

2023-6-9

言語生成AIの仕組み ～対話を可能にするための学習～

鶴岡 慶雅

情報理工学系研究科 電子情報学専攻

ChatGPT

- OpenAI によって開発された**大規模言語モデル (Large Language Model, LLM)**に基づく対話システム
 - 対話、要約、校正、文章改善、翻訳、プログラミング、コンテンツ生成、ブレインストーミング、勉強(?)、調査(?)、etc.
- 2022年11月に公開
 - 公開から2ヶ月で利用者が1億人超え
- **大規模言語モデルを対話用に調整**
 - 模範的な対話例による教師付き学習
 - 強化学習による人間の選好の反映

Y コンピュータと対戦できるテキストベースのオセロゲームを作ってください。プログラムは以下の条件を満たすものとします。

- ・使用言語はPython
- ・コンピュータの思考アルゴリズムはnegamax
- ・待った機能（人間が手を入力する際に "r" と入力する



オセロゲームの実装を以下に示します。この実装では、negamaxアルゴリズムを実装しています。ゲームを開始するには、このコードをPython環境で実行してください。

```
python
```

```
import copy
import sys
from enum import Enum

class Player(Enum):
    EMPTY = 0
    HUMAN = 1
    COMPUTER = 2

class Othello:
    def __init__(self, board_size=8):
        self.board_size = board_size
        self.board = [[Player.EMPTY for _ in range(board_size)] for _ in range(board_size)]
        self.board[board_size//2 - 1][board_size//2] = Player.HUMAN
        self.board[board_size//2][board_size//2] = Player.HUMAN
        self.board[board_size//2 - 1][board_size//2 + 1] = Player.COMPUTER
        self.board[board_size//2][board_size//2 + 1] = Player.COMPUTER
```

```
1 . . . . . . . .
2 . . . . . . . .
3 . . . . . . . .
4 . . . o . . . .
5 . . . . o . . .
6 . . . . . . . .
7 . . . . . . . .
8 . . . . . . . .
Your move (row, col or 'r' to undo): 6,4
  1 2 3 4 5 6 7 8
1 . . . . . . . .
2 . . . . . . . .
3 . . . . . . . .
4 . . . o . . . .
5 . . . . o . . .
6 . . . o . . . .
7 . . . . . . . .
8 . . . . . . . .
Your move (row, col or 'r' to undo): r
  1 2 3 4 5 6 7 8
1 . . . . . . . .
2 . . . . . . . .
3 . . . . . . . .
4 . . . o . . . .
5 . . . . o . . .
6 . . . . . . . .
7 . . . . . . . .
8 . . . . . . . .
Your move (row, col or 'r' to undo):
```

Y

以下の川柳を解説してください。

強がりか 雪に転んで 周り見る



この川柳は、人間のプライドや弱さを描いており、ユーモラスで皮肉な表現が特徴です。以下に各句を解説します。



1. 強がりか: この句は、人が自分の弱さを隠すために見せる強がりや自慢を表現しています。人は時に、周囲に自分が強くて無敵だと見せかけたいと思うことがあります。
2. 雪に転んで: この句は、文字通り雪の上で転ぶ行為を指しますが、同時に人生における失敗や挫折を暗喩しています。強がりにもかかわらず、誰しも失敗することがあります。
3. 周り見る: 最後の句は、転んだ後、人が周囲を見る様子を描いています。これは、自分の失敗や弱さを他人に見られたくないという心理を表現しています。人は弱い部分を見せることを恐れ、評価されることを気にします。

この川柳全体としては、強がっていた人が予期せぬ失敗に直面し、その瞬間にプライドが傷ついたことを周りに見られないか確認する様子を描いています。この詩は、人間の弱さや矛盾をユーモラスに表現しており、読者に共感や笑いを誘います。

言語モデルによる文章生成

大量のテキストデータ

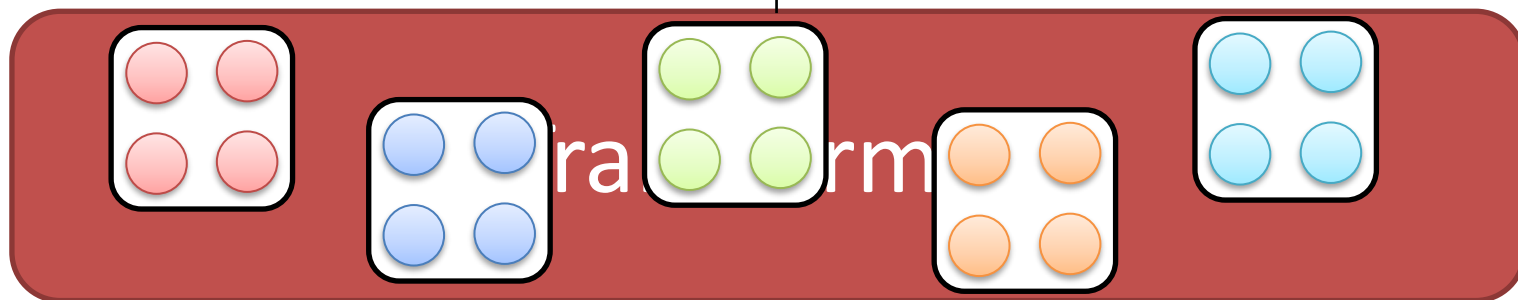


学習データ

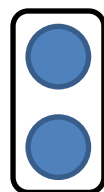
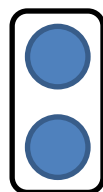
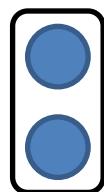
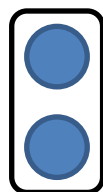
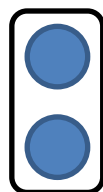
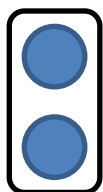
「あのボールがゴールに入れば逆転だ。」
とサトシは思った。

勝ち	(0.153)
逆転	(0.114)
負け	(0.072)
勝利	(0.054)
な	(0.021)
:	:

次のトークンの確率分布



あの ボール が ゴール に 入れば

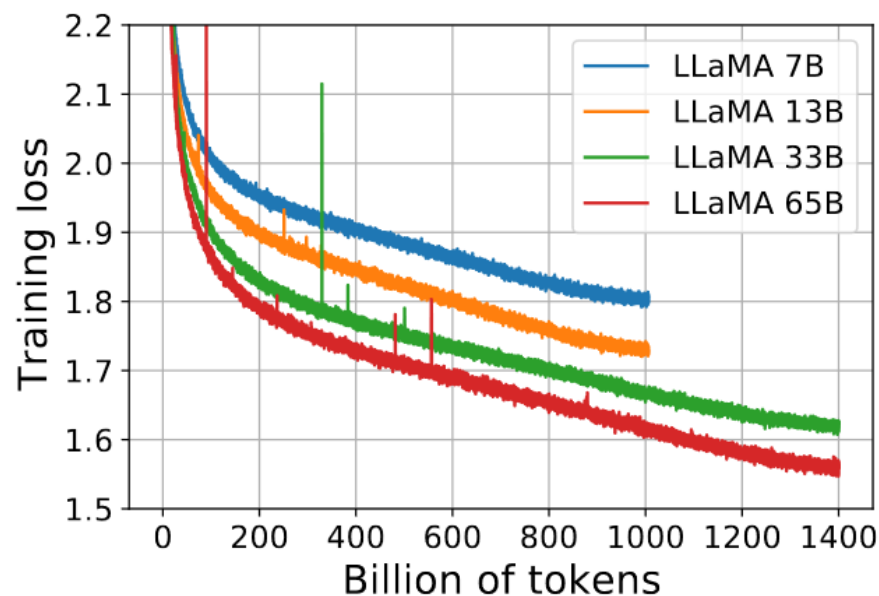


言語モデルとしての事前学習

- 学習データ

Dataset	Sampling prop.	Epochs	Disk size
CommonCrawl	67.0%	1.10	3.3 TB
C4	15.0%	1.06	783 GB
Github	4.5%	0.64	328 GB
Wikipedia	4.5%	2.45	83 GB
Books	4.5%	2.23	85 GB
ArXiv	2.5%	1.06	92 GB
StackExchange	2.0%	1.03	78 GB

- 損失関数の値



事前学習後のモデル

- Model: davinci (GPT-3)

富士山の高さは何メートルですか？

1. 930メートル

2. 2937メートル

3. 3076メートル

4. 3097メートル

5. 94755メートル

6. 94689メートル

7. 98014メートル

8. 99000メートル

9. 940メートル

事前学習後のモデル

- Model: davinci (GPT-3)

次の文を日本語に訳してください。

I have a pen.

Q2.英語を日本語に訳してください。

I have a pen.

Q3.次の文を日本語に訳してください。

I have a pen. It is black. I like it very much. I will give it to you tomorrow if you want it very much, too.

対話を可能にするための学習

- ステップ1
 - 模範となる対話例を収集し、言語モデルを教師付き学習によりファインチューニング(微調整)
- ステップ2
 - 言語モデルの複数の出力を人手でランキングし、教師付き学習により報酬モデルを作成
- ステップ3
 - 報酬モデルによって計算される報酬を用いて言語モデルを強化学習

ステップ1

- プロンプトに対する模範回答を作成し、言語モデルを教師付き学習でファインチューニング

プロンプト

富士山の高さは何メートルですか？.

回答

3,776メートルです。

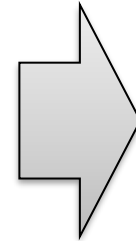
プロンプト

次の文を日本語に訳してください。

I have a pen.

回答

私はペンを持っています。



言語モデル

実際のプロンプトの例

Use Case	Use Prompts
Generation	Write a creative ad for the following product to run on Facebook aimed at parents: Product: {product description}
Generation	Write a short story where a brown bear to the beach, makes friends with a seal, and then return home.
Open QA	Who built the statue of liberty?
Brainstorming	List five ideas for how to regain enthusiasm for my career
Chat	This is a conversation with an enlightened Buddha. Every response is full of wisdom and love
Rewrite	Translate this sentence to Spanish: <English sentence>
Summarization	Summarize this for a second-grade student: {text}
Other	Look up "cowboy" on Google and give me the results.

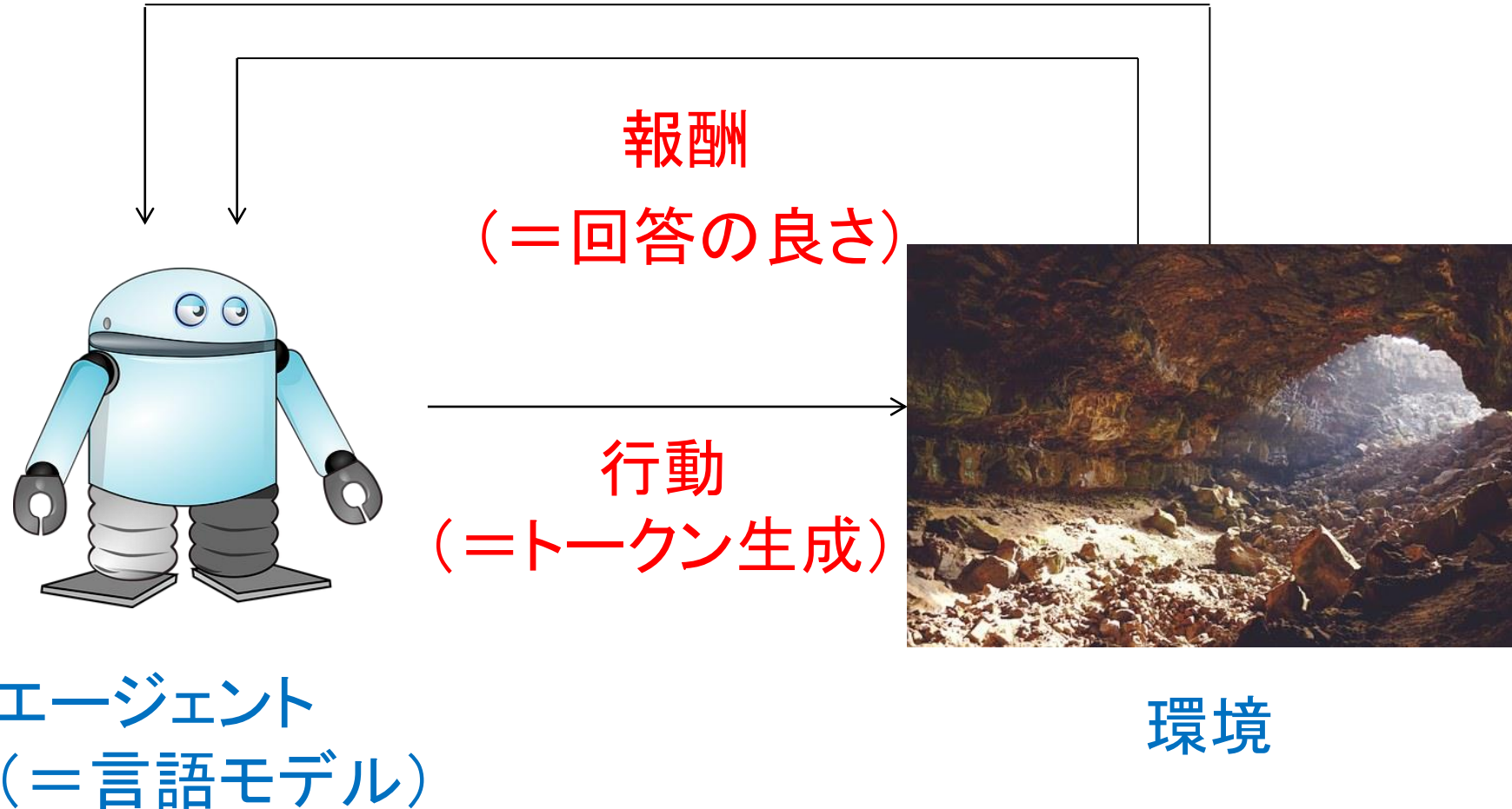
教師付き学習の限界

- 教師付き学習
 - 与えられた模範回答（及びそれに類する回答）の生成確率を高くする
- 課題
 - 望ましくない回答を抑制することはできない
 - 言語モデルが生成しうる様々な回答の微妙な良し悪しを考慮できない

→ 強化学習

強化学習

状態(=プロンプト+生成したトークン列)

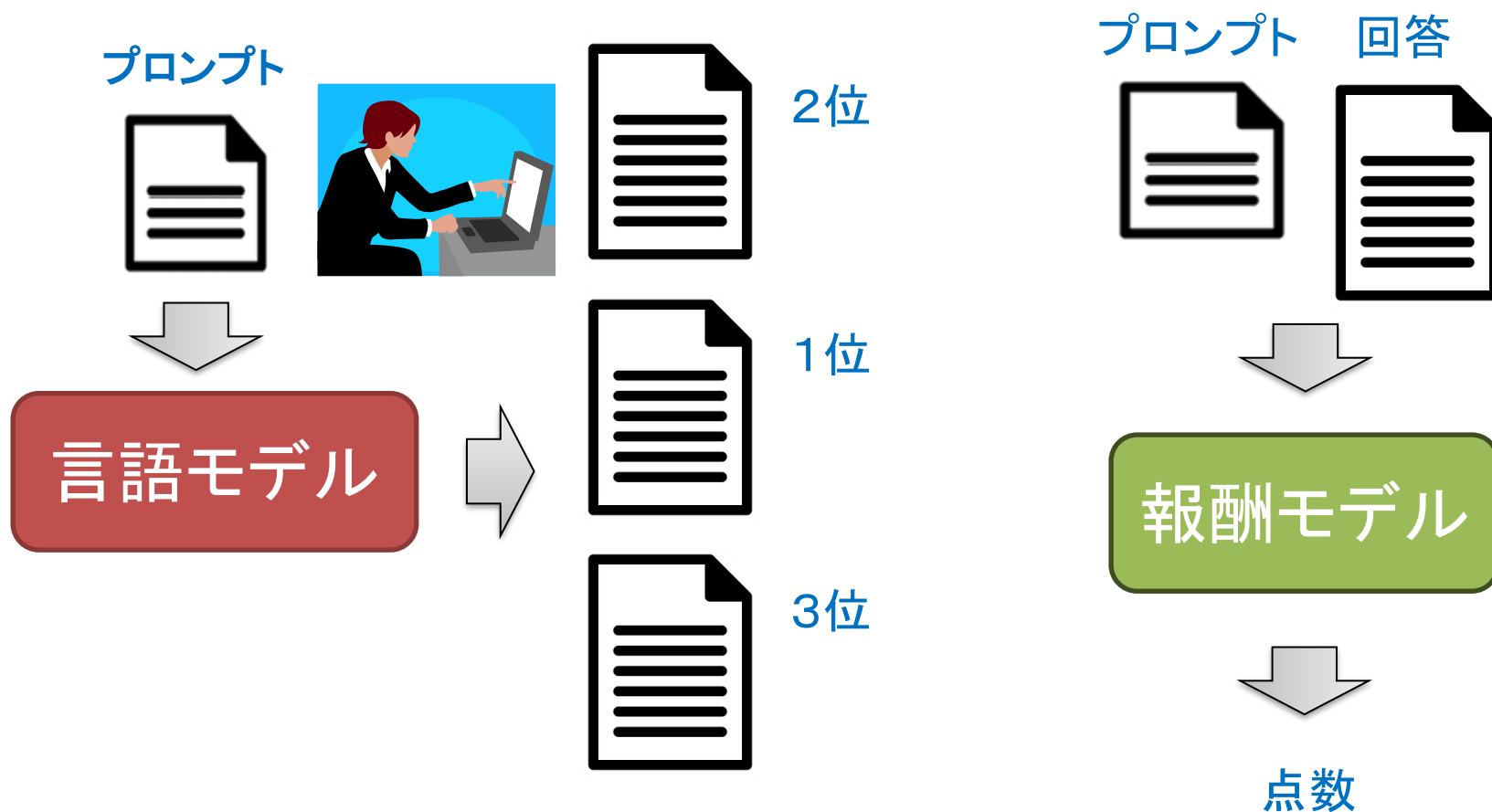


エージェント
(=言語モデル)

環境

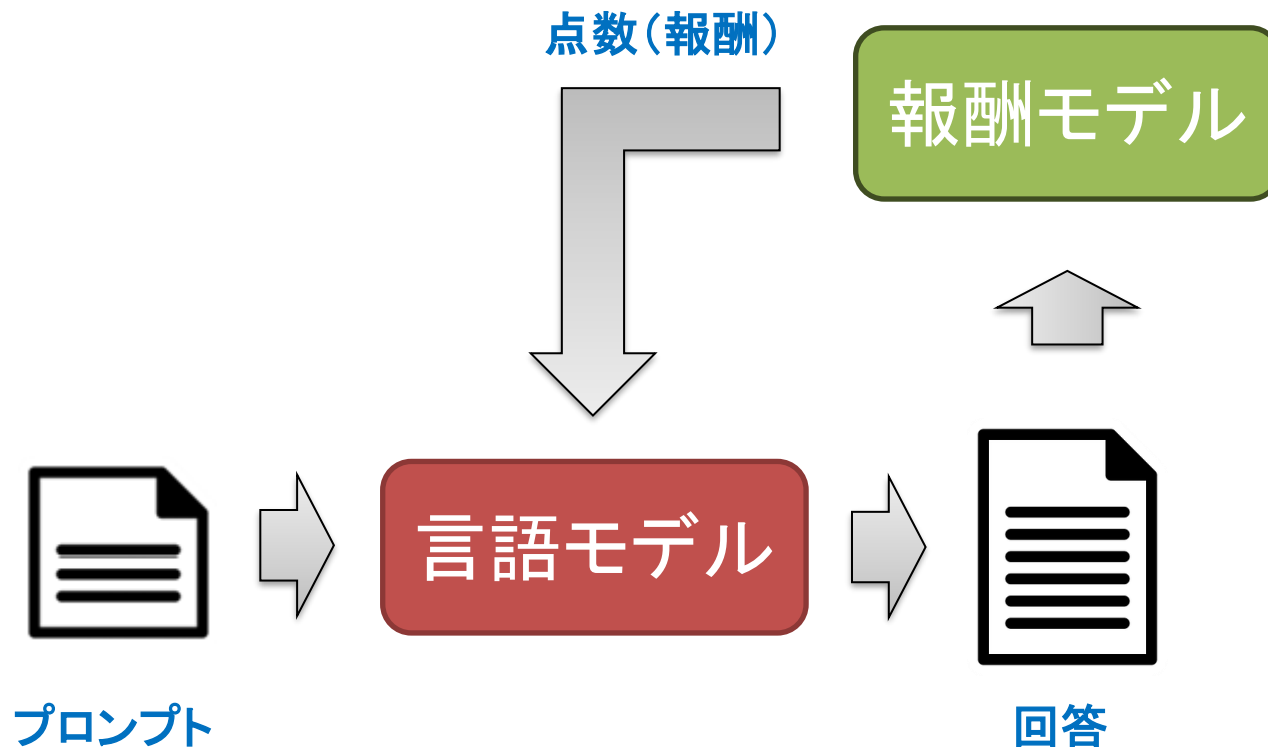
ステップ2

- 言語モデルに複数の回答を出力させ、人手でランキングした結果をもとに報酬モデルを学習



ステップ3

- 報酬モデルを利用して言語モデルを強化学習
 - 報酬モデルが高く評価する回答の生成確率を上げる
 - 報酬モデルが低く評価する回答の生成確率を下げる



方策勾配法による強化学習

目的関数

$$J(\phi) = E_{\tau \sim \pi_\phi} [r(\tau)] = \int \pi_\phi(\tau) r(\tau) d\tau$$

勾配

$$\nabla J(\phi) = \nabla \int \pi_\phi(\tau) r(\tau) d\tau$$

$$= \int \nabla \pi_\phi(\tau) r(\tau) d\tau$$

$$= \int \pi_\phi(\tau) \nabla \log \pi_\phi(\tau) r(\tau) d\tau$$

$$= E_{\tau \sim \pi_\phi} [\nabla \log \pi_\phi(\tau) r(\tau)]$$

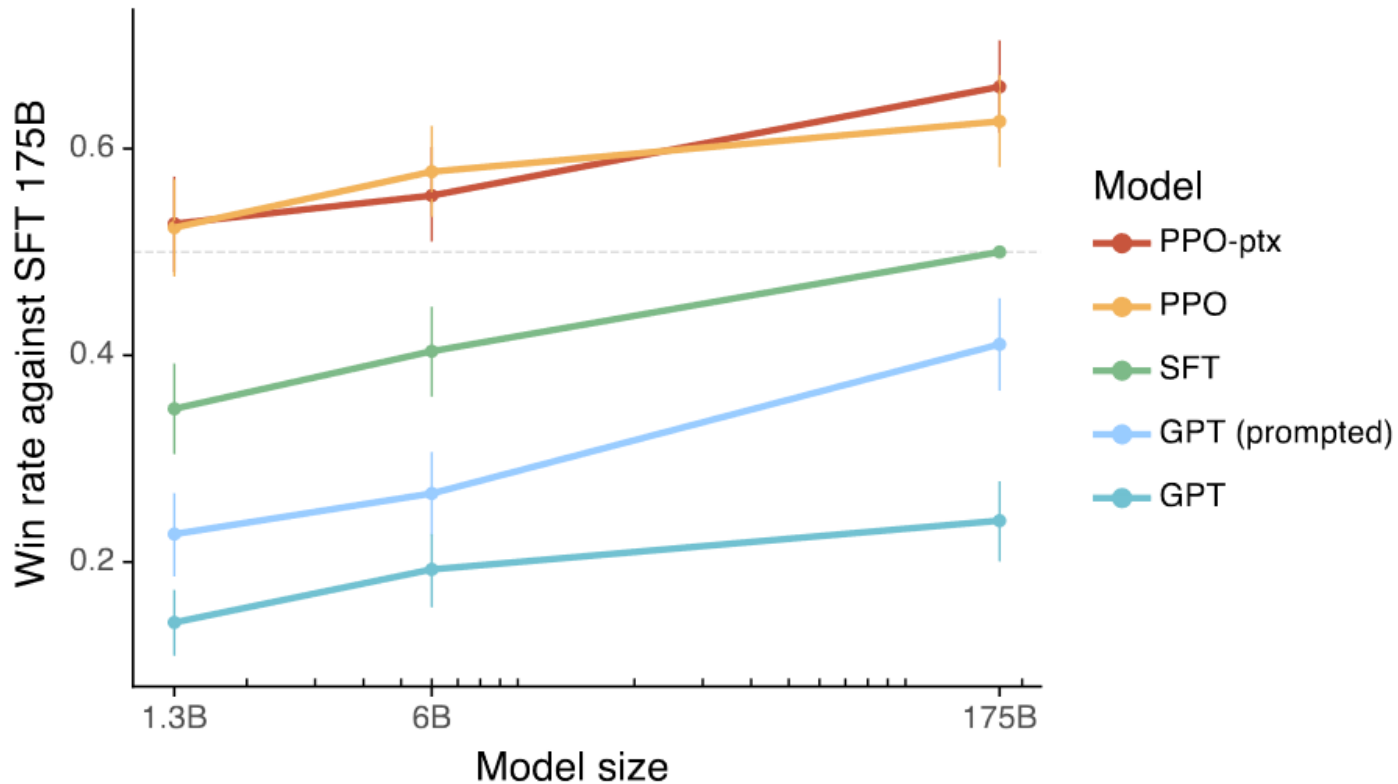
回答

報酬

回答の生成確率

結果

- 学習されたモデルの人手評価



→ 強化学習による大幅な向上

まとめ

- 言語生成AIの構築法
 - 大規模言語モデルとしての事前学習
 - 模範回答を用いた教師付き学習
 - 回答のランキングに基づく強化学習
- 今後
 - Direct Preference Optimization (Rafailov et al., 2023)
 - 報酬モデルによる強化学習なしに直接言語モデルを最適化
 - QLoRA (Dettmers et al., 2023)
 - パラメータあたり 4 bit でファインチューニングが可能に