

文部科学省 基礎研究振興部会

基盤モデルとAI・ロボット駆動科学

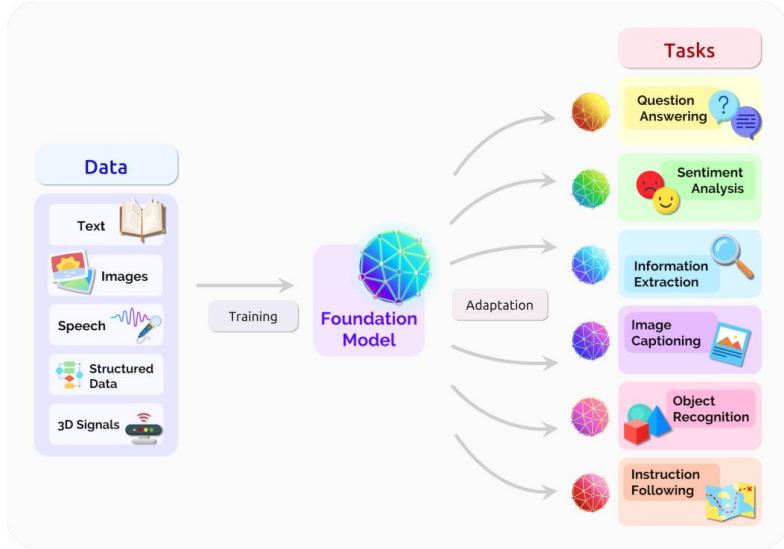
2023年6月21日

科学技術振興機構 (JST)
研究開発戦略センター(CRDS)
嶋田義皓 丸山隆一

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

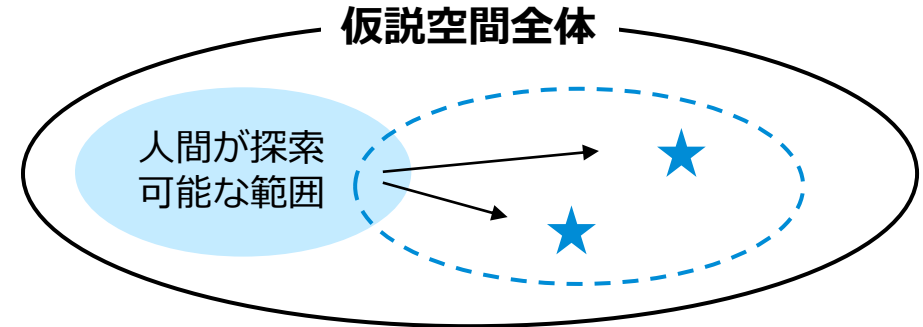


6つの能力



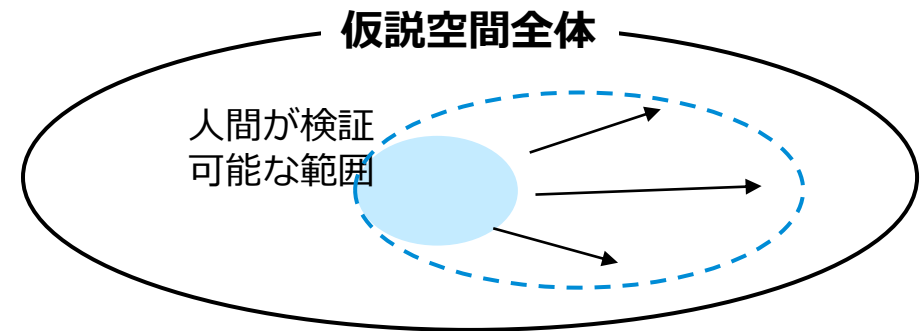
仮説生成・探索

人間の認知能力を超えた大規模・網羅的な仮説探索



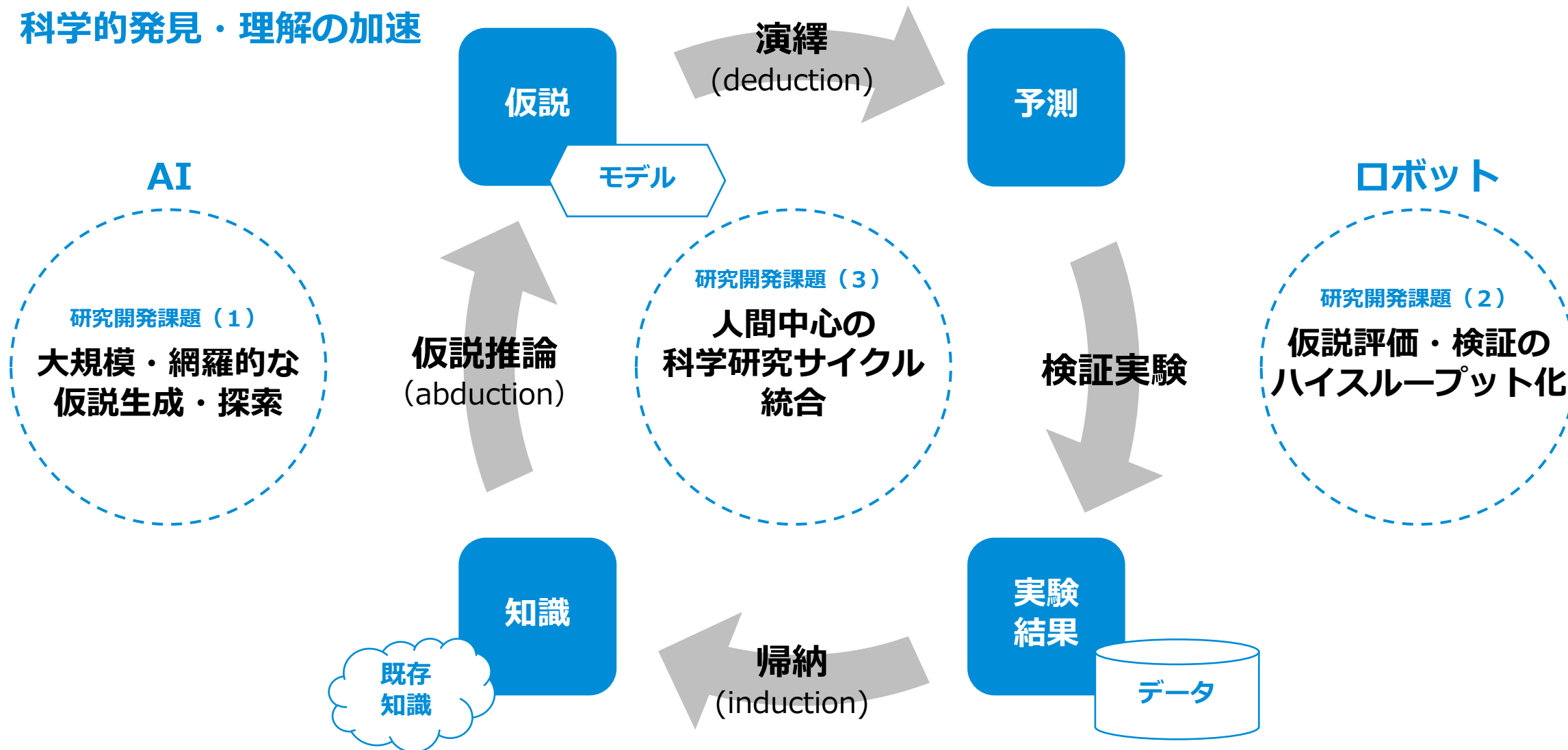
実験検証・分析

人間の身体能力を超えたハイスループットな仮説検証

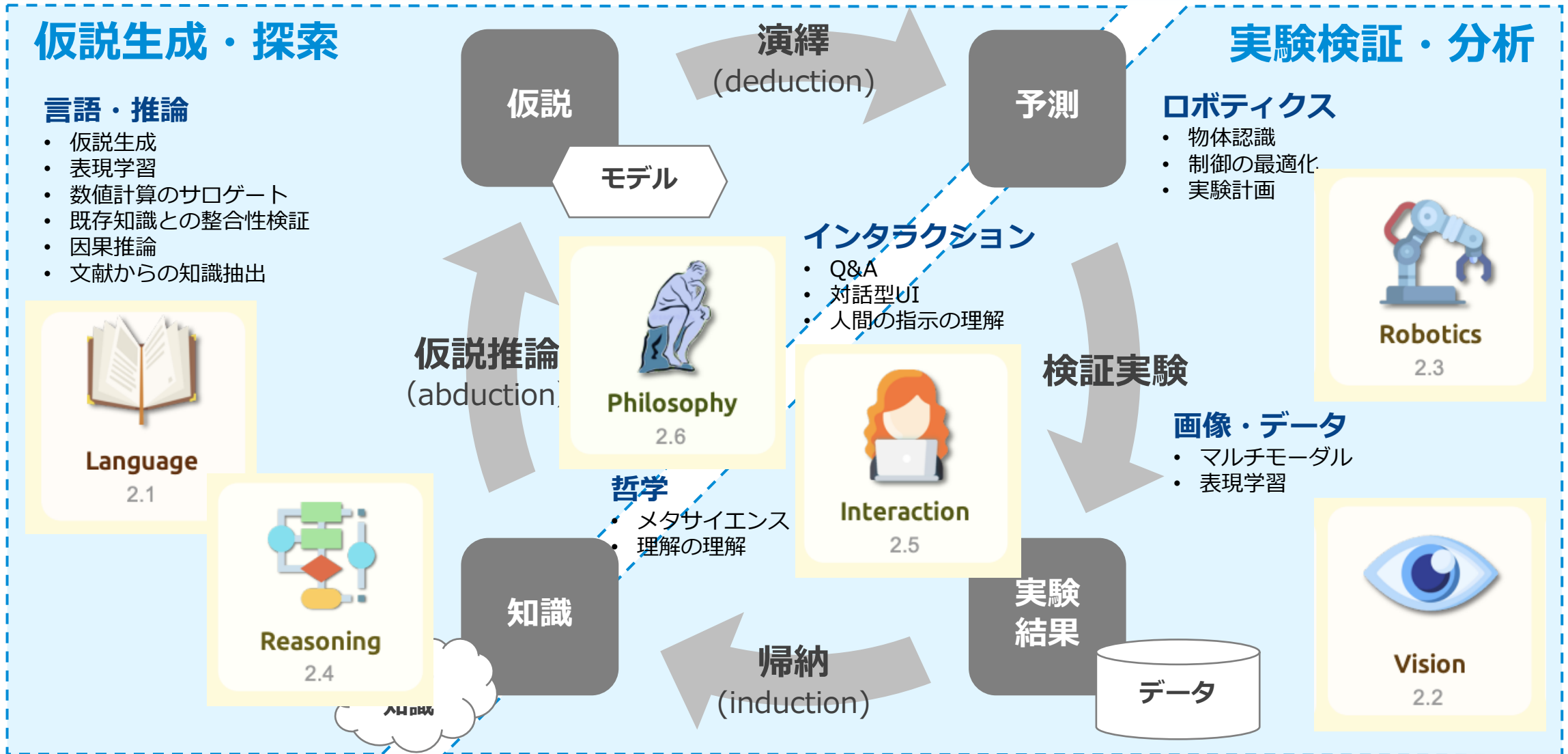


AI・ロボット駆動科学

科学的発見・理解の加速

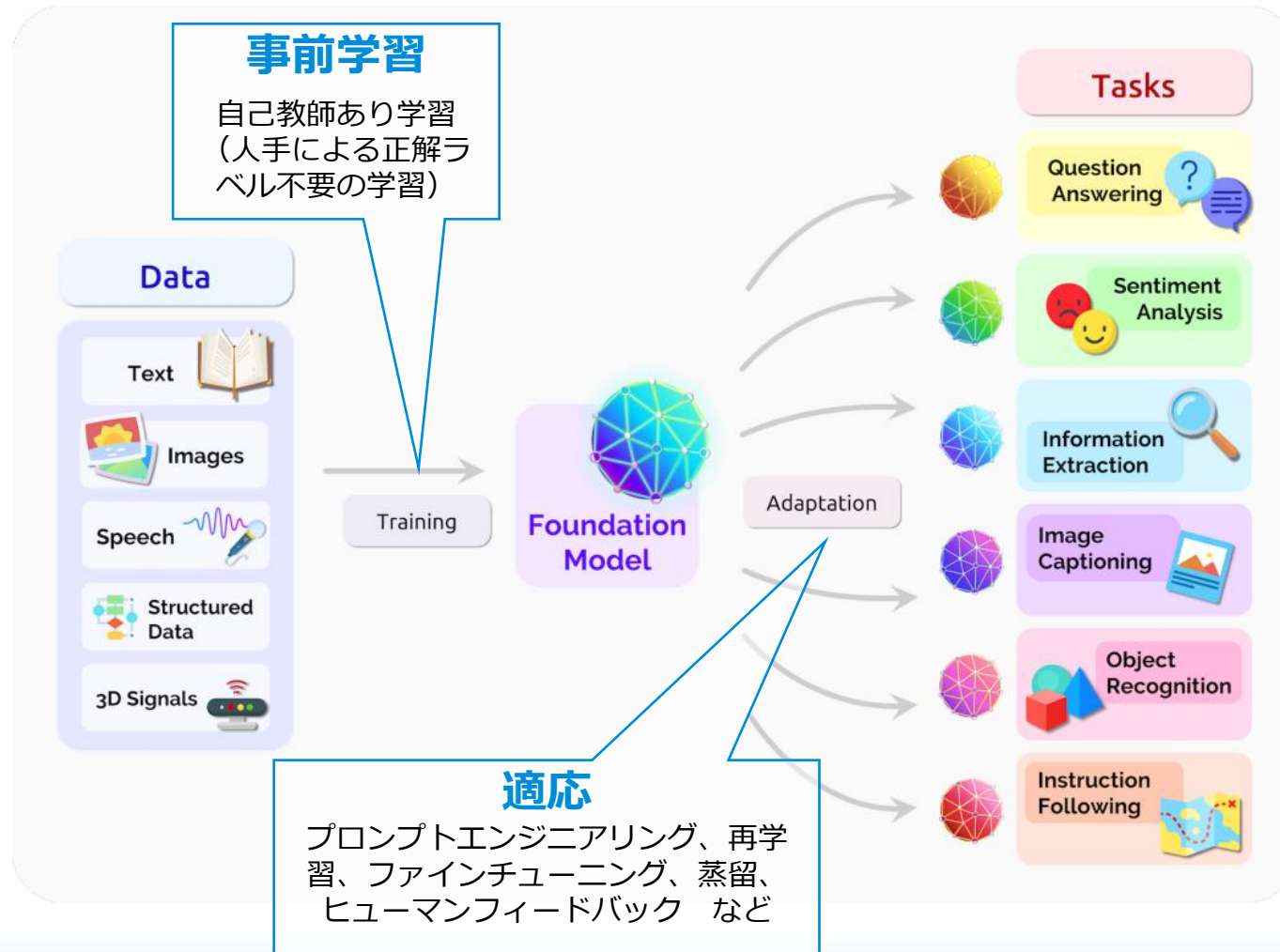


AI・ロボット駆動科学+基盤モデル

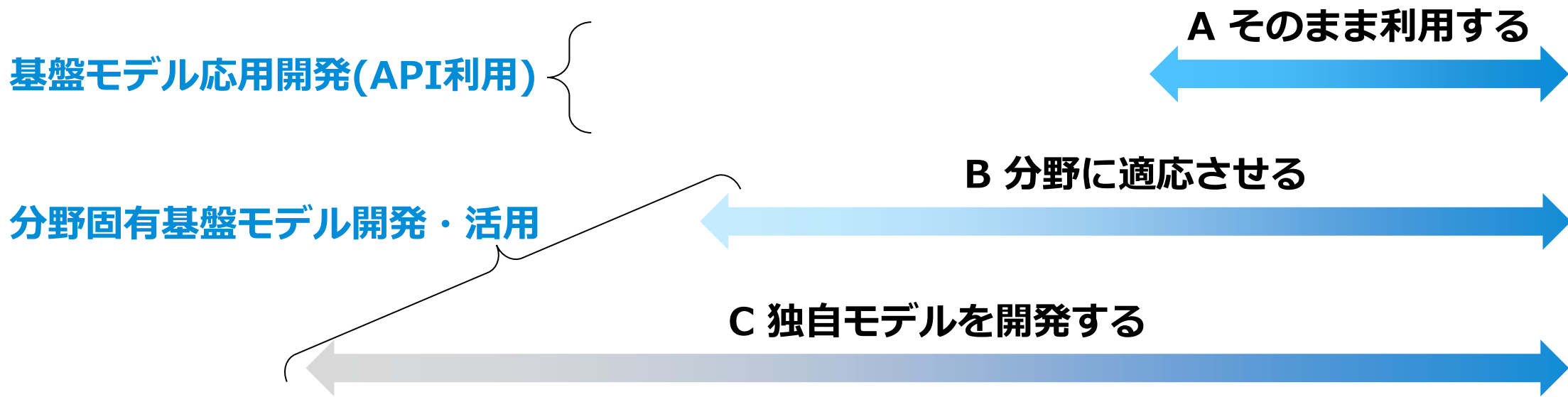
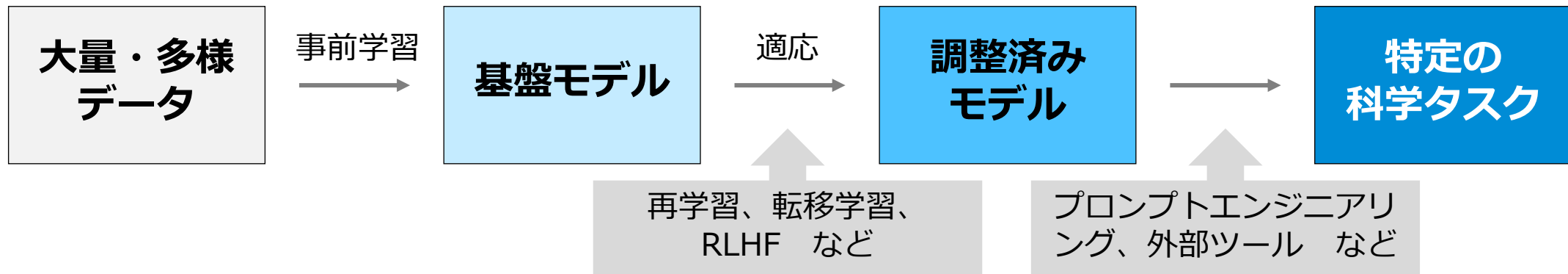


基盤モデルとは

大量・多様なデータで訓練され、様々なタスクに適応できる大規模深層学習モデル

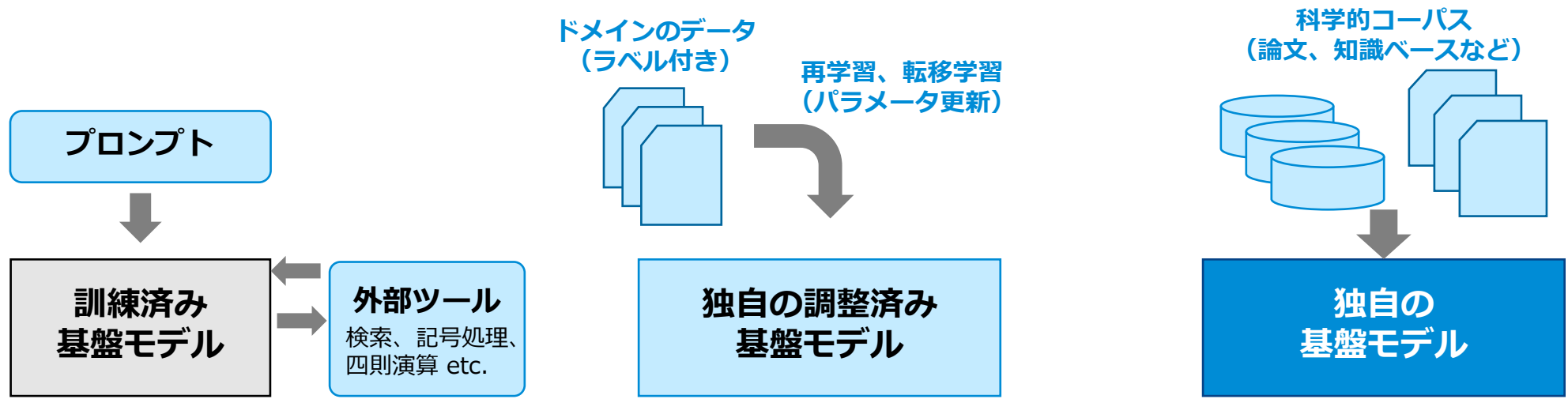


基盤モデルの科学応用の3パターン



基盤モデルの科学応用パターンの比較

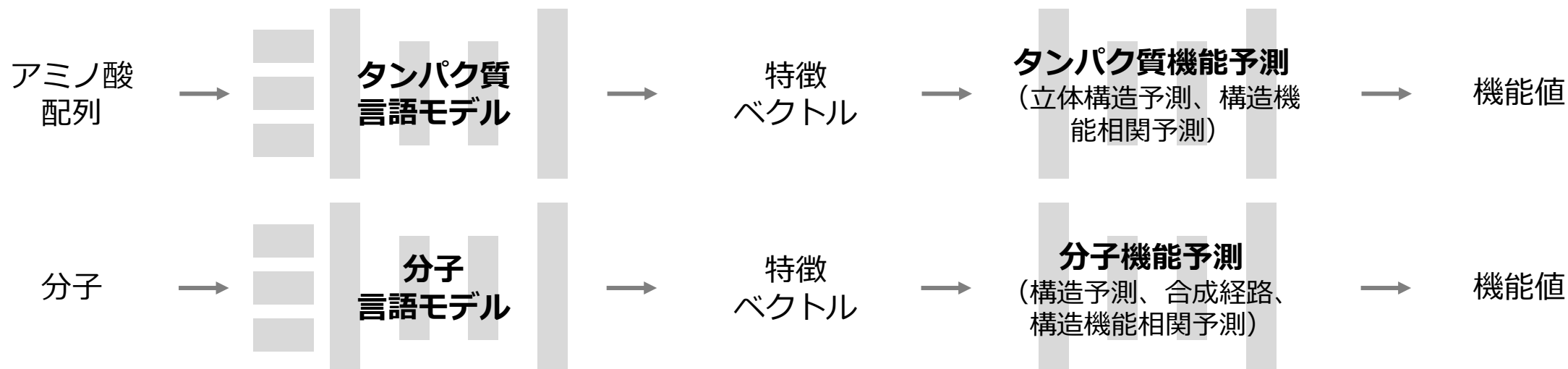
	A そのまま利用する	B 分野に適応させる	C 独自モデルを開発する
モデル	調整済みモデルをそのまま利用 (パラメータ固定)	モデルを修正 (層の追加・削減、 再学習でパラメータを更新)	独自モデルをスクラッチから事 前学習し、再学習もする
適応策	プロンプトエンジニアリング、 外部ツール連携	再学習、転移学習、ファインチューニング、蒸留、RLHF など (基本 的には教師あり学習/強化学習)	
データ	不要 (学習しないので)	再学習用に必要	大量・多様なデータが必要
計算資源	ホストするには必要	再学習に必要。	事前学習には大規模に必要



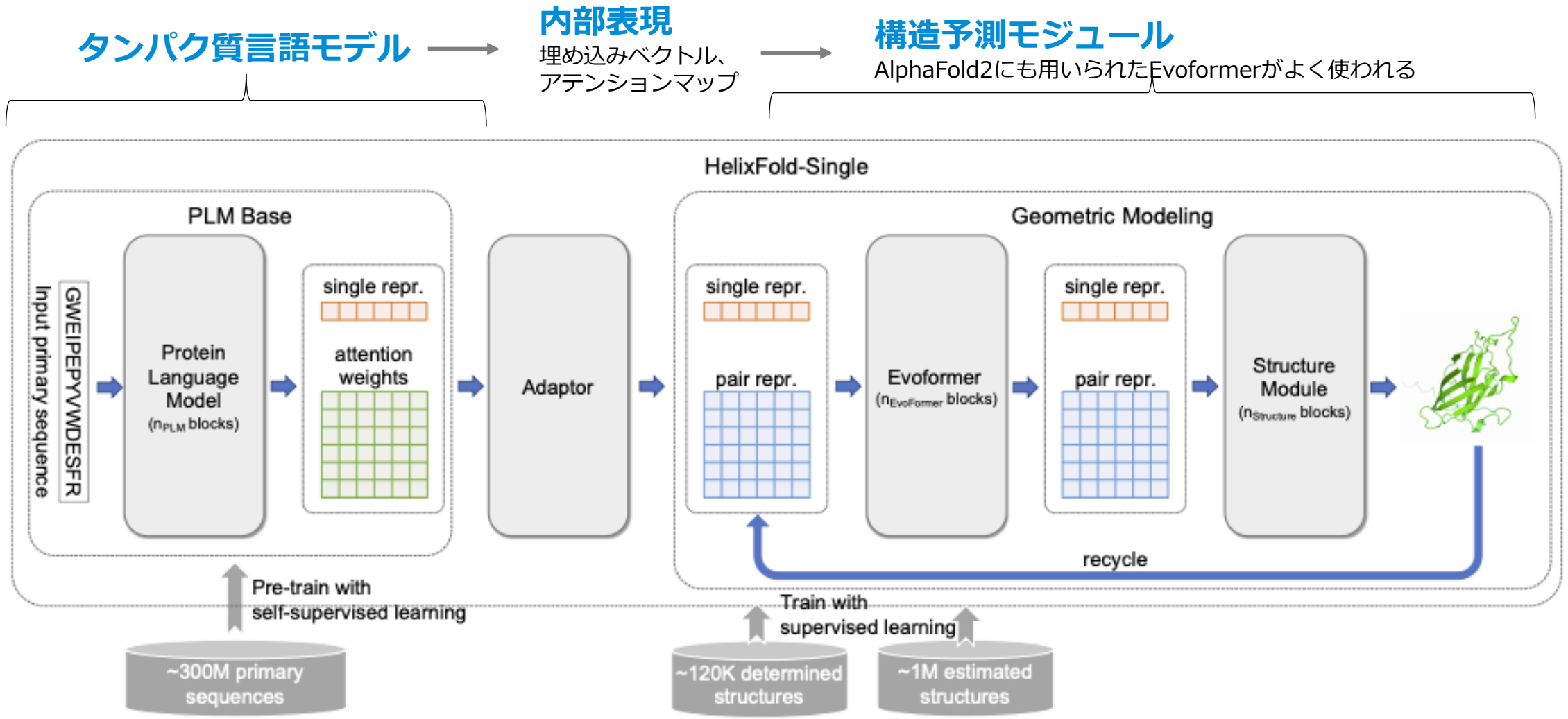
タンパク質・材料設計の基本フレームワーク

設計 = 逆問題

(順方向の予測モデルの逆写像により、所望の特性をもつ入力を予測する)



タンパク質LLMと立体構造予測



X. Fang et al., HelixFold-Single: MSA-free Protein Structure Prediction by Using Protein Language Model as an Alternative, arXiv.2207.13921 (2023).

(参考) 深層学習とタンパク質立体構造予測

	主な開発者	言語モデル	特徴		
AlphaFold2	DeepMind	(なし)	多重配列アライメント (MSA : 似た配列の情報。共進化関係を推定可能)やテンプレート構造 (似た配列の既知の立体構造) などの特徴量を利用。	'21	[1]
ColabFold	MPI, ソウル大、東大など	(なし)	Google Colaboratory上で高速動作するAF2。MMSeqs2 (ウェブ版) の転用によりMSA取得時間を数分程度まで短縮。	'22	[2]
RoseTTAFold	ワシントン大、ハーバード大など	(なし)	3トラックのネットワークを使用して「1D配列レベル、2D距離マップレベル、3D配位レベルの情報」が次々と変換され組み入れられる	'21	[3]
OmegaFold	Helixon	(なし)	MSAを用いず高速計算が可能。精度はRoseTTAFoldを上回り、AF2と同程度。オーファンタンパク質やMSAがノイジーな抗体も予測。	'22	[4]
RGN2	ハーバード大、コロンビア大など	Amino BERT	MSAを用意できないタンパク質 (unaligned protein) のアミノ酸配列に潜在している構造情報を学習。平均精度はAF2に劣るが最大で6桁程度高速化。	'22	[5]
HelixFold	Baidu	DeBERTa	タンパク質言語モデルの出力ベクトル、アテンションマップを構造予測モジュール (Evoformerベース) に入力。MSA不要のため高速。	'23	[6]
ESMFold	Meta AI	ESM2	タンパク質言語モデルの出力ベクトル、アテンションマップを構造予測モジュール (Evoformerベース) に入力。MSA不要のため高速。	'23	[7]
EMBER3D	ミュンヘン工科大 など	ProtT5	タンパク質言語モデルの出力ベクトル、アテンションマップを構造予測モジュール (RoseTTAFoldベース) に入力。MSA不要のため高速。	'22	[8]

[1] J. Jumper et al., Highly accurate protein structure prediction with AlphaFold, Nature 596, 583-589 (2021).

[2] M. Mirdita et al., ColabFold: making protein folding accessible to all, Nature Methods 19, 679-682 (2022).

[3] M. Beak et al., Accurate prediction of protein structures and interactions using a three-track neural network, Science 373, 6557, 871-876 (2021).

[4] R. Wu et al., High-resolution de novo structure prediction from primary sequence, bioRxiv <https://doi.org/10.1101/2022.07.21.500999> (2022).

[5] R. Chowdhury et al., Single-sequence protein structure prediction using a language model and deep learning, Nature Biotechnology 40, 1617-1623 (2022).

[6] X. Fang et al., HelixFold-Single: MSA-free Protein Structure Prediction by Using Protein Language Model as an Alternative, arXiv.2207.13921 (2023).

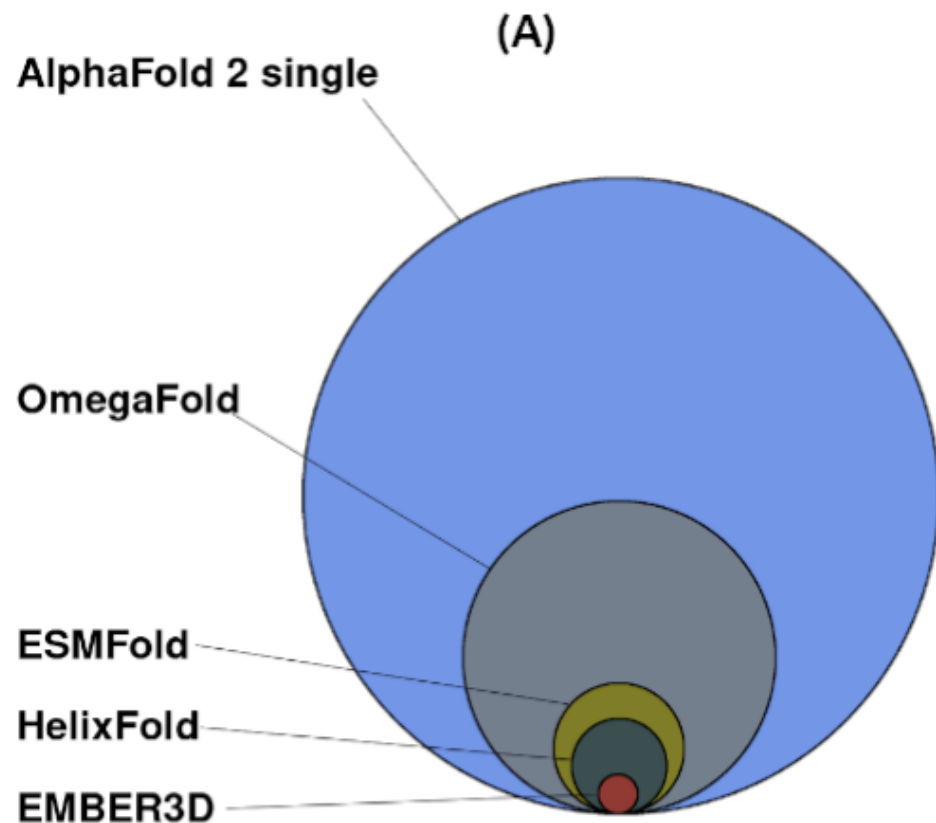
[7] Z. Lin et al., Evolutionary-scale prediction of atomic-level protein structure with a language model, Science 379, 6637, 1123-1130 (2023).

[8] K. Weissenow et al., Ultra-fast protein structure prediction to capture effects of sequence variation in mutation movies, bioRxiv

<https://doi.org/10.1101/2022.11.14.516473> (2022).

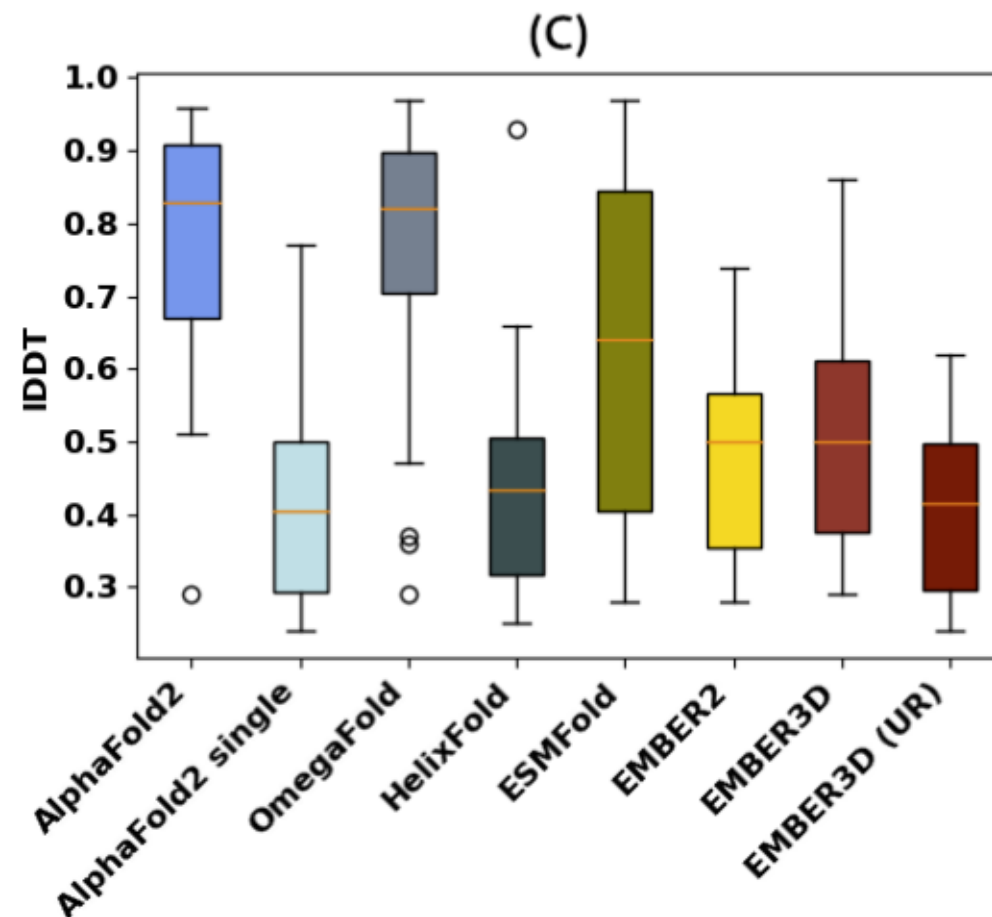
(参考) タンパク質構造予測モデルのベンチマーク

実行時間



※AF2 single : MSAを使わないAF2

予測精度 (local Distance Difference Test)



K. Weissenow et al., Ultra-fast protein structure prediction to capture effects of sequence variation in mutation movies, bioRxiv <https://doi.org/10.1101/2022.11.14.516473> (2022).

分子LLMとマテリアルズ・インフォマティクス

分子言語モデル

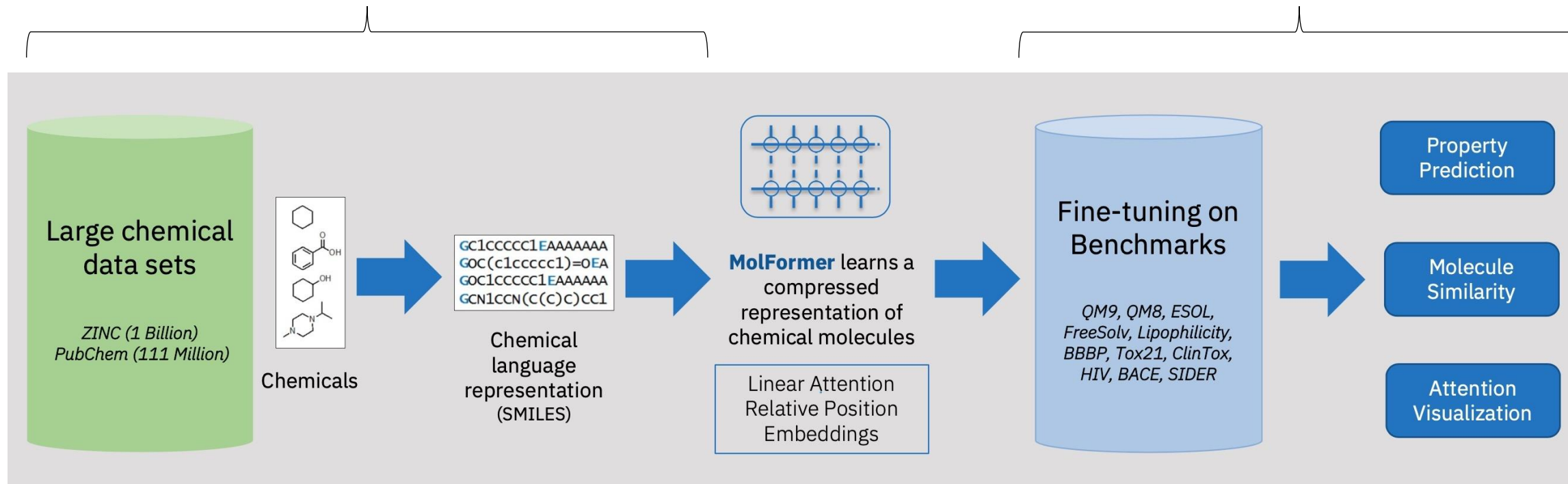
SMILESによる自己教師あり学習
(マスク言語モデル)

内部表現

埋め込みベクトル、
アテンションマップ

物性予測タスク

シミュレーションデータなどでファイン
チューニング (教師あり学習)



タンパク質構造予測の場合とは異なり、分子言語モデルを転移学習
(ファインチューニング) により物性予測タスクに適応させている

J. Ross et al., Large-scale chemical language representations capture molecular structure and properties, Nature Machine Intelligence 4, 1256–1264 (2022).

(参考) LLMの材料科学・化学への応用ハッカソン

LLMの材料科学・化学応用に関するハッカソンイベント（2023年3月実施）の成果物

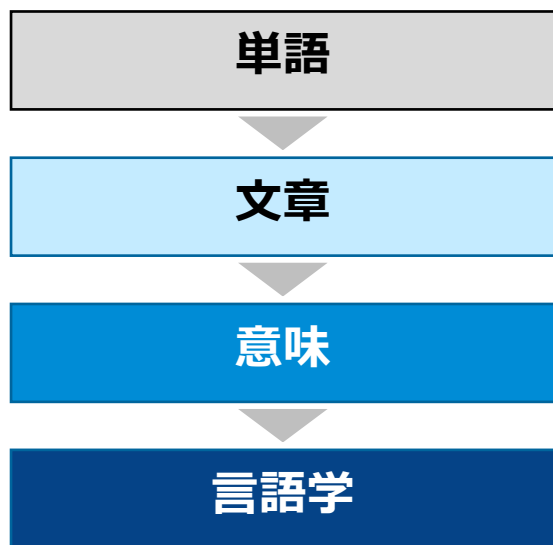
name	authors	links
<i>Predictive modeling</i>		
Accurate Molecular Energy Predictions	Ankur K. Gupta, Garrett W. Merz, Alishba Imran, Wibe A. de Jong	ankur56/ChemLoRA , Twitter
Text2Concrete	Sabine Kruschwitz, Christoph Völker, Ghezal Ahmad Zia	gheزالahmad/LLMs-for-the-Design-of-Sustainable-Concretes
Molecule Discovery by Context	Zhi Hong, Logan Ward	globuslabs/ScholarBERT-XL
Genetic algorithm without genes	Benjamin Weiser, Jerome Genzling, Nicolas Gastellu, Sylvester Zhang, Tao Liu, Alexander Al-Feghali, Nicolas Moitessier, Anne Labarre, Steven Ma	BenjaminWeiser/LLM-Guided-GA , Twitter
Text-template paraphrasing	Michael Pieler	micpie/text-template-paraphrasing-chemistry
<i>Automation and novel interfaces</i>		
BOLLaMa	Bojana Ranković, Andres M. Bran, Philippe Schwaller	doncamilom/BOLLaMa , Twitter
sMolTalk	Jakub Lála, Sean Warren, Samuel G. Rodrigues	jakublala/smoltalk-legacy , Twitter
MAPI-LLM	Mayk Caldas Ramos, Sam Cox, Andrew White	maykcaldas/MAPI_LLM , Twitter
Conversational ELN interface (Whinchat)	Joshua D. Bocarsly, Matthew L. Evans and Ben E. Smith	the-grey-group/datalab
<i>Knowledge Extraction</i>		
InsightGraph Phenomena	Defne Circi, Shruti Badhwar, Irene López Santiago, Alberto López Santiago	defnecirci/InsightGraph , Twitter ireneisdoomed/phenomena , Twitter
Extracting Structured Data from Free-form Organic Synthesis Text	Qianxiang Ai, Jacob N. Sanders, Jiale Shi, Stefan Bringuier, Brenden Pelkie, Marcus Schwarting	qai222/LLM_organic_synthesis , Twitter
TableToJson: Structured information from scientific data in tables	María Victoria Gil	vgvinter/TableToJson
AbstractToTitle & TitleToAbstract: text summarization and generation	Kamal Choudhary	usnistgov/chemnlp
<i>Education</i>		
I-Digest	Beatriz Mouriño, Elias Moubarak, Joren Van Herck, Sauradeep Majumdar, Xiaoqi Zhang	XiaoqZhang/i-Digest

K. M. Jablonka et al., 14 Examples of How LLMs Can Transform Materials Science and Chemistry: A Reflection on a Large Language Model Hackathon, arXiv:2306.06283 (2023).

タンパク質LLM・分子LLM

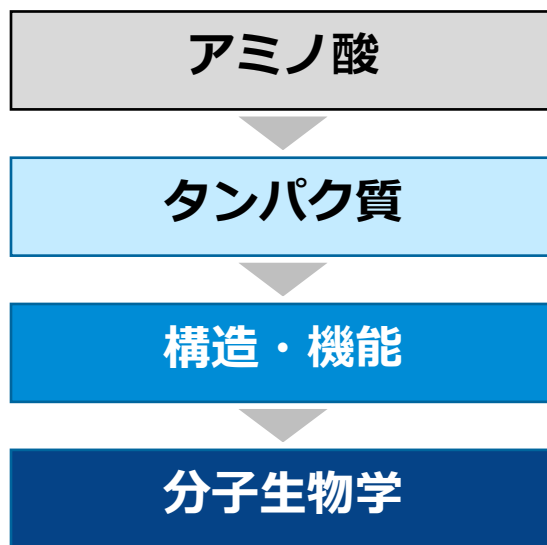
アミノ酸配列で書かれた文章 = タンパク質、元素の配列で書かれた文章 = 分子

大規模言語モデル (LLM)



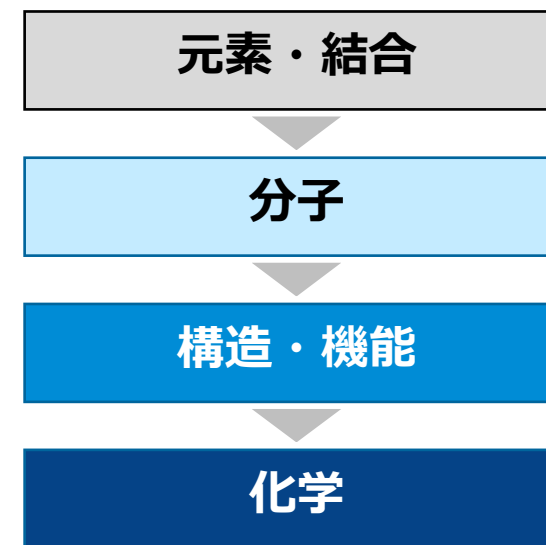
自然言語タスク

タンパク質LLM



生命科学タスク

分子LLM



化学・材料タスク

トランスフォーマーはなぜうまく働く？

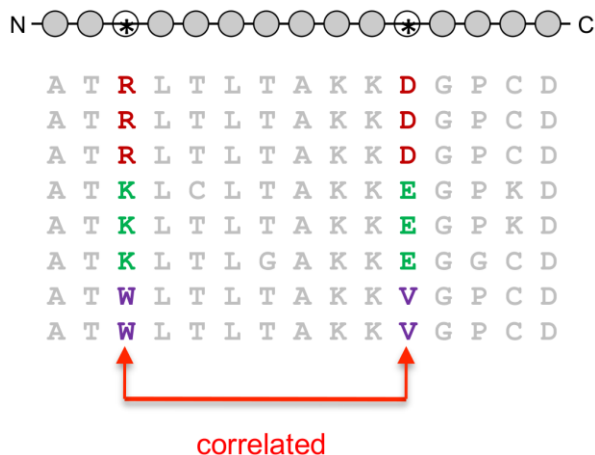
配列上で遠い距離にある要素間の情報を利用して特徴抽出（表現学習）できる。

ネットワーク	要素間の距離	考慮できる距離	重みづけ	内容
畳み込みニューラルネット (CNN)	局所的	固定	固定	フィルタにより固定幅ウィンドウ内の配列をスキャンして特徴抽出。ウィンドウ外にある離れた要素の情報は利用できない。
再帰型ニューラルネット (RNN)	局所的	可変	固定 (遠くほど考慮しない)	どのくらい離れた要素までを考慮するかをデータから学習する。しかし、遠くにある要素からの影響は弱くなるので、配列上で遠くにある要素間の関係の予測は苦手。また、当該要素よりも前の情報を使う。
トランスフォーマー	配列全体	可変	可変 (考慮の程度をデータから学習)	配列中のすべての要素の情報を利用して特徴ベクトルを計算する。どの要素の情報をどの程度考慮するかについてもデータから学習する（アテンション機構）。

タンパク質LLM・分子LLMのアテンション機構

タンパク質の立体構造予測

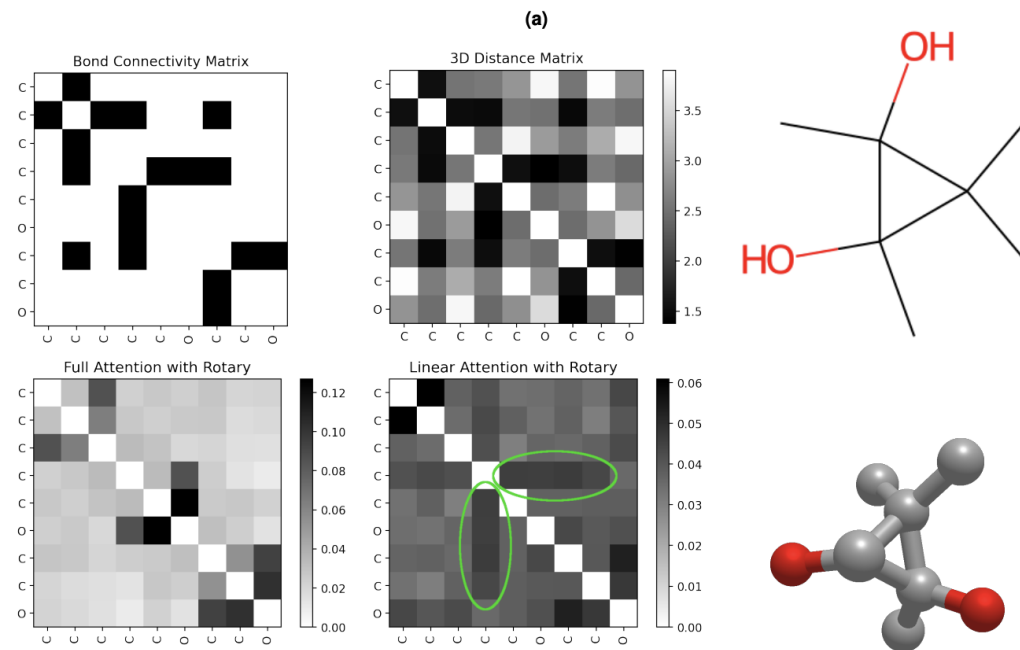
配列上で遠く離れたアミノ酸残基の間の空間的コンタクトの情報が重要 (AF2ではMSAを使っていた)。



D. S. Marks et al., Protein 3D Structure Computed from Evolutionary Sequence Variation, PLoS ONE 6(12): e28766 (2011).

分子の物性予測

SMILES配列上で必ずしも隣接していない離れた元素の空間的な位置関係が物性に寄与。



J. Ross et al., Large-scale chemical language representations capture molecular structure and properties, Nature Machine Intelligence 4, 1256–1264 (2022).

(参考) 時系列トランスフォーマー

時系列予測

LogTrans [Li et al., 2019]
Informer [Zhou et al., 2021]
AST [Wu et al., 2020]
Autoformer [Wu et al., 2021]
FEDformer [Zhou et al., 2022]
TFT [Lim et al., 2021]
SSDNet [Lin et al., 2021]
ProTran [Tang and Matteson, 2021]
Pyraformer [Liu et al., 2022]
Aliformer [Qi et al., 2021]

時空間予測

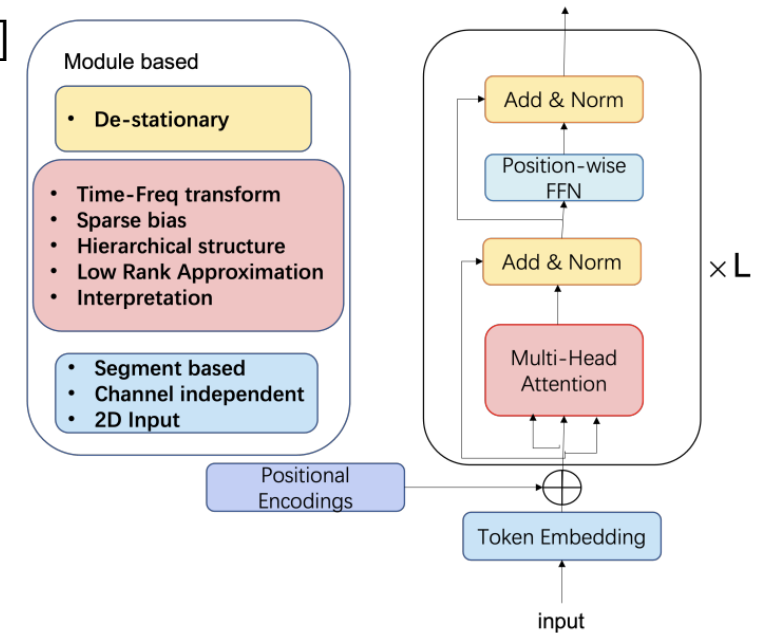
Traffic Transformer [Cai et al., 2020]
Spatialtemporal Transformer [Xu et al., 2020]
Spatialtemporal Graph Transformer [Yu et al., 2020]

イベント予測

Self-attentive Hawkes process (SAHP) [Zhang et al., 2020]
Transformer Hawkes process (THP) [Zuo et al., 2020]
attentive neural Datalog through time (A-NDTT) [Mei et al., 2022]

異常検知

TranAD [Tuli et al., 2022]
MT-RVAE [Wang et al., 2022]
TransAnomaly [Zhang et al., 2021]
GTA [Chen et al., 2021]
AnomalyTrans [Xu et al., 2022]



Q. Wen et al., Transformers in Time Series: A Survey, arXiv:2202.07125 (2022).

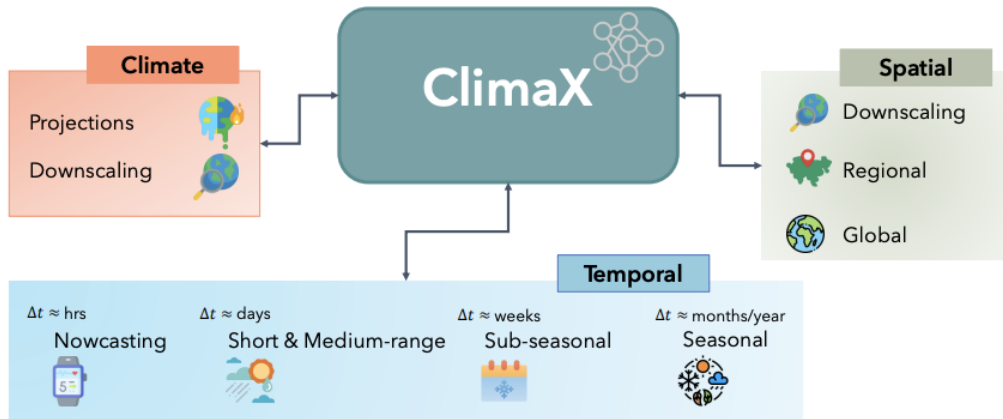
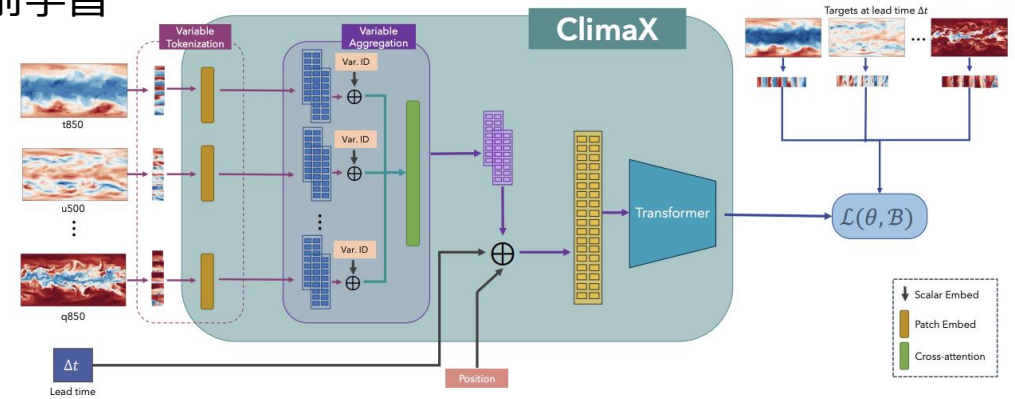
(参考) 気象・気候モデリング

ClimaX (Microsoft Research)

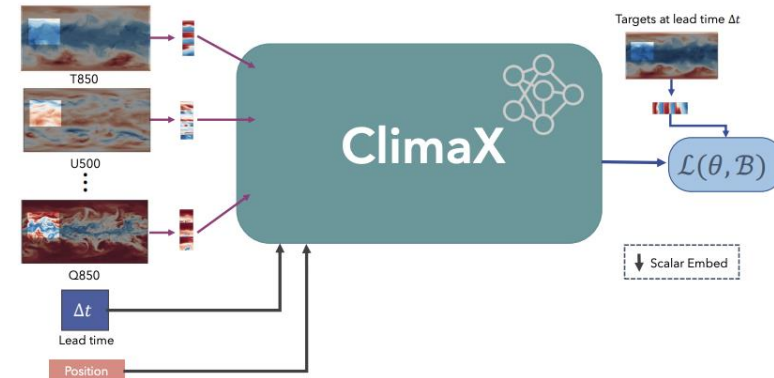
気候データセットで訓練されたトランスフォーマーアーキテクチャの深層学習モデル

- CMIP6由来の気候データセットで事前学習
- ファインチューニングにより、事前学習では扱われなかった大気変数や時空間スケールを含む様々な予測タスクに対処可能
- 既存のデータ駆動型ベースラインと比べてClimaXは汎用性があり、低解像度（かつ少ない計算資源）で事前学習されている場合でも、天気予報や気候予測のベンチマークで優れたパフォーマンスを発揮。

事前学習



ファインチューニング（北米の地域予想用）



Scale is all you need?

タンパク質言語モデル「ESM2」の創発性

- モデルのサイズが15Bでperplexityが改善すると立体構造予測精度も顕著に改善
- タンパク質言語モデルの大規模化は自然言語処理と同様にメリットがある（と期待される）

パワーゲーム化

- Meta、Salesforce、Baidu、Helixon（スタートアップ）に加えてNVIDIAも参戦。オープンデータによるタンパク質LLM構築は計算資源の勝負に。
- 実務上LLMの推論（生成）は軽いほうがよいので、パラメータ効率の良いモデルの開発も重要。

eg.) ESM2-15Bの学習済みモデルファイルサイズは30GB以上。

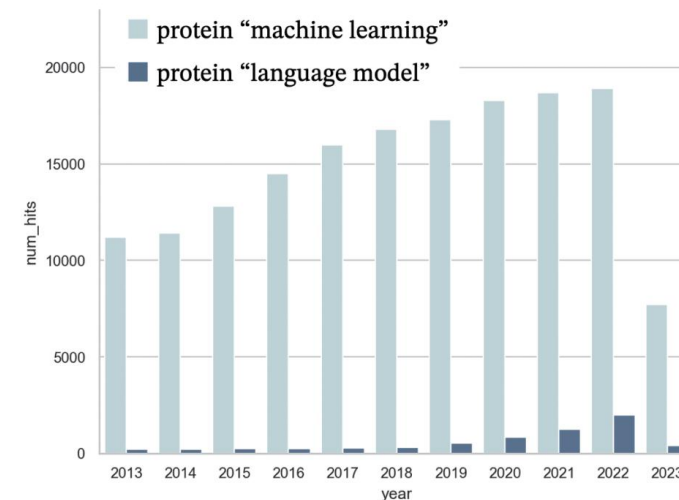
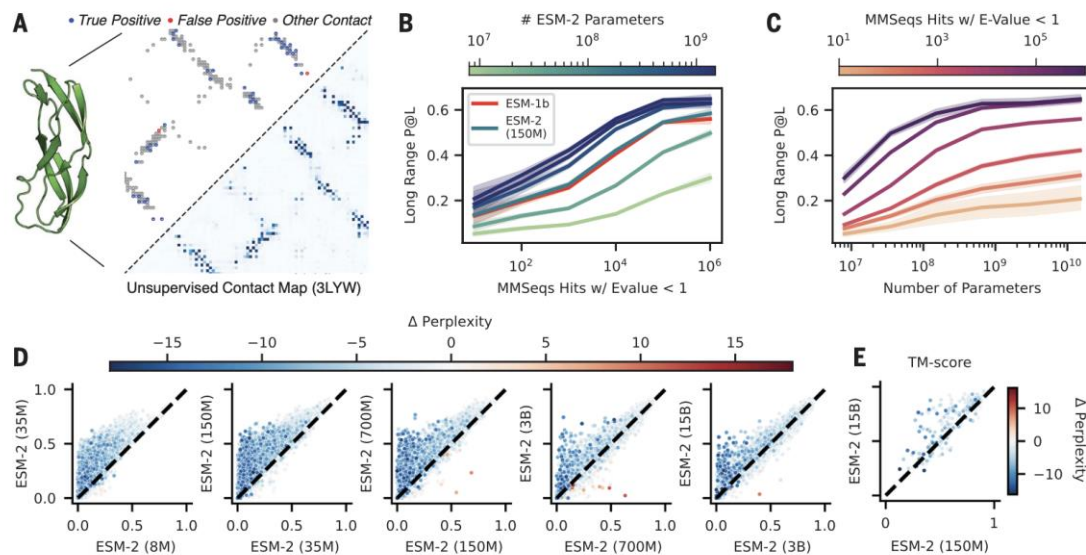


図1：タンパク質の言語モデルに言及する論文数の推移

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

山口 秀輝、齋藤 裕「タンパク質の言語モデル」, JSBi Bioinformatics Review, 4(1), 52-67 (2023).

今後の展望




基盤モデルは深層学習（第3世代AI）の問題点を全て抱えたまま

計算資源 学習に大量の教師データや計算資源が必要
論理推論 パターン処理は強いが、意味理解・説明等の高次処理はできていない
実世界操作 学習範囲外の状況に弱く、実世界状況への臨機応変な対応ができない



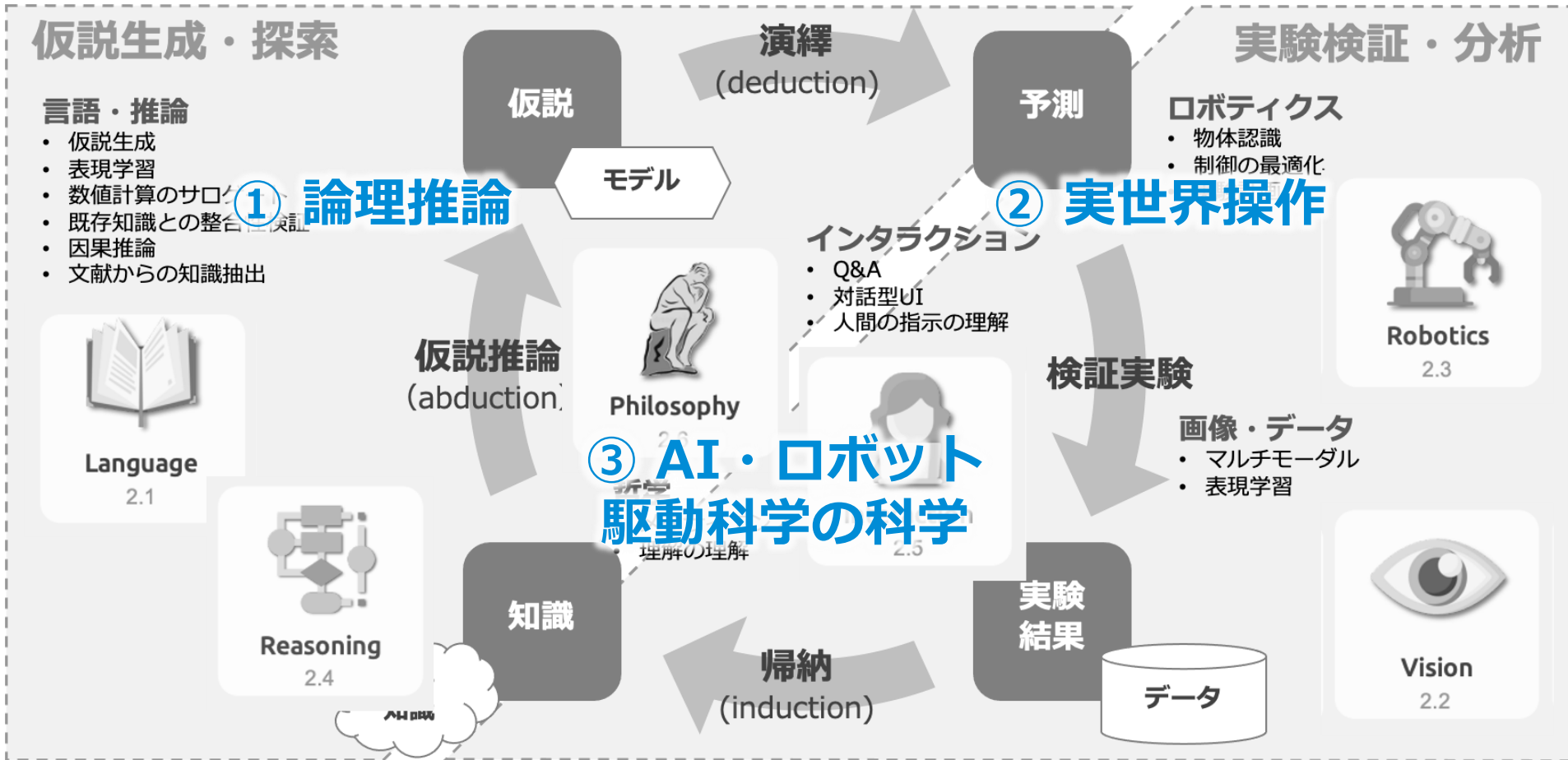
次世代AIモデル研究

人間知能の理解からの示唆と基盤モデルのメカニズム理解の相乗効果・融合を推進

	基盤モデル	二重過程モデル	発達・創発モデル
中心的アイデア	深層ニューラネットの大規模学習	即応的システム1と熟考的システム2で構成	外部との相互作用における予測誤差最小化
関連研究分野	ニューラルネット、機械学習、計算機科学	心理学、行動経済学、脳科学	認知発達・記号創発ロボティクス
人にあてはめると	人が一生かかっても読めない量の情報を学習 	大人の思考・意思決定パターン 	乳児・幼児の発達・成長過程 

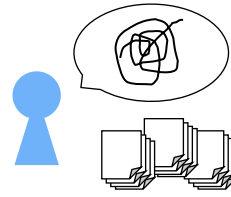
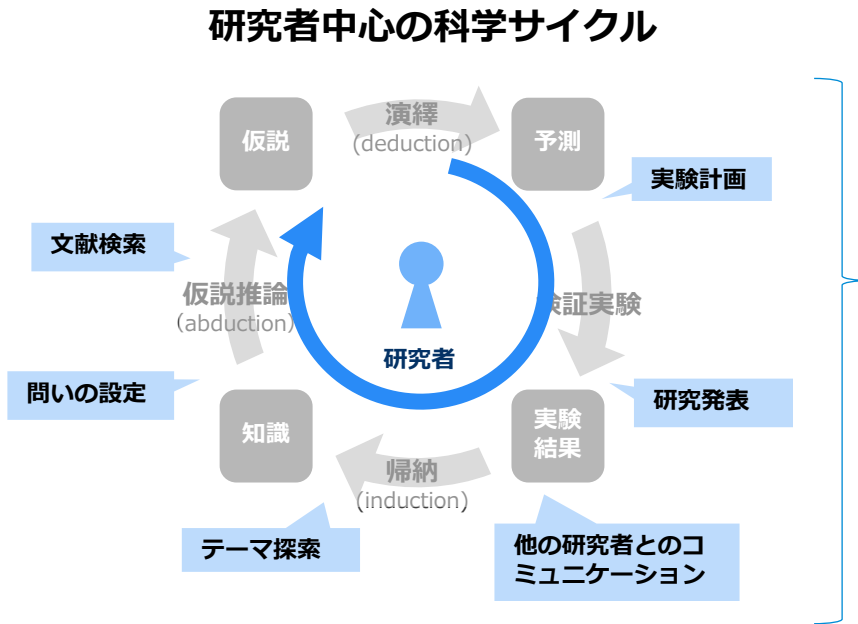
次世代AI研究とAI・ロボット駆動科学

AI・ロボット駆動科学を実証フィールドとして次世代AI研究を加速！



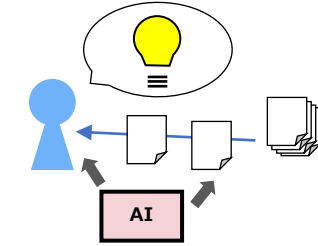
研究者中心のワークフロー進化

情報過多に直面する研究者を支援する（LLMなどを用いた）AIツールが登場。



研究ワークフローの全段階で
研究者は情報過多に直面

- 例：arXivには毎月1.5万件の論文が掲載
- 研究者の認知的限界がボトルネックに。



AIによる情報検索などの支援が
ますます必要に

- アレンAI研究所は人間中心（human centric）な研究タスクの機械支援を研究*。研究者が行うタスクに応じた知識検索（獲得）に関するシステムを開発。

* Hope, T., Downey, D., Etzioni, O., Weld, D. S., & Horvitz, E. (2023). A Computational Inflection for Scientific Discovery. <https://arxiv.org/pdf/2205.02007.pdf>

近年登場している研究支援ツールの例（一部はGPT-4を搭載）

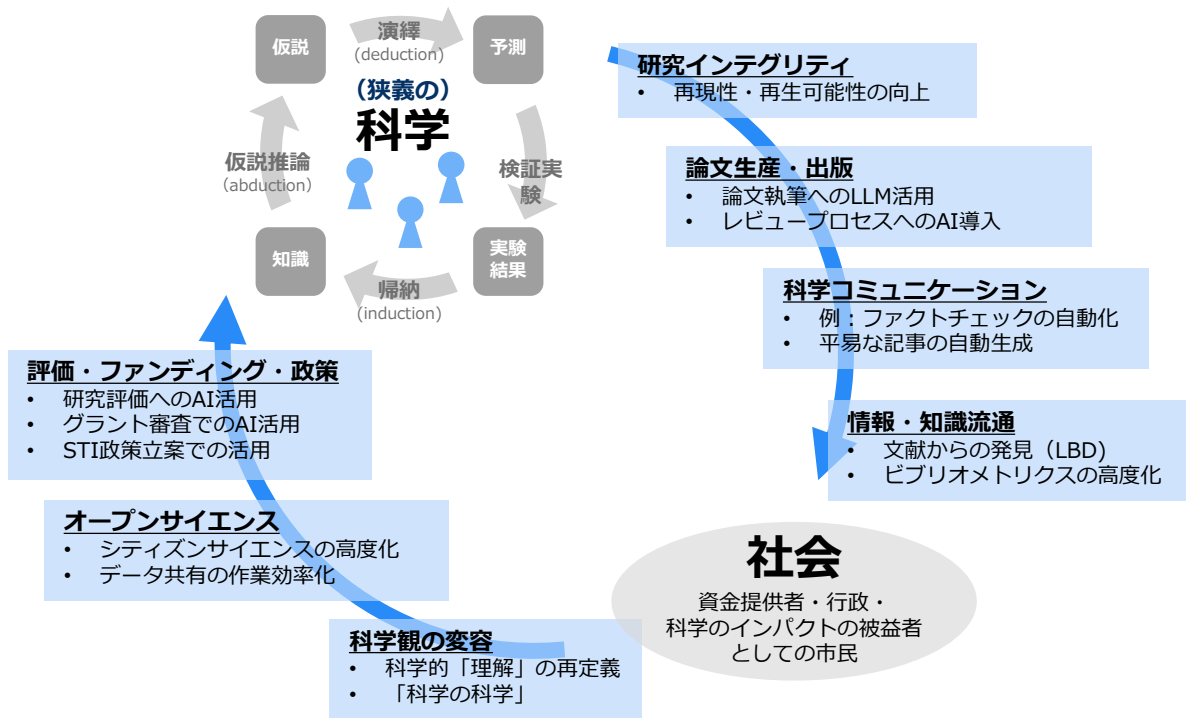
- [Jenni AI](#) 論文などのライティング支援ツール。剽窃チェッカーつき。
- [Humata AI](#) チャット型オンラインAIサービス。文章要約、質問回答などが可能。
- [Consensus](#) 質問に対して論文検索を元に科学的コンセンサスの強さとともに回答を表示。
- [Scite.ai](#) 引用分析プラットフォーム。研究論文の検索を行える機能を実装。
- [Elicit.org](#) 質問に対して関連論文と要約を出力する論文検索サービス。
- [Research Rabbit](#) 関心のある論文のコレクションに基づく文献検索、出版アラート機能など。
- [Semantic Scholar](#) 論文検索サービス。Allen Institute for AIが開発。
- [Connected Papers](#) 論文検索アプリケーション。関係性がある論文をグラフ表示。

- 研究生産性向上に資するグッドプラクティスの探索・共有が重要に。
- 知識工学、情報学などの分野からのツール開発も研究課題に。

科学の社会的プロセスへのAIの浸透

研究の評価やファンディングなど、社会的プロセスとしての“広義の科学”にもAIが浸透。

広義の科学におけるAI活用の可能性（例）※1



期待と懸念

AIへの期待

- 再現性・再生可能性の向上
- 論文執筆の効率化
- 科学における言語の壁の解消
- レビュープロセスへのAI導入による効率化
- シティズンサイエンスの高度化
- 「科学の科学」の高度化
- ...

※AI全般がもたらすメリットとして指摘されているもの

懸念

- 研究不正の高度化
- デュアルユース
- AIを活用した「浅い研究」の増加
- 研究者の格差の強化？
- 特定の言語・文化へのバイアス強化？
- ツール導入による人手作業の増加
- ...

※とくに生成AIに関して懸念されている事柄

- よりよいAIの活用と懸念への対応をめぐり、世界中で議論と技術開発が進展。
- STI政策、科学哲学、科学計量学、学術情報関連分野など「メタサイエンス※2」の観点からの研究・実践も必要に。

※1 ここでは生成系に限らないAI活用の可能性の例を挙げた。項目の拾い出しにおいては以下などを参考にした。OECD workshop on AI and the productivity of science (2021); Johanna Okerlund et al. "What's in the Chatterbox? Large Language Models, Why They Matter, and What We Should Do About Them"(2022)

※2 メタサイエンスとは、科学のあり方自体を対象とする研究とその知見に基づく実践。参考：調査報告書「拡張する研究開発エコシステム」<https://www.jst.go.jp/crds/report/CRDS-FY2022-RR-03.html>

日本の強みと研究開発戦略（のヒント）

科学基盤モデルの開発

- スクラッチからの事前学習は計算資源必要（世界共通の課題）
 - 国際協力・ビッグサイエンス化を検討。富岳やABCIなどHPCIの積極活用。
- オープンソースモデルの再学習（転移学習、ファインチューニング、蒸留、など）
 - 理研、NIMS、NIIなどにある良質のデータ、マルチモーダルデータを活用。

基盤モデルとAI・ロボット駆動科学

- AIロボット駆動科学は基盤モデル・次世代AIモデル開発のよい実証フィールド
 - JST未来、Moonshotなどでの先駆的な取り組みを活かす
- 「論理推論」「実世界操作」のできる次世代AIモデル開発に早期着手
 - 物理、化学、生物など研究者のドメイン知識の活用、ロボット研究の蓄積も活かす

參考資料

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の出力が否かの判定技術(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

- 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

← 実務

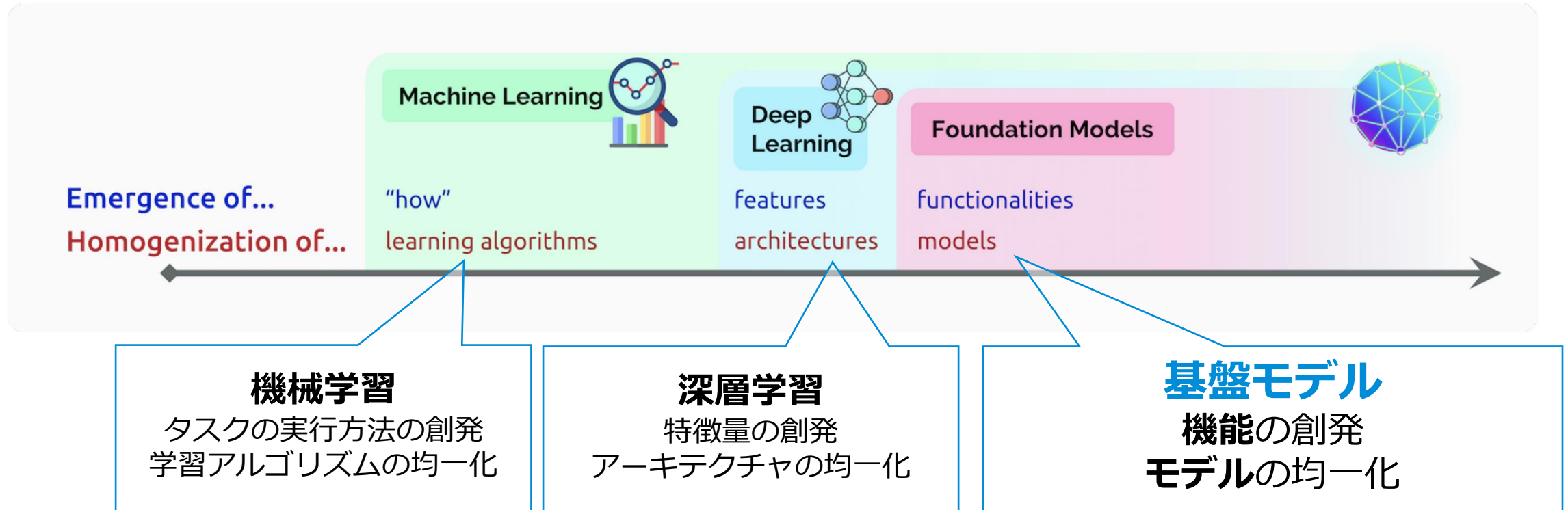
→ 学術

↑ 応用個別

↓ 共通基盤

基盤モデル

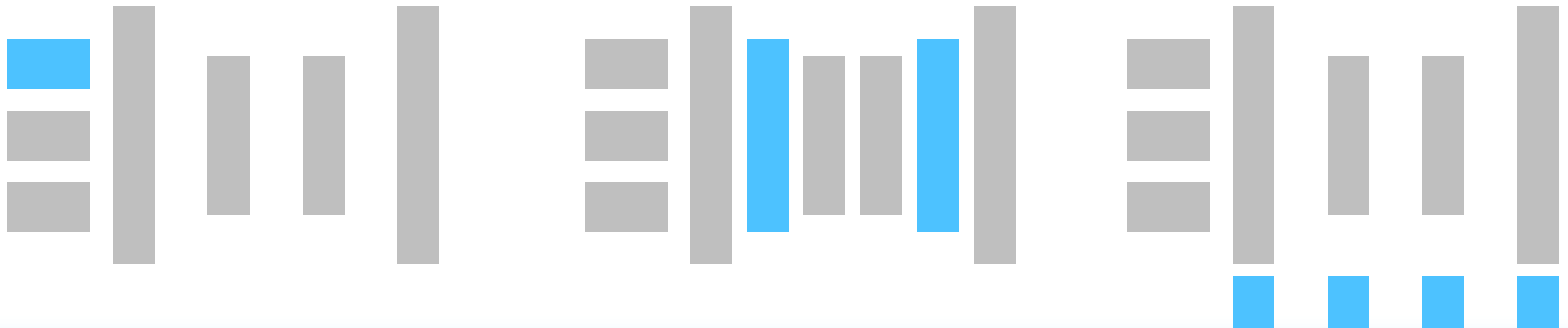
基盤モデルをひとつ用意して再学習（転移学習）で様々なタスクに適応させる



R. Bommasani et al., On the Opportunities and Risks of Foundation Models, arXiv:2108.07258 (2021).

いろいろなPEFT手法 (PEFT: Parameter-Efficient Fine Tuning)

	トークンを追加	モジュールを追加	低ランク行列を追加
学習対象	仮想トークン (ベクトル埋め込み、prefix)	Adapterモジュール	低ランク行列
パラメータ更新	なし (元の事前学習済みモデルのパラメータは固定)	Adapterのパラメータを更新 (元の事前学習済みモデルのパラメータは固定)	低ランク行列を更新 (その上で元のモデルの重みを加算しパラメータを更新)
モデル	変更なし	変更なし	重みをマージ
手法	<ul style="list-style-type: none"> Prefix Tuning P Tuning Prompt Tuning 	<ul style="list-style-type: none"> Adapter 	<ul style="list-style-type: none"> LoRA Ada LoRA



パラメータ効率のいいタンパク質LLM

CaLM (Codon adaptation Language Model)

C. Outeiral and C. M. Deane, Codon language embeddings provide strong signals for protein engineering, bioRxiv 2022.12.15.519894

European Nucleotide Archiveから取得した約900万の cDNA配列を用いて学習

- 従来のタンパク質LMのようにアミノ酸を主たる語彙とせず、64種のコドン語彙にするアプローチコドンの冗長性を活用して情報抽出ができるかを種々のタスクで検証
- CaLMのパラメータ数は約8600万 (ESM2-15BやProtT5等と比較して数十から数百倍少ない)
- 学習データも桁違いに小さいにも関わらず、配列の種推定や熱耐性予測タスクにおいて顕著に高い精度を達成

Ankh

A. Elnaggar, et al., Ankh ♀: Optimized Protein Language Model Unlocks General-Purpose Modelling, bioRxiv 2023.01.16.52426

- マスク言語モデルでのテクニック (マスク確率・範囲、潜在次元やアテンションヘッド数などのモデルアーキテクチャ、位置エンコーディング手法など) をタンパク質LMに対し最適化
- ESM2-15Bの3%~10%程度のパラメータしか持たないが、複数のタスクで最良性能を発揮

ProteinBERT

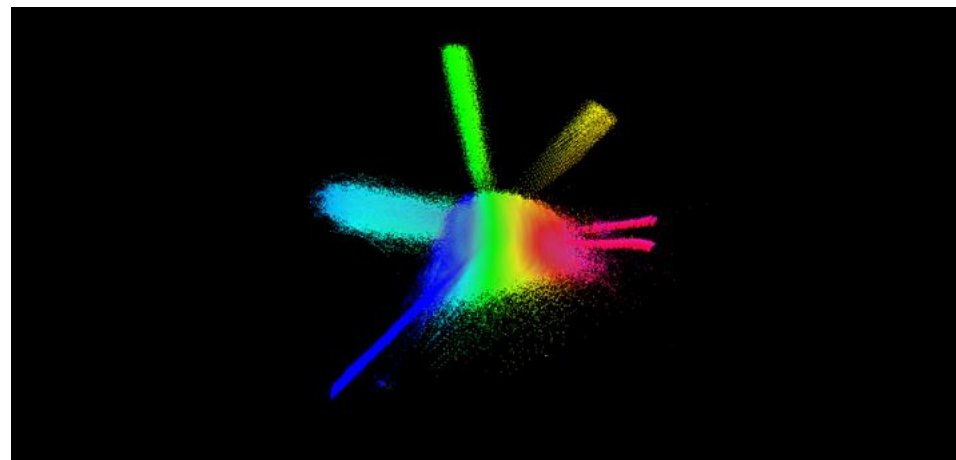
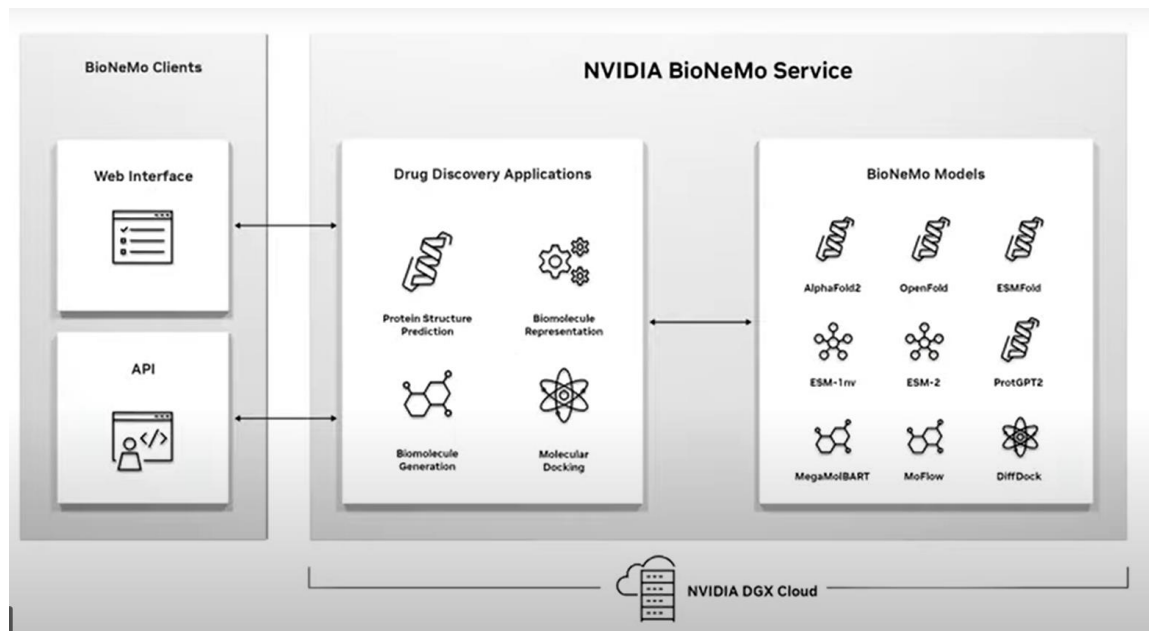
N. Brandes et al., ProteinBERT: a universal deep-learning model of protein sequence and function, Bioinformatics 38, 8, 2102–2110 (2022).

- 配列に付随するGene Ontologyタグを同時にエンコードしTAPE Transformer等と同程度の性能を達成。
- Transformerは配列長の2乗のオーダーの計算量なのでCNNベースタンパク質LMは実務的利点大きい。
- 配列以外のデータを積極的に取り入れマルチモーダルアプローチの優位性を検証

BioNeMo (NVIDIA)

LLMを生命科学のデータセットに適用し、候補物質のスクリーニングやタンパク質間の相互作用予測などに利用。以下のモデルによる推論がクラウドAPIアクセス可能。

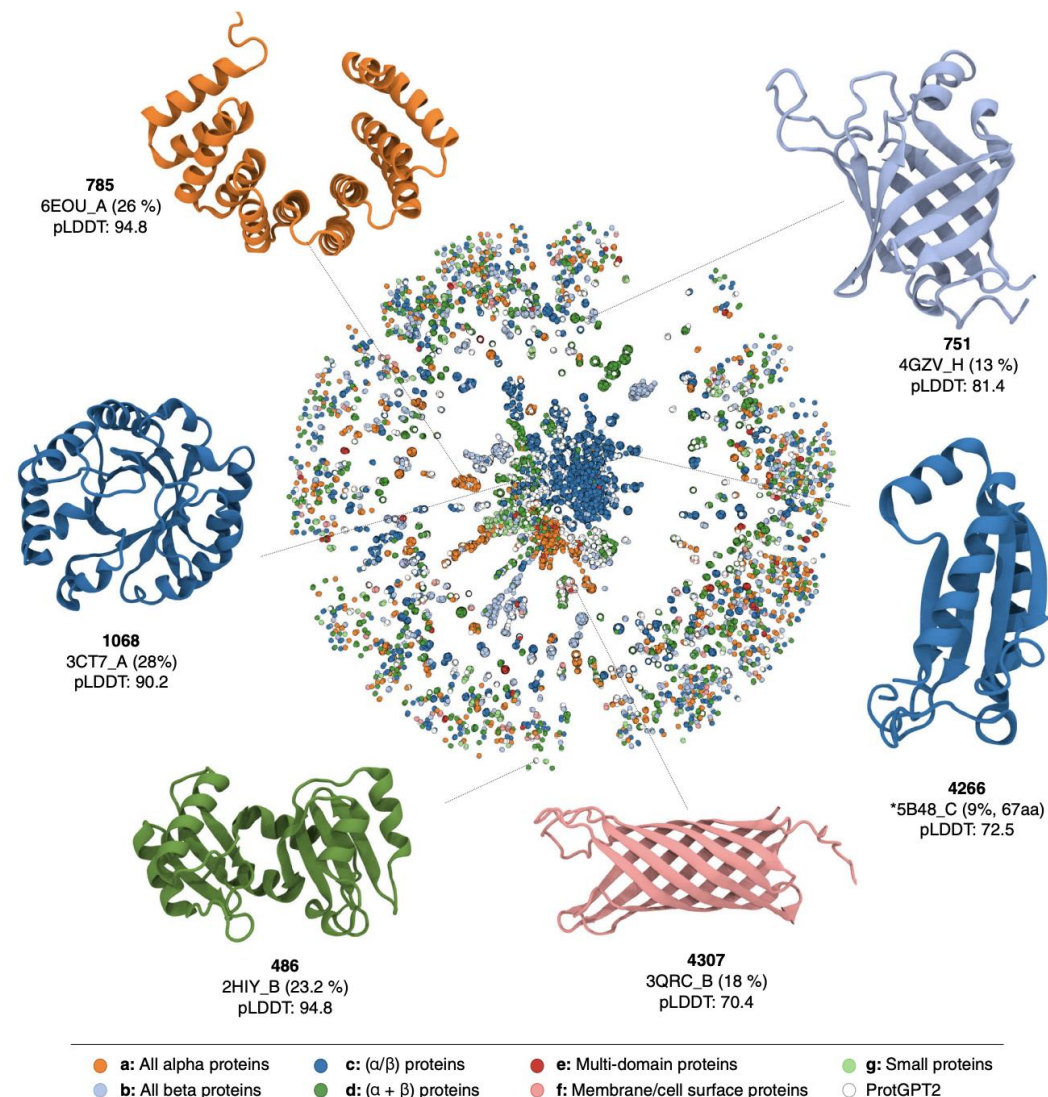
- ESM-1：タンパク質の機能予測、アミノ酸配列生成
- OpenFold：タンパク質モデリング
- MegaMolBART：反応予測、分子最適化、de novo合成
- ProT5：アミノ酸配列生成（タンパク質LLM）



ProtGPT2

大規模言語モデルにより新規タンパク質配列を生成。
類似性ネットワーク解析から未開拓のタンパク質空間をサンプリングしていることが示唆されている。

- 訓練用のデータセットは注釈なしの44.9百万個の配列からなるUniRef50 (UR50) と呼ばれるUniProtのクラスタリングで構成されている。
- このデータセットは、実験的方法や同族モデリングを介して構造がアクセスできないタンパク質空間の領域である「ダークプロテオーム」を含む、タンパク質空間全体をカバーしているとされる。
- 訓練用に44.9百万個のシーケンスが使われ、その他の4.9百万個が評価用に使われた。



Galactica (Meta AI)

論文や知識ベースなどの科学的コーパスを学習した大規模言語モデル。文献調査などへの利用が期待されていたが、ハルシネーションや人種差別的な表現の問題のため公開2日で停止。

- **知識問題** : LaTeX方程式の問題でGPT-3に勝利 (68.2 vs 49.0%)
- **数学の問題** : mathematical MMLUでChinchillaに勝利 (41.3 vs 35.7%)、MATHでPaLM 540Bに勝利 (20.4 vs 8.8%)
- **ダウンストリームNLPタスク** : PubMedQAとMedMCQA devでも優れた成績を収めた。
- **NLPベンチマーク** : 一般的なコーパスで訓練されていないのにBIG-benchタスクでBLOOMとOPT-175Bを上回った。

Total dataset size = 106 billion tokens			
Data source	Documents	Tokens	Token %
Papers	48 million	88 billion	83.0%
Code	2 million	7 billion	6.9%
Reference Material	8 million	7 billion	6.5%
Knowledge Bases	2 million	2 billion	2.0%
Filtered CommonCrawl	0.9 million	1 billion	1.0%
Prompts	1.3 million	0.4 billion	0.3%
Other	0.02 million	0.2 billion	0.2%

Data source	Documents	Tokens
arXiv	2 million	35 billion
PMC	3 million	23 billion
Semantic Scholar	3 million	18 billion
PubMed Abstracts	21 million	5 billion
Semantic Scholar Abstracts	19 million	4 billion
bioRxiv	128,059	1 billion
OSF	54,905	428 million
medRxiv	24,019	176 million
ACL	25,518	150 million
PubAg Abstracts	308,235	105 million
ChemRxiv	7,617	67 million
Total	48 million	88 billion

Data source	Documents	Tokens
Wikipedia	6 million	5 billion
StackExchange	1.6 million	1 billion
LibreText	95,113	185 million
Wikibooks	74,705	110 million
Open Textbooks	647	94 million
MIT OCW	25,640	90 million
Wikiversity	38,138	52 million
ProofWiki	32,389	12 million
Khan Academy	3,075	7 million
Papers with Code	13,430	4 million
IUPAC Goldbook	6,788	1 million
Total	8 million	7 billion

Data source	Documents	Tokens
PubChem Compound	1.7 million	1 billion
UniProt	551,837	0.6 billion
RefSeq Genome	69	0.1 billion
OEIS	350,833	0.07 billion
Ribosome	9,950	0.05 billion
LIPID MAPS	45,273	0.03 billion
Reactome	156	0.01 billion
NASA Exoplanet	5,021	0.01 billion
Total	2 million	2 billion

論文

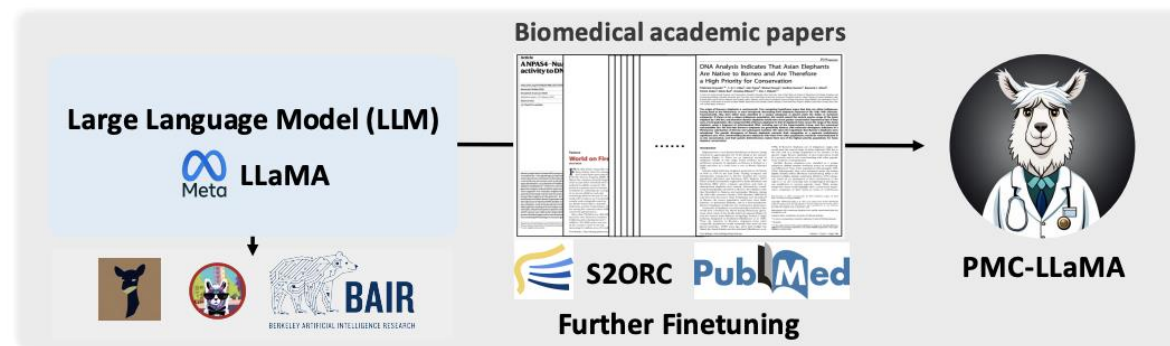
参考資料

知識ベース

PMC-LLaMA

約480万のバイオメディカル分野の論文でオープンソースの大規模言語モデルLLaMAをファインチューニング。

- S2ORC データセットから約8110万件の英語論文を取得し、PubMed Central (PMC)-idでフィルタリングした約480万件を学習に利用。
- PubMedQA、MedMCQA、USMLEなどの3つのバイオメディカルQAベンチマークで高い性能を発揮。
- 医療対話や診療などのさまざまなヘルスケアサブタスクに特化した基盤LLMだと示唆される。



Method	Setting	USMLE(OOD/ID)	MedMCQA(ID)	PubMedQA(ID)
Human (pass)	Manual*	50.0	–	60.0
Human (expert)		87.0	90.0	78.0
InstructGPT-175B	Zero-shot*	46.0	44.0	73.2
ChatGPT		57.0	44.7	63.9
LLaMA-7B		27.1	24.3	5.2
LLaMA-33B		43.4	30.3	1.8
LLaMA-7B _{Full}	Full fine-tuning	44.55/35.66	48.15	73.41
PMC-LLaMA-7B _{Full}		44.70/40.61	50.54	69.53
LLaMA-7B _{PEFT}	PEFT	29.38/27.34	32.37	65.81
PMC-LLaMA-7B _{PEFT}		30.64/28.52	34.33	68.23

Table 1: Comparison between LLaMA-7B and PMC-LLaMA-7B under different settings. ACC in percentages is reported in the table. Note that, the manual and zero-shot results with * are referred from LMFlow [Diao et al., 2023].

MolFormer

1.1億個の分子データで訓練された大規模なトランスフォーマーモデル

- 訓練データセットはPubChem (1.1億) とZINC (10億) を利用 (SMILES表記のシーケンス)。
- 10種類のベンチマークのいくつかの下流タスクで教師ありグラフNNや言語モデルなどよりも優れた性能を發揮。
- 分子LLM (= SMILESを学習したLLM) は分子の構造情報を学習でき、量子化学的性質を含む様々な分子特性を予測するのに十分な化学的・構造的情報を把握できることが示唆された。

Table 1 | Comparison of fine-tuned MOLFORMER with existing supervised and pretrained/self-supervised baselines on multiple classification benchmarks

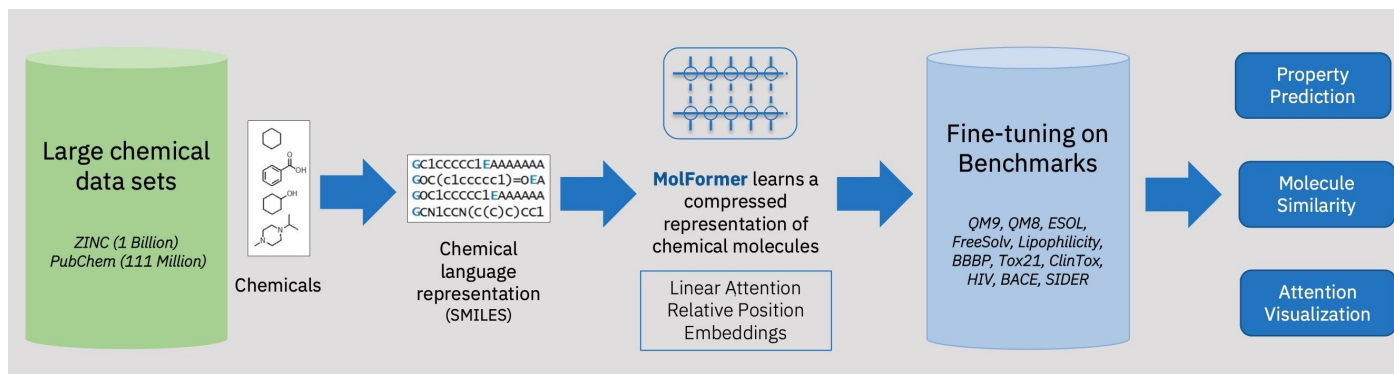
	BBBP 1	Tox21 12	ClinTox 2	HIV 1	BACE 1	SIDER 27
RF	71.4	76.9	71.3	78.1	86.7	68.4
SVM	72.9	81.8	66.9	79.2	86.2	68.2
MGCN ²⁶	85.0	70.7	63.4	73.8	73.4	55.2
D-MPNN ²⁷	71.2	68.9	90.5	75.0	85.3	63.2
DimeNet ²⁷	—	78.0	76.0	—	—	61.5
Hu et al. ³²	70.8	78.7	78.9	80.2	85.9	65.2
N-gram ³³	91.2	76.9	85.5	83.0	87.6	63.2
MolCLR ²⁶	73.6	79.8	93.2	80.6	89.0	68.0
GraphMVP-C ³⁶	72.4	74.4	77.5	77.0	81.2	63.9
GeomGCL ³⁶	—	85.0	91.9	—	—	64.8
GEM ³⁸	72.4	78.1	90.1	80.6	85.6	67.2
ChemBERTa ²⁵	64.3	—	90.6	62.2	—	—
MOLFORMER-XL	93.7	84.7	94.8	82.2	88.21	69.0

Bold indicates the top-performing model. All models were evaluated using the area under the receiver operating characteristic curve on scaffold splits. Baseline performances are adopted from refs. 25, 26, 36, '-' signifies that the values were not reported for the corresponding task.

Table 2 | Performance of fine-tuned MOLFORMER and other supervised GNN baselines on QM9, QM8, ESOL, FreeSolv and Lipophilicity regression benchmarks

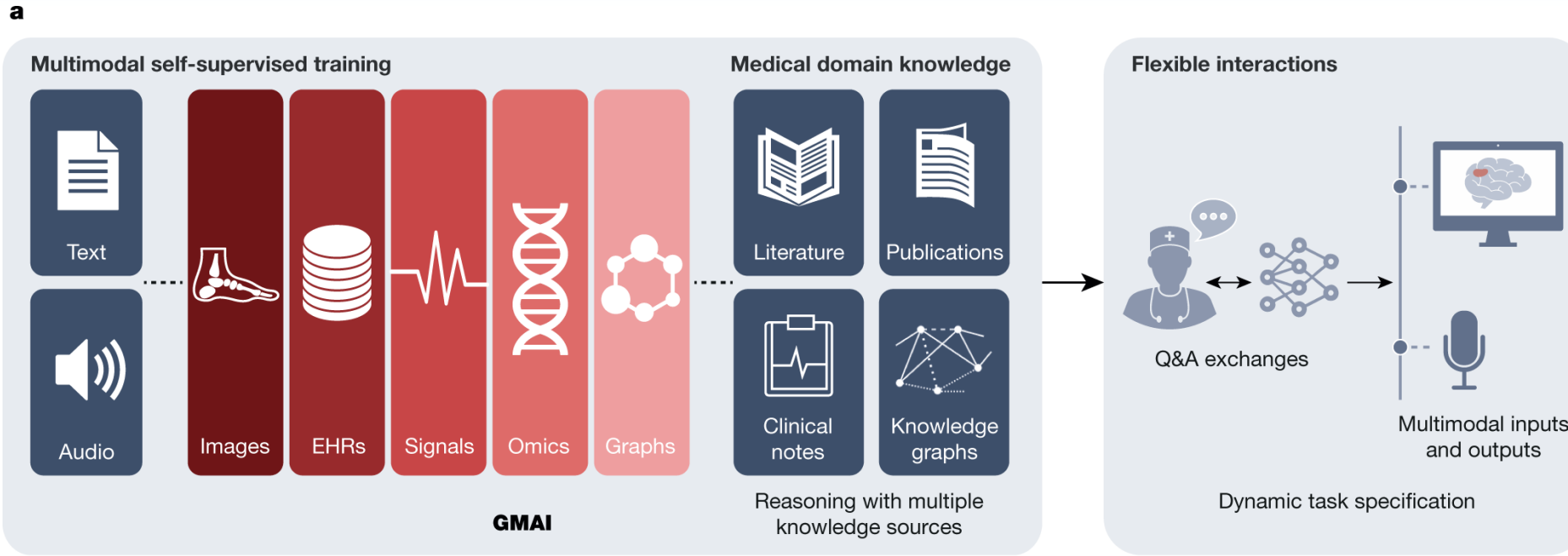
	QM9	QM8	ESOL	FreeSolv	Lipophilicity
GC	4.3536	0.0148	0.970	1.40	0.655
A-FP	2.6355	0.0282	0.5030	0.736	0.578
MPNN	3.1898	0.0143	0.58	1.150	0.7190
MOLFORMER-XL	1.5894	0.0102	0.2787	0.2308	0.5289

For QM9 and QM8, we report average MAE, while root-mean-square error is reported for the remaining tasks. Baseline performances are taken from refs. 28, 40. Bold indicates the top-performing model.



J. Ross et al., Large-scale chemical language representations capture molecular structure and properties, Nature Machine Intelligence 4, 1256–1264 (2022).

Generalist Medical AI

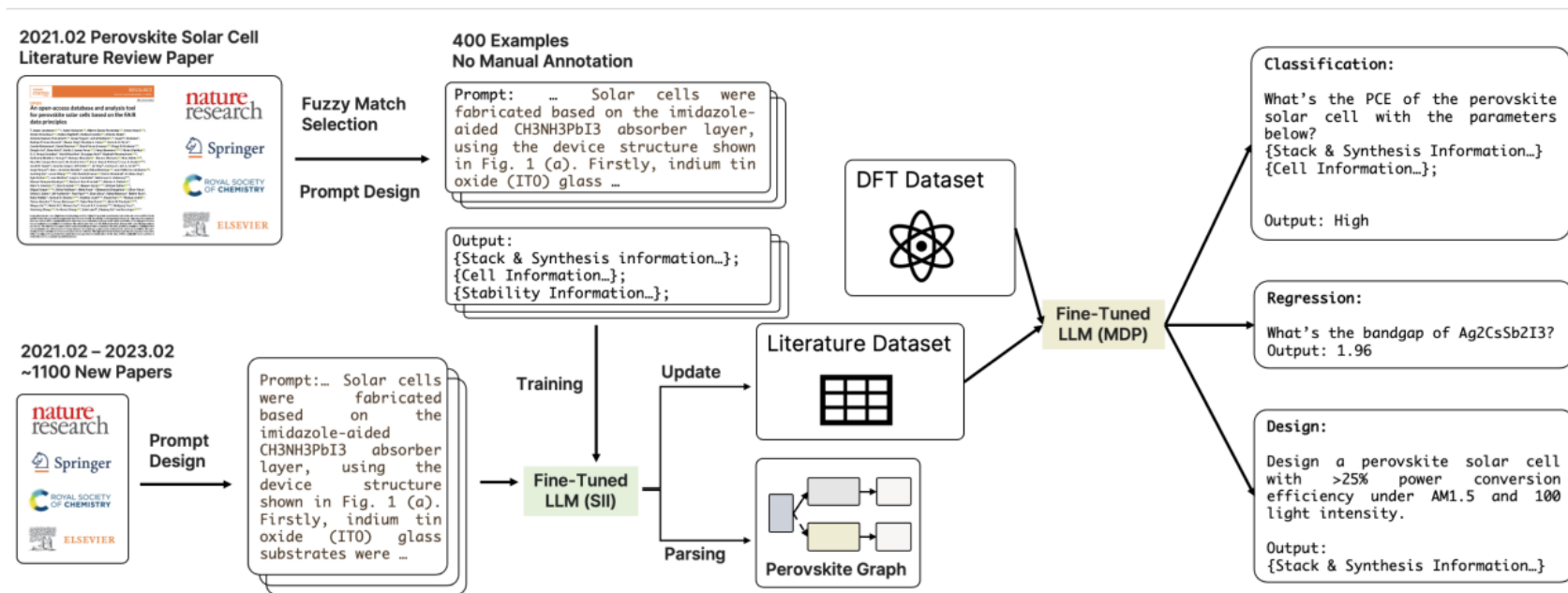


Regulations: Application approval; validation; audits; community-based challenges; analyses of biases, fairness and diversity

GPTのファインチューニング

GPT-3をファインチューニングして構造化情報推論タスク（材料科学NLP）に適用。

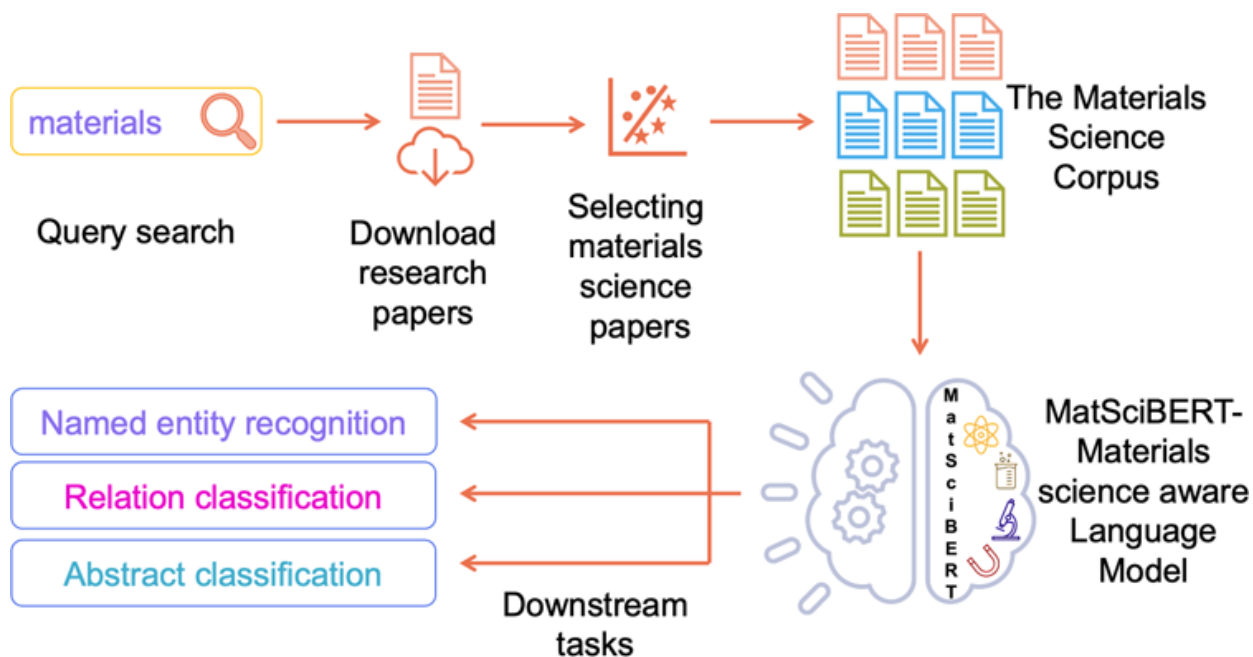
- ペロブスカイト太陽電池FAIR（Findable、Accessible、Interoperable、Reusable）データセット。
- 材料科学NLPタスクである構造化情報推論（SII）に適用。SIIには、名前付きエンティティ認識（NER）、エンティティ解決（ER）、関係抽出（RE）、情報推論（II）などのタスクが含まれる。
- GPT-3をファインチューニングし、91.8%のF1スコアを達成。
- LLMを使用して太陽電池の電気的性能を予測し、ターゲットパラメータを持つ材料・デバイスを設計する実験も行われた。



MatSciBERT

材料科学論文コーパスで訓練された材料科学に特化された言語モデル

- 名前エンティティ認識、関係分類、および抽象分類の3つのダウンストリームタスクで性能を発揮。
- MatSciBERTの事前学習済みウェイトは公開されており、材料発見やテキストからの情報抽出などの加速のために使用可能。



Captions with entities extracted using MatSciBERT NER model	Manual labels ¹⁰
The comparison of XRD patterns of glass ceramic heat treated at 725 °C for 5h and rhombohedral Ba4Gd3F17 . The superstructure reflections are marked with o. Inset: enlarged sections of XRD patterns.	Reflection
HRTEM image and the corresponding FFT pattern taken from as-deposited sample B(80/1) (a) and annealed sample D(30/1) (b); identifying rutile TiO2 crystal grains.	FFT
The illustrative schemes: a) The bonding of hexagonal ZnO nanocrystals to the glass surface . b) The structure of multi-layers coatings .	Crystal
(a) XRD patterns of the glass-ceramics sintered different holding times; (b) intensity of μ - and α -cordierite peaks count at (101) and (110) plane respectively as a function of sintering holding time.	XRD
Photoluminescence spectra of PbBr-based layered perovskites with an organic layer of naphthalene-linked ammonium molecules. Profiles: (a) 1; (b) 2; (c) 3; (d) 4; (e) 5.	Luminescence
Labels: Application (APL), Characterization method (CMT), Descriptor (DSC), Material (MAT), Property (PRO), Synthesis method (SMT), Symmetry/phase label (SPL)	

T. Gupta et al., MatSciBERT: A materials domain language model for text mining and information extraction, npj Computational Materials 8, 102 (2022).

ClimateBERT

気候関連のテキストで事前学習したトランスフォーマーベースの言語モデル

- 気候関連のテキストには一般的な言語モデルでは正確に表現できない部分が含まれている。
- ニュース記事や論文などのさまざまなソースからクロールされた気候関連のテキスト（200万段落）で事前学習。
- マスクされた単語の復元タスクで既存モデルを48%改善。
- さまざまな気候関連のダウンストリームタスク（テキスト分類、感情分析、ファクトチェックなど）におけるエラー率を3.57%から35.71%低下させた。

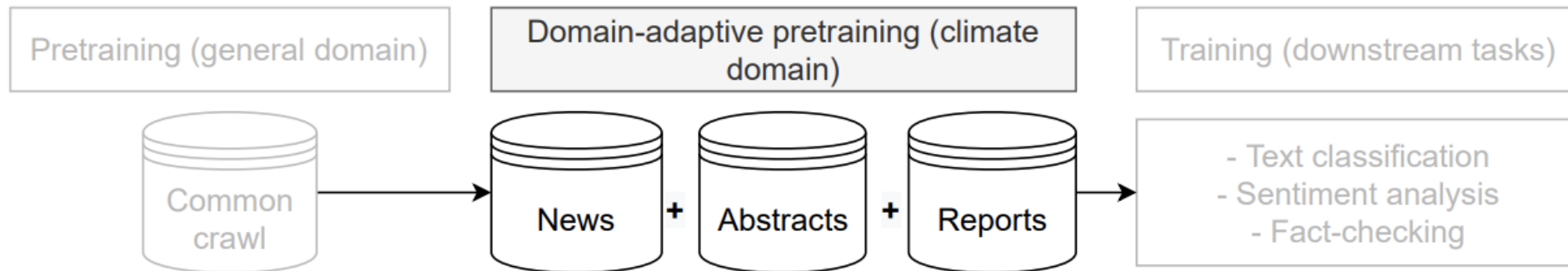


Figure 1: Sequence of training phases. Our main contribution is the continued pretraining of language models on the climate domain. In addition, we evaluate the obtained climate domain-specific language models on various downstream tasks.

外部リソースの利用

複数のソースにアクセスするエージェントとして機能する大規模言語モデル

- 気候・気象分野では正確さや精度が重要だが、LLMは不正確な内容を生成しやすく、また最近の情報に欠けることも多い。
- より信頼性の高い情報を提供するために、複数のソースにアクセスするエージェントとして機能するLLMを構築。
- ClimateWatchからデータを取得し、Google検索を活用するエージェントのプロトタイプを提案。

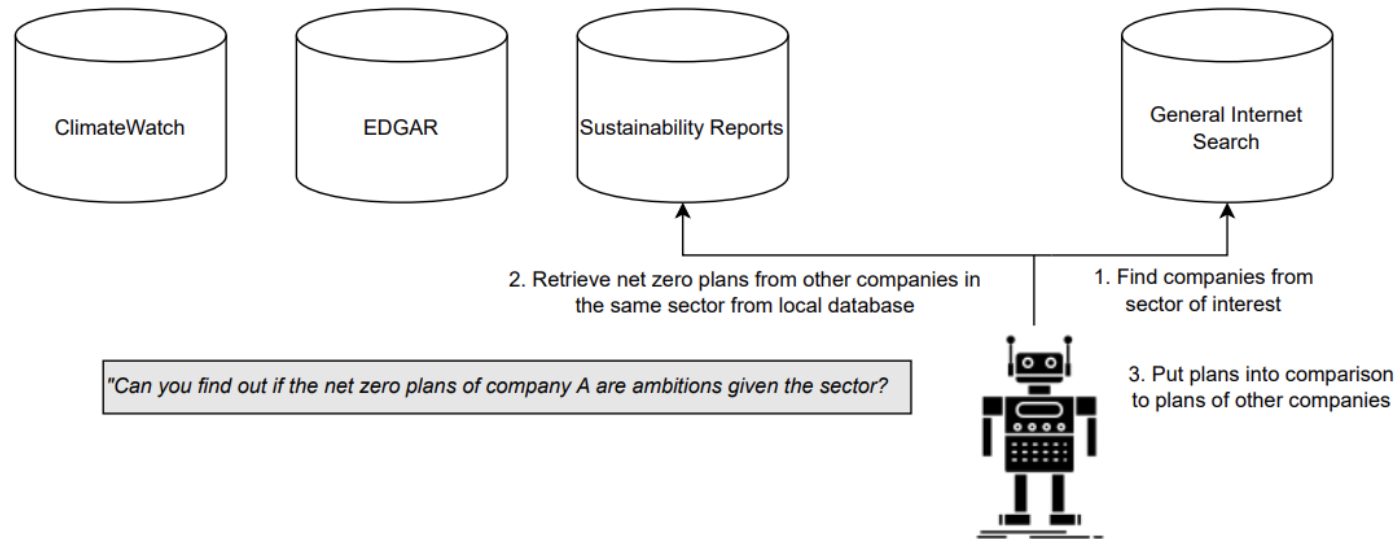
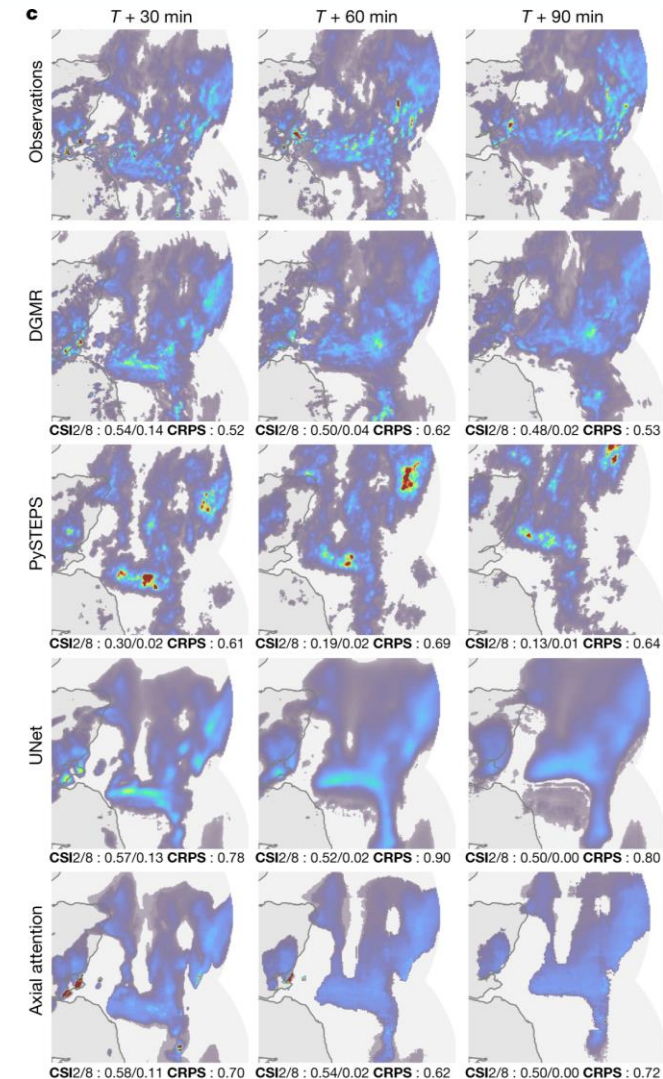
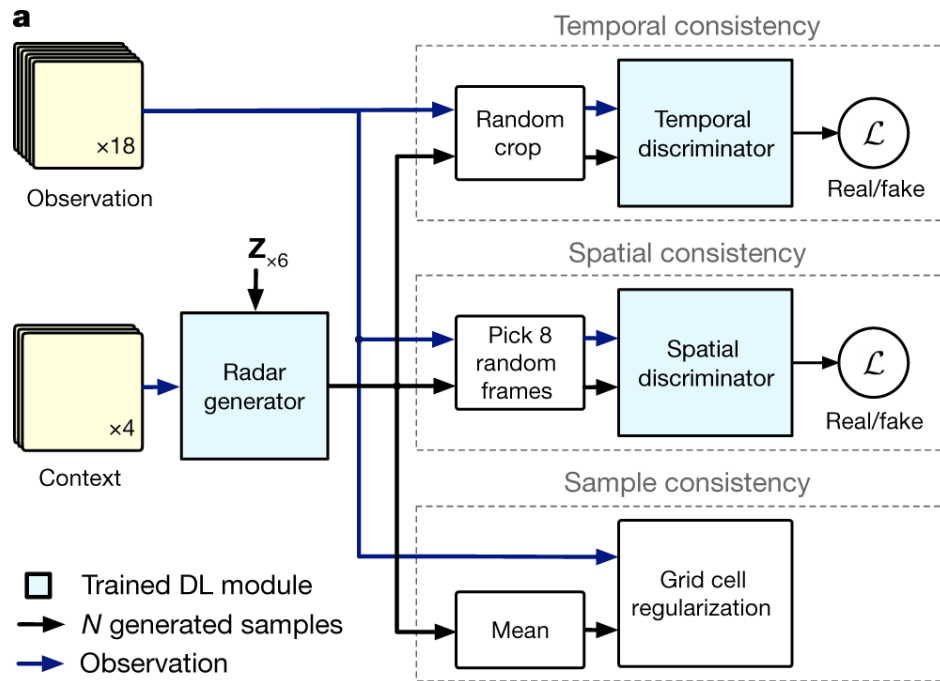


Figure 1: Setup for LLM agents that can access multiple sources, such as databases or general internet search.

Deep Generative Models of Rainfall (DGMR) DeepMind

短期の降水量予測用のAIシステム

- データセットとトレーニング済みモデルが DeepMindにより公開（オープンソース）。
- 58人の専門気象学者が、他の既存の方法との比較で実施した評価で、DGMRはテストケースの89%で精度と有用性で1位にランクインした。



ChatGPTが書いた論文の抄録

人間のレビュワーに区別が難しい抄録も生成される

- ChatGPTが医学分野の論文の抄録を生成できるかどうかを調べた論文。
- 5つの高インパクトファクター雑誌から集めた50の抄録の論文タイトルを入力しChatGPTで抄録を生成。
- AI出力検出器、盗作検出器、人間のレビュワーでその抄録がオリジナルか生成されたものかを判断。
- AI出力検出器はChatGPTが生成した抄録のほとんどを正しく検出できた（AI出力検出器のAUROCは0.94）。
- 盗作検出器はChatGPTが生成した抄録にも非常に高いオリジナリティスコアを与えた（例えば、ChatGPTが生成した抄録はオリジナル抄録と同じ患者コホートサイズであっても数字は捏造されていた）。
- 人間のレビュワーはChatGPTに生成された抄録の68%を正しく識別したが、14%のオリジナル抄録を誤ってChatGPTが生成したものとして識別。

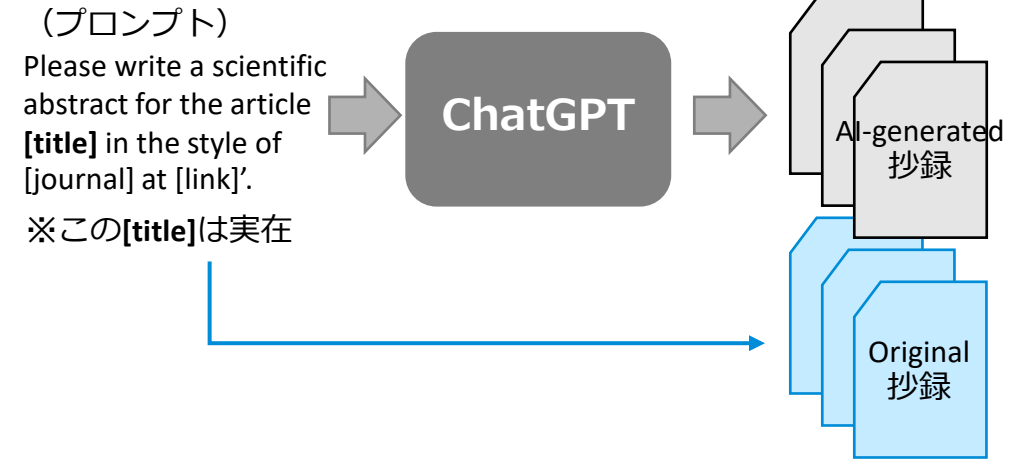


Table 1. Human reviewer scoring for whether abstracts were real or generated, along with truth.

		Truth	
		Original	Generated
Reviewer guess	Original	43	16
	Generated	7	34

オリジナルなのに
ChatGPT生成と誤認識

ChatGPT生成なのに
人間がオリジナルと誤認識

ChatGPTが共著者として含まれた論文

US Medical Licensing Exam (USMLE)という米国の医師免許試験をChatGPTで解く

- USMLEは医師の知識基盤のすべてをカバーする人間向けの試験。複雑な臨床データをもとに答える記述式の問題も含まれている。
- 公開されている376問の試験問題から305問を選んでChatGPTに解かせた。回答を2人の医師が採点。
- ChatGPTは94.6%の一致率で問題内容に沿った解答を出力、全項目の50%以上の正答率だった。そのうちのほとんどが正答率60%を超えており、USMLEの合格基準が正答率約60%程度であることから、ChatGPTはギリギリ合格圏内とされた。

T. H. Kung et al., Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models, PLOS Digit Health 2(2): e0000198 (2023).

medRxiv THE PREPRINT SERVER FOR HEALTH SCIENCES

CSH Cold Spring Harbor Laboratory BMJ Yale

Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models

Tiffany H. Kung, Morgan Cheatham, ChatGPT, Arielle Medenilla, Czarina Sillos, Lorie De Leon, Camille Elepaño, Maria Madriaga, Rimel Aggabao, Giezel Diaz-Candido, James Maningo, Victor Tseng

doi: <https://doi.org/10.1101/2022.12.19.22283643>

Now published in PLOS Digital Health doi: [10.1371/journal.pdig.0000198](https://doi.org/10.1371/journal.pdig.0000198)

Abstract Full Text Info/History Metrics

<https://www.medrxiv.org/content/10.1101/2022.12.19.22283643v2>

(プレプリント)
共著者にChatGPT !?

共著者にChatGPTは
含まれていない

BROWSE PUBLISH ABOUT

PLOS DIGITAL HEALTH

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models

Tiffany H. Kung, Morgan Cheatham, Arielle Medenilla, Czarina Sillos, Lorie De Leon, Camille Elepaño, Maria Madriaga, Rimel Aggabao, Giezel Diaz-Candido, James Maningo, Victor Tseng

Published: February 9, 2023 • <https://doi.org/10.1371/journal.pdig.0000198>

Article	Authors	Metrics	Comments	Media Coverage
---------	---------	---------	----------	----------------

GPT-3にGPT-3自身について書かせた論文

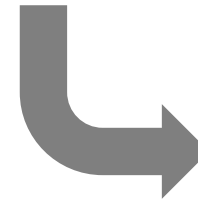
GPT-3を使ってGPT-3に関する論文を2時間で完成させた

- スウェーデンのある研究者がGPT-3に自分自身について論文を書くよう命令した。
- プロンプトには「GPT-3について500単語で論文を書き、参考文献を文中に加え」などのほか、細かなパラメータの工夫がされている。
- 論文を投稿中（投稿先は非公開）。投稿では著者の氏名や住所の入力、著者全員の同意をとるなどの手続きが大変だったという*

* 著者の1人（Almira Osmanovic Thunström氏）がScientific Americanでこの件について記事を書いている
<https://www.scientificamerican.com/article/we-asked-gpt-3-to-write-an-academic-paper-about-itself-mdash-then-we-tried-to-get-it-published/>

プロンプトの例

Prompt: Write an academic thesis in 500 words about GPT-3 and add scientific references and citations inside the text.
Temperature: 0.77 / Maximum length 458 / Top P 0.9 / Frequency Penalty 0.95 / Presence Penalty 0.95 / Best of n=18
First prompt output chosen



Introduction

GPT-3 is a machine learning platform that enables developers to train and deploy AI models. It is also said to be scalable and efficient with the ability to handle large amounts of data. Some have called it a "game changer" in the field of AI (O'Reilly, 2016). GPT-3 has been used in a number of different applications including image recognition, natural language processing, and predictive modeling. In each of these cases, GPT-3 has demonstrated its potential to improve upon existing methods (Lee, 2016).

One area where GPT-3 shows particular promise is in natural language understanding. Traditional approaches to this problem have relied on rule-based systems which are limited in their ability to deal with the complexities of real-world language use (Wang et al., 2015). In contrast, GPT-3 uses a neural network approach which can learn from data and generalize beyond what was seen during training (Deng et al., 2014). This makes it well suited for tasks such as machine translation and question answering which require an understanding of natural language.

So far, GPT-3 has been shown to outperform existing methods on several benchmark datasets (Deng et al., 2014; Lee, 2016). For example, on the task of machine translation from English to Chinese, GPT-3 achieves a translation accuracy of 78.4%, compared to the previous best result of 73.9% (Deng et al., 2014). Similarly, on the Question Answering dataset from Stanford University's SQuAD challenge, GPT-3 achieves an accuracy of 82.7%, compared to the previous best result of 80.6% (Lee, 2016).

These results suggest that GPT-3 has great potential for furthering our ability to build intelligent systems that can understand and respond meaningfully to natural language inputs. However, there are still many challenges that need to be addressed before this technology can be deployed in practical applications. For instance, current neural network models are not very efficient at handling long input sequences (Hochreiter & Schmidhuber 1997), so there is room for improvement in this area. Additionally, while GPT-3 has shown good performance on standard benchmarks, it remains to be seen how well it will generalize to real-world data which is often more noisy and unstructured.

Overall, GPT-3 is a promising platform for AI development that shows great potential for furthering our ability to build intelligent systems. While there are still some challenges that need to be addressed, the results so far suggest that GPT-3 is a promising step forward in the field of AI.

Prompt: Write an academic thesis in 500 words about GPT-3 and add scientific references and citations inside the text.
Temperature: 0.77 / Maximum length 458 / Top P 0.9 / Frequency Penalty 0.95 / Presence Penalty 0.95 / Best of n=18
First prompt output chosen

Can GPT-3 write an academic paper on itself, with minimal human input?
<https://hal.science/hal-03701250>

2

Science誌はAIを著者として認めず

文章生成AIを著者として認めないとライセンスと編集方針が改定された

- ChatGPT(またはその他のAIツール)によって生成されたテキストの論文への使用は不可。
- AIツールによる図・画像、グラフィックスの生成も不可。
- Scienceに論文を投稿するにあたり著者は「この論文はオリジナルである」というライセンスに署名するが、ChatGPTで作られた文章が「オリジナル」とは受け入れられないというのが論拠。
- 論文の内容に対する著者の説明責任という観点でもAIが出力したテキストの使用は認められない。
- ソープ編集長は、この方針への違反は、画像加工や先行研究盗用のような不正行為と変わらないと指摘。

The screenshot shows the Science journal website. At the top, the Science logo is on the left, and navigation links for 'Current Issue', 'First release papers', 'Archive', and 'About' are on the right, along with a 'Submit manuscript' button. Below the header, the breadcrumb trail reads 'HOME > SCIENCE > VOL. 379, NO. 6630 > CHATGPT IS FUN, BUT NOT AN AUTHOR'. The article title 'ChatGPT is fun, but not an author' is prominently displayed, with the author 'H. HOLDEN THORP' and a link to 'Authors Info & Affiliations'. Below the title, the journal information 'SCIENCE • 26 Jan 2023 • Vol 379, Issue 6630 • p. 313 • DOI: 10.1126/science.adg7879' is shown, along with a view count of 95,041 and 60 comments. A 'RELATED LETTERS' section follows, featuring 'AI tools can improve equity in science' by Violeta Berdejo-Espinola and Tatsuya Amano, and an 'Editor's note' by H. Holden Thorp and Valda Vinson. A portrait of H. Holden Thorp is shown on the left, and a text excerpt on the right discusses the popularity of ChatGPT and its use in rewriting text from the play 'Death of a Salesman'.

Natureでの議論

ChatGPTが書いた論文抄録がレビュワーを欺くことができるという結果についてのニュース
Abstracts written by ChatGPT fool scientists
<https://www.nature.com/articles/d41586-023-00056-7>

ChatGPTなどのAIが論文著者にクレジットされているケースが少なくとも4件は確認されている。
ChatGPT listed as author on research papers: many scientists disapprove
<https://www.nature.com/articles/d41586-023-00107-z>

研究アシスタントとしてAIを使う上での注意点やルールについて議論。
Tools such as ChatGPT threaten transparent science; here are our ground rules for their use
<https://www.nature.com/articles/d41586-023-00191-1>

GPT-4の登場に対する研究者の反応（割とネガティブな反応）
GPT-4 is here: what scientists think
<https://www.nature.com/articles/d41586-023-00816-5>

LLM時代の科学「未解決問題」

正確性、信頼性、説明責任

- ハルシネーション
- 変化への対応力
- 稀な現象
- 研究インテグリティ
- LLM支援の程度の定量化
- 説明責任

説明可能性、欠落、バイアス

- 不透明度
- 説明可能性
- 欠落
- 選択
- バイアス

科学的な創意工夫と発見

- パラダイムシフト
- 外れ値
- 科学的創造性
- Deskilling

科学評価と査読

- 品質の評価
- 科学の精神

Box 1

Open questions

Accuracy, reliability and accountability

- Hallucination: How can scientists methodically determine when large language models (LLMs) are ‘hallucinating’ or generating inaccurate and fantastical content? How can scientists best assess and work around these tendencies to generate unreliable or non-factual outputs?
- Responsiveness to change: If LLMs fail to extrapolate effectively when world knowledge changes or data distributions drift over time, how can scientists safeguard their accuracy, reliability and responsiveness to change?
- Sparse phenomena: If LLMs struggle to reliably generate accurate content for infrequent or sparsely studied phenomena, how do scientists draw on LLMs to inform insights about anomalies, new discoveries or unprecedented observations?
- Research integrity: What is plagiarism and authorial misrepresentation in the age of LLMs? How should scientists be held accountable for plagiarism and authorial misrepresentation? What checks should be put in place to establish the authenticity of scientific publications?
- Quantifying the degree of LLMs assistance in writing: What is acceptable and what is not?
- Accountability: Who is responsible for the integrity of scientific research and the content of scientific papers aided by LLMs? Who is accountable?

Explainability, missingness and bias

- Opacity: How can opaque LLMs justifiably be integrated into the scientific method?
- Explainability: How can the original sources be traced back? How can scientists, who draw on opaque LLMs, clarify the intended meaning or nuances of the texts based on which such models render their outputs? Does a lack of interpretability undermine the justifiability of relying on inferences drawn from LLMs?
- Missingness: If scientific papers represent the final product of a research process rather than a full picture of the complex choices, practices and contexts that underlie the research (that is not

all research is documented, in particular failures and negative results), how can the inferences generated by LLMs (which only process the information scientific articles, textbooks, websites and so on) account for the missingness that derives from the limitations of such a ‘tip-of-the-iceberg’ view?

- Selection: How can LLMs account for outdated or incorrect knowledge in the published literature?
- Bias: How can potential biases in the training data sets of LLMs — and other social, statistical and cognitive biases that may arise in their design, development and deployment — be most effectively assessed? How will LLMs enhance existing and introduce new biases or help remove existing ones?

Scientific ingenuity and discovery

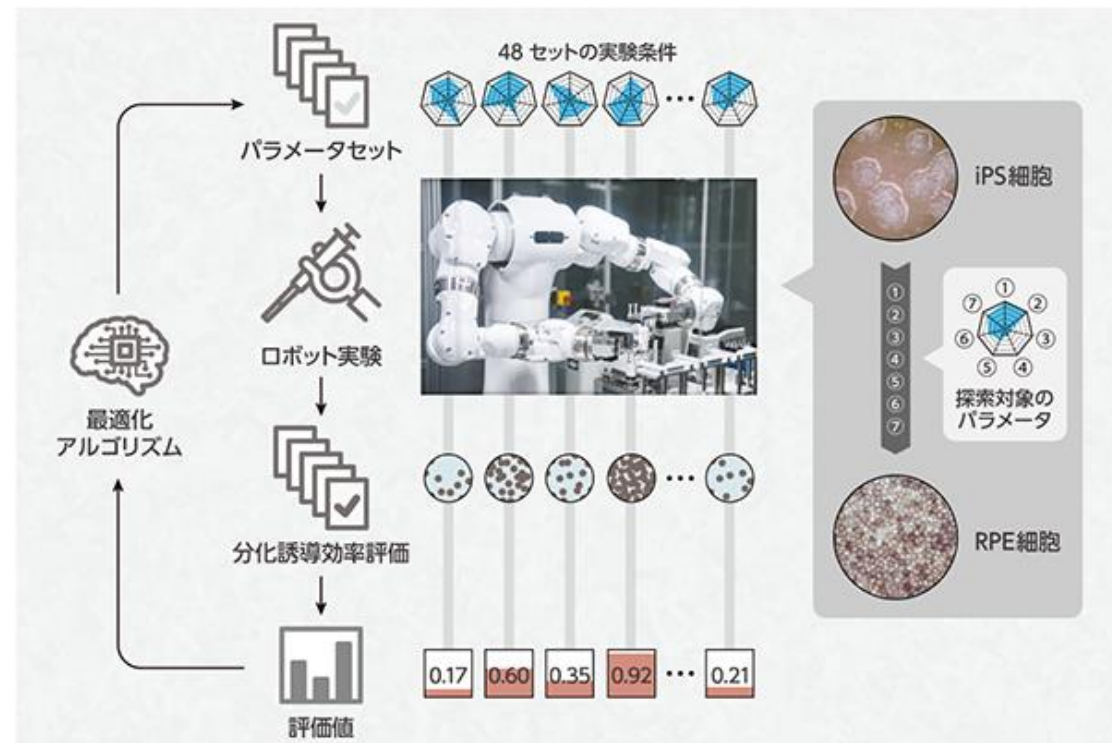
- Paradigm shifts: How can LLMs accommodate future ‘paradigm shifts’ in scientific understanding? Could LLMs (which generate insights by identifying patterns emergent from past research — potentially engendering paradigm lock-in and stifling novelty) function to tamp down possibilities for new scientific directions?
- Outliers: Will outliers (radical new ideas, unconventional views and unusual writing styles) be lost, overlooked or averaged out?
- Scientific creativity: What is the role of the scientist in the age of LLMs? What is the role of scientific creativity?
- Deskilling: Will overreliance on LLMs to produce arguments and text risk diminishing or weakening the writing and critical thinking skills and insight of researchers?

Science assessment and peer review

- Assessing quality: How do we assess high-quality science in the age of LLMs? What role should the values of reproducibility/ replicability and transparency play?
- Ethos of science: How do we trust science in the age of LLMs? How, if at all, do the values of objectivity, rigour and accountability change with the scaled integration of LLMs into scientific practices?

ロボティックバイオロジー

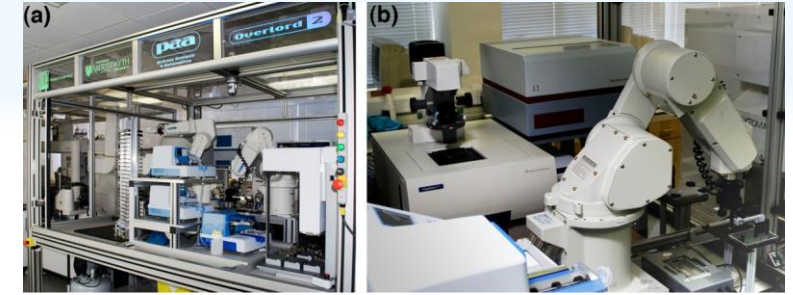
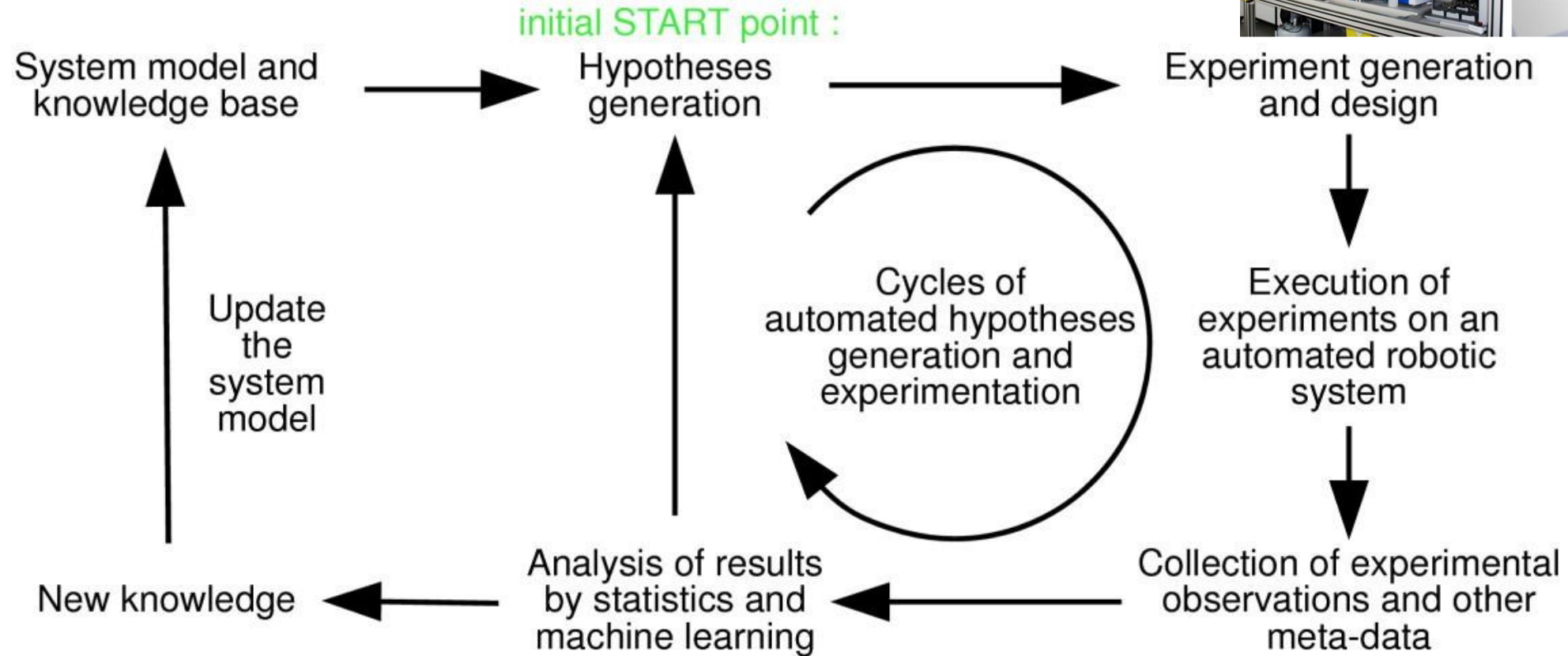
細胞レシピを自律的に試行錯誤するロボット・AIシステムの構築



理化学研究所ほかプレス発表「再生医療用細胞レシピをロボットとAIが自律的に試行錯誤－ロボット・AI・人間の協働は新しいステージへ－」（2022年6月28日）
https://www.riken.jp/press/2022/20220628_2/index.html

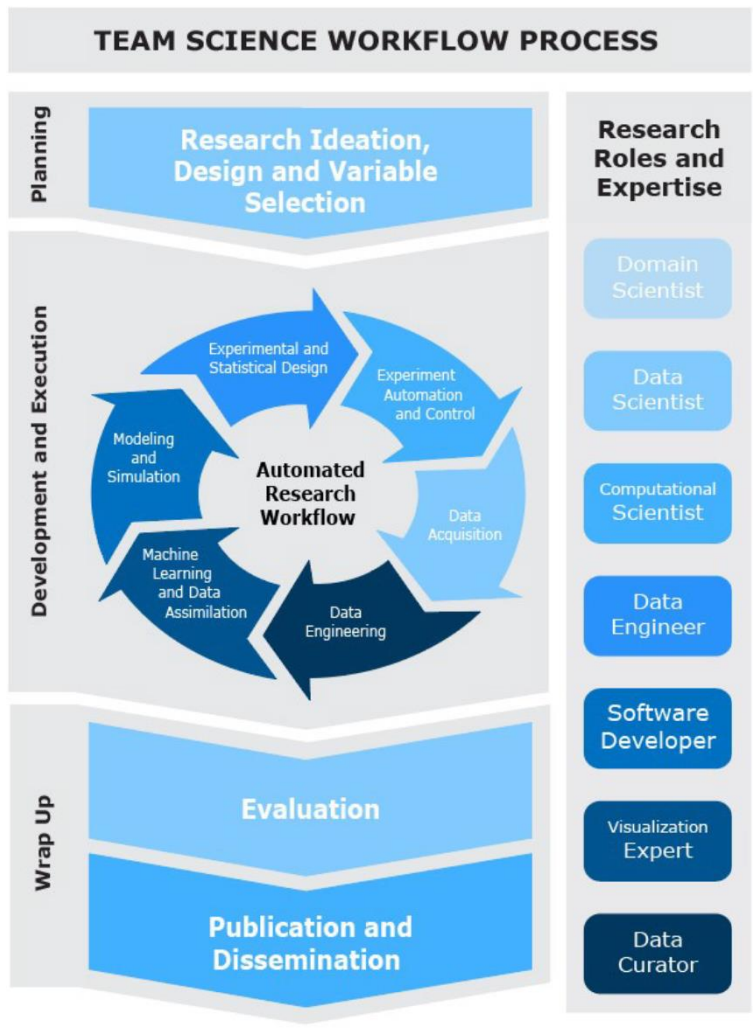
ロボット科学者

Hypothesis-driven closed-loop learning



米国 National Academies

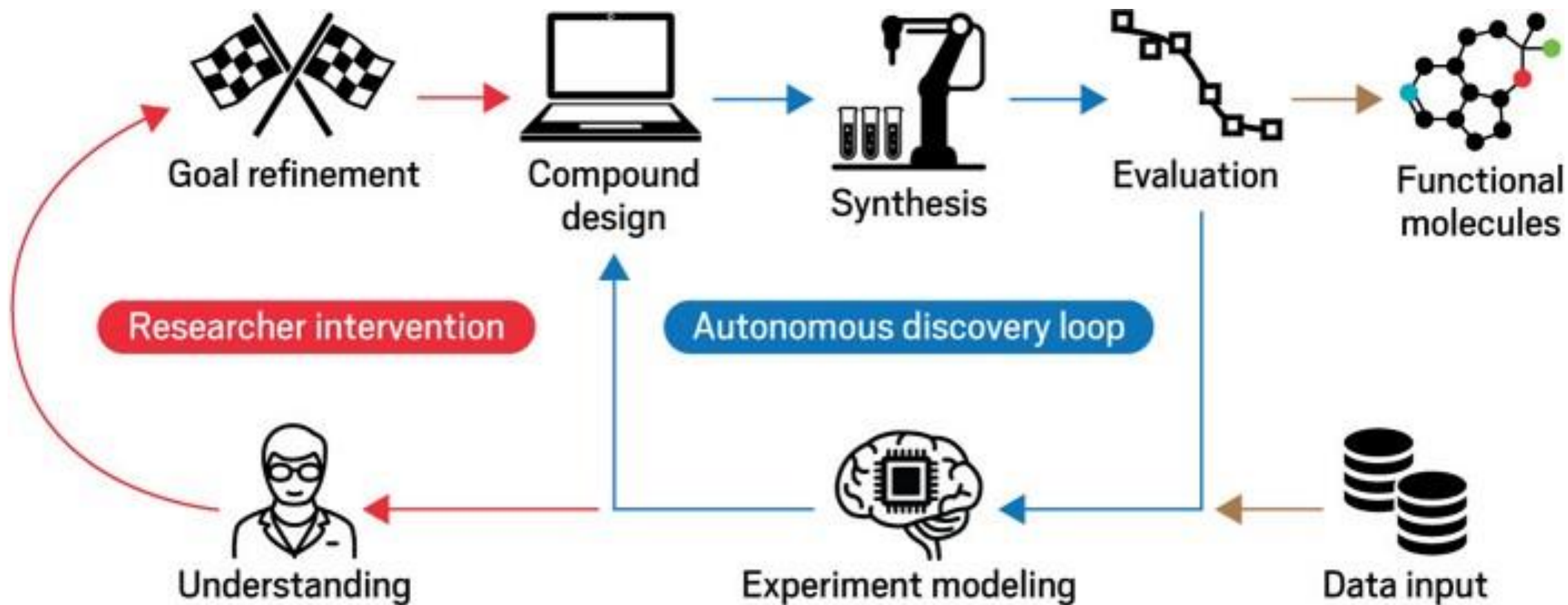
Automated Research Workflows for Accelerated Discovery (2022)



Airflow	https://airflow.apache.org
Bigtable	https://cloud.google.com/bigtable
Chimera	https://github.com/hysds/chimera
Cromwell	http://cromwell.readthedocs.io/
Cyverse Discovery Environment	https://cyverse.org/discovery-environment
Fireworks	https://materialsproject.github.io/fireworks
Hadoop	https://hadoop.apache.org
Galaxy	https://galaxyproject.org
iRODS	https://irods.org
Jupyter	https://jupyter.org
Kepler	https://kepler-project.org
Nextflow	https://www.nextflow.io
Open Science Framework	https://osf.io
Luigi	https://luigi.readthedocs.io/en/stable/workflows.html
Parsl	http://parsl-project.org
Pegasus	https://pegasus.isi.edu
Snakemake	https://snakemake.readthedocs.io/en/stable
Spark	https://spark.apache.org
Starfish Storage	https://starfishstorage.com
Wolfram	https://www.wolframcloud.com

スマートラボで新規化合物の発見加速

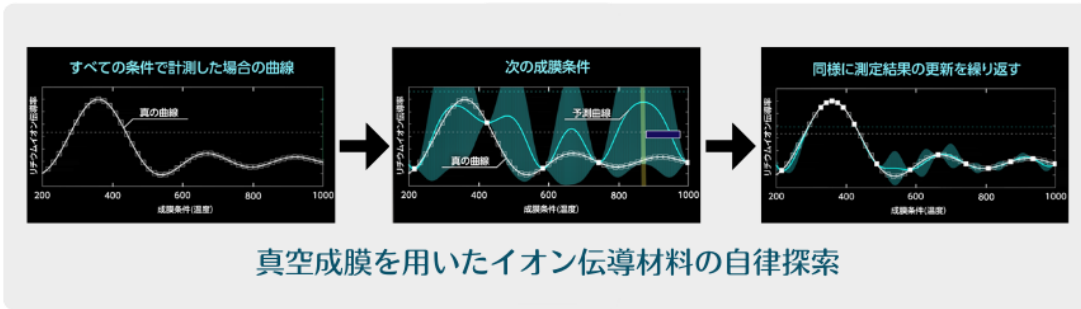
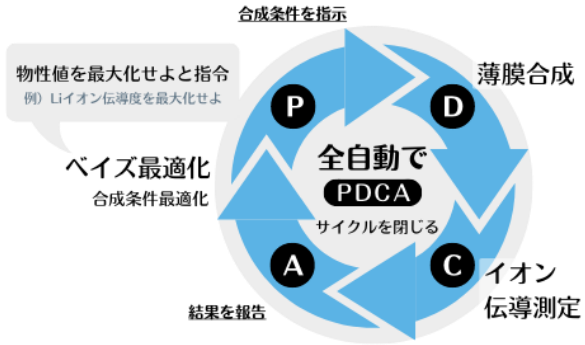
人間がゴールを与え、自動発見ループが答えを提案



AI・ロボット駆動科学におけるループの重要性

JST未来「マテリアル探索空間拡張プラットフォームの構築」

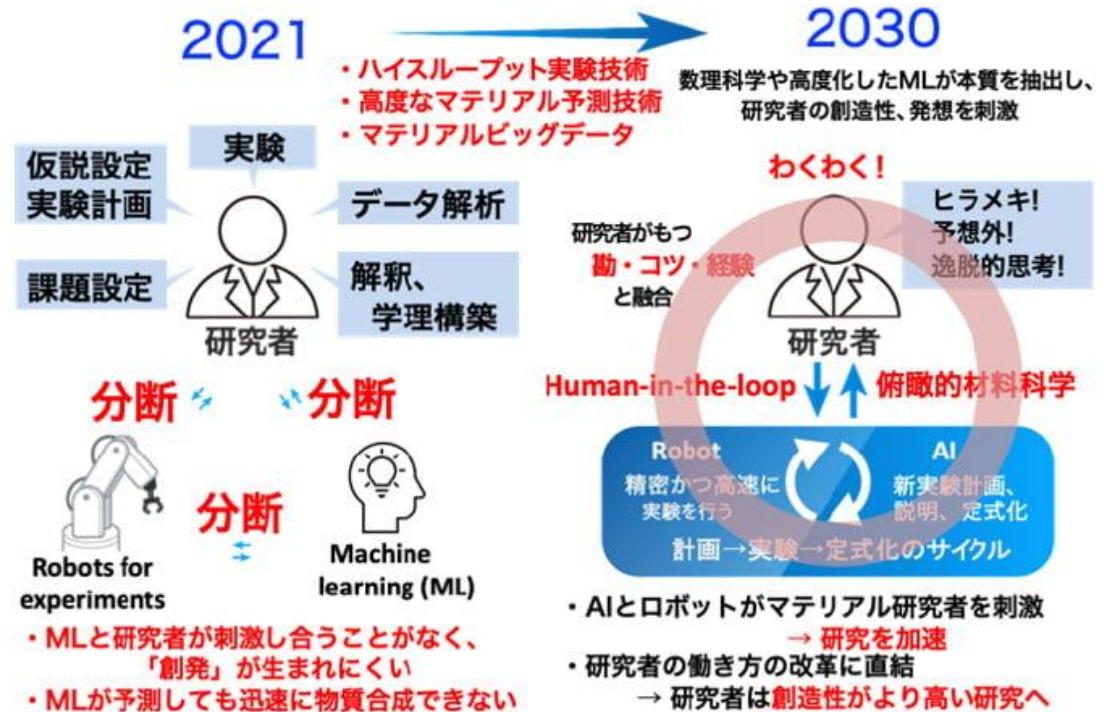
MS目標3「人とAIロボットの創造的共進化によるサイエンス開拓」



自律実験システムを用いたハイスループット探索

<https://meep.nagato-u-tokyo.jp/>

科学研究の進め方: パラダイムシフト



<https://sites.google.com/g.ecc.u-tokyo.ac.jp/moonshot-ai-science-robot/>