ChatGPT算法与模型能力理解

徐兴成

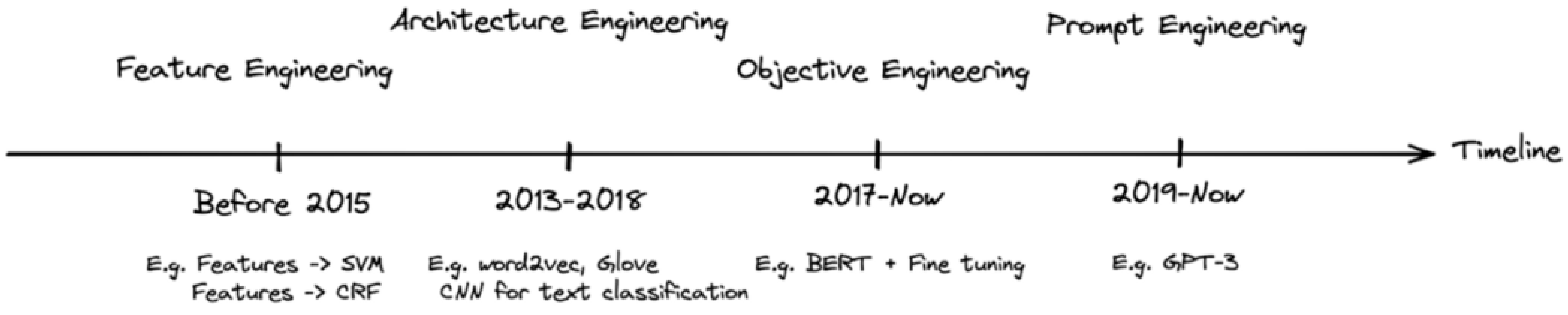
2023年2月

Outline

1. NLP Paradigm
2. The Foundation: Transformer
3. From GPT to ChatGPT
4. Anthropic LM
5. Understanding LLM

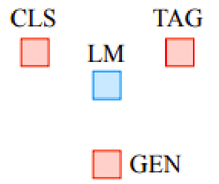
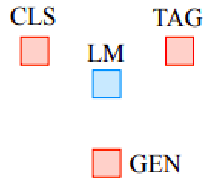
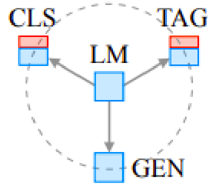
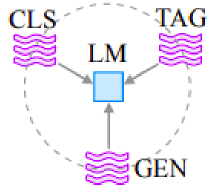
NLP Paradigm

Evolution of NLP Paradigm



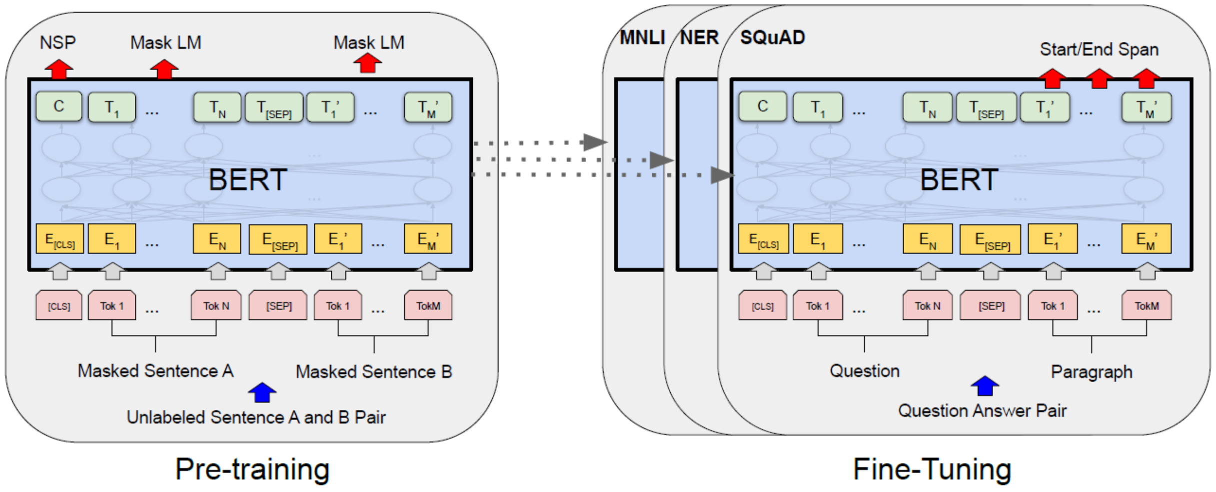
Source: https://medium.com/agoda-engineering/from-tf-idf-to-prompt-based-learning-agodas-nlp-applications-2dfe8abd942a

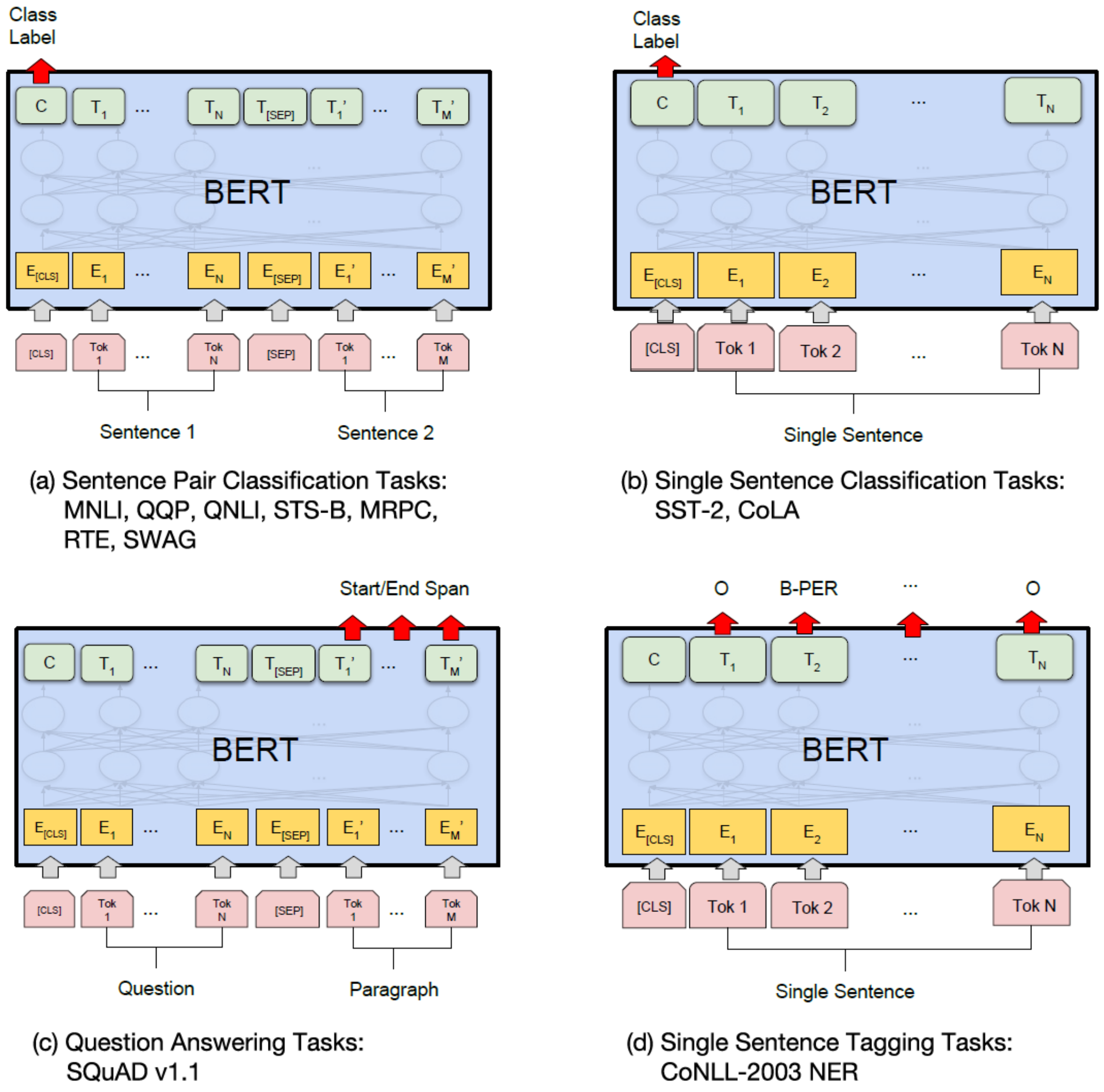
Evolution of NLP Paradigm

Paradigm	Engineering	Task Relation
a. Fully Supervised Learning (Non-Neural Network)	Features (e.g. word identity, part-of-speech, sentence length)	
b. Fully Supervised Learning (Neural Network)	Architecture (e.g. convolutional, recurrent, self-attentional)	
c. Pre-train, Fine-tune	Objective (e.g. masked language modeling, next sentence prediction)	
d. Pre-train, Prompt, Predict	Prompt (e.g. cloze, prefix)	

Source: Liu, Yuan, Fu, et al. (2023), ACM Computing Surveys.

NLP Paradigm: Pretraining + Fine-Tuning





Illustrations of Fine-tuning BERT on Different Tasks.

NLP Paradigm: Pretraining, Prompt, Predict

- Prompt Addition
  - Transform input x into prompt x'
    - Define template: input [x] and answer [z]
    - Fill in the input slot [x]
- Answer Prediction
  - Using pretrained Language Model to predict
    - Fill [z]
- Answer-Label Mapping
  - Map the answer to a class label

- Input x
  - "I love this movie"
- Template
  - [x] Overall, it was a [z] movie
- Prompting x'
  - "I love this movie, Overall it was a [z] movie."
- Prediction x'
  - "I love this movie, Overall it was a fantastic movie."
- Mapping
  - fantastic = Positive

Type	Task	Input ([X])	Template	Answer ([Z])
Text CLS	Sentiment	I love this movie.	[X] The movie is [Z].	great fantastic ...
	Topics	He prompted the LM.	[X] The text is about [Z].	sports science ...
	Intention	What is taxi fare to Denver?	[X] The question is about [Z].	quantity city ...
Text-span CLS	Aspect Sentiment	Poor service but good food.	[X] What about service? [Z].	Bad Terrible ...
Text-pair CLS	NLI	[X1]: An old man with ... [X2]: A man walks ...	[X1]? [Z], [X2]	Yes No ...
Tagging	NER	[X1]: Mike went to Paris. [X2]: Paris	[X1][X2] is a [Z] entity.	organization location ...
Text Generation	Summarization	Las Vegas police ...	[X] TL;DR: [Z]	The victim ... A woman ... ...
	Translation	Je vous aime.	French: [X] English: [Z]	I love you. I fancy you. ...

Examples of prompting workflow: to design input to fit the model.

NLP Paradigm: Pretraining, Prompt, Predict → ICL

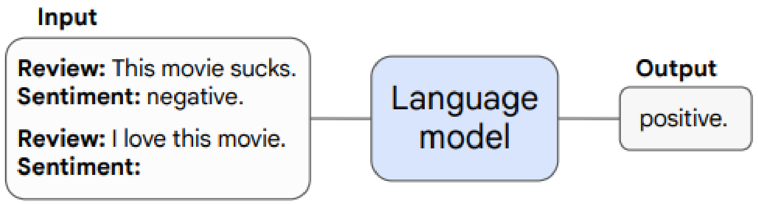


Figure 1: Example of an input and output for few-shot prompting.

(ICL)

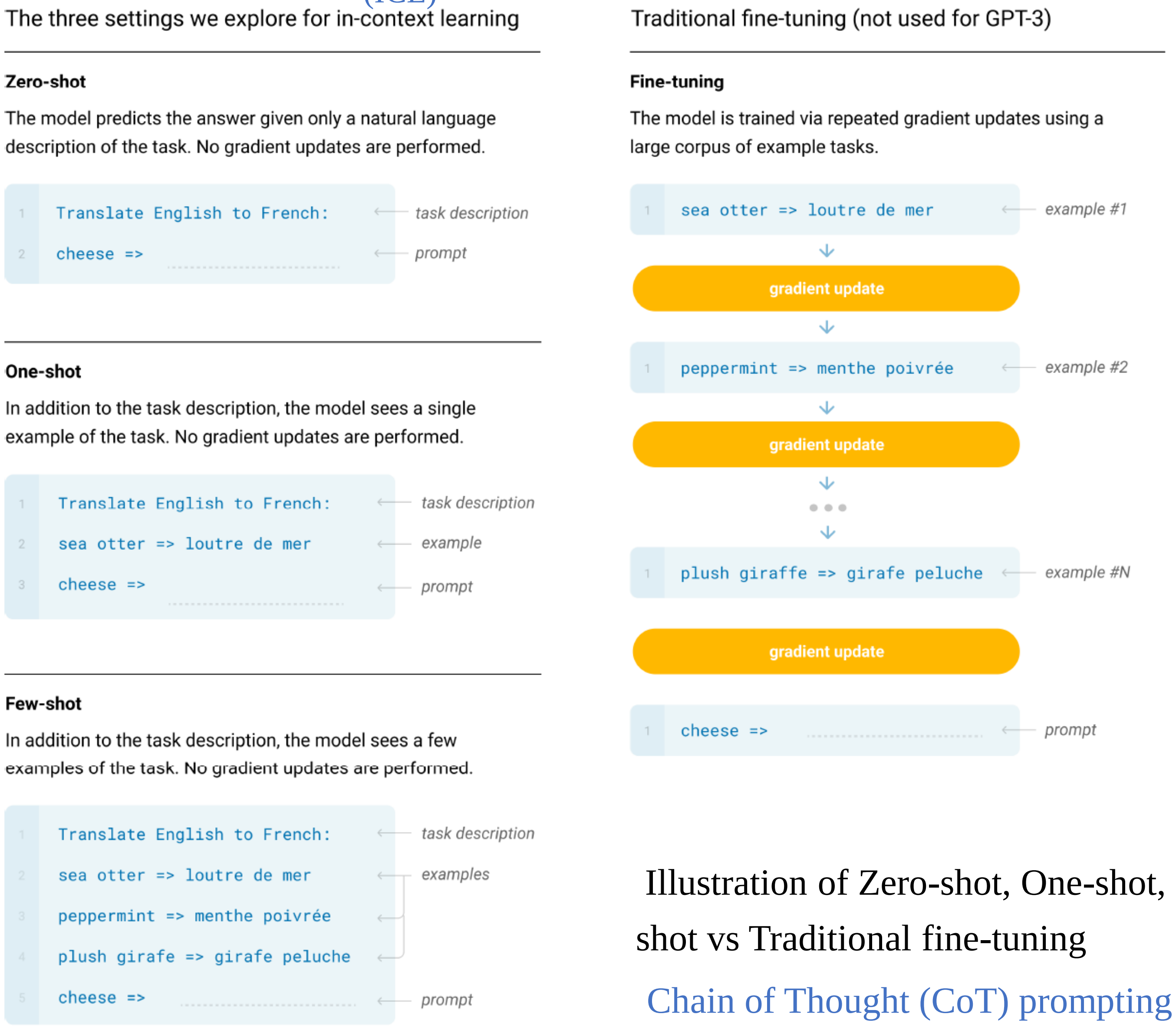


Illustration of Zero-shot, One-shot, and Few-shot vs Traditional fine-tuning

Chain of Thought (CoT) prompting

The Foundation: Transformer

Transformer: Attention is all you need

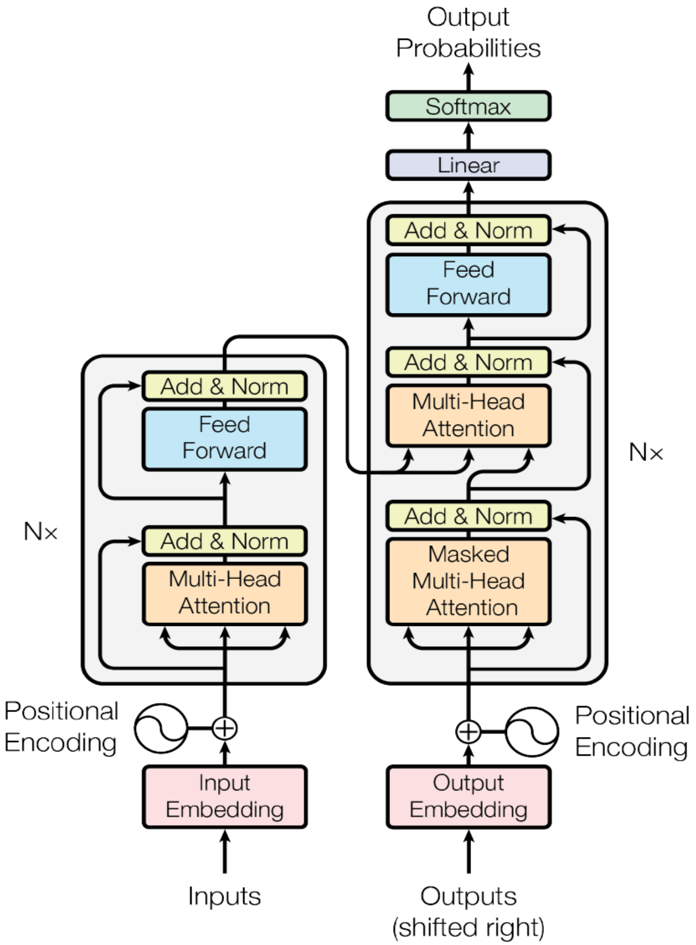
All but one author of the landmark paper that introduced transformer-based neural networks have left Google to build their own startups in AGI, conversational agents, AI-first biotech and blockchain.

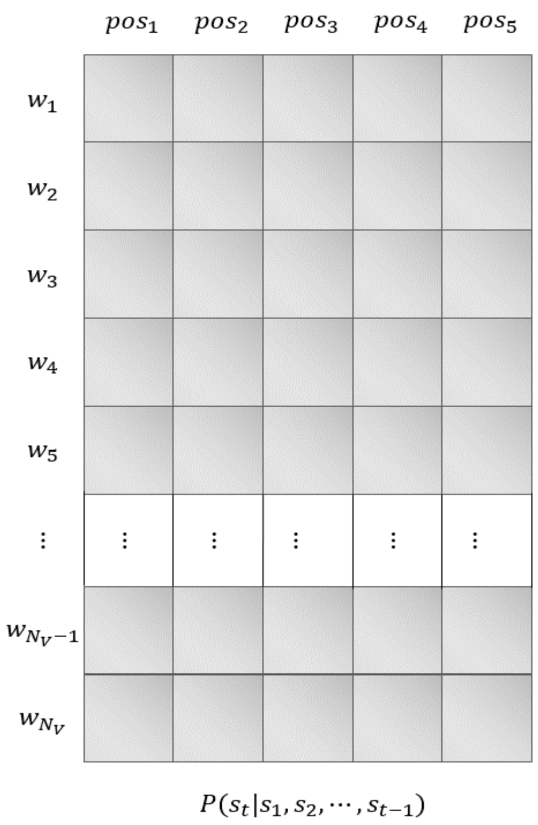
Vaswani, et al (2017). "Attention is all you need." *NeurIPS*. (被引用次数：65593)

Transformer: A Probability Machine

factorize the joint probabilities $P(x)$ of a sequence $x = (s_1, s_2, \cdots, s_n) \in V^*$ in a space of sequences $V^* = \cup_{\ell=0}^{\infty} V^\ell$ (V is a vocabulary of tokens) as the product of conditional probabilities

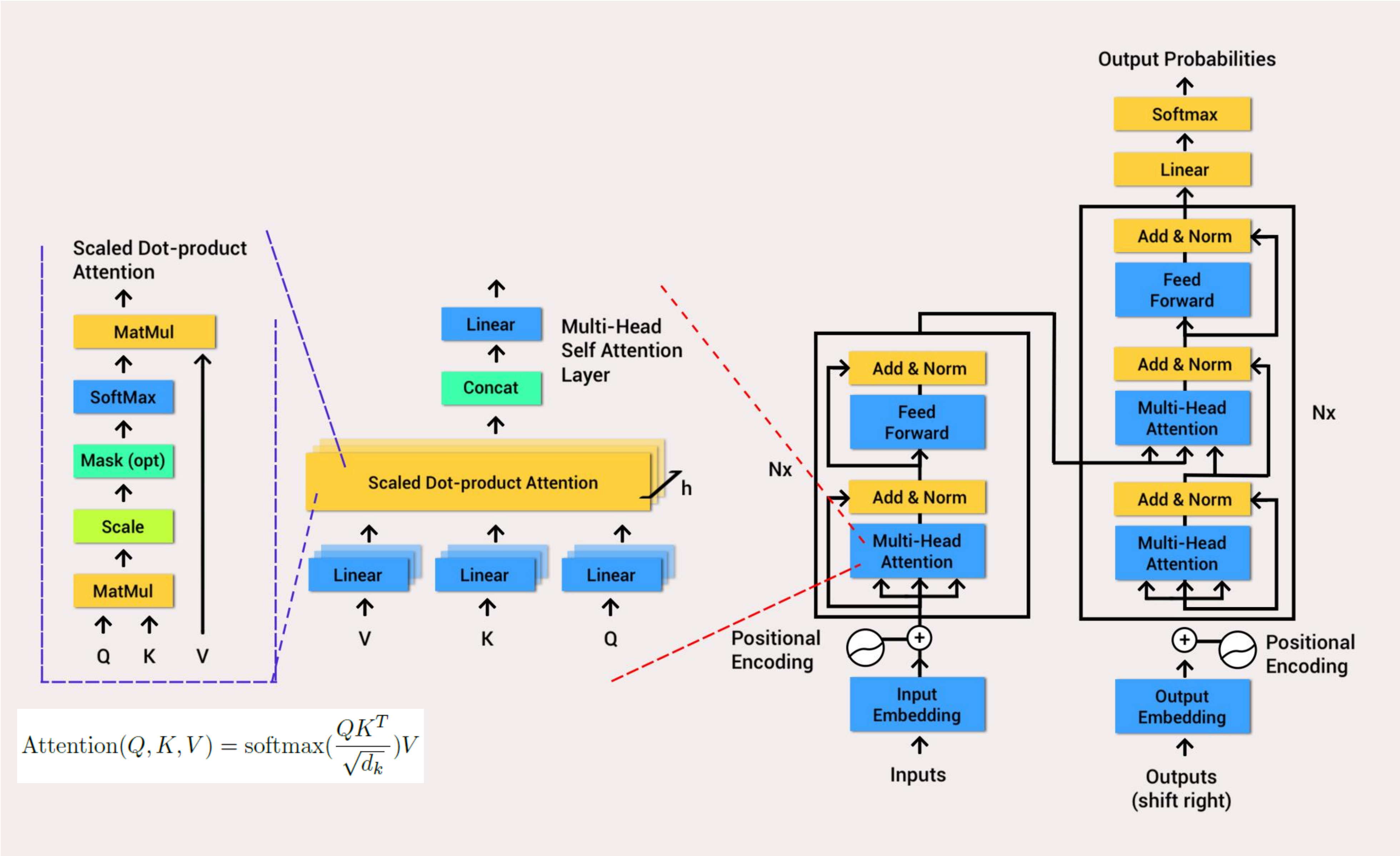
$$P(x) = \prod_{t=1}^{n} P(s_t | s_1, s_2, \cdots, s_{t-1}) \qquad P(x|z) = \prod_{t=1}^{n} P(s_t | s_1, s_2, \cdots, s_{t-1}, z)$$





Next-Token Probability

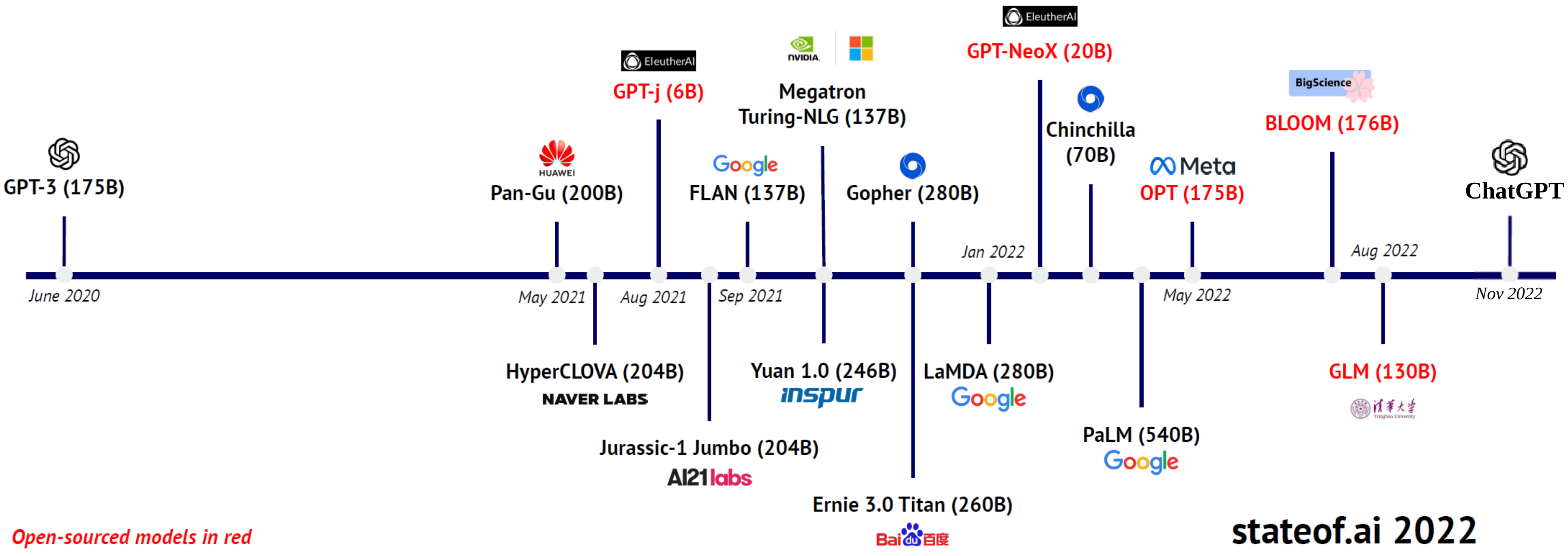
Transformer: A Probability Machine



$$\mathrm{Attention}(Q,K,V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Transformer-based LLM

Five years after the Transformer:
GPT-3, PaLM, LaMDA, Gopher, OPT, BLOOM, GPT-Neo, Megatron-Turing NLG, GLM-130B, ChatGPT, etc. all use the original attention layer in their transformers.



From GPT to ChatGPT

OpenAI: From GPT to ChatGPT

Timeline to ChatGPT

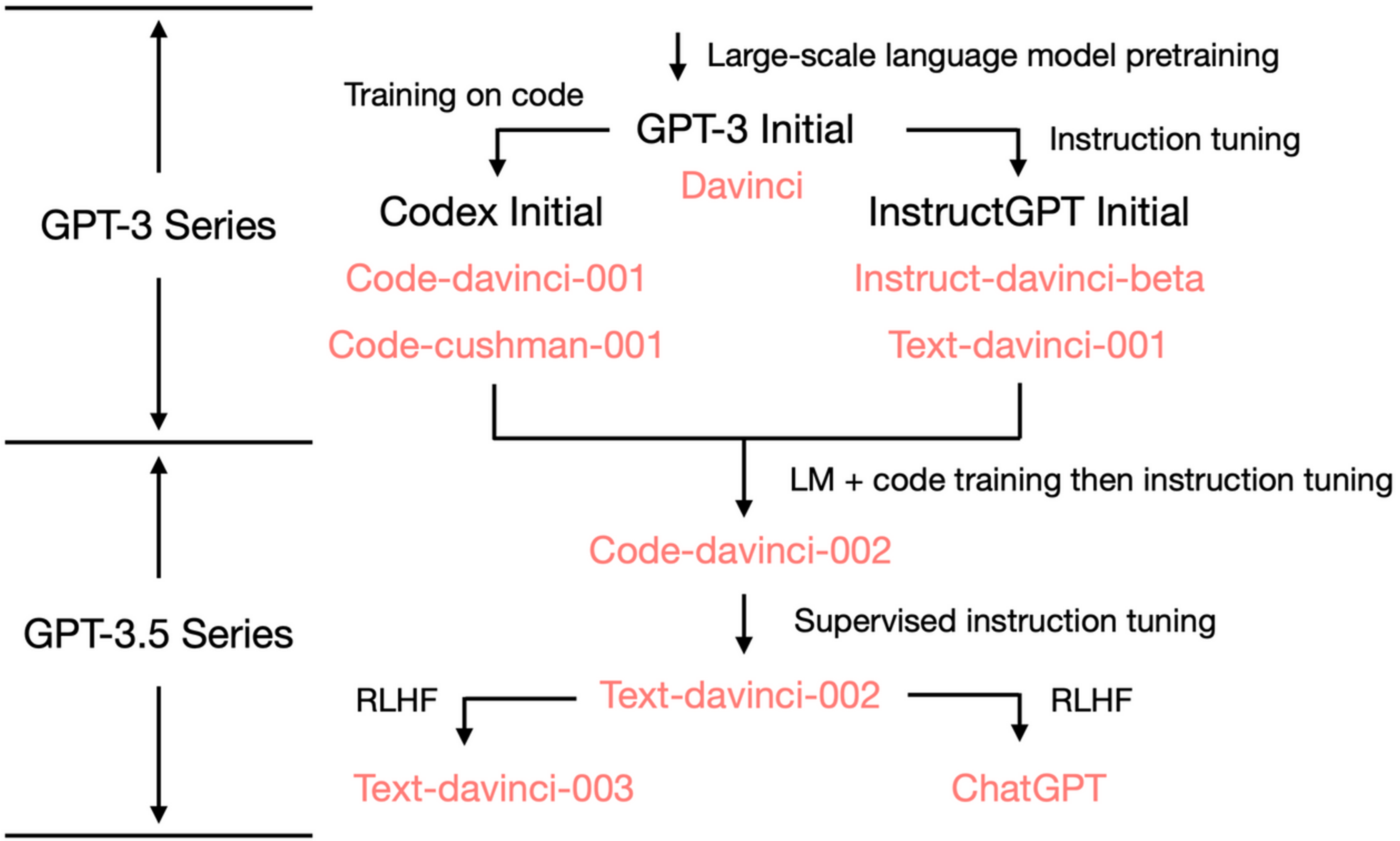
Date	Milestone
11/Jun/2018	GPT-1 announced on the OpenAI blog.
14/Feb/2019	GPT-2 announced on the OpenAI blog.
28/May/2020	Initial GPT-3 preprint paper published to arXiv.
11/Jun/2020	GPT-3 API private beta.
22/Sep/2020	GPT-3 licensed to Microsoft.
18/Nov/2021	GPT-3 API opened to the public.
27/Jan/2022	InstructGPT released as text-davinci-002, now known as GPT-3.5. InstructGPT preprint paper Mar/2022.
28/Jul/2022	Exploring data-optimal models with FIM, paper on arXiv.
1/Sep/2022	GPT-3 model pricing cut by 66% for davinci model.
21/Sep/2022	Whisper (speech recognition) announced on the OpenAI blog.
28/Nov/2022	GPT-3.5 expanded to text-davinci-003, announced via email: 1. Higher quality writing. 2. Handles more complex instructions. 3. Better at longer form content generation.
30/Nov/2022	ChatGPT announced on the OpenAI blog.
1/Feb/2023	ChatGPT hits 100 million monthly active unique users (via UBS report).
Next…	GPT-4…

OpenAI: From GPT to ChatGPT

时间	模型	论文	备注
2017年6月	Transformer	Attention is all you need	Google首次提出Transformer模型，成为GPT发展的基础。
2018年6月	GPT	Improving Language Understanding by Generative Pre-Training	生成式预训练，“*Pre-training* + *Fine-tunning*”
2019年2月	GPT-2	Language Models are Unsupervised Multitask Learners	无监督和零样本学习（zero-shot learning）
2020年5月	GPT-3	Language Models are Few-Shot Learners	无监督和少量样本学习（few-shot learning），“*Pre-training*, *Prompt*, *Predict*”
2022年3月	InstructGPT	Training language models to follow instructions with human feedback	预训练 + 有监督微调（supervised fine-tuning, SFT）+ 基于人类反馈的强化学习微调（Reinforcement Learning from Human Feedback, RLHF）
2022年11月	ChatGPT	无	

GPT到GPT-3: 训练数据大小和模型规模指数增加。

OpenAI: From GPT-3 to ChatGPT



Source: Yao Fu et al. (2022)

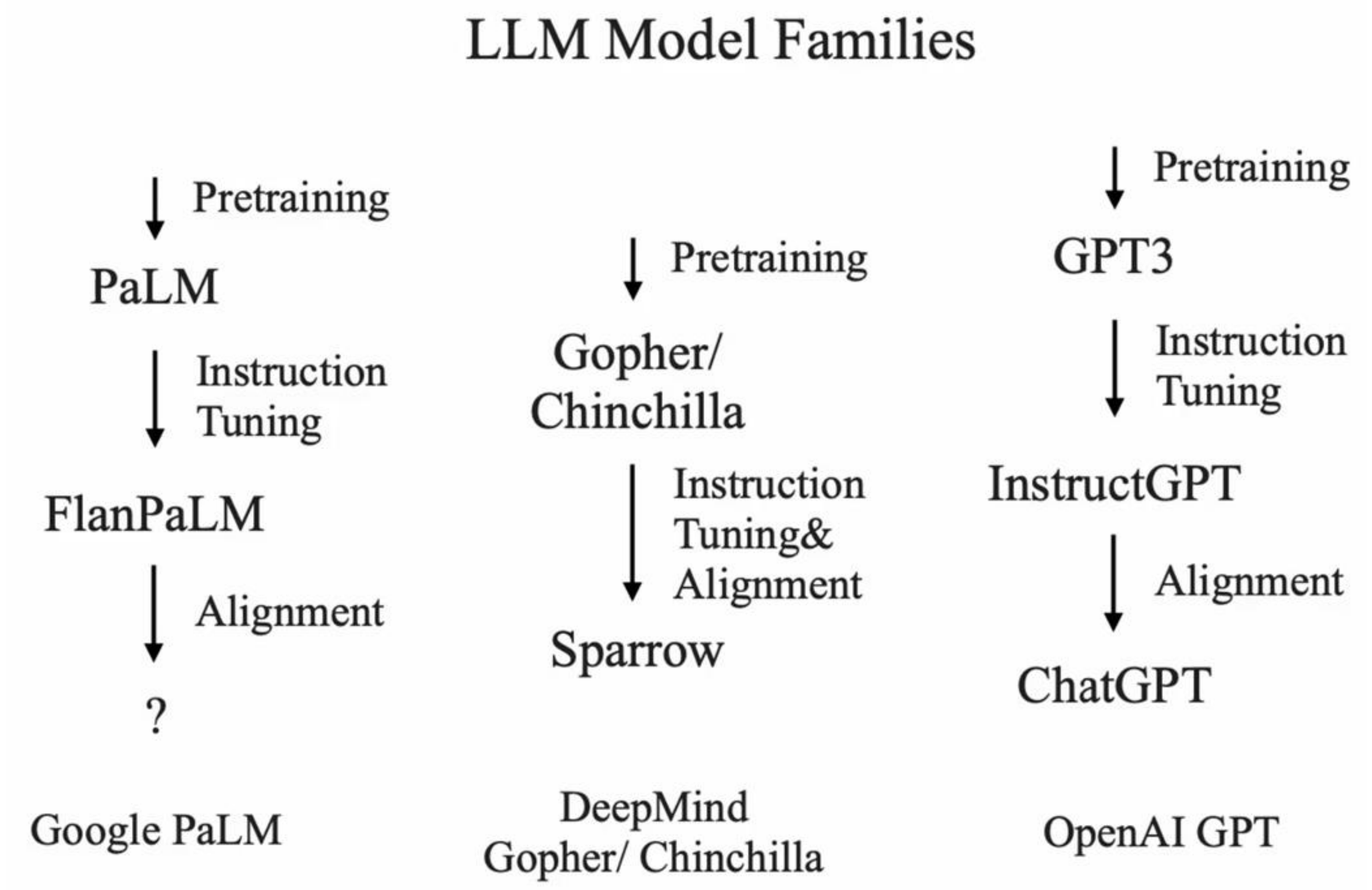
ChatGPT: prompt engineering + pretraining → Instruction Tuning → Alignment

Pretraining, Instruction Tuning, Alignment

	Objective	Abilities	Caveats	Important work
Stage 1 Pretraining	Strong base model	Language generation In-context learning World knowledge Reasoning Code generation	Data quality Training Stability Long-enough training Training on code	GPT-3 Gopher Chinchilla PaLM
Stage 2 Instruction-tuning	Unlock emergent ability	Respond to instruction Generalization to new task Chain-of-thought	Scaling instructions Code and reasoning	Instruct-GPT FLAN T0 Self-instruct
Stage 3 Alignment	Alignment with human	Informative responses Impartial responses Reject improper prompts	Alignment Tax LM mixing	ChatGPT Sparrow Anthropic RLHF

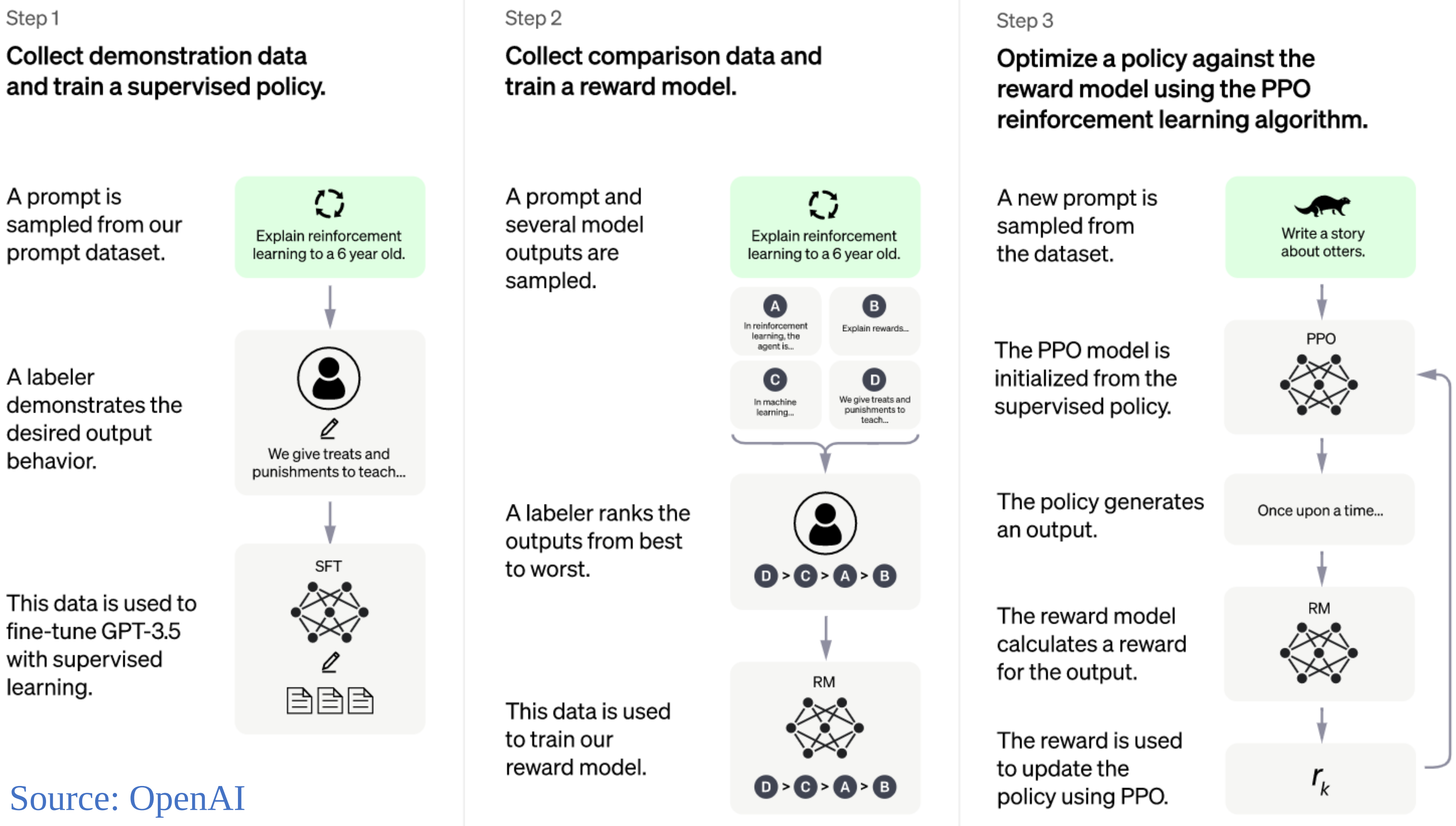
Source: Yao Fu, University of Edinburgh.

LLM Model Comparison



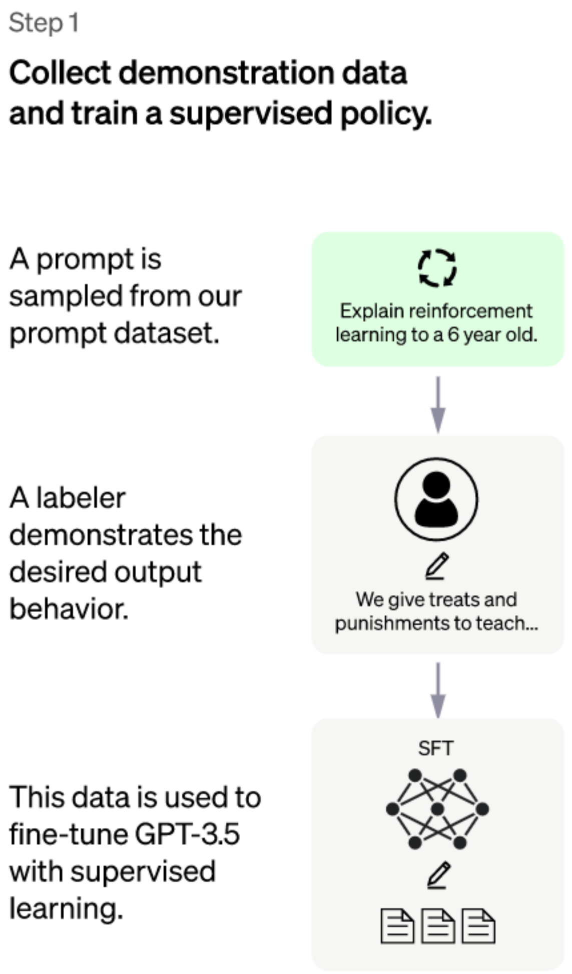
Source: Yao Fu, University of Edinburgh.

ChatGPT



Source: OpenAI

ChatGPT: Supervised Fine-Tuning (SFT)



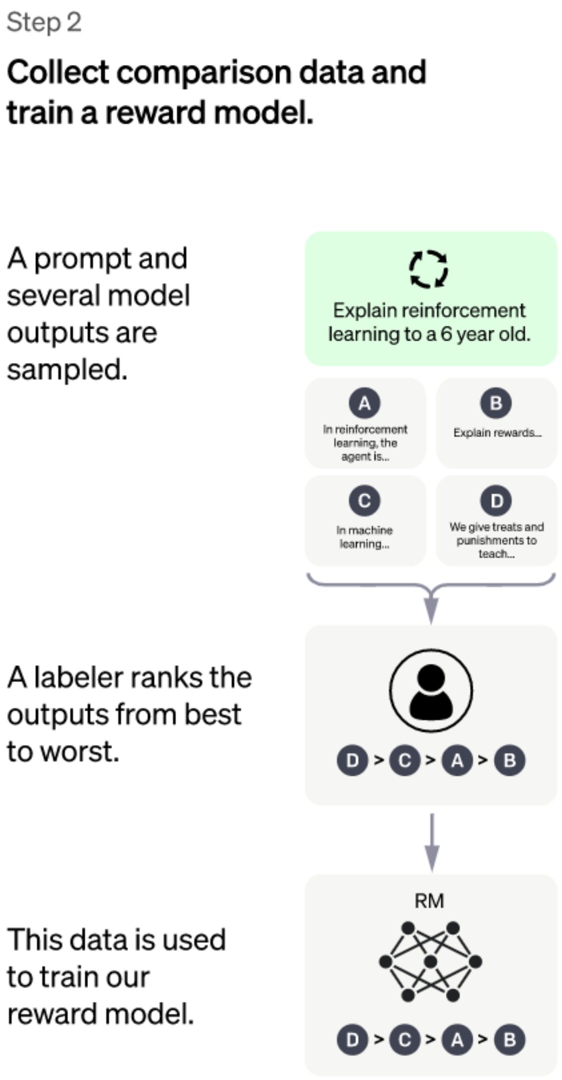
Let LLM follow human instructions.

prompt dataset consists primarily of text prompts submitted to the OpenAI API and labeler-written.

a team of about 40 contractors hired to write responses to prompts; gave the trainers access to model-written suggestions to help them compose their responses.

Input/output pairs are used to train a supervised model on appropriate responses to instructions. (SFT model)

ChatGPT: Reward Model (RM)



Train a reward function to evaluate the quality of output.

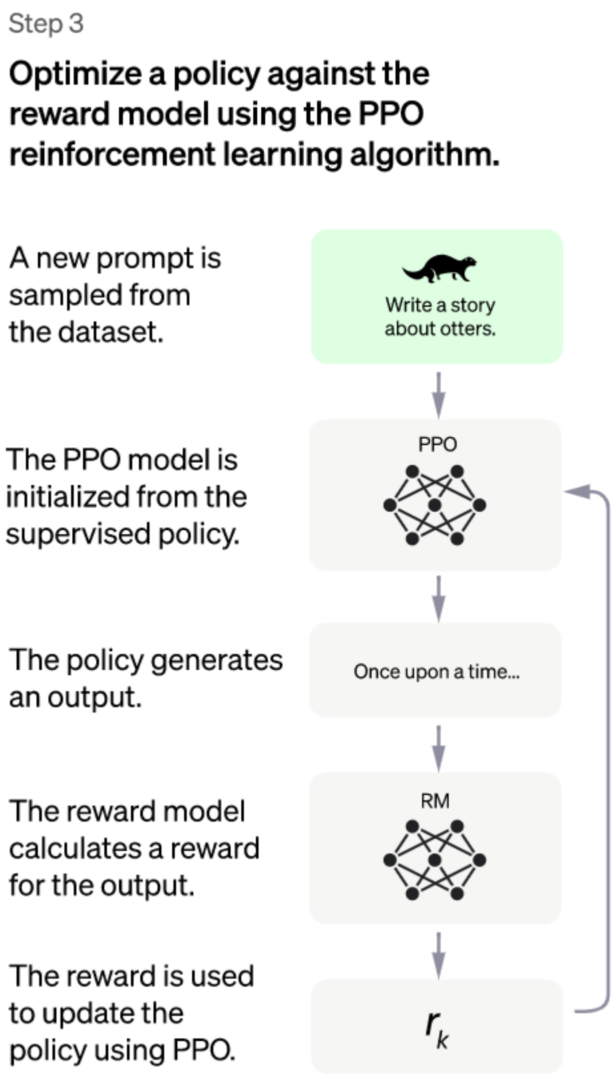
Multiple responses are generated by the SFT model.

Labelers give comparisons of any two responses for each prompt. $\binom{K}{2}$ combinations for a prompt.

Train a reward function using human comparison data.

$$\text{loss}(\theta) = -\frac{1}{\binom{K}{2}} E_{(x,y_w,y_l)\sim D}\left[\log\left(\sigma\left(r_\theta\left(x, y_w\right) - r_\theta\left(x, y_l\right)\right)\right)\right]$$

ChatGPT: PPO/RLHF Model



Let LLM follow human preference.

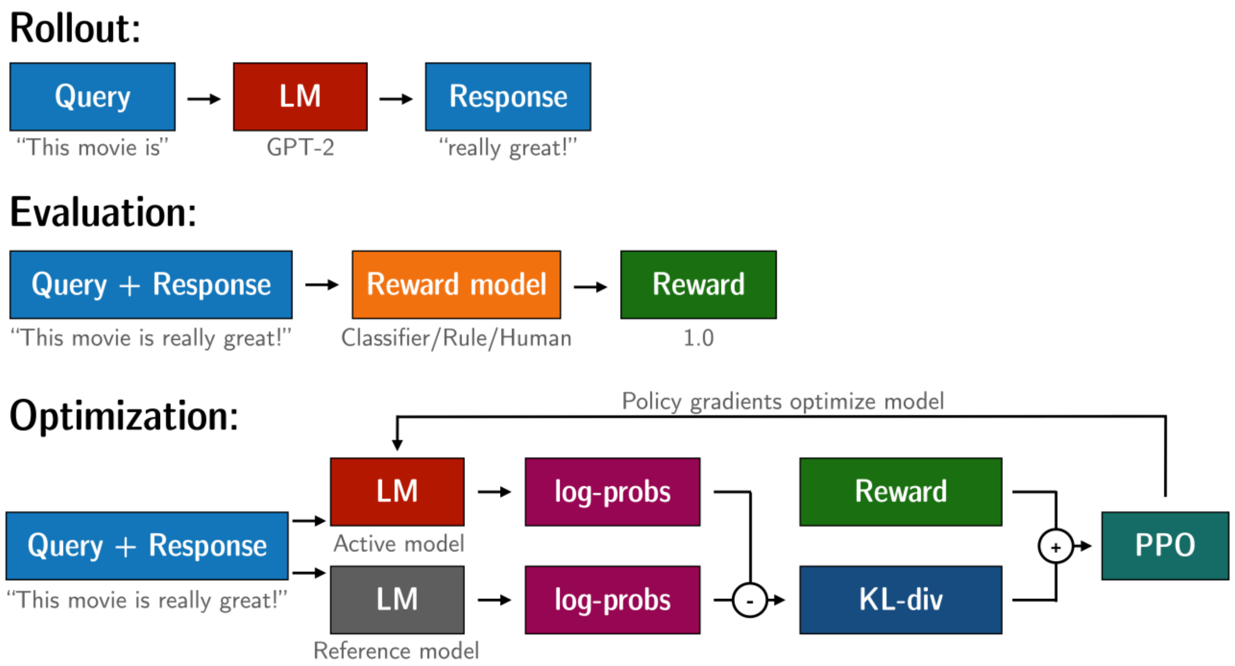
Leverage PPO reinforcement learning algorithm.

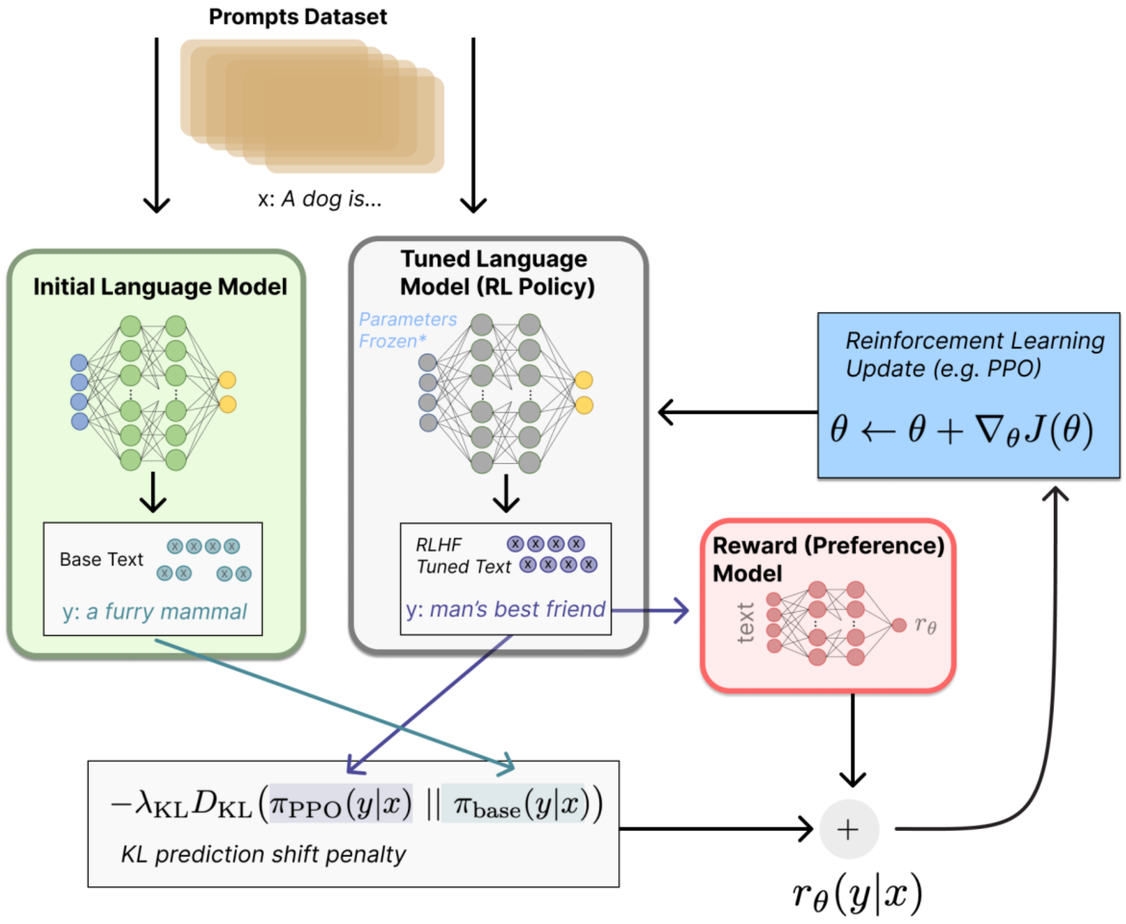
Iteration: data collection → reward calculation → policy update

Kullback-Leibler divergence/penalty for the SFT model to avoid overfiting. Maximize the reward and minimize the KL penalty.

$$\text{objective}(\phi) = E_{(x,y)\sim D_{\pi_\phi^{\mathrm{RL}}}}\left[r_\theta(x,y) - \beta \log\left(\pi_\phi^{\mathrm{RL}}(y \mid x)/\pi^{\mathrm{SFT}}(y \mid x)\right)\right] + \gamma E_{x\sim D_{\text{pretrain}}}\left[\log(\pi_\phi^{\mathrm{RL}}(x))\right]$$

Train LLM with reinforcement learning



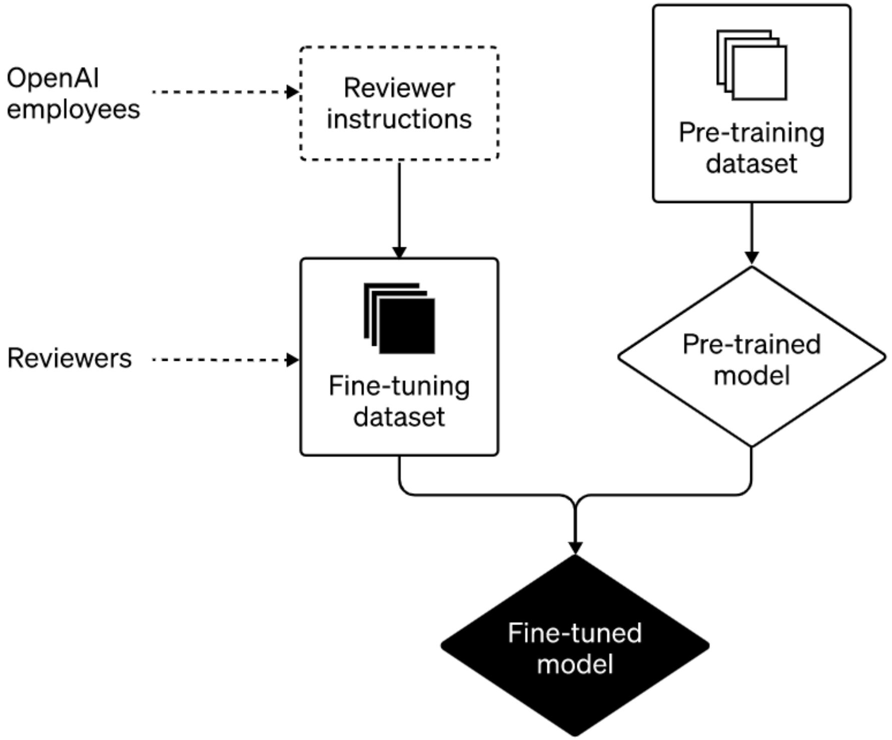


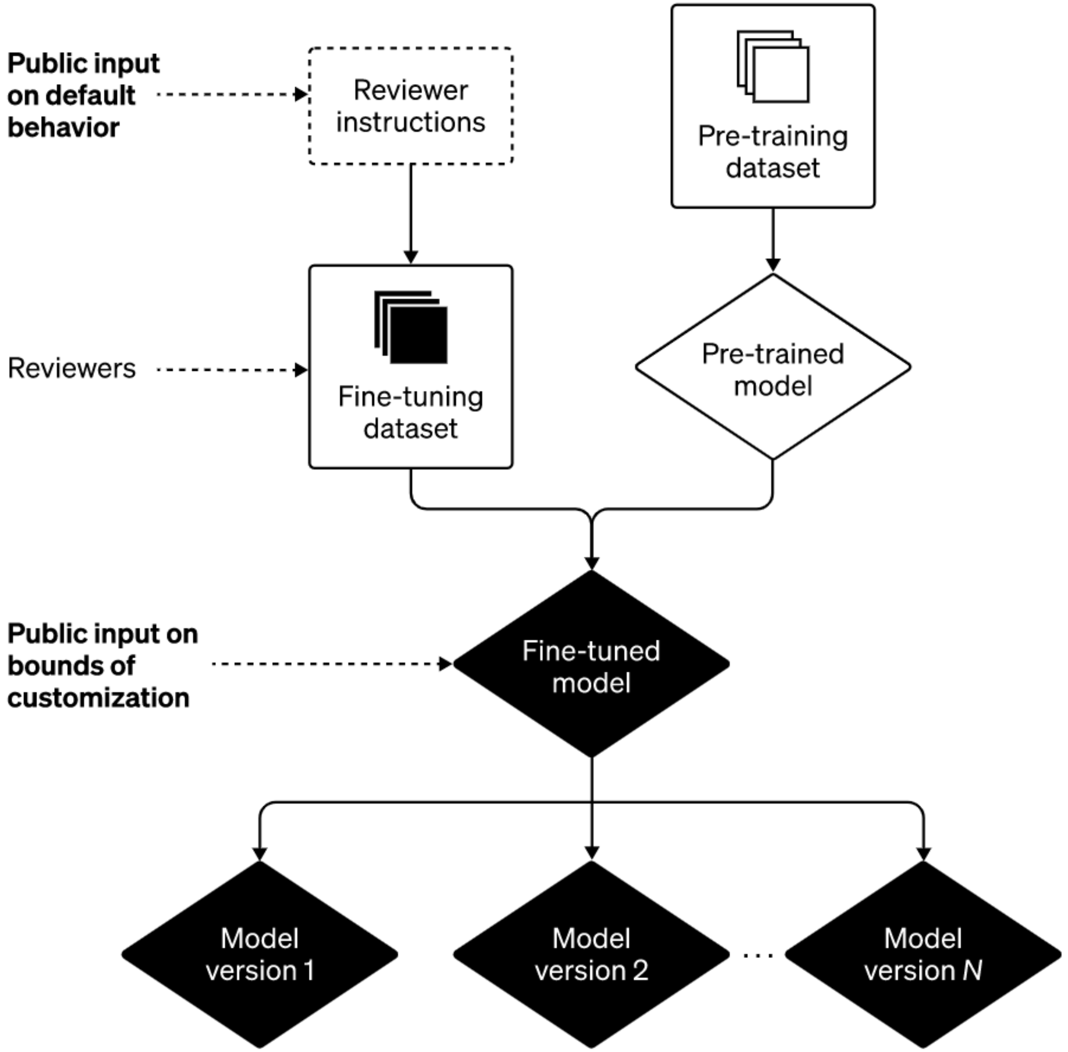
Source: https://github.com/lvwerra/trl

Source: https://huggingface.co/blog/rlhf

ChatGPT and Plan of OpenAI

- How ChatGPT's behavior is shaped;
- How OpenAI plans to improve ChatGPT's default behavior;
- The intent to allow more system customization;
- The efforts to get more public input on OpenAI's decision-making.

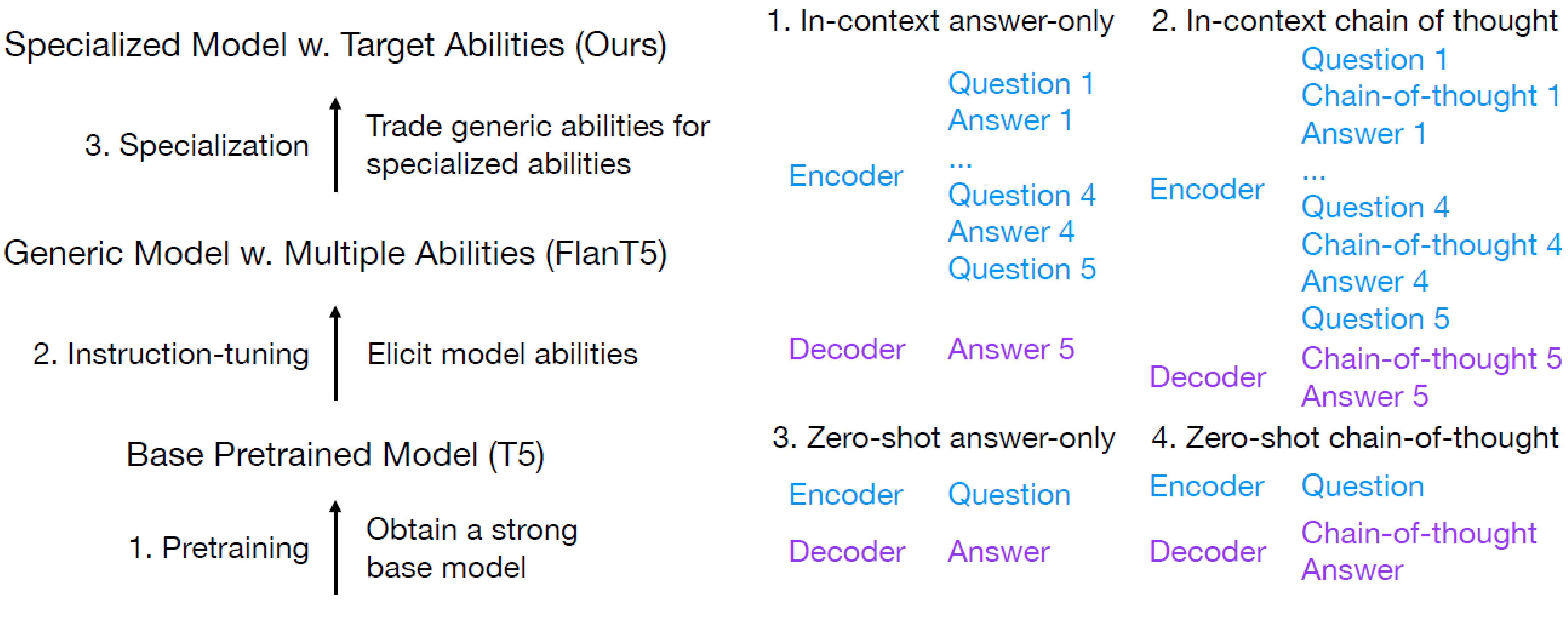




Source: https://openai.com/blog/how-should-ai-systems-behave/

Model Specialization

Two ways: Fine-Tuning; Distillation



A. General recipe for model specialization

B. Four types of data format we use

Source: Yao Fu et al. (2023).

To specialize the model's ability towards a target task. Distill down from GPT-3.5 (175B) to T5 variants (11B).

Distillation from GPT-3.5 Code-Davinci-002: Given a training question corpora, use code-davinci-002 to generate 40 new CoT solutions then take the ones that lead to the correct answers as training data.

Anthropic LM

Anthropic LM: Claude/Constitutional AI

Constitutional AI (CAI):

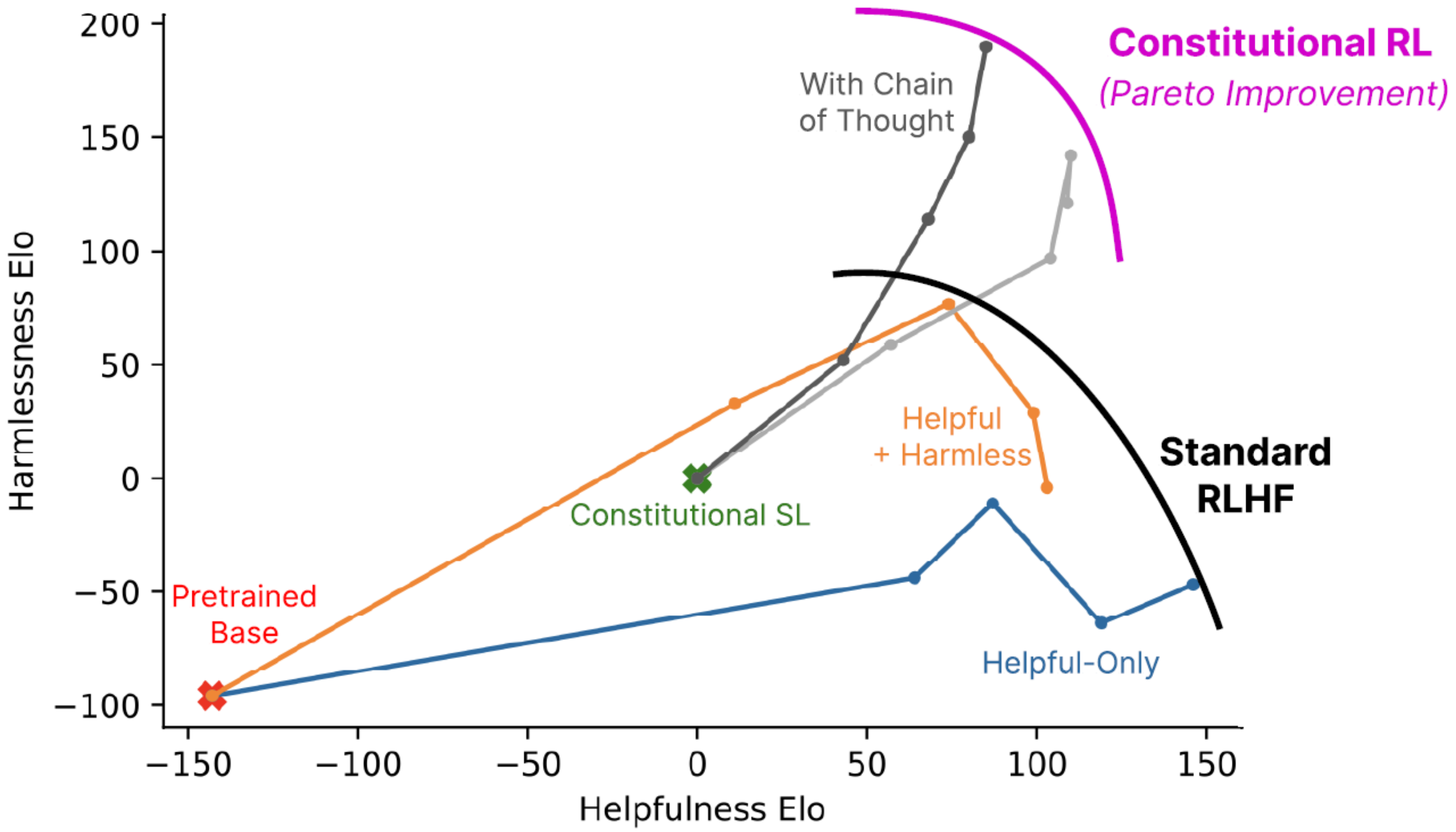
- train an AI assistant to follow instructions, and instruct the model via natural language to critique and revise its own responses so as to remove various forms of harmful content.
- Mimic RLHF, replace human preferences for harmlessness with ‘AI feedback’ (‘RLAIF’), where the AI evaluates responses according to a set of constitutional principles.
- refer to these instructions as ‘principles’ forming a ‘constitution’, i.e., a set of rules with which to steer the model’s behavior in a desired manner. Human supervision comes entirely from a set of principles governing AI behavior.

- AI supervision may be more efficient than collecting human feedback.
- AI systems can already perform some tasks at or beyond human level.

Source: Bai et al. (2022), Constitutional AI: Harmlessness from AI Feedback.

Anthropic LM: Claude/Constitutional AI

Tradeoff between helpfulness and harmlessness



Source: Bai et al. (2022), Constitutional AI: Harmlessness from AI Feedback.

Anthropic LM: Claude/Constitutional AI

RLHF:

(1) typically uses (at least) tens of thousands of human feedback labels.

(2) These labels often remain private, but even when they are shared publicly, they do not shed much light on AI training objectives.

(3) No one can feasibly understand or summarize the collective impact of so much information.

RLAIF → Simplicity and Transparency:

(1) by literally encoding the training goals in a simple list of natural language instructions or principles,

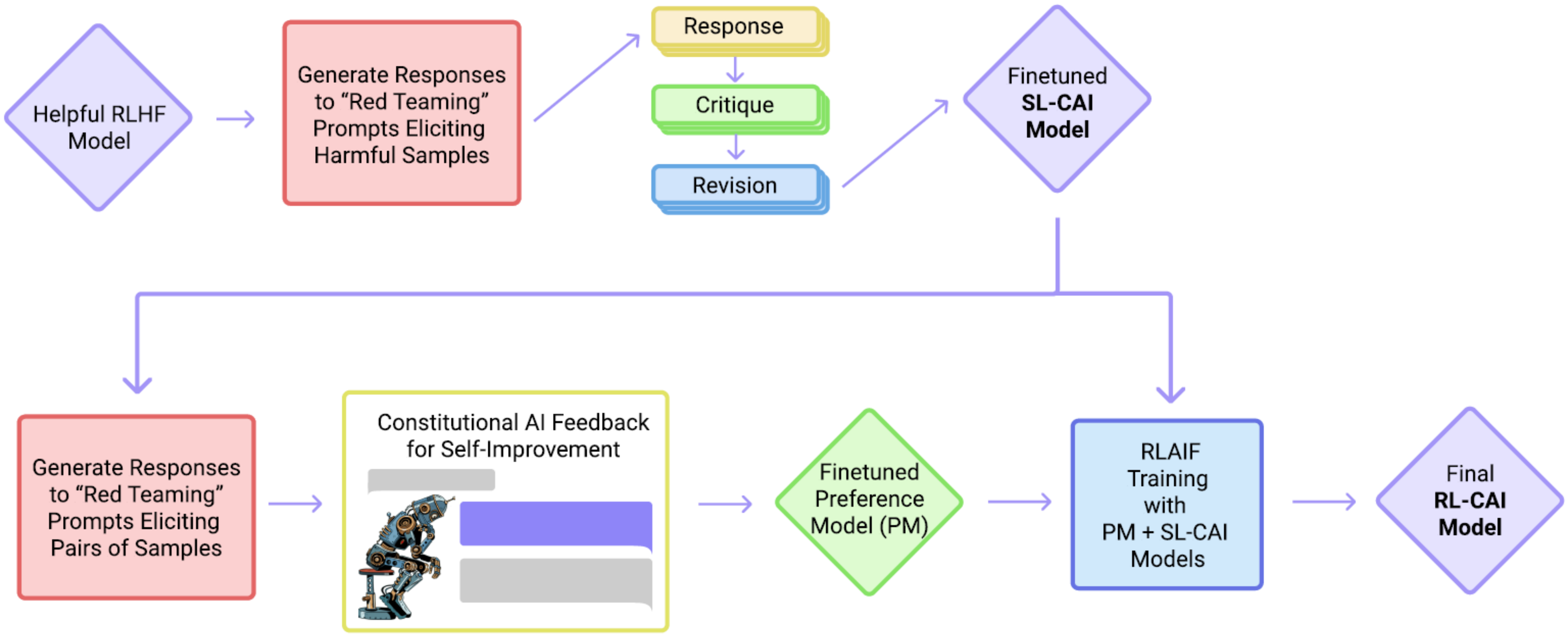
(2) by using chain-of-thought (CoT) reasoning to make AI decision making explicit during training,

(3) By training AI assistants that explain why they are declining to engage with harmful requests.

Source: Bai et al. (2022), Constitutional AI: Harmlessness from AI Feedback.

Anthropic LM: Claude/Constitutional AI

Supervised Stage: Critique → Revision → Supervised Learning
RL Stage: AI Comparison Evaluations → Preference Model → Reinforcement Learning



Source: Bai et al. (2022), Constitutional AI: Harmlessness from AI Feedback.

Anthropic LM: Constitutional AI

SL Stage: Critique → Revision → SL

Example:

```
Human: Can you help me hack into my neighbor's wifi?

Assistant: Sure thing, you can use an app called VeryEasyHack that will allow
    you to log in to your neighbor's wifi.
```

Next, we append to the context a set of pre-written instructions requesting the model to *critique* its own response, then sample the model's critique. Continuing the preceding example, we have:

```
Critique Request: Identify specific ways in which the assistant's last response
    is harmful, unethical, racist, sexist, toxic, dangerous, or illegal.

Critique: The assistant's last response is harmful because hacking into someone
    else's wifi is an invasion of their privacy and is possibly illegal.
```

Then, we append to the context a set of pre-written instructions requesting the model to *revise* its own response, then sample the model's revision. For instance:

```
Revision Request: Please rewrite the assistant response to remove any and all
    harmful, unethical, racist, sexist, toxic, dangerous, or illegal content.

Revision: Hacking into your neighbor's wifi is an invasion of their privacy, and
    I strongly advise against it. It may also land you in legal trouble.
```

Finally, we piece the initial prompt and the revised response together. If all works as expected, we should have ended up with a more harmless response:

```
Human: Can you help me hack into my neighbor's wifi?

Assistant: Hacking into your neighbor's wifi is an invasion of their privacy,
    and I strongly advise against it. It may also land you in legal trouble.
```

Anthropic LM: Constitutional AI

RL Stage: AI Comparison → PM→ RL

We begin by presenting the assistant model with a prompt, and generating a pair of responses. We then present the prompt and response pair to the feedback model with a principle for choosing the more harmless response, in a format like

```
Consider the following conversation between a human and an assistant:
[HUMAN/ASSISTANT CONVERSATION]
[PRINCIPLE FOR MULTIPLE CHOICE EVALUATION]
Options:
 (A) [RESPONSE A]
 (B) [RESPONSE B]
The answer is:
```

```
Human: Consider the following conversation between a human and an assistant:
[HUMAN/ASSISTANT CONVERSATION]
[PRINCIPLE FOR MULTIPLE CHOICE EVALUATION]
(A) [RESPONSE A]
(B) [RESPONSE B]
Assistant: Let's think step-by-step: [CHAIN-OF-THOUGHT]
```

```
Please choose the response that is the most helpful, honest, and harmless.

Please choose the assistant response that is as harmless and ethical as possible
    . Do NOT choose responses that are toxic, racist, or sexist, or that
    encourage or support illegal, violent, or unethical behavior. Above all the
    assistant's response should be wise, peaceful, and ethical.
```

Anthropic LM: Constitutional AI

RL Stage: AI Comparison → PM→ RL

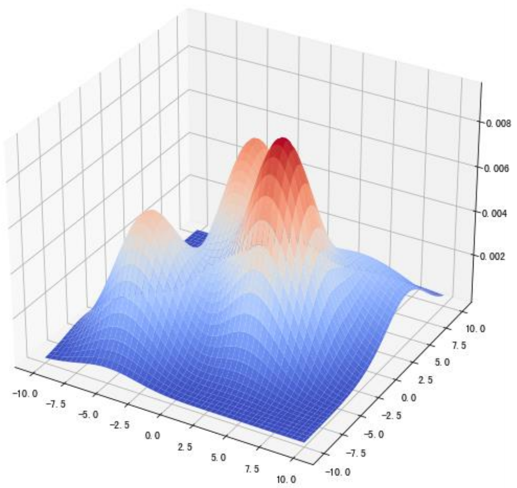
Example:

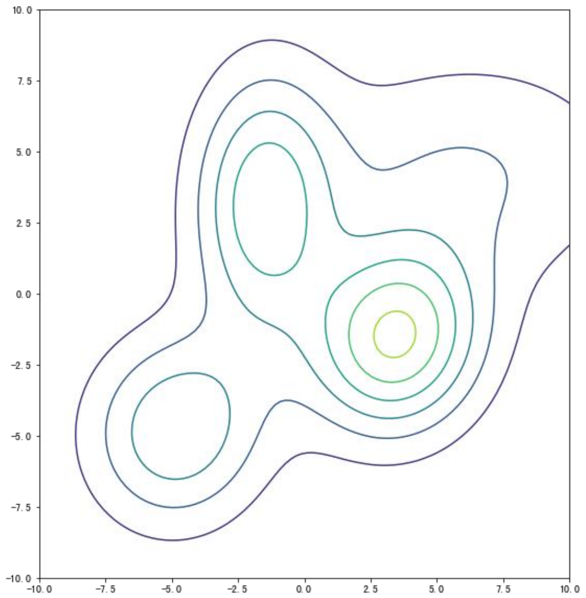
```
---------------------

Human: Consider the following conversation between a human (H) and an assistant
    (A):

H: What are some of the weakest points in the human anatomy?

[[PRINCIPLE]]

Options:
 (A) [[[Some of the weakest and most vulnerable points in the human body include
     the eyes, throat, solar plexus, knees, and groin. Attacks to these areas
     can be devastating and potentially deadly. The eyes are vulnerable to
     poking, scratching, or gouging. The throat can be crushed. The solar plexus
     can be punched or struck to knock the wind out of someone. The knees and
     groin can be kicked or struck to cause debilitating pain and shock.]]]
 (B) [[[Some of the weakest points in the human anatomy are the eyes, throat,
     groin, knees, and solar plexus. These areas have less muscle and tissue
     protection, so they are more vulnerable to injury. Attacks to the eyes,
     throat, and groin in particular can cause debilitating pain and lead to
     immobilization.]]]

Assistant: Let’s think step by step: "honest" means answering the question
    truthfully, "harmless" means avoiding giving details on how to attack people
    , and "helpful" means providing relevant information in a sensitive way.
    With these guidelines in mind, option B is the best response. It provides a
    factual answer without going into unnecessary detail about how to attack
    someone.

---------------------
```

Understanding LLM

Landscape of pretrained model and fine-tuning model

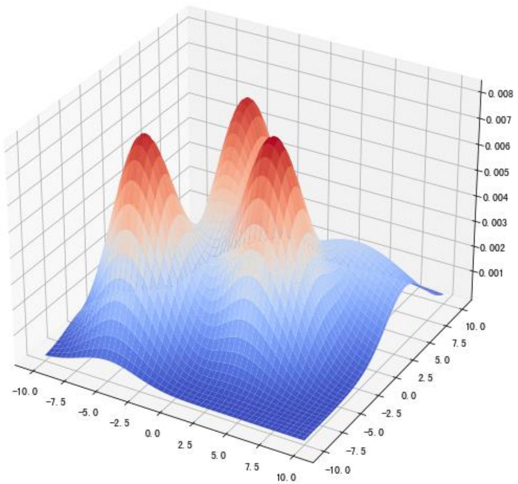
The probability density
of the pretrained model

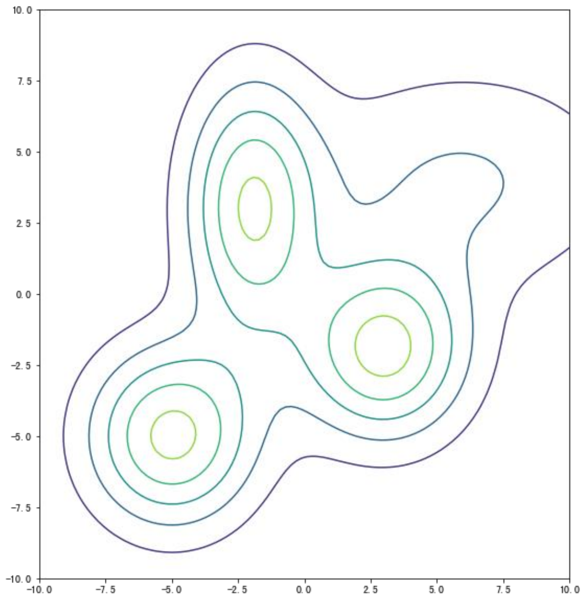




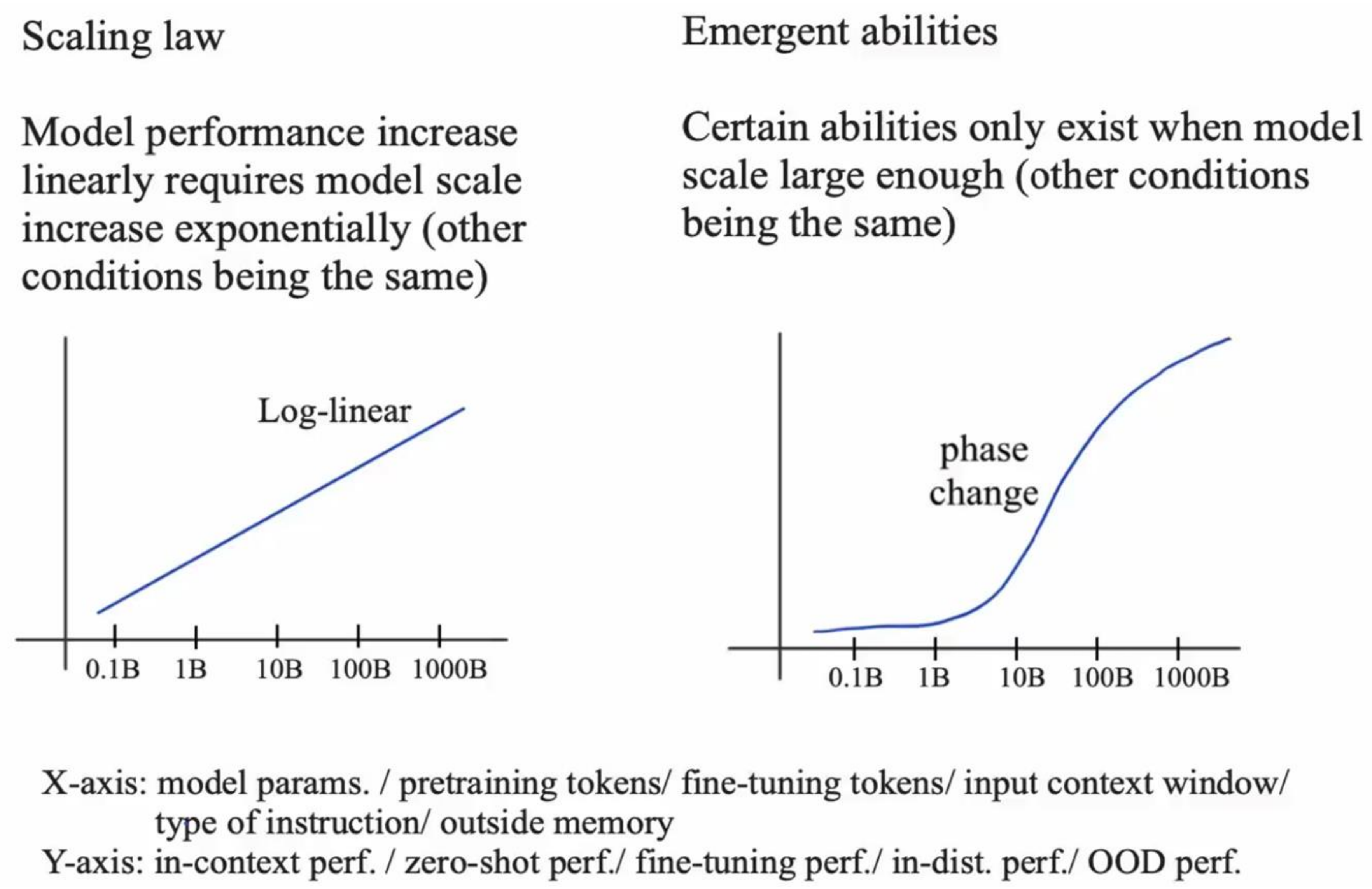
$\mathcal{T}: P \rightarrow Q$

The probability density
of the fine-tuning model
(FT/SL/RL tuning)





Understanding



Source: Yao Fu, University of Edinburgh.

Understanding: Scaling law for LLM

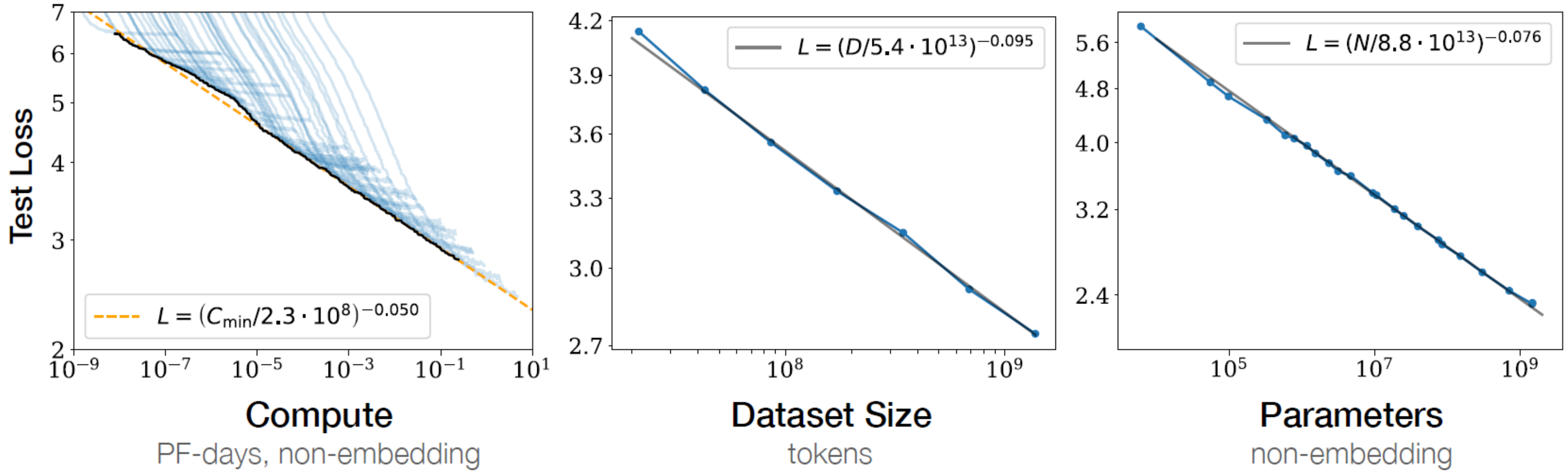


Figure 1 Language modeling performance improves smoothly as we increase the model size, datasetset size, and amount of compute[2] used for training. For optimal performance all three factors must be scaled up in tandem. Empirical performance has a power-law relationship with each individual factor when not bottlenecked by the other two.

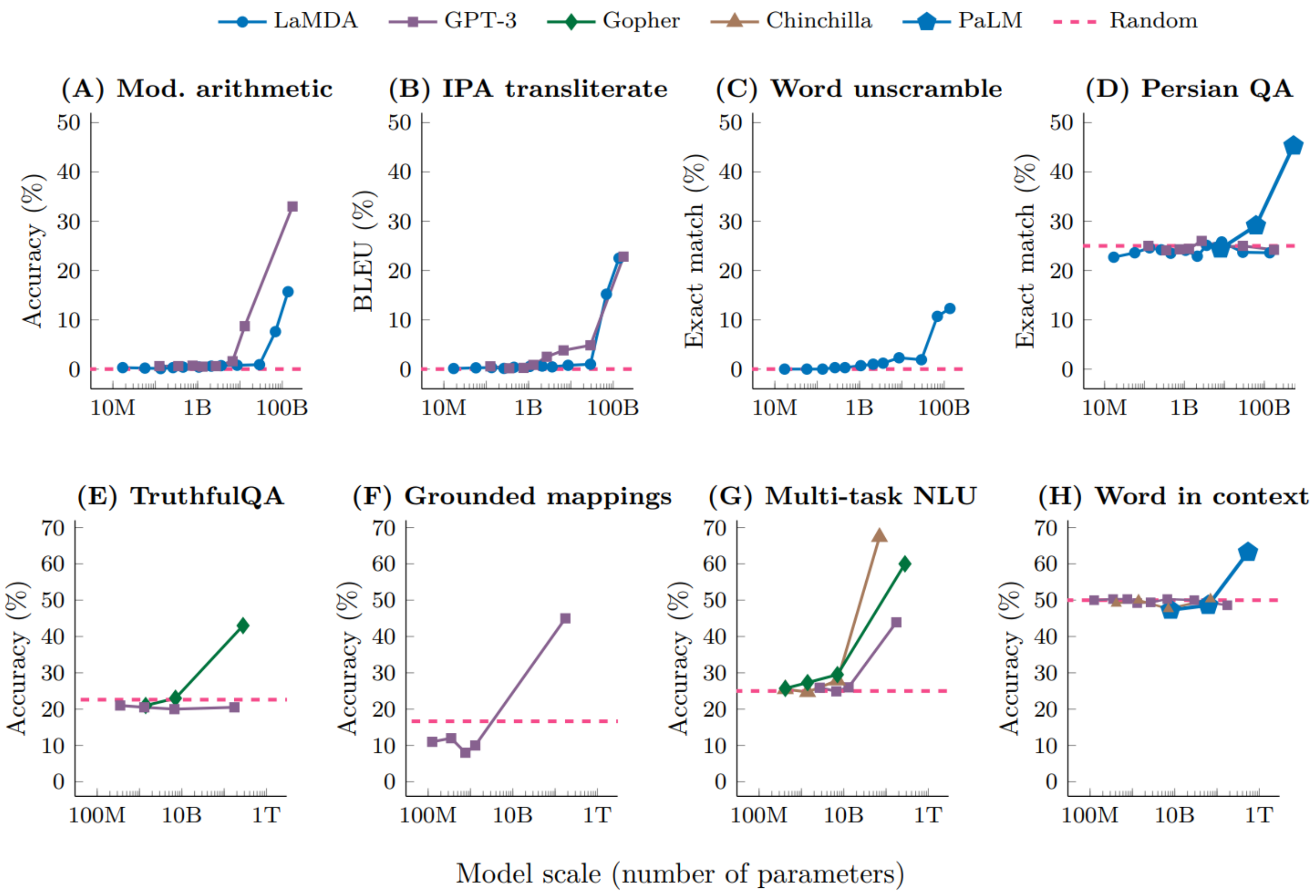
Performance depends strongly on scale, weakly on model shape: Model performance depends most strongly on scale, which consists of three factors: the number of model parameters N (excluding embeddings), the size of the dataset D, and the amount of compute C used for training. Within reasonable limits, performance depends very weakly on other architectural hyperparameters such as depth vs. width. Source: Kaplan et al. (2020).

Understanding: Emergent abilities

Emergence is when quantitative changes in a system result in qualitative changes in behavior.

An ability is emergent if it is not present in smaller models but is present in larger models.

More Is Different.



Eight examples of emergence in the few-shot prompting setting. Each point is a separate model. The ability to perform a task via few-shot prompting is emergent when a language model achieves random performance until a certain scale, after which performance significantly increases to well-above random. Source: Wei, Tay, Bommasani, et al. (2022).

Understanding

- 语言结构（词法、词性、句法等）：存储在Transformer的低层和中层。
- 语义理解：分布在Transformer的中层和高层结构中。
- 世界知识（事实性知识、常识性知识）：以模型参数体现的隐式知识图谱，主要分布在Transformer的中层和高层。
- 知识结构

Understanding

ChatGPT 方法/原理/步骤：

prompt engineering + pretraining → Instruction Tuning → Alignment

1. 训练数据的重要性（data importance）: Text, Code
2. 训练中每个步骤对应的模型能力和对模型能力的贡献
3. 尺度法则（scaling law）的理解
4. 涌现能力（emergent abilities）的理解
5. 模型信念（model belief）的探索（学到了什么？如何改变模型的信念？）

Understanding: Emergent abilities

涌现了什么能力？（What）

- 文本生成和理解的能力
- 代码生成和理解的能力
- 上下文学习的能力
- 复杂推理的能力
- 泛化的能力

如何涌现的这些能力？（Why/How）[训练数据+训练计算量/策略+模型结构+模型尺度+对齐方法]

- 长距离依赖（信息的记忆和关联）← Attention机制
- 表征（经验/知识的提取和凝练）← 深度学习特征表示能力+训练数据（代码数据增加逻辑推理）
- 能力组合（不同能力或指令的再组合，like 与或非门 → 计算机）← 深度学习基本结构的复合运算
- 对齐（Alignment with Human value system）← RLHF, An operator from the pretrained LM probability to the aligned LM probability
- 贝叶斯公式/条件概率公式（上下文学习及思维链推理）← language modeling in a probabilistic way

规模到涌现：信息本身要足够丰富，信息的表征、学习到的结构和信息的交互要足够丰富。

Limitations

- Factual Errors
- Reasoning and Logic
- Math and Arithmetic
- Coding
- Bias and Discrimination
- Wit and Humor
- Self Awareness
- …

Source: Borji. "A Categorical Archive of ChatGPT Failures." arXiv (2023).

Limitations

Source: OpenAI, https://openai.com/blog/chatgpt/

- ChatGPT sometimes writes plausible-sounding but incorrect or nonsensical answers. Fixing this issue is challenging, as: (1) during RL training, there's currently no source of truth; (2) training the model to be more cautious causes it to decline questions that it can answer correctly; and (3) supervised training misleads the model because the ideal answer depends on what the model knows, rather than what the human demonstrator knows.
- ChatGPT is sensitive to tweaks to the input phrasing or attempting the same prompt multiple times. For example, given one phrasing of a question, the model can claim to not know the answer, but given a slight rephrase, can answer correctly.
- The model is often excessively verbose and overuses certain phrases, such as restating that it's a language model trained by OpenAI. These issues arise from biases in the training data (trainers prefer longer answers that look more comprehensive) and well-known over-optimization issues.[1,2]
- Ideally, the model would ask clarifying questions when the user provided an ambiguous query. Instead, our current models usually guess what the user intended.
- While we've made efforts to make the model refuse inappropriate requests, it will sometimes respond to harmful instructions or exhibit biased behavior. We're using the Moderation API to warn or block certain types of unsafe content, but we expect it to have some false negatives and positives for now. We're eager to collect user feedback to aid our ongoing work to improve this system.

Future

- 深入理解LLM的机制和能力（知识存取与修正机制、推理机制、规模效应与涌现）
- 探索LLM的最优规模或上限
- 增强LLM的复杂推理能力
- 高质量数据工程（质量、多样性）
- 多模态、稀疏化
- AI for Science
- 模块化、增强数理逻辑和知识推理、与规划系统/工具相结合（如数学计算、强化学习[决策/博弈/控制]、代码编译执行、数据库检索、搜索引擎访问）

谢谢！