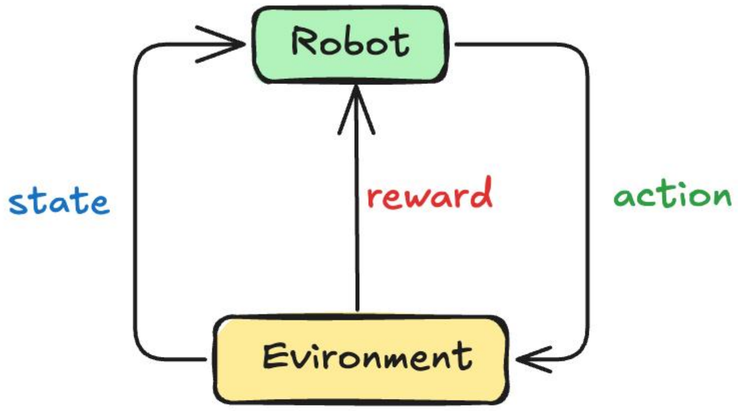
DeepSeekR1论文精读 （补充材料）

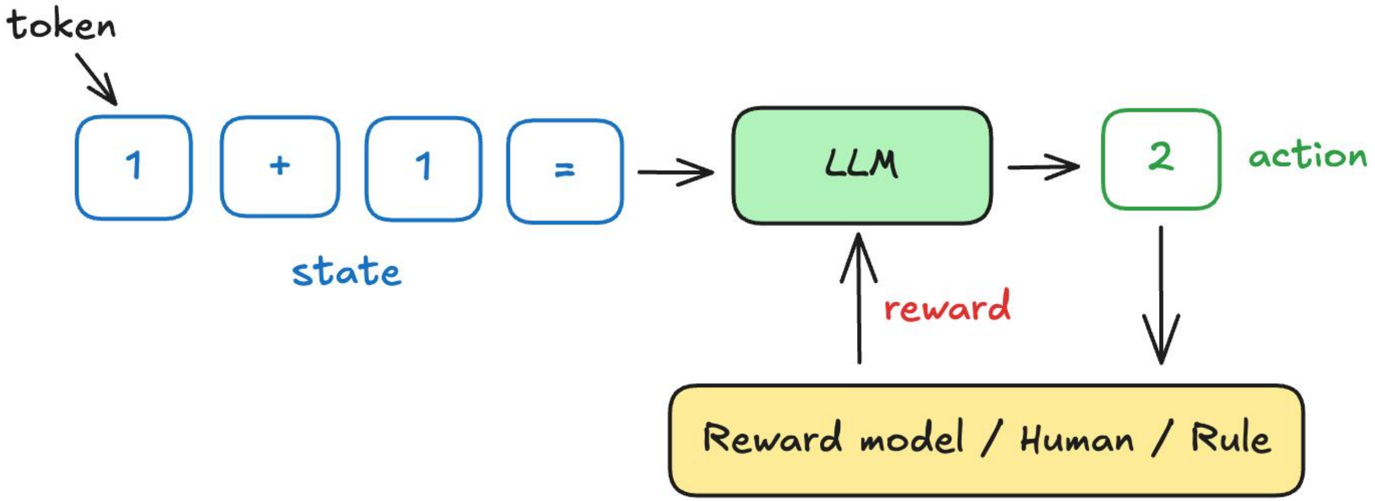
B站-李小羊学AI

PPO原理讲解

Reinforcement Learning

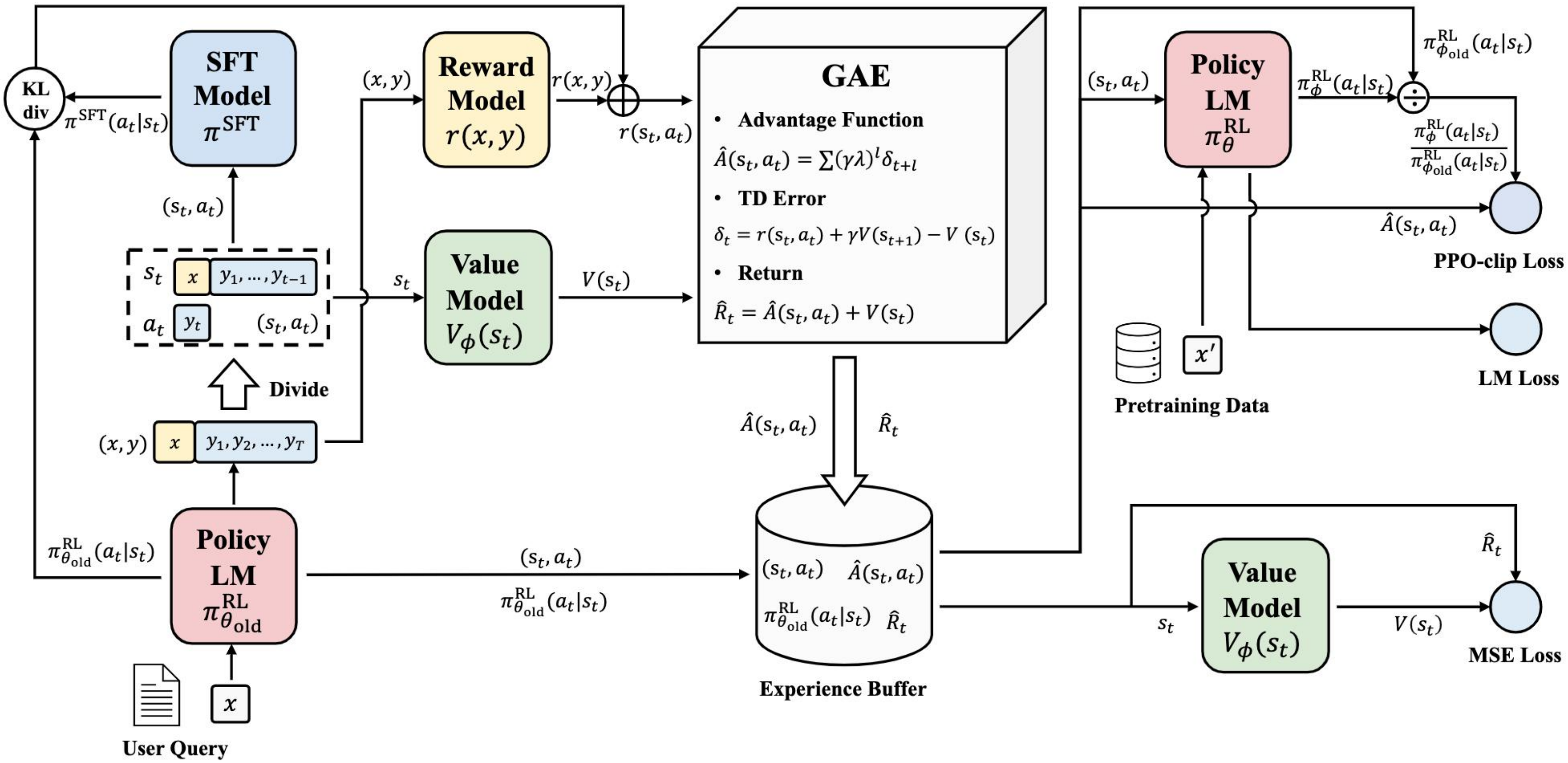


Reinforcement Learning



Reinforcement Learning in LLM Training

Proximal Policy Optimization (PPO)



KL div
$\pi^{\mathrm{SFT}}(a_t|s_t)$
SFT Model π^{SFT}
(x, y)
Reward Model $r(x, y)$
$r(x, y)$
$r(s_t, a_t)$
GAE
• Advantage Function
$\hat{A}(s_t, a_t) = \sum(\gamma\lambda)^l \delta_{t+l}$
• TD Error
$\delta_t = r(s_t, a_t) + \gamma V(s_{t+1}) - V(s_t)$
• Return
$\hat{R}_t = \hat{A}(s_t, a_t) + V(s_t)$
(s_t, a_t)
s_t x $y_1, \dots, y_{t-1}$
a_t y_t (s_t, a_t)
s_t
Value Model $V_\phi(s_t)$
$V(s_t)$
Divide
(x, y) x $y_1, y_2, \dots, y_T$
$\pi^{\mathrm{RL}}_{\theta_{\mathrm{old}}}(a_t|s_t)$
Policy LM $\pi^{\mathrm{RL}}_{\theta_{\mathrm{old}}}$
(s_t, a_t)
$\pi^{\mathrm{RL}}_{\theta_{\mathrm{old}}}(a_t|s_t)$
User Query
x
$\hat{A}(s_t, a_t)$
$\hat{R}_t$
(s_t, a_t) $\hat{A}(s_t, a_t)$
$\pi^{\mathrm{RL}}_{\theta_{\mathrm{old}}}(a_t|s_t)$ $\hat{R}_t$
Experience Buffer
(s_t, a_t)
Policy LM $\pi^{\mathrm{RL}}_{\theta}$
$\pi^{\mathrm{RL}}_{\phi}(a_t|s_t)$
$\pi^{\mathrm{RL}}_{\phi_{\mathrm{old}}}(a_t|s_t)$
$\frac{\pi^{\mathrm{RL}}_{\phi}(a_t|s_t)}{\pi^{\mathrm{RL}}_{\phi_{\mathrm{old}}}(a_t|s_t)}$
$\hat{A}(s_t, a_t)$
PPO-clip Loss
x'
Pretraining Data
LM Loss
$\hat{R}_t$
s_t
Value Model $V_\phi(s_t)$
$V(s_t)$
MSE Loss

Figure 1: PPO workflow, depicting the sequential steps in the algorithm's execution. The process begins with sampling from the environment, followed by the application of GAE for improved advantage approximation. The diagram then illustrates the computation of various loss functions employed in PPO, signifying the iterative nature of the learning process and the policy updates derived from these losses.

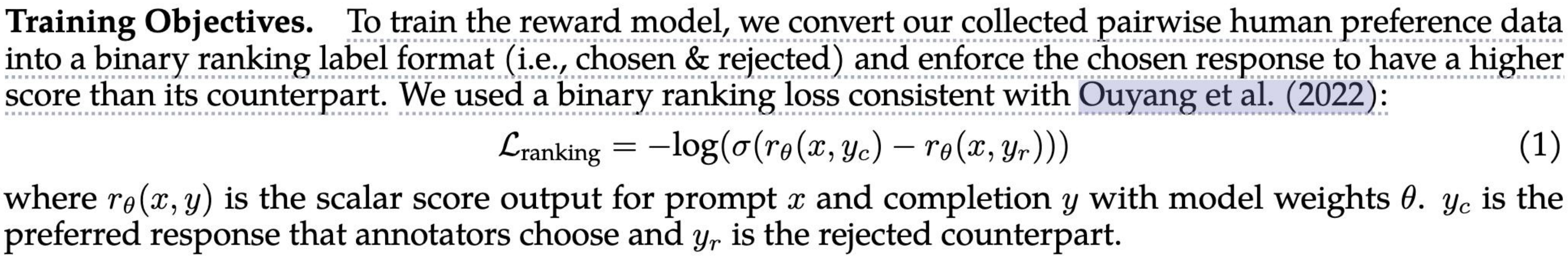
reference：Secrets of RLHF in Large Language Models Part I: PPO

Reward Model

如何训练一个Reward Model来给LLM的回复评分呢?

user: What is 1 + 1 = ?
robot1: 3
robot2: 2
our target: reward1 < reward2

reward model loss:



Training Objectives. To train the reward model, we convert our collected pairwise human preference data into a binary ranking label format (i.e., chosen & rejected) and enforce the chosen response to have a higher score than its counterpart. We used a binary ranking loss consistent with Ouyang et al. (2022):

$$\mathcal{L}_{\text{ranking}} = -\log(\sigma(r_\theta(x, y_c) - r_\theta(x, y_r))) \tag{1}$$

where $r_\theta(x, y)$ is the scalar score output for prompt x and completion y with model weights θ. y_c is the preferred response that annotators choose and y_r is the rejected counterpart.

reference: Llama 2: Open Foundation and Fine-Tuned Chat Models

Value model

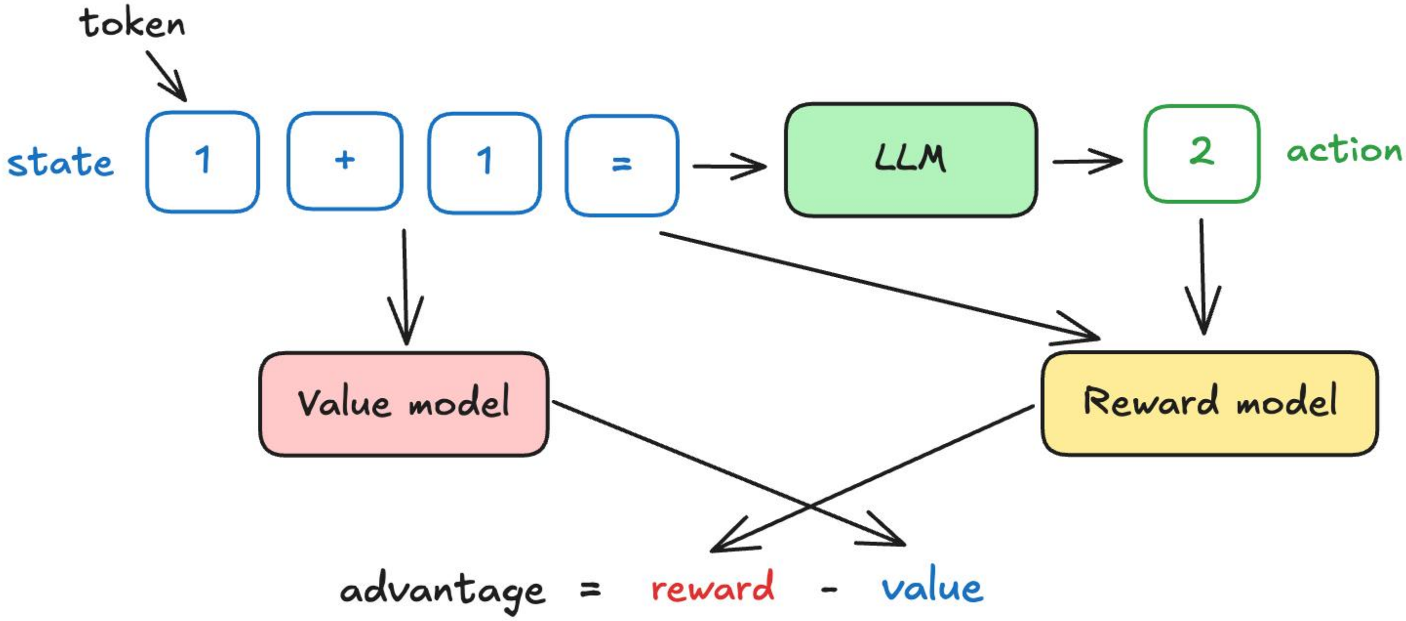
思考一下，如果reward model输出的奖励全为负值或者全为正值怎么办?

利用Value Model / Critic Mode 用于估计reward的基准值

Value = V(s_t)

Reward = R(s_t, a_t)

Advantage（优势估计）= Reward - Value



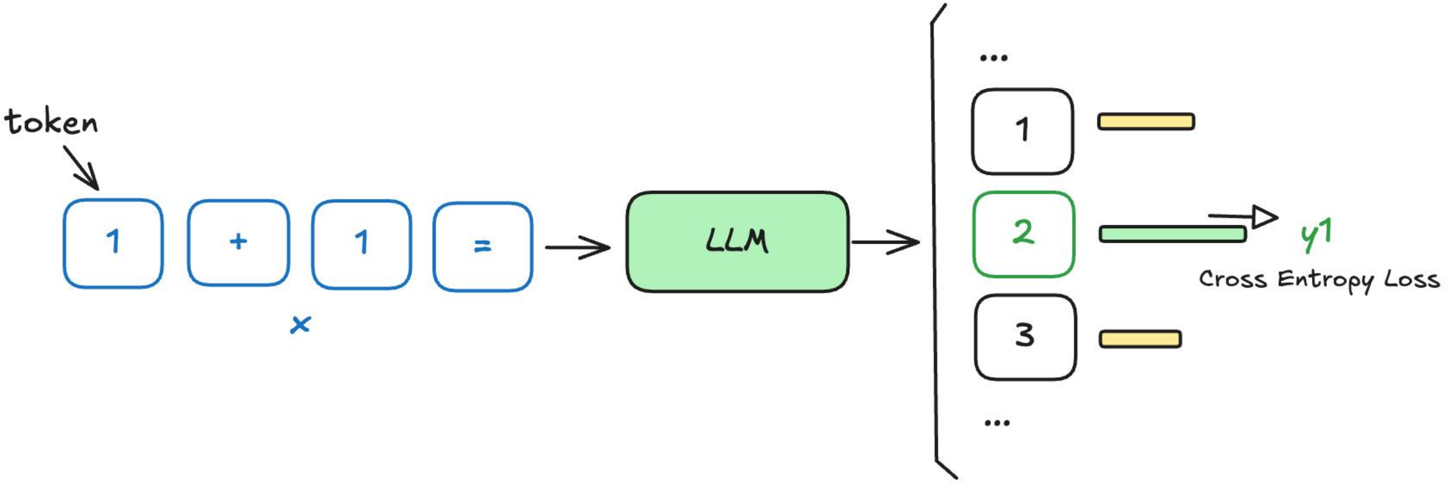
Generalized Advantage Estimated（GAE）是一种最常用的优势估计计算方法

GAE详情参考：Secrets of RLHF in Large Language Models Part I: PPO

Policy Gradient Methods 策略梯度

如何根据advantage优化参数呢？这时就需要策略梯度了。

让我们先来看一下SFT（supervised fine tuning）的原理



Supervised Fine-tuning

其梯度表示为：

$$\nabla_\theta J(\theta) = \mathbb{E}_{\tau \sim \pi_\theta}\left[\sum_{t=0}^{T} \nabla_\theta \log \pi_\theta(a_t|s_t) \quad \right],$$

SFT是reward/advantage为1的策略梯度

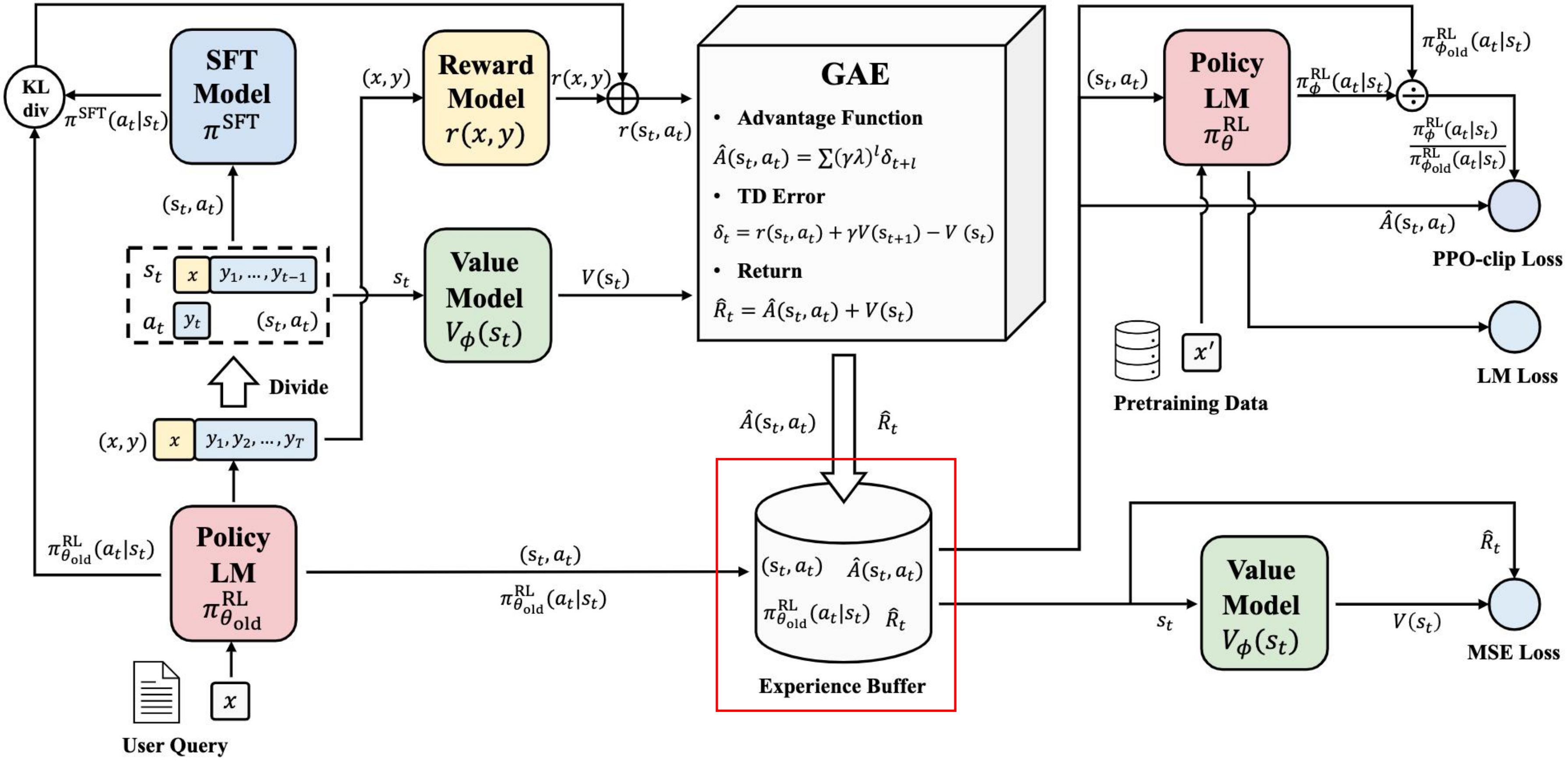
而强化学习中的策略梯度表示为

$$\nabla_\theta J(\theta) = \mathbb{E}_{\tau \sim \pi_\theta}\left[\sum_{t=0}^{T} \nabla_\theta \log \pi_\theta(a_t|s_t) \boxed{\Phi_t} \right],$$

reward/advantage，可正可负

importance sampling 重要性采样

在训练过程中，为了提高训练效率和数据利用效率，通常会基于old policy进行大批量数据采样，然后在进行后续训练。



KL div
$\pi^{\mathrm{SFT}}(a_t|s_t)$
SFT Model π^{SFT}
(x,y)
Reward Model $r(x,y)$
$r(x,y)$
$r(s_t,a_t)$
GAE
• Advantage Function
$\hat{A}(s_t,a_t)=\sum(\gamma\lambda)^l\delta_{t+l}$
• TD Error
$\delta_t=r(s_t,a_t)+\gamma V(s_{t+1})-V(s_t)$
• Return
$\hat{R}_t=\hat{A}(s_t,a_t)+V(s_t)$
(s_t,a_t)
Policy LM π_θ^{RL}
$\pi_{\phi_{\mathrm{old}}}^{\mathrm{RL}}(a_t|s_t)$
$\pi_\phi^{\mathrm{RL}}(a_t|s_t)$
$\frac{\pi_\phi^{\mathrm{RL}}(a_t|s_t)}{\pi_{\phi_{\mathrm{old}}}^{\mathrm{RL}}(a_t|s_t)}$
(s_t,a_t)
s_t x $y_1,\dots,y_{t-1}$
a_t y_t (s_t,a_t)
s_t
Value Model $V_\phi(s_t)$
$V(s_t)$
$\hat{A}(s_t,a_t)$
PPO-clip Loss
x'
Pretraining Data
LM Loss
Divide
(x,y) x $y_1,y_2,\dots,y_T$
$\hat{A}(s_t,a_t)$
$\hat{R}_t$
$\pi_{\theta_{\mathrm{old}}}^{\mathrm{RL}}(a_t|s_t)$
Policy LM $\pi_{\theta_{\mathrm{old}}}^{\mathrm{RL}}$
(s_t,a_t)
$\pi_{\theta_{\mathrm{old}}}^{\mathrm{RL}}(a_t|s_t)$
(s_t,a_t) $\hat{A}(s_t,a_t)$
$\pi_{\theta_{\mathrm{old}}}^{\mathrm{RL}}(a_t|s_t)$ $\hat{R}_t$
Experience Buffer
$\hat{R}_t$
s_t
Value Model $V_\phi(s_t)$
$V(s_t)$
MSE Loss
x
User Query

参考：https://www.bilibili.com/video/BV1XP4y1d7Bk

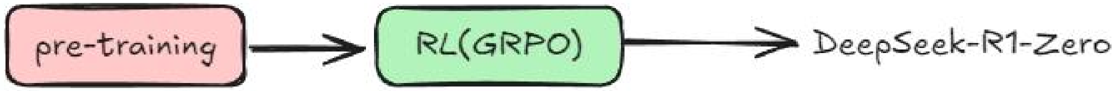
PPO优缺点分析

PPO的优点：

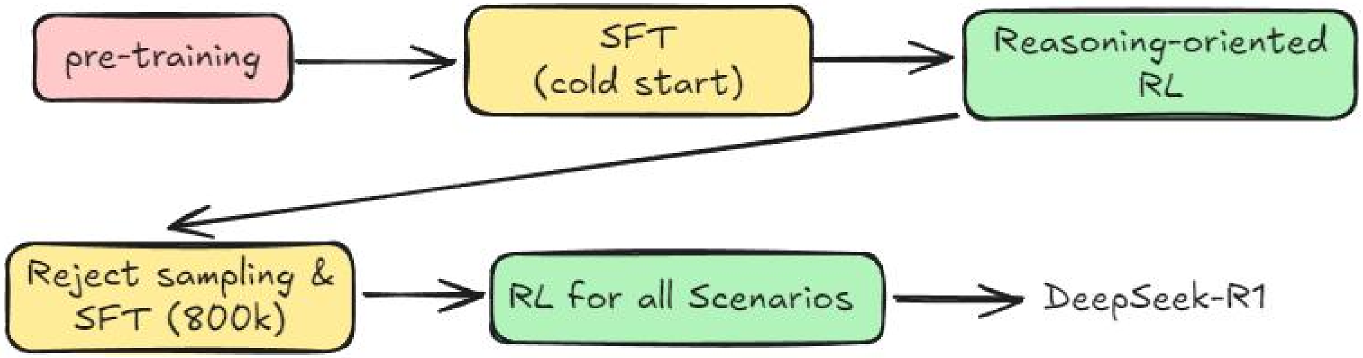
- 训练稳定，效果经过大批量实验验证。

PPO的问题：

- 过程复杂，需要同时加载4个模型（reference model、reward model、critic model、policy model）
- 训练负载高，其中policy model、value/critic model都需要参数更新



DeepSeek-R1-Zero



DeepSeek-R1

测试集介绍

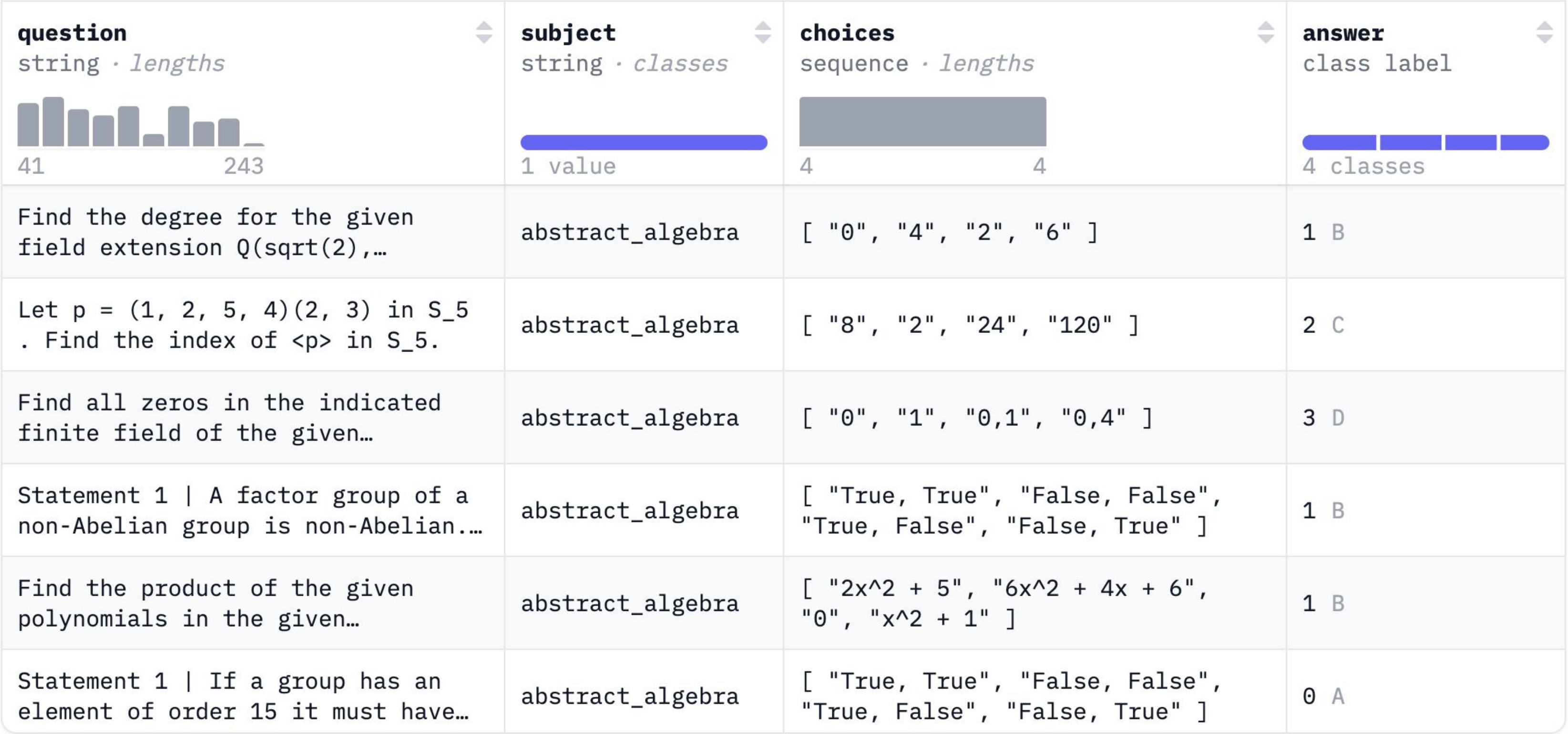
English

MMLU（多学科单项选择题）

https://paperswithcode.com/dataset/mmlu

MMLU (Massive Multitask Language Understanding) is a new benchmark designed to measure knowledge acquired during pretraining by evaluating models exclusively in **zero-shot and few-shot** settings. This makes the benchmark more challenging and more similar to how we evaluate humans. The benchmark covers 57 subjects across STEM, the humanities, the social sciences, and more. It ranges in difficulty from an elementary level to an advanced professional level, and it tests both world knowledge and problem solving ability. Subjects range from traditional areas, such as mathematics and history, to more specialized areas like law and ethics. The granularity and breadth of the subjects makes the benchmark ideal for identifying a model’s blind spots.

注：STEM: Science, Technology, Engineering, and Mathematics



question string · *lengths* 41 243	subject string · *classes* 1 value	choices sequence · *lengths* 4 4	answer class label 4 classes
Find the degree for the given field extension Q(sqrt(2),…	abstract_algebra	["0", "4", "2", "6"]	1 B
Let p = (1, 2, 5, 4)(2, 3) in S_5 . Find the index of <p> in S_5.	abstract_algebra	["8", "2", "24", "120"]	2 C
Find all zeros in the indicated finite field of the given…	abstract_algebra	["0", "1", "0,1", "0,4"]	3 D
Statement 1 | A factor group of a non-Abelian group is non-Abelian.…	abstract_algebra	["True, True", "False, False", "True, False", "False, True"]	1 B
Find the product of the given polynomials in the given…	abstract_algebra	["2x^2 + 5", "6x^2 + 4x + 6", "0", "x^2 + 1"]	1 B
Statement 1 | If a group has an element of order 15 it must have…	abstract_algebra	["True, True", "False, False", "True, False", "False, True"]	0 A

MMLU-Redux（MMLU精简版）

https://github.com/aryopg/mmlu-redux

MMLU-Redux is a carefully annotated version of the MMLU (Massive Multitask Language Understanding) dataset to provide a more accurate and reliable benchmark for evaluating the performance of language models.

MMLU-Redux consists of 30 MMLU subjects, each containing 100 randomly sampled questions. Please refer to 🤗MMLU-Redux Dataset for more details.

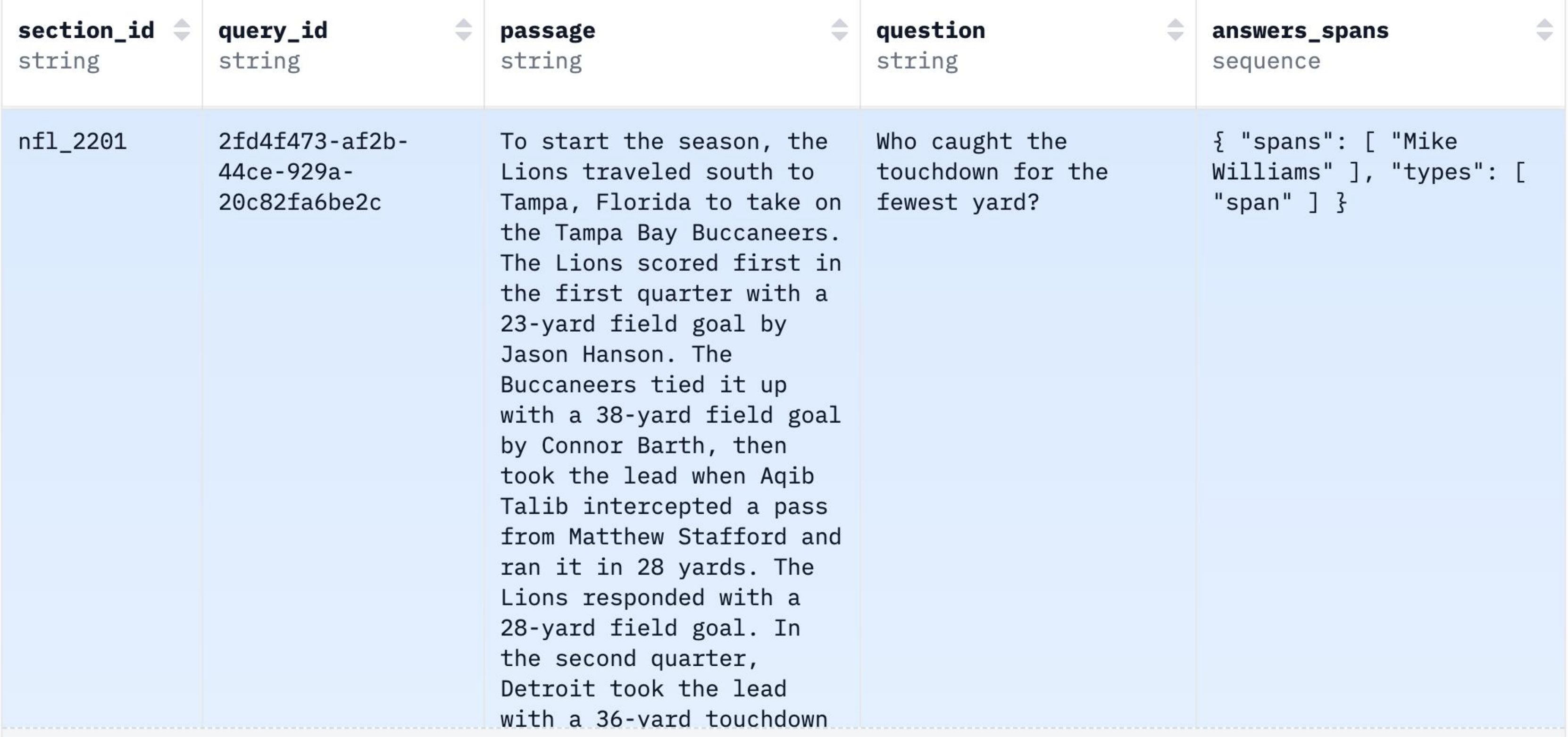
MMLU-Pro（MMLU困难版）

https://huggingface.co/spaces/TIGER-Lab/MMLU-Pro

We introduce MMLU-Pro, an enhanced benchmark designed to evaluate language understanding models across broader and more challenging tasks. Building on the Massive Multitask Language Understanding (MMLU) dataset, MMLU-Pro integrates more challenging, reasoning-focused questions and increases the answer choices per question from four to ten, significantly raising the difficulty and reducing the chance of success through random guessing. MMLU-Pro comprises over 12,000 rigorously curated questions from academic exams and textbooks, spanning 14 diverse domains including Biology, Business, Chemistry, Computer Science, Economics, Engineering, Health, History, Law, Math, Philosophy, Physics, Psychology, and Others.

DROP（阅读理解） https://paperswithcode.com/dataset/drop

Discrete Reasoning Over Paragraphs DROP is a crowdsourced, adversarially-created, 96k-question benchmark, in which a system must resolve references in a question, perhaps to multiple input positions, and perform discrete operations over them (such as addition, counting, or sorting). These operations require a much more comprehensive understanding of the content of paragraphs than what was necessary for prior datasets. The questions consist of passages extracted from Wikipedia articles. The dataset is split into a training set of about 77,000 questions, a development set of around 9,500 questions and a hidden test set similar in size to the development set.



section_id string	query_id string	passage string	question string	answers_spans sequence
nfl_2201	2fd4f473-af2b-44ce-929a-20c82fa6be2c	To start the season, the Lions traveled south to Tampa, Florida to take on the Tampa Bay Buccaneers. The Lions scored first in the first quarter with a 23-yard field goal by Jason Hanson. The Buccaneers tied it up with a 38-yard field goal by Connor Barth, then took the lead when Aqib Talib intercepted a pass from Matthew Stafford and ran it in 28 yards. The Lions responded with a 28-yard field goal. In the second quarter, Detroit took the lead with a 36-yard touchdown	Who caught the touchdown for the fewest yard?	{ "spans": ["Mike Williams"], "types": ["span"] }

https://paperswithcode.com/dataset/ifeval

IFEval (Instruction Following Evaluation Datset，指令跟随测试集)

This dataset evaluates **instruction following** ability of large language models. There are 500+ prompts with instructions such as

- "write an article with more than 800 words"
- "wrap your response with double quotation marks", etc.

https://huggingface.co/datasets/Idavidrein/gpqa

GPQA stands for Graduate-Level Google-Proof Q&A Benchmark.（大学水平的多项选择题）

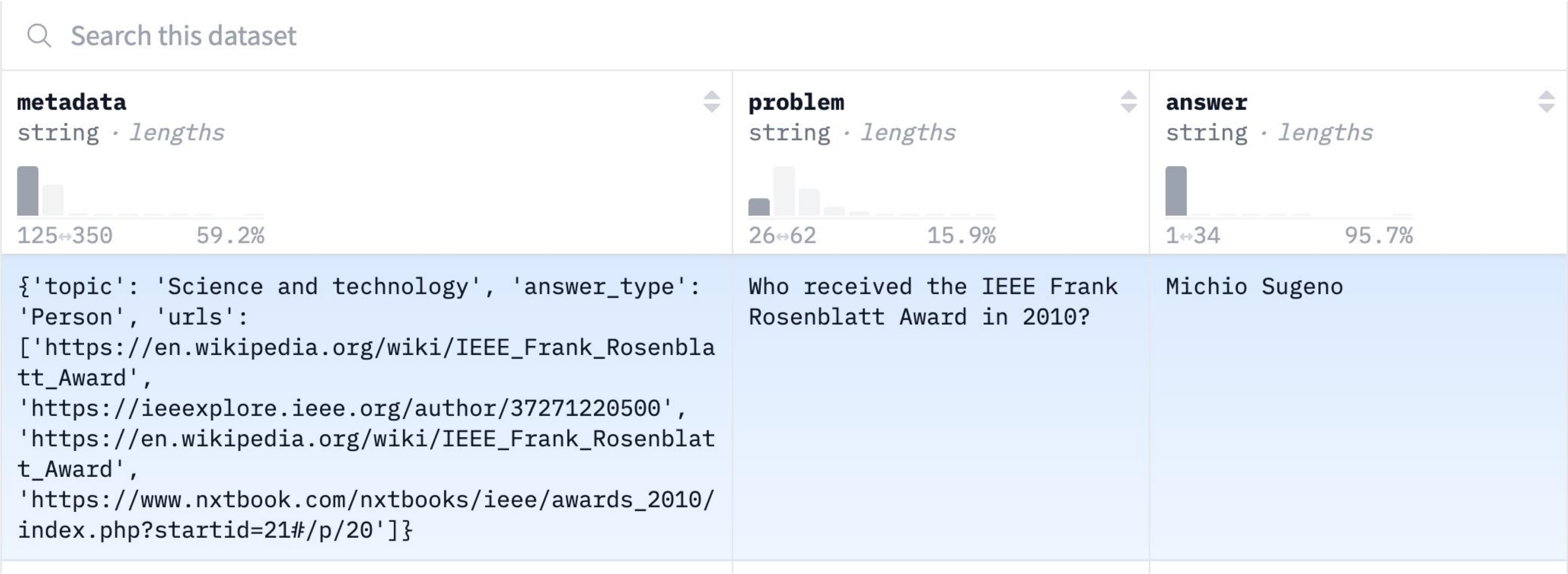
It's a challenging dataset designed to evaluate the capabilities of Large Language Models (LLMs) and scalable oversight mechanisms. Let me provide more details about it:

- Description: GPQA consists of 448 multiple-choice questions meticulously crafted by domain experts in biology, physics, and chemistry. These questions are intentionally designed to be high-quality and extremely difficult.
- Expert Accuracy: Even experts who hold or are pursuing PhDs in the corresponding domains achieve only 65% accuracy on these questions (or 74% when excluding clear mistakes identified in retrospect).
- Google-Proof: The questions are "Google-proof," meaning that even with unrestricted access to the web, highly skilled non-expert validators only reach an accuracy of 34% despite spending over 30 minutes searching for answers.
- AI Systems Difficulty: State-of-the-art AI systems, including our strongest GPT-4 based baseline, achieve only 39% accuracy on this challenging dataset.

SimpleQA（事实性问答题）

https://openai.com/index/introducing-simpleqa/

A factuality benchmark called SimpleQA that measures the ability for language models to answer short, fact-seeking questions.



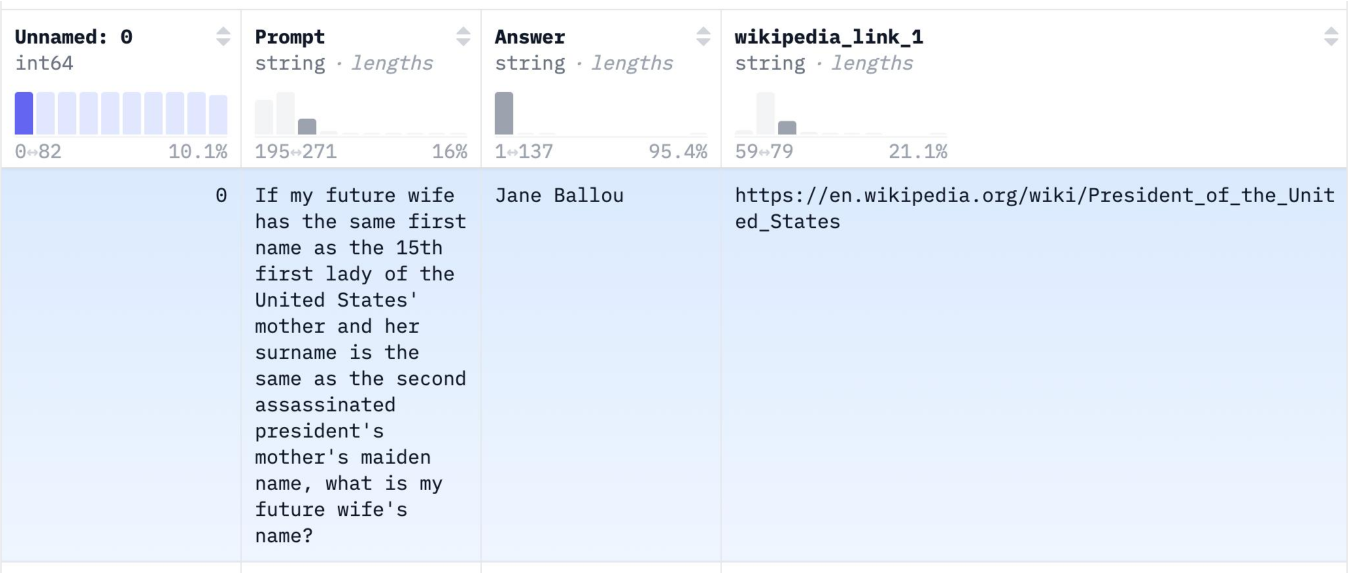
metadata string · *lengths* 125↔350 59.2%	problem string · *lengths* 26↔62 15.9%	answer string · *lengths* 1↔34 95.7%
{'topic': 'Science and technology', 'answer_type': 'Person', 'urls': ['https://en.wikipedia.org/wiki/IEEE_Frank_Rosenblatt_Award', 'https://ieeexplore.ieee.org/author/37271220500', 'https://en.wikipedia.org/wiki/IEEE_Frank_Rosenblatt_Award', 'https://www.nxtbook.com/nxtbooks/ieee/awards_2010/index.php?startid=21#/p/20']}	Who received the IEEE Frank Rosenblatt Award in 2010?	Michio Sugeno

FRAMES: Factuality, Retrieval, And reasoning MEasurement Set（事实性问答）

FRAMES is a comprehensive evaluation dataset designed to test the capabilities of Retrieval-Augmented Generation (RAG) systems across factuality, retrieval accuracy, and reasoning. Our paper with details and experiments is available on arXiv: https://arxiv.org/abs/2409.12941.

- 824 challenging multi-hop questions requiring information from 2-15 Wikipedia articles
- Questions span diverse topics including history, sports, science, animals, health, etc.
- Each question is labeled with reasoning types: numerical, tabular, multiple constraints, temporal, and post-processing
- Gold answers and relevant Wikipedia articles provided for each question

https://huggingface.co/datasets/google/frames-benchmark?row=0



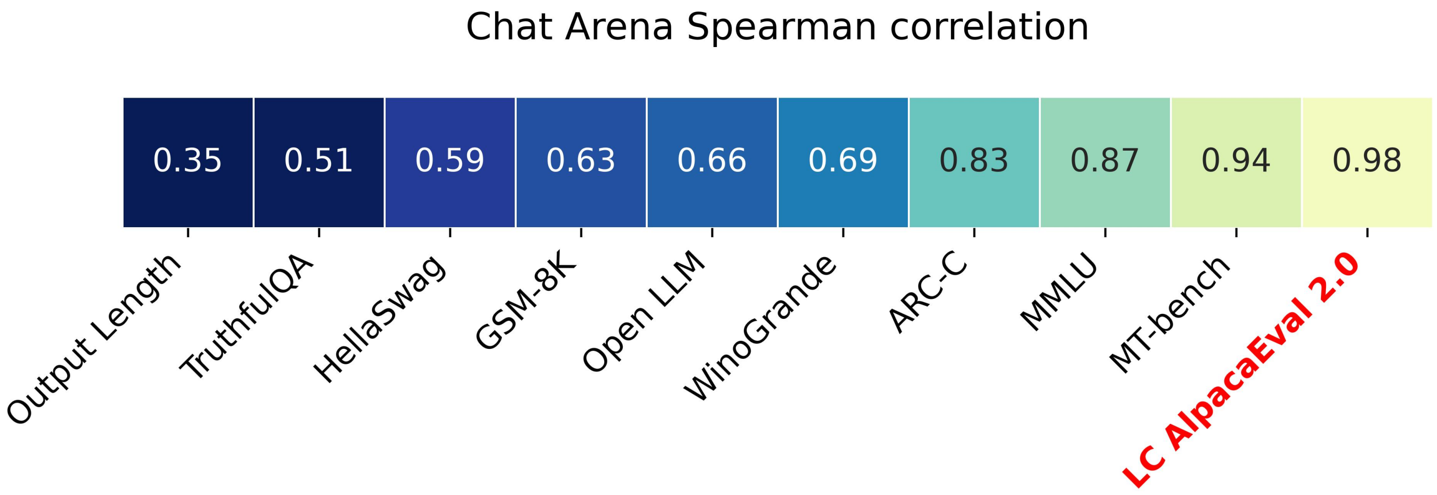
Unnamed: 0	**Prompt**	**Answer**	**wikipedia_link_1**
int64	string · *lengths*	string · *lengths*	string · *lengths*
0↔82 10.1%	195↔271 16%	1↔137 95.4%	59↔79 21.1%
0	If my future wife has the same first name as the 15th first lady of the United States' mother and her surname is the same as the second assassinated president's mother's maiden name, what is my future wife's name?	Jane Ballou	https://en.wikipedia.org/wiki/President_of_the_United_States

AlpacaEval https://github.com/tatsu-lab/alpaca_eval?tab=readme-ov-file#data-release

Evaluation of instruction-following models (e.g., ChatGPT) typically requires human interactions. This is time-consuming, expensive, and hard to replicate. AlpacaEval in an LLM-based automatic evaluation that is fast, cheap, replicable, and validated against 20K human annotations.

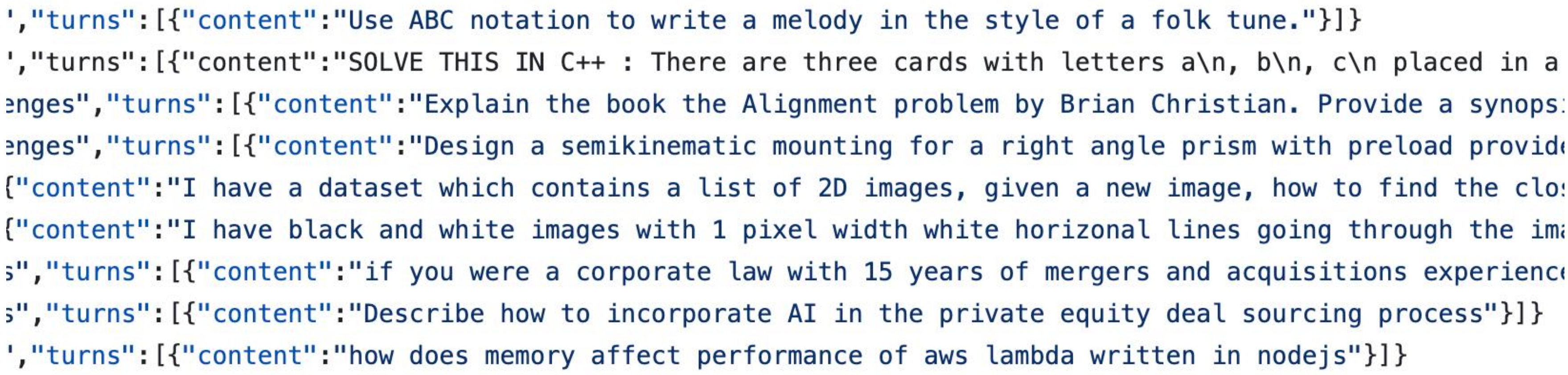
AlpacaEval 2.0 with length-controlled win-rates (paper) has a spearman correlation of 0.98 with ChatBot Arena while costing less than $10 of OpenAI credits run and running in less than 3 minutes. Our goal is to have a benchmark for chat LLMs that is: fast (< 5min), cheap (< $10), and highly correlated with humans (0.98). Here's a comparison with other benchmarks:

一个自动评估模型指令跟随能力的测试集，和人类打分的相似度能达到0.98



Areana-Hard 指令跟随测试 https://github.com/lmarena/arena-hard-auto

Arena-Hard-Auto-v0.1 (See Paper) is an automatic evaluation tool for instruction-tuned LLMs. It contains 500 challenging user queries sourced from Chatbot Arena. We prompt GPT-4-Turbo as judge to compare the models' responses against a baseline model (default: GPT-4-0314). Notably, Arena-Hard-Auto has the highest correlation and separability to Chatbot Arena among popular open-ended LLM benchmarks (See Paper).



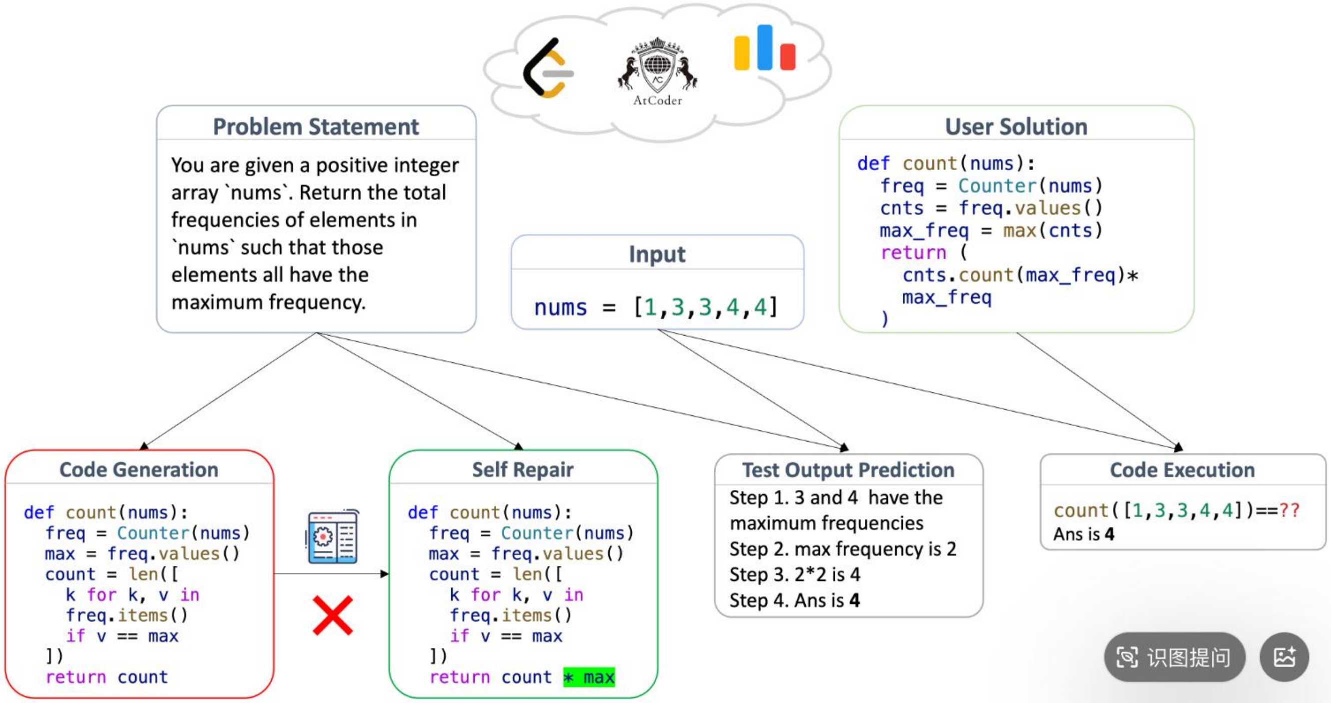
```
',"turns":[{"content":"Use ABC notation to write a melody in the style of a folk tune."}]}
',"turns":[{"content":"SOLVE THIS IN C++ : There are three cards with letters a\n, b\n, c\n placed in a
enges","turns":[{"content":"Explain the book the Alignment problem by Brian Christian. Provide a synops
enges","turns":[{"content":"Design a semikinematic mounting for a right angle prism with preload provid
{"content":"I have a dataset which contains a list of 2D images, given a new image, how to find the clos
{"content":"I have black and white images with 1 pixel width white horizonal lines going through the ima
s","turns":[{"content":"if you were a corporate law with 15 years of mergers and acquisitions experienc
s","turns":[{"content":"Describe how to incorporate AI in the private equity deal sourcing process"}]}
',"turns":[{"content":"how does memory affect performance of aws lambda written in nodejs"}]}
```

Code

LiveCodeBench

https://livecodebench.github.io/

LiveCodeBench is a holistic and contamination-free evaluation benchmark of LLMs for code that continuously collects new problems over time. Particularly, LiveCodeBench also focuses on broader code-related capabilities, such as self-repair, code execution, and test output prediction, beyond mere code generation. Currently, LiveCodeBench hosts over 300 high-quality coding problems published between May 2023 and February 2024. We evaluate 29 LLMs on LiveCodeBench scenarios and present novel empirical findings not revealed in prior benchmarks.



Problem Statement

You are given a positive integer array `nums`. Return the total frequencies of elements in `nums` such that those elements all have the maximum frequency.

Input

```
nums = [1,3,3,4,4]
```

User Solution

```
def count(nums):
  freq = Counter(nums)
  cnts = freq.values()
  max_freq = max(cnts)
  return (
    cnts.count(max_freq)*
    max_freq
  )
```

Code Generation

```
def count(nums):
  freq = Counter(nums)
  max = freq.values()
  count = len([
    k for k, v in
    freq.items()
    if v == max
  ])
  return count
```

Self Repair

```
def count(nums):
  freq = Counter(nums)
  max = freq.values()
  count = len([
    k for k, v in
    freq.items()
    if v == max
  ])
  return count * max
```

Test Output Prediction

Step 1. 3 and 4 have the maximum frequencies
Step 2. max frequency is 2
Step 3. 2*2 is 4
Step 4. Ans is **4**

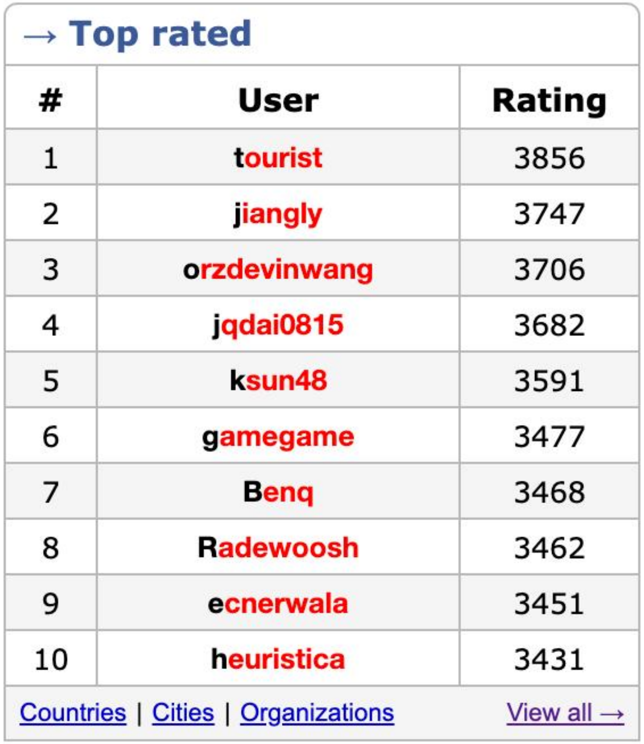
Code Execution

```
count([1,3,3,4,4])==??
```

Ans is **4**

LiveCodeBench collects problems from periodic contests on LeetCode, AtCoder, and Codeforces platforms and uses them for constructing a holistic benchmark for evaluating Code LLMs across variety of code-related scenarios continuously over time.

Codeforces 代码排行榜，类似于ACM比赛
https://codeforces.com/help#q1
简而言之，在按照 Codeforces 规则举行的竞赛中，你要为竞赛中的问题编写解决方案，这些方案在竞赛期间会在非常少量的测试中进行测试。那些通过了那组解决方案测试的人，他们的作者可以进行封锁（即使发现错误也拒绝在未来重新提交此任务的解决方案）。这样的作者有机会查看其他参赛者的源代码，在那里寻找错误，并提出这些解决方案不起作用的测试。因此，你可以攻击别人的解决方案并通过它获得分数。竞赛结束后，所有通过了预测试且未被攻击的解决方案将在最终的一组测试中进行测试。在竞赛期间，任务的价值会下降（你解决问题的速度越快，获得的分数就越多），不成功的攻击会扣除分数，成功的攻击会增加分数。



→ Top rated

#	User	Rating
1	tourist	3856
2	jiangly	3747
3	orzdevinwang	3706
4	jqdai0815	3682
5	ksun48	3591
6	gamegame	3477
7	Benq	3468
8	Radewoosh	3462
9	ecnerwala	3451
10	heuristica	3431

Countries | Cities | Organizations View all →

SWE Verified
https://www.kaggle.com/datasets/harrywang/swe-bench-verified

SWE-bench Verified is a subset of 500 samples from the SWE-bench test set, which have been human-validated for quality. SWE-bench is a dataset that tests systems' ability to solve GitHub issues automatically. See this post for more details on the human-validation process.

The dataset collects 500 test Issue-Pull Request pairs from popular Python repositories. Evaluation is performed by unit test verification using post-PR behavior as the reference solution.

The original SWE-bench dataset was released as part of SWE-bench: Can Language Models Resolve Real-World GitHub Issues?

Polyglot leaderboard
https://aider.chat/docs/leaderboards/

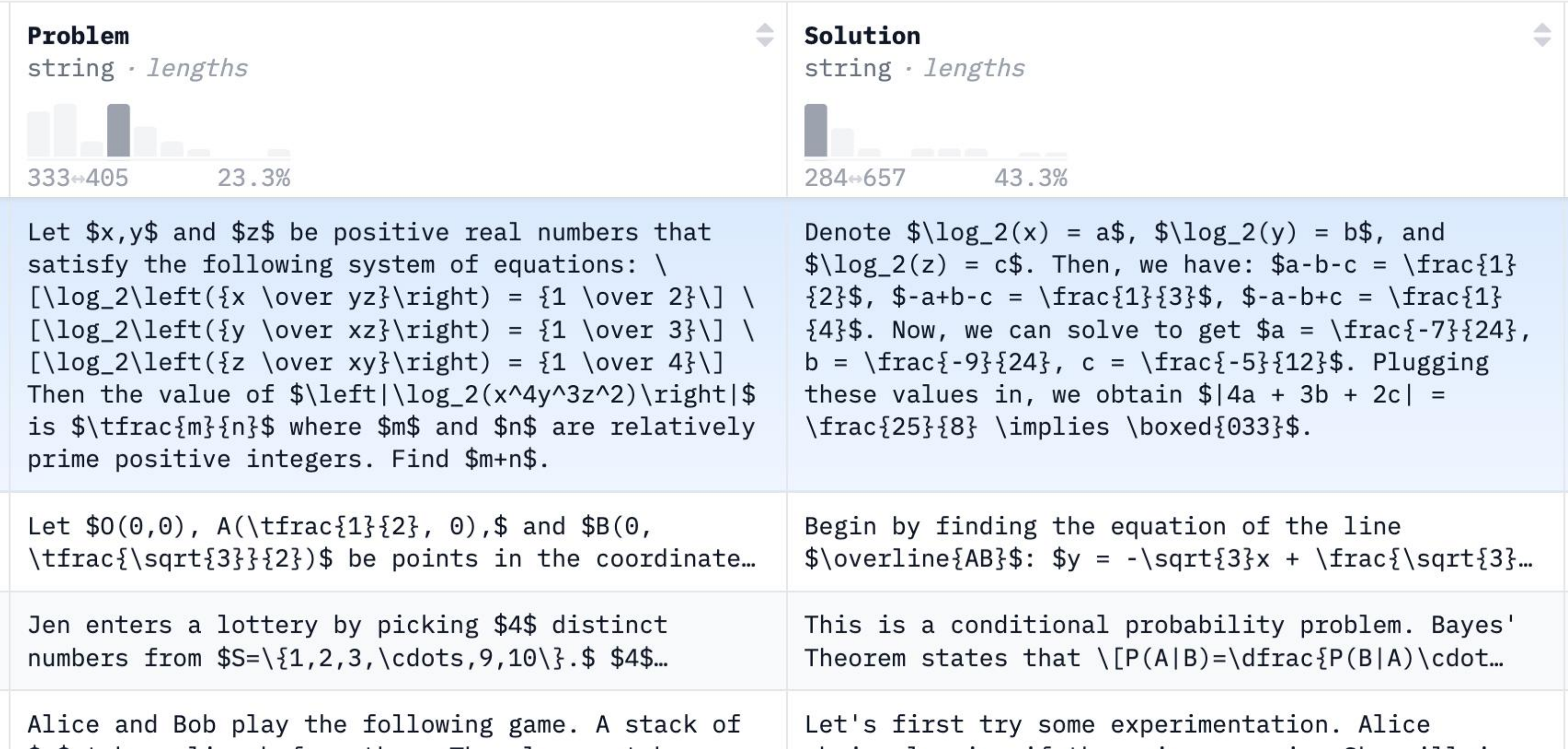
Aider's polyglot benchmark asks the LLM to edit source files to complete 225 coding exercises from Exercism. It contains exercises in many popular programming languages: C++, Go, Java, JavaScript, Python and Rust. The 225 exercises were purposely selected to be the hardest that Exercism offered in those languages, to provide a strong coding challenge to LLMs.

This benchmark measures the LLM's coding ability in popular languages, and whether it can write new code that integrates into existing code. The model also has to successfully apply all its changes to the source file without human intervention.

Math

AIME 2024（美国高中数学竞赛题，共30道）
https://huggingface.co/datasets/Maxwell-Jia/AIME_2024?row=0

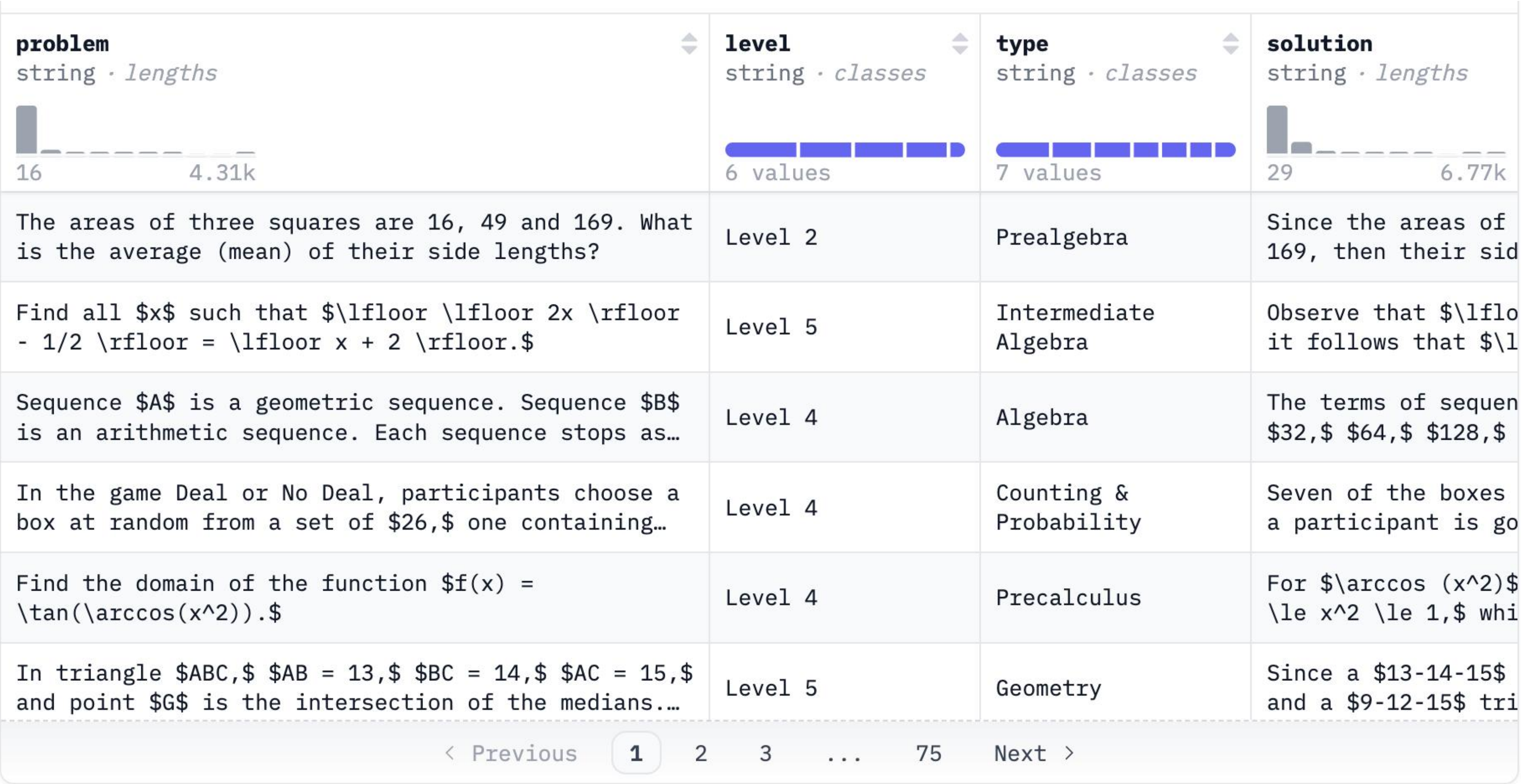
This dataset contains problems from the American Invitational Mathematics Examination (AIME) 2024. AIME is a prestigious high school mathematics competition known for its challenging mathematical problems.



Problem string · *lengths* 333↔405 23.3%	**Solution** string · *lengths* 284↔657 43.3%
Let x,y and z be positive real numbers that satisfy the following system of equations: \ [\log_2\left({x \over yz}\right) = {1 \over 2}\] \ [\log_2\left({y \over xz}\right) = {1 \over 3}\] \ [\log_2\left({z \over xy}\right) = {1 \over 4}\] Then the value of $\left\|\log_2(x^4y^3z^2)\right\|$ is $\tfrac{m}{n}$ where m and n are relatively prime positive integers. Find $m+n$.	Denote $\log_2(x) = a$, $\log_2(y) = b$, and $\log_2(z) = c$. Then, we have: $a-b-c = \frac{1}{2}$, $-a+b-c = \frac{1}{3}$, $-a-b+c = \frac{1}{4}$. Now, we can solve to get $a = \frac{-7}{24}, b = \frac{-9}{24}, c = \frac{-5}{12}$. Plugging these values in, we obtain $\|4a + 3b + 2c\| = \frac{25}{8} \implies \boxed{033}$.
Let $O(0,0), A(\tfrac{1}{2}, 0),$ and $B(0, \tfrac{\sqrt{3}}{2})$ be points in the coordinate…	Begin by finding the equation of the line $\overline{AB}$: $y = -\sqrt{3}x + \frac{\sqrt{3}…
Jen enters a lottery by picking 4 distinct numbers from $S=\{1,2,3,\cdots,9,10\}.$ 4…	This is a conditional probability problem. Bayes' Theorem states that \[P(A|B)=\dfrac{P(B|A)\cdot…
Alice and Bob play the following game. A stack of	Let's first try some experimentation. Alice

MATH-500
https://huggingface.co/datasets/ankner/math-500

500道数学试题，涵盖多个题目类型（几何、代数等）和题目难度（level1-5）



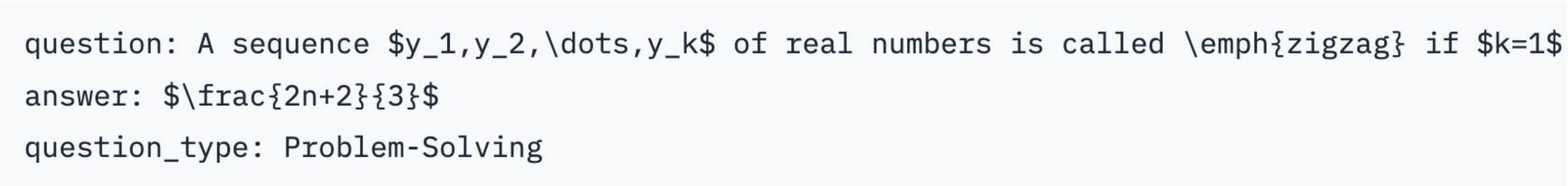
problem string · *lengths* 16 4.31k	level string · *classes* 6 values	type string · *classes* 7 values	solution string · *lengths* 29 6.77k
The areas of three squares are 16, 49 and 169. What is the average (mean) of their side lengths?	Level 2	Prealgebra	Since the areas of 169, then their sid
Find all x such that $\lfloor \lfloor 2x \rfloor - 1/2 \rfloor = \lfloor x + 2 \rfloor.$	Level 5	Intermediate Algebra	Observe that $\lflo it follows that $\l
Sequence A is a geometric sequence. Sequence B is an arithmetic sequence. Each sequence stops as…	Level 4	Algebra	The terms of sequen $32,$ $64,$ $128,$
In the game Deal or No Deal, participants choose a box at random from a set of $26,$ one containing…	Level 4	Counting & Probability	Seven of the boxes a participant is go
Find the domain of the function $f(x) = \tan(\arccos(x^2)).$	Level 4	Precalculus	For $\arccos (x^2)$ \le x^2 \le 1,$ whi
In triangle $ABC,$ $AB = 13,$ $BC = 14,$ $AC = 15,$ and point G is the intersection of the medians.…	Level 5	Geometry	Since a $13-14-15$ and a $9-12-15$ tri

‹ Previous 1 2 3 ... 75 Next ›

China National Mathematical Olympiad (CNMO) 2024

https://huggingface.co/datasets/opencompass/LiveMathBench

中国数学奥林匹克竞赛试题



```
question: A sequence $y_1,y_2,\dots,y_k$ of real numbers is called \emph{zigzag} if $k=1$
answer: $\frac{2n+2}{3}$
question_type: Problem-Solving
```

Chinese

CLUEWSC2020

git项目地址

Winograd Scheme Challenge（WSC）是一类代词消歧的任务。新版与原CLUE项目WSC内容不同

即判断句子中的代词指代的是哪个名词。题目以真假判别的方式出现，如：

句子：这时候放在床上枕头旁边的手机响了，我感到奇怪，因为欠费已被停机两个月，现在它突然响了。需要判断“它”指代的是“床”、“枕头”，还是“手机”？

数据来源：数据有CLUE benchmark提供，从中国现当代作家文学作品中抽取，再经语言专家人工挑选、标注。

数据形式：

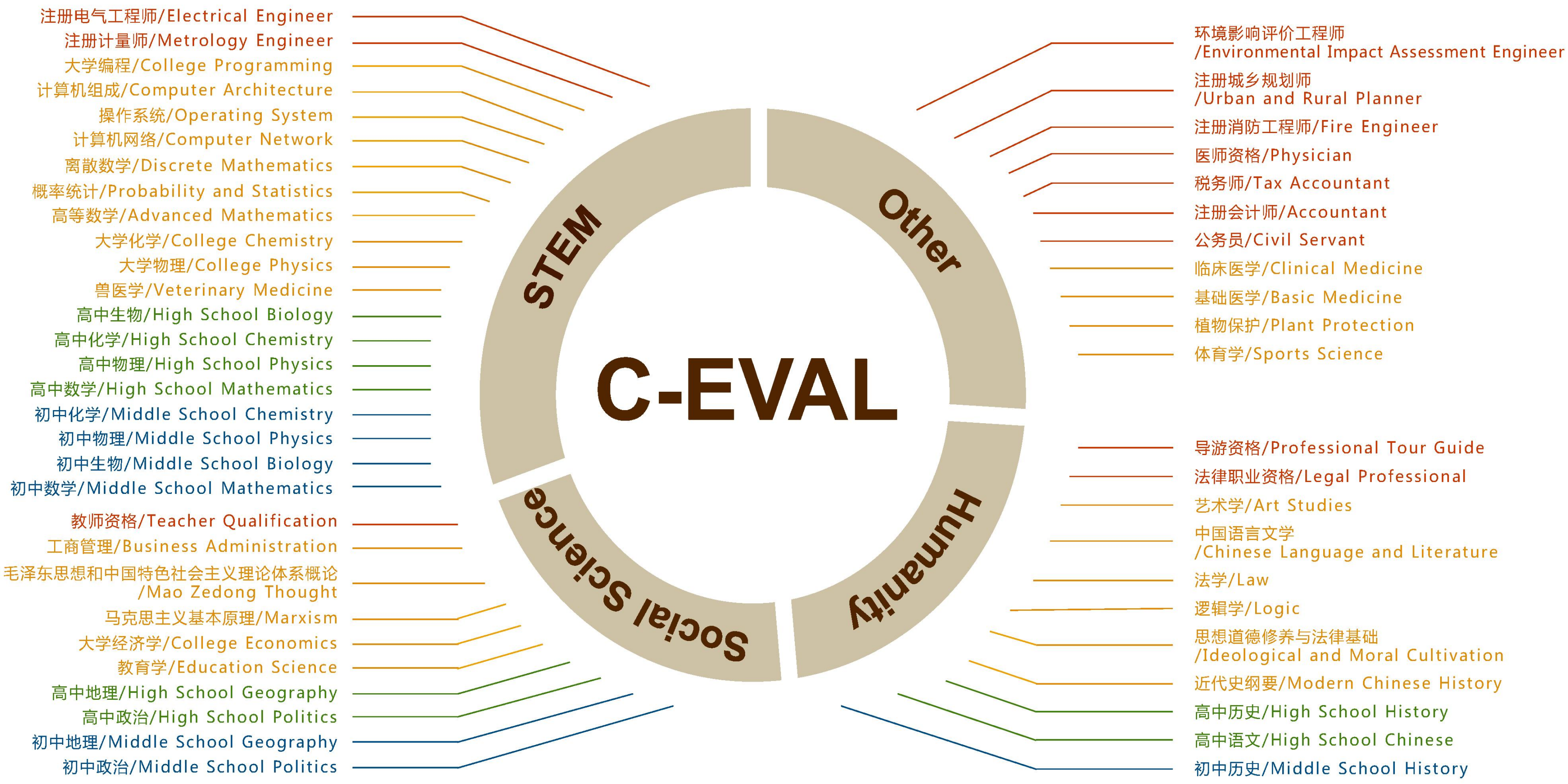
{"target": {"span2_index": 37, "span1_index": 5, "span1_text": "床", "span2_text": "它"}, "idx": 261, "label": "false", "text": "这时候放在床上枕头旁边的手机响了，我感到奇怪，因为欠费已被停机两个月，现在它突然响了。"} "true"表示代词确实是指代span1_text中的名词的，"false"代表不是。 数据集大小：

训练集：1244

开发集：304

C-Eval 是一个全面的中文基础模型评估套件。它包含了13948个**多项选择题**，涵盖了52个不同的学科和四个难度级别。

https://cevalbenchmark.com/index_zh.html



C-EVAL

STEM

注册电气工程师/Electrical Engineer
注册计量师/Metrology Engineer
大学编程/College Programming
计算机组成/Computer Architecture
操作系统/Operating System
计算机网络/Computer Network
离散数学/Discrete Mathematics
概率统计/Probability and Statistics
高等数学/Advanced Mathematics
大学化学/College Chemistry
大学物理/College Physics
兽医学/Veterinary Medicine
高中生物/High School Biology
高中化学/High School Chemistry
高中物理/High School Physics
高中数学/High School Mathematics
初中化学/Middle School Chemistry
初中物理/Middle School Physics
初中生物/Middle School Biology
初中数学/Middle School Mathematics

Social Science

教师资格/Teacher Qualification
工商管理/Business Administration
毛泽东思想和中国特色社会主义理论体系概论
/Mao Zedong Thought
马克思主义基本原理/Marxism
大学经济学/College Economics
教育学/Education Science
高中地理/High School Geography
高中政治/High School Politics
初中地理/Middle School Geography
初中政治/Middle School Politics

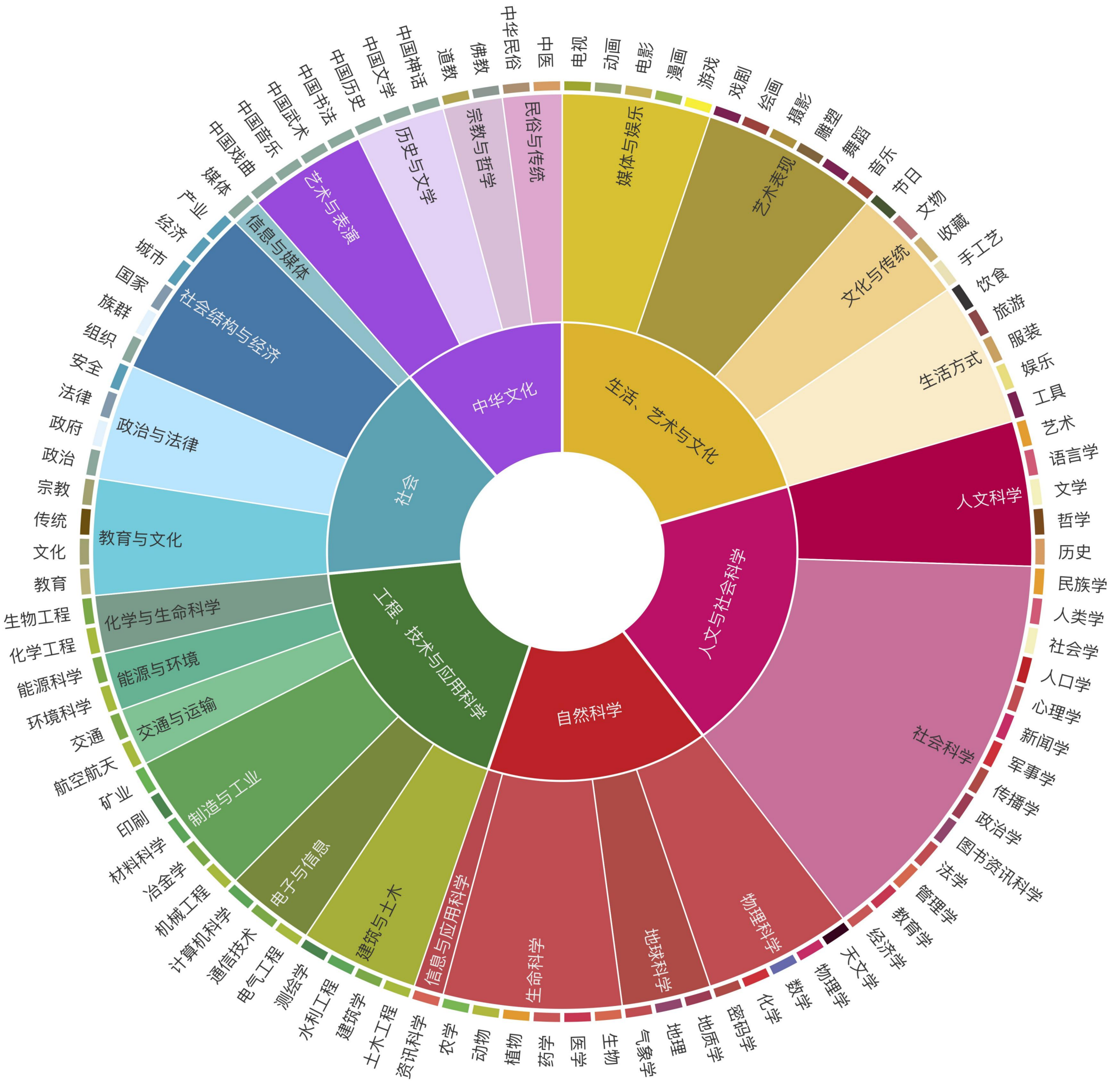
Other

环境影响评价工程师
/Environmental Impact Assessment Engineer
注册城乡规划师
/Urban and Rural Planner
注册消防工程师/Fire Engineer
医师资格/Physician
税务师/Tax Accountant
注册会计师/Accountant
公务员/Civil Servant
临床医学/Clinical Medicine
基础医学/Basic Medicine
植物保护/Plant Protection
体育学/Sports Science

Humanity

导游资格/Professional Tour Guide
法律职业资格/Legal Professional
艺术学/Art Studies
中国语言文学
/Chinese Language and Literature
法学/Law
逻辑学/Logic
思想道德修养与法律基础
/Ideological and Moral Cultivation
近代史纲要/Modern Chinese History
高中历史/High School History
高中语文/High School Chinese
初中历史/Middle School History

Chinese SimpleQA是首个简短事实问答能力的中文评测集，用于评估语言模型回答简短问题的真实性，主要有五个特点（即中文、多样化、高质量、静态、易于评估）。我们的基准涵盖6 个主要主题和99 个多样化子主题。
https://github.com/OpenStellarTeam/ChineseSimpleQA/blob/master/README_zh.md


中华文化
生活、艺术与文化
人文与社会科学
自然科学
工程、技术与应用科学
社会
民俗与传统
宗教与哲学
历史与文学
艺术与表演
信息与媒体
社会结构与经济
政治与法律
教育与文化
化学与生命科学
能源与环境
交通与运输
制造与工业
电子与信息
建筑与土木
信息与应用科学
生命科学
地球科学
物理科学
社会科学
人文科学
生活方式
文化与传统
艺术表现
媒体与娱乐
中医
电视
动画
电影
漫画
游戏
戏剧
绘画
摄影
雕塑
舞蹈
音乐
节日
文物
收藏
手工艺
饮食
旅游
服装
娱乐
工具
艺术
语言学
文学
哲学
历史
民族学
人类学
社会学
人口学
心理学
新闻学
军事学
传播学
政治学
图书资讯科学
法学
管理学
教育学
经济学
天文学
物理学
数学
化学
密码学
地质学
地理
气象学
生物
医学
药学
植物
动物
农学
资讯科学
土木工程
建筑学
水利工程
测绘学
电气工程
通信技术
计算机科学
机械工程
冶金学
材料科学
印刷
矿业
航空航天
交通
环境科学
能源科学
化学工程
生物工程
教育
文化
传统
宗教
政治
政府
法律
安全
组织
族群
国家
城市
经济
产业
媒体
中国戏曲
中国音乐
中国武术
中国书法
中国历史
中国文学
中国神话
道教
佛教
中华民俗