

# 自然言語処理 —大規模言語モデル概観—

<https://satoyoshiharu.github.io/nlp/>

# ChatGPT

トランスフォーマー言語モデル  
を巨大化

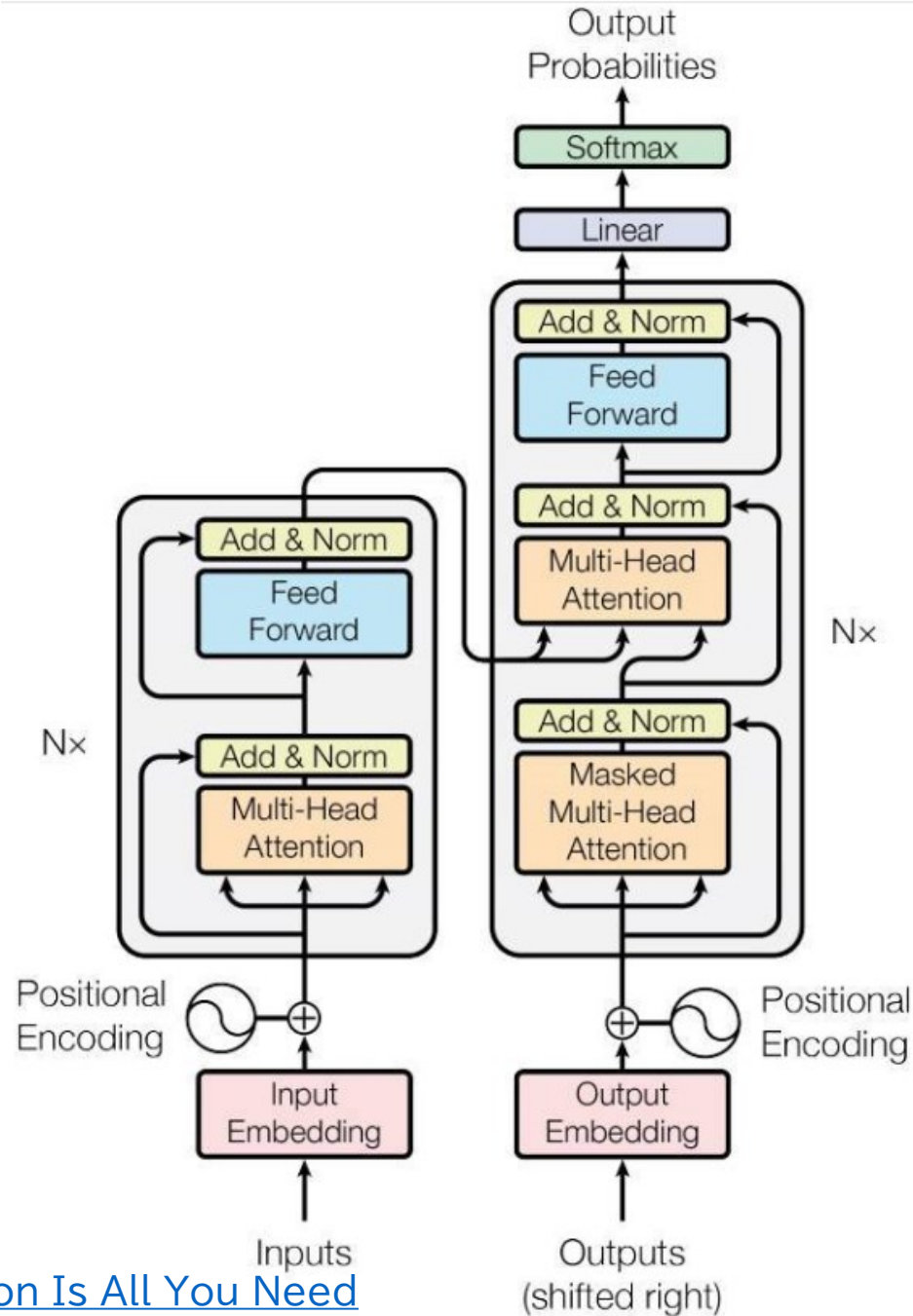
強化学習で人とのやり取り向き  
に訓練

# BERT

Encoderは、self-attention、全結合、Residual結合などを何層も重ねる

所々穴にして、穴の単語を推理させる  
Masked LMとして訓練

[Attention Is All You Need](#)

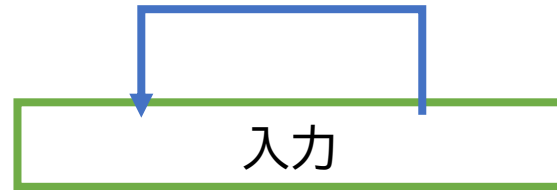


Decoderは、self-attention、全結合、Cross-Attention、Residual結合などを何層も重ねる

単語列で、次の単語を推理させる  
Causal LMとして訓練



# Self-Attention



「鳥がナク」

「ナク」は鳥に注目すれば、「泣く」でなくて「鳴く」。

**ある入力文の中で、ある単語とほかの単語の関連度**

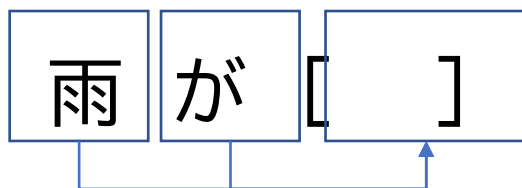


# Causal Language Model

雨 が 激しく 降る

雨 が 激しく 降る

先行単語を与えて、  
後続単語をマスク

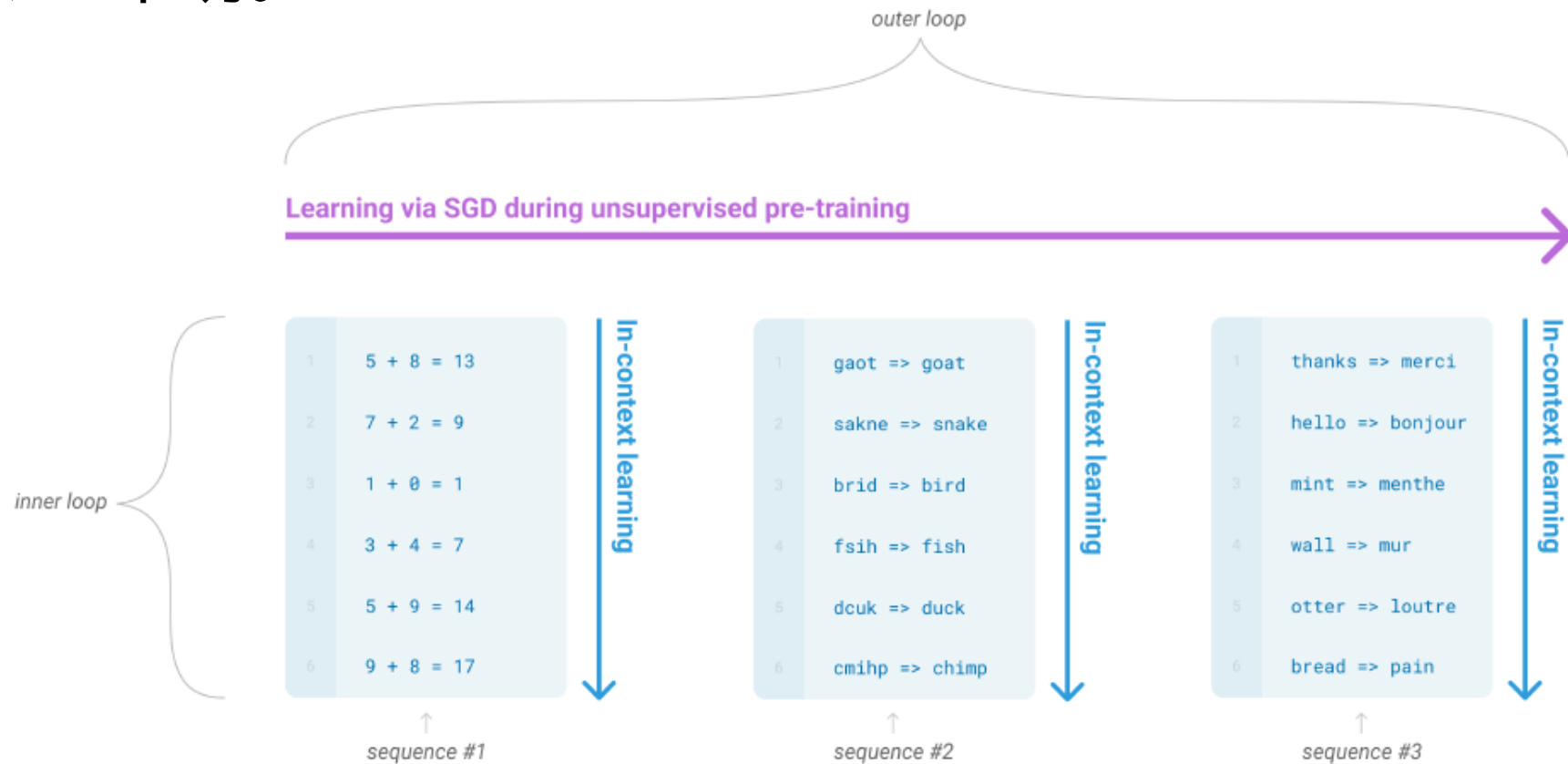


先行単語から次単語を  
予測

**先行単語との関連度から次の単語を予測する**



In-Context Learning: 言語モデルを巨大化するに伴い、ある問題解決をしている系列が訓練例に含まれる。その結果、実行時にいろいろな問題解決を類推できるようになったらしい。詳しい仕組みは不明。

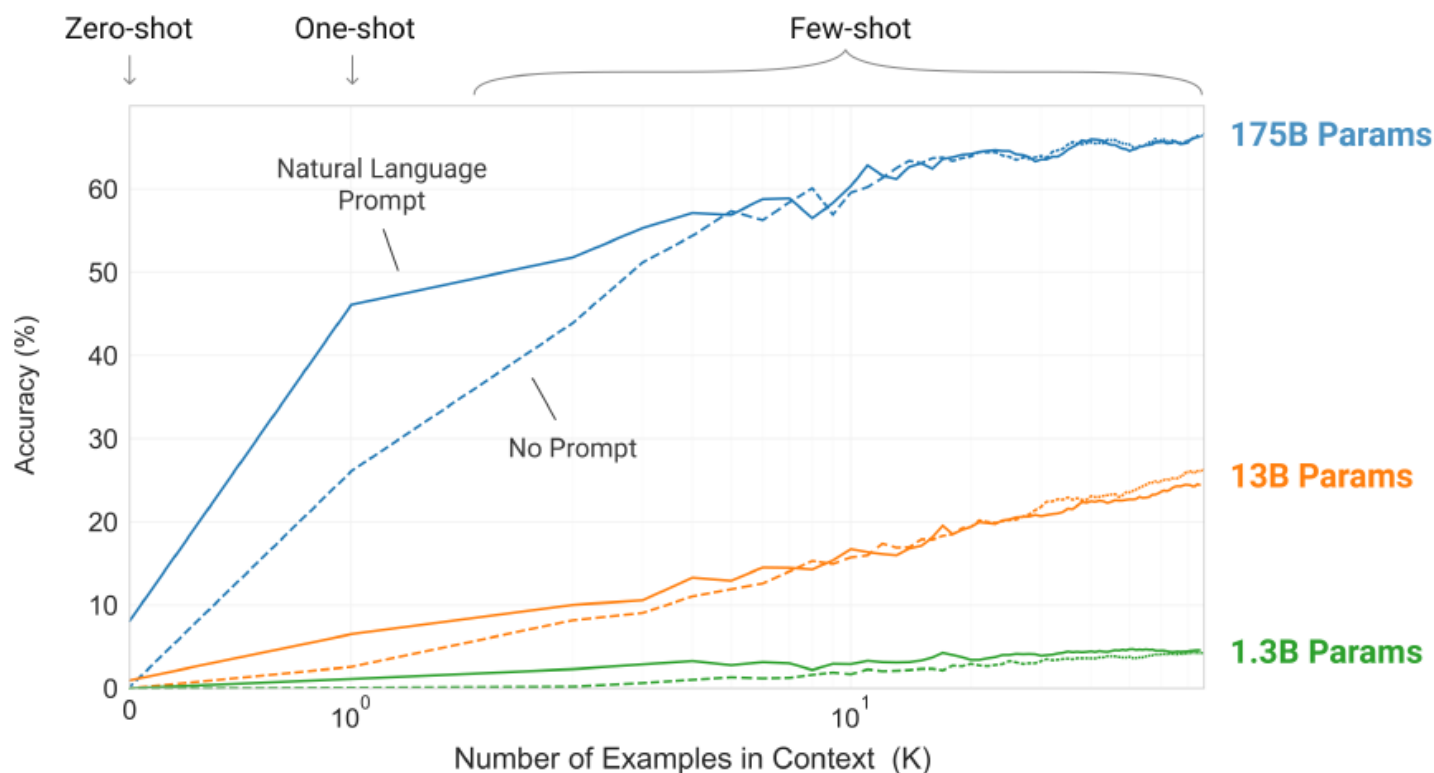


特定のタスク  
領域への転移  
学習・ファイン  
チューニング  
しなくていい

Language Models are Few-Shot Learners

# In-Context Learning: ニューラルネットのパラメータ数をでかくすると、In-Context Learningがどんどんうまくできるようになる。

Few-shot学習というのは、In-Context学習のことで、いくつか先行お手本(デモ)事例を、問に入れたもの



Language Models are Few-Shot Learners

# GPT-3にRLHF (Reinforcement Learning from Human Feedback)したらうまくいったよ。

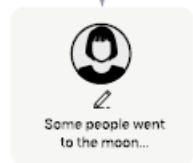
1. 人手で、こんなやり取りをしたいという例を作り、教師あり学習した。
2. 複数の解答を吐き出させて、望ましさでランクづけ。そのデータで、報酬モデルを作成した。
3. その報酬モデルで、解答を導くポリシーモデルを作成。報酬とポリシーを相互に改善してブート。

Step 1  
Collect demonstration data,  
and train a supervised policy.

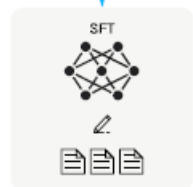
A prompt is  
sampled from our  
prompt dataset.



A labeler  
demonstrates the  
desired output  
behavior.

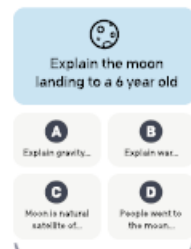


This data is used  
to fine-tune GPT-3  
with supervised  
learning.



Step 2  
Collect comparison data,  
and train a reward model.

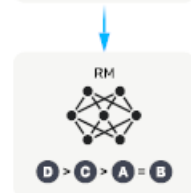
A prompt and  
several model  
outputs are  
sampled.



A labeler  
ranks the  
outputs from  
best to worst.



This data is used  
to train our  
reward model.

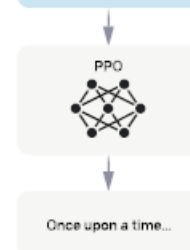


Step 3  
Optimize a policy against  
the reward model using  
reinforcement learning.

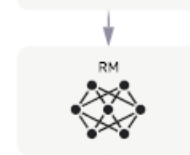
A new prompt  
is sampled from  
the dataset.



The policy  
generates  
an output.



The reward model  
calculates a  
reward for  
the output.



The reward is  
used to update  
the policy  
using PPO.



Training language models to follow instructions with human feedback



# 分かりやすい技術概観

- 話題爆発中のAI「ChatGPT」の仕組みにせまる！

# ChatGPT API入門

- [ChatGPT APIの使い方入門 - PythonからAIを使おう](#)