

## NIIオープンハウス 2023 基調講演

(対話型文章生成AI)

# ChatGPTを支える技術



東北大学  
自然言語処理  
研究グループ



Fai Lab

東北大学  
データ駆動科学・AI教育研究センター  
大学院情報科学研究科 人工知能基礎学講座  
鈴木潤

# [注意事項] 事前のお断り

- 「ChatGPT」という用語
  - 個別のモデルではなく「対話型文章生成AI」という意図での総称として用いられる場合がある  
=> 後続で発表された「GPT-4」なども含む
  - 本資料では以降 ChatGPT は個別のモデルを表す
  - 総称は「対話型文章生成AI」「文章生成AI」と記載
- 本資料の情報の正確性について
  - 対話型文章生成AI / ChatGPT 関連の情報は頻繁に更新
  - あくまで2023.06.02時点の情報

本資料の情報が時間経過と共に正しくなくなる可能性があることに注意

# [注意事項] ChatGPTに関する情報

## GPT-4 Technical Report

<https://cdn.openai.com/papers/gpt-4.pdf>

### 2 Scope and Limitations of this Technical Report

This report focuses on the capabilities, limitations, and safety of the GPT-4 Transformer-style model [39] pre-trained to predict the next token in a document, using both publicly available data (such as internet data) and data licensed from third-party providers. The model was then fine-tuned using Reinforcement Learning from Human Feedback (RLHF) [40]. **Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar.**

### 詳細不明 !!

[簡単な訳] GPT-4のような大規模モデルの競争環境と安全性を考慮し本報告書では、アーキテクチャ (モデルサイズを含む)、計算環境、学習計算量、データ、学習方法などに関する詳細な情報を記載しない。

ChatGPT (OpenAI) や Bard (Google) も同様に詳細は公表しない方針

今後、大手IT企業がサービスとして利用する**対話型文章生成AI**の詳細な情報は出てこない見込み

本資料はChatGPTの解説ではなく、そこで利用されている技術に関する解説

# ChatGPT (対話型文章生成AI) とは？

一言で言うと...

[一般向け]

(人が人に話すような) **対話形式の指示**を受け付け  
その指示に適した文章を生成する **文章生成器**

# [ChatGPT] なぜ注目されているのか？

- 月間アクティブユーザー数**1億人**達成：**2ヶ月**

比較参考：

TikTok  
**9ヶ月**

Instagram  
**2年半**

これらは画像や映像のサービス  
それに対して、高々文章を入力し  
たら文章を返すだけのサービス

=> まさに驚異的と言える

- 根拠資料
  - ChatGPT sets record for fastest-growing user base

(Reutersの記事から取得)

<http://reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>

=> 何がそんなにユーザーの心にささったのか？

# [ChatGPT] なぜ注目されているのか？

- 一般に普及するのはずっと先とされていたクオリティのサービスがいきなり無料で利用可
    - 相手が人間の場合と同じ感覚で文章を入力してよい
      - チャット的な対話形式に対応
      - 専門知識が不要
    - ユーザーのあらゆる要求/指示に対応 (しようと頑張る)
      - 例：質問応答，翻訳，文書要約，文章構成，スピーチ/レポート作成，ブレスト相手，語学学習，プログラム生成. . .
- ⇒ (全ての指示に適切に回答できるわけではないが) そこそこ妥当な回答

# 利用例：翻訳

j

最近のAI、特にChatGPTのような生成AIの最近の発展には驚かされます。という文を英語と中国語とドイツ語とロシア語とアラビア語とタイ語に翻訳して



英語:

"I am amazed by the recent advancements in AI, especially in the development of generative AIs like ChatGPT."

中国語 (簡体字):

"最近人工智能的发展让我感到惊讶，特别是像ChatGPT这样的生成型人工智能的发展。"

中国語 (繁體字):

"最近人工智能的發展讓我感到驚訝，特別是像ChatGPT這樣的生成型人工智能的發展。"

ドイツ語:

"Ich bin beeindruckt von den jüngsten Fortschritten in der KI, insbesondere bei der Entwicklung von generativen KIs wie ChatGPT."

ロシア語:

"Недавние достижения в области ИИ, особенно в развитии генеративных ИИ, таких как ChatGPT, меня поразили."

アラビア語:

"ChatGPT أنا مدهش من التطورات الأخيرة في مجال الذكاء الاصطناعي، خاصة في تطور الذكاء الاصطناعي الانتاجي مثل"

タイ語:

"ฉันประหลาดใจกับความก้าวหน้าล่าสุดในการพัฒนา AI โดยเฉพาะในการพัฒนา AI สร้างเช่น ChatGPT."

# 利用例：レポート

j

人工知能の講義の最終レポートがあるのだけど、200字でまとめてほしい。特にChatGPTなどの文章の生成系AIと、Stable Diffusionなどの画層の生成系AIについて言及してほしい



人工知能の最終レポートのまとめを200字で提供します。

ChatGPTや他の文章生成AIは、自然言語処理技術を使用して、人間のように対話したり文章を生成したりすることができます。これらのモデルは、大規模なトレーニングデータセットを使用して学習し、さまざまなタスクに対して優れた結果を示すことがあります。

一方、Stable Diffusionなどの画像生成AIは、生成モデルを使用して高品質な画像を生成します。これらのモデルは、学習時に大量の画像データを使用し、ランダムなノイズからリアルな画像を生成することができます。

これらのAI技術は、現在の技術革新の一環として非常に興味深いものです。自然言語処理や画像生成の分野において、より高度な応用が期待されています。しかし、倫理的な問題や課題も存在し、適切な使用法や社会的な影響についての考察も必要です。AIの進化と発展に伴い、これらの技術はますます重要性を増していくでしょう。

# 利用例：スピーチ

j 明日、言語処理学会のプログラム委員のキックオフミーティングがあり、委員長として挨拶を1分ぐらい話す必要があるのだけど、その原稿の草稿を作ってください、できるだけ最近のAIブームやChatGPTについて触れながら前向きな感じの内容をお願いします。

 皆さん、こんにちは！

この度は言語処理学会のプログラム委員のキックオフミーティングにお集まりいただき、誠にありがとうございます。私は委員長として、今後の活動に向けて非常にワクワクしています。

まずは、最近のAIブームについてお話ししたいと思います。AIの進歩は私たちの生活や社会に大きな変革をもたらしています。特に言語処理の領域では、様々な進展があり、私たちの研究分野がますます重要性を増しています。

ChatGPTのような最新の自然言語処理モデルは、私たちが以前に想像もしなかったようなタスクにも成功を取っています。例えば、機械翻訳、要約、質問応答、会話エージェントなど、多岐にわたる応用が可能です。これにより、私たちはより多くの人々とより自然な形でコミュニケーションを取ることができます。

AIが持つ潜在能力は驚くべきものですが、私たちの役割はそれを活用し、さらなる進歩をもたらすことです。私たちの研究分野では、AIが人々の生活を豊かにし、社会全体の課題に取り組む手助けとなる可能性があります。例えば、教育分野での個別化された学習支援、医療分野での効率的な診断や治療支援、情報アクセスの改善など、数多くの可能性が広がっています。

私たちの使命は、AI技術の進歩を追いかけるだけでなく、その社会への波及効果を考え、倫理的な観点からも責任を持つことです。私たちの研究や開発においては、公平性、透明性、個人情報保護などの重要な要素を念頭に置く必要があります。

明るい未来への一歩として、私たちが果たすべき役割は大きいです。互いに協力し、知識とアイデアを共有しながら、言語処理の分野をより前進させていきましょう。新たなテクノロジーや新しい研究トピックが

j 続けて

 私たちの前に広がっていますが、その中で私たちは創造性と革新を追求し、社会にポジティブな影響を与えることができます。

今回のプログラム委員会の活動においても、AIの最新の動向や成果を反映させることが重要です。私たちは革新的な研究を発表し、学会を通じて情報を共有し、共同研究やコラボレーションを促進することができます。また、新しい人材を育成するために、若手研究者の支援や学生の参加を積極的に推進することも考えていきましょう。

さらに、AI技術の進歩は速いペースで進んでいますので、私たち自身も常に学び続ける姿勢が求められます。最新の研究トピックや手法について常に情報を収集し、国内外の研究者との交流を通じて知識を深めていきましょう。

最後に、私たちの研究や活動が社会に与える影響についても考えていきましょう。AI技術の導入には様々な課題や倫理的な問題が伴いますが、私たちが主体的に関与し、その解決に向けて努力することが重要です。公平性や個人のプライバシー保護など、社会的な側面を考慮しながら研究を進め、AI技術の社会への適切な展開に寄与していきましょう。

皆さんと共に、AIの進化と言語処理の未来を築いていくことにワクワクしています。一緒に素晴らしい成果を生み出し、前向きな変化をもたらしましょう！

ありがとうございました。

# 利用例：文章の書き換え

j 以下のスピーチをもう少しカジュアルな場で話す感じに仕上げて

皆さん、こんにちは！

この度は言語処理学会のプログラム委員のキックオフミーティングにお集まりいただき、誠にありがとうございます。私は委員長として、今後の活動に向けて非常にワクワクしています。

まずは、最近のAIブームについてお話ししたいと思います。AIの進歩は私たちの生活や社会に大きな変革をもたらしています。特に言語処理の領域では、様々な進展があり、私たちの研究分野がますます重要性を増しています。

ChatGPTのような最新の自然言語処理モデルは、私たちが以前に想像もしなかったようなタスクにも成功を収めています。例えば、機械翻訳、要約、質問応答、会話エージェントなど、多岐にわたる応用が可能です。これにより、私たちはより多くの人々とより自然な形でコミュニケーションを取ることができます。

AIが持つ潜在能力は驚くべきものですが、私たちの役割はそれを活用し、さらなる進歩をもたらすことです。私たちの研究分野では、AIが人々の生活を豊かにし、社会全体の課題に取り組む手助けとなる可能性があります。例えば、教育分野での個別化された学習支援、医療分野での効率的な診断や治療支援、情報アクセスの改善など、数多くの可能性が広がっています。

私たちの使命は、AI技術の進歩を追いかけるだけでなく、その社会への波及効果を考え、倫理的な観点からも責任を持つことです。私たちの研究や開発においては、公平性、透明性、個人情報保護などの重要な要素を念頭に置く必要があります。



みんな、やっほー！

今日は言語処理学会のプログラム委員のキックオフミーティングに来てくれて、ありがとう！委員長の俺は、これからの活動がすごく楽しみだぜ。

まずは、最近のAIブームについてちょっと話そうと思うんだ。AIの進歩は、めちゃくちゃ私たちの生活や社会に影響を与えてるんだよ。特に言語処理の分野では、いろんな進展があって、俺たちの研究分野がますます重要になってきてるんだ。

最新の自然言語処理モデル、ChatGPTみたいなやつは、前には考えられなかったようなタスクもやっちゃってるんだ。例えば、機械翻訳や要約、質問応答、会話エージェントなんか、いろんな応用ができるんだ。そうすることで、もっとたくさんの人々と自然な感じでコミュニケーションを取れるようになるんだ。

AIの持つ潜在能力はすごいんだけど、俺たちの役割はそれを使ってさらに進歩させることなんだ。俺たちの研究分野では、AIが人々の生活を豊かにするのに役立ち、社会の問題にも取り組むことができるんだ。例えば、教育の分野で個別化された学習支援をしたり、医療の分野で効率的な診断や治療をサポートしたり、情報のアクセスを改善したり、いろんな可能性があるんだ。

俺たちの使命は、AIの技術の進歩を追いかけるだけじゃなくて、それが社会にどんな影響を与えるかを考えて、倫理的な観点からも責任を持つことなんだ。俺たちの研究や開発では、公平さや透明性、個人情報保護とか、大事な要素を忘れずに考える必要があるんだ。

# 利用例：ブレインストーミング

**j** 今までにない新しいお菓子を新開発したいのだけれど、アイデアが浮かばないんです。今までにない食べた時に見た目のインパクトがあるお菓子のアイデアのブレインストーミングをしたいのですが、まずたたき台になるアイデアをだしてもらえないですか？

 もちろんです！以下に、見た目のインパクトがある新しいお菓子のアイデアのいくつかをご提案します。これらはブレインストーミングのたたき台としてご利用いただければと思います。

1. カラフルな宇宙ケーキ: レインボーカラーのスポンジケーキやクリームを使い、宇宙のような美しい模様を作り出します。食べるときに宇宙をイメージできるような、幻想的なデザートです。
2. ショコラテリウム: チョコレートで作られた小さなテリウム（植物の鉢）を作ります。内部には様々なフレーバーのチョコレートやカラフルなキャンディが詰まっており、食べるときにテリウムが開いて中のお菓子が現れる仕掛けになっています。
3. 立体迷路クッキー: 立体的な迷路の形をしたクッキーを作ります。クッキーの中には果物やチョコチップが隠れており、食べながら迷路を解いていく楽しみがあります。
4. ジュエルドーナツ: キラキラと輝くようなドーナツを作ります。透明なグレースや飴でコーティングし、表面にはカラフルな食用ジュエルを散りばめます。食べるときにまるで宝石を食べているような感覚が味わえます。
5. ピクセルクッキー: クッキーの表面に小さなピクセルアートを作ります。各ピクセルには違う味や色のフロスティングが使われており、食べるたびに違う組み合わせの味が楽しめます。

これらのアイデアは、見た目にインパクトがあり、食べる楽しみや驚きを提供することを目指しています。お好みや発展させたいアイデアに合わせてカスタマイズいただければと思います。どのアイデアが気に入りましたか？

# 利用例：ギャグの説明/生成

j ふとんがふっとんだ、というギャグがあるんだけど、この意味を説明して

「ふとんがふっとんだ」というギャグは、日本語の言葉遊びや音の響きを利用したユーモアの一つです。

「ふとん」は日本語で「布団」という寝具を指します。「ふっとんだ」という表現は、日本語の動詞「ふっとぶ」（風に吹かれて飛ぶ）から派生しており、「ふっとんだ」は「ふとんが飛んだ」という意味になります。

このギャグでは、「ふとん」と「ふっとんだ」の音の響きが似ていることを利用しています。つまり、言葉の響きのギャップや語呂合わせによって笑いを生み出すことを目的としています。

例えば、以下のようなシチュエーションで使われることがあります：

人A: 「昨日、家の布団がふっとんだんだよ！」

人B: 「えっ、本当に？ どうして？」

人A: 「いや、ふっとんだと言ったらふとんがふっとんだってことだよ！ 笑」

このギャグは短いポケとツッコミの形式で使われ、言葉遊びや意外性を楽しむことが特徴です。日本の落語やコント番組などでよく見られるジャンルの一つです。

j これに似た新しいギャグを作ってみて

もちろんです！ 以下に新しいギャグを作ってみます：

人A: 「昨日、家のテレビがテレポートしたんだよ！」

人B: 「えっ、本当に？ どうして？」

人A: 「いや、テレポートしたと言ったら、テレビがテレポートしたってことだよ！ 笑」

このギャグでは、「テレビ」と「テレポートした」の音の響きを利用しています。言葉の意味と響きのギャップによって笑いを生み出しています。

他の例としては：

人A: 「今日、ペンがペンギンになったんだよ！」

人B: 「えっ、本当に？ どうして？」

人A: 「いや、ペンギンになったと言ったら、ペンがギンになったってことだよ！ 笑」

このギャグでは、「ペン」と「ペンギンになった」の音の響きを利用しています。また、意外性も加えています。

これらの例からもわかるように、新しいギャグを作る際には、言葉の響きや意外性を活かした面白さを追求することがポイントです。ただし、ギャグの面白さは人によって異なるため、笑いのセンスや文化背景によって受け取り方が異なるかもしれません。

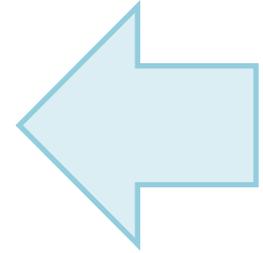
- 技術の説明

# ChatGPTを支える技術

- 基盤
  - 言語モデル
- 成功の要因
  - ① 多層ニューラルネットワーク (DNN) の利用
  - ② 大規模化 (パラメタ数/データ量)
  - ③ 指示文設計 (プロンプトエンジニアリング)
  - ④ 指示文 (+対話文) チューニング
  - ⑤ 人手点数付け結果の活用

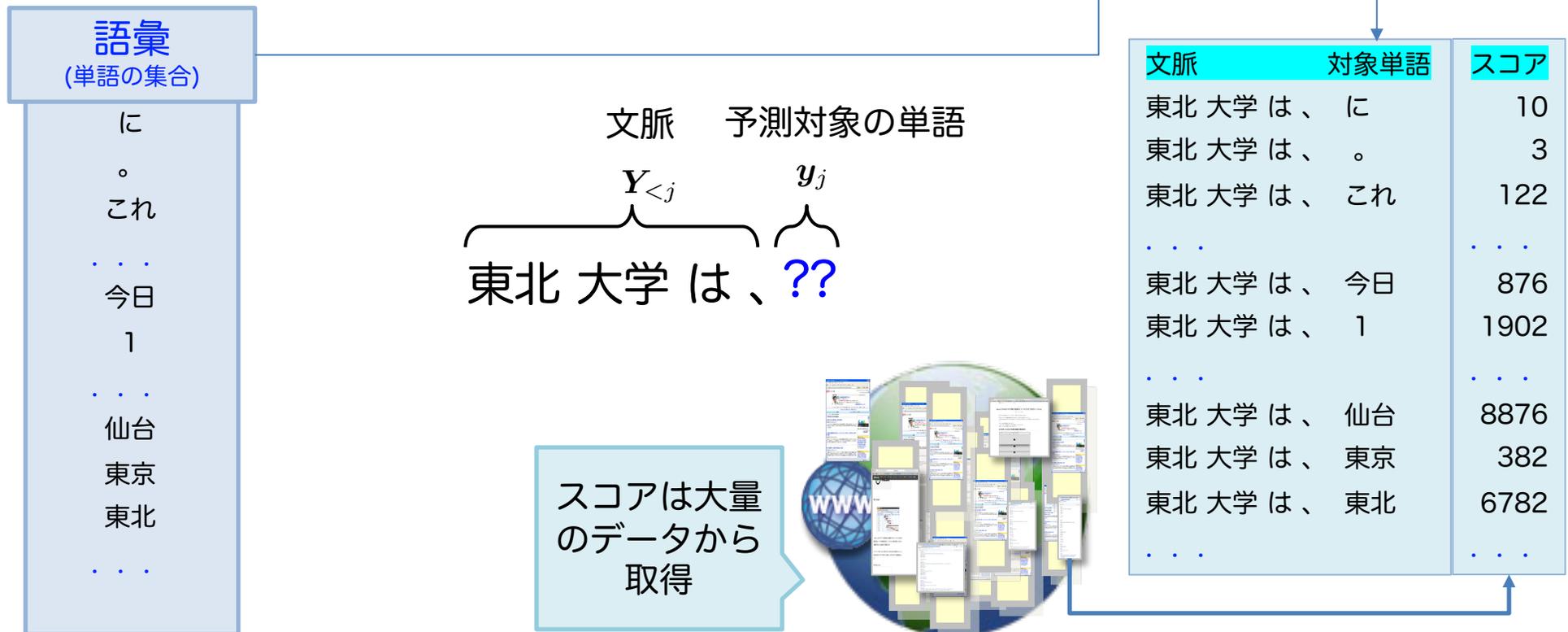
# ChatGPTを支える技術

- 基盤
  - 言語モデル
- 成功の要因
  - ① 多層ニューラルネットワーク (DNN) の利用
  - ② 大規模化 (パラメタ数/データ量)
  - ③ 指示文設計 (プロンプトエンジニアリング)
  - ④ 指示文 (+対話文) チューニング
  - ⑤ 人手点数付け結果の活用



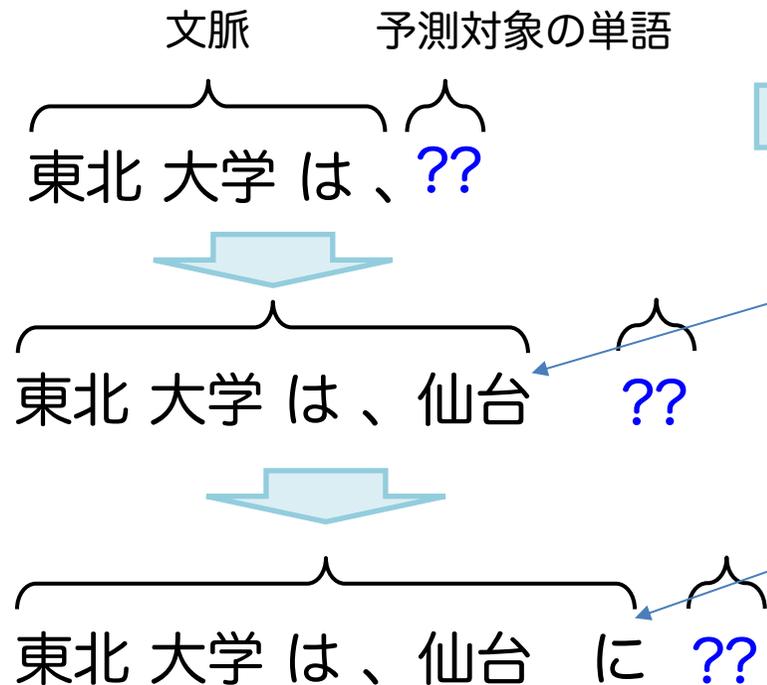
# 基盤：言語モデル

- 文章の出現確率を予測する**確率モデル**
  - 主に与えられた文脈に対して次の単語の予測確率をモデル化
    - $j - 1$  単語の文脈  $Y_{<j}$  が与えられた条件で  $j$  番目の単語  $y_j$  の出やすさの確率分布  $P(y_j | Y_{<j})$  をモデル化



# 言語モデルによる文生成

- 文章の先頭から1単語ずつ次の単語を予測  
=> 予測した単語は文脈として再利用  
生成が終わるまで繰り返す



| 文脈     | 対象単語 | スコア  |
|--------|------|------|
| 東北大学は、 | に    | 10   |
| 東北大学は、 | 。    | 3    |
| 東北大学は、 | これ   | 122  |
| ...    | ...  | ...  |
| 東北大学は、 | 今日   | 876  |
| 東北大学は、 | 1    | 1902 |
| ...    | ...  | ...  |
| 東北大学は、 | 仙台   | 8876 |
| 東北大学は、 | 東京   | 382  |
| 東北大学は、 | 東北   | 6782 |
| ...    | ...  | ...  |

| 文脈        | 対象単語 | スコア |
|-----------|------|-----|
| 東北大学は、仙台に | に    | 873 |
| 東北大学は、仙台  | 。    | 3   |
| 東北大学は、仙台  | これ   | 12  |
| ...       | ...  | ... |
| 東北大学は、仙台  | 今日   | 8   |
| 東北大学は、仙台  | 1    | 126 |
| ...       | ...  | ... |
| 東北大学は、仙台  | 仙台   | 15  |
| 東北大学は、仙台  | 東京   | 2   |
| 東北大学は、仙台  | 東北   | 89  |
| ...       | ...  | ... |

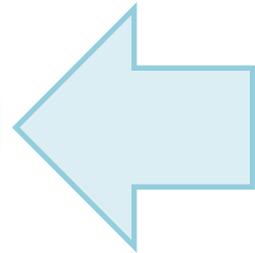
# ChatGPTを支える技術

- 基盤

- 言語モデル

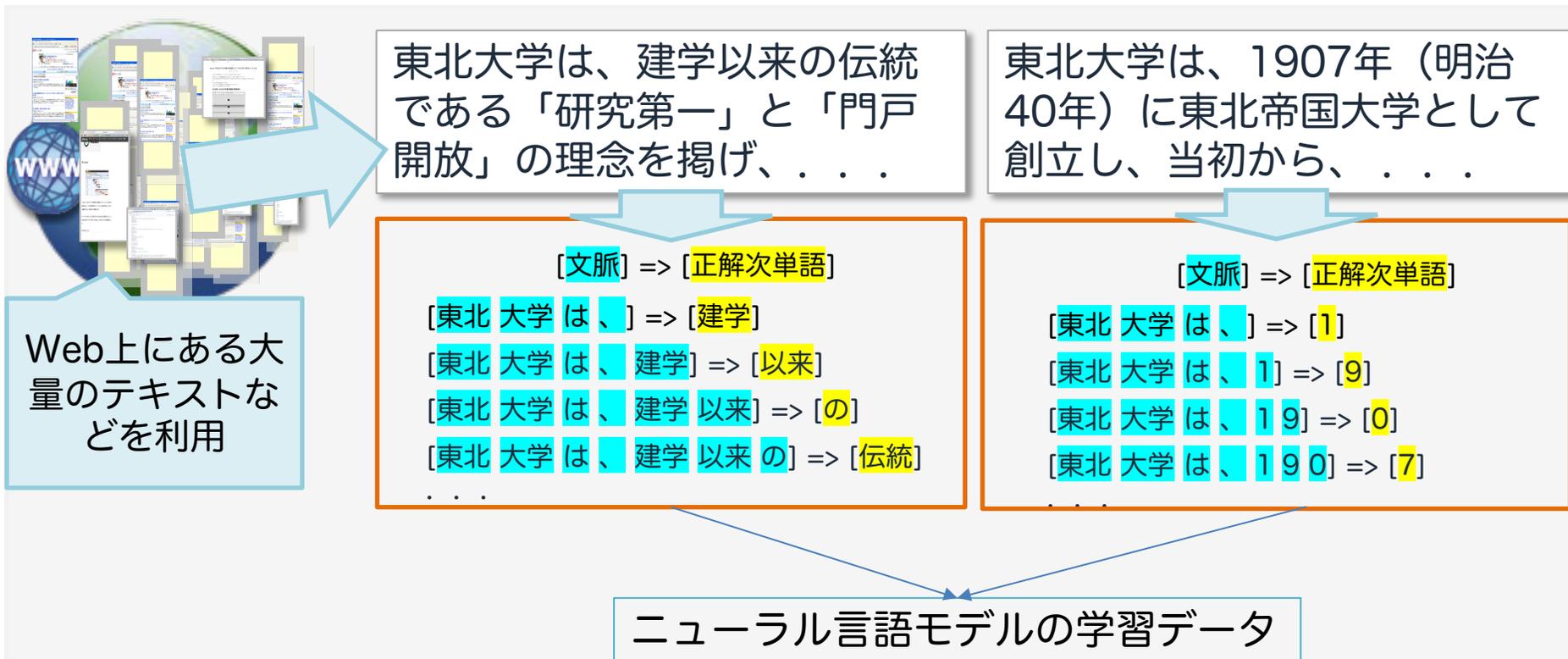
- 成功の要因

- ① 多層ニューラルネットワーク (DNN) の利用
- ② 大規模化 (パラメタ数/データ量)
- ③ 指示文設計 (プロンプトエンジニアリング)
- ④ 指示文 (+対話文) チューニング
- ⑤ 人手点数付け結果の活用



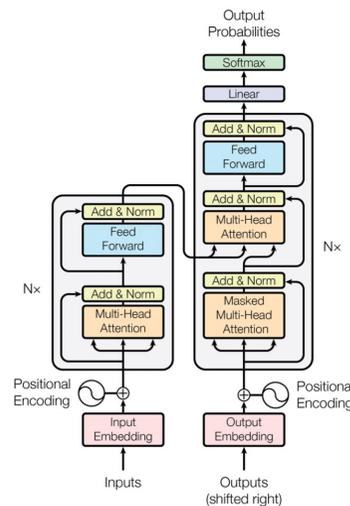
# ① DNNの利用：ニューラル言語モデル

- 言語モデルを多層ニューラルネットワークで構築
  - [文脈] => [正解次単語] を分類問題として学習
  - 文章中の**全単語を対象**として学習



# [参考] Transformer [本講演での詳細な説明は割愛]

- 深層ニューラルネットワークの種類の一つ
  - 2017年：元々は主に機械翻訳用として提案



詳細な解説はこちらで

IT Text 自然言語処理の基礎  
岡崎 直観, 荒瀬 由紀, 鈴木 潤,  
鶴岡 慶雅, 宮尾 祐介  
2022.08 発売



Transformerの  
章を執筆

[https://papers.nips.cc/paper\\_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html](https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html)

- 有名なニューラル言語モデルのほぼ**全て**で採用
  - 例：GPT-3 (ChatGPT/GPT-4), PaLM, LLaMA, OPT, BLOOM, ...
- 画像/音声/信号処理などでも広く普及
- ニューラルネットに基づくAI技術において**最重要技術**の一つ

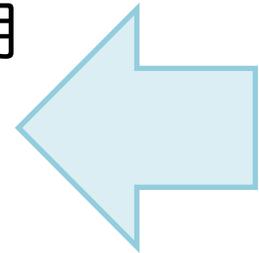
# ① DNNの利用：ニューラル言語モデル 効果

- ニューラル言語モデルの良い特性
  - 意味的に似ている単語や表現 (句) を捉えることが可能
  - 長い文脈を考慮可能

=> 学習データに**存在しない文脈**でも類似文脈から次単語を予測可能

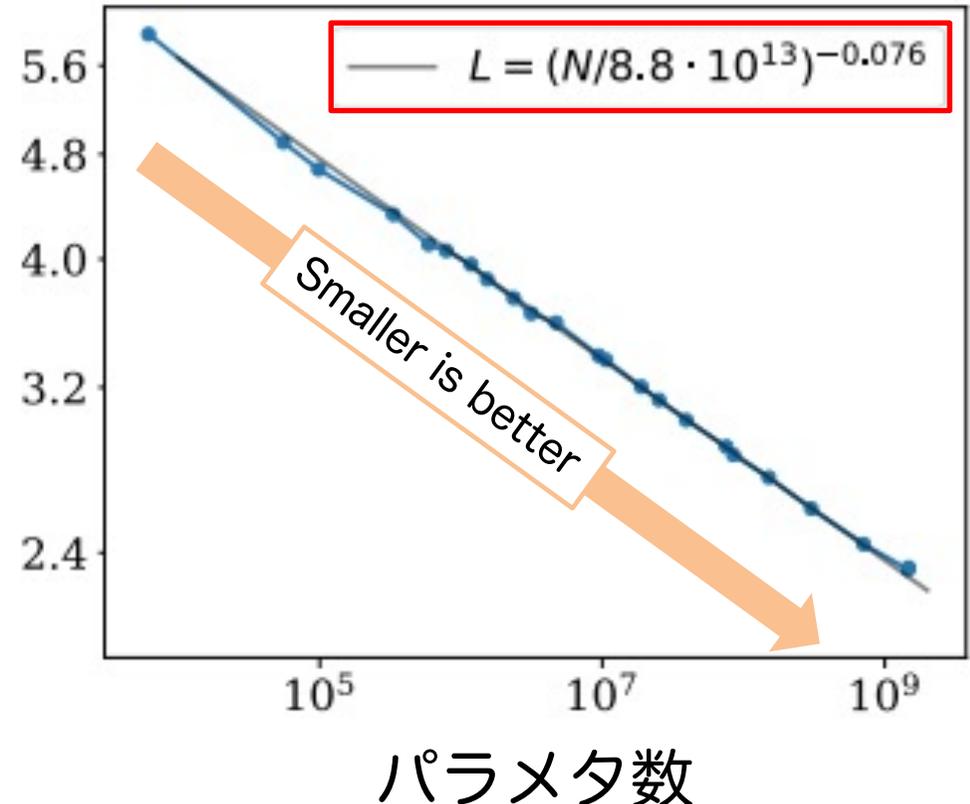
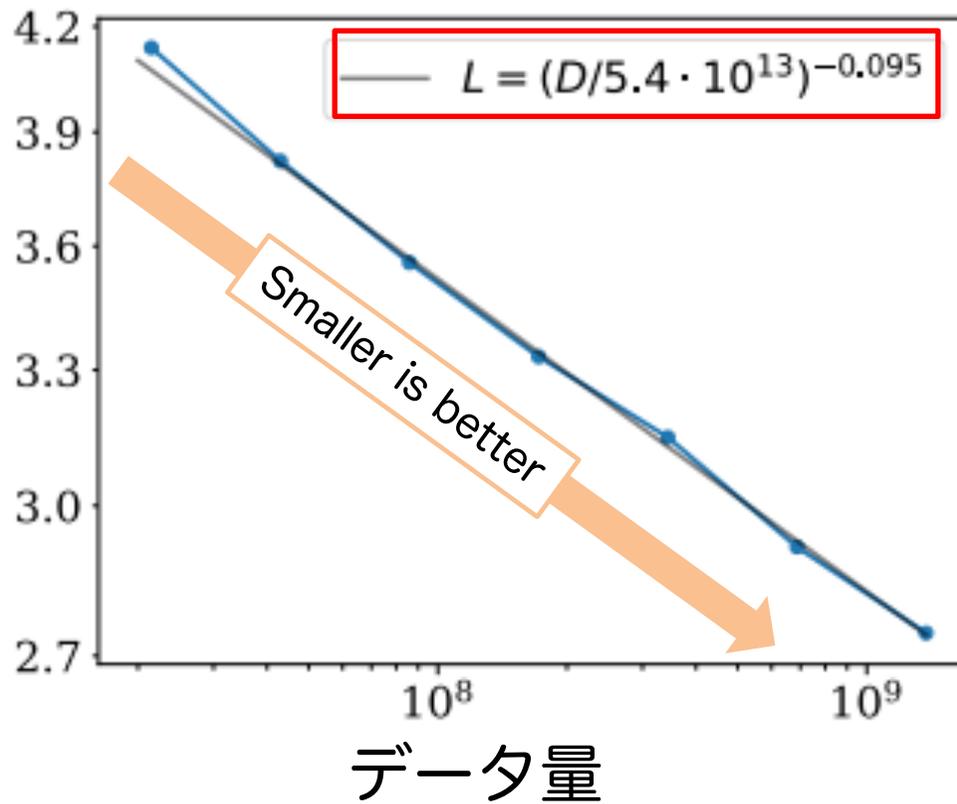
# ChatGPTを支える技術

- 基盤
  - 言語モデル
- 成功の要因
  - ① 多層ニューラルネットワーク (DNN) の利用
  - ② 大規模化 (パラメタ数/データ量)
  - ③ 指示文設計 (プロンプトエンジニアリング)
  - ④ 指示文 (+対話文) チューニング
  - ⑤ 人手点数付け結果の活用



## ② 大規模化 (データ量/パラメタ数)

- 学習データ量 / パラメタ数 が増えれば性能向上  
=> Scaling Laws

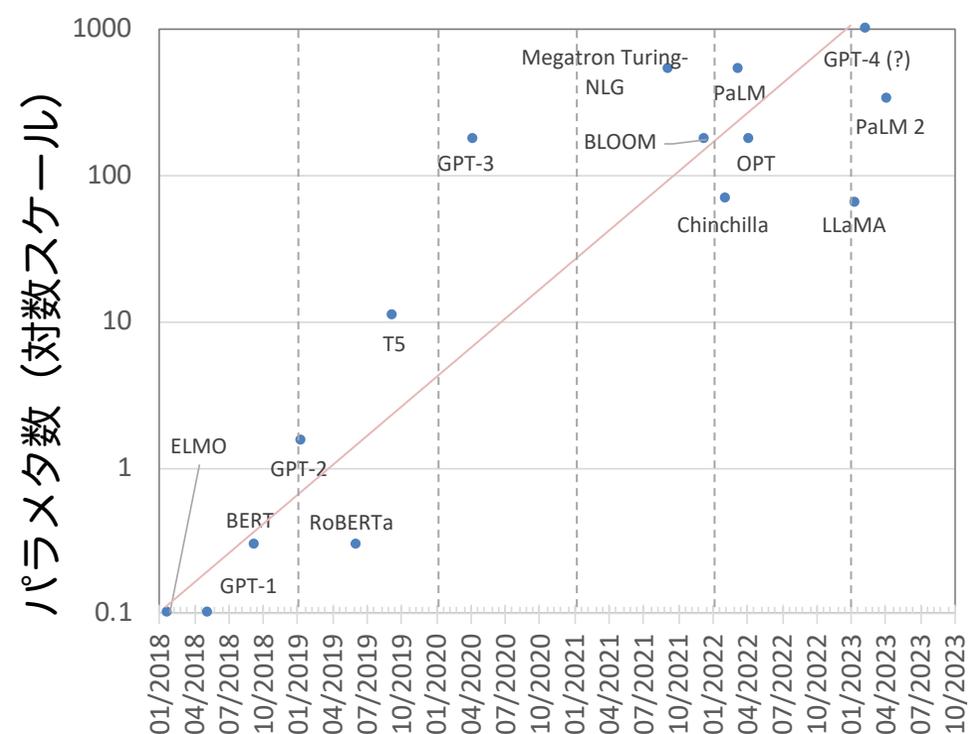
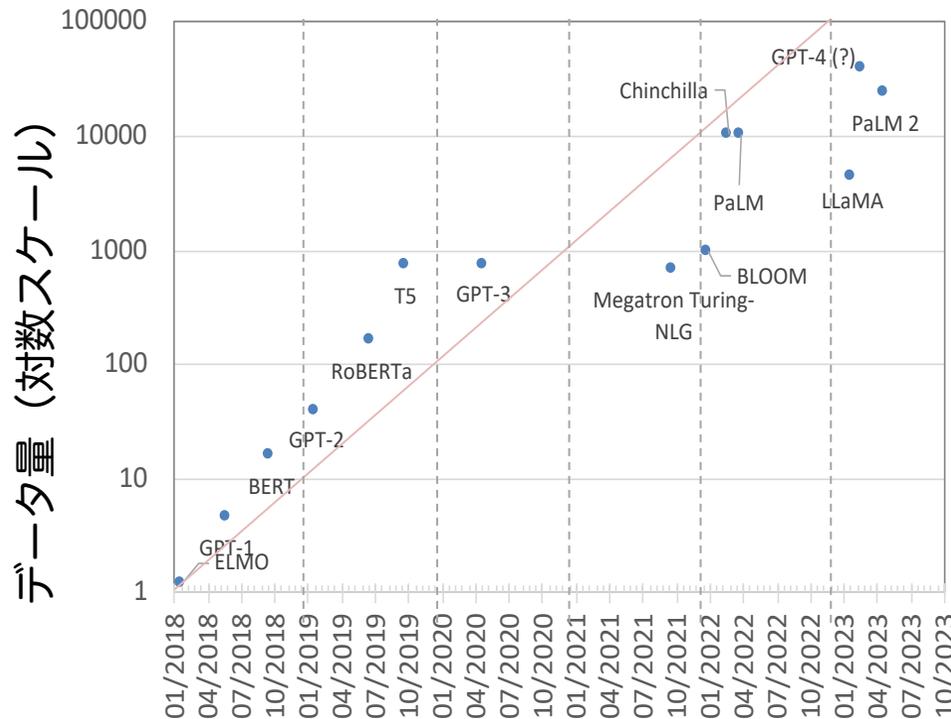


## ② 大規模化 (パラメタ数/データ量) 年毎の大規模化

### ● 学習データ量 / パラメタ数 : 年々順調に増加

|        | トークン数  | ファイル容量     |
|--------|--------|------------|
| GPT-2: | 0.01 B | 0.04 TB    |
| GPT-3: | 0.5 B  | 0.75 TB    |
| GPT-4: | 20 T   | 40 TB (憶測) |

|        | パラメタ数   | 必要メモリ量           |
|--------|---------|------------------|
| GPT-2: | 1.5 B   | 6 GB             |
| GPT-3: | 175 B   | 700 GB (ChatGPT) |
| GPT-4: | 1,000 B | 4,000 GB (憶測)    |



## ② 大規模化 (パラメタ数/データ量)

# 効果

- 人が作成する文章と同等レベルの流暢な文章を生成
  - 長文でも文章内の矛盾があまり発生しない

Title: United Methodists Agree to Historic Split

Subtitle: Those who oppose gay marriage will form their own denomination

Article: After two days of intense debate, the United Methodist Church has a one that is expected to end in the creation of a new denomination, one that will be "socially conservative," according to The Washington Post. The majority of delegates at the church's annual General Conference in May voted to strengthen a ban on the ordination of transgender clergy and to write new rules that will "discipline" clergy who officiate at same-sex weddings. But those who opposed these measures have a new plan: They say they will form a separate denomination by 2020, calling their church the Christian Methodist denomination.

The Post notes that the denomination, which claims 12.5 million members, was in the early 20th century the "largest Protestant denomination in the U.S.," but that it has been shrinking in recent decades. The new split will be the second in the church's history. The first occurred in 1968, when roughly 10 percent of the denomination left to form the Evangelical United Brethren Church. The Post notes that the proposed split "comes at a critical time for the church, which has been losing members for years," which has been "pushed toward the brink of a schism over the role of LGBTQ people in the church." Gay marriage is not the only issue that has divided the church. In 2016, the denomination was split over ordination of transgender clergy, with the North Pacific regional conference voting to ban them from serving as clergy, and the South Pacific regional conference voting to allow them.

言語モデルによる生成：与えられた文脈の次単語予測を繰り返して文章の続きを生成 (文章の補完)

Figure 3.14: The GPT-3 generated news article that humans had the greatest difficulty distinguishing from a human written article (accuracy: 12%).

<https://arxiv.org/abs/2005.14165>

② 大規模化 (パラメタ数/データ量)

# [推測] 流暢な文章を作れる要因

- 大量の文章を (DNN内で数値的に) **丸覚え** (しようと努力)

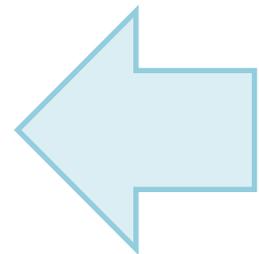


主にDNNの能力を活用

- 文章の特徴 (文, 句, 単語, 構文など) を数値の世界で学習
- 意味的な類似度, 模倣能力, 組合せ能力

# ChatGPTを支える技術

- 基盤
  - 言語モデル
- 成功の要因
  - ① 多層ニューラルネットワーク (DNN) の利用
  - ② 大規模化 (パラメタ数/データ量)
  - ③ 指示文設計 (プロンプトエンジニアリング)
  - ④ 指示文 (+対話文) チューニング
  - ⑤ 人手点数付け結果の活用



# ここまでの現状と課題

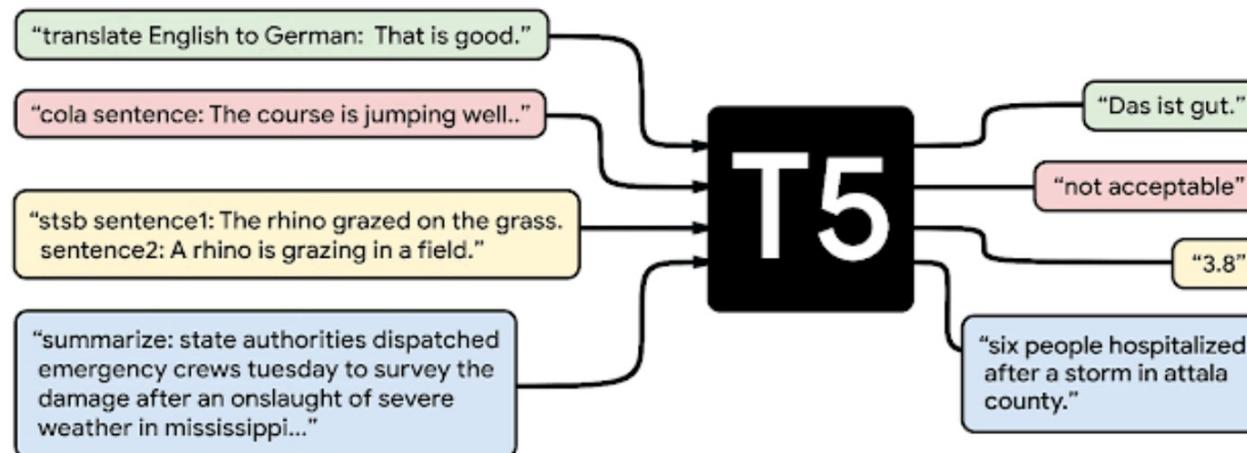
---

- 文章の補完はかなりのレベル
  
- (ここまでの) ChatGPTのように様々な要望に応じられる感じはしない. . .

# [参考] 言語処理タスクの解き方

- 従来：それぞれのタスクに特化したモデル/方法の構築
  - 特定のタスクのみ効果的に解ける
- 一つのモデルであらゆる言語処理タスクを解くことを目指す
  - 現在の対話型文章生成AIの先駆け (GPT-2, T5)

例：



Copied from <https://ai.googleblog.com/2020/02/exploring-transfer-learning-with-t5.html>

# 指示文 (プロンプト) とは？

文脈 = 指示文 (プロンプト) と考えてよい

- 言語モデル：与えられた文脈に基づいて文章の続きを生成 (文章の補完)
- 文脈を工夫すると様々な文章を生成することが判明
- なぜ特別な用語を用いるのか  
=> 単に「文脈」と呼ぶ以上の何かがありそう

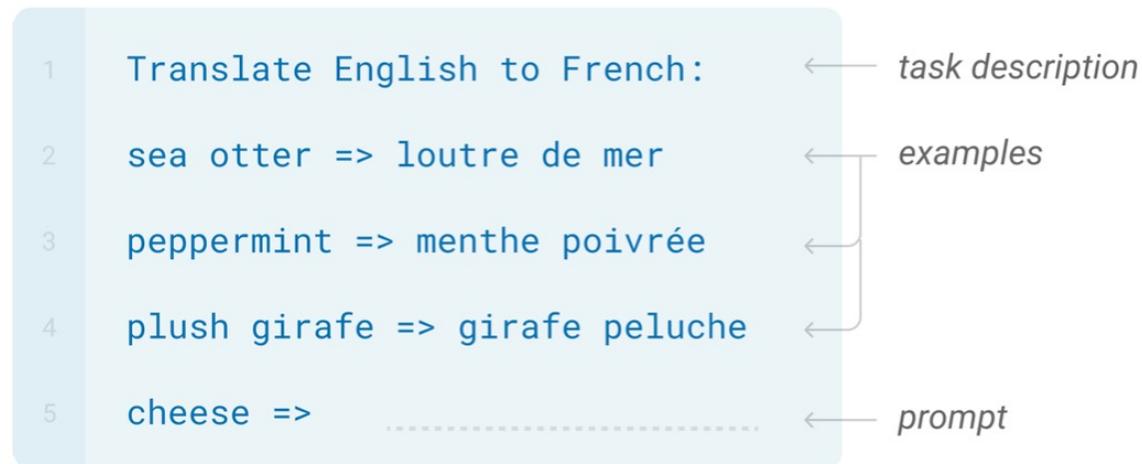
# ③ 指示文設計 (プロンプトエンジニアリング)

文脈 = 指示文 (プロンプト) と考えて良い

## ● 指示文 (プロンプト) 例

### Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.



### Chain-of-Thought Prompting

#### Model Input

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?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls.  $5 + 6 = 11$ . The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

#### Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had  $23 - 20 = 3$ . They bought 6 more apples, so they have  $3 + 6 = 9$ . The answer is 9. ✓

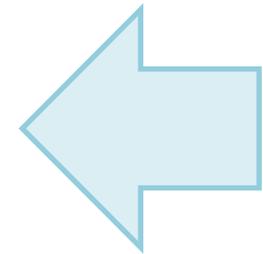
# ChatGPTを支える技術

- 基盤

- 言語モデル

- 成功の要因

- ① 多層ニューラルネットワーク (DNN) の利用
- ② 大規模化 (パラメタ数/データ量)
- ③ 指示文設計 (プロンプトエンジニアリング)
- ④ 指示文 (+対話文) による微調整学習
- ⑤ 人手点数付け結果の活用



# ここまでの現状と課題

- 文脈を工夫すると生成する文章を (ある程度) 制御可能
- 課題
  - うまくいかない場合も結構ある
  - うまくいくプロンプトを探すのはコスト大
  - うまくいくプロンプトを探す指針を見つけるのも困難

=> 試行錯誤が必要



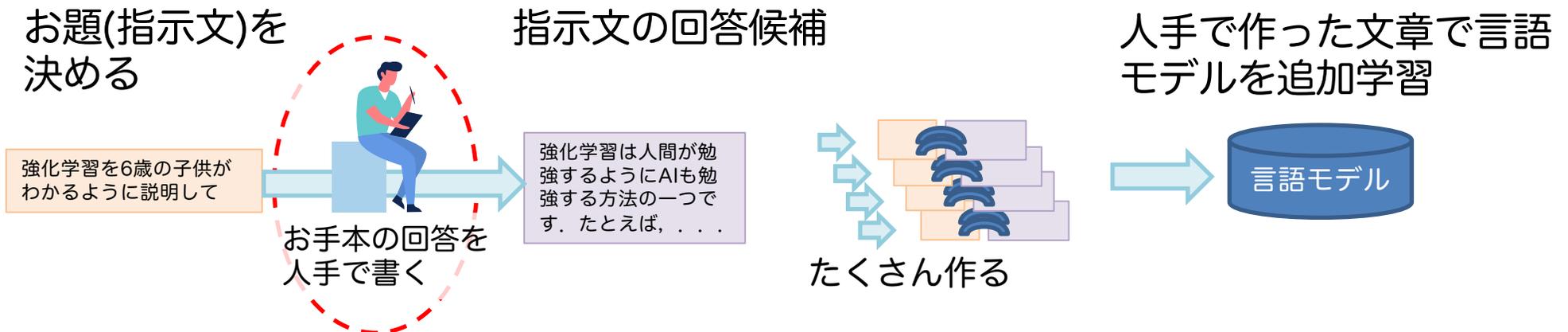
この課題への対応を考える

# ④ 指示文 (+対話文) による微調整学習

- 人が人に出す指示に近い文章で直接学習
  - 対話的な指示
  - 特に解き方などを教えることで適切度が向上

=> 人間が文章を作成

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



#### ④ 指示文 (+対話文) による微調整学習

# 効果

- 対話的な指示文 / より複雑な要望  
に対して流暢な文章を生成可能

=> 学習していない類似タスクにも対応

主にDNNの能力を活用

- 文章の特徴 (文, 句, 単語, 構文など) を数値の世界で学習
- 意味的な類似度, 模倣能力, 組合せ能力

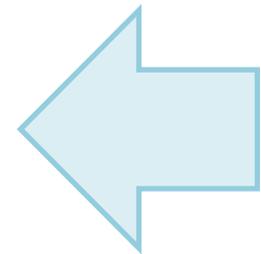
# ChatGPTを支える技術

- 基盤

- 言語モデル

- 成功の要因

- ① 多層ニューラルネットワーク (DNN) の利用
- ② 大規模化 (パラメタ数/データ量)
- ③ 指示文設計 (プロンプトエンジニアリング)
- ④ 指示文 (+対話文) による微調整学習
- ⑤ 人手点数付け結果の活用



## ここまでの現状と課題

- 対話的な指示文 / より複雑な要望  
に対して流暢な文章を生成可能
- 課題
  - 稀に不適切な文章を生成
    - 悪意があると捉えられる文章
    - 社会通念/社会問題的に不適切と考えられる文章
    - 不快に感じる文章, 特定の人物, 思想, 組織に偏った意見

=> 一般利用する上で一定以上の解決が必要



この課題への対応を考える

## ⑤ 人手点数付け結果の活用

# アイデア

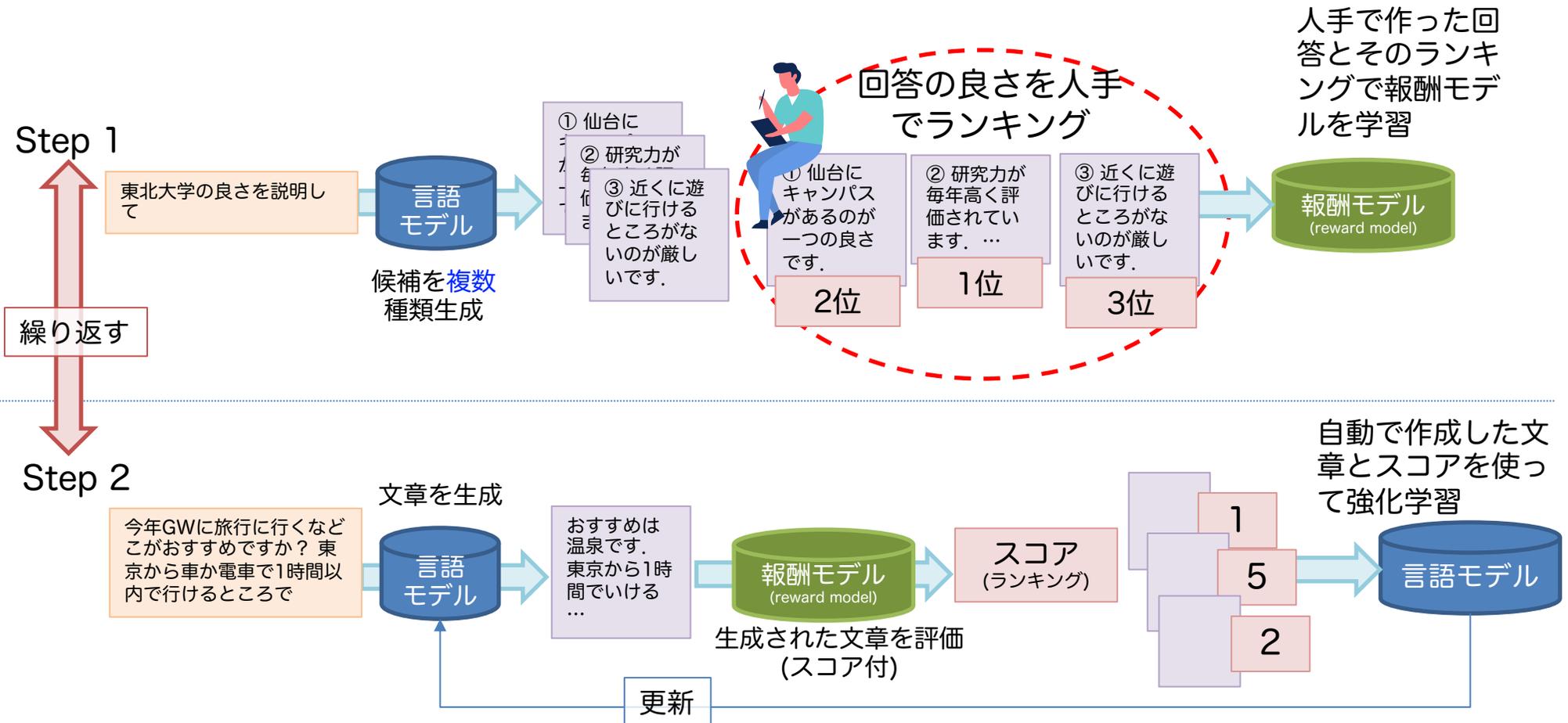
- 全ての不適切な文章生成を網羅して抑制するのは困難
  - 例 (前ページの再掲):
    - 悪意があると捉えられる文章
    - 社会通念/社会問題的に不適切と考えられる文章
    - 不快に感じる文章, 特定の人物, 思想, 組織に偏った意見

## => 人間からのフィードバックを活用

- 人間が生成文章を評価
- 複数の候補を生成
  - => 指示文に対してより適切な文章かを点数付け
- 人手点数付け結果データ
  - 指示文に対して人間が考える良し悪しが反映されたデータ

# ⑤ 人手点数付け結果の活用

- 人間が考える良し悪しを直接学習
  - 主に生成されたくない文章の生成を抑制



# ChatGPTを支える技術：まとめ図

## 2階部分 (指示文の活用/学習)

指示文設計  
指示文 (+対話文) 微調整学種  
人手点数付けの利用



獲得するスキル?

指示文の理解

対話的なやり  
とりの理解

不適切発言の  
抑制

## 1階部分 (言語モデルの学習)

ニューラルネット  
巨大言語モデル  
大規模データ  
(教師あり学習, 確率モデル)



文章の流暢さ  
世界の知識



# ChatGPT/GPT-4の到達点

- 流暢さは人間と同等かそれ以上
  - 不自然な間違いをほとんどしない
  - 長い生成文章中でも理論的破綻がかなり少なくい
- かなり広範かつ雑多な指示に適切に回答
  - 長文, 文章の構造化, 推論, 計算, 要約, 役割認識, 多言語
  - 論述など複数の回答があり得るような問題は一部人間レベル
  - (もちろんできないこともある)
- ただし, 人間の作った文章を頑張って模倣するという基本原理
  - 全人類がこれまで生成してきた文章
  - 人間が作った{指示, 模範回答}のデータ

# 文章生成AI：既知の課題

- 情報の正確性 (Hallucination)
  - 自信をもって間違ったことを言う
  - 現在情報の真偽を判断する機構がない
- バイアス (Bias)
  - 性別や人種など社会通念上の配慮が困難
  - 特定の個人や団体の利益になる応答が出るように調整可能
- プライバシー/コンプライアンス
  - 入力したデータは学習に利用される可能性 (opt-out可能)
- データ汚染 (Contamination)
  - 将来ChatGPTの出力がweb上に溢れかえると、それが間違っているにもかかわらず真実として信じられるようになる可能性

#### Does OpenAI train on my content to improve model performance?

For non-API consumer products like ChatGPT and DALL-E, we may use content such as prompts, responses, uploaded images, and generated images to improve our services. Please refer to this [article](#) to understand how this content may be used to improve model performance and how you can opt-out. You can request to opt out of having your content used to improve our services at any time by filling out this [form](#). This opt out will apply on a going-forward basis only.

Please note that for our API product, OpenAI will not use data submitted by customers via our API to train or improve our models, unless you explicitly decide to [share your data](#) with us for this purpose.

- まとめ

# まとめ (1/2)

- ChatGPT (対話型文章生成AI) とは？  
=> 対話形式の指示を受け付け  
その指示に適した文章を生成する文章生成器
- 技術
  - 基盤：言語モデル
  - ① 多層ニューラルネットワーク (DNN) の利用
  - ② 大規模化 (パラメタ数/データ量)
  - ③ 指示文設計 (プロンプトエンジニアリング)
  - ④ 指示文 (+対話文) チューニング
  - ⑤ 人手点数付け結果の活用

# まとめ (2/2)

- 到達点
  - 流暢さは人間と同等かそれ以上
  - かなり広範かつ雑多な指示に適切に回答
  - ただし，人間の作った文章を模倣するという基本原理
- 既知の課題
  - 情報の正確性
  - バイアス
  - プライバシー
  - データ汚染