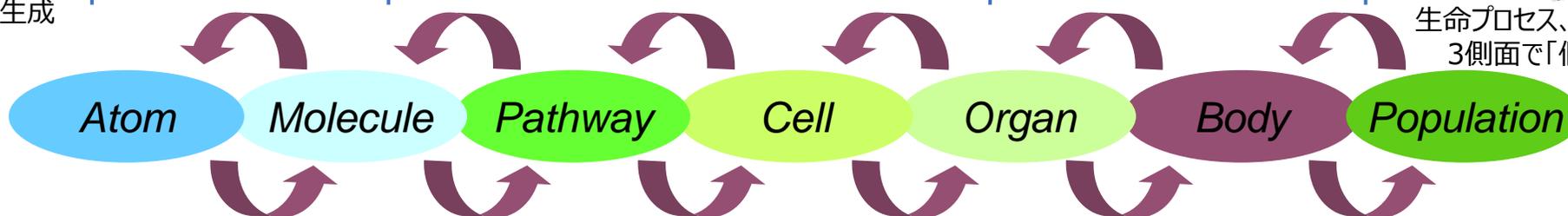
個体レベル行動・特性モデル



創薬ターゲット同定と  
薬物分子の生成

細胞状態の自在な制御を実現

生命プロセス、形態、ゲノムの  
3側面で「個体」を理解



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

**目標** ◆ 目指す材料機能を実現するための**候補物質と合成法**の提案。

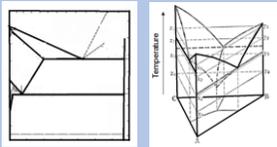
◆ 基礎科学及び産業において、ユーザーが独自データを用いた追加学習を行うことにより、**材料・物性科学研究を加速・強化**。

### 新たな取組

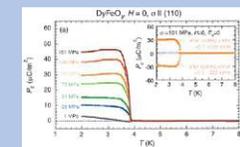
文献・実験データ、  
シミュレーション等による計算データ

所要の物性を持つ材料を提案、  
その合成法を予測

- 物質・材料作製方法に関する文献情報（>10<sup>6</sup>件を想定）

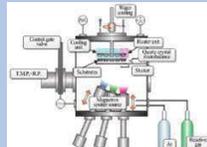


合成に関する情報



数値化されたグラフ情報

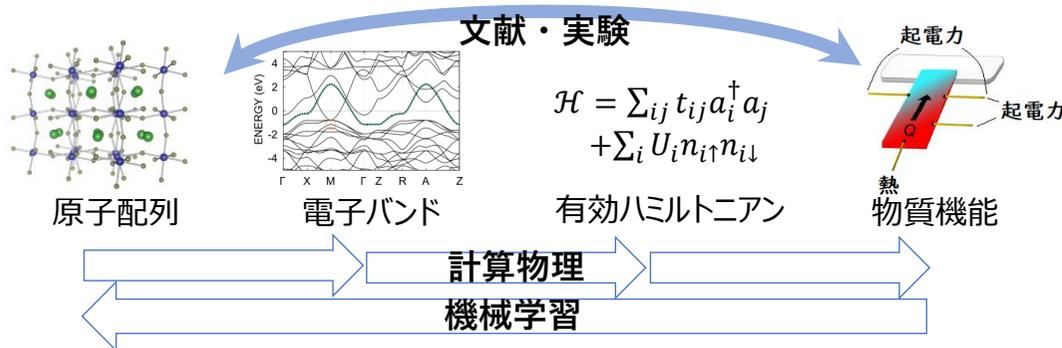
- 新規に取得する良質な実験データ



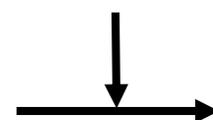
多サンプル同時合成手法による磁性薄膜の物性データ

- 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



①の取組で開発する科学研究向け基盤モデルのベース

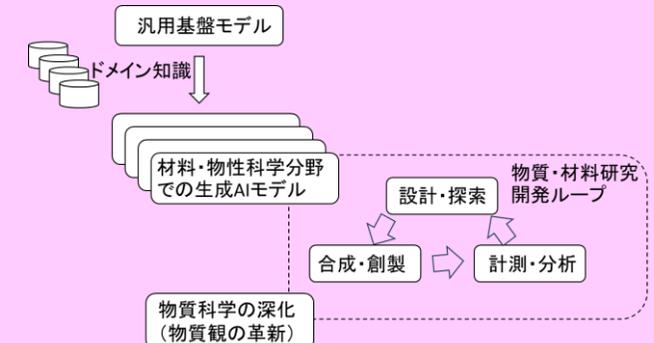


計算物理と連携（物理的な法則等に基づく正確な物性予測）

### 材料・物性科学分野指向 基盤モデル開発

- ◆ 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基盤モデルによる生成

→ 目指す機能を実現する材料開発を加速



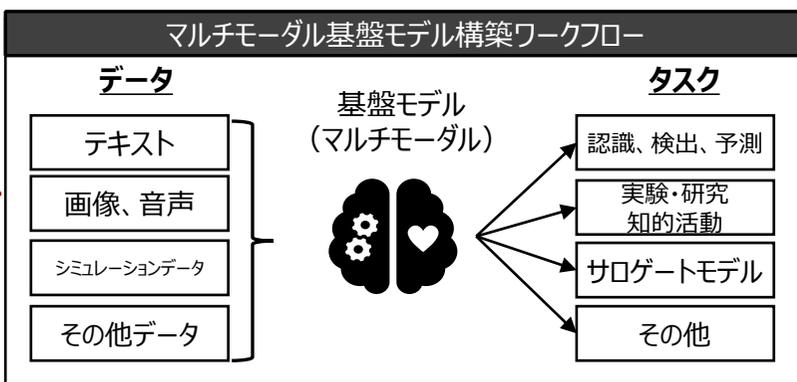
### ③ 革新的な計算基盤の開拓

#### 目標

- ◆ 科学研究向け生成AIモデルの開発や利活用に最適な計算環境や運用体制を構築。科学研究に革新をもたらし、**新たな価値の創出**に貢献。
- ◆ 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・省エネ化を実現する**新たな計算機原理を確立**し、**信頼性の高いAI開発・利活用につなげる**。

#### 新たな取組

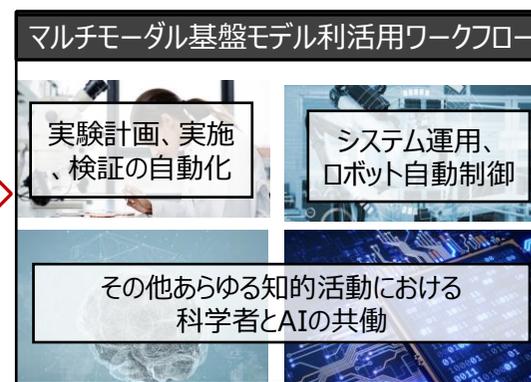
##### 学習



基盤モデル構築

効率的な“基盤モデル構築”ワークフロー

##### 推論



基盤モデル利活用

効率的な“基盤モデル推論”ワークフロー

#### AI for Science プラットフォーム

##### 複雑な科学の問題に対応可能な計算環境の構築



富岳

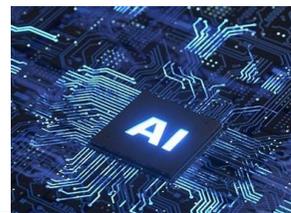
相補的連携



高性能AI計算資源

科学研究向け基盤モデルのための高性能AI計算資源整備

##### AI向けの新たな計算機原理の確立



AI向けプロセッサ

- 生成AIモデル構築と利活用の高度化に向けた新しい方式の研究開発
- AI向け演算処理の高度化・高速化・省エネ化を実現する新たな計算機構成原理の研究開発

AIモデルとAI向け計算機の高度化に向けた研究開発