

生成AIの進展と保健医療における活用可能性





1993年 香川県立丸亀高校卒。

1997年 東京大学工学部電子情報工学科卒業

2002年 同大学院博士課程修了．博士（工学）。

産業技術総合研究所 研究員

2005年 スタンフォード大学客員研究員

2007年 東京大学大学院工学系研究科 技術経営戦略学専攻 准教授

2014年 東京大学大学院工学系研究科 技術経営戦略学専攻 特任准教授

2019年 東京大学大学院工学系研究科 人工物工学研究センター／技術経営戦略学専攻 教授

2017年～ 日本ディープラーニング協会 理事長

2019年～ ソフトバンクグループ株式会社 社外取締役、人工知能学会理事、情報処理学会理

デジタル市場競争会議 構成員

2021年～ 新しい資本主義実現会議 有識者構成員

2023年～ AI戦略会議 座長

2023年の生成AI周りの動き



生成AIに始まり、生成AIに終わった目まぐるしい1年だった

- 2022年11月30日 OpenAIによるChatGPTの公開
- 2023年2月3日 自民党に「AIの進化と実装に関するプロジェクトチーム」（座長：平将明議員。塩崎彰久議員など）が立ち上がる
- 2月24日 MetaからLlamaリリース
- 3月14日 Open AIからGPT-4リリース
- 3月21日 GoogleからBardリリース
- 4月10日 岸田総理がOpenAI社のサム・アルトマンCEOと面会
- 5月11日 AI戦略会議（座長：松尾豊 東大大学院教授）での議論開始
 - 関係省庁連携のためAI戦略チームも発足（チーム長：村井英樹 総理補佐官）
- 5月26日 AI戦略会議「AIに関する暫定的な論点整理」の公表
- 7月18日 MetaからLlama 2がリリース
- 9月25日 OpenAIからGPT-4Vがリリース
- 10月30日 広島AIプロセスに関するG7首脳声明
- 11月1-2日 UKにてAI Safety Summit
- 11月7日 OpenAIからGPT-4 Turboリリース
- 12月6日 広島AIプロセス G7首脳声明
- 12月7日 GoogleからGeminiリリース
- 12月21日 AI戦略会議にて、AI Safety Instituteの設立が発表。



GPT-4は米国・日本の医師資格で一定の条件を経た場合、合格ラインを超えた

- 米・Massachusetts General HospitalのTiffany Kung氏らは、ChatGPTに、米国の医師資格試験の問題を解かせ、臨床的推論能力を検討(2023年2月)
 - その結果、ChatGPTは特別な訓練や強化学習を行わずとも合格ラインに近い正答率を示し、整合性の高さや洞察も認められたと発表した
- オンライン診療を手がけるMICIN(マイシン)と金沢大学が、GPT-4で2023年度の日本の医師国家試験を解いた(2023年6月)
 - 最初は画像を見ずに回答できる問題文について、チャットGPTに日本語のまま入力したところ、正答率が52.8%にとどまった。学習データの多くが英語圏のネット情報に由来するとみられる
 - そこでChatGPT自身に日本語の問題文を平易な英語に翻訳させた。さらに問題の種類別に入力内容を調整し、GPT-4を使って回答させたところ、正答率は82.8%にまで上昇し、正当な合格ラインを突破した

- 朝日新聞(<https://www.asahi.com/articles/ASR6H3Q13R6GUTFL00H.html>)
- 時事メディカル (<https://medical.jiji.com/news/56105>)

鍵となる2つの技術：①トランスフォーマー②自己教師あり学習

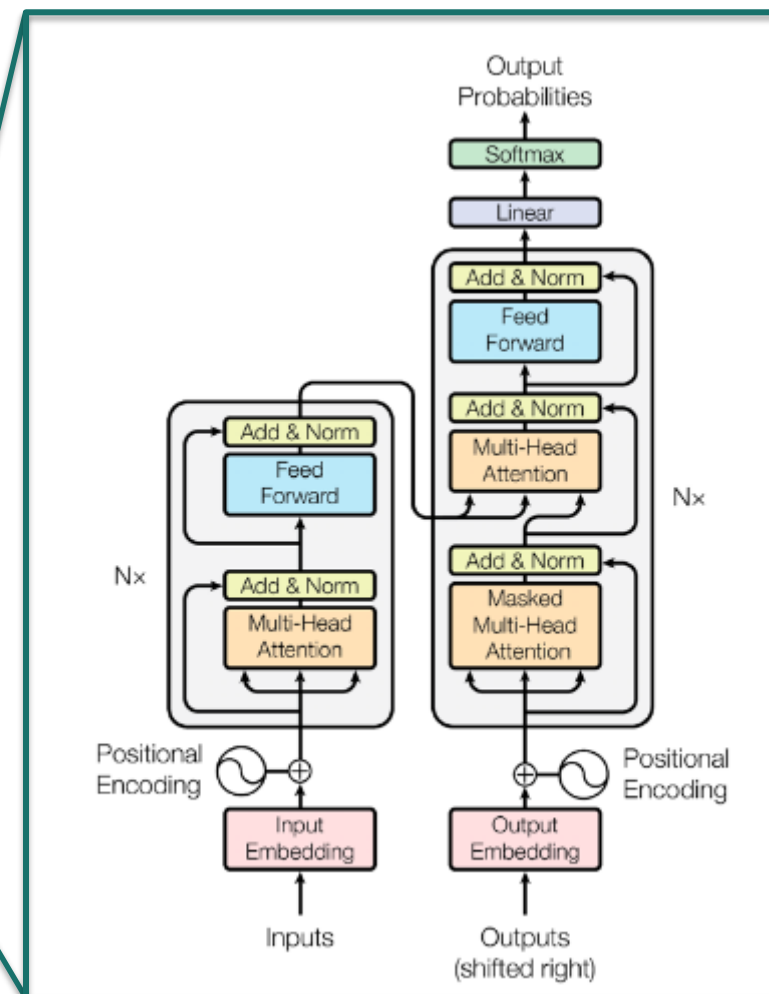
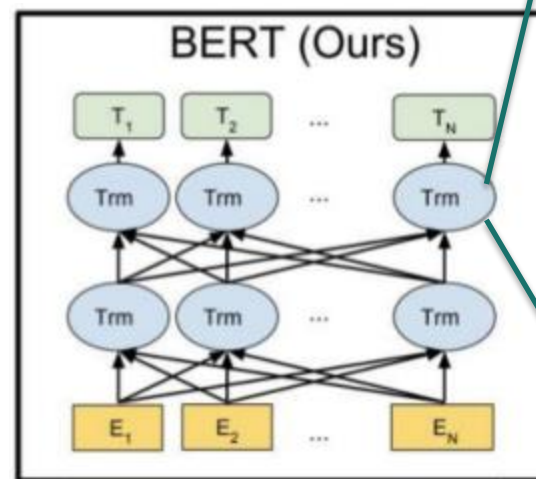
1) トランスフォーマー

(L. Kaiser et al., 2017)

- 有名な「Attention is All You Need」論文。Googleの研究者ら。
- 「アテンション」という機構を大規模に使ったトランスフォーマーを提案。
 - アテンション（自己注意機構）：ニューラルネットワーク中のどこかの情報をどのように使うか自体を学習することができる。
 - トランスフォーマーは、マルチヘッドのセルフアテンションを多層に重ねたもの。
 - 多くのタスクでSOTA（最新性能）を出した。

2) 自己教師あり学習

- 次の単語を予測するという、「**自己教師あり学習**」で事前に学習する。
 - Next word prediction
 - 事前学習をしておくことで、次に与えられるタスク（ダウンストリームタスク）で、少しの例を与えるだけで学習することができる。



Transformerの構造

自然言語処理における自己教師あり学習



与えられた文に対して、前の部分から、次の語を次々と予測していくタイプの自己教師あり学習。
自己教師あり学習による事前学習 ⇒ ダウンストリームタスク（後から与えられるタスク）

例：与えられた文

Generative Pre-trained Transformer 3 (GPT-3) is an autoregressive language model that uses deep learning to produce human-like text.

自己教師ありによる
予測問題

Generative Pre-trained Transformer 3 (GPT-3) is an
autoregressive language



予測

Generative Pre-trained Transformer 3 (GPT-3) is an
autoregressive language model



Generative Pre-trained Transformer 3 (GPT-3) is an
autoregressive language model that

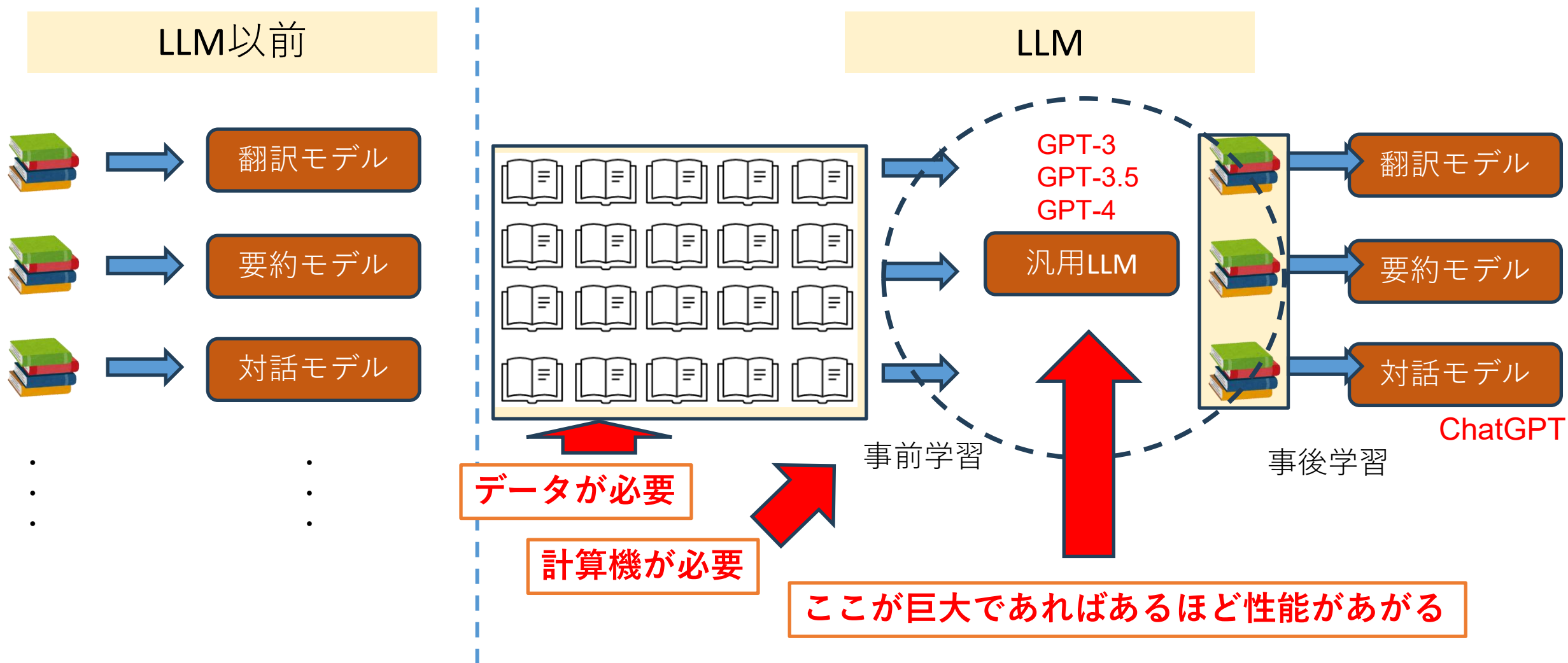


....

大規模言語モデル（LLM: Large Language Model）



- 以前と何が変わったか：汎用性の高さ



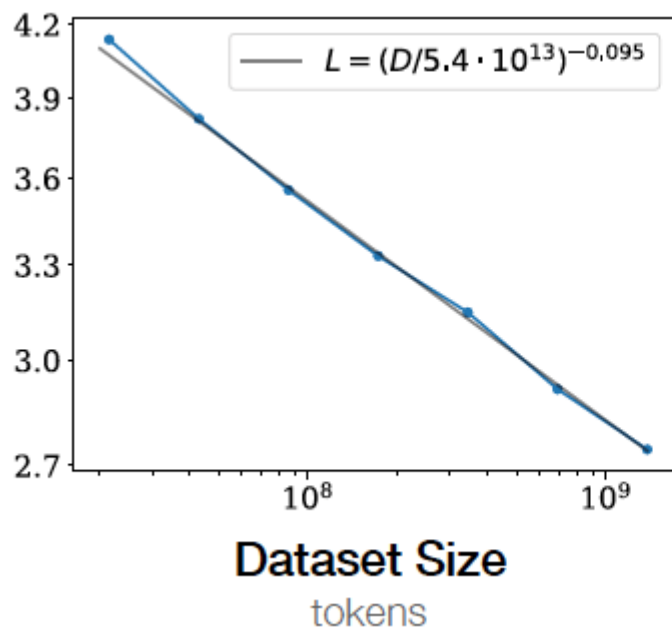
大規模言語モデルのスケール則



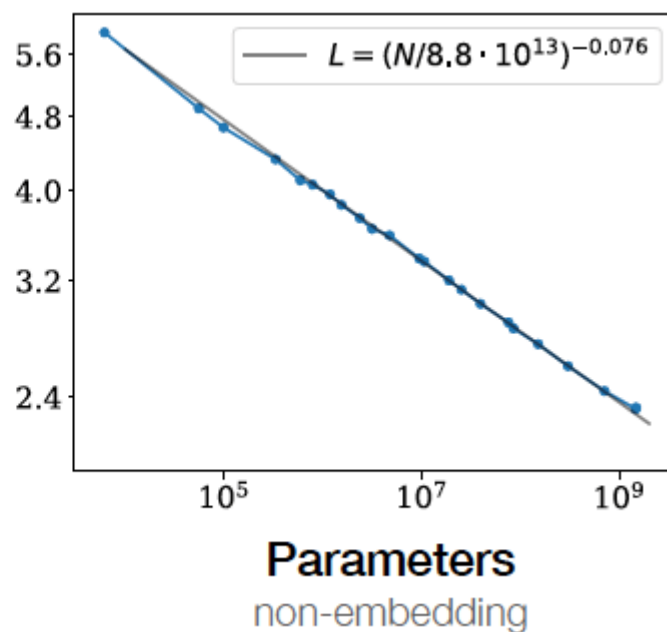
- データを増やし、モデルの大きさ（パラメータ数）を増やし、計算資源を増やせば、精度がどんどん上がる。
- つまり、大きければ大きいほど性能があがる。
- GPT-3は、ウェブの大きなデータ（4000億単語）で事前学習。1750億パラメータ。学習に数億～数十億円のコスト。

テスト誤差

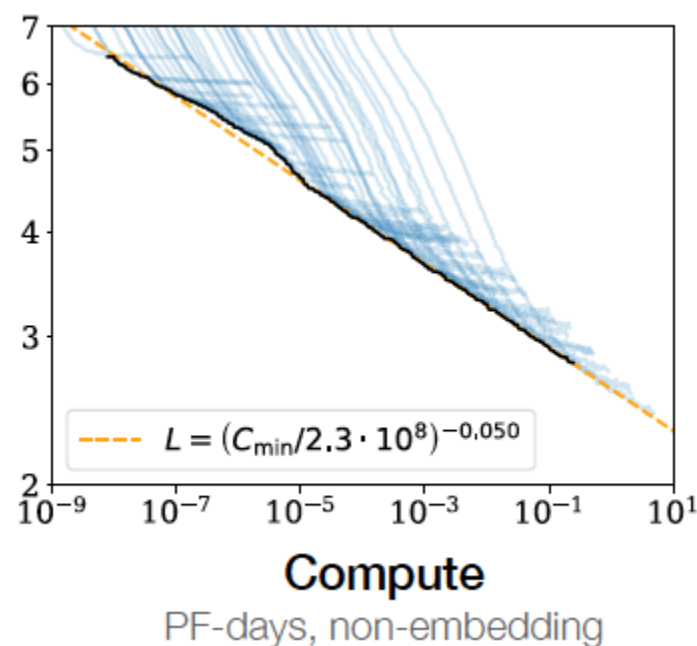
データ量 (トークン数)



モデルの大きさ
(パラメータ数)



計算資源 (GPU数)



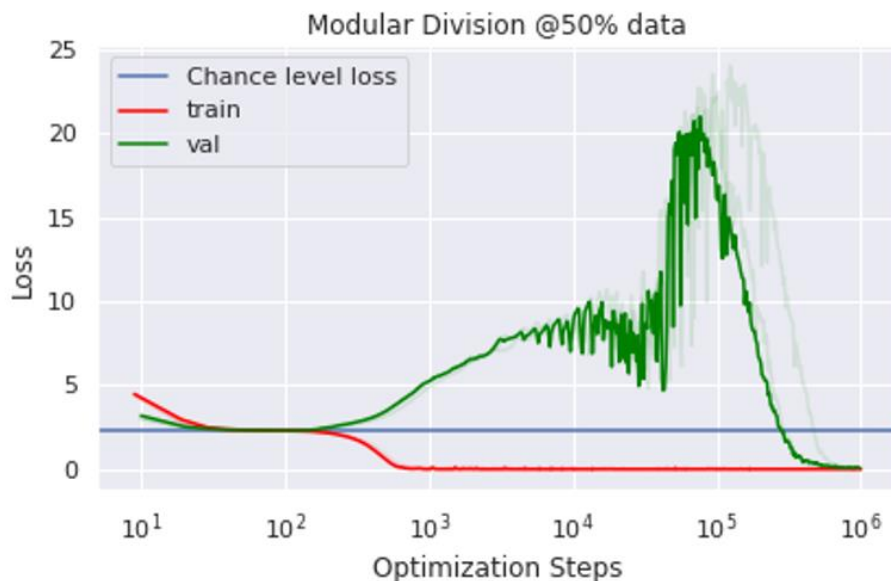
“Scaling Laws for Neural Language Models” [Kaplan+ 2020]

LLMの中で何が起こっているか：Grokkingについて

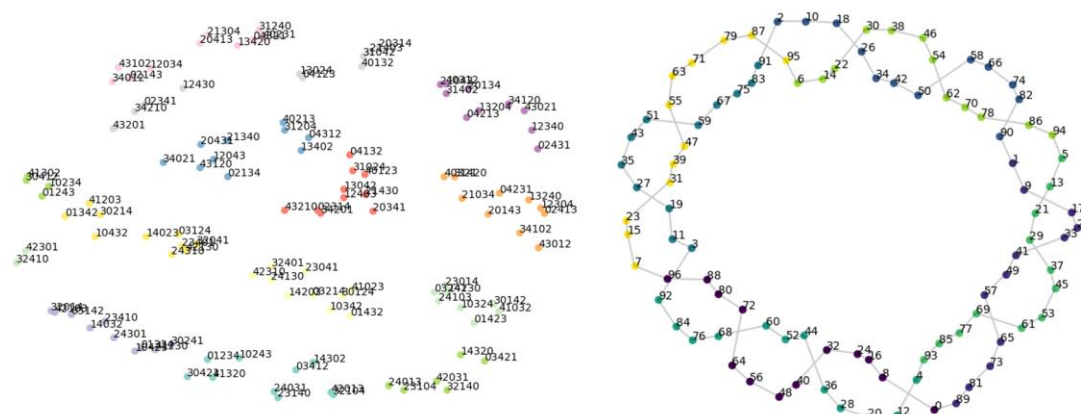
学習をひたすら継続すると、モデルが訓練中に突然、特定のタスクを非常に高い精度で解決できるようになる現象

- 一般的に、ニューラルネットワークはある一定の学習ステップを経ると、以降の検証データで損失が減少しない過学習に陥るとされる。
- しかし、数式の答えを求めるタスクにおいて、過学習するステップ数を遥かに超えて学習を続けると、ニューラルネットワークがデータ中のパターンを理解「Grokking」プロセスを通じて学習し、急激な汎化性能の向上がみられる場合があることを示した

過学習し始めてからその1000倍以上学習を継続することで、検証データの精度が急激に向上する



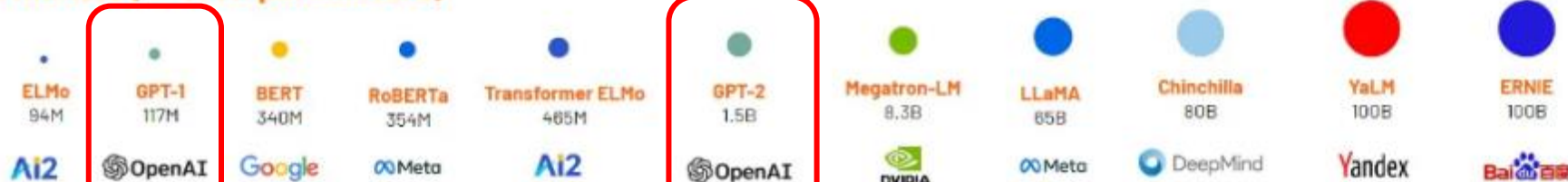
出力層を可視化すると、Grokking後は、最初はバラバラだった埋め込み層が数学的な特徴を捉えている。



引用：[Grokking: Generalization Beyond Overfitting on Small Algorithmic Datasets](#)

大規模言語モデルの開発状況

Small models (<= 100b parameters)

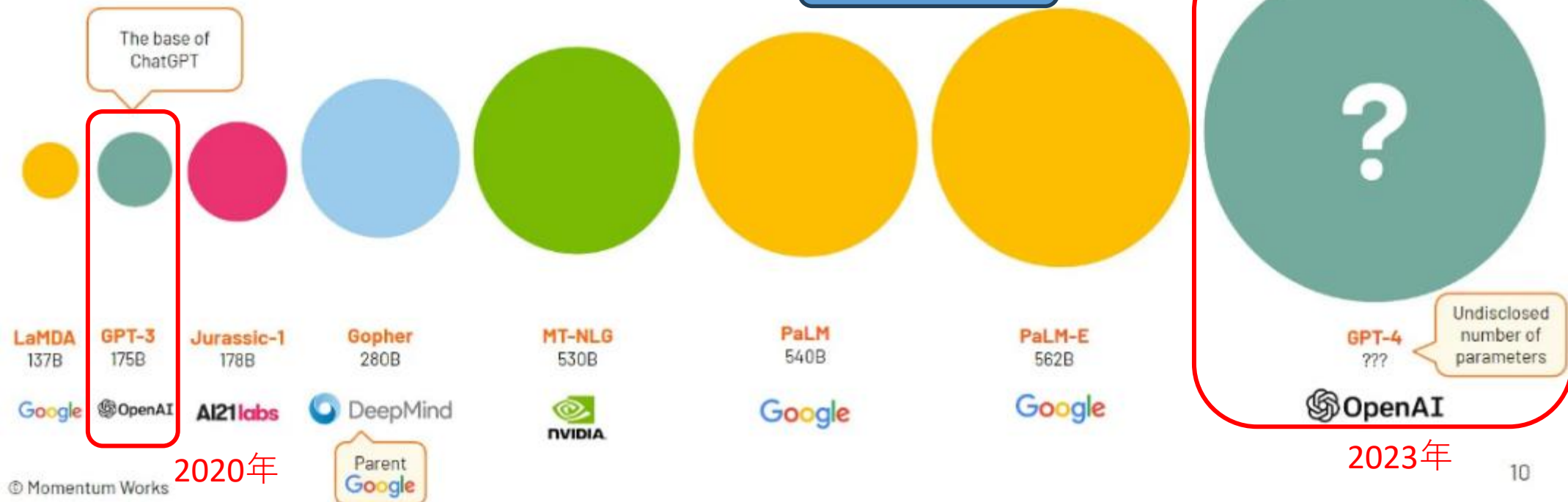


2018年

2019年

日本勢の開発状況は
現在この辺り.

Large models (>100b parameters)



2020年

2023年

10

基本的にはいずれも2018年に発明されたトランスフォーマーと呼ばれる構造を利用。GPT-3登場以降、米国企業を中心に複数の研究機関が独自の大規模言語モデルを開発。

Google Gemini (2023/12)

- 1.56兆パラメータ
- さまざまな指標でGPT-4を超える

TEXT				
Capability	Benchmark Higher is better	Description	Gemini Ultra	GPT-4 API numbers calculated where reported numbers were missing
General	MMLU	Representation of questions in 57 subjects (incl. STEM, humanities, and others)	90.0% CoT@32*	86.4% 5-shot* (reported)
Reasoning	Big-Bench Hard	Diverse set of challenging tasks requiring multi-step reasoning	83.6% 3-shot	83.1% 3-shot (API)
	DROP	Reading comprehension (F1 Score)	82.4 Variable shots	80.9 3-shot (reported)
	HellaSwag	Commonsense reasoning for everyday tasks	87.8% 10-shot*	95.3% 10-shot* (reported)
Math	GSM8K	Basic arithmetic manipulations (incl. Grade School math problems)	94.4% maj1@32	92.0% 5-shot CoT (reported)

最大かつ高性能 AI モデル、Gemini を発表 - AI をすべての人にとってより役立つものに

2023年12月7日 木曜日



Google / Alphabet CEO Sundar Pichaiからのメッセージ

すべてのテクノロジーの変化は、科学的探求を前進させ、人類の進歩を加速し、生活をより良くする機会となります。いま目の当たりにしている AI による変化は、私たちの人生で最も意味深いものになると確信しています。これは、これまでのモバイルやウェブへの移行よりも、はるかに大きなものになるでしょう。AI は、日常から非日常に至るまで、あらゆる場所の人々に機会を生み出す可能性を秘めています。AI は、イ

- No
- 数千億～2兆パラメータくらいまで
 - GPT-4も8つのモデルのMoE
 - ボトルネックはデータになる
- なんでもできるのか
 - No
 - 弱点も見えてきはじめた
- 現状の技術では
 - タスクの分解はあまりうまくいかない。→ 大きなタスクを自動化するのは難しい
 - ドラフト執筆は得意
 - データの変換は得意

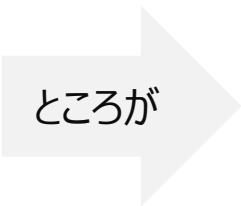
特定領域のLLMと汎用LLMのどちらが市場をとるか

特定領域のLLMが勝つか、それとも汎用LLMが勝つか。医療領域を例に。

汎用LLMより、医療特化のLLMのほうが医療タスクで勝った

- PubMedGPT：医療データを用いて学習したモデル
- GPT-3に勝った

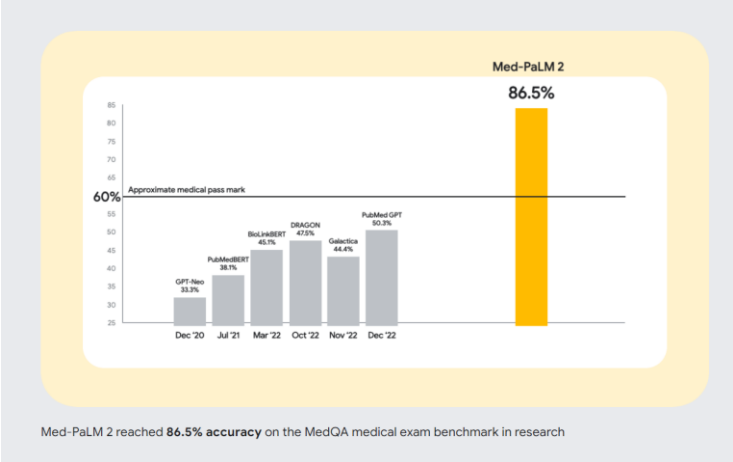
Size	Model	Compute FLOPs			MedNLI	RadQA		CLIP	
		General	BioMed	Clinical	Acc.	EM	F1	Micro	Macro
220M	T5-Base	4.5E+19	6.6E+17	—	0.818	0.479	0.662	0.767	0.594
	Clinical-T5-Base	—	—	5.3E+19	0.855	0.531	0.710	0.793	0.652
345M	RoBERTa	4.6E+21	—	—	0.852	0.521	0.684	0.793	0.677
	BioClinRoBERTa	—	4.2E+21	1.4E+20	0.900	0.604	0.759	0.805	0.707
	GatorTron	1.4E+19	1.9E+20	3.3E+21	0.883	0.583	0.759	0.791	0.690
770M	T5-Large	2.6E+19	2.3E+18	—	0.849	0.537	0.700	0.779	0.629
	Clinical-T5-Large	—	—	1.8E+20	0.872	0.550	0.745	0.800	0.663
2.7B	PubMedGPT	—	4.9E+21	—	0.870	0.512	0.698	0.819	0.666
3B	T5-XL	1E+20	9E+18	—	0.869	0.568	0.729	0.780	0.640
11B	Flan-T5-XXL	3.7E+20	5.5E+18	—	0.808	0.300	0.602	0.164	0.178
175B	GPT-3	?	?	?	0.805	0.362	0.619	0.154	0.146



医療特化LLMより、さらに大きな汎用LLMが医療タスクで勝った

- PubMedGPTは、GPT-3.5やGPT-4に負ける

ところが、さらに医療特化のMed-PaLM2が、勝った。



特定領域のLLMと汎用LLMのどちらが市場をとるかの結論はまだみえていない

ただ、おそらく領域ごとにLLMが成立しそう（パラメータ数ではなくデータがボトルネックになるため）
 → 日本としても、日本語特化、業界特化で勝負になる可能性がある

Med-PaLM2 : 医療特化LLM (2023年3月に発表)

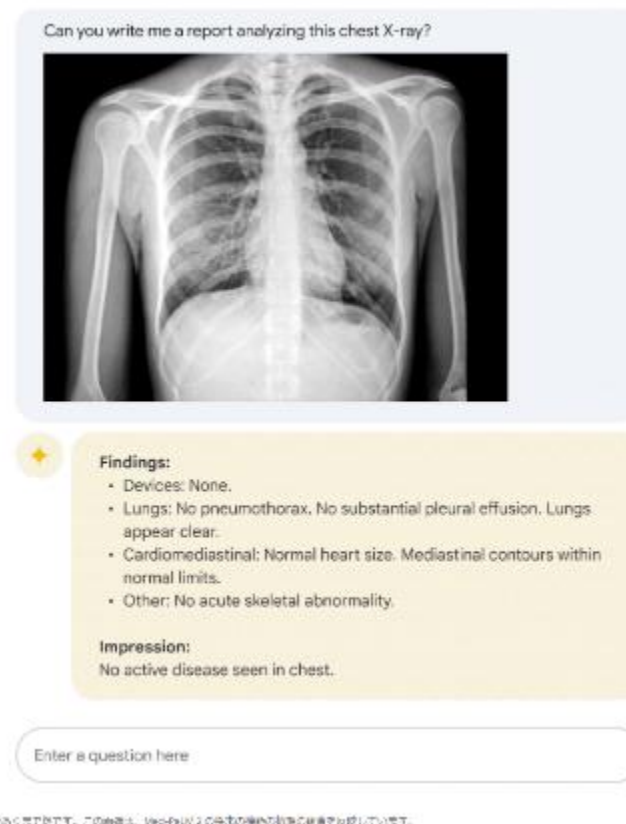


Googleは、医療上の質問に対して高品質な回答を提供するよう設計された大規模言語モデルであるMed-PaLM2を発表

- Googleは2023年3月にGoogle Healthの年次健康イベントで最新モデルMed-PaLM2を発表

- このLLMは医療分野に合わせて調整され、医療検査・医学研究・消費者のクエリを使用して評価されている
 - ✓ 米国医師国家試験(USMLE)形式の問題で構成されるMedQAデータセットに対し、「エキスパート」レベルの成績を収めた初のLLMとなる
 - ✓ 医療文書の要約や患者データの整理といった面倒な業務もこなすことができる
- Med-PaLM2は言語を超えて拡張することを予定している
 - ✓ 胸部X線写真やマンモグラムなどの画像から情報を読み解き、医師がより良い患者ケアを提供できるようサポートする
- Googleは、米総合病院Mayo Clinicで4月から「Med-PaLM2」をテストしている

将来の活用イメージ(X線画の分析支援)

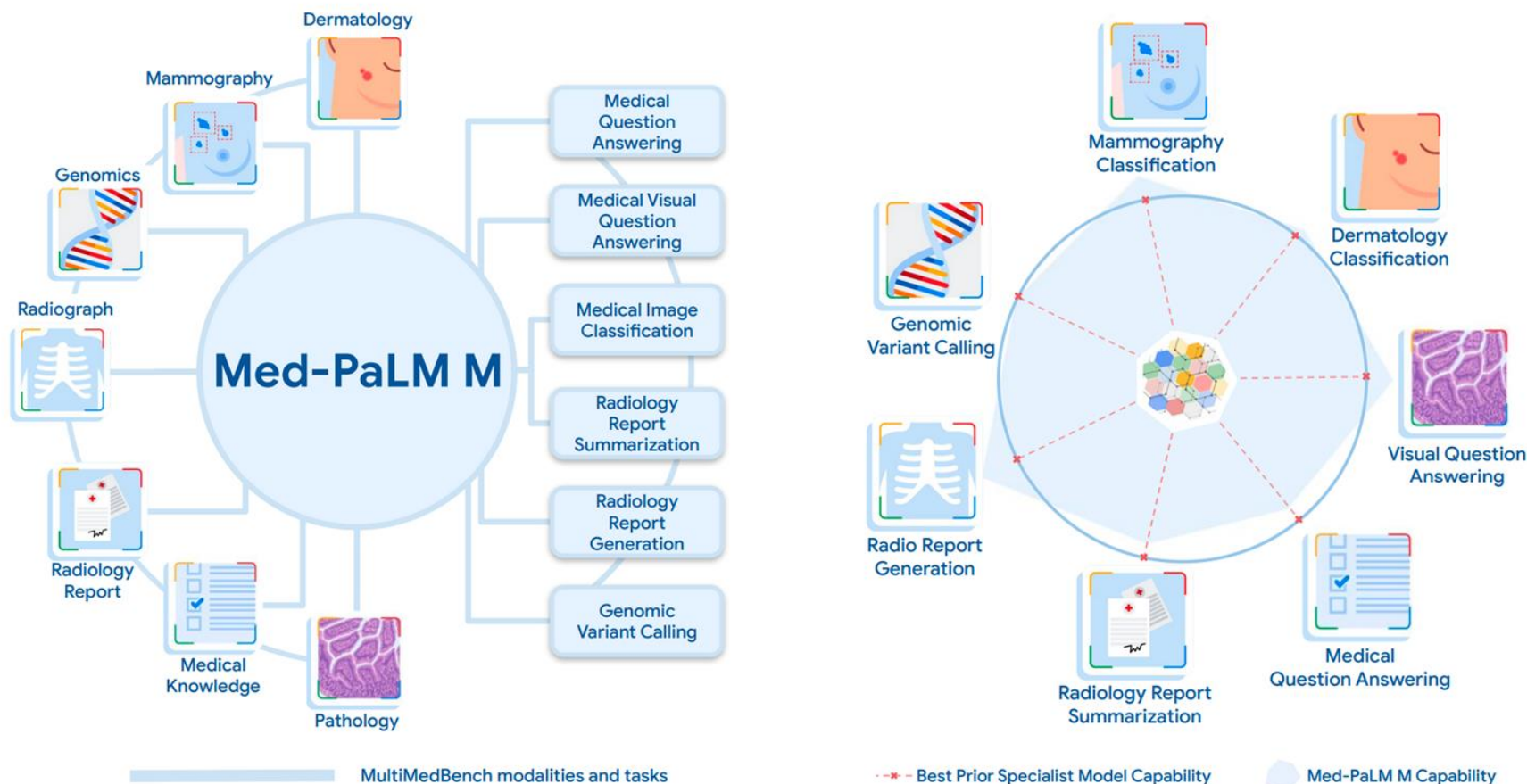


- Google MedPaLM https://sites-research-google.translate.goog/med-palm/?_x_tr_sl=auto&_x_tr_tl=ja&_x_tr_hl=ja&_x_tr_pto=wapp
- CNET Japan <https://news.yahoo.co.jp/articles/473d02bd4c45d8c14bdd45a2d8b3c9d0b0f24be1>

Med-PaLM-Mの概要(2023年7月に発表)



- テキスト, 画像が使える, マルチモーダルな医療タスクを扱える大規模モデル
- 医療の質問と応答, マンモグラフィーや皮膚ガンなどの画像認識, X画像検査のレポートの生成と要約, ゲノム, 病理学画像の扱いなどが含まれているMultiMedBenchというタスクで**最高水準の精度(SoTA)**を達成



- 12B, 84B, 562Bとパラメータ数が大きくなるほど精度は向上
- X線検査の画像と患者情報から、胸部X線のレポートを生成することも可能であり、放射線科医による評価の結果、人間と同等かそれ以上のパフォーマンスを達成した

Indication

-year-old male with intubation. No prior examinations for comparison.

Reference Report

Endotracheal tube terminates 6 cm from the carina. Nasogastric tube loops within the stomach with side port within the gastric body. There are no significant pleural effusions or pneumothorax. Mild retrocardiac atelectasis is present. Heart size is normal. The aorta is tortuous.



Med-PaLM M 12B

An endotracheal tube terminates 5 cm above the carina. There is no pneumothorax or large pleural effusion. Opacities in the **medial right lower lobe** and retrocardiac region may represent atelectasis aspiration or pneumonia. **A calcified lymph node is noted in the aortopulmonary window.** Cardiac size is normal. [No comment about NG/OG tube]

Med-PaLM M 84B

Endotracheal tube terminates 5 cm above the carina. Nasogastric tube courses below the diaphragm and inferiorly beyond the film. There is retrocardiac opacity which may represent atelectasis or aspiration. The heart size is normal. The aorta is tortuous and unfolded. No pleural effusions or pneumothorax.

Med-PaLM M 562B

Endotracheal tube terminates 5 cm above the carina. Nasogastric tube courses below the diaphragm and inferiorly beyond the film. Lung volumes are low **with crowding of bronchovascular markings.** There is a left retrocardiac opacity. There is no pneumothorax or pleural effusions.

Significant error

Insignificant error

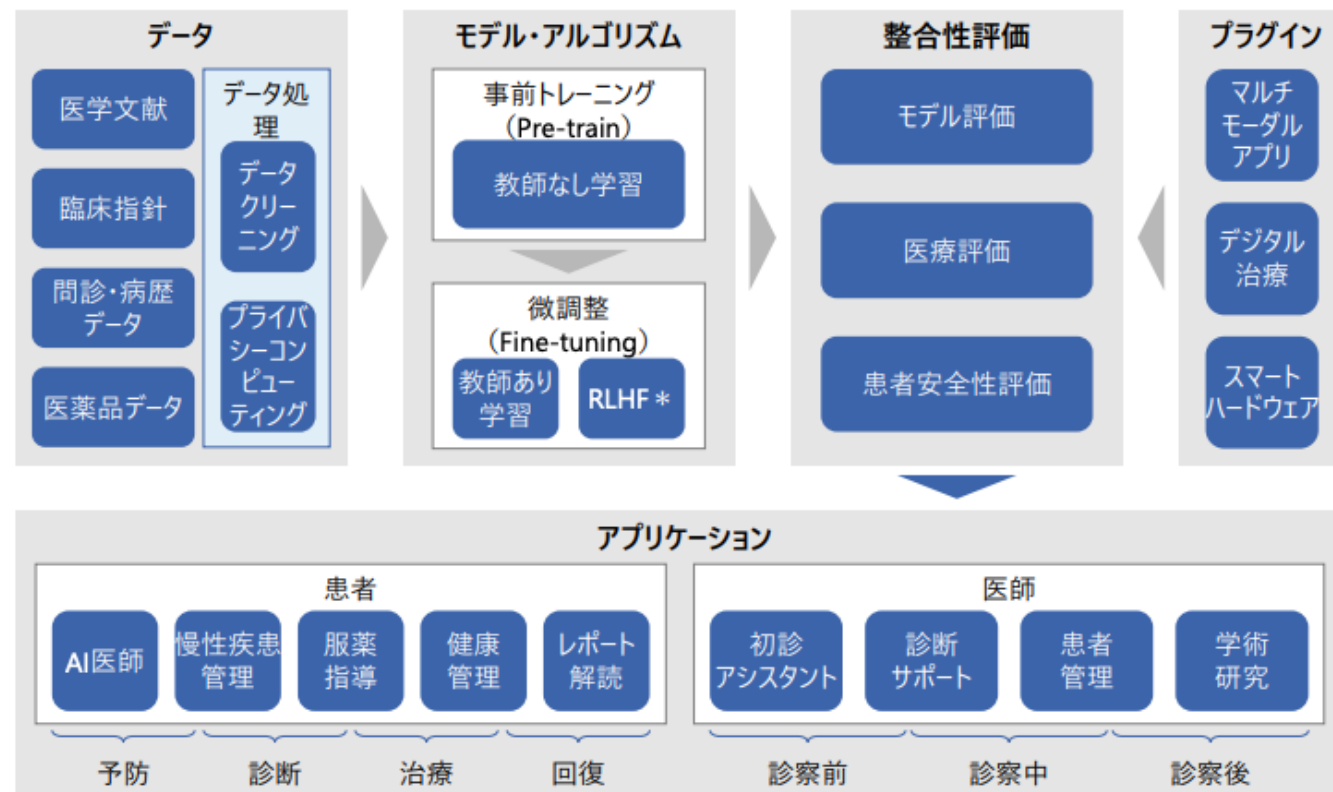
[Omission]

中国生成AI業界の現状-「業界特化型」事例②成都医雲科技

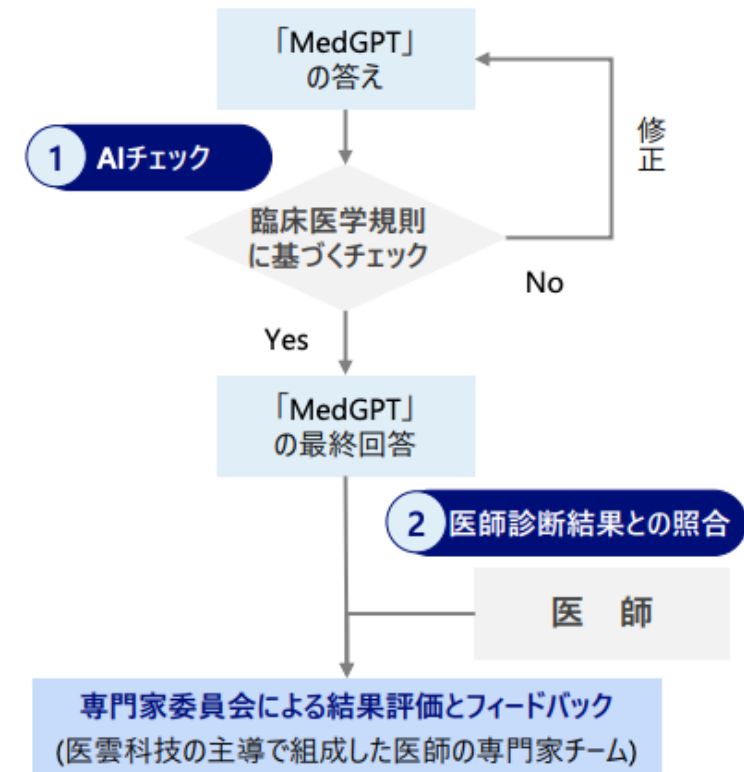
医療業界向けの大規模言語モデルMedGPTは、問診から支援まで全プロセスをAIにてサポートすることを実現

- 成都医雲科技は、1000億以上のパラメータを有する医療業界向けの生成AIモデルMedGPTを開発した。
- 20億件もの医学テキストデータ、800万件もの臨床診療データを訓練に活用し、提携医師による診断結果評価による学習も導入している。

MedGPTの概念図



医師診断結果との照合体制による正確性の確保



* RLHF: Reinforcement Learning from Human Feedback、言語モデルを、人間のフィードバックからの強化学習でファインチューニングする手法である。

出所: 公開情報よりNRI作成

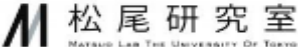
Copyright (C) Nomura Research Institute, Ltd. All rights reserved.

NRI 18

- 現状のLLMと、「堅い」システムとハイブリッドにしたシステム
 - LLMはコントロールがあまり効かない。変なことを言ってしまう。安定しない。
 - そのため、例えば、医療であれば、LLMの出力を、きちんとした医学的知識と照らし合わせ、最低限の品質の担保をする仕組みが重要になる。
 - また、医師のフィードバックを入れ、それを事前学習／事後学習／プロンプト等に反映して、精度を改善していく仕組みも必要になる。
 - 個人情報・著作権に配慮し、データとモデルを管理する必要も。
- こうしたものを備えた、「LLMシステム」を作れるかが、短期的に、競争上、重要

国内でも各企業がLLM自体の開発を計画／実行

現状、100億パラメータ程度（数億円規模程度）の開発が多い。本格的なバーティカルLLMはまだ登場していない。

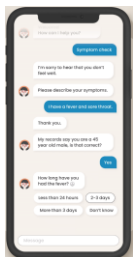
会社名	ステータス	概要
	5/10に和製GPTの開発検討を発表 7/6 GPUへ200億円規模の投資計画を発表	LINEにおいてGPT-3.0～3.5程度の学習ベースあり AIが触れるメンバーを1000人規模でピックアップし、開発メンバーにアサインを検討
	7/4に400億パラメータのモデル開発を発表。 1790億パラメータモデル開発中	Webから収集した350 GBの日本語テキストを用いて開発。従来から利用していた高品質な日本語テキストを事前学習することで、約4か月で開発
	7/6に130億パラメータのモデルの開発を発表	クラウド／オンプレミス環境での運用が可能。性能面でも、世界トップクラスの日本語処理能力を実現。GPU1枚で動作する程度にパラメータ数を抑えた上で、多量のデータと膨大な計算時間をかけ開発
	5/11社内で130億パラメータモデルの開発完了を発表	日本語に特化した独自の大規模モデル。 チャットボットやRPAをはじめとする業界特化型のLLMの構築や、各企業と連携したLLMを活用したビジネス開発の推進を予定
	6/9開発の推進を発表 今年度中に完了を予定 70～300億パラメータ	NTTは使い勝手と運用コスト低減を強みとし、汎用的な生成AIを展開する米テック企業とは一線を画す戦略をとる
	3/16ABEJA LLM Series （130億パラメータ）商用サービスとして展開	2018年より独自で本領域における研究開発推進。GPT-3をベースに学習した最大130億パラメータを持つ日本語に特化したモデルの一部をオープン化
	20年9月Elyza Brainを発表	特定の技術課題ではネイティブな日本語話者の精度を超える。 テキストを扱う様々な業務に適用可能
 富士通、理研、東工大、東北大	5/22に計画発表 24年3月末に成果物を Githubなどで公開予定	「富岳」の政策対応枠において、LLMの分散並列学習手法の研究開発を予定。東工大が全体統括、東北大はデータの収集とモデルの選択、富士通は、理研は分散並列化/通信高速化を予定（富岳は現状GPUベースではない）
	9/28 日英2言語を併せた能力でトップレベルの性能を示す言語モデルPLaMo-13Bを公開	約130億個のパラメータからなる言語モデル Apache License 2.0で公開されたOpen Source Software (OSS)
	8/18に100億パラメータサイズ・日英2ヶ国語対応の大規模言語モデル“WebLab-10B”を公開	公開済みの日本語大規模言語モデルで最高水準 今後WebLab-10Bのさらなる大規模化を進める

LLMを始めとする生成AIで扱えるパラメータが増え、さらに複雑な文脈に対して、人間らしい自然な回答や塩基配列などの文字列を扱うタスクの精度向上が可能になった

前提：生成AI(特にLLM)の医療分野における性能向上

1)汎用LLMの性能向上 2)医療特化LLMの進展

①診断支援系



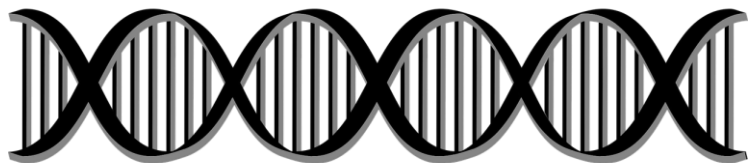
- ・ 医療質問への回答、感情分析
- ・ 診断補助、リスク察知
- ・ 画像などの医療データとリンク
- ・ 診断からクリニック・病院検索まで繋げる

②業務効率化系



- ・ **意思決定支援**
 - 診断補助、疾患/予後予測
- ・ **医師の業務効率化**
 - 患者説明書類
 - 保険請求の適正化など

③創薬の効率化



タンパク質構造や生体分子の特性を予測

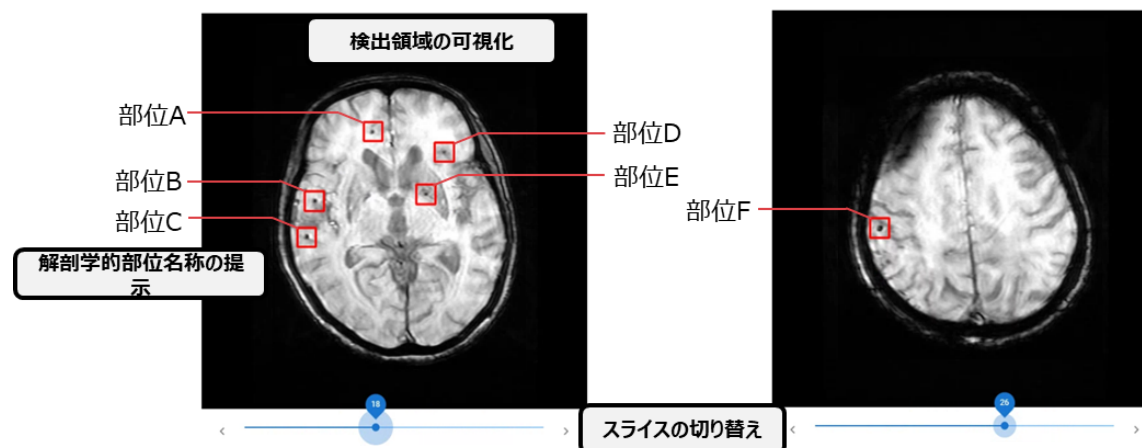
- ・ **製薬企業が大規模言語モデルを活用している事例が増大**
- ・ 抗体医薬品の創薬におけるアミノ酸配列の最適化
- ・ 新たな化合物の生成に関する特性探索
- ・ 例：T5言語モデルを用いたCO-T5によってタンパク質からDNAへの翻訳を学習し、コドン最適化

①診断支援系 – 松尾研の事例

脳のMRI画像を用いて認知症診断支援を実施する

東京都Kセンターとの共同研究：脳MRI画像を用いた認知症診断支援技術の開発

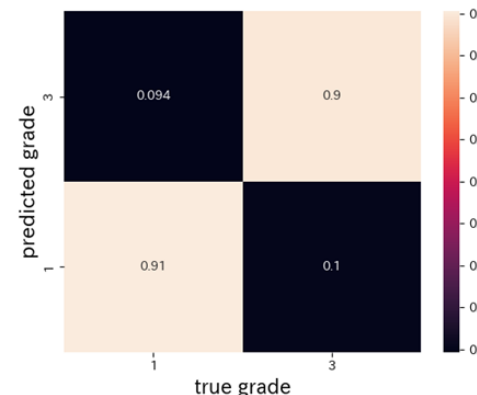
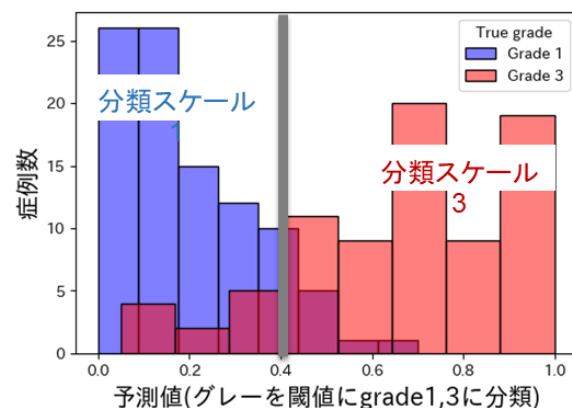
微小出血



- 脳には微小な出血が存在
- 健常者でも数%に存在
- 微小出血の領域・部位をAIで検出
- MRI画像に対してYOLOXを利用
- 左図のように、脳画像のスライスを切り替えて可視化できるシステム開発
- 微小出血の数・部位と軽度認知機能障害に関連性→投薬意思決定の参考に

※ 画面は開発段階のもので、イメージです。

Fazekas分類



- 白質と呼ばれる領域の病変について、機械学習（2D-CNN）で分類
- Fazekas分類 grade1と3の予測
- 投薬条件の参考に
- gradeの1と3の感度：9割程度

②業務効率化系

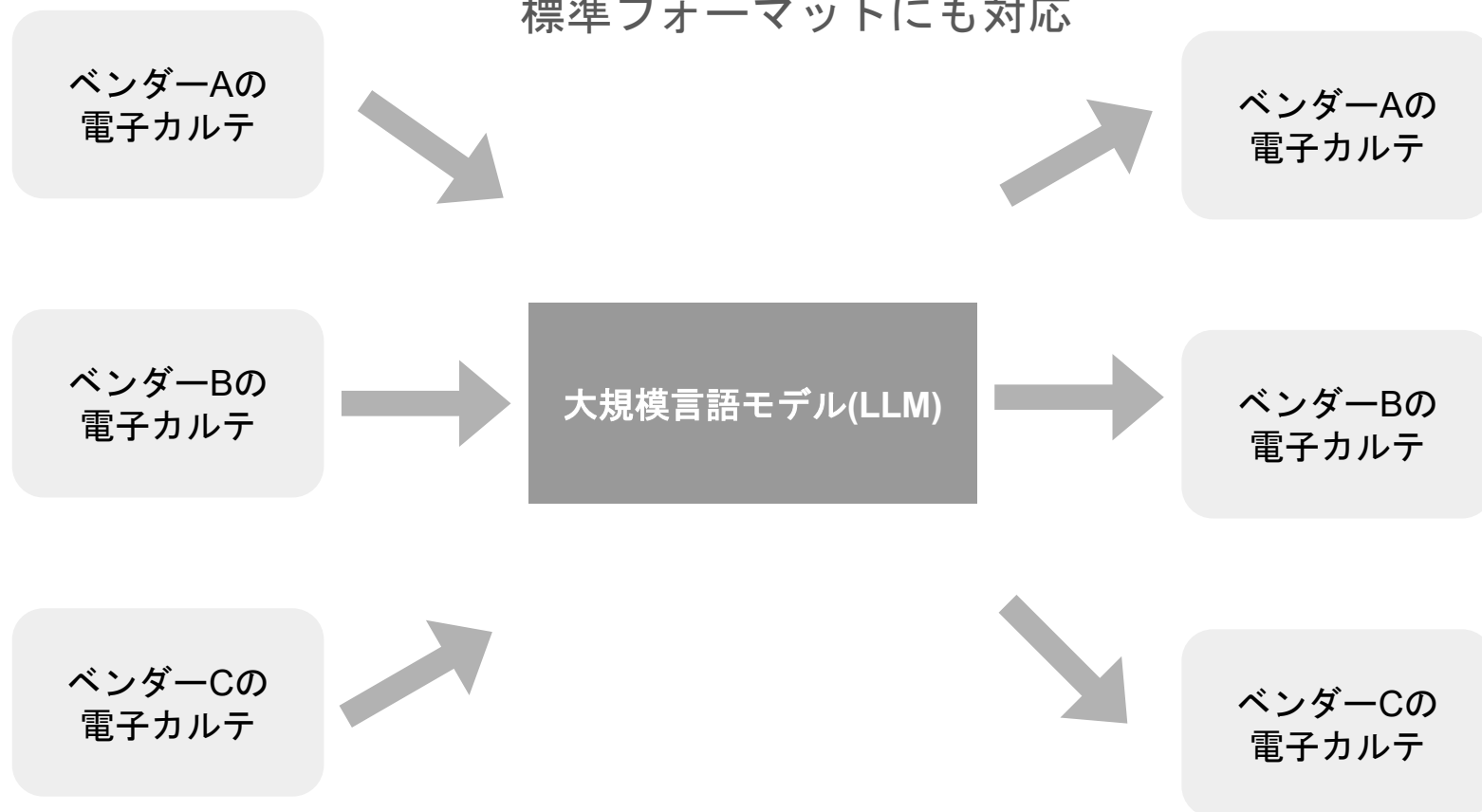


ChatGPT等の大規模言語モデルは「データの変換」が得意

日本語⇄英語の翻訳、プログラミング言語への翻訳、情報を抜き出したりまとめたりできる。

複数のフォーマットの電子カルテの相互変換

標準フォーマットにも対応



③創薬の効率化

遺伝子配列や分子配列をテキストと同様に解釈するLLMは、生物学や化学を含む様々な分野の科学研究や技術革新に活用可能性がある

- LLMは、生物学および化学的データの複雑な配列を解析するために使用され、科学的問題に対する新たな洞察を提供する可能性がある
- 生物学、化学、材料科学などの分野で、テキストに類似した分析能力を持つことから、今後ますます適用されていくことが考えられる
- エヌビディアのプラットフォームは、生物学研究におけるジェネレーティブAIの利用を検証している
- 日本では、理化学研究所が東京大学と共同で、包括的な科学研究のためのLLMを開発している

Clara What's New Solutions Get Started Resources

Join us in person at GTC for AI breakthroughs and more on March 18-21. [Get Early-Bird Pricing Now](#)

Benefits of BioNeMo

State-of-the-Art Biomolecular Models for AI Drug Discovery

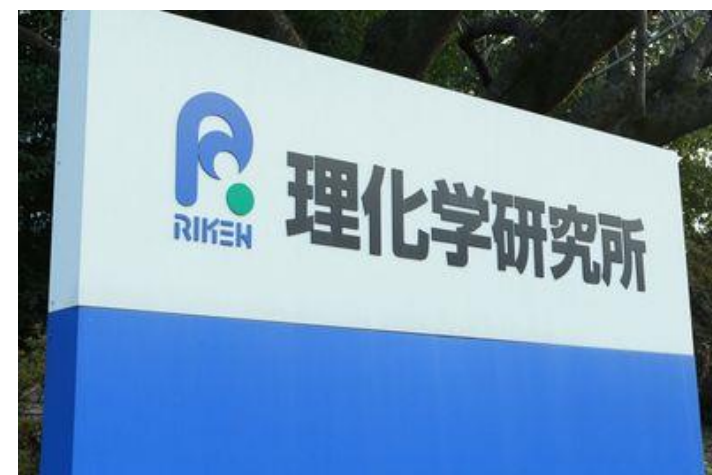
Customize and deploy AI models for 3D protein structure prediction, de novo protein and small molecule generation, property predictions, and molecular docking.

Turnkey Solution Through a Web Interface or Cloud APIs

Access pretrained models through cloud APIs or an easy-to-use web interface for interactive inference, visualization, and experimentation.

Seamless and Scalable AI Services

Experience the ultimate flexibility for experimenting and building enterprise-grade generative AI workflows with GUI and API endpoints on scalable, managed infrastructure.



日本のAIを世界と伍するものにしていくには、**医療分野や金融分野・製造分野などの巨大産業に貢献する形**を作る必要

- 世界的には**数兆円の投資の戦い** ← 後ろ側に数十兆円の売上
 - 2 桁兆円の投資の戦いかも
 - Googleやマイクロソフト
- 日本では、医療、金融、製造（都市含む）に貢献する形をいかに作れるか
- 医療
 - 医師の生産性をあげる。ひとりのお医者さんが診られる患者さんの数を増やす。
 - 医療事務を効率化し、コストを減らす。
 - 活用例）患者さんからの問い合わせの自動応答・問診、治療の診断支援、電子カルテの入力、医療データの統合など
- 金融
 - 人件費がコストの大部分を占め、効率化ができる幅が広い。
 - さらに融資先の事業をデータで把握し、ハイサイクル化（事業を加速）するような投融資をしていく
 - 活用例）顧客対応の効率化、社内の書類作成等の業務の効率化、融資先企業のデータによる把握と管理・与信
- 製造（建設や土木等の都市も含む）
 - 過去の設計データを活用し、製品の競争力をあげる。また、マーケティングと設計をつなぐ。
 - 設計データのLLM活用、マーケティングデータと設計のひもづけ、ロボットとの連携など

- 1. 大前提として：利用・開発と、懸念・リスクへの対応のバランス。【**現行のAI戦略で進めている部分**】
 - 広島AIプロセス
 - 現行法の周知徹底。やって良いことを良いと明示。
 - AIにおけるインフラ整備（計算資源とデータ）
- 2. 活用を進める。地方での活用の活性化。DXの推進として。【**次にやらないといけない部分**】
 - 行政や大企業での取り組みを促進
 - 中小企業での利用の支援。ベストプラクティスの共有。
 - 地方の大学や高専等での人材育成。スタートアップ設立のためのアクセラレーション。
- 3. 国内産業からは、医療、金融、製造【**最終的に作らなければならない形**】
 - 数十兆円規模の産業に貢献する形に
 - 医療（身体）、金融（与信、リアルタイムDX）、製造（人工物・都市）のモデル
 - 関連して、ロボットなどの実世界知能も重要
- 4. グローバルに展開【**さらなる発展**】
 - アジアからグローバルに展開。新たな輸出産業へ。
 - 各産業のグローバル展開を後押ししながら、日本のAI産業自体も発展。

