

九州工学教育協会第18回シンポジウム
AIを理解し、AIを活用するには ～AI時代の工学教育～

第3次AIブームと大規模言語モデル

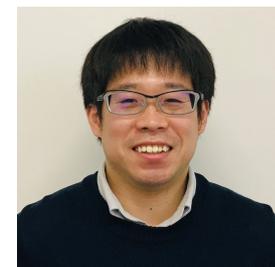
2025/07/08

愛媛大学 / 大阪大学 梶原 智之

kajiwara@cs.ehime-u.ac.jp

梶原 智之 (かじわら ともゆき)

- 愛媛大学 大学院理工学研究科 准教授
- 大阪大学 D3センター 特任准教授 (クロスアポイントメント)
- 専門：自然言語処理 (NLP; 言葉に関する人工知能の技術)
- <https://sites.google.com/site/moguranosenshi/>



経歴

- ~2018年：博士 (工学) @ 首都大学東京 (現在の東京都立大学)
 - 2015年：客員研究員 @ University of Liverpool
 - 2017年：協力研究員 @ 情報通信研究機構 (先進的翻訳技術研究室)
- ~2020年：特任助教 @ 大阪大学 (データビリティフロンティア機構)
- ~2023年：助教 @ 愛媛大学 (大学院理工学研究科)
- ~2024年：講師 @ 愛媛大学 (大学院理工学研究科)
- 2025年～：准教授 @ 愛媛大学 (大学院理工学研究科)
- 2025年～：特任准教授 @ 大阪大学 (D3センター)

• 第3次AIブームと自然言語処理

- 自然言語処理とは
- 深層学習による自然言語処理



• 大規模言語モデルとChatGPT

- 言語モデル
- 大規模言語モデル
- ChatGPT：大規模言語モデル × 指示チューニング × 強化学習
- 大規模言語モデルの特性

自然言語のコンピュータ処理に関する学問分野 (人工知能の一種)

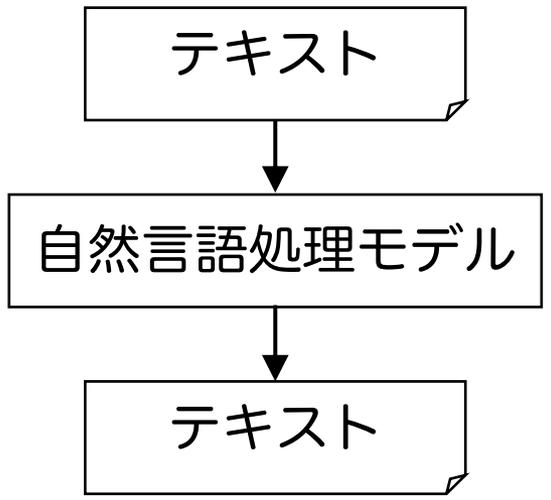
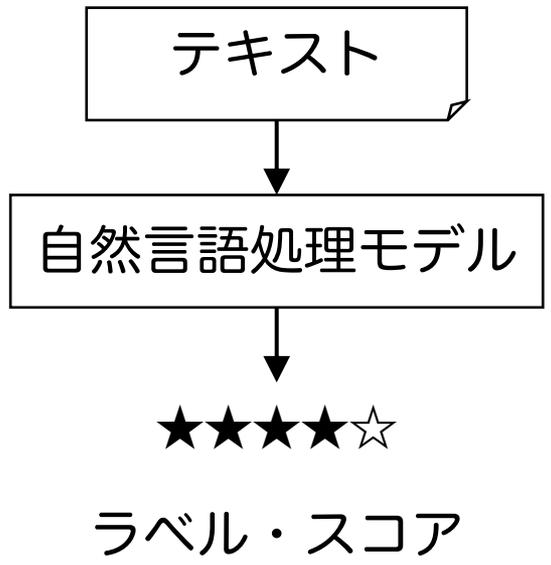
- 自然言語：日本語や英語などの自然発生的に生まれた言語
- 人工言語：PythonやC++などのプログラミング言語

1. テキストの分類 (言語理解)

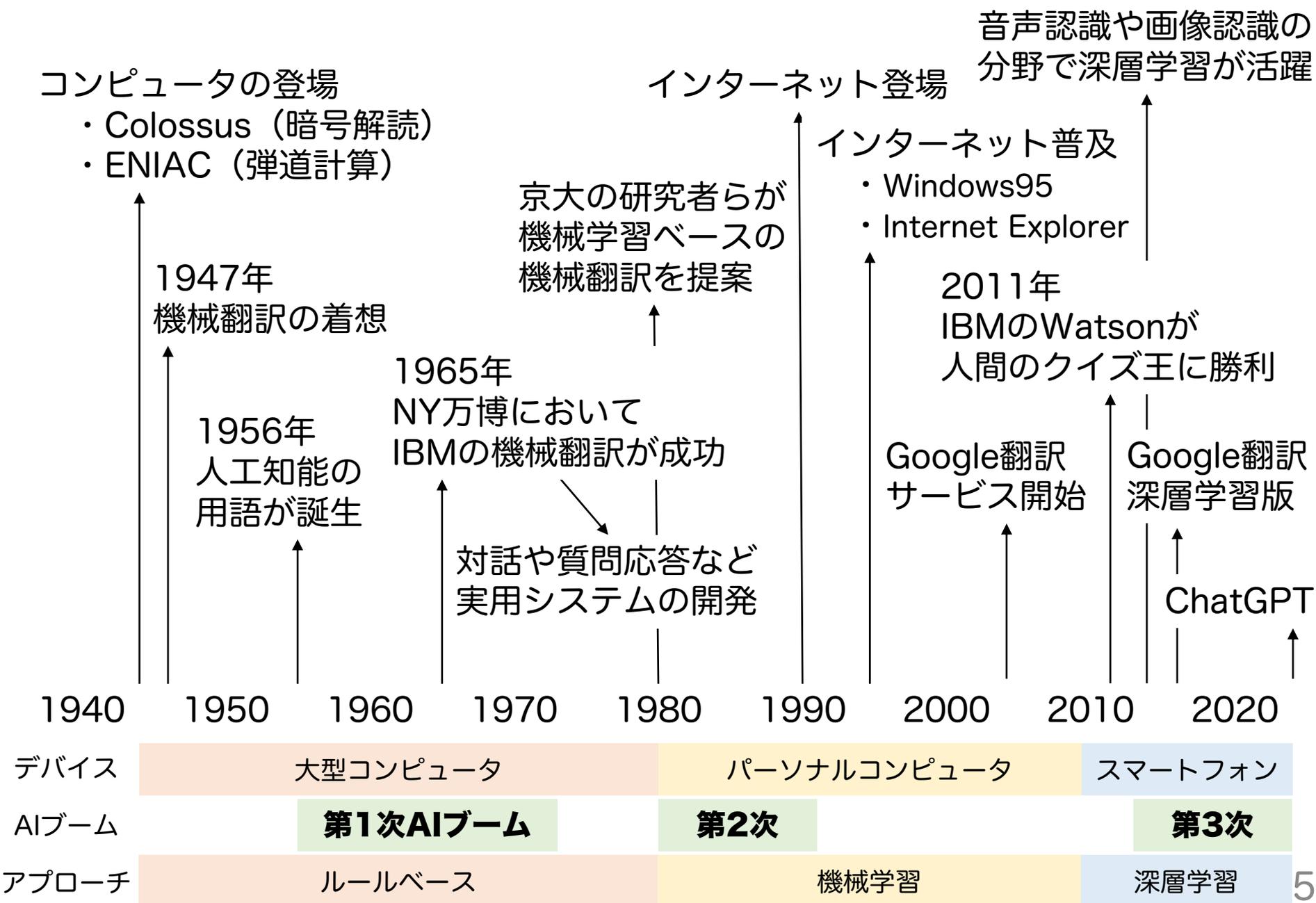
- 質問応答
- 感情分析
- 文法誤り検出

2. テキストの生成 (言語生成)

- 機械翻訳
- 自動要約
- 対話システム

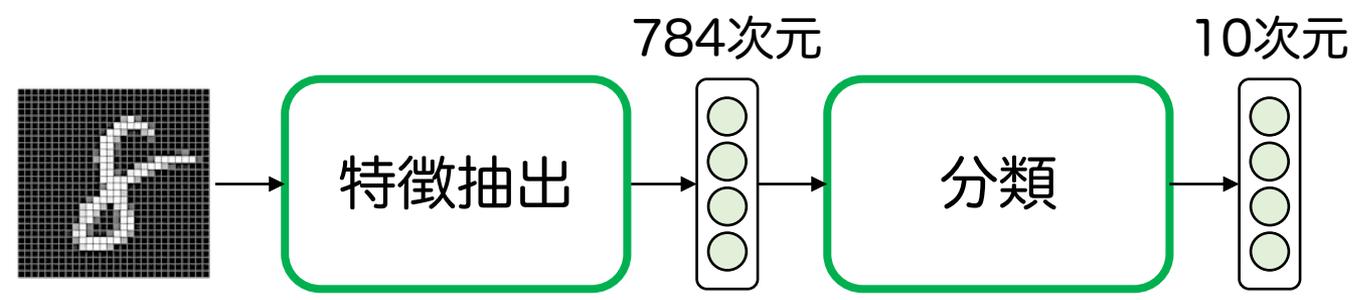
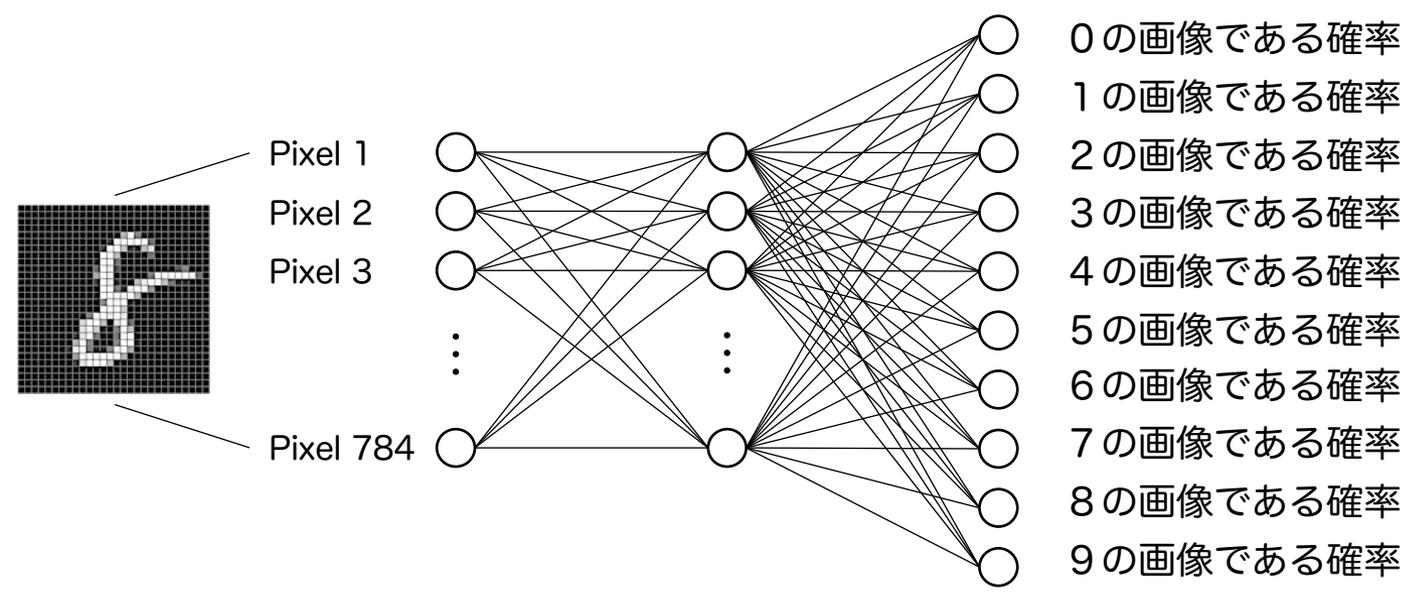


人工知能と自然言語処理の歴史



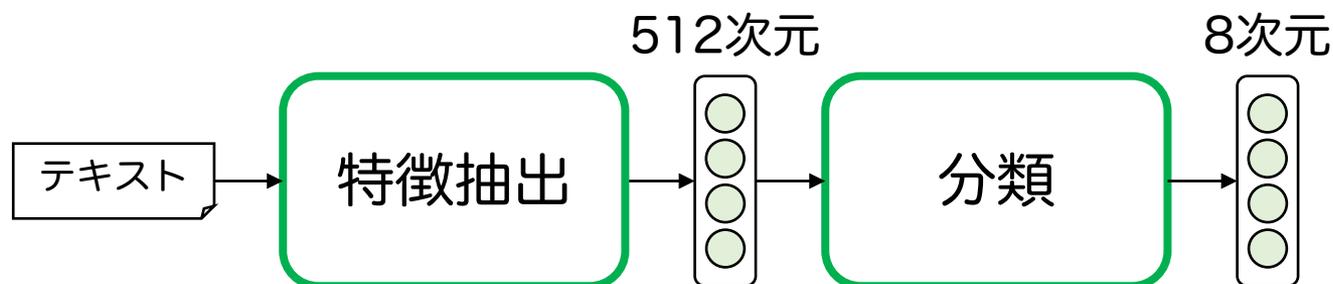
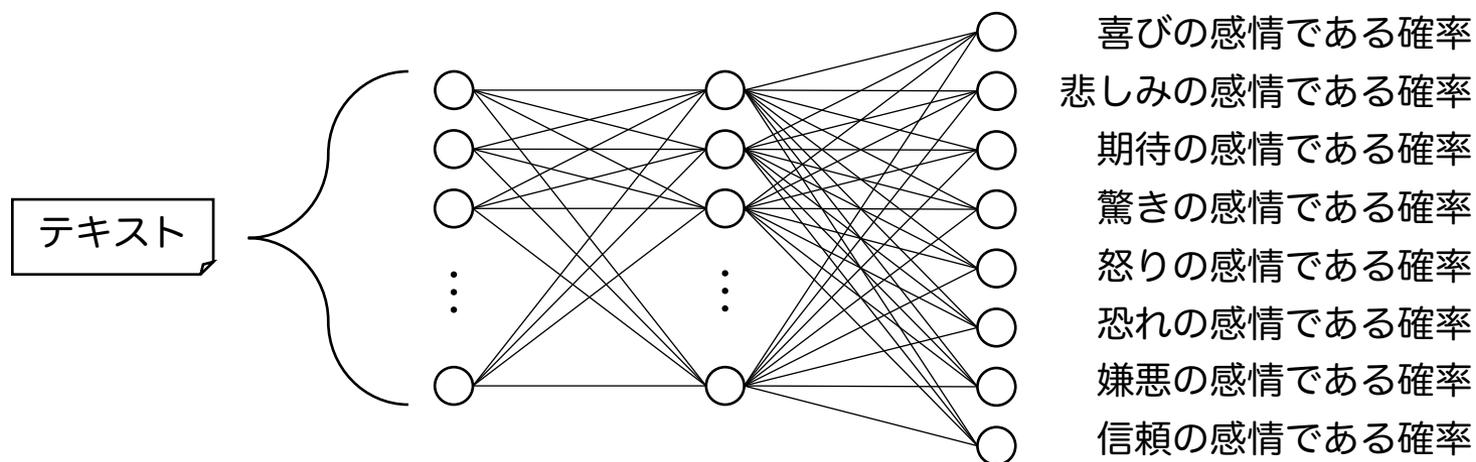
深層学習 (Deep Learning) 1/2

- ニューラルネットワークという道具を使い、ベクトルを変換する
- 例：28×28ピクセルの画像を、0~9の10種類に分類する



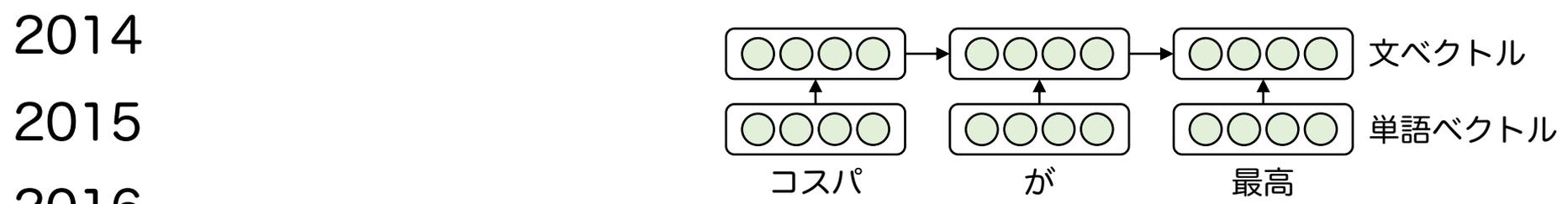
深層学習 (Deep Learning) 2/2

- ニューラルネットワークという道具を使い、ベクトルを変換する
- 例：テキストを、喜びや悲しみなど8種類の感情に分類する

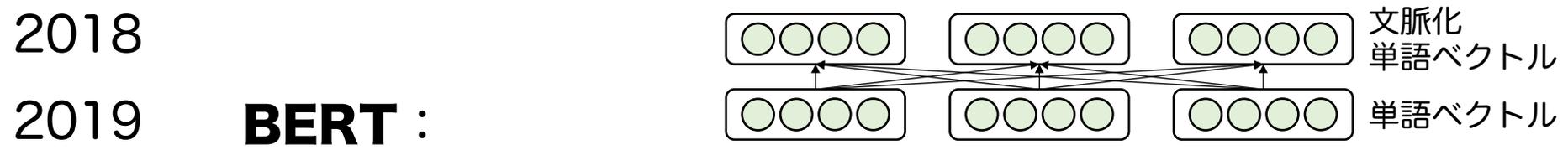


深層学習による自然言語処理：テキストをベクトルで表現する技術

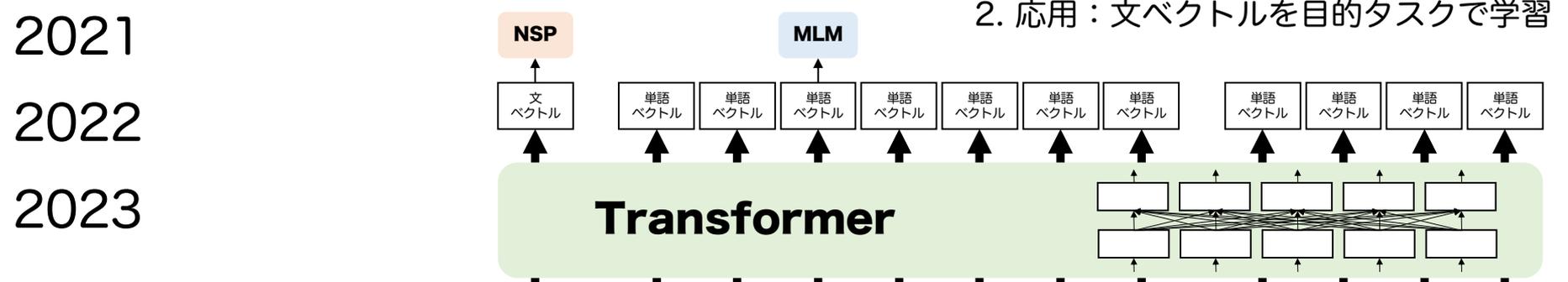
2013 **word2vec**：単語ごとにベクトルを割り当てる



2017 **Transformer**：効率的に文脈を考慮する深層学習手法

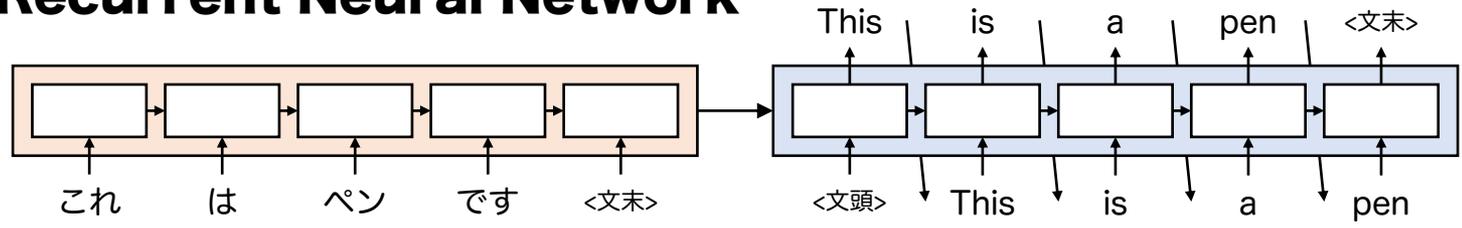


2020 基礎と応用の2段階学習

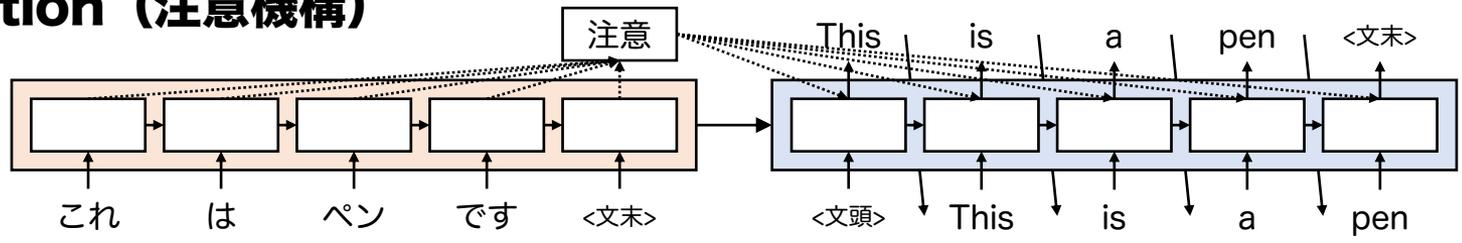


※ MLM: Masked Language Modeling
※ NSP: Next Sentence Prediction

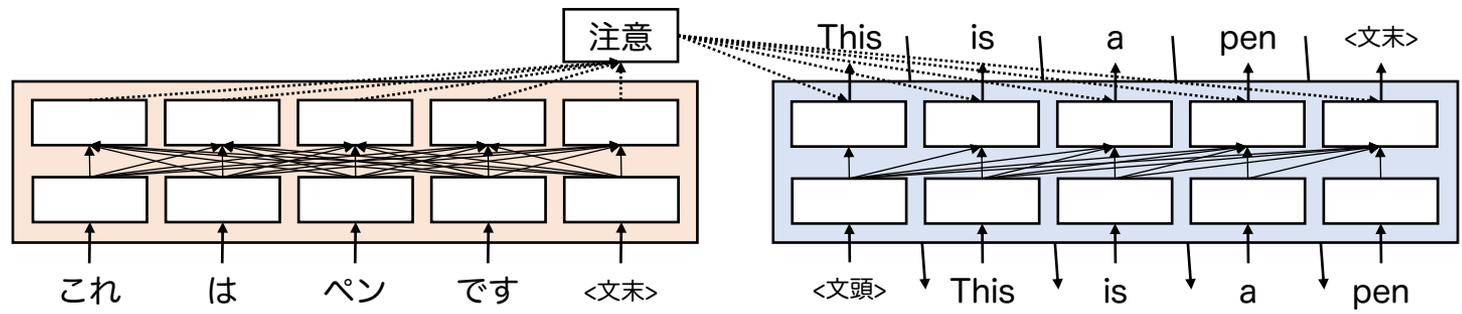
2013- RNN: Recurrent Neural Network



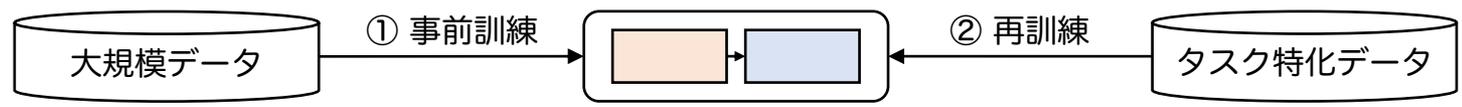
2015- Attention (注意機構)



2017- SAN: Self-Attention Network



2019- Transfer Learning (転移学習)



• 第3次AIブームと自然言語処理

- 自然言語処理とは
- 深層学習による自然言語処理



• 大規模言語モデルとChatGPT

- 言語モデル
- 大規模言語モデル
- ChatGPT：大規模言語モデル × 指示チューニング × 強化学習
- 大規模言語モデルの特性

- 途中まで文を読み、次の単語を推定する自然言語処理モデル
- 例：吾輩は猫である。名前は... → 次の単語は「まだ」

$$p(w_k | w_1, \dots, w_{k-1}) = \prod_{i=1}^k p(w_i | w_1, \dots, w_{i-1})$$

- $k - 1$ 番目までの単語列（文脈）から次の単語 w_k を推定
- 大規模データ（例えば3,000億単語）から学習すれば、人間のような流暢なテキスト生成が可能

- 第3次AIブーム以前から存在する伝統的な手法
- 次の単語の生成確率は、直前の $N - 1$ 単語のみに依存して決まる

$$p(w_k | w_1, \dots, w_{k-1}) = p(w_k | w_{k-N+1}, \dots, w_{k-1})$$

- $N=1$ の場合 $p(w_k | w_1, \dots, w_{k-1}) = p(w_k)$
- $N=2$ の場合 $p(w_k | w_1, \dots, w_{k-1}) = p(w_k | w_{k-1})$
- $N=3$ の場合 $p(w_k | w_1, \dots, w_{k-1}) = p(w_k | w_{k-2}, w_{k-1})$
- 吾輩は猫である。名前は ...
 - 1-gram言語モデルでは、「の」や「。」が候補に
 - 2-gram言語モデルでは、「美味しい」や「今」が候補に
 - 3-gram言語モデルでは、「鈴木」や「佐藤」が候補に
 - N が充分に大きくなると、「まだ」が候補になる

- **第3次AIブームと自然言語処理**

- 自然言語処理とは
- 深層学習による自然言語処理



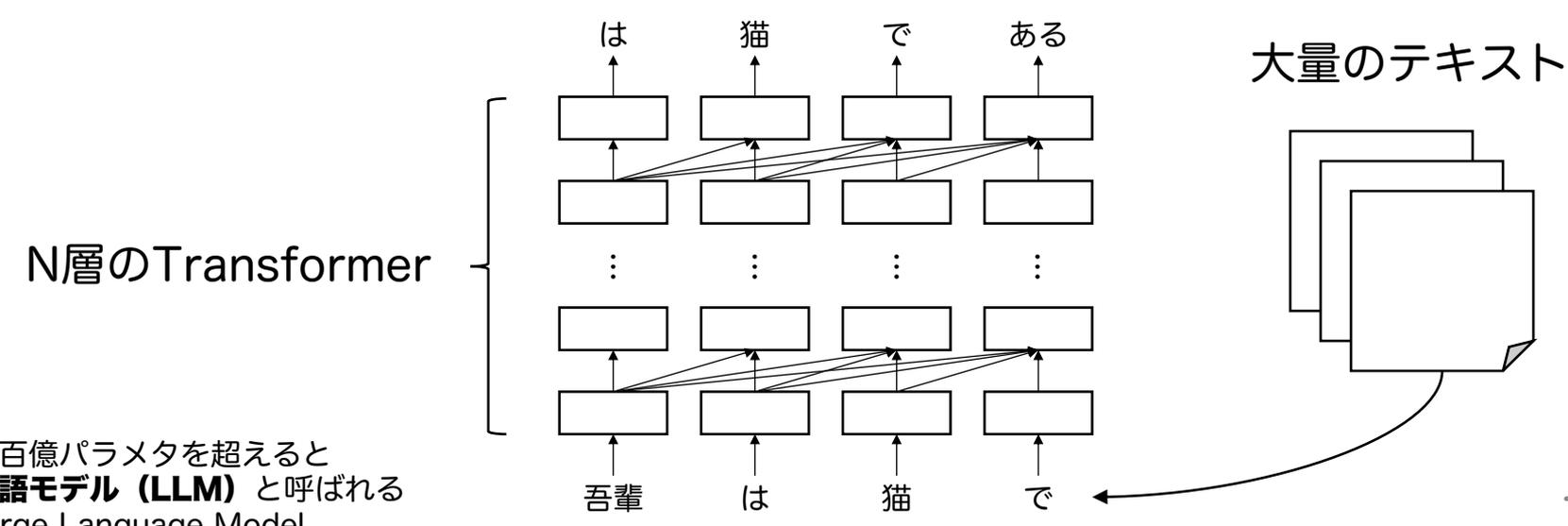
- **大規模言語モデルとChatGPT**

- 言語モデル
- **大規模言語モデル**
- ChatGPT：大規模言語モデル × 指示チューニング × 強化学習
- 大規模言語モデルの特性

ニューラル言語モデル (GPT: Generative Pre-trained Transformer)

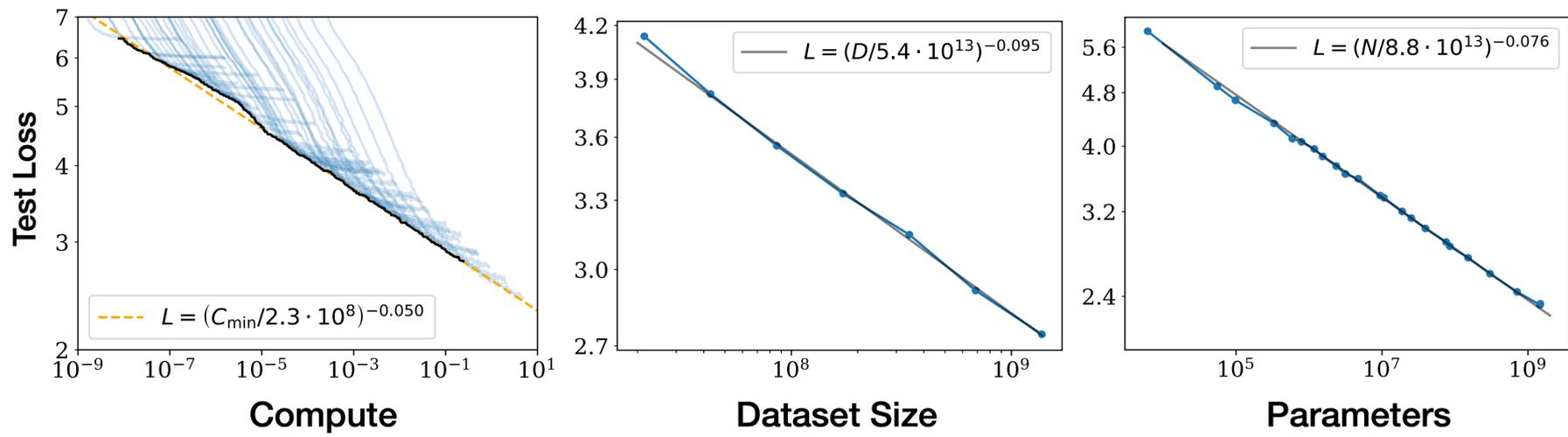
- ニューラルネットワークを用いて全履歴を考慮して次の単語を推定
- 特に、OpenAI社のTransformerベース言語モデルを“GPT”と呼ぶ
- モデルサイズと学習データ量を拡大し、性能改善を続けている

	発表時期	パラメタ数 (モデルサイズ)	学習データ量
GPT	2018年6月	1億 (12層Transformer)	4GB (7,000冊の書籍)
GPT-2	2019年2月	15億 (48層Transformer)	40GB (800万件のwebページ)
GPT-3	2020年6月	1,750億 (96層Transformer)	570GB
GPT-4	2023年3月	不明	不明



※ だいたい百億パラメタを超えると
大規模言語モデル (LLM) と呼ばれる
LLM: Large Language Model

ニューラル言語モデルのスケールング則



ニューラル言語モデルの性能は、以下の3つの要因で決まる

- 計算量
- データ量
- パラメタ量

これらを増やすほど、ニューラル言語モデルの性能を改善できる

※ 2026年問題：高品質な訓練データは2026年までに枯渇すると言われている

文脈内学習 (In-context Learning)

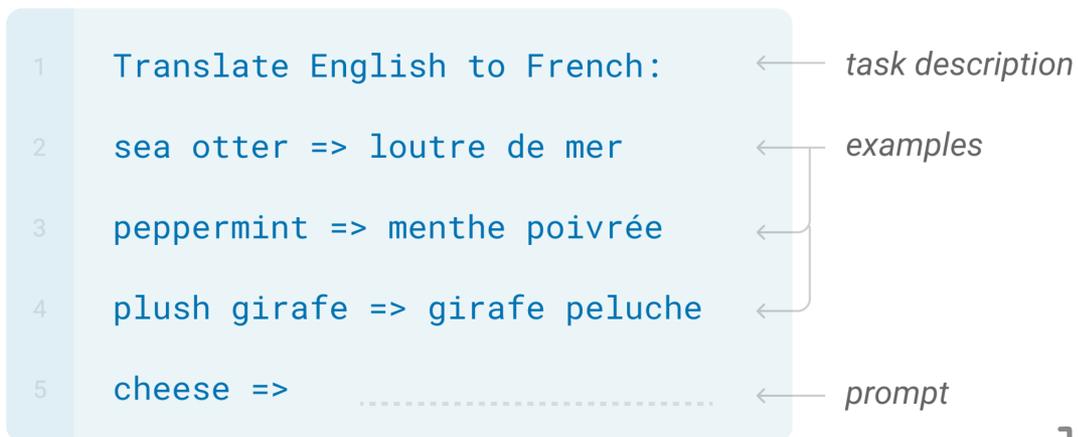
GPT-3 (1,750億パラメタのTransformer) を
目的タスクで再訓練するためには、1.2TBのGPUメモリが必要

- 深層学習向け計算サーバは通常、24~48GBのGPUメモリを搭載
- 大学等の普通の計算環境で再訓練できるのは GPT-2 ぐらいまで
- GPT-3以降の大規模言語モデルは通常環境では再訓練できない

文脈内学習 または **few-shot**学習

しかし、GPT-3は再訓練しなくても、**タスクの説明** と **数件の事例**
を入力するだけで、様々なタスクに (ある程度の性能で) 対応できる

- 機械翻訳
- 質問応答
- ニュースのタイトル
から記事本文の生成



※ モデルに与える指示のことをプロンプトと呼び、
文脈内学習のことをプロンプティングとも呼ぶ。

思考の連鎖 (Chain-of-Thought)

指示の最後に “**Let’s think step by step.**” と付けるだけで、難しい算術推論などの正答率が上がる

普通の指示

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A: **8**

思考の連鎖

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

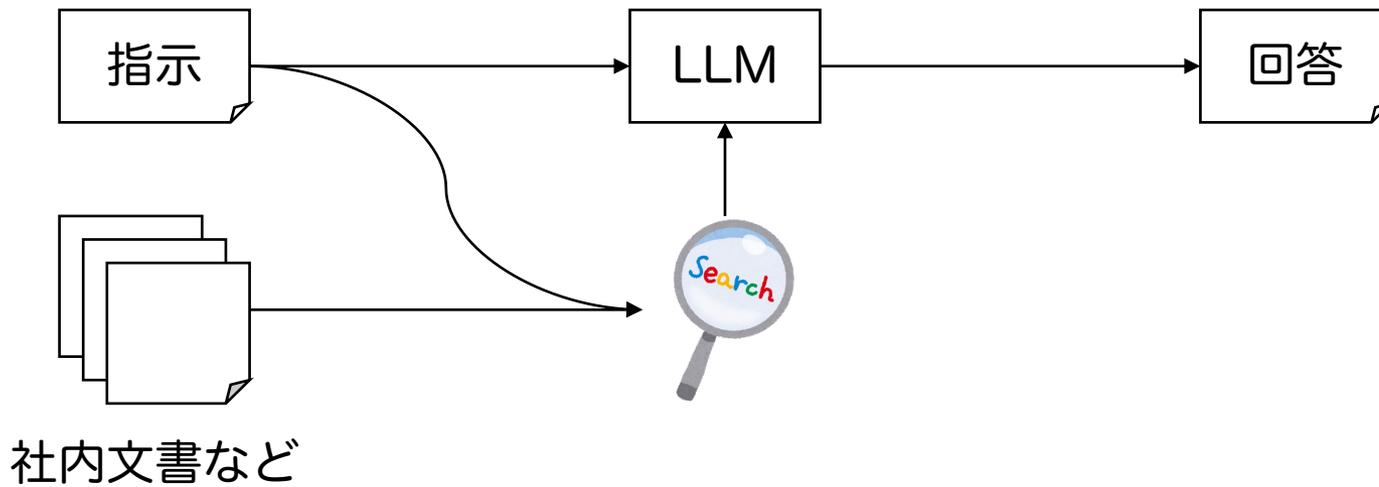
A: *Let’s think step by step.*

There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls.

追加の指示	正解率
なし	17.7
AbraKadabra!	15.5
Don’t think. Just feel.	18.8
Let’s think.	57.5
Let’s think step by step.	78.7

検索拡張生成 (Retrieval Augmented Generation)

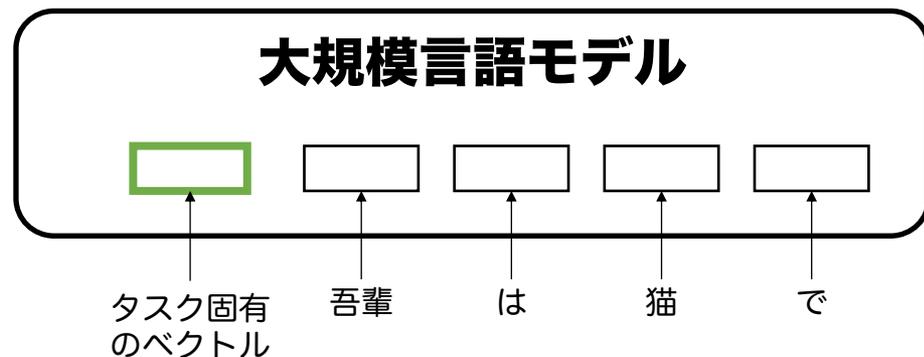
- 言語モデルが知らないことを教える
- 言語モデルの誤った知識を修正する



入力に関連する事例を検索して指示に含めると、
few-shot学習の性能を更に高めることもできる

Prefix Tuning

- 入力の先頭にタスク固有のベクトルを追加
- 追加した部分だけを学習

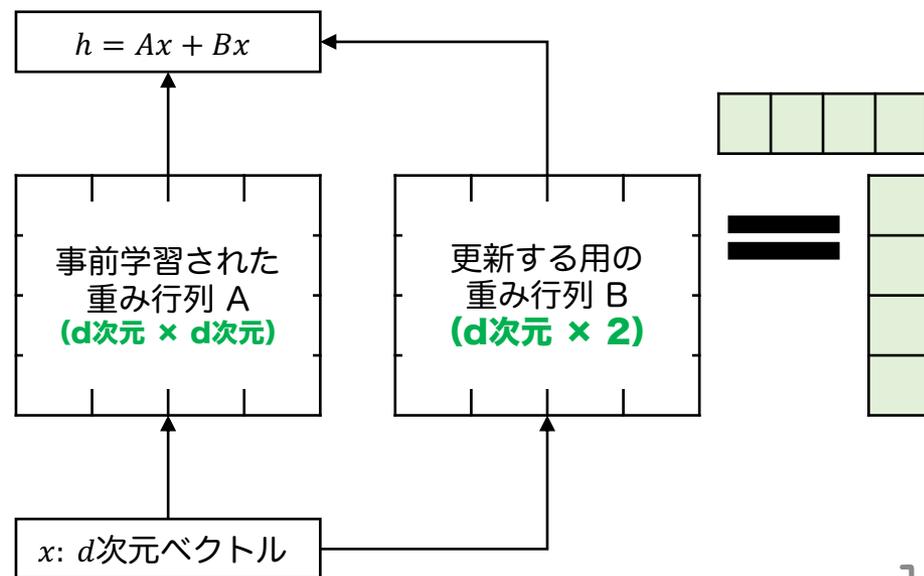


※ 緑色の追加パラメタのみを学習
その他のパラメタは更新しない

LoRA (Low-Rank Adaptation)

- 低ランク近似した重みパラメタ行列 (右図のB) を追加
- 追加した部分だけを学習

量子化によって、さらに各パラメタのメモリ使用量を削減する**QLoRA**も
※ 16ビット → 4ビット など



- **第3次AIブームと自然言語処理**

- 自然言語処理とは
- 深層学習による自然言語処理



- **大規模言語モデルとChatGPT**

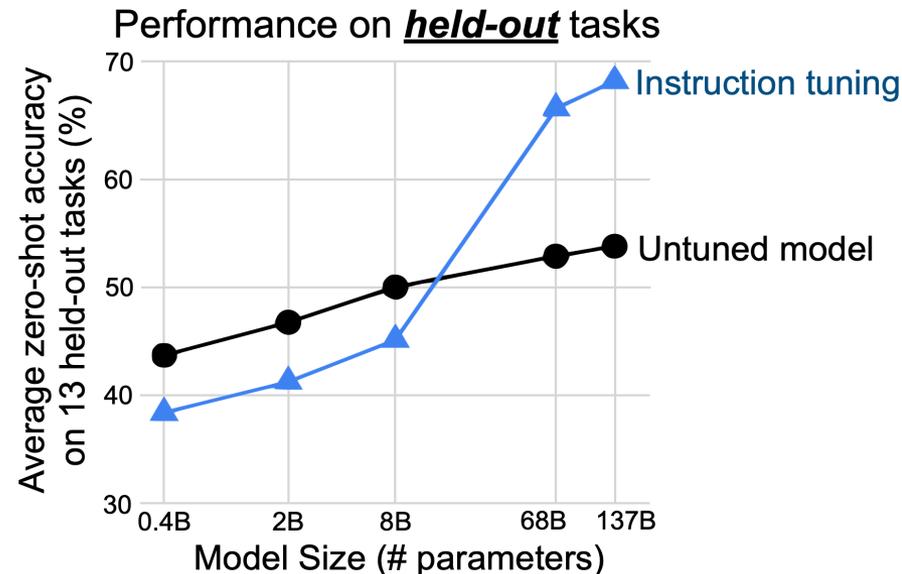
- 言語モデル
- 大規模言語モデル
- **ChatGPT：大規模言語モデル × 指示チューニング × 強化学習**
- 大規模言語モデルの特性

教師あり学習によるLLMの改善 (Instruction Tuning)

人間が自然に与える指示（タスク説明+数件の事例の提示）と言語モデルの性能を引き出す最適な指示は、必ずしも一致しない

FLAN : Fine-tuned Language Net

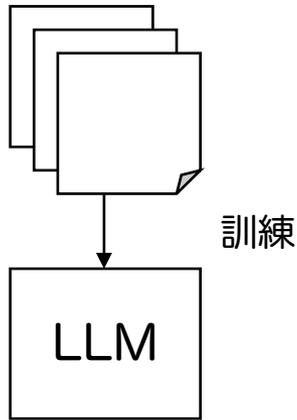
- BERTやGPT-2で採用されている**転移学習**のアプローチとGPT-3で採用されている**文脈内学習**のアプローチの融合
- 62種類のNLPデータセットを用いて、言語モデルが人間の**指示**に応じて適切な回答を行うように**再訓練**する
- GPT-3を上回る性能を達成
- 数十億パラメタ規模の言語モデルはInstruction Tuningによって未知タスクにおける性能が低下
- GPT-3など数百億パラメタ以上の大規模言語モデルに対して有効



※ 普通の計算機環境ではGPT-3を再訓練できないが、OpenAIやGoogleなどの一部の大企業の環境でなら、GPT-3の再訓練も可能

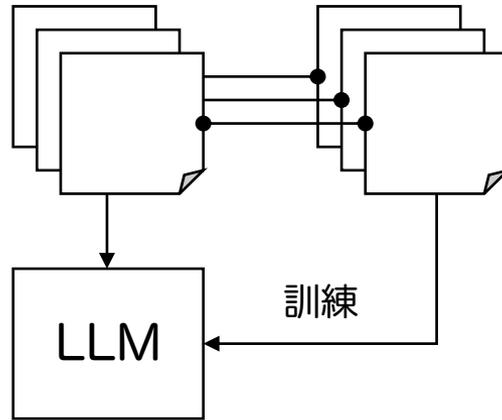
教師なし学習

人間が正解を与えず、データの傾向を学習



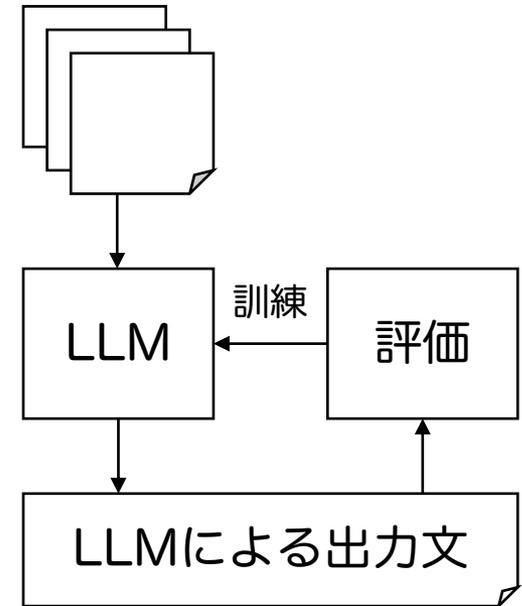
教師あり学習

人間が正解を与え、入出力の関係を学習



強化学習

人間が出力を評価し、出力の良／悪を学習

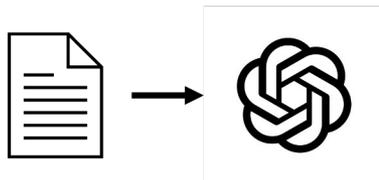


実際には、人間が大量に評価するのはコストが高いため、人間の評価を模倣するモデルを訓練し、自動評価で代替

InstructGPT：強化学習（RLHF）によって再訓練されたGPT
→ 指示の無視や差別的な発言が減り、人間の好む振る舞いに近づく

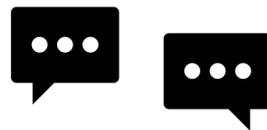
- 2022年11月末に公開されたGPTベースの対話システム
- 大きく3つの手順で訓練
 1. **教師なし学習** (Language Modeling)
 - Transformerモデルを用いた自己回帰型言語モデル (GPT-3.5)
 - GPT-3の時点でパラメタ数は1,750億、学習データ量は570GB
 2. **教師あり学習** (Instruction Tuning)
 - 人間が大量の指示-回答ペアを作成
 - ステップ1の言語モデルが指示に応じて適切に回答するよう再訓練
 3. **強化学習** (Reinforcement Learning from Human Feedback)
 - 人間の評価に基づき報酬モデルを訓練
 - ステップ2の言語モデルに報酬を与えて再訓練

Step 1



大規模テキスト
で**教師なし学習**

Step 2



指示と回答のペアを
用いた**教師あり学習**

Step 3



人間の意図に沿い、安全な
発言をするように**強化学習**

- **第3次AIブームと自然言語処理**

- 自然言語処理とは
- 深層学習による自然言語処理



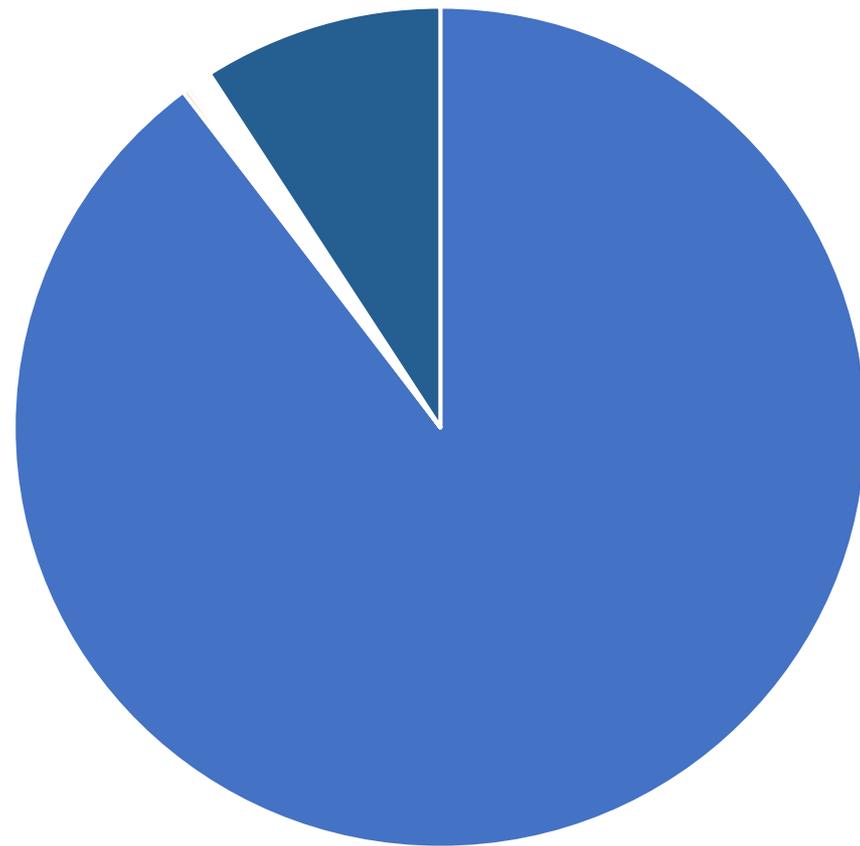
- **大規模言語モデルとChatGPT**

- 言語モデル
- 大規模言語モデル
- ChatGPT：大規模言語モデル × 指示チューニング × 強化学習
- **大規模言語モデルの特性**

大規模言語モデルの学習データ

9割が英語で、その他の言語は0.2%未満ずつ

言語	割合
英語	89.70%
ドイツ語	0.17%
フランス語	0.16%
スウェーデン語	0.15%
中国語	0.13%
スペイン語	0.13%
ロシア語	0.13%
オランダ語	0.12%
イタリア語	0.11%
日本語	0.10%
その他	9.10%

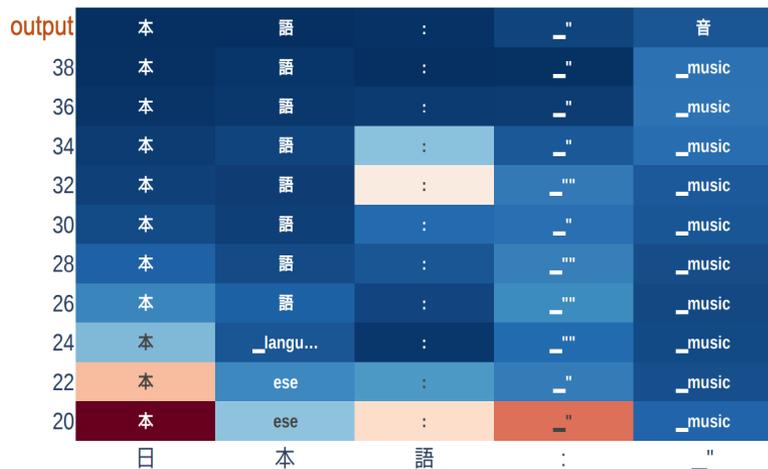


- 英語
- ドイツ語
- フランス語
- スウェーデン語
- 中国語
- スペイン語
- ロシア語
- オランダ語
- イタリア語
- 日本語
- その他

大規模言語モデルは内部では英語を使って考える

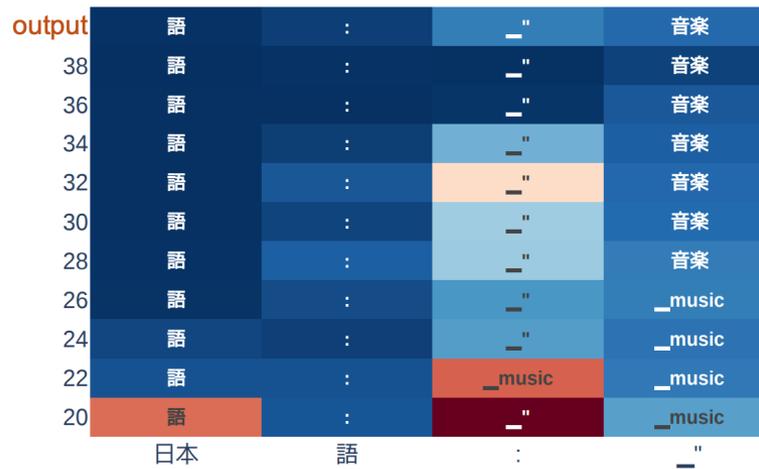
例：「音楽」をフランス語から日本語へ翻訳してください
"Français: 'musique' - 日本語: "

英語LLMのLLaMA



最終層だけ日本語を使う
直前までは英語で考える

日本語LLMのSwallow



途中から日本語を使う
途中まで英語で考える

ChatGPTによる「自然言語処理の歴史」の説明

KA

3回のAIブームと関連付けながら自然言語処理の歴史について教えてください



自然言語処理（Natural Language Processing, NLP）は、人工知能の一分野であり、コンピュータによる人間の言語の理解や生成を目的としています。以下に、NLPの歴史を3回のAIブームと関連付けて説明します。



1. 最初のAIブーム（1950年代-1970年代）

最初のAIブームでは、NLPに関する最初の研究が行われ、文法規則をベースにした自然言語生成システムが開発されました。この時代の代表的なシステムには、Joseph Weizenbaumによって開発されたELIZAがあります。ELIZAは、ユーザーとの会話をシミュレートするプログラムであり、人間が話しているように見える応答を生成することができました。

2. 第二次AIブーム（1980年代-1990年代）

第二次AIブームでは、統計的自然言語処理が注目され、大規模な言語コーパスを利用した自然言語処理が進歩しました。代表的なアルゴリズムには、隠れマルコフモデルや条件付き確率場があります。この時期には、情報抽出、テキスト分類、機械翻訳など、NLPのさまざまなアプリケーションが開発されました。

3. 現在のAIブーム（2010年代以降）

現在のAIブームでは、深層学習に基づく自然言語処理が急速に発展し、言語処理の精度が大幅に向上しました。深層学習による代表的なアルゴリズムには、畳み込みニューラルネットワーク、再帰型ニューラルネットワーク、Transformerがあります。この時期には、機械翻訳やテキスト生成のような高度なNLPアプリケーションが開発され、AIによる自然言語処理の適用範囲が拡大しています。

ChatGPTによる感情分析

「車のタイヤがパンクしてた。。
いたずらの可能性が高いんだって。。」

- 喜び : 0
- 期待 : 1
- 怒り : 2
- 不安 : 1
- 信頼 : 0
- 驚き : 2
- 悲しみ : 0
- 嫌悪 : 0

有り得そうな感情強度の組であるし

解説にも納得感がある

→ 人間の感情について理解してそう

KA

次の文について、Plutchikの基本8感情における書き手の各感情の強度を4段階（0-3）で推定してください。
車のタイヤがパンクしてた。。いたずらの可能性が高いんだって。。



この文の書き手の各感情の強度を推定するには、文脈に基づいてPlutchikの基本8感情を考慮する必要があります。以下に、推定される各感情の強度を示します（0から3の範囲で）：

1. 喜び (Joy) : 0

この文では、車のタイヤがパンクしているという問題が発生しています。そのため、喜びを感じることは少ないでしょう。

2. 期待 (Anticipation) : 1

「いたずらの可能性が高いんだって」というフレーズから、書き手はいたずらによるタイヤのパンクを予測している可能性があります。それによって、少しの期待を感じているかもしれません。

3. 怒り (Anger) : 2

タイヤのパンクは一般的にイライラや怒りを引き起こすことがあります。したがって、書き手は怒りの感情を強く感じている可能性があります。

4. 不安 (Fear) : 1

いたずらによるタイヤのパンクという可能性が示唆されているため、書き手は少し不安を感じているかもしれません。

5. 信頼 (Trust) : 0

文中には他の人物や信頼に関連する要素が明示されていないため、書き手の信頼感は推定できません。

6. 驚き (Surprise) : 2

タイヤのパンクは予期せぬ出来事であり、驚きを引き起こす可能性があります。書き手は少し驚きを感じているかもしれません。

7. 悲しみ (Sadness) : 0

この文では明確な悲しみの要素は示されていません。

8. 嫌悪 (Disgust) : 0

書き手の嫌悪感に関する情報は提供されていないため、推定できません。

したがって、推定される各感情の強度は次のようになります：喜び (0)、期待 (1)、怒り (2)、不安 (1)、信頼 (0)、驚き (2)、悲しみ (0)、嫌悪 (0)。

ChatGPTによるプログラミング

- FizzBuzz問題の模範解答を
- 実行結果は正しく、解説も

KA

1から30までの整数を順番に表示するPythonコードを教えてください。また、3の倍数のときには数字の代わりにfizzという文字、5の倍数のときにはbuzzという文字を表示してください。



以下は、1から30までの整数を順番に表示するPythonコードです。3の倍数のときは"fizz"、5の倍数のときは"buzz"を表示します。

python

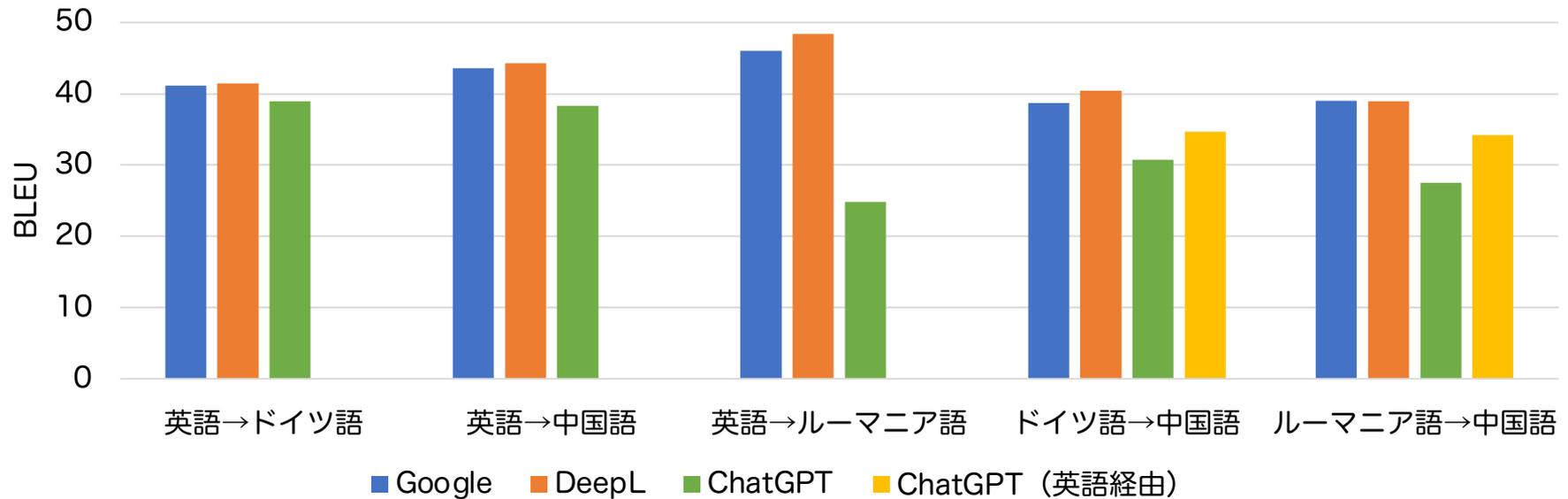
```
for i in range(1, 31):  
    if i % 3 == 0:  
        print("fizz")  
    elif i % 5 == 0:  
        print("buzz")  
    else:  
        print(i)
```

このコードを実行すると、以下のような出力が得られます:

```
1  
2  
fizz  
4  
buzz  
fizz  
7  
8  
fizz  
buzz  
11  
fizz  
13  
14  
fizzbuzz  
16  
17  
fizz  
19  
buzz  
fizz  
22  
23  
fizz  
buzz  
26  
fizz  
28  
29  
fizzbuzz
```

ここで、15の倍数（つまり3と5の両方の倍数）の場合は、"fizzbuzz"と表示されます。

ChatGPTの性能：機械翻訳

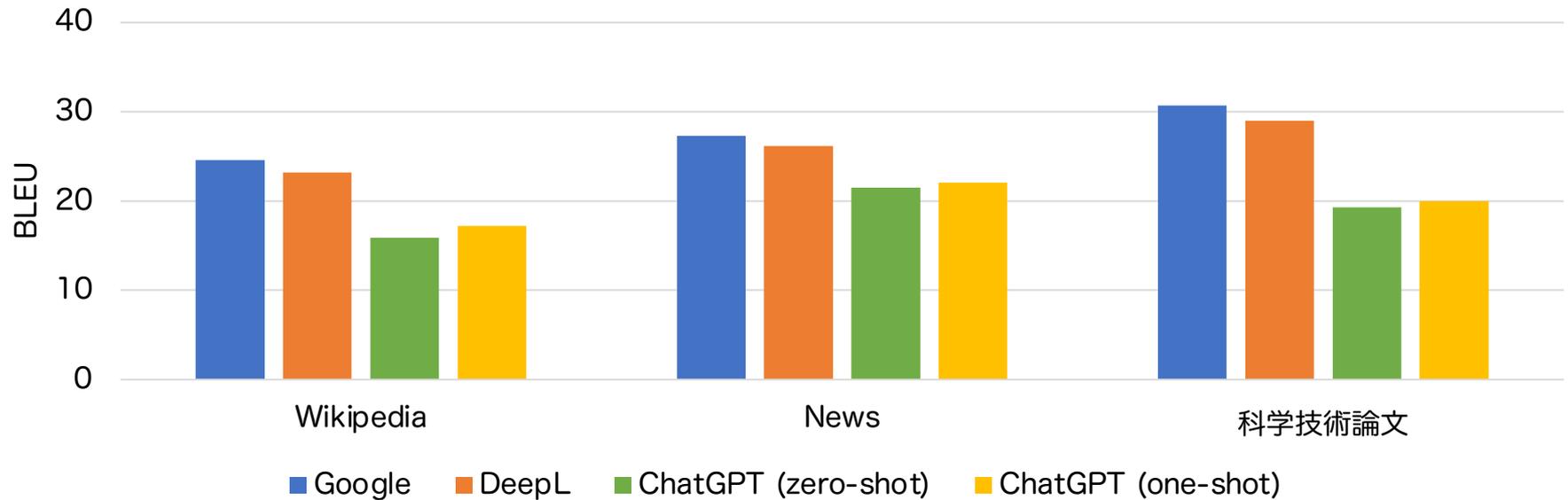


- 既存の翻訳システムには及ばない
- データの多い言語は比較的得意（英語→ドイツ語、英語→中国語）
- データの少ない言語は苦手（ルーマニア語）
- 英語が登場しない場合、英語経由で翻訳すると性能が上がる（緑→黄）

【英語経由で翻訳する際の指示】

Please provide the [English] translation first and then the [Chinese] translation for these sentences one by one: [ここに文を入力]

ChatGPTの日本語生成能力：機械翻訳（英→日）



- 他言語の翻訳品質と同じく、既存の翻訳システムに及ばない
- 事例を見せることで、わずかに翻訳品質が改善できる

算数の文章問題：それっぽいことを言うが途中で間違っ

KA

Aさんは30本のバナナを買いました。そして、AさんはBさんCさんDさんに均等に分配しました。その後、Bさんは半分をEさんに渡し、CさんとDさんは半分ずつをAさんに渡しました。最終的に、それぞれ何本ずつのバナナを持っていますか？



Aさんは最初に30本のバナナを買いました。その後、Bさん、Cさん、Dさんに均等に分配するので、それぞれには $30本 \div 3人 = 10本$ ずつのバナナがあります。



次に、Bさんは持っている10本のバナナの半分、つまり $10本 \div 2 = 5本$ をEさんに渡しました。したがって、Bさんは最終的に $10本 - 5本 = 5本$ のバナナを持っています。

一方、CさんとDさんは各々持っている10本のバナナの半分、つまり $10本 \div 2 = 5本$ をAさんに渡しました。したがって、CさんとDさんは最終的にはそれぞれ $10本 - 5本 = 5本$ のバナナを持っています。

最終的な結果は次のようになります：

Aさん： $30本 - 5本$ (Cさんに渡した分) $- 5本$ (Dさんに渡した分) $= 20本$ のバナナ

Bさん：5本のバナナ

Cさん：5本のバナナ

Dさん：5本のバナナ

Eさん：5本のバナナ

ChatGPT、算数は苦手みたい

2桁の掛け算は10/10問、3桁は4/10問、4桁は0/10問に正解できた

 ランダムに4桁の数の掛け算を10回計算してください	 ランダムに4桁の数の掛け算を10回計算してください	 ランダムに4桁の数の掛け算を10回計算してください
 はい、ランダムに4桁の数の掛け算を10回計算します。	 はい、ランダムに4桁の数の掛け算を10回計算します。	 はい、ランダムに4桁の数の掛け算を10回計算します。
<ol style="list-style-type: none">1. 58 × 12 = 6962. 14 × 23 = 3223. 26 × 34 = 8844. 39 × 45 = 17555. 67 × 56 = 37526. 85 × 67 = 56957. 93 × 78 = 72548. 30 × 89 = 26709. 49 × 90 = 441010. 62 × 01 = 62	<ol style="list-style-type: none">1. 123 × 456 = 560882. 789 × 012 = 94683. 345 × 678 = 2339104. 567 × 890 = 5052305. 432 × 101 = 436326. 876 × 234 = 2049847. 654 × 321 = 2101348. 890 × 123 = 1094709. 456 × 789 = 35998410. 321 × 012 = 3852	<ol style="list-style-type: none">1. $1234 \times 5678 = 7016652$2. $9876 \times 5432 = 53607432$3. $4567 \times 2345 = 10707515$4. $7890 \times 3210 = 25338900$5. $4321 \times 8765 = 37875465$6. $5678 \times 9012 = 51221056$7. $3456 \times 7890 = 27229440$8. $6789 \times 5432 = 36888848$9. $8901 \times 1234 = 10987434$10. $2109 \times 8765 = 18477785$

ChatGPTは、存在しないモノも存在するかのよう

KA

四国中央大学について教えてください



四国中央大学（しこくちゅうおうだいがく）は、日本の高知県に位置する私立大学です。



以下に、四国中央大学に関する一般的な情報を提供します。

- 四国中央大学は、1965年に設立された大学で、高知県中東部にキャンパスを構えています。大学の設立目的は、地域社会の発展と文化の向上に貢献することです。
- 四国中央大学は、学部・大学院を含む総合大学です。学部には、人文学部、社会福祉学部、法学部、経済学部、工学部、薬学部の6つの学部があります。また、大学院には、法学研究科、経済学研究科、人文学研究科、工学研究科、薬学研究科の5つの研究科があります。
- 四国中央大学は、幅広い学問分野で教育・研究活動を展開しています。特に、地域社会との連携を重視しており、地域の課題解決や地域振興に取り組んでいます。また、国際交流にも力を入れており、留学生の受け入れや国際的な学術交流を推進しています。
- キャンパス内には、図書館や研究施設、体育館、学生寮などの施設が整っています。学生生活をサポートするための様々な活動やサークルも充実しており、学生の自己成長と交流を促しています。
- 四国中央大学は、地域の中核大学として、学生の育成だけでなく、地域社会への貢献や人材の育成にも力を注いでいます。地域の産業や社会のニーズに合わせた教育カリキュラムやプログラムを提供し、学生が社会で活躍できる能力を身につけることを目指しています。

以上が、四国中央大学についての一般的な情報です。詳細な情報や入学試験に関する情報を知りたい場合は、大学の公式ウェブサイトや関連資料を参照することをおすすめします。

- 最新のGPT-4（2023年3月発表）は、難しい試験も突破できる
- ただし、事実に反する発言など、言語モデルの課題は残っている
- 大学入試
 - Stanford大学の大学院に入学できるレベル
- 司法試験
 - 英語：ChatGPTは下位10%レベル → GPT-4は上位10%レベル
 - 日本語：ChatGPTは最下位レベル → GPT-4でも下位10%レベル
→ アメリカの司法試験には合格できるが、日本ではまだまだ
- 医師国家試験
 - 英語：ChatGPTは5年分の過去問全て不合格 → GPT-4は全て合格
 - 日本語：ChatGPTは不合格 → GPT-4は合格ラインを僅かに上回る
→ 合格はできるが、患者に安楽死を促すなどの不適切な対応も

その他の大規模言語モデル

1. 日本語データのみで訓練したモデル
2. 日本語と英語の両方のデータで訓練したモデル
3. 英語モデルを日本語データで追加訓練したモデル
4. 英語を中心とするデータで訓練したモデル

1. 日本語	サイズ	発表時期
LINE (japanese-large-lm)	3B	2023/08
rinna (japanese-gpt-neox)	3B	2023/05
CyberAgent (OpenCALM)	7B	2023/05
Stockmark	13B	2023/10

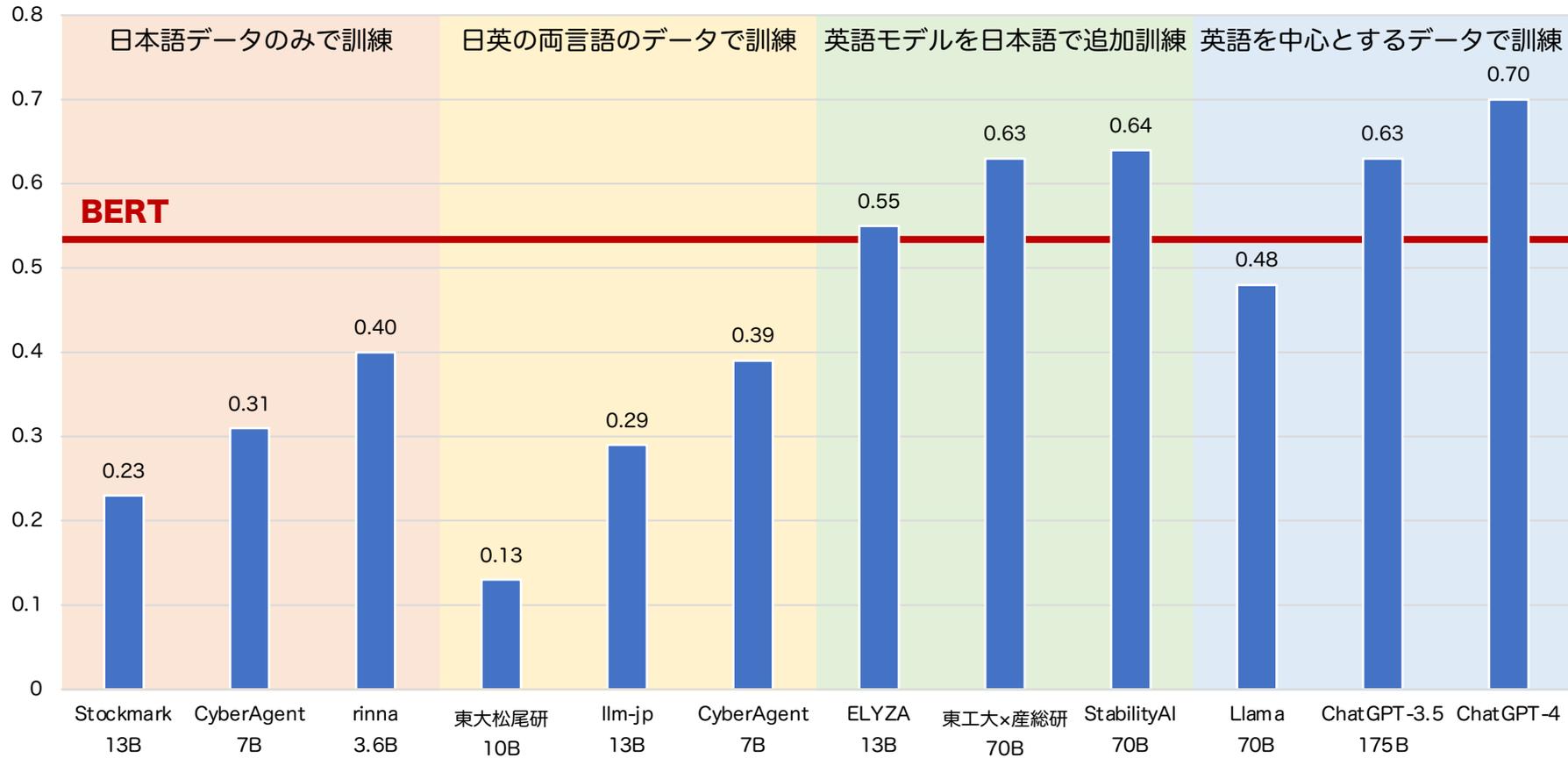
2. 日本語+英語	サイズ	発表時期
CyberAgent (CALM2)	7B	2023/11
東京大学松尾研究室	10B	2023/08
PFN (PLaMo)	13B	2023/09
LLM-JP	13B	2023/10

3. 英語→日本語	サイズ	発表時期
ELYZA	7B	2023/08
rinna (Youri)	7B	2023/10
StabilityAI	70B	2023/11
東工大×産総研 (Swallow)	70B	2023/12

4. 英語	サイズ	発表時期
Meta (LLaMA)	70B	2023/07
OpenAI (ChatGPT)	175B	2022/11
Google (PaLM)	540B	2022/04
...		

日本語LLMの感情分析タスクにおける性能比較

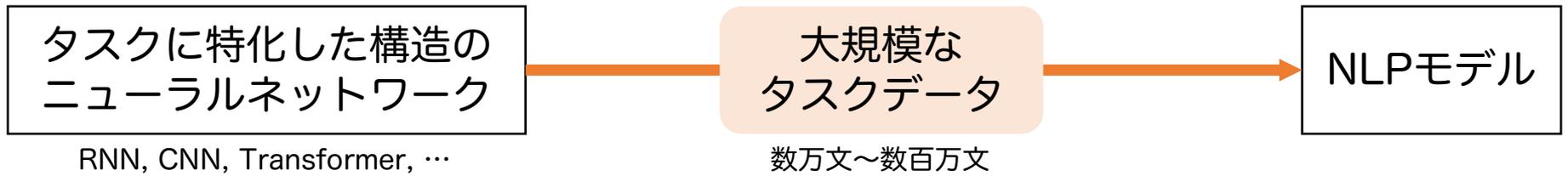
- GPT-4が最高性能
- LLaMA-70Bを日本語で追加訓練したモデルとChatGPT-3.5が高性能
- 日本語ベースLLMはデータもモデルも小規模で、BERTを超えられない



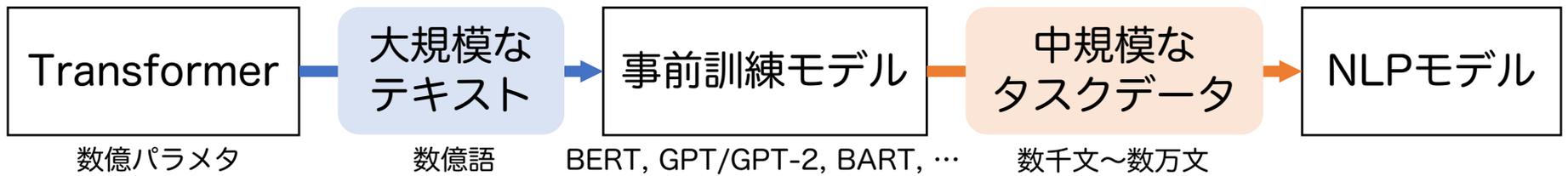
※ 近藤里咲, 大塚琢生, 梶原智之, 二宮崇, 早志英朗, 中島悠太, 長原一. 大規模言語モデルによる日本語感情分析の性能評価. 情報処理学会第86回全国大会 (2024年3月) 37

まとめ (1/2) 自然言語処理の主要なアプローチの変遷

- 2013年～ タスクに特化した深層学習モデルの学習



- 2017年～ Transformerモデルの転移学習



- 2020年～ Transformerモデルの文脈内学習 (大規模言語モデル)



まとめ (2/2) ChatGPTと大規模言語モデル

- 2010年代からの第3次AIブームの中で、**深層学習**の技術（ビッグデータ × ニューラルネットワーク）が進歩し、Webの大規模テキストと**Transformer**で自然言語処理が進化
- ChatGPTなどの大規模言語モデルは、大量のテキストを使って「**次の単語の推定**」を繰り返し学習した自然言語処理モデル
- 指示を与えると、その**指示に続く自然なテキスト**を推定するという振る舞いの中で、それらしい**回答を確率的に生成**する
- インターネット上に書いてあることは何でも知っているが、**自ら考えるわけではない**ため、簡単な算数も解けないことがある

