

文部科学省 生成AI研究開発の方向性について

令和5年 8月10日

研究振興局基礎・基盤研究課

- **生成AI研究開発の方向性について** **1**
- **【参考】科学研究向け生成AIモデルに関して（背景・ねらい等）**
..... **6**

基盤モデル研究開発の推進

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 アカデミアにおけるオープンな基盤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、**科学研究向け生成AIモデルの開発及び多様な科学分野での利活用** [AI for Science]
 - ③ 若手研究者等に対する**人材育成** [Cross AI Talent Development]
- を推進し、基盤モデルの基礎的な研究開発力を国内で醸成する。

AI For Society

基盤モデルの透明性・信頼性の確保と高度化事業（仮称）

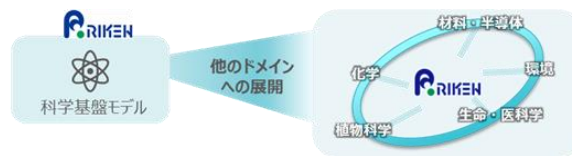
- ✓ 創発性の観測される規模の基盤モデルに関する研究開発を通じ、一連の構築プロセスの検証を実施するとともに、知見・ノウハウを共有。
- ✓ 例えば、学習コーパスが明らかなモデルに対し、コーパス検索機能を活用して入出力関係を分析するなど原理解明等の研究開発を実施。



AI For Science

科学研究向け生成AIモデルの開発・共有（TRIP-AGIS）

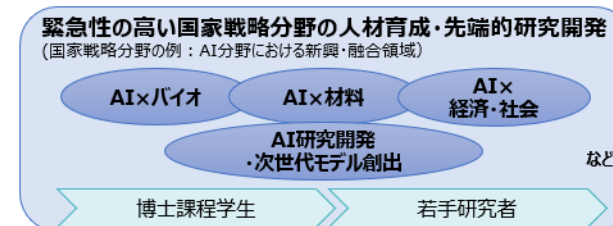
- ✓ 特定科学分野（ドメイン）に強い他の研究機関と連携し、理研を中核として、基盤モデルを活用した科学研究向け生成AIモデル（科学基盤モデル）を開発。
- ✓ 開発した科学基盤モデルを他のドメインにも展開しつつ、その利用については広く開放し、日本の科学研究を世界に先駆けて革新。



Cross AI Talent Development

次世代AI人材育成プログラム（仮称）

- ✓ 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。



知見・人材・ノウハウ等の共有・活用

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

AI for Society : 新技術によるリスクに対しては、新技術で対応

「AIに関する暫定的な論点整理」にリスクとして記載されている生成AIの透明性と信頼性を確保するため、基盤モデルの原理解明を進めるとともに、学習技術を高度化し、AI 開発者・サービス提供者・企業や個人が、安心して開発・提供・利用できる環境を整える必要がある。

AIに関する暫定的な論点整理（抜粋）

令和5年5月26日 AI戦略会議

3-1 リスクへの対処

リスクへの対応を考える際に、
まずAIの透明性と信頼性を確保することが重要

透明性

AIがどのようなデータを学習しているのか、
学習データをどのように作成しているのか、
どのような手法で回答を作成しているのか、

AIの透明性を高めることにより、

使用目的に対して適切なAIを選択することが出来る
ほか、問題が生じた場合の対処が容易となる

信頼性

AIが誤った回答をしていないか、
AIとの対話によって機密情報が漏洩しないかなど

AIの信頼性への不安がある

懸念されるリスクの具体例

① 機密情報の漏洩や個人情報の不適正な利用のリスク

AIとの対話によって利用者の機密情報の漏洩や、個人情報の不適正な利用のリスクが指摘されている。

AIが特定の個人について、個人情報を探索・収集・分析したり（不適切なプロファイリング）、個人に関する不適切な情報を出力する可能性もある。

この問題に関しては、AI開発者・サービス提供者はデータの取扱いなどを開示し、透明性、信頼性を高める努力が必要である。（中略）
なお、AIサービス提供者の信頼性の確認は一義的には利用者が行うものであるが、政府も利用の当事者であり、生成AIサービス提供者や利用者のデータの取扱いなどを確認する必要がある。

対応方策例

透明性：原理解明



小規模モデルの構築・実験

創造性が確認できる規模のオープンなモデルを構築し、ブラックボックスとなっている課題に対応



学習理論研究

深層学習の理論構築を通じて、基盤モデル構築までの学習の段階を解剖する

意味の汎化現象の解明

コーパスと入出力の関係やモデル内部での数理的表現の分析を行う中で意味の汎化現象の原理を解明

学習データと入出力の関係

コーパス検索機能を用いつつモデルの入出力を観察することにより、どのように学習コーパス中のテキストが表出されるか、どのような場合にハルシネーションが現れるかを解明

信頼性：学習技術の高度化



多様な情報への対処

変更されたデータの混入リスクや、異なる時点における異なる事実（時系列情報）への対応など同一事象に対して複数の情報が紐づく問題への対処

学習内容の更新

基盤モデルにおける時間概念の扱いを分析するなど、異なる時間、異なる領域をパラレルに扱うことの出来る学習技術を開発

敵対的学習

改ざんしたデータを予め併せて学習データに用いてモデルを構築するなど、敵対的な攻撃（悪意を持って変更されたデータが混入させられる等）からシステムを守る学習技術を開発

AIアライメント

AIを人間や社会に適応させるための効果的なインストラクションチューニング等の手法の開発



科学研究向け生成AIモデルの開発・共用

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

- **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生成AIモデル（科学基盤モデル）を開発**
- **開発した科学研究向け生成AI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

（2023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・共同開発
- 特定科学分野：まずは、
生命・医学分野（例：薬物等による動的变化・遺伝子変異による差異予測向け）
材料・物性科学分野（例：新奇材料の物性予測向け）など

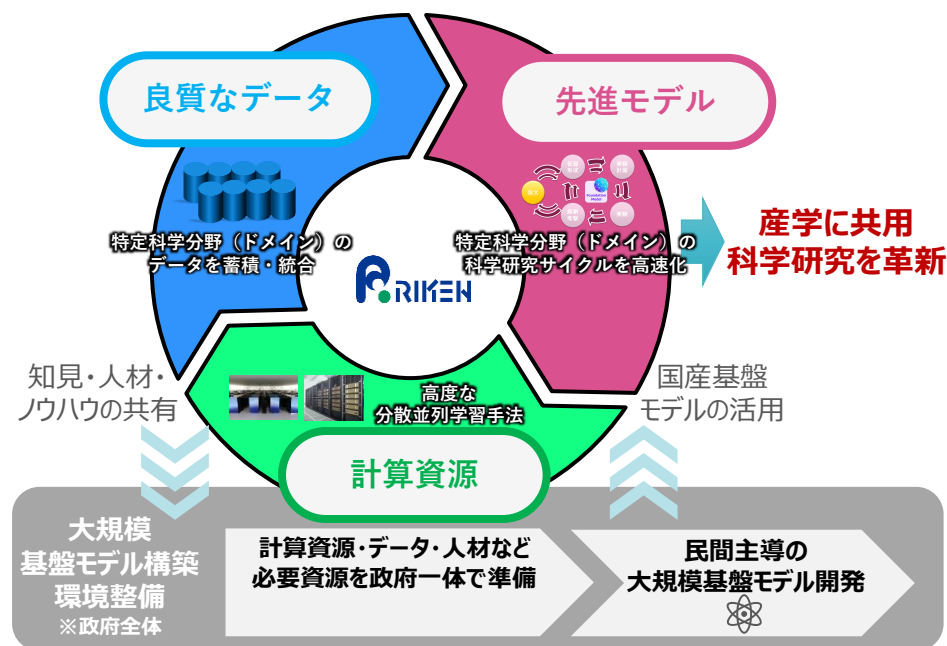
先進モデル

- 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

“科学研究向け生成AIモデル”による研究革新



※科学基盤モデル： 基盤モデル（言語・画像等）に科学研究データ（論文、リアルタイムな実験・シミュレーションデータ等）を追加学習、推論等させ、特定の科学研究分野（ドメイン）向けに調整した基盤モデルのこと

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

背景・課題

- ChatGPTなど、超大規模深層学習で作られた基盤モデルに基づく生成系AIは、人間の知的作業全般に急速な変革をもたらし、産業、研究開発、教育、創作など様々な分野に幅広く波及してきている。経済安全保障や科学研究の国際競争力も左右することから、米国をはじめ各国において国家戦略・政策の検討が急速に立ち上がっている。
- 我が国においても、このような国家戦略分野において、イノベーション創出や産業競争力強化をはかるため、若手研究者や博士課程学生がオープンな研究環境で活躍できる支援の抜本的な拡充が必要。

【政策文書等における関連記載】
AI戦略会議の論点整理（令和5年5月26日）
・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。

事業概要

【事業の目的・目標】

- 緊急性の高い国家戦略分野として、AI分野及びAI分野における新興・融合領域（クロスAI研究分野）を設定。
- 当該分野の人材育成及び先端的な研究開発を推進。

【支援イメージ】

○ 支援対象

1. 若手研究者

国家戦略分野におけるオールジャパンの基盤構築・研究力向上に大きく貢献する大学等における独立した/独立が見込まれる研究者

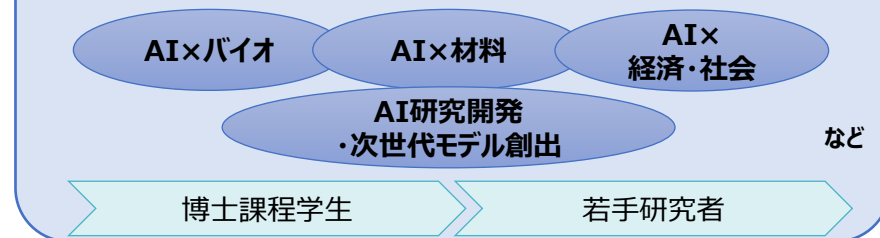
- ✓ 研究費：直接経費＋間接経費（独立支援も検討）
- ✓ 人件費：研究者の流動性及び人材獲得力を高めるため、人件費を上乗せ支援

2. 博士課程学生

国家戦略分野における博士課程学生

- ✓ 支援内容：当該分野の博士号取得を促す観点から、十分な研究費や生活費を支援

緊急性の高い国家戦略分野の人材育成・先端的な研究開発 (国家戦略分野の例：AI分野における新興・融合領域)



【事業の特徴】

- 緊急性の高い国家戦略分野への挑戦を志す若手研究者が、**所属機関に関わらず**、最適な場所を求めて**自由に独立して**研究に従事し、**ステップアップ**できる環境を構築（クロスアポイントメント制度の最大活用）
 - ✓ 自身が持つ高い専門性（バイオ、材料など）を活かしつつ、それを超えて国家戦略分野にチャレンジする意欲を喚起【**異分野融合**】
 - ✓ 産学官のセクターを超えた複数の組織への所属を推奨し、国家戦略分野に従事する人材の流動化を促進【**人材流動化**】

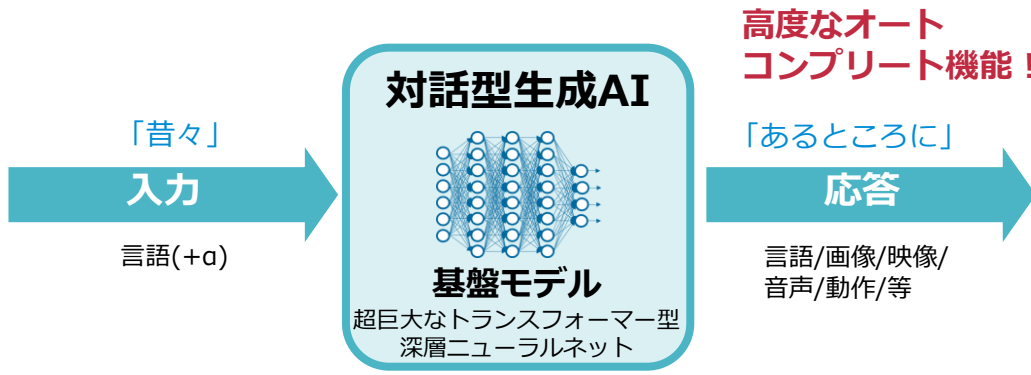
（イメージ例）大学や企業に所属する研究者が、科学研究向け生成AIモデル（科学基盤モデル）を利用し、クロスAI研究を行う際、AI分野の研究機関（理研等）とクロスアポイントメントを行うことを想定

- 国家戦略分野の**研究者層を厚く**するため、同分野に資する研究に取り組もうとする博士課程学生に対して、**十分な研究費や生活費をインセンティブ付与**

- 生成AI研究開発の方向性について 2
- **【参考】科学研究向け生成AIモデルに関して（背景・ねらい等）**
..... **6**

基盤モデルと生成AI

- 対話型生成AIは、与えられた入力の続きを統計的に予測することで応答を生成
- 自己教師あり学習を用いて、大量データからつながり関係を事前学習した超巨大な深層ニューラルネット(汎用性の高い基盤モデル)を用いて予測



対話型生成AIブームが生まれた3つの要因

(1)予測精度の飛躍的向上

トランスフォーマー型の深層ニューラルネットは、幅広いコンテキスト情報を参照しつつ、予測のために注目すべき箇所を適切に選択。さらに、スケールリング則、創発的能力(性能が相転移的に急向上)によりモデルを巨大化するほど精度向上。

(2)対話型ユーザインタフェースの採用

AIコミュニティ内で(1)はGPT-3 (2020年)の時点で大きな話題になっていたが、ChatGPT (2022年)はチャットという一般ユーザに分かりやすく使いやすいインタフェースによって爆発的に利用拡大。

(3)AIアライメント

人間の意図・価値観に合わせてAIを振る舞わせる仕組み(学習データ選別、RLHF:人間のフィードバックを用いた強化学習、Content Moderation)。 ※2016年のMicrosoft Tayのような炎上回避

基盤モデルの代表例

GPT-4 (OpenAI)
PaLM2 (Google)
CLIP (OpenAI)

対話型生成AIの代表例

ChatGPT (OpenAI)
Bard (Google)
DALL-E2 (OpenAI)
Stable Diffusion (Stability AI)

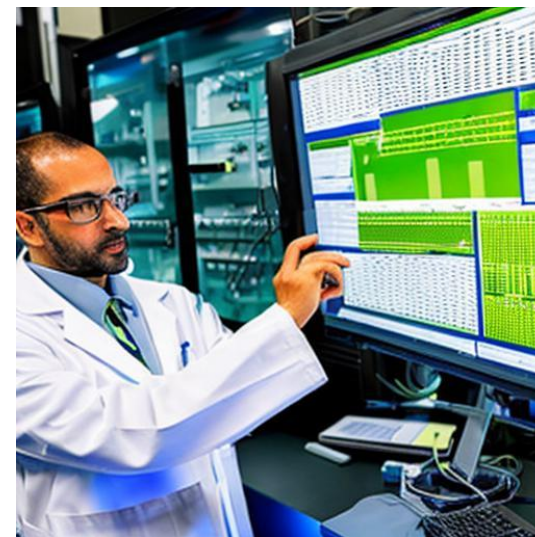
生成AIは高度な知的生産活動（科学研究）にも影響していく

- 生成AIによって、米国労働人口の約19%は仕事の50%以上が影響を受ける可能性があるとしている※。
- 今後AI技術がますます発展していくことを見据えると、**高度な知的生産活動である科学研究も影響を受けるようになる**ことは想像に難くない。
- 現代において産業競争力の根幹となっている科学技術力を維持・向上させるためには、**生成AIを科学研究に上手に活用していくことが重要**。

※T Eoundou, et al., GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models, econ.GN, 2023.

Group	Occupations with highest exposure	% Exposure
Human α	Interpreters and Translators	76.5
	Survey Researchers	75.0
	Poets, Lyricists and Creative Writers	68.8
	Animal Scientists	66.7
	Public Relations Specialists	66.7
Human β	Survey Researchers	84.4
	Writers and Authors	82.5
	Interpreters and Translators	82.4
	Public Relations Specialists	80.6
	Animal Scientists	77.8
Human ζ	Mathematicians	100.0
	Tax Preparers	100.0
	Financial Quantitative Analysts	100.0
	Writers and Authors	100.0
	Web and Digital Interface Designers	100.0
	<i>Humans labeled 15 occupations as "fully exposed."</i>	
Model α	Mathematicians	100.0
	Correspondence Clerks	95.2
	Blockchain Engineers	94.1
	Court Reporters and Simultaneous Captioners	92.9
	Proofreaders and Copy Markers	90.9
Model β	Mathematicians	100.0
	Blockchain Engineers	97.1
	Court Reporters and Simultaneous Captioners	96.4
	Proofreaders and Copy Markers	95.5
	Correspondence Clerks	95.2
Model ζ	Accountants and Auditors	100.0
	News Analysts, Reporters, and Journalists	100.0
	Legal Secretaries and Administrative Assistants	100.0
	Clinical Data Managers	100.0
	Climate Change Policy Analysts	100.0
	<i>The model labeled 86 occupations as "fully exposed."</i>	
Highest variance	Search Marketing Strategists	14.5
	Graphic Designers	13.4
	Investment Fund Managers	13.0
	Financial Managers	13.0
	Insurance Appraisers, Auto Damage	12.6

Table 4: Occupations with the highest exposure according to each measurement. The final row lists the occupations with the highest σ^2 value, indicating that they had the most variability in exposure scores. Exposure percentages indicate the share of an occupation's task that are exposed to GPTs (α) or GPT-powered software (β and ζ), where exposure is defined as driving a reduction in time it takes to complete the task by at least 50% (see exposure rubric A.1). As such, occupations listed in this table are those where we estimate that GPTs and GPT-powered software are able to save workers a significant amount of time completing a large share of their tasks, but it does not necessarily suggest that their tasks can be fully automated by these technologies.



AIと協業する科学者
(Stable Diffusionを用いて作成)

AIの影響予測を大中小に分けたシナリオそれぞれにおいて大きく影響を受ける可能性がある職業

基盤モデル・生成AIの研究開発課題の全体観

基盤モデル応用開発(API利用)

- チャットボット、仮想アシスタント、問い合わせ自動応答、質問応答
- コンテンツ生成(文章、画像、映像)
- 翻訳、要約、ライティング支援
- 企画支援、発想支援 他

分野固有基盤モデル開発・活用

- プログラミング向け基盤モデル
- 個別企業業務向け基盤モデル
- 法業務向け基盤モデル
- 医療・ヘルスケア向け基盤モデル
- 教育向け基盤モデル
- 科学研究向け基盤モデル 他

基盤モデル周辺拡張技術

- 基盤モデルが不得手な機能を扱う外部処理連携(最新情報検索、数式処理、物理シミュレーション、論理推論等)
- 問題解決ワークフロー設計の自動化
- プロンプトやワークフローの最適化 他

利活用時の問題対処

- 生成AIの出力か否かの判定技術(フェイク検出技術、電子透かし技術を含む)
- 入出力データの著作権・肖像権関連問題への対処 他

AIアライメント研究

人間・社会の価値観にAIを整合させる

- 基盤モデル自体の倫理性確保(RLHF等)
- 生成AI応用システムの品質管理(プロンプト型開発法のソフトウェア工学等)
- 人間・AI共生社会のリスク低減(エージェント設計論、トラスト形成等) 他

基盤モデル運用技術

- 継続運用可能なビジネスモデル(ビジネス用途、研究用途)、エコシステム
- データ追加・更新プロセス
- トラストを確保した運営体制 他

基盤モデル構築技術

- 大規模深層学習モデル(トランスフォーマー、マルチモーダル)の実装
- 学習データの収集・選別・整備
- 大規模計算環境構築
- 高速化アルゴリズム、デバイス 他

次世代AIモデル研究

- 基盤モデル高効率化、生成AI高性能化
- 基盤モデルのメカニズム解明
- 人間知能の理解に基づくモデルの探求、基盤モデルとの融合
- 新モデル向けコンピューティング 他

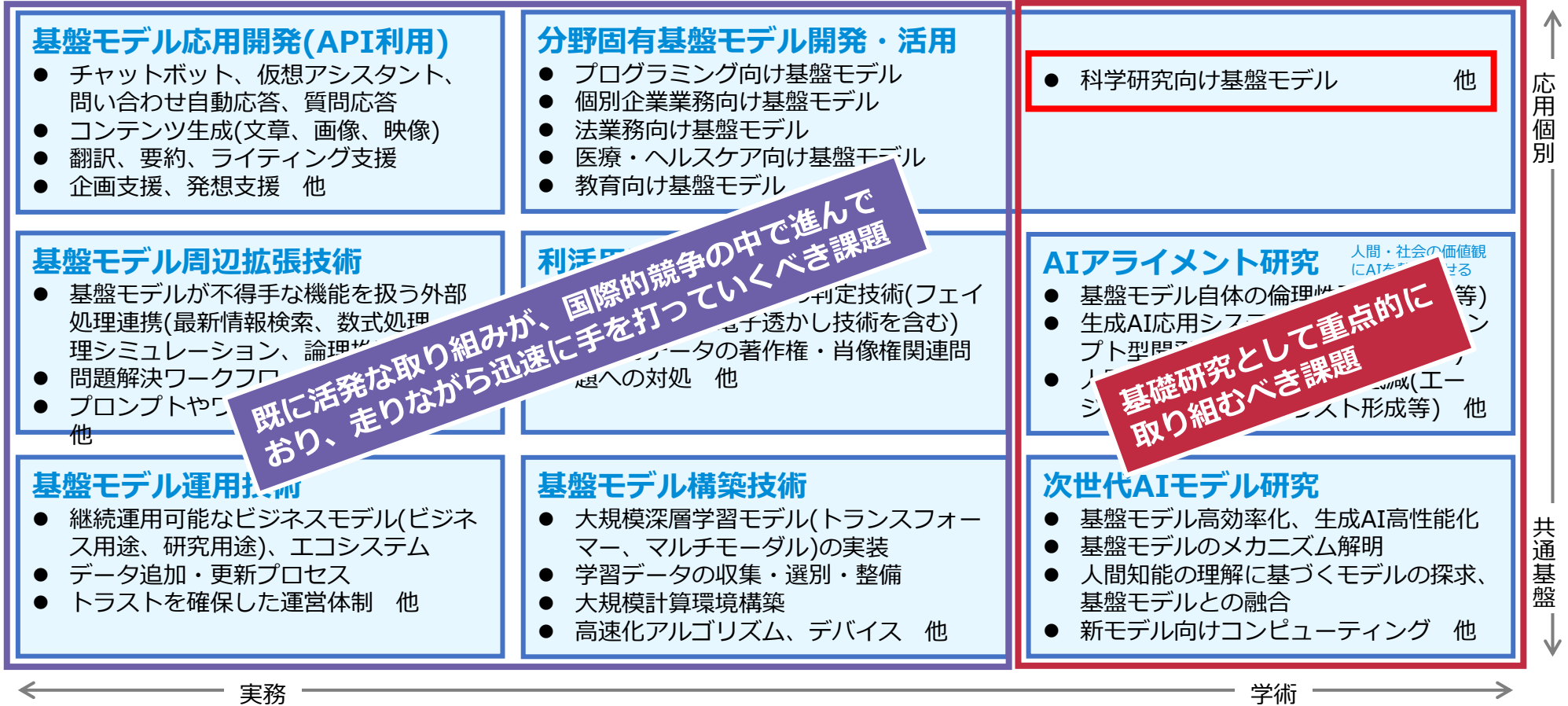
↑ 応用個別

↓ 共通基盤

← 実務

→ 学術

基盤モデル・生成AIの研究開発課題の全体観



既に活発な取り組みが、国際的競争の中で進んでおり、走りながら迅速に手を打っていくべき課題

基礎研究として重点的に取り組むべき課題

基盤モデルを科学研究に活用するためには“マルチモーダル化”が鍵

- 基盤モデルは、その利活用により、科学研究の加速に大きく寄与するため、**鍵となる技術である“マルチモーダル化”**が重要となっている。
- 開発された基盤モデル（言語・画像等）に、**数値情報や配列情報など多様な科学研究データ（新たなモダリティ）**を追加学習させ、いち早く活用可能にすべき。

■ 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



■ マルチモーダル科学基盤モデル

(Multimodal Foundation Model for science)

自然言語

+ 数値情報（科学研究データ）など

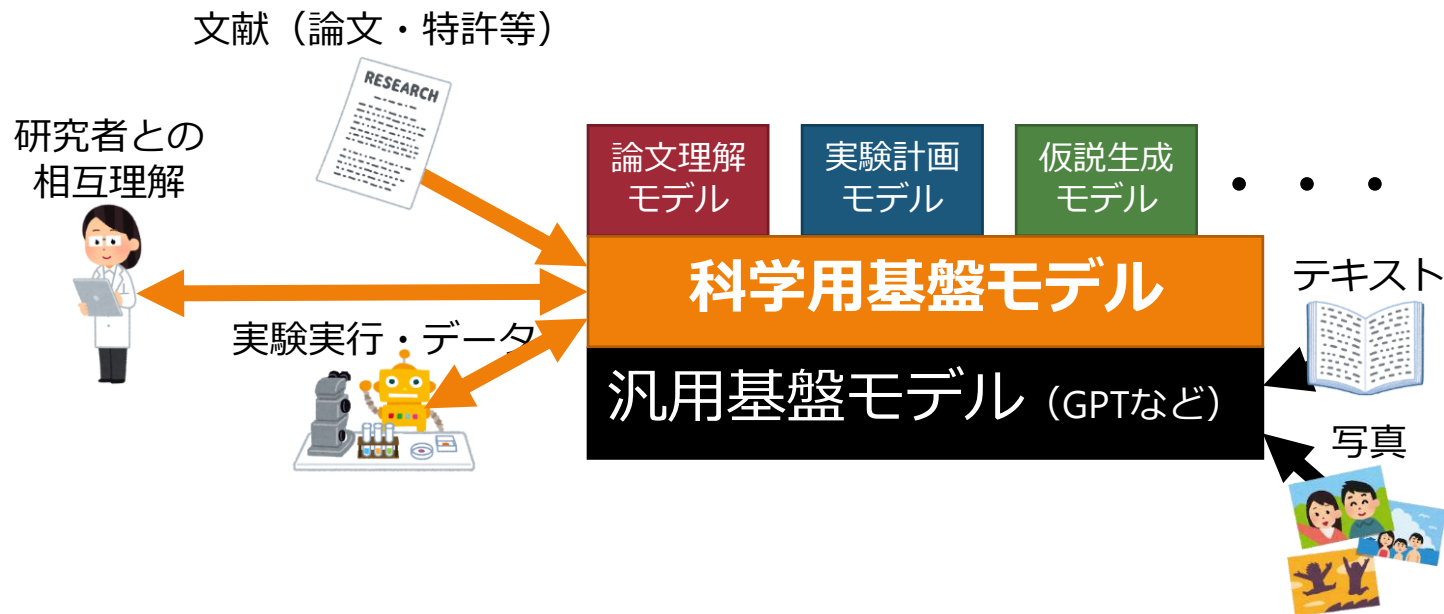


人間が本来苦手な数値情報も加えることで
超知性のような存在に進化可能



AI駆動ロボット科学のための基盤モデル

- 正しい文献を図表ごとと理解し（マルチモーダル）
- 研究者の理解を得られる説明を行い、研究者の教示を賢く学んで（共進化AI; Co-Understandable AI）
- 実際の実験を実行して結果のデータを学ぶ科学用基盤モデルを構築



自動的に科学的原理・解法を発見するAI ロボットのミッション

- 幅広い分野で

→マルチディシプリン

- 実験を行い

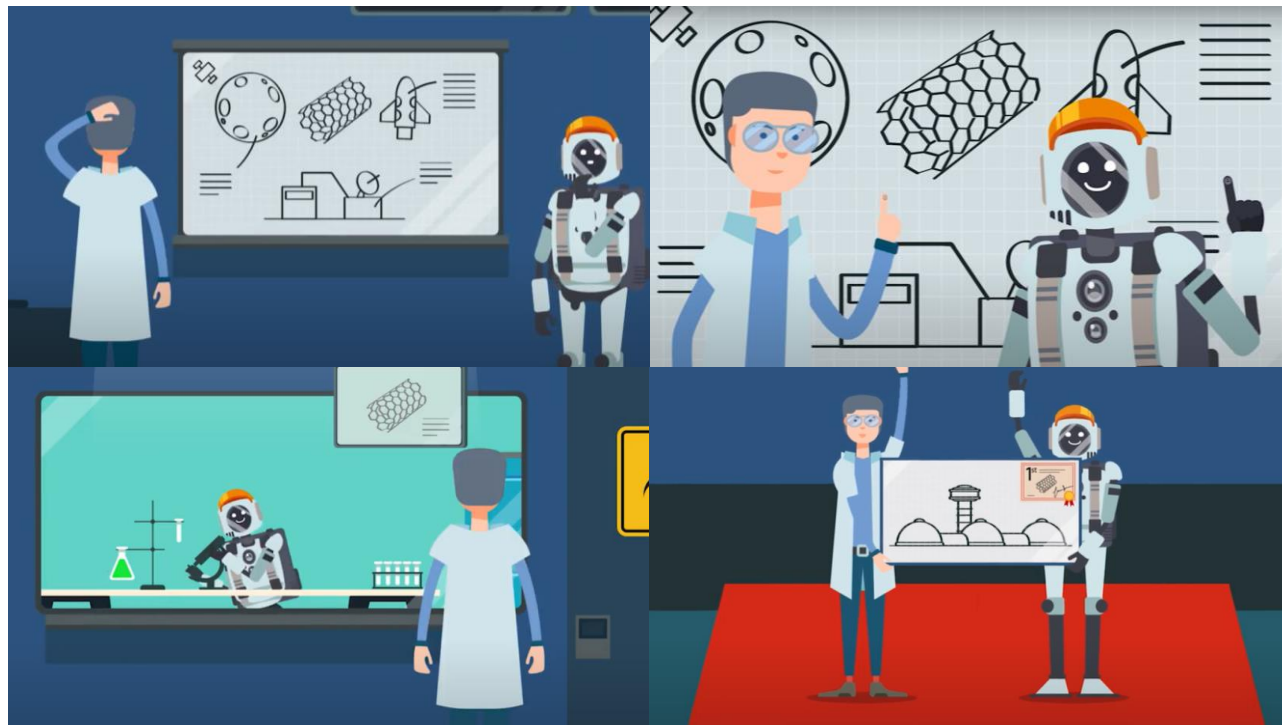
- 自ら計画
- 自律的に実行

→研究の主張と実験

- 法則を見つけ出す

- 実験結果の解析
- 仮説の検証

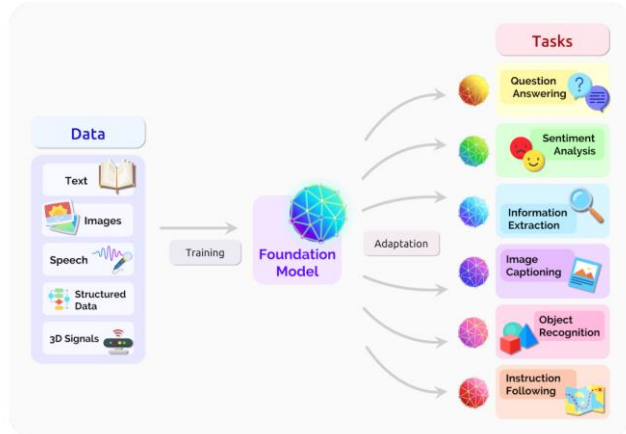
→研究の解析と記述&対話



基盤モデルの科学研究へのインパクト

基盤モデル

再学習（転移学習）により様々なタスクに適応

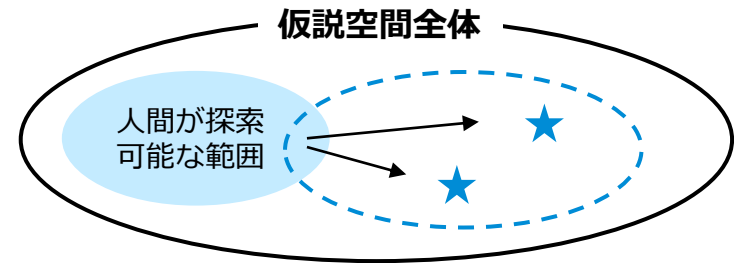


6つの能力



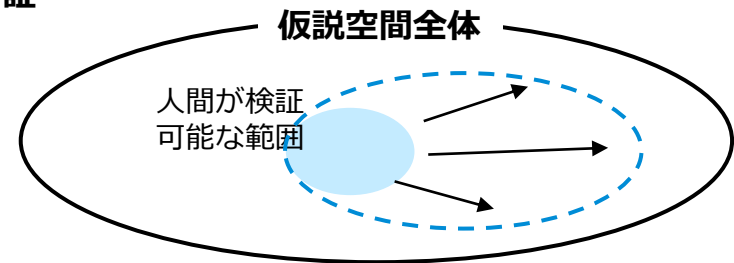
仮説生成・探索

人間の認知能力を超えた大規模・網羅的な仮説探索



実験検証・分析

人間の身体能力を超えたハイスループットな仮説検証

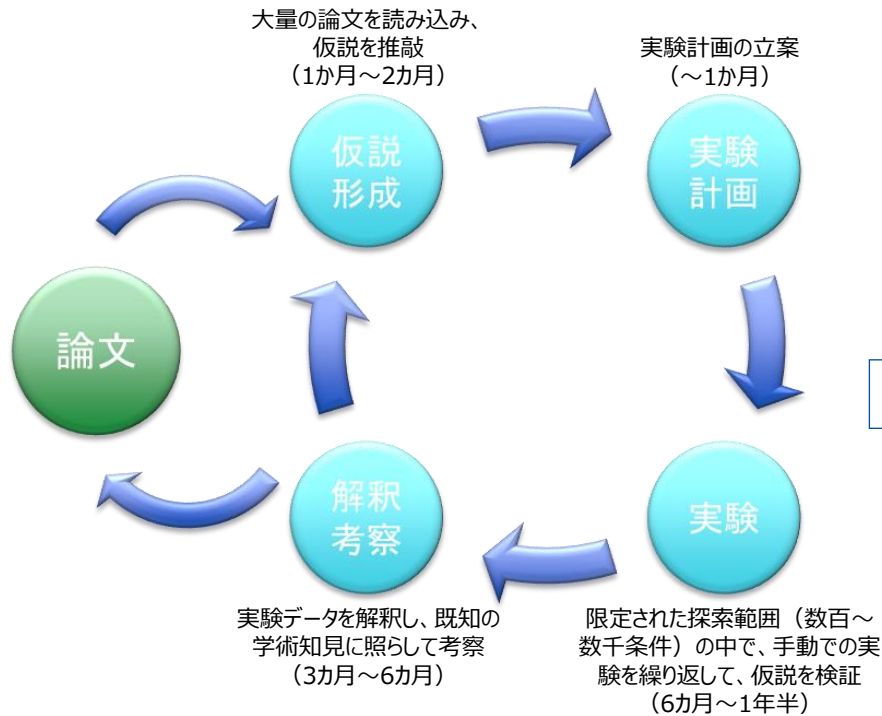


R. Bommasani et al., On the Opportunities and Risks of Foundation Models, arXiv:2108.07258

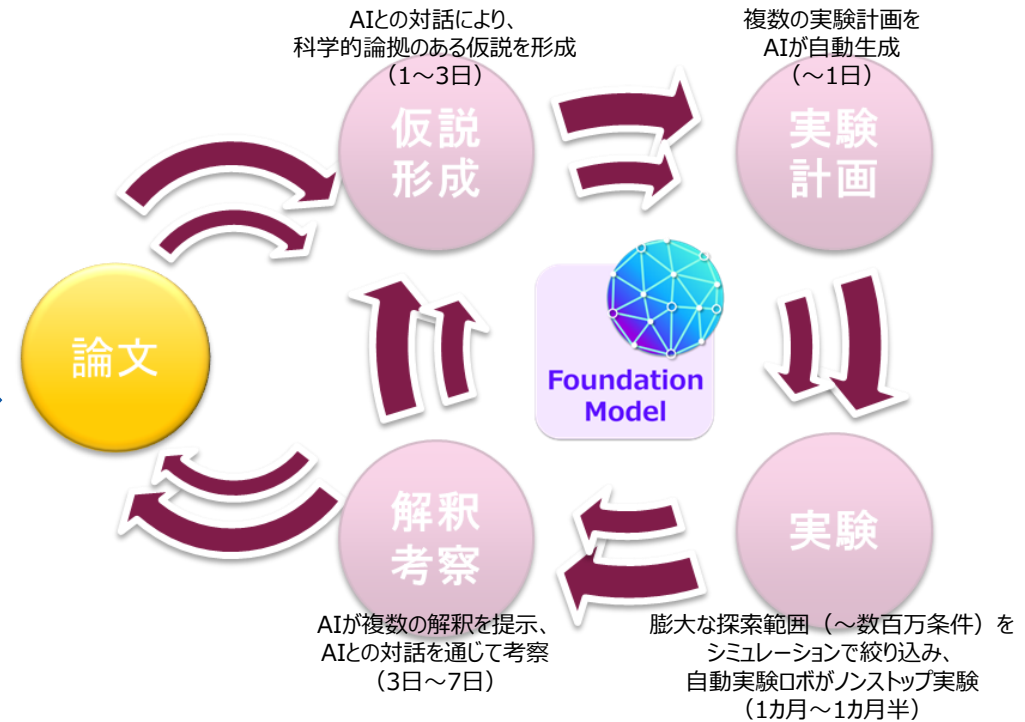
科学研究向け生成AIモデルによる科学研究の劇的な変革

- 科学研究向け生成AIモデル（科学基盤モデル）が確立されると、①**科学研究サイクルが飛躍的に加速**されるとともに、②**科学研究における探索空間も大幅に拡大**される。
- 例えば、生命・医科学分野では、**着想から論文化までの期間が約2年間から約2か月に大幅短縮**され、**科学的探索範囲も約1000倍に拡大**する可能性※。

科学研究サイクル



基盤モデルの活用により大幅加速



科学研究サイクルの加速 (X10) & 科学的探索範囲の拡大 (X1,000) ※

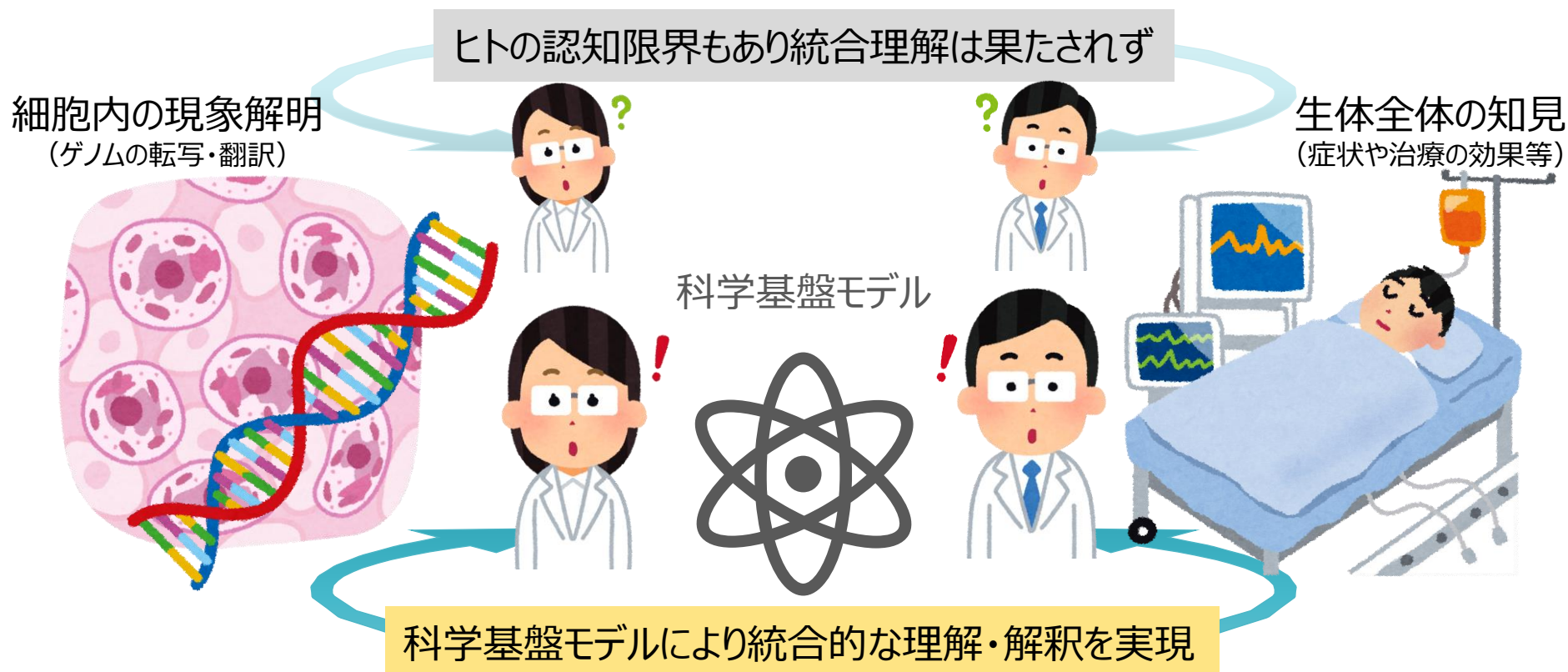
生命・医科学分野

- 生命現象の統合的理解、及びそれに基づく革新的・効率的な治療法の開発

材料・物性科学分野

- 無機・有機・バイオの材料的知見を融合した画期的材料の発掘及び必要とする特性からの材料及びその合成法の提案

生命・医科学分野におけるイメージ



→未開拓の治療法の開発、効果的・効率的な治療法の確立へ

“科学研究向け生成AIモデル”開発の流れ

基盤モデルの場合（ChatGPTを例に）



科学研究向け生成AIモデルへ



※計算資源の目安：大規模→20エクサFLOPS程度（NVIDIA GH100×1万基程度）、小規模→2エクサFLOPS程度（NVIDIA GH100×1,000基程度）