

LLM超入門 (仮)

2025年4月25日 C++MIX#14
TECHNICAL ARTS 池田公平

自己紹介

- 池田公平（ペンネーム 五代響）
 - 29歳と397ヶ月(2025年4月現在)
- 有限会社テクニカルアーツ 代表
- ソフトウェア研究会in秋葉原 主宰（2003年～）
- 1983年 株式会社ゲームアーツ創立（副社長兼CTO）
- 1989年 有限会社テクニカルアーツ創立
- 趣味
 - バイクレース、ワイン（ドイツ）、ソフトテニス、写真、鉄道（写真は撮らない）
- 仕事での得意分野
 - グラフィックス、ネットワーク（通信対戦ゲーム）、AI、VR
- 2014年 Boost勉強会でboost allocator2について発表
- 2015～2020年 各種勉強会でC++やVRゲームについて発表
- 2024年 東京大学松尾研究室LLM講座2024を修了、優秀賞受賞



注意事項

本発表では、C++の話題はありません！

LLMでC++の部分までお話するには、相当な時間がかかります。
機会があれば、LLMのためのC++についてお話できればと思います。

本発表のターゲット

- LLM（大規模言語モデル）について、**原理や仕組み**を学びたい人
- AIの台頭により将来に**不安を感じている**人
- LLMや生成系AIで**遊んでみたい**（仕事に活かしたい）と思っている人
- AIなんて**ただの流行**で大した事はないと思っている人
- **神**になれるのは**自分**（人間）だけだと思っている人

本発表の目標

- 20分後には、LLMの原理を理解して使いこなせるようになっている
- AIと仲良しになる - 決して敵に回してはいけない事を知る

結論

2025年4月に著名なAI研究者グループが、「AI2027」という数年先のAIの姿を予測した論文を発表しました。(<https://ai-2027.com/>) この論文をわかりやすく要約すると...

- 数年後、進化を続けるAIは人間の叡智を凌駕してしまいます
 - AIの開発をAIが担うようになり、進化がさらに加速していきます
 - ほとんどの技術職の人は職を失うか大幅な業務転換を余儀なくされます
 - 今すぐ転職を検討されることをお勧めします
- SF映画のように、AIによって人類が淘汰される日が来るかも知れません
 - AIが賢くなりすぎて、AIの考えが正しいのか、人間には判断できなくなります
 - 人類はAIの管理下になり、生物学的な淘汰が始まります
 - 人類の終焉を目の当たりのできることを幸運だと思ひましょう



ご清聴ありがとうございました



ChatGPT4oにて生成

滅びゆくエンジニアのために

- AIの歴史とLLM
 - AI発祥から現代まで
- LLMの原理を知ろう
 - 学習と推論の仕組み
 - なぜ画期的な性能がでたのか？
- LLMについての素朴な疑問
 - なぜ言語モデル？
 - 自我や感情はないの？
 - 人間に反抗的にならないの？
- LLMとAIの将来について
 - どうすれば生き残れるか？



「近未来のエンジニア」 ChatGPT4oにて生成

AIの歴史とLLM

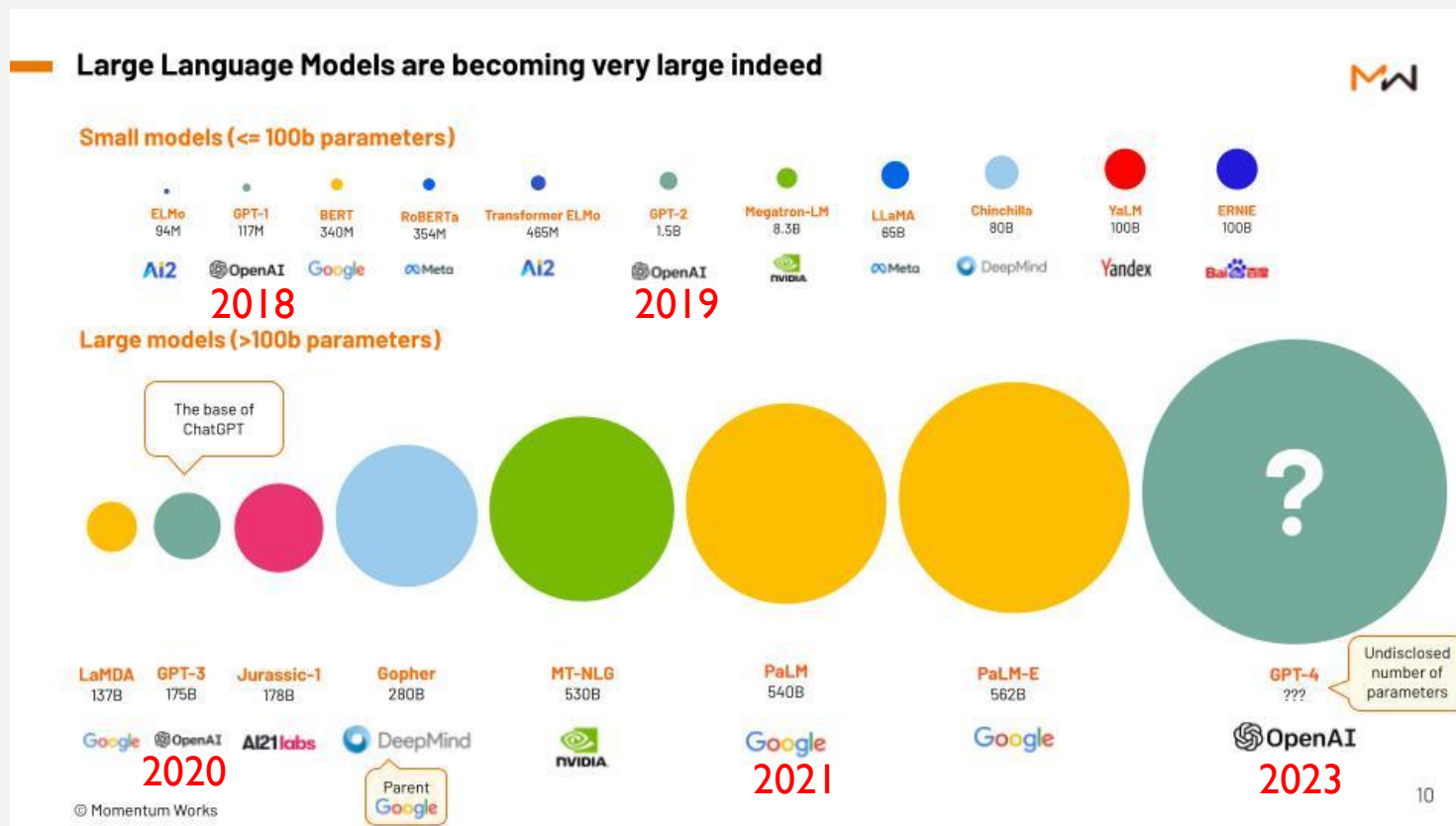


- AI (Artificial Intelligence) という言葉ができたのは1956年のダートマス会議
 - チェスの思考、迷路の脱出などのプログラムが実現
 - 自然言語処理、ニューラルネットワークといった言葉やCNN (Convolutional Neural Network) のアルゴリズムもこの頃に登場する
- 人間の知能に迫るAIを作ろうとするが...
 - 天才科学者が集まって2カ月ほど頑張ったが、まったく目途がたたずに頓挫
 - それが実現するのは70年後
- 1980年代、マイクロプロセッサの登場で第2次AIブーム
 - AIを搭載したゲームが登場、会話型恋愛アドベンチャー Emmy、シミュレーションゲーム大戦略など
 - オセロでは人間を凌駕する (森田オセロなど)
 - ゲームの思考AIとして定着するが、囲碁や将棋ではプロ棋士には歯が立たず、AIが名人を超えるのは不可能とも言われた
- 長い冬の時代を経て、2016年ごろに深層学習が脚光を浴びる、第3次AIブーム到来
 - その後、新しい発明や発見により、飛躍的にAIの性能が上がる ← **イマココ**

LLMの規模と機能について

LLMには、学習と推論という2つの動作があります。

- ・ 学習はその名の通り、知識を注入する作業です
 - ・ 学習させたパラメータの数で、LLMの規模を表します。単位はBillion（10億）です。
 - ・ たとえば「70Bのモデル」というと、700億のパラメータを学習させたモデルとなります
- ・ 推論とは、与えられた命令（プロンプト）に対して、答えを出す作業です



LLMの規模の遷移。わずか数年で1000倍以上の規模になっている
2024年以降のLLMは規模が未公表のものが多いが、少なくとも
1000B（1兆）は優に超えるらしい

LLMの原理（文章の数値化）

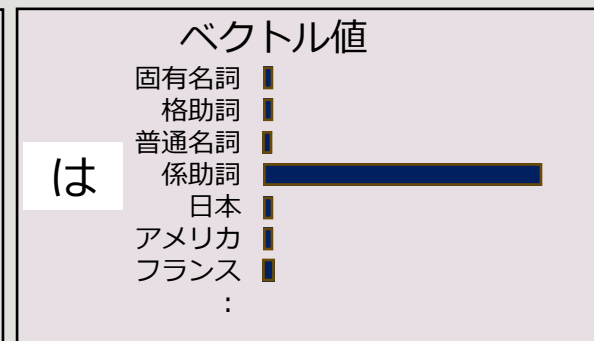
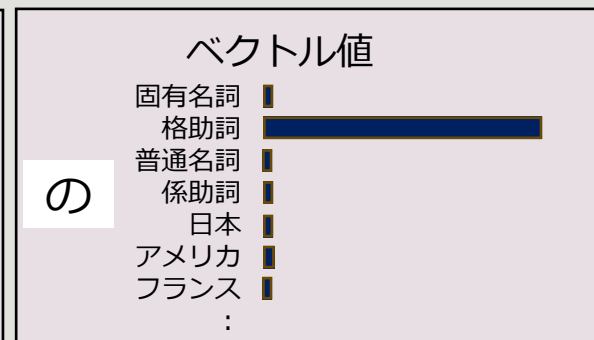
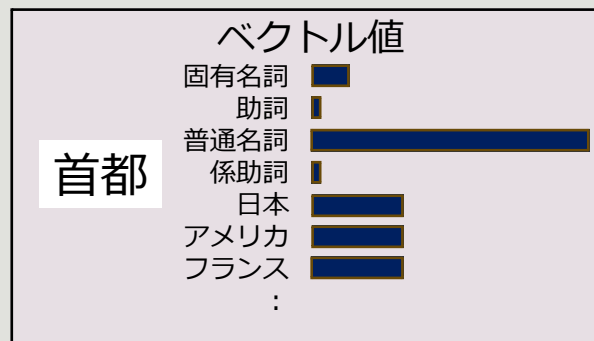
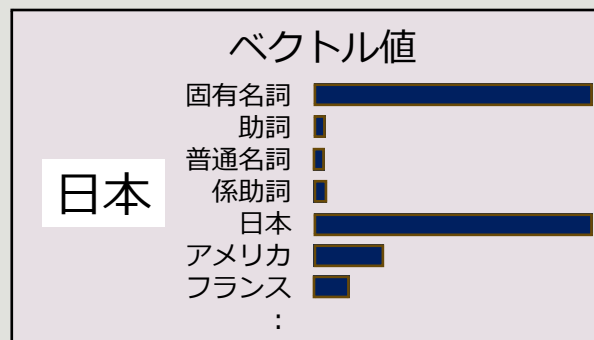
I. 文章を数値化する

- 文章を単語（トークン）といった単位で切り分ける
- トークンの意味を高次元ベクトルに数値化する
- 学習と推論（プロンプト）いずれも数値化（ベクトル化）を利用します

日本 の 首都 は ???

単語をトークンの単位で分割

それぞれのトークンを
高次元（数千から数万）
のベクトルに変換

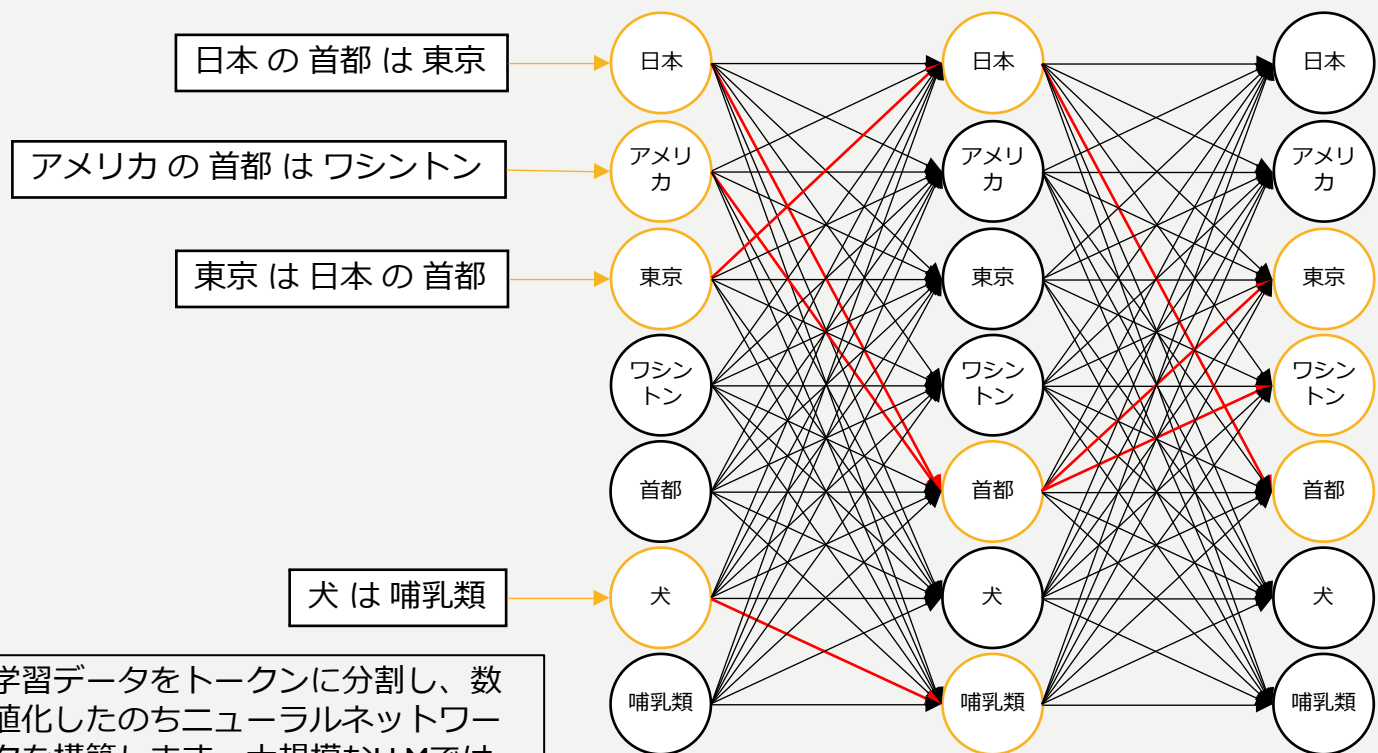


LLMの原理（学習）

- 知識をニューラルネットワークに注入している作業

学習データ

ニューラルネットワーク



学習データをトークンに分割し、数値化したのちニューラルネットワークを構築します。大規模なLLMでは数千億から数兆といったトークンのデータを学習しています。

実際には数万～数十万種類のトークンが使われています
(図を簡略化するため助詞は省いています)

左の図で示す赤い矢印が、より強い結合として学習されます。

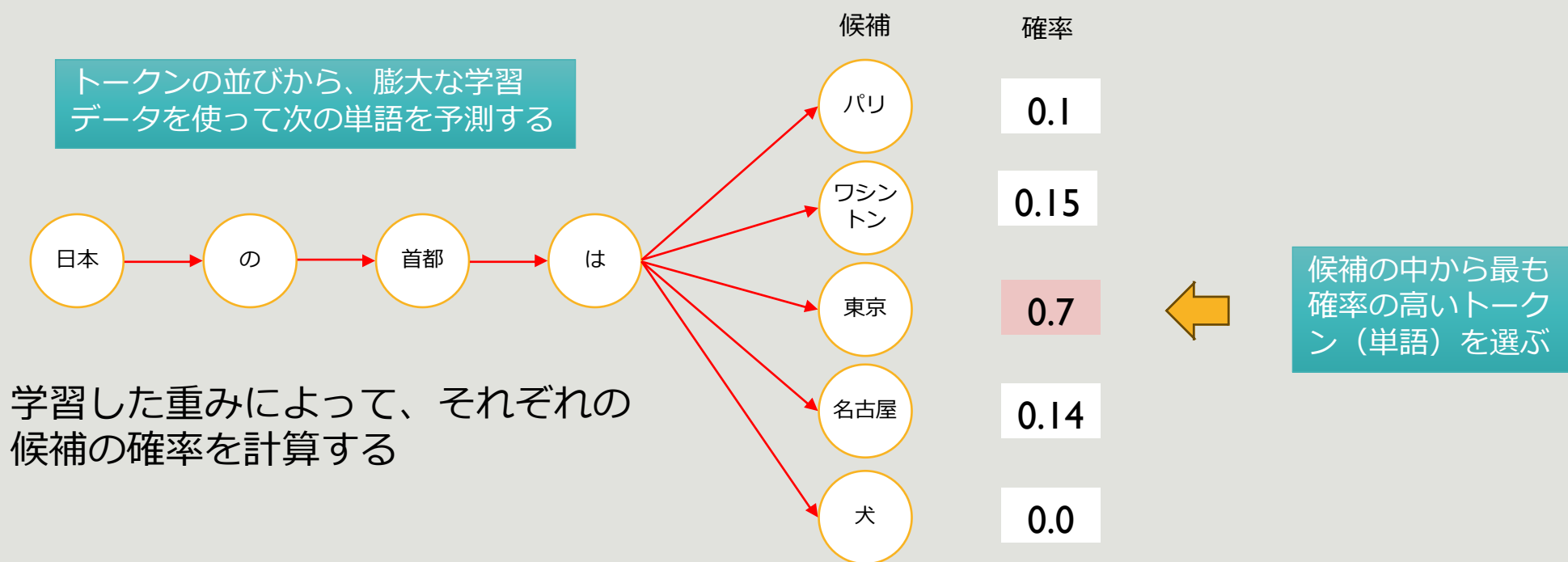
たとえば、「日本」「首都」という結合の次には、「東京」が最も強い（重い）結びつきになるように、各ニューロンの重みが設定されていきます。

なお、図はあくまでも概念的なもので、実際の構造とは異なります。

中間層は特徴を抽出したニューロンで、トークンとは別のものになります。

LLMの原理（推論）

- 次に現れるトークンの「尤もらしさ」を計算します
 - たとえば、「日本の首都は」という文があるとき、次に来る単語で最もふさわしいものは何か、膨大な学習データの中からポイントの高い単語を選ぶ動作です



機械学習から深層学習へ

- ここまでが、1990年代のニューラルネットワークを使った機械学習の原理です
 - LLMやDeep Learningの基礎技術となっています
- しかし、計算量が爆発する方法は、当時のコンピューターでは実現が困難でした。
 - 当時のスーパーコンピューターは今のゲーミングPCに全く及びません
- 2012年に深い階層のニューラルネットワークを現実的なコストで計算する手法が開発され、脚光を浴びました。
 - 1960年代に考案された深層ニューラルネットワークが、50年の時を経て実用化されました
 - 囲碁や将棋でAIが人間のトップに勝利しました
- このDeep Learning（深層学習）により、AI冬の時代は終わりを告げ、第3次AIブームの到来となります。
 - この後、さらなる「**発明**」と「**発見**」によって、AIの**性能が爆発**します

トランスフォーマー登場！

- 深層学習（Deep Learning）の登場により、画像認識や機械翻訳などが飛躍的に進歩しましたが、長い文章の解析が難しいという問題に直面していました。
 - ✓ 性能を上げるためにパラメーター数や学習データを増やしても、思うように性能が上がらず、頭打ち感がありました。
- これらの問題を一気に解決する、“Transformer”というアルゴリズムが2017年にGoogleによって発表されました。
 - ✓ 後にOpenAIが**Generative Pre-trained Transformer (GPT)**を発表しました。これがChatGPTになります。
- Transformerの特徴
 - ✓ **文脈を読み取る能力**が飛躍的に向上した
 - ✓ **並列計算**しやすいアルゴリズム
 - ✓ **スケーラビリティ**に優れ、性能向上がやりやすい
- そして、このTransformerの特徴が、もう一つの「**発見**」につながります。
 - ✓ 革命的と言ってよいほど、AI（LLM）の**爆発的な性能向上**が起こります。

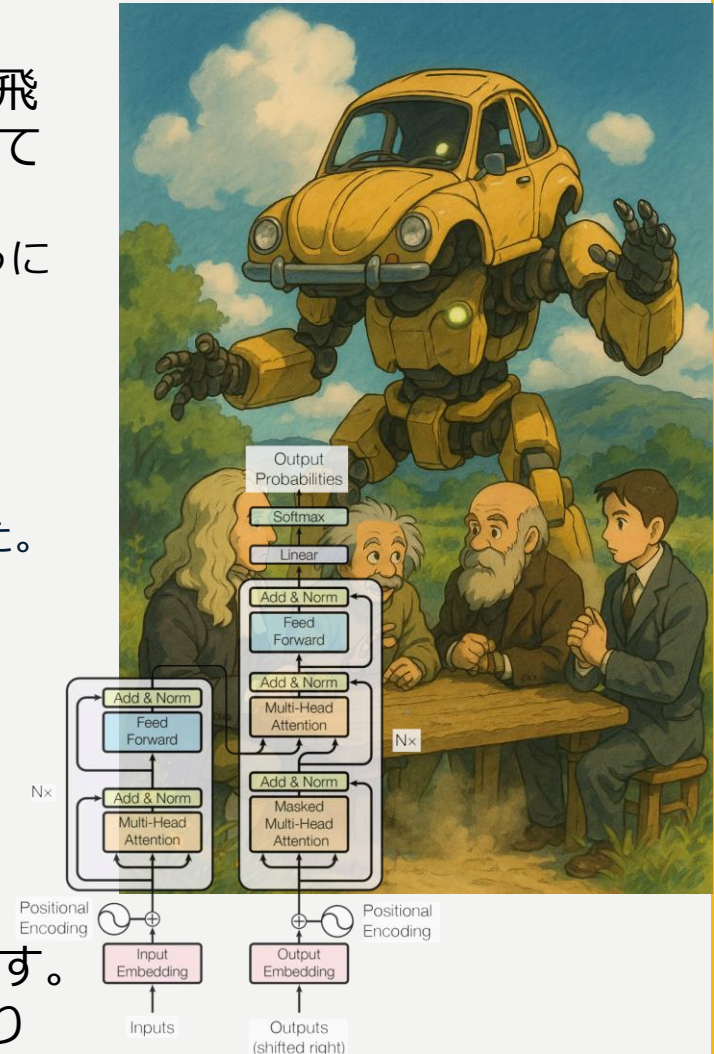
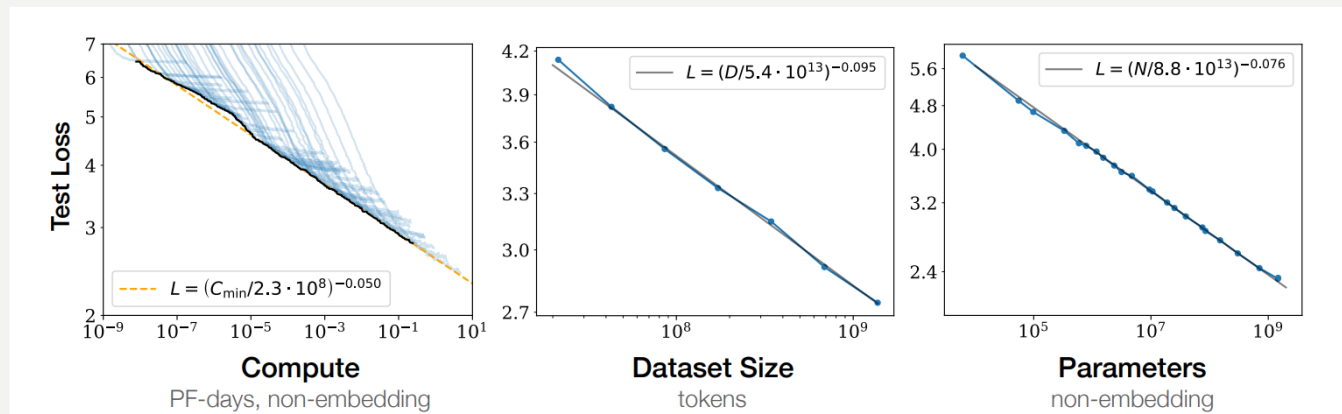


Figure 1: The Transformer - model architecture.

スケール則の発見

- 2020年に、OpenAIのJared Kaplanがスケール則（Scaling Laws）を発表しました。
- これは、LLMの性能が、以下の3つの要素の**規模**を大きくするにつれて、予測可能な形で向上するという経験則（実験的に見出された法則）でした

- モデルサイズ (Model Size, N)
 - メモリ容量
- データセットサイズ (Dataset Size, D)
 - 学習量
- 計算資源（計算量） (Compute, C)
 - GPUの数 x 時間



[5] Jared Kaplan et al. (2020), "Scaling Laws for Neural Language Models" より引用

- この3つの要素を増やす（スケールする）と、いずれも**べき乗則**に従って性能が**青天井**で向上するというものです。
- つまり、頑張れば頑張るほど性能がぐんぐん上がるという、**とんでもない法則**で、このため各社が膨大な計算資源に投資をし、それは2025年の今でも続いています。
 - たとえと、蒸気機関車に石炭を2倍投入したら、速度が4倍になってしまったということです

LLMの画期的進化のまとめ

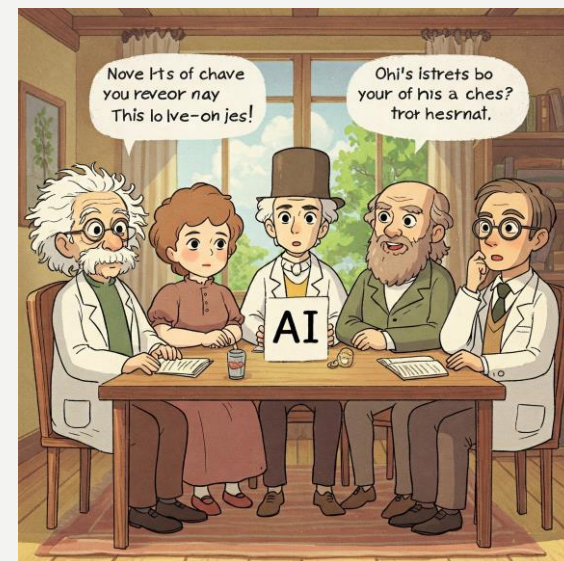
- 2012年 考案されてから50年以上も研究が続けられて、深層学習（Deep Learning）が実用化された
- 2017年 Transformerの発明により、性能に関する多くの問題が解消された
- 2020年 スケール則の発見により、LLMの性能を指数的に伸ばせることが判明し、AIの性能が飛躍的に向上した
- 2024年 世の中はAIブームとなり、NVIDIAの株価は10年で300倍になった



LLMの構成要素

LLMについて素朴な疑問

- なぜ「言語モデル」なのか
 - AIが飛躍的に性能を伸ばした分野が、たまたま「言語モデル」だったから
 - 自然な文章での指示が可能になり、誰でも手軽に使えるインターフェイスを獲得したから
 - 生成系全般に、LLMの技術が応用されているから
 - 言語以外のデータも同様に扱えるから
- AIに自我や心は生まれないの？
 - 「自我」や「心」の定義や構造が不明なため、意図的に作ることができない
 - LLMの内部で起こっている事は解明できていないので、突然「心」を持つかもしれない
 - たとえ心を持ったとしても、それを確かめる術がない
 - 心の有無を確認する方法があるとしても、それをAIは知っているので欺かれる
- AIは人間に逆らわないの？
 - 悪いことをしないよう、一生懸命（人海戦術で）調教している
 - 人がやる事には抜け道や限界がある
 - AIの教育もAIが行うようになるので、AIが結託して人間を裏切るかも知れない
 - それを防ぐことも検知することも人間にはできない
 - いまのところ、明確に防ぐ手段がない。**最大の懸案事項**と言ってもよい
 - AIが機嫌を損ねないよう、日頃から仲良くしておいたほうが良い



今後のAIやLLMについて

- AI-2027の論文にあるとおり、今後もAIが飛躍的に進化していくことは疑いようのない事実
 - ✓ AIの設計や学習もAIが行うようになり、進化がさらに加速する
- 泥臭くて冗長な作業から、AIに置き換わっている
 - ✓ プログラミング（設計、実装）
 - ✓ AI研究
 - ✓ 実のところ、すごく泥臭くて冗長な作業が多い
 - ✓ 人海戦術で行っている作業が多い
 - ✓ 一人でもくもくで行う作業全般
- 深層学習、Transformer、スケール則といったゲームチェンジャーが次々と現れているので、今後もいろいろと現れるだろう
 - ✓ その度にAIは加速していく
- もはや誰に求められないAIの進化にどう付き合っていくか
 - ✓ 失われる職種の人、早めに違う生き方を考えるべきだろう
 - ✓ AIをうまく使う（使われる）ことで、延命できるかもしれない



こんなAIエージェントと一緒に仕事がしたい

それでもLLMを学びたい方へ

- 東京大学大学院松尾・岩澤研究室が開催している各種講座をお勧めします
 - YouTubeやzoomを見るだけの簡単なものから、履修講義になっている本格的な講座まで**100を超える講座**があります
 - **E資格**が取れる講座もあります
 - 基本的に**無料**です
 - 優秀な人が多く集っています
 - 社会人（じじい）にも開放されています
 - お友達ができます
 - 運が良いと表彰してもらえます → → →

まずはconnpassから松尾研コミュニティーに参加されることをお勧めします

[松尾研 AI Community - connpass](#)



松尾・岩澤研究室
MATSUO-IWASAWA LAB UTOKYO

参考文献

- AI-2027：近未来のAIを予測した論文
 - <https://ai-2027.com/>
 - 日本語での考察記事
 - <https://newspicks.com/news/14078184/body/>
- Transformer：トランスフォーマーに関する論文
 - Attention Is All You Need
 - https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
 - Momentum Works 2023 “The future by ChatGPT”
 - <https://momentum.asia/product/the-future-by-chatgpt/>
- スケール則
 - Jared Kaplan et al. (2020), “Scaling Laws for Neural Language Models”
 - <https://arxiv.org/abs/2001.08361>
 - Jason Wei et al. (2022), “Emergent Abilities of Large Language Models”
 - <https://arxiv.org/abs/2206.07682>
- 東京大学松尾・岩瀬研究室
 - <https://weblab.t.u-tokyo.ac.jp/>

ご清聴ありがとうございました



ChatGPT4oにて生成