

# 生成AIの仕組みと社会への影響

東京大学 大学院工学系研究科 教授 川原圭博

# 本日の講演のポイント

生成AIは、基本的な深層学習の発展に加え、いくつかのブレークスルーに繋がる諸技術が重なり誕生した。本講演では、その技術的な構成要素を紹介し、科学者、国、社会にとっての意義についての私見を共有する。

1. 技術的な意義
2. 各研究者にとっての意義
3. 国としての意義

# ChatGPTなどの大規模言語モデルができること

AI技術の進化に伴い、大規模言語モデルが登場

文章の翻訳・言い換えの精度改善に加え、対話・要約・生成などで**従来不可能であったレベルを達成**

ChatGPTや生成系AIと呼ばれる技術の用途



文章の翻訳・言い換え



問い合わせ等の自動化  
内容の整理・要約



文章や絵の生成



✓ 従来よりも精度が向上



✓ **従来不可能であったレベルを達成**

# 現在、研究者は何を思っているのか

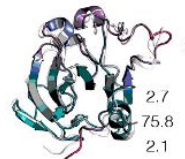
## 驚異

- **言語モデルでそこまでできるのか！**
  - 言語モデルは原理上「テキストの続き」を予測するだけなのに
- **データから法則性を導き出し、具体化させる脅威的能力**
  - 科学技術全体に不可欠なツールになる

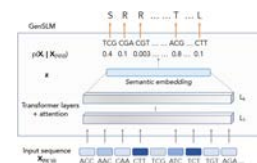
## 脅威

- 大量の計算資源と人的資源が必要で**大学・公的機関では今は手が出ない**
- 独占・非公開のため性能の検証や再現が困難

## サイエンス分野での基盤モデルも登場しつつある



タンパク質構造  
予測: ESM-2  
(Meta社)



ゲノム解析  
: GenSLMs  
コロナウィルスの  
変異を解析

化合物生成や画像診断、ロボットなど他分野へのインパクトも大きく、まだまだ「大発見」が続くだろう

# AIの変遷

第一次AIブーム:

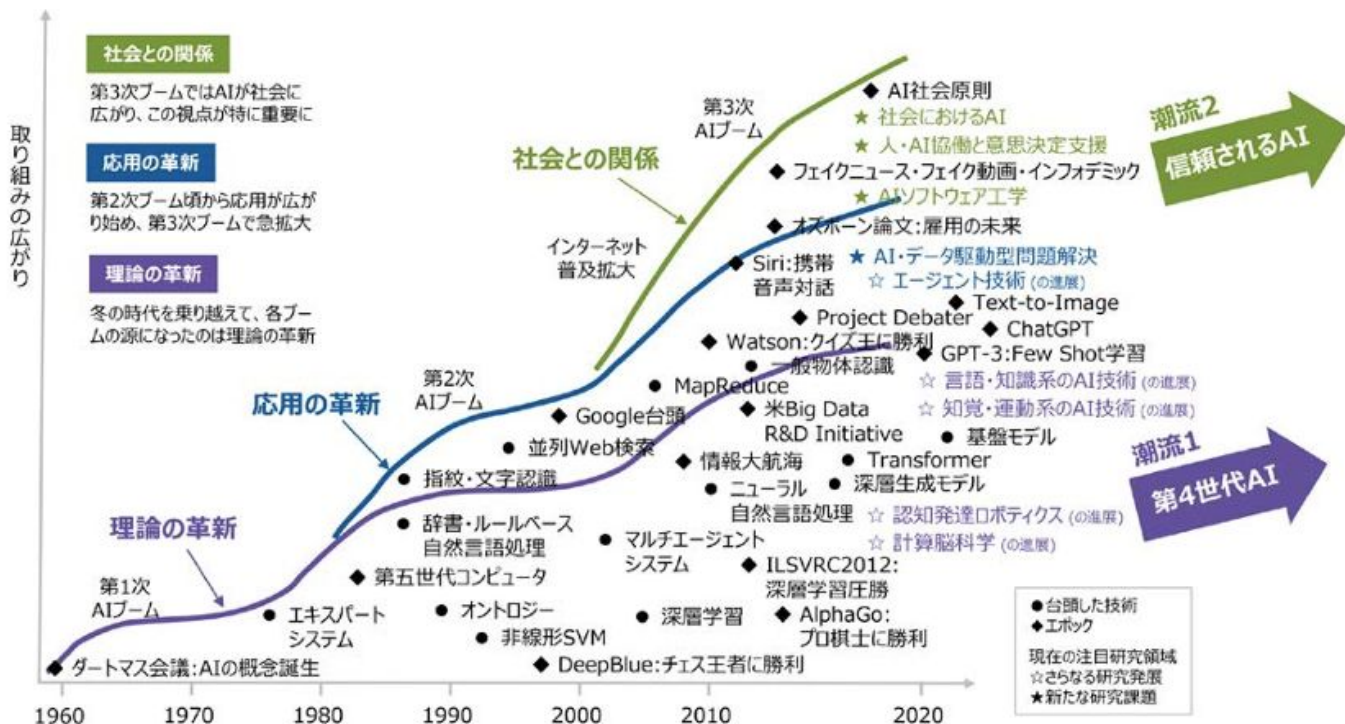
概念の誕生

第二次AIブーム:

ルールベース

第三次AIブーム:

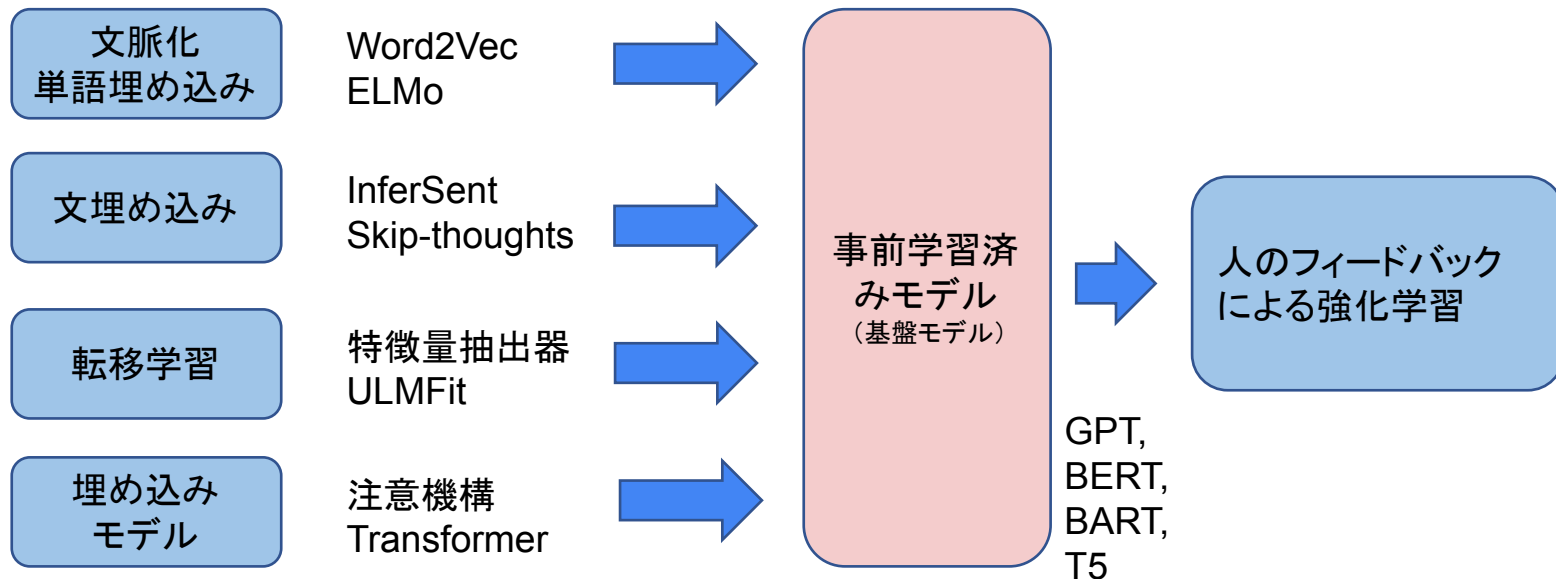
ビッグデータ、  
機械学習



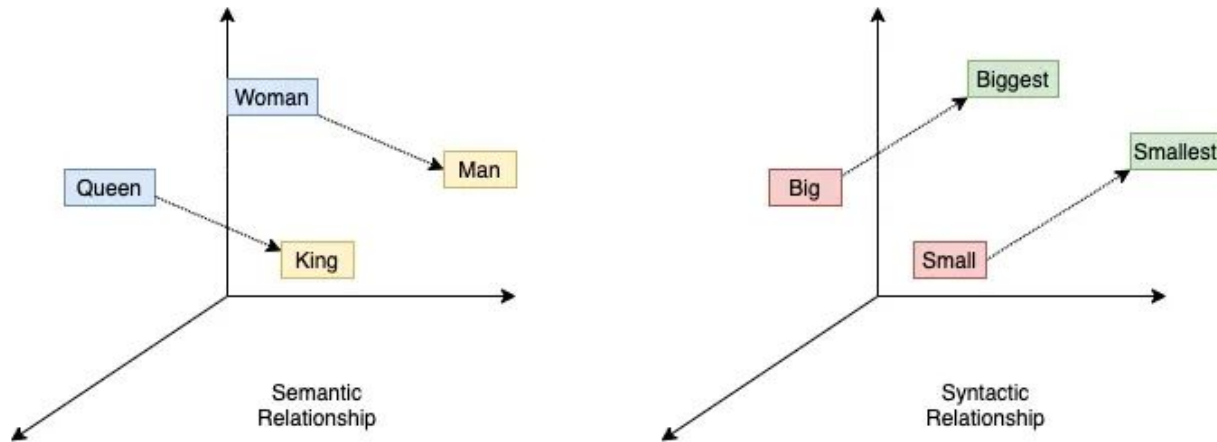
JST CRDS 人工知能研究の新潮流2 ~基盤モデル・生成AIのインパクト~より

# そもそも、自然言語処理は、なぜここまで発達したのか？

- 基本的な深層学習の発展に加え、ブレークスルーに繋がる諸技術が発明され、データから抽象的意味を取り扱えるようになった。
- GPUも元はゲーム用のアクセラレータであり、AIに使えられてなかった。



# 単語の埋め込み表現



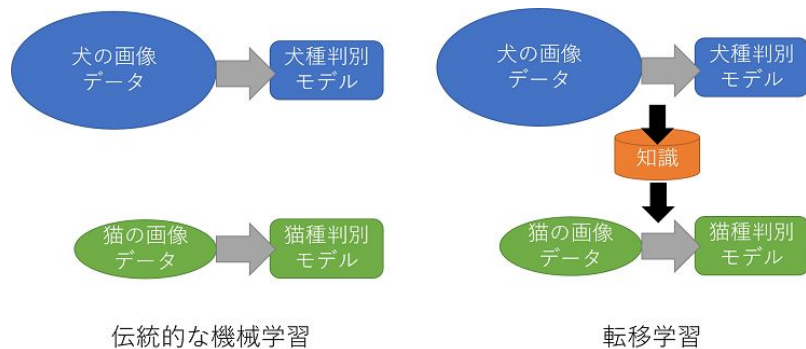
Vector("King")-Vector("Man")= Vector("Queen")-Vector("Woman").

7

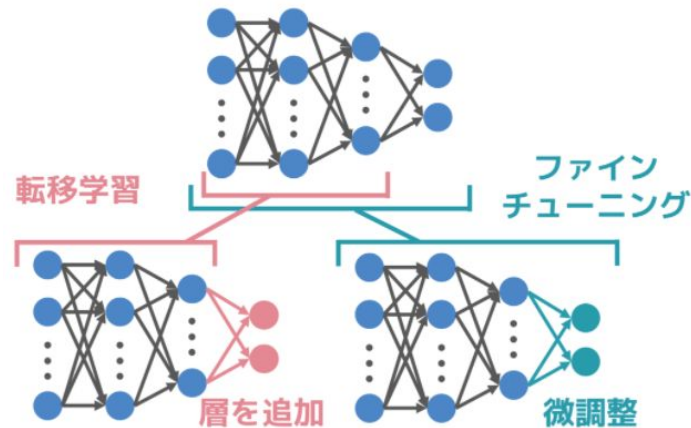
- Word2vecで単語の意味や関係を数値的な表現として扱うことが可能に
- ELMoではさらに文脈に沿った埋め込み表現を獲得

# 転移学習・ファインチューニング

- 大量のトレーニングデータが必要なタスクにおいて、既存のモデルを利用することで、学習時間の短縮や精度の向上などを行う手法
- 当初自然言語処理では転移学習はうまくいかなかったがULMFitで改善



<https://ledge.ai/transfer-learning/>

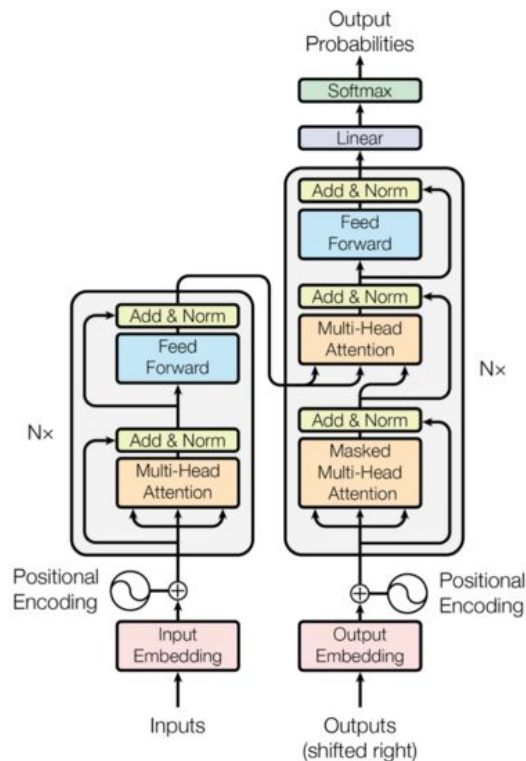


[https://jinbeizame.hateblo.jp/entry/kelpnet\\_transfer](https://jinbeizame.hateblo.jp/entry/kelpnet_transfer)

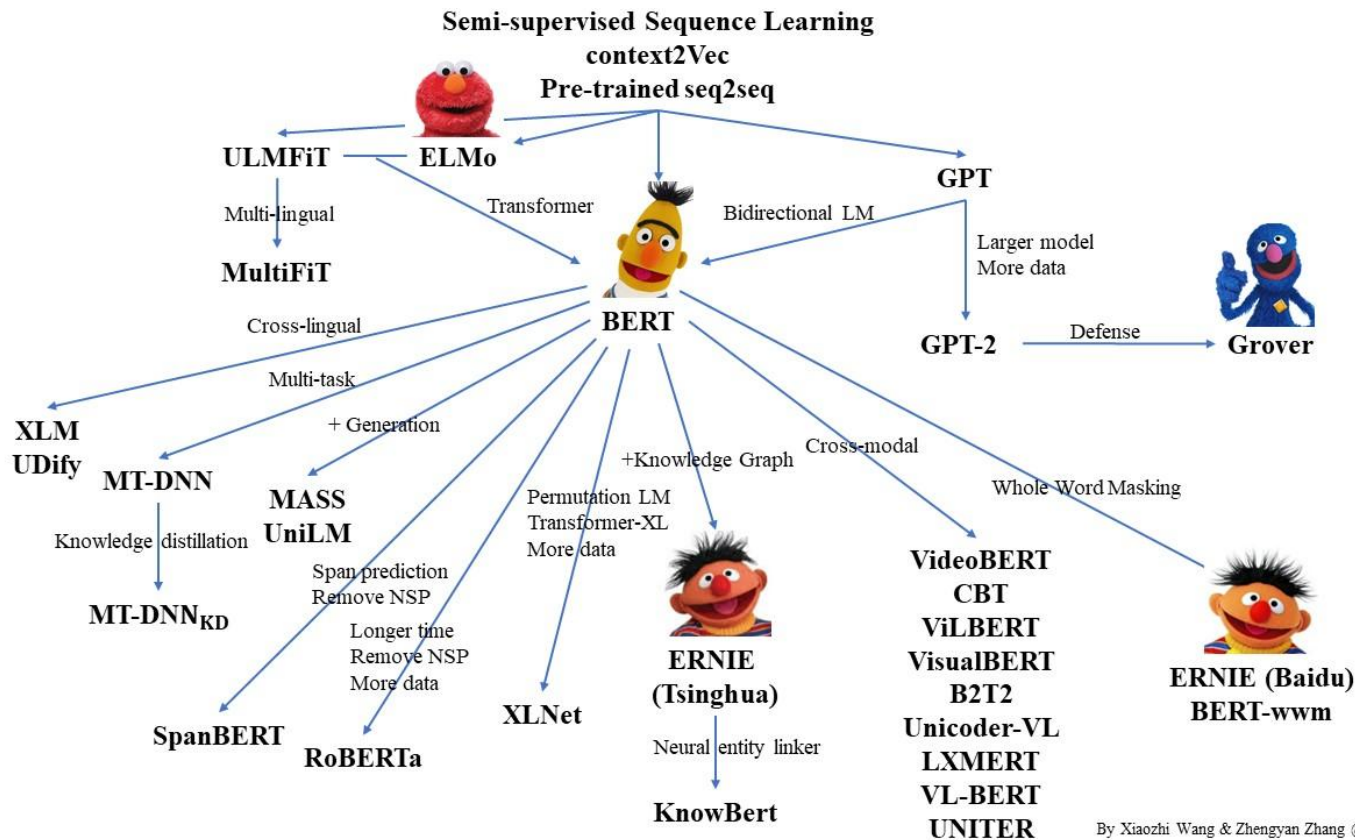


# 注意機構 (attention)

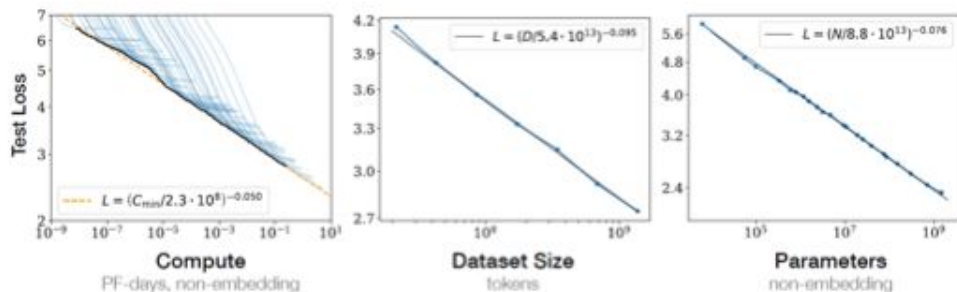
- 自然言語処理タスクにおいて、重要な単語や文脈を特定するための手法
- 文章中の単語に対する重みを付け、重要な単語に注目することで、文章の意味を正確に把握する
- Transformerでは、注意を非常にシンプルな数式に一般化。高速化が可能に。
  - 加法注目: 各単語の重要度を計算し、重要な単語に注目することで、文章の意味を把握する。
  - 内積注目: 各単語とクエリベクトルの内積に基づいて、重要な単語に注目することで、文章の意味を把握する。
  - ソースターゲット注目: 入力文章とターゲット文章を同時に考慮して、重要な単語に注目することで、文章の意味を把握する。
  - 自己注目: 各単語が他の単語に対する影響を計算し、重要な単語に注目することで、文章の意味を把握する。



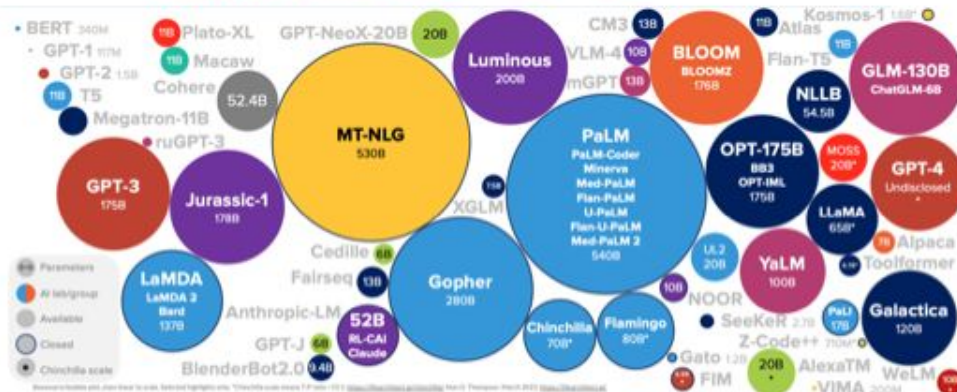
# これにより、良質なコーパスと計算資源を持つ者による事前学習済み言語モデルの開発が加速



# 言語モデルの超大規模化



言語モデルの性能と**計算能力**、**訓練データ量**、**パラメータ数**の間にべき乗則 (Kaplan+ 2020)



GPT-3以降、超大規模な言語モデルの構築が試みられる (Thompson 2023)

J Kaplan et al. (2020) Scaling Laws for Neural Language Models. *arXiv:2001.08361*.  
 A D Thompson. (2023) Language model sizes. LifeArchitect.ai. <https://lifearchitect.ai/models/>

大規模言語モデルの  
驚異と脅威  
岡崎 直観

# InstructGPT (Ouyang+ 2022)

- ChatGPTの基となるモデル（ChatGPTの詳細に関する論文は公開されていない）
- 人間のフィードバックによる強化学習（RLHF）により、プロンプトで与えられた人間の指示に従い、詳細な応答を返すように調整されたモデル

Step 1  
**Collect demonstration data,  
and train a supervised policy.**

A prompt is  
sampled from our  
prompt dataset.



A labeler  
demonstrates the  
desired output  
behavior.



This data is used  
to fine-tune GPT-3  
with supervised  
learning.



Step 2  
**Collect comparison data,  
and train a reward model.**

A prompt and  
several model  
outputs are  
sampled.



A labeler ranks  
the outputs from  
best to worst.



This data is used  
to train our  
reward model.



Step 3  
**Optimize a policy against  
the reward model using  
reinforcement learning.**

A new prompt  
is sampled from  
the dataset.



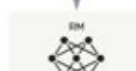
The policy  
generates an  
output.



The reward model  
calculates a  
reward for  
the output.



The reward is  
used to update  
the policy  
using PPO.



# ここまでやると何が起るか

## プロンプト

- テキストに続く単語を予測させる(通常の言語モデルの使い方)
  - 英国の首都は [ロンドン]
- タスクの説明によるプロンプト(zero-shot)
  - 次の質問に教えてください。英国の首都はどこでしょうか？
- タスクの説明と解き方の例を連結したプロンプト(few-shot)
  - 次の質問に教えてください。
  - 日本の首都は東京です。
  - 英国の首都はどこでしょうか？
- 
- ☺ タスクの解き方をテキストで与えるため、言語モデルが汎用的に振舞うように見える

## 思考の連鎖

- 解答例をモデルに与えるときに「考え方」を含めるようにすると数学問題、常識推論、記号推論などのタスクの性能が大幅に改善

解き方の例

Standard Prompting	Chain-of-Thought Prompting
<p><b>Model Input</b></p> <p>Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?</p> <p>A: The answer is 11.</p> <p>Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?</p>	<p><b>Model Input</b></p> <p>Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?</p> <p>A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. <math>5 + 6 = 11</math>. The answer is 11.</p> <p>Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?</p>
<p><b>Model Output</b></p> <p>A: The answer is 27. ❌</p>	<p><b>Model Output</b></p> <p>A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had <math>23 - 20 = 3</math>. They bought 6 more apples, so they have <math>3 + 6 = 9</math>. The answer is 9. ✅</p>

解答例に考える

# 弱点・問題点

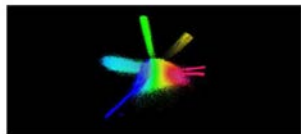
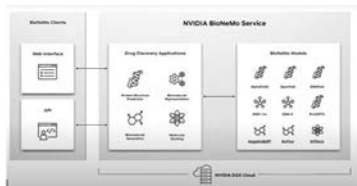
- 嘘 (Hallucination)
  - 確率的にもっともらしい回答をするだけなので、事実に基づかない回答をすることがある
- バイアス
  - データセットに含まれるデータの偏りが出力にそのまま反映される
  - 幅広く学んだモデルであっても意図的に出力の傾向を変えることができる
- 個人情報の暴露
  - データセットに含まれていた個人情報を出力してしまう可能性がある
- 人が騙される
  - 強化学習により、人が好みそうな回答をするようにチューニングされている

# さまざまなサイエンスへの応用が始まっており、世界中で次の戦いが始まる

## 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)



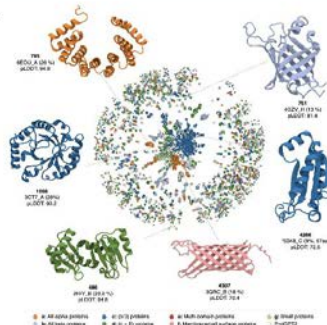
<https://blogs.nvidia.co.jp/2022/09/26/bionemo-large-language-models-drug-discovery/>



## 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のクラスター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。



<https://www.nature.com/articles/s41467-022-32007-7>



## 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
The composition of XYZ proteins of <b>ABC</b> is 123 °C for 1h and <b>DEF</b> for 2h. The <b>ABC</b> and <b>DEF</b> are marked with a heat enlarged section of <b>XYZ</b> protein.	Refinement
<b>ABC</b> design and the corresponding <b>DEF</b> protein takes 100h. <b>ABC</b> sample 1000(1) (a) and <b>DEF</b> sample 200(1) (b), identifying <b>ABC</b> and <b>DEF</b> genes.	FFT
The structure scheme of the bonding of <b>ABC</b> and <b>DEF</b> is shown in the figure. <b>ABC</b> is the <b>DEF</b> structure.	Crystal
(a) <b>ABC</b> protein of the <b>DEF</b> and <b>ABC</b> different building times. (b) intensity of <b>ABC</b> and <b>DEF</b> peaks count at (111) and (110) plane respectively as a function of <b>DEF</b> building time.	XRD
<b>ABC</b> occurs of <b>DEF</b> bond <b>ABC</b> with an <b>DEF</b> of sulphurated linked amine molecule. Profile: (a) 1, (b) 2, (c) 3, (d) 4, (e) 5.	Luminescence
Labels: Application (CAPS), Chemical element (C), Chemical element (O), Material (DEF), Protein (ABC), Chemical element (N), Chemical element (S).	

<https://www.nature.com/articles/s41524-022-00784-w>



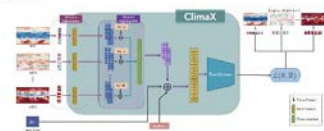
## ClimaX (Microsoft Research)

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

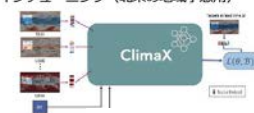
- CMIP6由来の気候データセットで事前学習済み (自己教師あり学習)
- ファインチューニングにより、事前学習では扱われなかった大気変数や時間空間スケールを含む様々な予測タスクに対処可能
- 既存のデータ駆動型ベースラインと比べてClimaXは汎用性があり、低解像度 (かつ少ない計算資源) で事前学習されている場合でも、天気予報や気候予測のベンチマークで優れたパフォーマンスを発揮。



事前学習



ファインチューニング (北米の地域予想用)

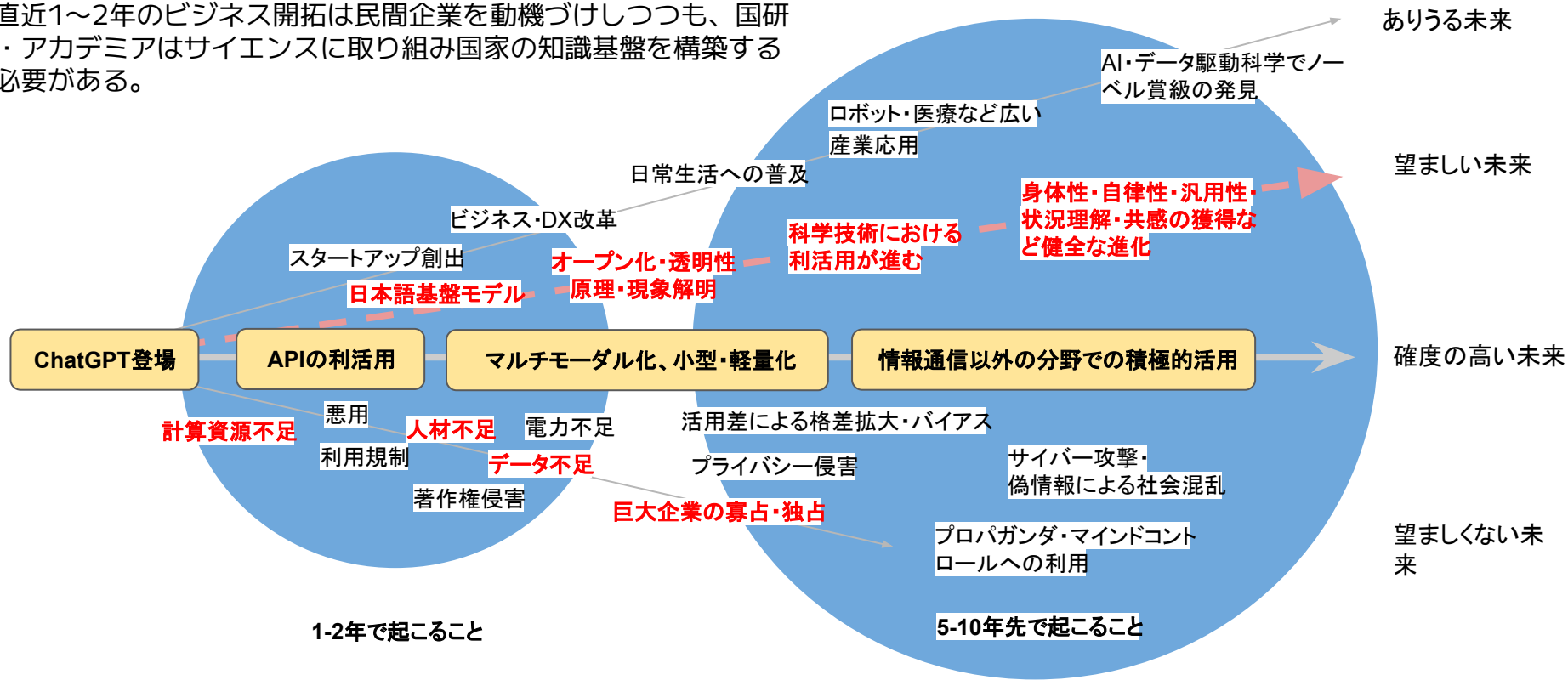


JST CRDS提供資料



# 5年先を見据えた産業界とアカデミア(国研・大学)の役割分担と連携

直近1~2年のビジネス開拓は民間企業を動機づけしつつも、国研・アカデミアはサイエンスに取り組み国家の知識基盤を構築する必要がある。



産業界

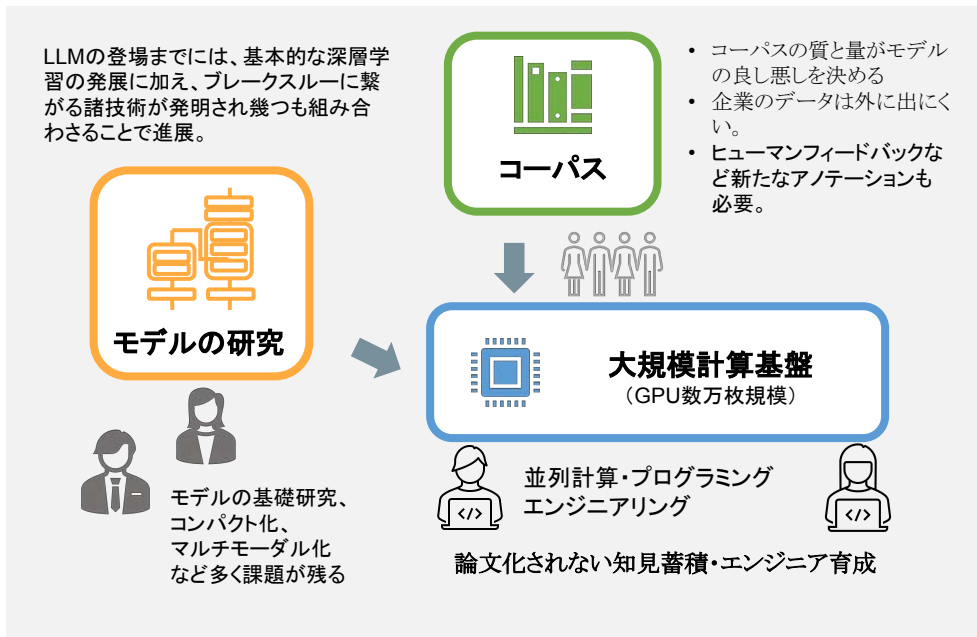
国研・アカデミアがサイエンスにおけるAIを開拓し、産業界に橋渡し



# 基盤モデルの作成に必要なこと

- 基盤モデルの研究環境は (1)モデル研究、(2)コーパス、(3)大規模計算基盤、(4)それを支える多くのエンジニアが不可欠。

## AI研究開発と計算基盤の整備



※ 国内のAI・情報研究者数十名へのヒアリングをもとに作成

# 日本においても、民間企業や大学、国研で急ピッチにキャッチアップを目指している



## LLM 勉強会

本勉強会では、自然言語処理および計算機システムの研究者が集まり大規模言語モデルの研究開発について定期的に情報共有を行っています。

具体的には、以下の目的で活動しています。

- オープンソースかつ日本語に強い大規模モデルの構築とそれに関連する研究開発の推進
- 上記に関心のある自然言語処理および関連分野の研究者によるモデル構築の知見や最近の研究の発展についての定期的な情報交換
- データ・計算資源等の共有を前提とした組織横断的な研究者間の連携の促進
- モデル・ツール・技術資料等の成果物の公開

国立情報学研究所を中心としたオープンソース開発



トップ > 広報 > プレスリリース > 2023年 > 日本語に特化した大規模言語モデル（生成AI）を試作

## 日本語に特化した大規模言語モデル（生成AI）を試作

～日本語のWebデータのみで学習した400億パラメータの生成系大規模言語モデルを開発～

### ポイント

- 400億パラメータの大規模言語モデル（生成AI）をWebから収集した350 GBの日本語テキストを用いて開発
- 従来から利用していた高品質な日本語テキストを事前学習することで、約4か月で開発
- 現在1,790億パラメータのモデルも学習中。今後は民間企業、国研、大学等と共同研究等を通して更に発展

NICTによる400億パラメータのモデル開発

# JST 先端科学技術委員会 AI・情報分野委員会 調査・提言 (2023年3月作成)

- ・ オープンな基盤モデルはビジネス・学術研究・国民生活にも今後大きく影響するものの、アカデミアや日本企業単体での研究開発は厳しい。
- ・ 基盤モデルの研究環境は (1)モデル研究、(2)コーパス、(3)大規模計算基盤、(4)それを支える多くのエンジニアが不可欠。
- ・ イノベーションに繋げるには、(5)ルール・倫理面での整理と(6)ビジネス創出・スタートアップ創出・人材育成も必要。

## AI研究開発と計算基盤の整備

LLMの登場までには、基本的な深層学習の発展に加え、ブレークスルーに繋がる諸技術が発明され幾つも組み合わさることで進展。



モデルの研究



モデルの基礎研究、コンパクト化、マルチモーダル化など多く課題が残る



コーパス



- ・ コーパスの質と量がモデルの良し悪しを決める
- ・ 企業のデータは外に出にくい。
- ・ ヒューマンフィードバックなど新たなアノテーションも必要。



大規模計算基盤  
(GPU数万枚規模)



並列計算・プログラミング  
エンジニアリング

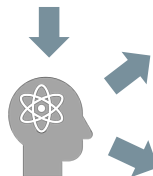


論文化されない知見蓄積・エンジニア育成



ELSI

- ・ 倫理
- ・ ルールメイキング



各種  
基盤  
モデル

## ビジネス創出・スタートアップ支援

改良・カスタマイズ

- ・ プロンプトエンジニアリング
- ・ 事業ドメインごとのファインチューニング
- ・ 日本語精度向上



ビジネス創出

活用

- ・ 文章生成(要約、言い換え、情報整理・応答、創作)、RPA(ロボティック・プロセス・オートメーション)・コンテンツ制作・カスタマーサポート

言語モデルに限らず基盤モデルは他研究分野にも大きく影響



多様な研究

# AI戦略会議

<https://www8.cao.go.jp/cstp/ai/index.html>

## AI 戦略会議 構成員名簿

江間有沙	東京大学未来ビジョン研究センター 准教授
岡田 淳	森・濱田松本法律事務所 弁護士
川原圭博	東京大学大学院工学系研究科 教授
北野宏明	株式会社ソニーリサーチ 代表取締役 CEO
佐渡島庸平	株式会社コルク 代表取締役社長
田中邦裕	さくらインターネット株式会社 代表取締役社長
松尾 豊	東京大学大学院工学系研究科 教授【座長】
山口真一	国際大学グローバル・コミュニケーション・センター 准教授

## (政府側参加者)

岸田文雄	内閣総理大臣 (第 1 回目のみ)
高市早苗	科学技術政策担当大臣
尾崎正直	デジタル大臣政務官 (代理出席)
谷 公一	国務大臣 (サイバーセキュリティ戦略本部副本部長)
松本剛明	総務大臣
築 和生	文部科学副大臣 (代理出席)
西村康稔	経済産業大臣
(中谷真一	経済産業副大臣)
村井英樹	内閣総理大臣補佐官 (司会進行)

# AI戦略会議「AIに関する暫定的な論点整理」(2023/5/26)

## 開発力強化に向けた考え方

- 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
- AI の進化を促す知識基盤研究は、将来に渡った革新的なイノベーションの創出にも貢献する
- 計算資源とデータの整備・拡充を行うことが最も重要
- 生成AI自体の開発は、スピード感を持って行うことが重要 / 民間の活力を十分に活用すべきである。
- 政府による AI 開発の支援先には、透明性・信頼性の確保やその説明など、リスク対応に関する一層の責任を求める

## 計算資源

- 可及的速やかに計算資源の整備・拡充が必要
- 電力調達が大きな課題 / 省エネ半導体等の開発を促し、早期に社会実装すべき

## データ

- 著作権等に留意しつつ、公的機関が保有するデータについて、我が国の民間企業・アカデミア等に対し開発用にアクセス可能となる仕組みを構築すべき
- 日本語を中心とするデータの整備・拡充を国立の研究所が中心となって、進めるべき

## 従来型ではない開発促進策

- 特に生成AIの技術革新のスピードや予見可能性の低さを踏まえると、従来の政府による開発促進策では対応が難しい。市場原理を最大限尊重し、迅速、柔軟かつ集約的にプレイヤーの取組を加速するような支援を政府としても行っていくことが期待
- 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

# オープンな開発環境の構築と高度人材育成の方針

1. フェアな競争を促す環境整備
2. スピードを促す環境整備
3. 優秀人材の呼び込みと循環を促す環境整備

## インフラ整備と利用の分離

- 計算基盤の整備を政府が後押しする。
- 多様なデータ整備を国策として取り組む。
- 計算基盤とデータの割り当てはフェアな競争に基づいて行う。

## ビジネス支援と、共通の知識基盤の並走

- 2-3年先のビジネス開拓だけでなく、5-10年先を見据えた取り組みにも予算措置をすることを勧める。
- 国研や大学が中心となり、オープンソースの日本語に強い言語モデルを作成し、国際的な諸問題を協力して解決することを国がサポートする。

## 優秀人材の抜擢と流動性向上

- 若手研究者のエンパワーメントを進め、優秀人材が最適な場所で自由に研究に従事できる環境を整備する。
- 若手研究者の待遇改善に加えて、兼業や副業を推奨し、大学と企業の両方に所属しながら研究開発を進めることを推奨する。

# 参考資料

松尾豊 [AIの進化と日本の戦略](#)、自民党PT発表資料

岡崎直観 [大規模言語モデルの驚異と脅威](#)、AIPシンポジウム

JST 福島俊一 CRDSフェロー 資料

理研 泰地真弘人 副センター長 提供スライド