

Attention Is All You Need

生成系AIの深層学習での革新

Generative Pre-trained Transformer

本日のテーマ：GPT-4の仕組みの概要を理解する

自然言語処理NLPが対話に特化したLLMを実現するまで

◆ 2012年～

深層学習の普及：画像処理にはCNN (Convolutional Neural Network)が使われるケースが多かった
自然言語処理にはRNN (Recurrent Neural Network)が使われるケースが多かった

◆ 2017年～

深層学習の革新：畳み込みや再帰を使わない（CNNやRNNを完全に不要にする）
シンプルなネットワークTransformer（言わば変換器）のみを使う
Transformerはデコーダー/エンコーダー構造を基本としている
TransformerはSelf-Attention層とPosition-Wise層の組み合わせで動く

◆ OpenAI社の独自成果

深層学習の進化：教師ありファインチューニング

Reward Model学習

RLHF (Reinforcement Learning from Human Feedback)

学習規模と質の進化：GPT-3 ⇒ InstructGPT ⇒ ChatGPT ⇒ ChatGPT Plus

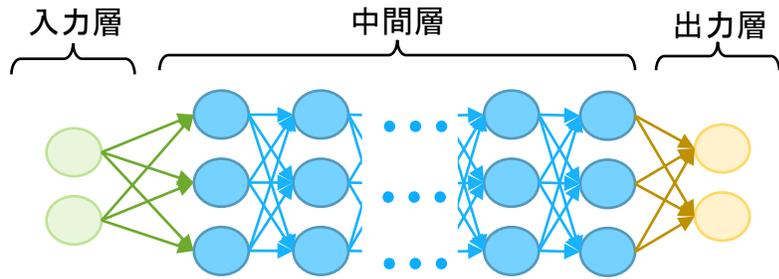
対話（チャット）に特化：間違いを認めることができる

正しくない前提に対する異議を唱えることもできる

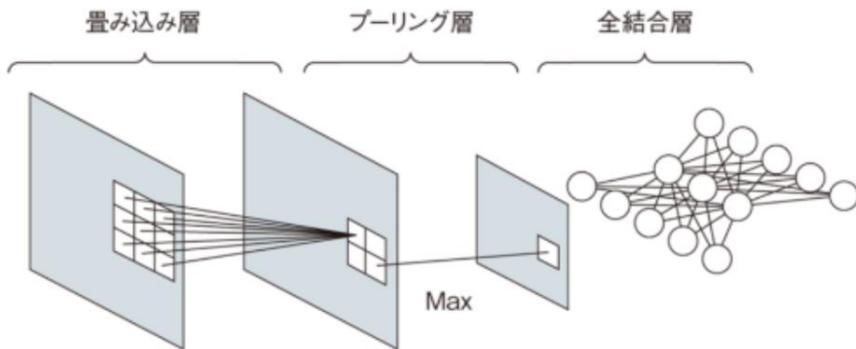
不適切なリクエストには応じない

2012年～ : CNN/RNNとその課題

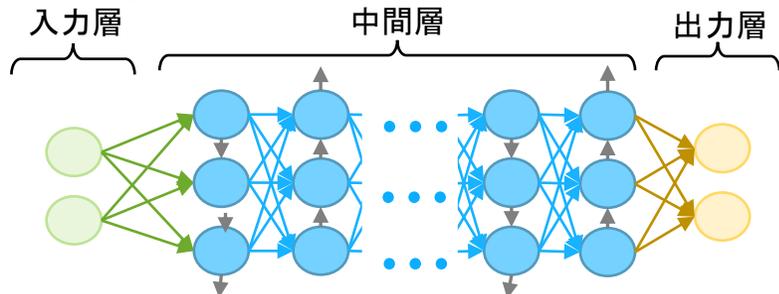
◆ NN (ニューラルネットワーク) の基本形



◆ CNNの構造の概要



◆ RNNの構造の概要



Model: GPT-4

WE

CNNやRNNにはどのような課題や問題があったのか、箇条書きで教えてください。



CNN (畳み込みニューラルネットワーク) とRNN (リカレントニューラルネットワーク) は、それぞれ異なるタイプのニューラルネットワークであり、異なる課題や問題が存在します。以下に、それぞれの課題や問題を箇条書きで示します。

CNNの課題や問題：

1. 局所的な特徴の捉え方に制限がある
2. 固定サイズの入力データにしか対応できない
3. 計算量が大きいため、学習や推論に時間がかかることがある
4. 時系列データや自然言語処理において、効果的な情報の捉え方が難しい
5. 過学習 (overfitting) のリスクがある

RNNの課題や問題：

1. 長期依存性 (long-term dependencies) の問題：長いシーケンスにおいて、過去の情報を効果的に利用できないことがある
2. 勾配消失 (vanishing gradient) 問題：勾配が小さくなりすぎて、学習が進まなくなることがある
3. 勾配爆発 (exploding gradient) 問題：勾配が大きくなりすぎて、学習が不安定になることがある
4. 計算量が大きく、学習や推論に時間がかかることがある
5. 並列化が難しく、GPUを活用した高速化が制限される

2017年：CNN/RNNと決別する論文が登場

◆ 論文"Attention Is All You Need"

- 複数のGoogleエンジニアが担当した論文です。
- Google社内での研究成果を一つにまとめた体裁となっています。
- Google社はこれを積極的に事業展開に結び付ける動きはしませんでした。（翻訳とかに使う程度...）

Model: GPT-4

WE 論文"Attention Is All You Need"を200文字程度で要約してください。

 "Attention Is All You Need"は、Vaswaniらによって2017年に発表された論文で、自然言語処理(NLP)分野に革新的なモデルであるTransformerを紹介しました。従来のリカレントニューラルネットワーク(RNN)や畳み込みニューラルネットワーク(CNN)に依存せず、アテンション機構を中心に据えたネットワーク設計を提案しました。アテンション機構により、長距離の依存関係を効率的に学習でき、並列化による高速化が可能になります。Transformerは、エンコーダ-デコーダ構造を持ち、自己アテンションと位置的エンコーディングを用いて、単語の順序情報を捉えます。論文は、機械翻訳タスクでTransformerが従来のモデルを大幅に上回る性能を発揮し、NLPの新たな基準を確立したことを示しています。

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaier@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

Abstract

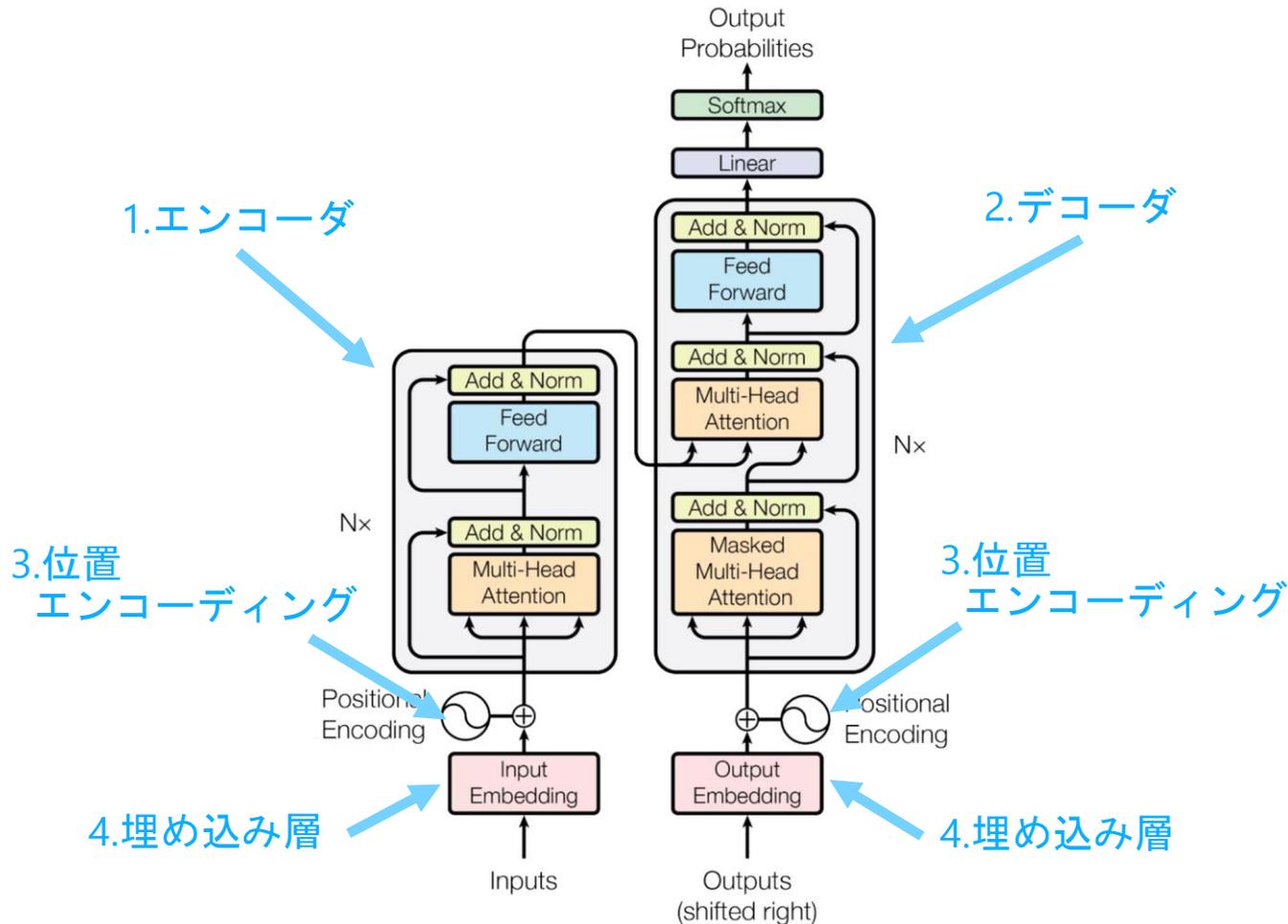
The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

1 Introduction

Recurrent neural networks, long short-term memory [13] and gated recurrent [7] neural networks in particular, have been firmly established as state of the art approaches in sequence modeling and

Transformerの構成要素

◆ Transformerのモデルアーキテクチャ



Model: GPT-4

WE Transformerの構成要素を簡潔に説明してください。

トランスフォーマーは、自然言語処理タスクにおいて高い性能を発揮する深層学習モデルで、以下の主要な構成要素から成り立っています。

1. エンコーダ: 入力シーケンスを固定長のベクトル表現に変換します。エンコーダは、以下のサブコンポーネントで構成されています。

- セルフアテンション層: 入力トークン間の関係を捉え、より有益な情報を持つ表現を生成します。
- フィードフォワードネットワーク: 非線形変換を適用し、各トークンの表現を変換します。
- 正規化層と残差接続: ネットワークの学習を安定化し、より深い層の追加を容易にします。

2. デコーダ: エンコーダからのベクトル表現を使用して、ターゲットシーケンスを生成します。デコーダは、以下のサブコンポーネントで構成されています。

- セルフアテンション層: デコーダ内の各トークン間の関係を捉えます。
- エンコーダ-デコーダアテンション層: 入力シーケンスの情報をデコーダの各ステップに適用します。
- フィードフォワードネットワーク: 各トークンの表現を変換します。
- 正規化層と残差接続: 学習の安定化と深い層の追加を容易にします。

3. 位置エンコーディング: トークンの位置情報をモデルに追加し、シーケンス内の相対的な位置関係をキャプチャします。

4. 埋め込み層: 入力および出力トークンをベクトル表現に変換します。

これらの構成要素が組み合わされ、トランスフォーマーは入力シーケンスをターゲットシーケンスに効果的に変換できるようになります。トランスフォーマーは、自然言語処理のさまざまなタスクに適用可能であり、多くの派生モデルやアーキテクチャが開発されています。

Self-Attention層の役割と優位点

Model: GPT-4

WE

Transformerに組み込まれているアテンションの役割を詳しく教えてください。



トランスフォーマーは、自然言語処理タスクにおいて非常に効果的なディープラーニングアーキテクチャです。トランスフォーマーは、アテンションメカニズムを基本要素として構築されており、これがその性能と柔軟性に大きく寄与しています。アテンションの役割は、主に以下の3つの点で重要です。

1. 重要度に応じた情報の抽出:

アテンションメカニズムは、シーケンス中の各単語やトークンの重要度を評価し、関連性の高い情報を抽出する役割を果たします。アテンションの計算では、入力シーケンス中のすべてのトークン間の関係を同時に考慮するため、文脈を適切に捉えることができます。

2. 長距離依存関係の解決:

RNNやLSTMのような従来のシーケンシャルなアーキテクチャでは、長距離の依存関係を捉えるのが困難でした。アテンションメカニズムは、シーケンス内のすべての位置間の関係を同時に考慮するため、長距離依存関係を効果的に解決することができます。

3. 平行化による計算効率の向上:

アテンションメカニズムは、シーケンス内のトークン間の関係を同時に計算することができるため、計算の平行化が可能で、これにより、トランスフォーマーは、RNNやLSTMのような逐次的なアーキテクチャと比較して、計算効率が大幅に向上します。

トランスフォーマーでは、アテンションメカニズムは主にセルフアテンションと呼ばれる形で実装されています。セルフアテンションは、入力シーケンス内の各トークンが他のすべてのトークンとどの程度関連しているかを評価し、その情報を利用して新しい表現を生成します。これにより、トランスフォーマーは文脈に応じて重要な情報を適切に捉えることができます。

Model: GPT-4

WE

Self-Attentionが従来のRNNで使われていたAttentionよりどこが優れているのか教えてください。

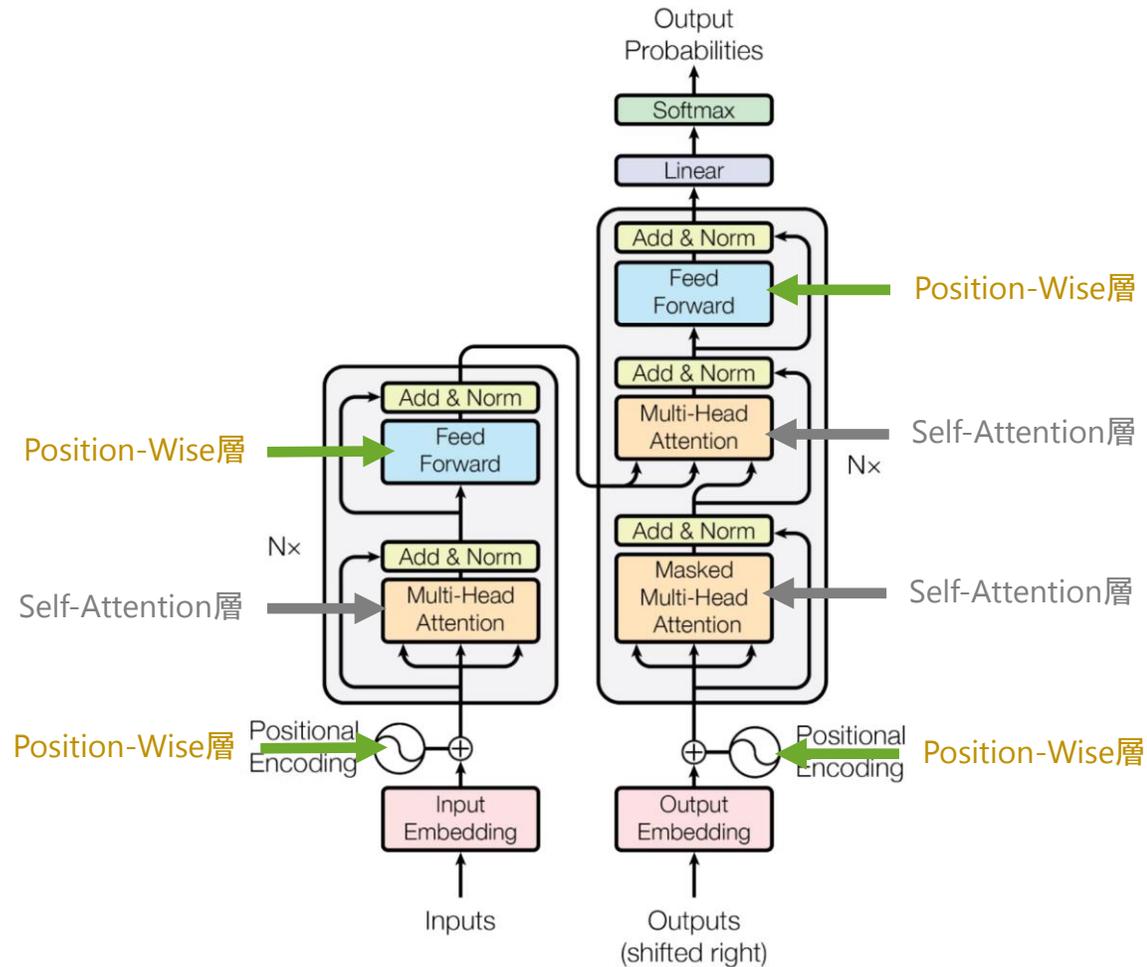


Self-Attention（セルフアテンション）と従来のRNN（Recurrent Neural Network）で使われていたAttentionにはいくつかの違いがあり、それぞれの利点があります。Self-Attentionが従来のRNNで使われていたAttentionに対して優れている点は以下の通りです。

- 並列化の容易さ: Self-Attentionは、入力シーケンス内のすべての要素が互いに直接関連付けられるため、計算を並列化しやすくなります。一方、RNNはシーケンスの要素を逐次的に処理するため、並列化が難しいです。
- 長距離依存関係の扱い: Self-Attentionは、シーケンス内の全ての要素間の関連性を直接計算するため、長距離の依存関係を捉えることが容易になります。RNNでは、シーケンスの長さが増すと、長距離依存関係を捉える能力が低下することが一般的です。
- メモリ効率: Self-Attentionは、入力シーケンスの長さに関わらず、固定サイズのメモリを使用するため、メモリ効率が非常に高いです。RNNでは、シーケンスの長さに応じてメモリ使用量が増加するため、長いシーケンスの処理には不向きです。
- 解釈可能性: Self-Attentionは、各入力要素間の関連性を明示的に計算するため、モデルの出力に対する解釈が容易になります。これに対して、RNNは隠れ状態を通じて情報を伝播させるため、解釈が困難になることがあります。

ただし、Self-AttentionとRNNは、状況に応じて相補的に使用することができるため、どちらが絶対的に優れているわけではありません。適切なアーキテクチャを選択することが重要です。

Position-Wise層の役割



Model: GPT-4

WE Transformerに組み込まれているPosition-Wise層の役割は何ですか？



トランスフォーマー (Transformer) モデルにおけるポジションワイズ (position-wise) 層は、主に入力シーケンス内の各要素の位置に関する情報を処理します。これにより、トランスフォーマーは順序付けられた情報を保持し、シーケンス内の要素間の相対的な位置を学習できます。

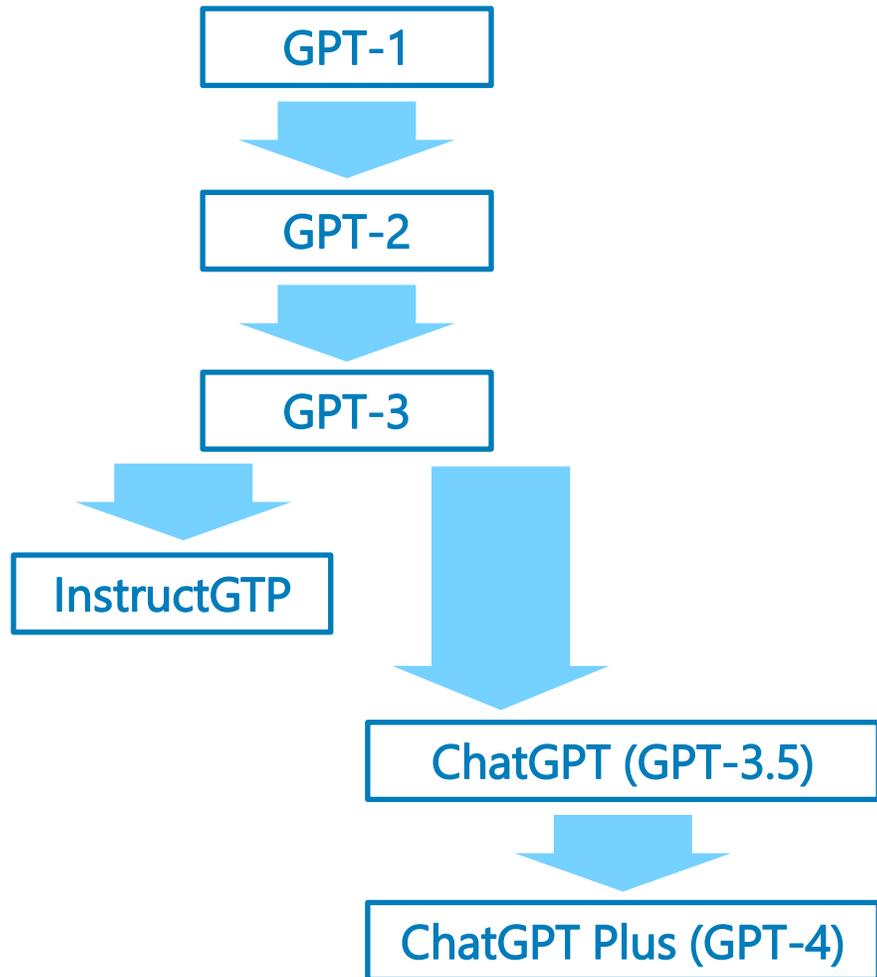
トランスフォーマーは、Self-Attentionメカニズムを使用して、文脈に応じた単語やトークンの表現を捉えます。ただし、このメカニズム自体は位置情報に関して不変であるため、シーケンス内の各トークンの位置情報を追加する必要があります。これが、ポジションワイズ層の主な役割です。

具体的には、ポジションワイズ層にはポジションエンコーディング (position encoding) が含まれます。ポジションエンコーディングは、各トークンの位置情報を表現するベクトルで、入力シーケンスのトークン埋め込みに加算されます。この結果、シーケンス内の各トークンは、その内容と位置情報の両方を表現するようになります。

さらに、ポジションワイズ層には、フィードフォワードニューラルネットワーク (Feed-Forward Neural Network, FFNN) が含まれます。この層は、シーケンス内の各位置において独立して働く全結合層です。FFNNは、Self-Attention層からの出力を処理し、非線形性を導入して、モデルの表現力を向上させます。

要約すると、ポジションワイズ層は、トランスフォーマーモデルがシーケンス内の各要素の位置情報を学習し、処理するために重要な役割を果たしています。ポジションエンコーディングにより位置情報が付加され、フィードフォワードニューラルネットワークが非線形性を導入し、モデルの表現力を強化します。

GPT-4までの進化の概要



- **GPT-1 (2018年)**
 - 1.17億個のパラメータ数です。
 - 単語の文脈を理解する能力を持っていました。
 - 生成したテキストの一貫性や精度には制限や問題がありました。
- **GPT-2 (2019年)**
 - 15億個のパラメータ数です。
 - 文章の生成、質問応答、翻訳、要約など、さまざまなタスクに適用可能です。
 - 情報の正確性や偏りに問題がありました。
- **GPT-3 (2020年)**
 - 1,750億個のパラメータ数です。
 - より深い文脈理解が可能になりました。
 - 多くのタスクを少数の例で学習結果から導き出します。(Few-Shot Learning)
- **InstructGPT (2022年)**
 - GPT-3がベースとなっています。
 - より具体的な指示に従ったテキスト生成能力が強化されています。
- **GPT-3.5ベースのChatGPT (2022年)**
 - 自然な対話を生成して、ユーザーの質問や要求に応答することに特化しています。
 - 2022年11月に一般公開しました。
 - APIが利用可能となりました。
- **GPT-4ベースのChatGPT (2023年)**
 - 約100兆個のパラメータ数です。
 - プラグインが使えるようになりました。

GPT-3はスケールアウトした学習で精度を上げた

Model: GPT-4

WE GPT-3の学習は従来のTransformerを使った学習事例よりはるかパラメータが増えていますが、従来方法から進化した部分がありますか？

 GPT-3は、GPT-2と同様に、基本的なTransformerアーキテクチャに基づいています。ただし、いくつかの進化があります。

1. パラメータ数の大幅な増加: GPT-3は、約1750億のパラメータを持っており、GPT-2や従来のTransformerモデルと比較してはるかに大規模です。この増加により、モデルはより複雑で幅広いタスクを実行できるようになりました。
2. 事前学習データセットの拡大: GPT-3の事前学習には、インターネット上のさまざまなドメインから取得したテキストデータが含まれており、モデルは多様な知識を持っています。これにより、より広範なタスクでより正確な応答を生成できます。
3. 転移学習の改善: GPT-3は、転移学習のパフォーマンスを向上させるために、いくつかの技術的改良が施されています。例えば、モデルは、事前学習時に獲得した知識をうまく活用し、より少ないファインチューニングデータで高性能を発揮できるようになっています。
4. ゼロショット学習とフューショット学習: GPT-3は、ゼロショット学習とフューショット学習の能力が大幅に向上しています。これにより、モデルは新しいタスクに対しても短い学習期間で適応できるようになりました。

GPT-3の進化により、従来のTransformerモデルよりも、より多様で複雑なタスクを効果的に実行できるようになっています。しかし、基本的なアーキテクチャは、従来のTransformerと同様であり、主な進歩はパラメータ数の増加と事前学習データセットの拡大によるものです。

WE ゼロショット学習とフューショット学習について、もう少し詳しく教えてください。

 ゼロショット学習とフューショット学習は、機械学習モデルが新しいタスクを学習する際のアプローチです。これらのアプローチは、トレーニングデータが限られている場合や、新しいタスクに対応するためにモデルを素早く適応させる必要がある場合に特に有用です。

1. ゼロショット学習 (Zero-Shot Learning):

ゼロショット学習では、モデルは新しいタスクを学習するために、事前に用意されたトレーニングデータを全く使わずに、事前学習された知識と推論能力を活用します。つまり、モデルは新しいタスクのための例を見ずに、そのタスクを遂行する方法を学ぶことが期待されます。これは、モデルが一般的な知識と論理的な推論能力を獲得し、それらを新しい問題に適用できるようになることを意味します。

2. フューショット学習 (Few-Shot Learning):

フューショット学習では、モデルは新しいタスクを学習するために、ごく少量のトレーニングデータ（通常は数個から数十個のデータポイント）を利用します。このアプローチは、新しいタスクのデータが限られている場合や、モデルを迅速に適応させる必要がある場合に特に有益です。フューショット学習では、モデルは既存の知識を活用しながら、新しいタスクに関する追加情報を学習することで、そのタスクに対応します。

GPT-3は、ゼロショット学習とフューショット学習の能力が向上しています。これは、大規模な事前学習データセットと多数のパラメータを使用して、モデルが幅広い知識と理解を獲得することができるためです。この結果、GPT-3は、新しいタスクに対しても、事前学習時に獲得した知識と推論能力を適用して、効果的に対応できるようになっています。

GPT-3では570GB以上のコーパスを学習

Model: GPT-4

WE 自然言語におけるコーパスとは何か、またその目的を200文字以内で要約してください。

 コーパスとは、自然言語処理（NLP）や言語学研究において用いられる、言語のサンプルを集めたデータセットです。コーパスは、テキストや音声データなど多様な形式で存在し、異なる言語やジャンルのもがあります。コーパスの目的は、言語パターンや統計的性質を分析し、言語理解や生成技術を開発するための基盤となるデータを提供することです。また、コーパスを用いて、自然言語処理アルゴリズムの学習や評価が行われます。

- **Common Crawlから抽出したコーパス（n個の単語で1コーパス）を用意します**

$$\mathcal{U} = \{u_1, u_2, \dots, u_n\}$$

- **以下の目的関数（対数）でコーパスの文脈を生成できるように ϕ を最適化します**

$$L_1(\mathcal{U}) = \sum_i \log P(u_i | u_{i-k}, \dots, u_{i-1}; \Theta)$$

Model: GPT-4

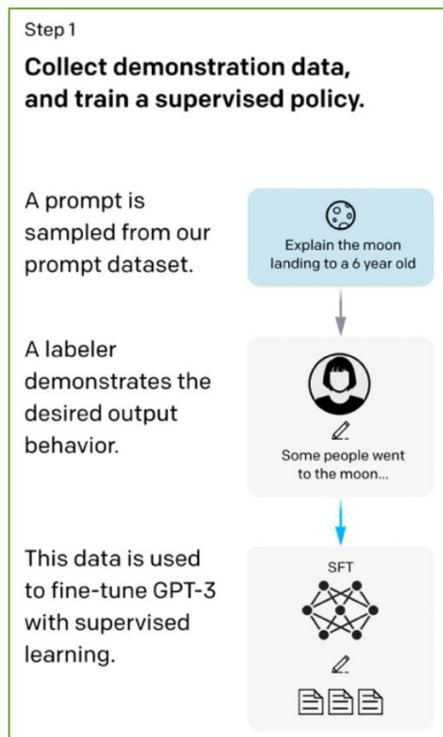
WE GPT-3で使われたコーパスの特徴を教えてください。

 GPT-3（OpenAIの第三世代汎用前処理トランスフォーマー）は、インターネット上の大規模なテキストデータセットから学習しています。以下に、GPT-3で使われたコーパスの主な特徴を示します。

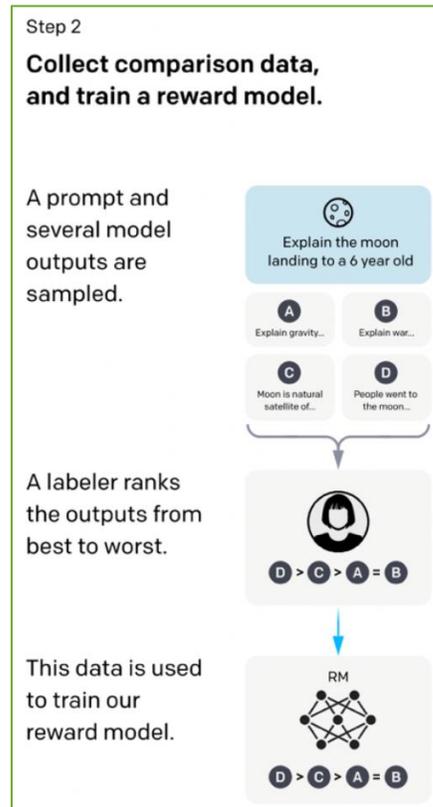
1. 大規模：GPT-3のトレーニングデータは、インターネット上の45テラバイトのテキストデータから抽出されています。これにより、GPT-3は非常に幅広い知識を持つことができます。
2. 多様性：GPT-3のコーパスは、ウェブページ、ブログ、ニュース記事、書籍、論文、フォーラム、百科事典、そしてソーシャルメディアなど、様々なジャンルやドメインからのテキストデータを含んでいます。これにより、GPT-3は多様なトピックや文体に対応できるようになっています。
3. 時代の範囲：GPT-3のデータセットには、過去数十年間のテキストが含まれており、時代や文化的背景に関する知識も持っています。ただし、学習データの大部分は近年のものであり、古典的な文献や歴史的な文書に関する知識は限定的です。また、2021年以降の情報は含まれていません。
4. 言語の範囲：GPT-3は主に英語のテキストデータを学習していますが、他の言語のデータも一部含まれています。そのため、GPT-3は多言語のタスクにも対応できることがありますが、英語以外の言語に関しては、精度や質が低くなる可能性があります。
5. バイアス：GPT-3のコーパスはインターネット上のテキストデータから構成されているため、社会的なバイアスやステレオタイプが反映されることがあります。OpenAIは、モデルのバイアスを軽減する取り組みを行っていますが、完全にバイアスを排除することは難しいとされています。

InstructGPTはアラインメント問題をほぼ解決できた

- InstructGPTは、GPT-3のLLMで「人が好まない文を生成してしまう」というアラインメントの問題を解決するためのプロジェクトです。
- PPO（Proximal Policy Optimization）という3ステップのオンポリシー強化学習を実施しました。
 - 学習ステップ1：教師ありファインチューニング（Supervised Fine-Tuning, SFT）を学習する
 - 学習ステップ2：あらかじめファインチューニングされたReward Modelを学習する
 - 学習ステップ3：強化学習RLHF (Reinforcement Learning from Human Feedback) で報酬関数（reward function）を最大化するように学習する

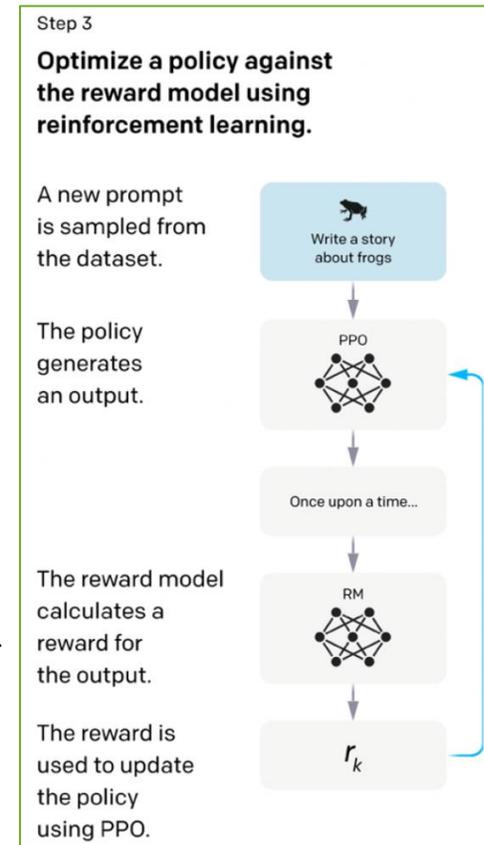


スーパーバイザーポリシーとして1.3万ケースほどを学習



評価軸は：
Truthfulness（真実性）
Harmlessness（無害性）
Helpfulness（有益性）

評点ではなくランキングを学習する



ChatGPTの学習メカニズムはInstructGPTと同じ

November 30, 2022

Authors
OpenAI ↓

[Product, Announcements](#)

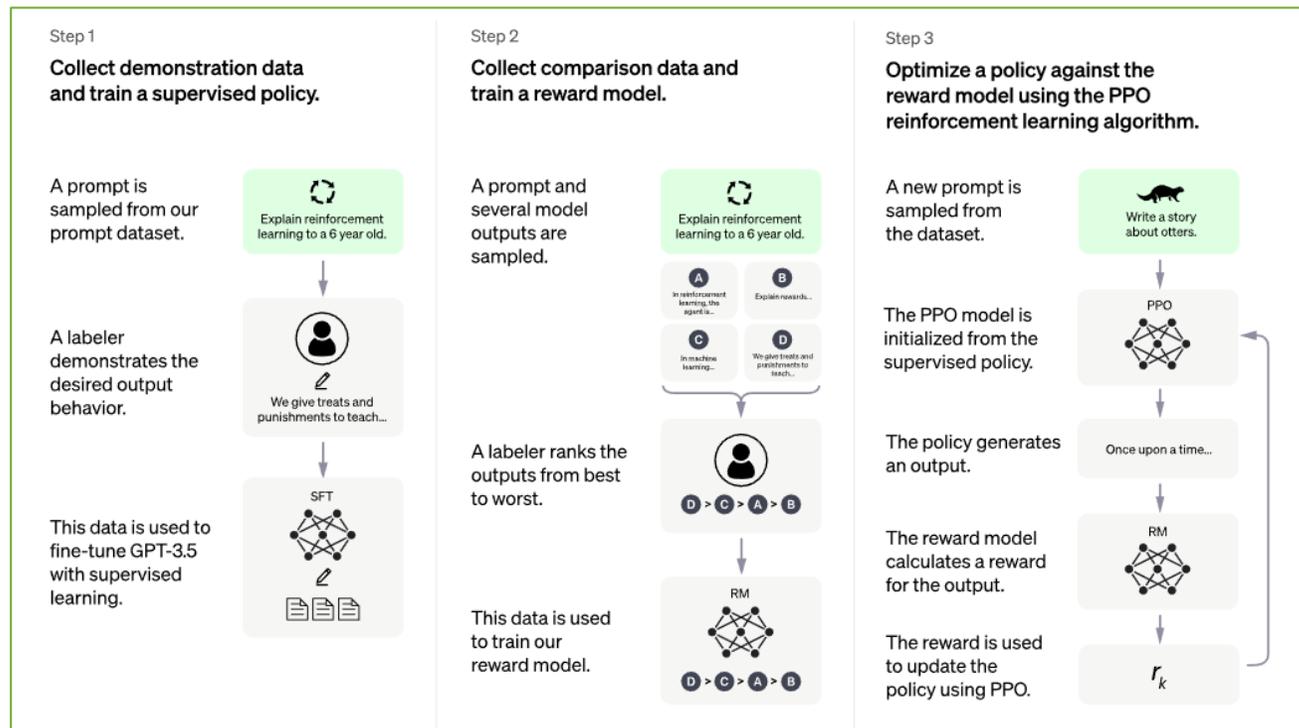
ChatGPT is a sibling model to [InstructGPT](#), which is trained to follow an instruction in a prompt and provide a detailed response.

We are excited to introduce ChatGPT to get users' feedback and learn about its strengths and weaknesses. During the research preview, usage of ChatGPT is free. Try It now at chat.openai.com.

Open AI社の昨年11月30日付のアナウンスによると...

『ChatGPTはInstructGPTの兄弟モデルであり、プロンプトの指示に従い、詳細な応答を提供するようにトレーニングされています。』

学習メカニズムの説明も...
InstructGPTと同等でした。



ご清聴、ありがとうございました。

生成系AIは...

まず2021年に画像生成AIでブレイクしました。

そして2022年からChatGPTのようなテキスト生成AIがブレイクしています。

どちらも依頼を書いて投入すると...

専門の人が対応したかのような高品質の結果が生成されて返ってきます。

専門の知識を必要とするような仕事をAIが肩代わりできるようになっていくわけで...
これからも加速度的に精度が上がって既存ビジネスでの人の作業を置き換えていくでしょう。

これはまだほんの序の口です。

現在のAIの応答出力はコンピュータスクリーン上にしか出力できませんが...

しかしAPIとPluginsが登場して半年を待たずにText2Textを超えて続々と派生AIが登場...

Auto-GPT、AgentGPT、GitHub Copilot、Creative Reality Studio、HuggingGPT...

これからの『リスキリング』はプロンプトエンジニア？

参考資料 : Transformerを提案した論文

[Attention Is All You Need](#)

<https://arxiv.org/pdf/1706.03762.pdf>

arXiv:1706.03762v5 [cs.CL] 6 Dec 2017

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaizer@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

1 Introduction

Recurrent neural networks, long short-term memory [13] and gated recurrent [7] neural networks in particular, have been firmly established as state of the art approaches in sequence modeling and

参考資料 : chainarによるTransformerサンプル

[Attention Is All You Need](https://github.com/soskek/attention_is_all_you_need)

https://github.com/soskek/attention_is_all_you_need

The screenshot shows the GitHub repository page for 'soskek/attention_is_all_you_need'. The repository is public and has 14 watchers, 70 forks, and 307 stars. The main content area displays a list of files and their commit history:

File Name	Commit Message	Commit Date
.gitignore	Create .gitignore	6 years ago
.gitmodules	transform forked repo	6 years ago
LICENSE	Create LICENSE	6 years ago
Observe_Position_Encoding.ipynb	add note for position encoding	6 years ago
README.md	fix	6 years ago
download_wmt.sh	improve usability	6 years ago
net.py	fix	6 years ago
preprocess.py	fix readme and rename	6 years ago
subfuncs.py	fix much for well optimization	6 years ago
train.py	fix	6 years ago

The README.md file is expanded, showing the following content:

Transformer - Attention Is All You Need

Chainer-based Python implementation of Transformer, an attention-based seq2seq model without convolution and recurrence.
If you want to see the architecture, please see [net.py](#).

See "Attention Is All You Need", Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, arxiv, 2017.

This repository is partly derived from my [convolutional seq2seq](#) repo, which is also derived from Chainer's official [seq2seq example](#).

The right sidebar provides additional information:

- About:** Transformer of "Attention Is All You Need" (Vaswani et al. 2017) by Chainer. Tags include google, deep-neural-networks, deep-learning, neural-network, chainar, and attention-mechanism.
- Readme:** Present.
- License:** BSD-3-Clause license.
- Stars:** 307 stars.
- Watching:** 14 watching.
- Forks:** 70 forks.
- Releases:** No releases published.
- Packages:** No packages published.
- Languages:** Jupyter Notebook (91.2%), Python (8.8%).

参考論文：GPT-3のLLMに関する論文

[Language Models are Few-Shot Learners](https://papers.nips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf)

https://papers.nips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Language Models are Few-Shot Learners

Tom B. Brown* Benjamin Mann* Nick Ryder* Melanie Subbiah*
Jared Kaplan† Prafulla Dhariwal Arvind Neelakantan Pranav Shyam
Girish Sastry Amanda Askell Sandhini Agarwal Ariel Herbert-Voss
Gretchen Krueger Tom Henighan Rewon Child Aditya Ramesh
Daniel M. Ziegler Jeffrey Wu Clemens Winter
Christopher Hesse Mark Chen Eric Sigler Matusz Litwin Scott Gray
Benjamin Chess Jack Clark Christopher Berner
Sam McCandlish Alec Radford Ilya Sutskever Dario Amodei

Abstract

We demonstrate that scaling up language models greatly improves task-agnostic, few-shot performance, sometimes even becoming competitive with prior state-of-the-art fine-tuning approaches. Specifically, we train GPT-3, an autoregressive language model with 175 billion parameters, 10x more than any previous non-sparse language model, and test its performance in the few-shot setting. For all tasks, GPT-3 is applied without any gradient updates or fine-tuning, with tasks and few-shot demonstrations specified purely via text interaction with the model. GPT-3 achieves strong performance on many NLP datasets, including translation, question-answering, and cloze tasks. We also identify some datasets where GPT-3's few-shot learning still struggles, as well as some datasets where GPT-3 faces methodological issues related to training on large web corpora.

1 Introduction

NLP has shifted from learning task-specific representations and designing task-specific architectures

参考論文：InstructGPTのRLHF等に関する論文

[Training language models with RLHF](https://arxiv.org/pdf/2203.02155.pdf)

<https://arxiv.org/pdf/2203.02155.pdf>

arXiv:2203.02155v1 [cs.CL] 4 Mar 2022

Training language models to follow instructions with human feedback

Long Ouyang* Jeff Wu* Xu Jiang* Diogo Almeida* Carroll L. Wainwright*

Pamela Mishkin* Chong Zhang Sandhini Agarwal Katarina Slama Alex Ray

John Schulman Jacob Hilton Fraser Kelton Luke Miller Maddie Simens

Amanda Askell† Peter Welinder Paul Christiano*†

Jan Leike* Ryan Lowe*

OpenAI

Abstract

Making language models bigger does not inherently make them better at following a user’s intent. For example, large language models can generate outputs that are untruthful, toxic, or simply not helpful to the user. In other words, these models are not *aligned* with their users. In this paper, we show an avenue for aligning language models with user intent on a wide range of tasks by fine-tuning with human feedback. Starting with a set of labeler-written prompts and prompts submitted through the OpenAI API, we collect a dataset of labeler demonstrations of the desired model behavior, which we use to fine-tune GPT-3 using supervised learning. We then collect a dataset of rankings of model outputs, which we use to further fine-tune this supervised model using reinforcement learning from human feedback. We call the resulting models *InstructGPT*. In human evaluations on our prompt distribution, outputs from the 1.3B parameter InstructGPT model are preferred to outputs from the 175B GPT-3, despite having 100x fewer parameters. Moreover, InstructGPT models show improvements in truthfulness and reductions in toxic output generation while having minimal performance regressions on public NLP datasets. Even though InstructGPT still makes simple mistakes, our results show that fine-tuning with human feedback is a promising direction for aligning language models with human intent.

1 Introduction

Large language models (LMs) can be “prompted” to perform a range of natural language process-

参考調査資料：LLMがもたらす労働市場への潜在的影響

[Labor Market Impact Potential of LLM](https://arxiv.org/pdf/2303.10130.pdf)

<https://arxiv.org/pdf/2303.10130.pdf>

WORKING PAPER

GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models

Tyna Eloundou¹, Sam Manning^{1,2}, Pamela Mishkin*¹, and Daniel Rock³

¹OpenAI

²OpenResearch

³University of Pennsylvania

March 27, 2023

Abstract

We investigate the potential implications of large language models (LLMs), such as Generative Pre-trained Transformers (GPTs), on the U.S. labor market, focusing on the increased capabilities arising from LLM-powered software compared to LLMs on their own. Using a new rubric, we assess occupations based on their alignment with LLM capabilities, integrating both human expertise and GPT-4 classifications. Our findings reveal that around 80% of the U.S. workforce could have at least 10% of their work tasks affected by the introduction of LLMs, while approximately 19% of workers may see at least 50% of their tasks impacted. We do not make predictions about the development or adoption timeline of such LLMs. The projected effects span all wage levels, with higher-income jobs potentially facing greater exposure to LLM capabilities and LLM-powered software. Significantly, these impacts are not restricted to industries with higher recent productivity growth. Our analysis suggests that, with access to an LLM, about 15% of all worker tasks in the US could be completed significantly faster at the same level of quality. When incorporating software and tooling built on top of LLMs, this share increases to between 47 and 56% of all tasks. This finding implies that LLM-powered software will have a substantial effect on scaling the economic impacts of the underlying models. We conclude that LLMs such as GPTs exhibit traits of general-purpose technologies, indicating that they could have considerable economic, social, and policy implications.

1 Introduction

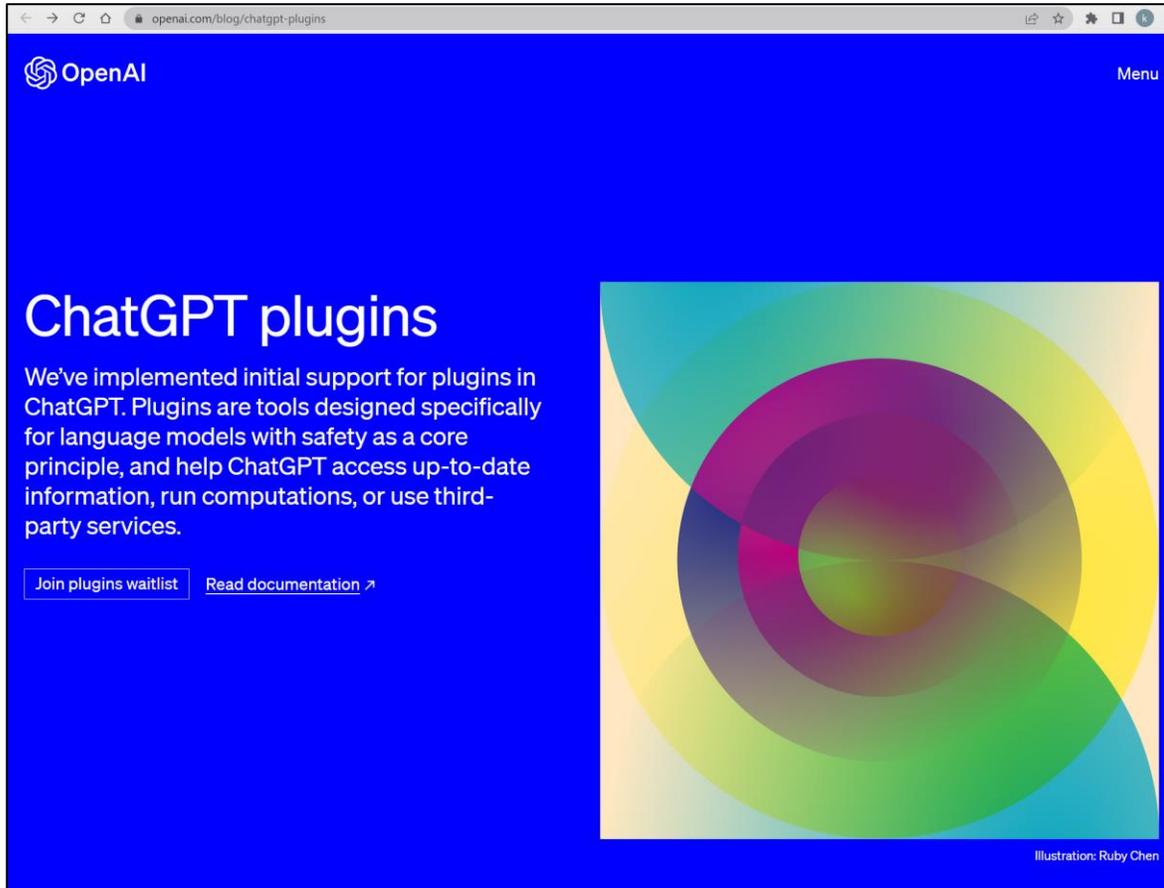
As shown in Figure 1, recent years, months, and weeks have seen remarkable progress in the field of generative AI and large language models (LLMs). While the public often associates LLMs with various iterations of the

arXiv:2303.10130v4 [econ.GN] 23 Mar 2023

参考 : ChatGPT plugins

[ChatGPT plugins](https://openai.com/blog/chatgpt-plugins)

<https://openai.com/blog/chatgpt-plugins>



The screenshot shows the OpenAI waitlist form for plugins. The form is titled "How do you want to use plugins?*" and has two radio button options: "I want to try plugins in ChatGPT" and "I am a developer and want to build a plugin". The second option is selected. Below the options, there is a note: "We will be granting access to a limited number of users and developers at launch." The form then asks "What do you want to build a plugin for?*" and has a text input field containing "I'm a freelance developer. I want to set up and try out a development environment using your code interpreter." Below the text input field, there is a note: "Please include what the plugin will do, if you already have an API that will work with the plugin, and any other relevant info you want to share with our team." The form then has several input fields: "Company name" (containing "freelance developer"), "Company size" (a dropdown menu with "1 - 20" selected), "Twitter profile URL" (containing "https://twitter.com/Hj_weasel"), and "GitHub profile URL" (containing "https://github.com/Kazuaki2018"). Below the input fields, there is a note: "Please include the full link to your profile." At the bottom of the form, there is a "Join waitlist" button.