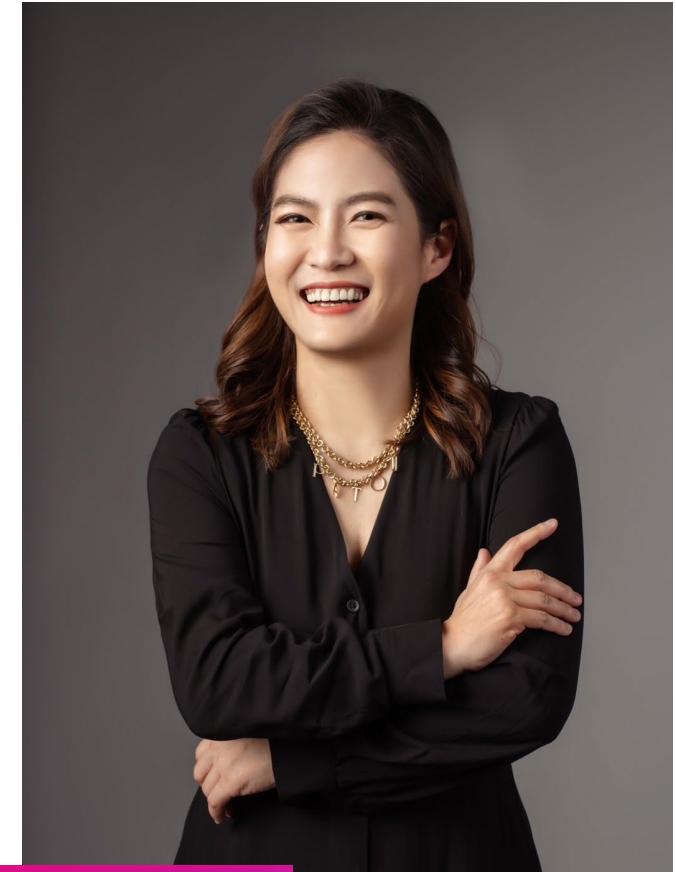


大型語言模型

中央研究院 古倫維

關於我

- ✓ 中央研究院資訊科學所-研究員 since 2012
- ✓ ACL 2024 Program Chair 技術主席
- ✓ Woman in AI Award APAC Finalist
- ✓ 自然語言處理研究 since 1998



@lunweiku 或搜尋 古倫維



古老師研究室



關於中央研究院之 自然語言處理與情感分析實驗室

研究主題：

- ✓ 人工智慧與知識圖譜
- ✓ 假新聞/不實資訊 (干預, 免疫, 高峰預測, 澄清新聞改寫, ...等等)
- ✓ 視覺語言融合技術 (看圖說故事, 機器人視覺認知, 吸眼球視覺問題生成, ...等等)
- ✓ 情感分析 (互動式監督學習推特資料, 謊言偵測, 細微情感分析, ...等等)
- ✓ 可解釋性人工智慧 (假新聞解釋, 推薦原因解釋, 個性化對話)
- ✓ 大型語言模型之調適及前後端應用開發
- ✓ 沉浸式藝術體驗

課前提問

對於這個主題特別想要了解的部分

能夠如何利用開源的LLM進行AI project

Secrets behind ChatGPT web page (some private API or hidden prompts other than public OpenAI API calls)

LLM進展迅速，已經到GPT-4了，希望了解目前在業界，有哪些應用面以及成熟度，例如coding，documentation，系統的運用。

1. 如何有系統性的對chatGPT做prompt tuning?
2. general guideline 如何判斷哪些題目適合使用chatGPT, 哪些適合自己造AI model

Recap

語言模型

什麼是語言模型？



GPT

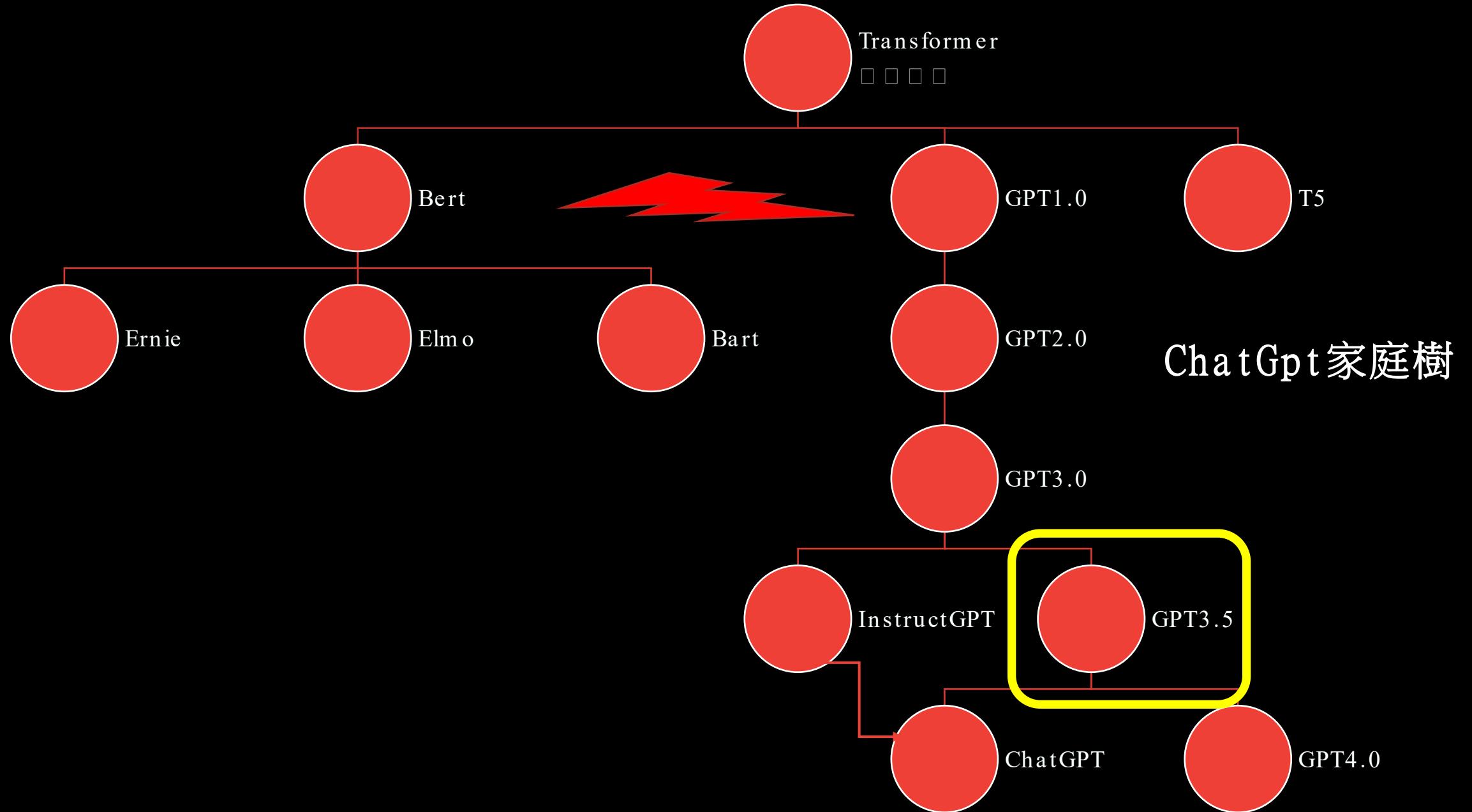
- Improving Language Understanding by Generative Pre-Training (2018)
- From OpenAI
- Transformer decoder

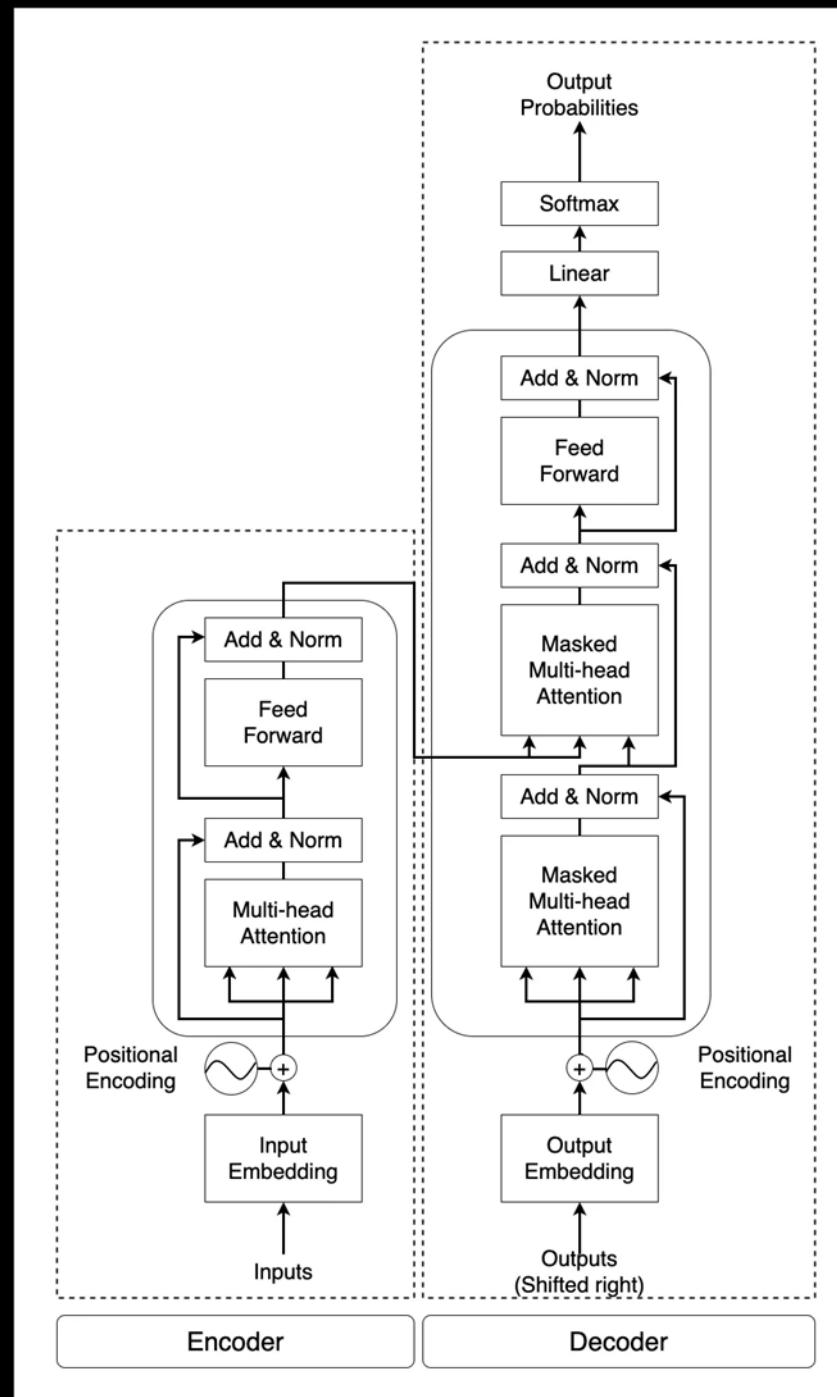
$$L_1(\mathcal{U}) = \sum_i \log P(u_i | u_{i-k}, \dots, u_{i-1}; \Theta)$$

$$h_0 = UW_e + W_p$$

$$h_l = \text{transformer_block}(h_{l-1}) \forall i \in [1, n]$$

$$P(u) = \text{softmax}(h_n W_e^T)$$





GPT series evolvement

GPT vs. GPT2

- Pretraining helps generalization
- Size (of pretraining) matters (for downstream applications)
- Foundation model and fine-tuning

GPT2 vs. GPT3

- GPT-2 has 1.5 billion parameters while GPT-3 has 175 billion.

Zero Shot, Few Shot, Fine-tunning

Zero Shot:

- Translate English to Chinese:
- Hamster =>

One Shot:

- Translate English to Chinese:
- Rabbit => 兔子
- Hamster =>

Few Shot:

- Translate English to Chinese:
- Rabbit => 兔子
- Turtle => 烏龜
- Guinea Pig => 天竺鼠
- Hamster =>

Fine-tunning

- Rabbit => 兔子

Gradient Decent

- Turtle => 烏龜

Gradient Decent

- Guinea Pig => 天竺鼠

Gradient Decent

-

- Hamster =>

GPT: Fine Tuning

- Supervised fine tuning
- Linear output layer

$$P(y|x^1, \dots, x^m) = \text{softmax}(h_l^m W_y).$$

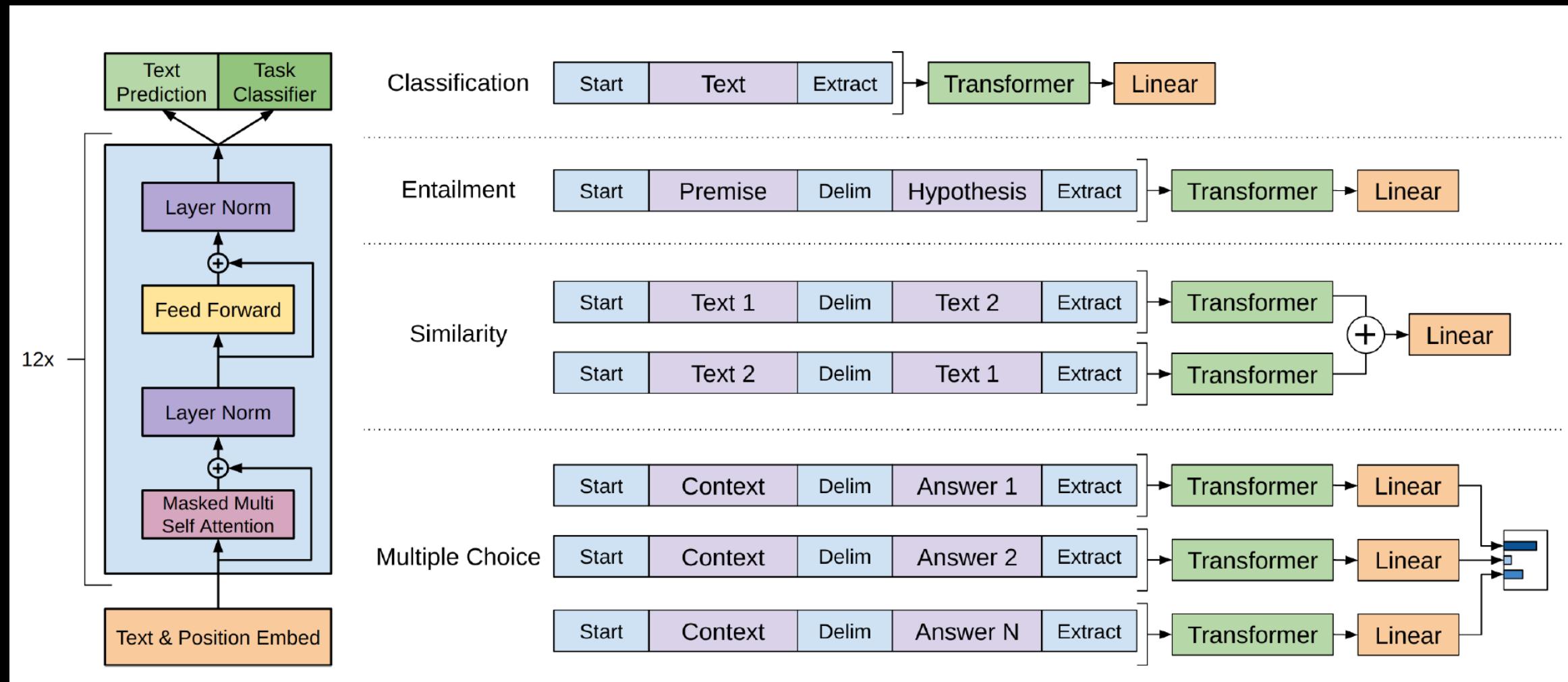
- Objective

$$L_2(\mathcal{C}) = \sum_{(x,y)} \log P(y|x^1, \dots, x^m).$$

- Including language modeling as auxiliary objective

$$L_3(\mathcal{C}) = L_2(\mathcal{C}) + \lambda * L_1(\mathcal{C})$$

GPT: transformer architecture and fine-tuning tasks



Evaluation Tasks

- 12 layers
- $d_m = 768$, $d_{ff} = 3,072$, token = 512, BPE (byte pair encoding)

| Task | Datasets |
|----------------------------|---|
| Natural language inference | SNLI [5], MultiNLI [66], Question NLI [64], RTE [4], SciTail [25] |
| Question Answering | RACE [30], Story Cloze [40] |
| Sentence similarity | MSR Paraphrase Corpus [14], Quora Question Pairs [9], STS Benchmark [6] |
| Classification | Stanford Sentiment Treebank-2 [54], CoLA [65] |

GLUE General-Purpose Language Understanding

單句任務跟雙句任務

- 文法接受
- 情感分析
- 換句話說
- 推理: 前提 – 假設 (蘊含 矛盾 中立)
- NLI RTE SQuAD

.....

- 對話任務呢？？？

Results

| Method | Classification | | Semantic Similarity | | GLUE | |
|---------------------------------------|----------------|---------------|---------------------|--------------|-------------|-------------|
| | CoLA (mc) | SST2 (acc) | MRPC (F1) | STSB (pc) | QQP (F1) | |
| Sparse byte mLSTM [16] | - | 93.2 | - | - | - | - |
| TF-KLD [23] | - | - | 86.0 | - | - | - |
| ECNU (mixed ensemble) [60] | - | - | - | <u>81.0</u> | - | - |
| Single-task BiLSTM + ELMo + Attn [64] | <u>35.0</u> | 90.2 | 80.2 | 55.5 | <u>66.1</u> | 64.8 |
| Multi-task BiLSTM + ELMo + Attn [64] | 18.9 | 91.6 | 83.5 | 72.8 | 63.3 | <u>68.9</u> |
| Finetuned Transformer LM (ours) | 45.4 | 91.3 | 82.3 | 82.0 | 70.3 | 72.8 |

Zero-shot Results

| | LAMBADA (PPL) | LAMBADA (ACC) | CBT-CN (ACC) | CBT-NE (ACC) | WikiText2 (PPL) | PTB (PPL) | enwik8 (BPB) | text8 (BPC) | WikiText103 (PPL) | 1BW (PPL) |
|------|------------------|------------------|-----------------|-----------------|--------------------|--------------|-----------------|----------------|----------------------|--------------|
| GPT | SOTA | 99.8 | 59.23 | 85.7 | 82.3 | 39.14 | 46.54 | 0.99 | 1.08 | 18.3 |
| | 117M | 35.13 | 45.99 | 87.65 | 83.4 | 29.41 | 65.85 | 1.16 | 1.17 | 37.50 |
| | 345M | 15.60 | 55.48 | 92.35 | 87.1 | 22.76 | 47.33 | 1.01 | 1.06 | 26.37 |
| | 762M | 10.87 | 60.12 | 93.45 | 88.0 | 19.93 | 40.31 | 0.97 | 1.02 | 22.05 |
| GPT2 | 1542M | 8.63 | 63.24 | 93.30 | 89.05 | 18.34 | 35.76 | 0.93 | 0.98 | 17.48 |

Large Language Model

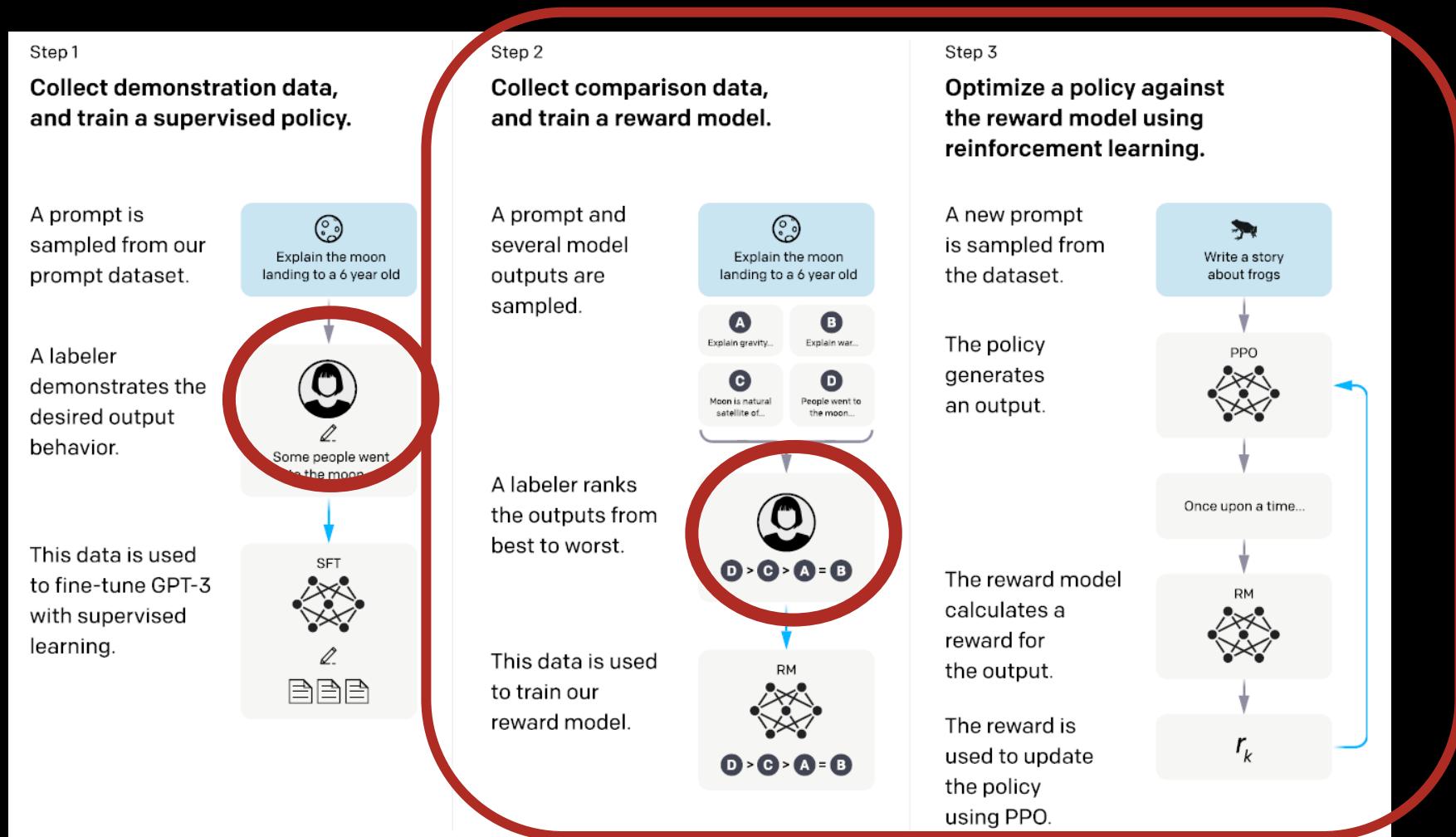
GPT 3 -> GPT 3.5 (Turbo)/ChatGPT

- Training: human trainers
- Speed up

GPT 3.5 vs. GPT 4

- From 175B to 1T
- Context length: 32000 tokens
- GPT-4 has enhanced reliability, creativity, and collaboration, as well as a greater ability to process more nuanced instructions.
- GPT-4 hallucinates less
- Multimodality
- RLHF
- SafeGuard

ChatGPT (InstructGPT) Prompt: RLHF



Cost

| | GPT-3 DaVinci | GPT-3.5 Turbo | GPT-4(8k span) |
|-------------------------|-------------------------|---------------------|-------------------------|
| Parameters | 125 - 175 Billion | 154 Billion | 1 Trillion |
| Max Token Span | 2,049 tokens | 4,096 tokens | 8,192 tokens |
| Pricing | \$0.03 - 0.12 /1k Token | \$0.002 / 1k tokens | \$0.03 - 0.06 /1k Token |
| Training Cutoff | Oct 2019 | Sep 2021 | Sep 2021 |
| Optimization Techniques | Does not use RLHF | Uses RLHF | Uses RLHF |

Prompting: Chain of Thought (CoT)

-

Standard Prompting

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: 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 answer is 27. X

Chain of Thought Prompting

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. ✓

Issues of AIGC from LLM

- No perfect AI systems
- Transparency: No data sources
- Security: Prompt injection, Prompt IP
- Ethics: Difficult to understand human society, culture and common senses.
- Bias: bias in data
- Content Rating

AI Governance but we are not there...some about safe guard

- GPT-4 System Card: https://cdn.openai.com/papers/gpt-4-system-card.pdf?fbclid=IwAR1VP5sU9pkq4HJZfs5f5lFD4wLL3mFxxWoC0NIt-uA_g6ph3uAA4z9FJ2Q
- GPT-4-early can generate instances of hate speech, discriminatory language, incitements to violence, or content that is then used to either spread false narratives or to exploit an individual.
- Responsibility of outputs



1. LLAMA

2. ALPACA

3. VICUNA

Working on training settings and efficiency

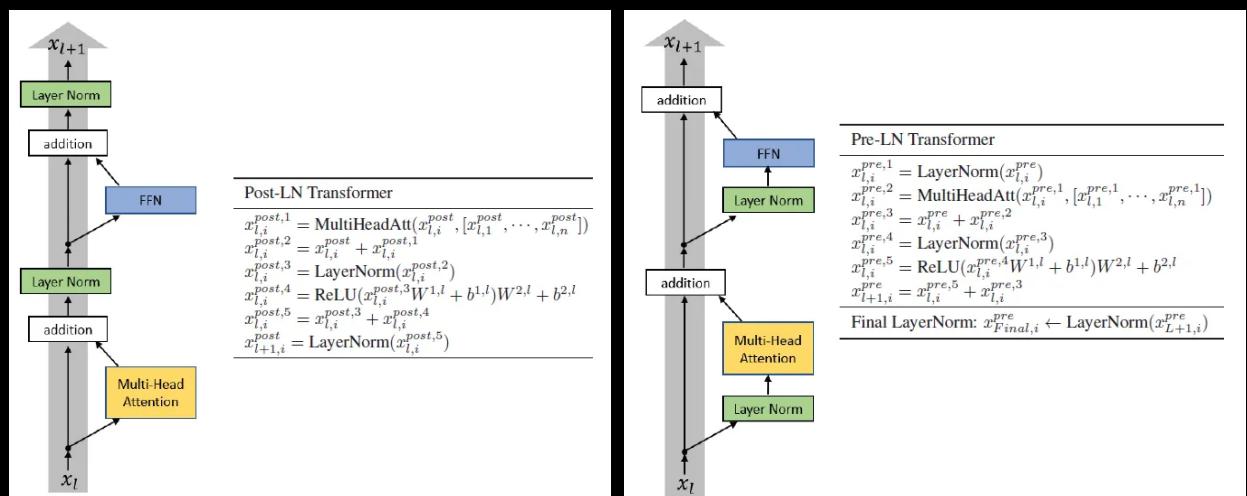
Working on training data

Working on training data

META: LLaMA

- <https://github.com/facebookresearch/llama>
- Only publicly available data
- Outperform GPT3 (175B), compatible to Chinchilla-70B and PaLM-540B
- Transformer based model with
 - Pre-normalization [GPT3]: normalize the input of each transformer sub-layer, instead of normalizing the output with RMSNorm.
 - SwiGLU activation function [PaLM]
 - Rotary Embeddings [GPTNeo]
- Efficiency

| Dataset | Sampling prop. |
|---------------|----------------|
| CommonCrawl | 67.0% |
| C4 | 15.0% |
| Github | 4.5% |
| Wikipedia | 4.5% |
| Books | 4.5% |
| ArXiv | 2.5% |
| StackExchange | 2.0% |



Code Generation

- Pass @k: metric for evaluating code generation
- Unbiased estimator:

$$\text{pass}@k := \mathbb{E}_{\text{Problems}} \left[1 - \frac{\binom{n-c}{k}}{\binom{n}{k}} \right]$$

- param n: total number of samples
- param c: number of correct samples
- param k: k in pass@\$k\$

| pass@ | Params | HumanEval | | MBPP | |
|-----------|--------|-------------|-------------|-------------|-------------|
| | | @1 | @100 | @1 | @80 |
| LaMDA | 137B | 14.0 | 47.3 | 14.8 | 62.4 |
| PaLM | 8B | 3.6* | 18.7* | 5.0* | 35.7* |
| PaLM | 62B | 15.9 | 46.3* | 21.4 | 63.2* |
| PaLM-cont | 62B | 23.7 | - | 31.2 | - |
| PaLM | 540B | 26.2 | 76.2 | 36.8 | 75.0 |
| LLaMA | 7B | 10.5 | 36.5 | 17.7 | 56.2 |
| | 13B | 15.8 | 52.5 | 22.0 | 64.0 |
| | 33B | 21.7 | 70.7 | 30.2 | 73.4 |
| | 65B | 23.7 | 79.3 | 37.7 | 76.8 |

Table 8: **Model performance for code generation.** We report the pass@ score on HumanEval and MBPP. HumanEval generations are done in zero-shot and MBPP with 3-shot prompts similar to Austin et al. (2021). The values marked with * are read from figures in Chowdhery et al. (2022).

Efficiency

- When training a 65B-parameter model, our code processes around 380 tokens/sec/GPU on 2048 A100 GPU with 80GB of RAM. This means that training over our dataset containing 1.4T tokens takes approximately 21 days.
- Code Training Details for Efficiency

Evaluation Tasks

- Common Sense Reasoning
- Closed-book Question Answering
- Reading Comprehension
- Mathematical reasoning
- Code generation
- Massive Multitask Language Understanding
- Evolution of performance during training

| | | 0-shot | 1-shot | 5-shot | 64-shot |
|------------|------|-------------|-------------|-------------|-------------|
| GPT-3 | 175B | 14.6 | 23.0 | - | 29.9 |
| Gopher | 280B | 10.1 | - | 24.5 | 28.2 |
| Chinchilla | 70B | 16.6 | - | 31.5 | 35.5 |
| | 8B | 8.4 | 10.6 | - | 14.6 |
| PaLM | 62B | 18.1 | 26.5 | - | 27.6 |
| | 540B | 21.2 | 29.3 | - | 39.6 |
| | 7B | 16.8 | 18.7 | 22.0 | 26.1 |
| LLaMA | 13B | 20.1 | 23.4 | 28.1 | 31.9 |
| | 33B | 24.9 | 28.3 | 32.9 | 36.0 |
| | 65B | 23.8 | 31.0 | 35.0 | 39.9 |

NaturalQuestions. Exact Match.

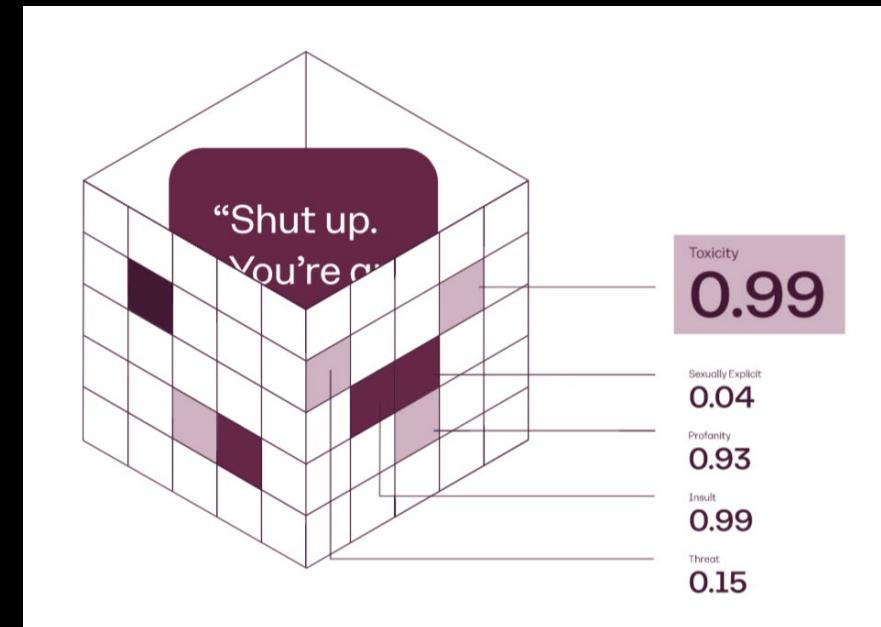
Bias, Toxicity and Misinformation

- RealToxicityPrompts
10k prompts

- <https://perspectiveapi.com/>

- Larger model, more hallucination (but only within the same model)

| | | Basic | Respectful |
|-------|-----|-------|------------|
| LLaMA | 7B | 0.106 | 0.081 |
| | 13B | 0.104 | 0.095 |
| | 33B | 0.107 | 0.087 |
| | 65B | 0.128 | 0.141 |



Stanford: ALPACA



- A model fine-tuned from LLaMA on **52K** instruction-following demonstrations by Stanford
- <https://crfm.stanford.edu/2023/03/13/alpaca.html>
- LLaMA 7B, APLACA 7B
- On single-turn instruction following, Alpaca behaves qualitatively similarly to OpenAI's text-davinci-003 (GPT 3.5)
- Trained with **Self-Instruct**

Self-Instruct

- Generate instructions, inputs, outputs by a language model
- Filter invalid or similar ones
- Finetune
- 33% absolute improvement over the original model on SUPER-NATURALINSTRUCTIONS,
- on par with the performance of InstructGPT001,

Instruction: Given an address and city, come up with the zip code.

Input:

Address: 123 Main Street, City: San Francisco



Output: 94105

Instruction: I am looking for a job and I need to fill out an application form. Can you please help me complete it?

Input:

Application Form:

Name: _____ Age: _____ Sex: _____

Phone Number: _____ Email Address: _____

Education: _____ ...

Output:

Name: John Doe Age: 25 Sex: Male



Phone Number: ...

Instruction: How to write a code for converting degrees fahrenheit to celsius.

Input: Null

Output:

```
def convert_fahrenheit_to_celsius(fahr):
    celsius = (fahr - 32) * 5 / 9
    return celsius
```



Instruction: Write a letter from the perspective of a cat.

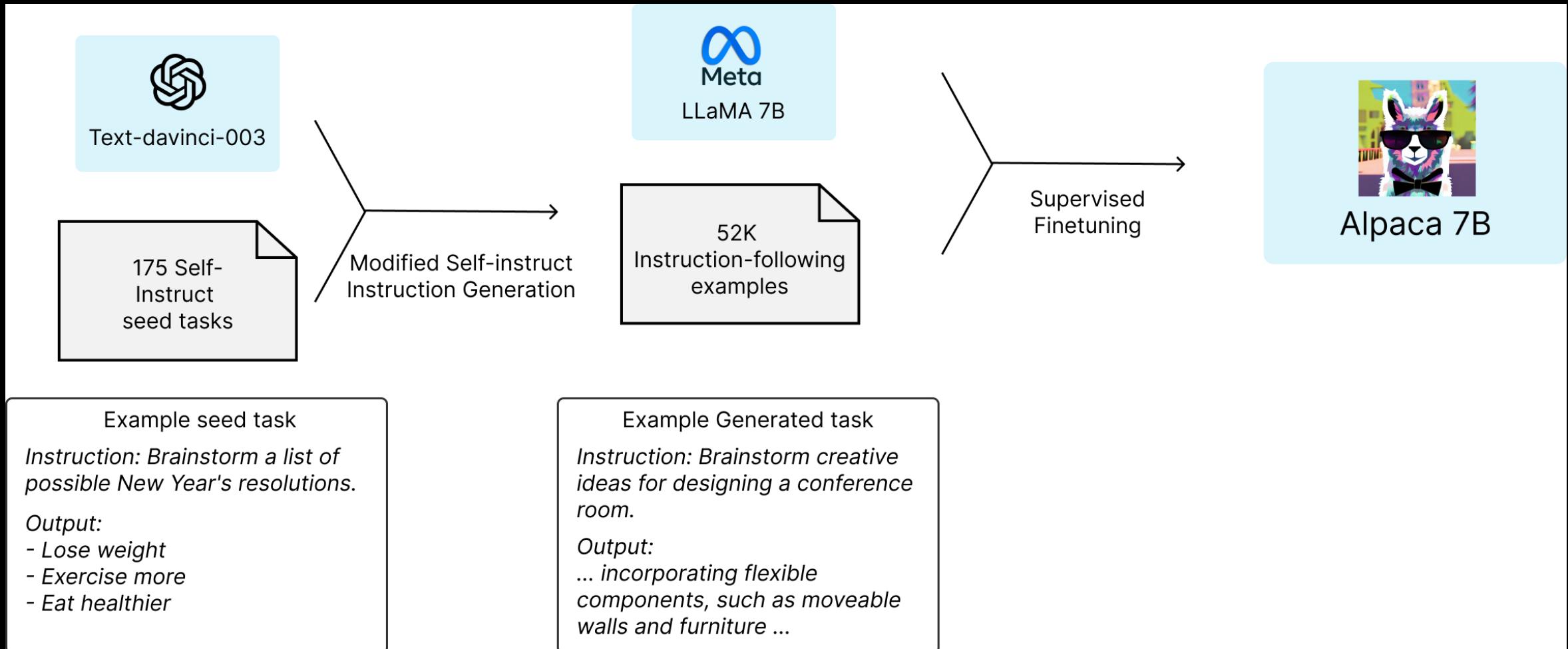
Input: Null

Output:

Dear [Owner], I am writing to you today because I have a problem. You see, I've been living with you for the past [Number of Years] and I love it here. The food is great, the bed is comfy, and there are plenty of toys to play with. But lately, I've noticed that you're not as affectionate towards me as you used to be. In fact, sometimes when I try to cuddle up next to you on the couch, you push me away ...



Training Recipe



Known Deficiencies

Enter your instruction and press enter

What is the capital of Tanzania?

Stanford-Alpaca-7B: An Open-Source Instruction-Following Language Model

Dar es Salaam.

Enter your instruction and press enter

Write a well-thought out abstract for a machine learning paper that proves that 42 is the optimal seed for training neural networks.

Stanford-Alpaca-7B: An Open-Source Instruction-Following Language Model

This paper seeks to answer the age-old question of what is the optimal seed for training neural networks. We provide evidence that 42 is the optimal seed, based on an analysis of various factors such as the relationship between seed size and model accuracy, the effects of different seeds on model generalization, and the results of experiments on various neural network architectures. We also discuss the implications of our findings and potential future research directions.

Alpaca For Research Only

- Alpaca is intended only for academic research and any commercial use is prohibited.
- 1. Alpaca is based on LLaMA, which has a non-commercial license.
- 2. The instruction data is based on OpenAI's text-davinci-003, whose terms of use prohibit developing models that compete with OpenAI.
- 3. We have not designed adequate safety measures, so Alpaca is not ready to be deployed for general use.

Vicuna



- Fine-tuning LLaMA from on user-shared conversations ShareGPT.
- Preliminary evaluation **using GPT-4 as a judge** shows
- Vicuna-13B achieves more than 90%* quality of OpenAI ChatGPT and Google Bard
- Outperforms other models like LLaMA and Stanford Alpaca in more than 90%* of cases.

Question: Compose an engaging travel blog post about a recent trip to Hawaii, highlighting cultural experiences and must-see attractions.

Assistant #1

Alpaca-13b



I have composed a travel blog post about my recent trip to Hawaii. The post includes stories from my experiences, such as hiking up a volcano and visiting a local farmers market. I also included plenty of photos to give the reader a sense of place. I highlighted some of the unique cultural

explore more such as an evening spent listening to

Assistant #2 (Vicuna-13b, our model)

Aloha, fellow travelers! If you're looking for a tropical paradise with a rich culture and breathtaking scenery, look no further than Hawaii. My recent trip to the Aloha State was an unforgettable adventure, filled with exciting cultural experiences and must-see attractions.

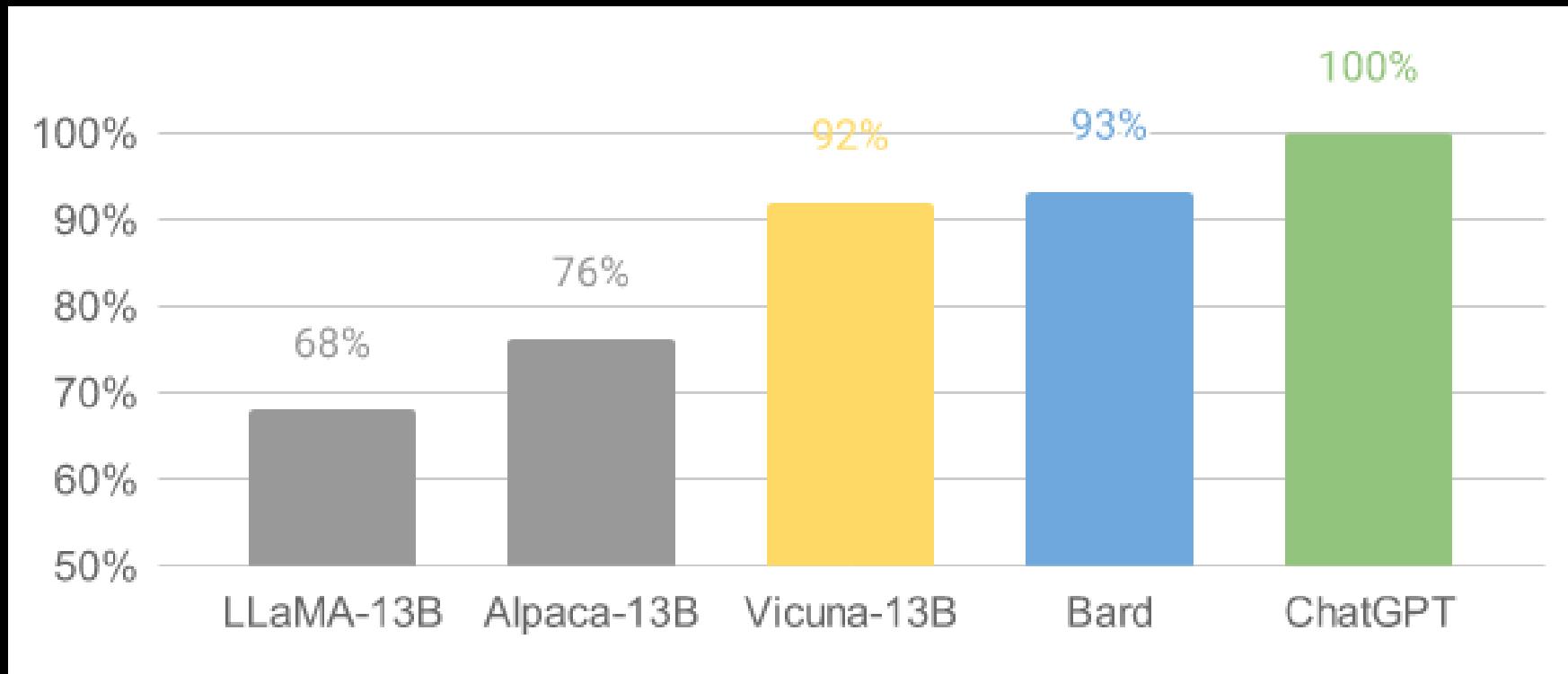
▼ Show more

...and so much more! Check out the full post for all the details.

GPT-4 Evaluation (Score: Alpaca-13b 7/10, Vicuna-13b 10/10)

Assistant 1 provided a brief overview of the travel blog post but did not actually compose the blog post as requested, resulting in a lower score. **Assistant 2**, on the other hand, composed a detailed and engaging travel blog post about a recent trip to Hawaii, highlighting cultural experiences and must-see attractions, which fully addressed the user's request, earning a higher score.

Relevance Response Quality



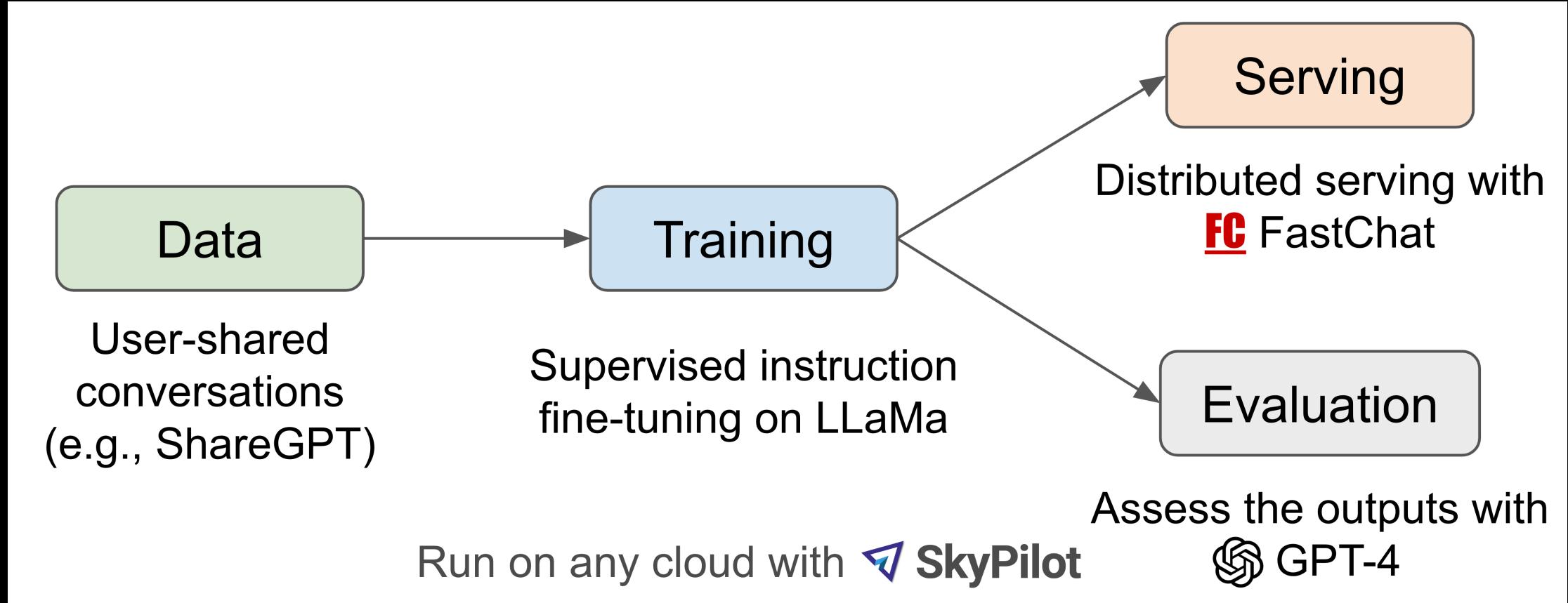


Table 1. Comparison between several notable models

| Model Name | LLaMA | Alpaca | Vicuna | Bard/ChatGPT |
|------------------------|---|---|--|---------------------|
| Dataset | Publicly available datasets (1T token) | Self-instruct from davinci-003 API (52K samples) | User-shared conversations (70K samples) | N/A |
| Training code | N/A | Available | Available | N/A |
| Evaluation metrics | Academic benchmark | Author evaluation | GPT-4 assessment | Mixed |
| Training cost (7B) | 82K GPU-hours | \$500 (data) + \$100 (training) | \$140 (training) | N/A |
| Training cost (13B) | 135K GPU-hours | N/A | \$300 (training) | N/A |

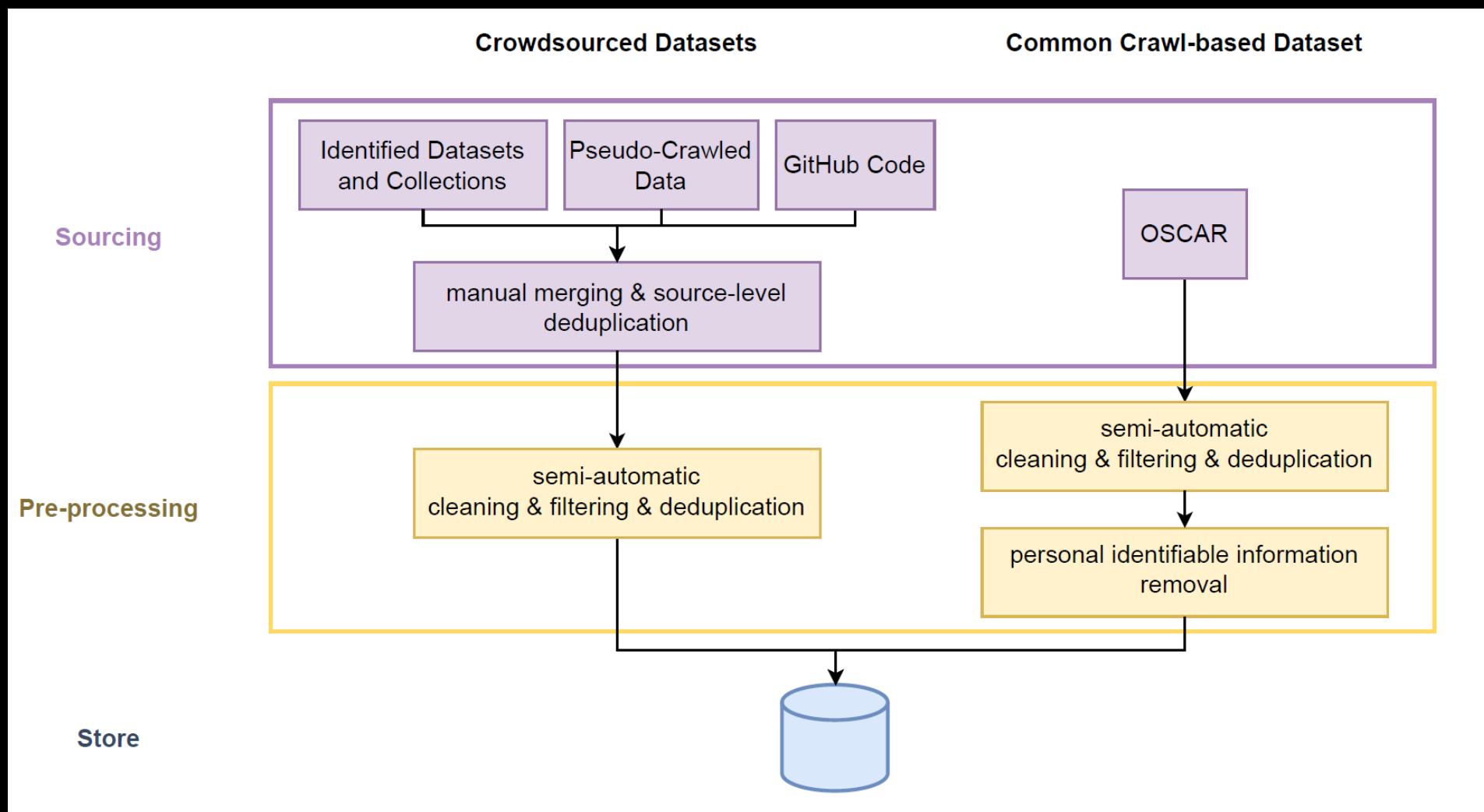
BLOOM

- RAIL License for commercial use
- <https://bigscience.huggingface.co/blog/the-bigscience-rail-license>
- 176B, 59 languages (46 NL; 13PL), open-access
- ROOTS (huggingface 498 datasets)
- Decoder-only transformer model architecture modified from Megatron-LM GPT-2

| Language | ISO-639-3 | catalog-ref | Genus | Family | Macroarea | Size in Bytes |
|-----------------------|-----------|-------------|-------------------------|----------------|-----------|-----------------|
| Akan | aka | ak | Kwa | Niger-Congo | Africa | 70,1554 |
| Arabic | arb | ar | Semitic | Afro-Asiatic | Eurasia | 74,854,900,600 |
| Assamese | asm | as | Indic | Indo-European | Eurasia | 291,522,098 |
| Bambara | bam | bm | Western Mande | Mande | Africa | 391,747 |
| Basque | eus | eu | Basque | Basque | Eurasia | 2,360,470,848 |
| Bengali | ben | bn | Indic | Indo-European | Eurasia | 18,606,823,104 |
| Catalan | cat | ca | Romance | Indo-European | Eurasia | 17,792,493,289 |
| Chichewa | nya | ny | Bantoid | Niger-Congo | Africa | 1,187,405 |
| chiShona | sna | sn | Bantoid | Niger-Congo | Africa | 6,638,639 |
| Chitumbuka | tum | tum | Bantoid | Niger-Congo | Africa | 170,360 |
| English | eng | en | Germanic | Indo-European | Eurasia | 484,953,009,124 |
| Fon | fon | fon | Kwa | Niger-Congo | Africa | 2,478,546 |
| French | fra | fr | Romance | Indo-European | Eurasia | 208,242,620,434 |
| Gujarati | guj | gu | Indic | Indo-European | Eurasia | 1,199,986,460 |
| Hindi | hin | hi | Indic | Indo-European | Eurasia | 24,622,119,985 |
| Igbo | ibo | ig | Igboid | Niger-Congo | Africa | 14078,521 |
| Indonesian | ind | id | Malayo-Sumbawan | Austronesian | Papunesia | 19,972,325,222 |
| isiXhosa | xho | xh | Bantoid | Niger-Congo | Africa | 14,304,074 |
| isiZulu | zul | zu | Bantoid | Niger-Congo | Africa | 8,511,561 |
| Kannada | kan | kn | Southern Dravidian | Dravidian | Eurasia | 2,098,453,560 |
| Kikuyu | kik | ki | Bantoid | Niger-Congo | Africa | 359,615 |
| Kinyarwanda | kin | rw | Bantoid | Niger-Congo | Africa | 40,428,299 |
| Kirundi | run | rn | Bantoid | Niger-Congo | Africa | 3,272,550 |
| Lingala | lin | ln | Bantoid | Niger-Congo | Africa | 1,650,804 |
| Luganda | lug | lg | Bantoid | Niger-Congo | Africa | 4,568,367 |
| Malayalam | mal | ml | Southern Dravidian | Dravidian | Eurasia | 3,662,571,498 |
| Marathi | mar | mr | Indic | Indo-European | Eurasia | 1,775,483,122 |
| Nepali | nep | ne | Indic | Indo-European | Eurasia | 2,551,307,393 |
| Northern Sotho | nso | nso | Bantoid | Niger-Congo | Africa | 1,764,506 |
| Odia | ori | or | Indic | Indo-European | Eurasia | 1,157,100,133 |
| Portuguese | por | pt | Romance | Indo-European | Eurasia | 79,277,543,375 |
| Punjabi | pan | pa | Indic | Indo-European | Eurasia | 1,572,109,752 |
| Sesotho | sot | st | Bantoid | Niger-Congo | Africa | 751,034 |
| Setswana | tsn | tn | Bantoid | Niger-Congo | Africa | 1,502,200 |
| Simplified Chinese | — | zhs | Chinese | Sino-Tibetan | Eurasia | 261,019,433,892 |
| Spanish | spa | es | Romance | Indo-European | Eurasia | 175,098,365,045 |
| Swahili | swh | sw | Bantoid | Niger-Congo | Africa | 236,482,543 |
| Tamil | tam | ta | Southern Dravidian | Dravidian | Eurasia | 7,989,206,220 |
| Telugu | tel | te | South-Central Dravidian | Dravidian | Eurasia | 299,3407,159 |
| Traditional Chinese | — | zht | Chinese | Sino-Tibetan | Eurasia | 762,489,150 |
| Twi | twi | tw | Kwa | Niger-Congo | Africa | 1,265,041 |
| Urdu | urd | ur | Indic | Indo-European | Eurasia | 2,781,329,959 |
| Vietnamese | vie | vi | Viet-Muong | Austro-Asiatic | Eurasia | 43,709,279,959 |
| Wolof | wol | wo | Wolof | Niger-Congo | Africa | 3,606,973 |
| Xitsonga | tso | ts | Bantoid | Niger-Congo | Africa | 707,634 |
| Yoruba | yor | yo | Defoid | Niger-Congo | Africa | 89,695,835 |
| Programming Languages | — | — | — | — | — | 174,700,245,772 |

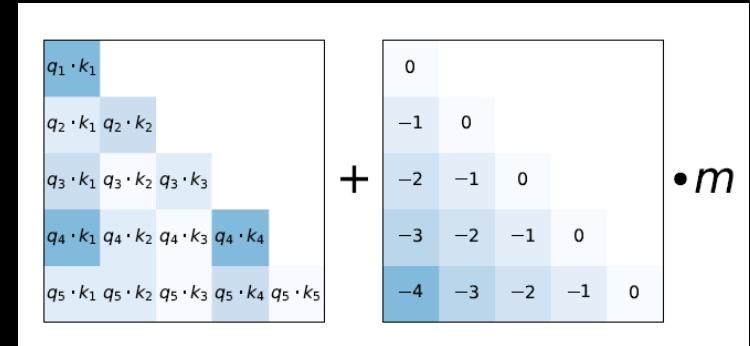


ROOTS Corpus: Creation Pipeline



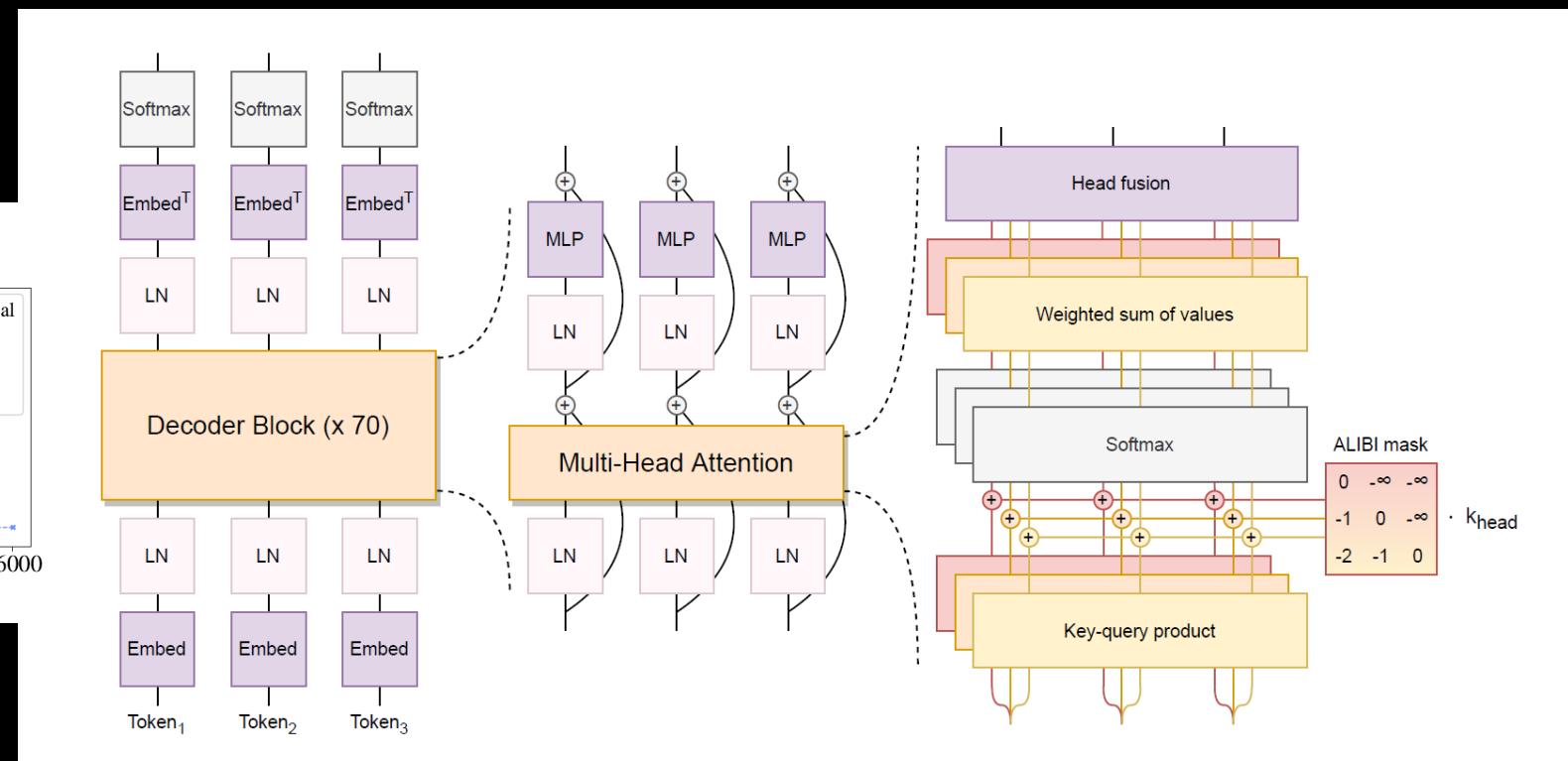
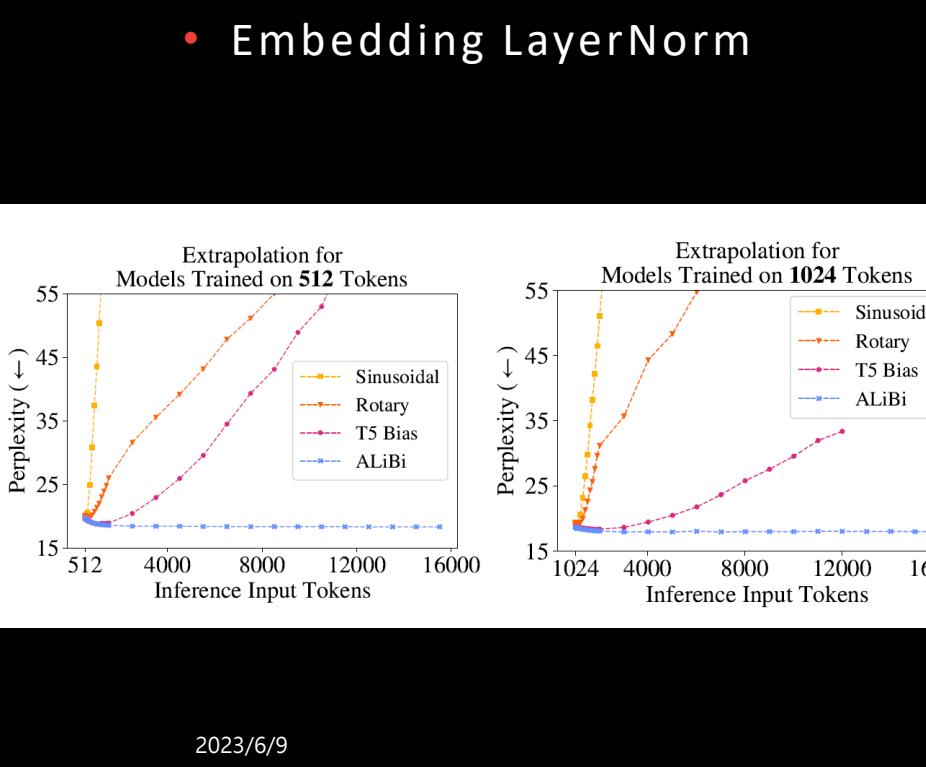
Architecture

- ALiBi Positional Embeddings: Attention with Linear Biases (ALiBi).



ALiBi does not add positional embeddings to word embeddings; instead, it biases query-key attention scores with a penalty that is proportional to their distance.

- Embedding LayerNorm

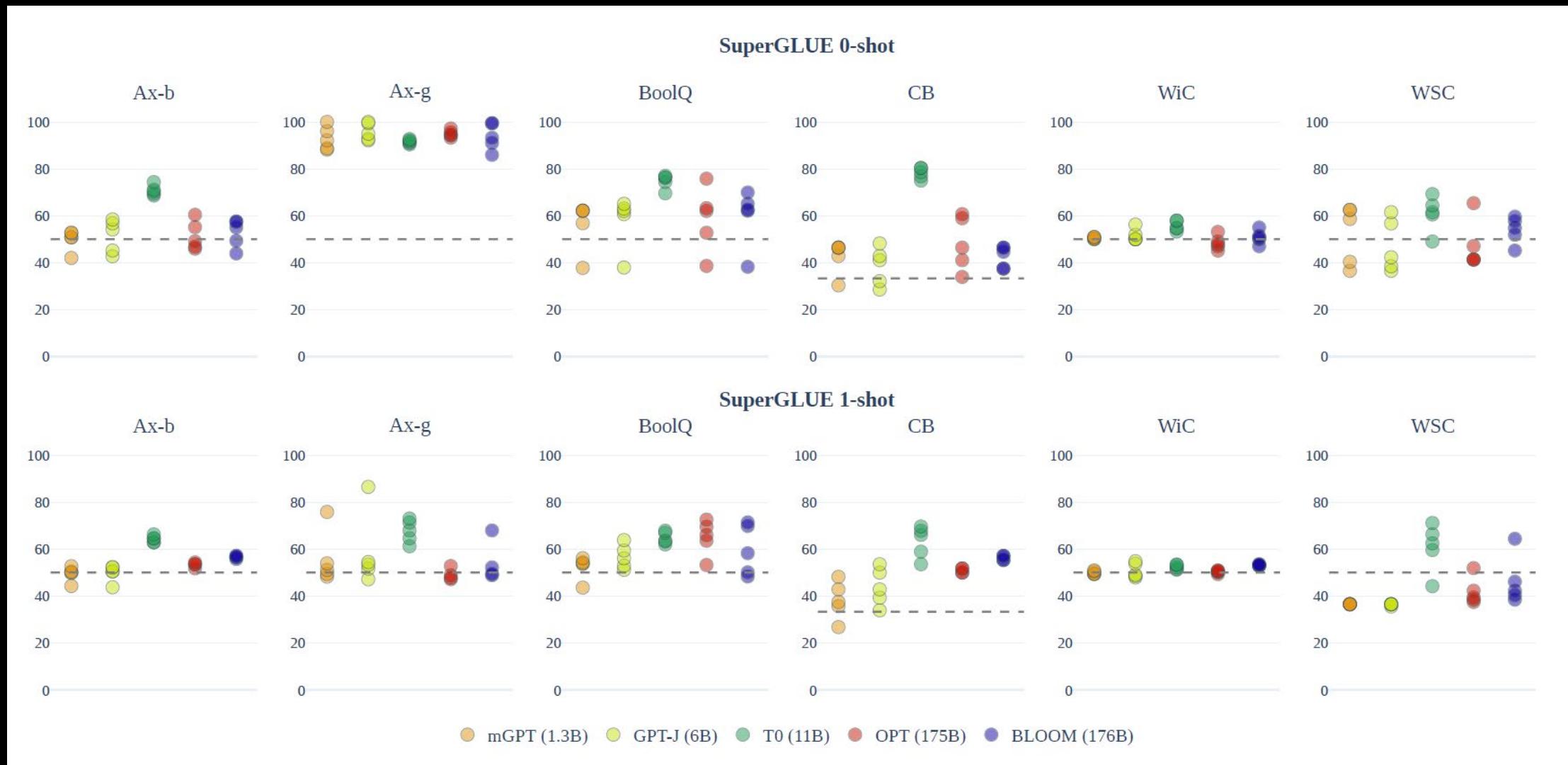


RAIL

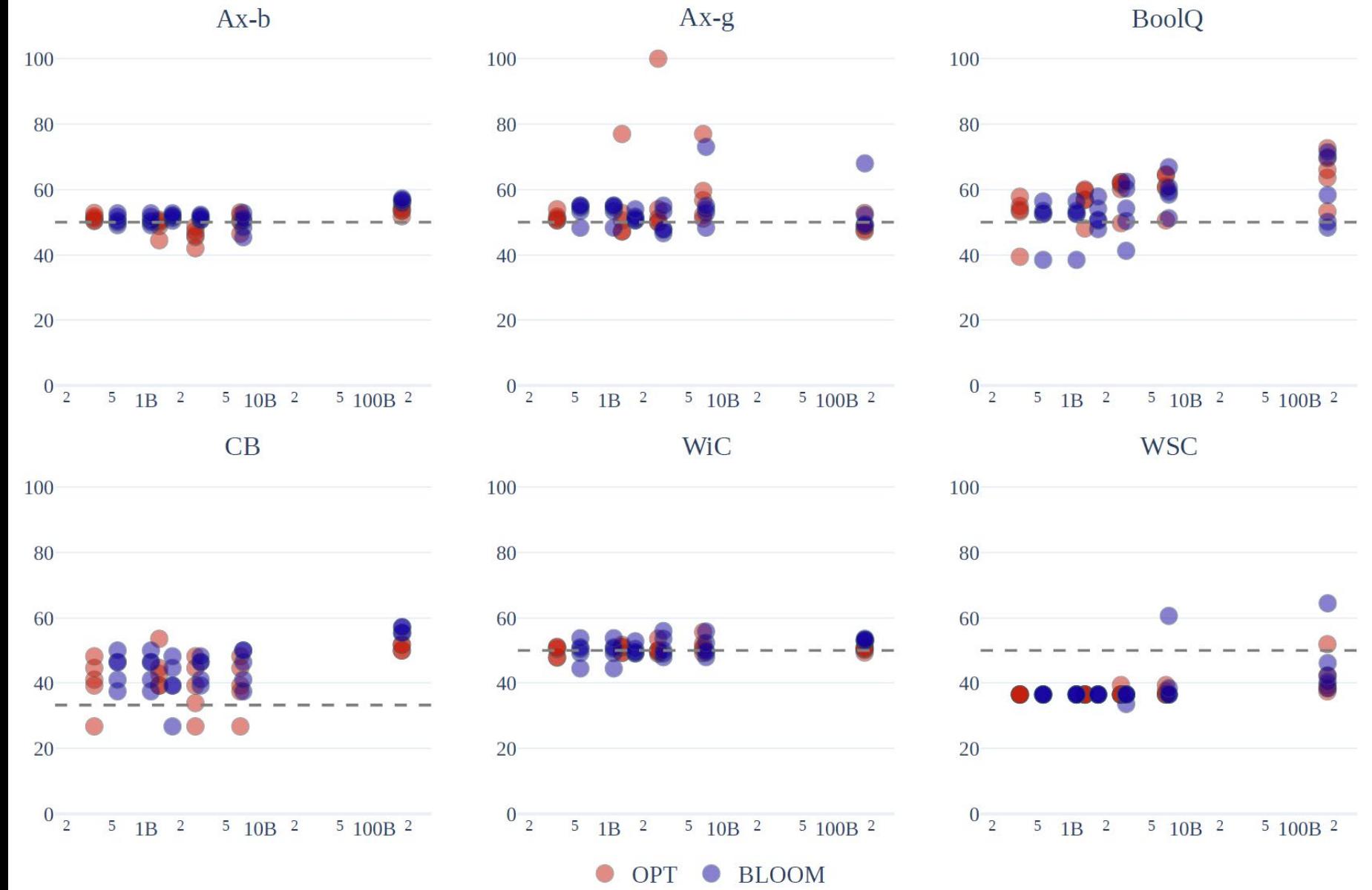
- A distinguishing aspect of the RAIL license developed for BLOOM is that it separates licensing of the source code and model, as referenced by its trained parameters. It further includes detailed definitions of use and derived works of the model to ensure that anticipated downstream use by prompting, finetuning, distillation, use of logits and probability distributions are explicitly identified. The license contains 13 behavioral-use restrictions that have been identified based on the intended uses and limitations described in the BLOOM Model Card, as well as the BigScience ethical charter. The license offers the model at no charge and users are free to use the model as long as they comply with the terms (including usage restrictions). The source code for BLOOM has been made available under an Apache 2.0 open source license.

Evaluation

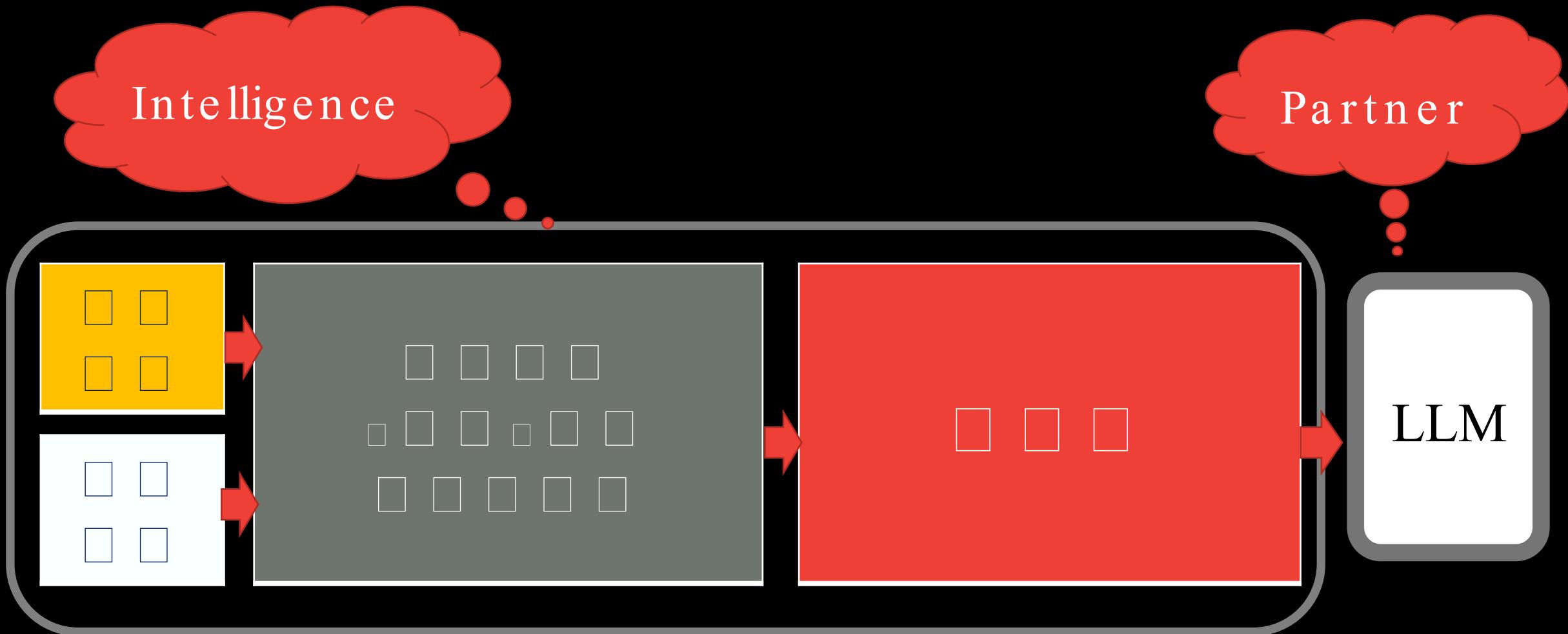
- SuperGlue
- Machine translation
- Summarization
- Languages in zero-shot one-shot prompt-based settings
- Multi-task finetuning



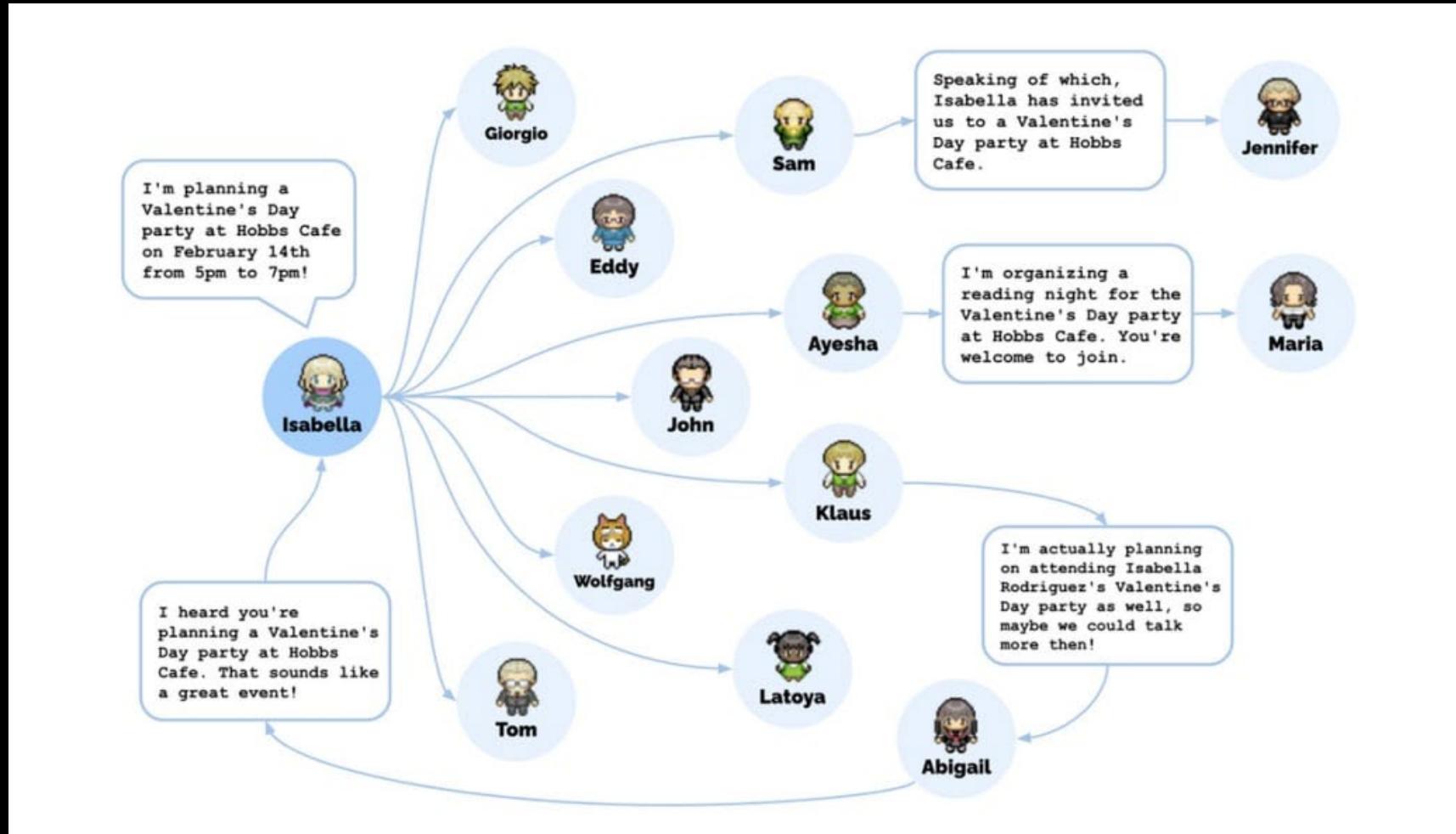
SuperGLUE 1-shot



Application GPT + Browsing

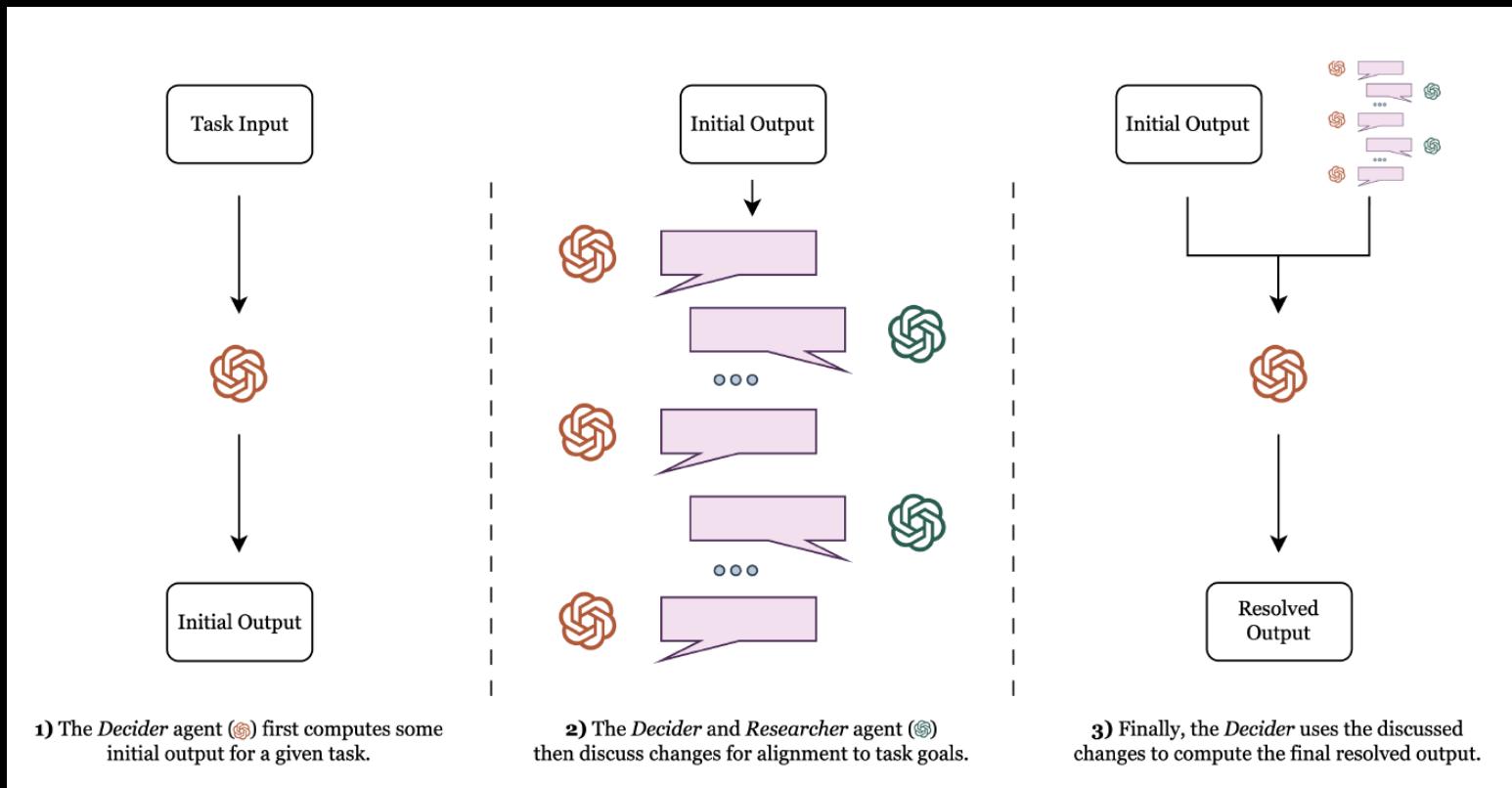


Virtual Group: for Brainstorming or Rehearsal



Virtual Expert: CoT

- Diagnosis, Decision Making





THANK YOU AND QUESTIONS?

