

## s1: Simple test-time scaling

Niklas Muennighoff<sup>\*134</sup> Zitong Yang<sup>\*1</sup> Weijia Shi<sup>\*2</sup> Xiang Lisa Li<sup>\*1</sup> Li Fei-Fei<sup>1</sup>  
Hannaneh Hajishirzi<sup>23</sup> Luke Zettlemoyer<sup>2</sup> Percy Liang<sup>1</sup> Emmanuel Candès<sup>1</sup> Tatsunori Hashimoto<sup>1</sup>

### Abstract

Test-time scaling is a promising new approach to language modeling that uses extra test-time compute to improve performance. Recently, OpenAI’s o1 model showed this capability but did not publicly share its methodology, leading to many replication efforts. We seek the simplest approach to achieve test-time scaling and strong reasoning performance. First, we curate a small dataset **s1K** of 1,000 questions paired with reasoning traces relying on three criteria we validate through ablations: difficulty, diversity, and quality. Second, we develop budget forcing to control test-time compute by forcefully terminating the model’s thinking process or lengthening it by appending “Wait” multiple times to the model’s generation when it tries to end. This can lead the model to double-check its answer, often fixing incorrect reasoning steps. After supervised finetuning the Qwen2.5-32B-Instruct language model on **s1K** and equipping it with budget forcing, our model **s1-32B** exceeds o1-preview on competition math questions by up to 27% (MATH and AIME24). Further, scaling **s1-32B** with budget forcing allows extrapolating beyond its performance without test-time intervention:

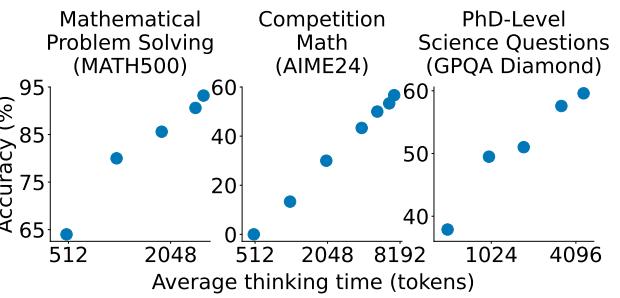


Figure 1. Test-time scaling with **s1-32B**. We benchmark **s1-32B** on reasoning-intensive tasks and vary test-time compute.

from 50% to 57% on AIME24. Our model, data, and code are open-source at <https://github.com/simplescaling/s1>.

### 1. Introduction

Performance improvements of language models (LMs) over the past years have largely relied on scaling up train-time compute using large-scale self-supervised pretraining (Kaplan et al., 2020; Hoffmann et al., 2022). The creation of these powerful models has set the stage for a new scaling paradigm built on top of them: *test-time scaling*. The aim of this approach is to increase the compute at test time to get better results. There has been much work exploring this idea (Snell et al., 2024; Welleck et al., 2024), and the viability of this paradigm was recently validated by OpenAI o1 (OpenAI, 2024). o1 has demonstrated strong reasoning performance with consistent gains from scaling test-time compute. OpenAI describes their approach as using large-scale reinforcement learning (RL) implying the use of sizable amounts of data (OpenAI, 2024). This has led to various attempts to repli-

<sup>\*</sup>Equal contribution. ZY and NM started the project. WS, NM and ZY collected the prompts, XL, ZY and NM, built the data pipeline, LZ and WS proposed using a 1K subset and ZY and NM built budget forcing. <sup>1</sup> Stanford University. <sup>2</sup> University of Washington, Seattle. <sup>3</sup> Allen Institute for AI. <sup>4</sup> Contextual AI.

## s1: 简单的测试时缩放

Niklas Muennighoff<sup>\*134</sup> Zitong Yang<sup>\*1</sup> Weijia Shi<sup>\*2</sup> Xiang Lisa Li<sup>\*1</sup> Li Fei-Fei<sup>1</sup>  
Hannaneh Hajishirzi<sup>23</sup> Luke Zettlemoyer<sup>2</sup> Percy Liang<sup>1</sup> Emmanuel Candès<sup>1</sup> Tatsunori Hashimoto<sup>1</sup>

### Abstract

\*警 告：该PDF由GPT-Academic开源项目调用大语言模型+Latex翻译插件一键生成，版权归原文作者所有。翻译内容可靠性无保障，请仔细鉴别并以原文为准。项目Github地址 [https://github.com/binary-husky/gpt\\_academic/](https://github.com/binary-husky/gpt_academic/)。项目在线体验地址 <https://auth.gpt-academic.top/>。当前大语言模型: Qwen2.5-72B-Instruct, 当前语言模型温度设定: 0.3。为了防止大语言模型的意外谬误产生扩散影响, 禁止移除或修改此警告。

测试时扩展是语言建模的一种有前景的新方法, 它利用额外的测试时计算资源来提高性能。最近, OpenAI 的 o1 模型展示了这种能力, 但没有公开其方法, 导致了许多复制尝试。我们寻求实现测试时扩展和强大推理性能的最简单方法。首先, 我们整理了一个包含 1,000 个问题及其推理轨迹的小型数据集 **s1K**, 这些数据集基于我们通过消融验证的三个标准: 难度、多样性和质量。其次, 我们开发了预算强制技术来控制测试时的计算资源, 通过强制终止模型的思考过程或在模型尝试结束时多次追加“等待”来延长其生成过程。这可以引导模型重新检查其答案, 通常会修正错误的推理步骤。在 **s1K** 上对 Qwen2.5-32B-Instruct 语言模型进行监督微调并配备预算强制后, 我

<sup>\*</sup>Equal contribution. ZY and NM started the project. WS, NM and ZY collected the prompts, XL, ZY and NM, built the data pipeline, LZ and WS proposed using a 1K subset and ZY and NM built budget forcing. <sup>1</sup> Stanford University. <sup>2</sup> University of Washington, Seattle. <sup>3</sup> Allen Institute for AI. <sup>4</sup> Contextual AI.

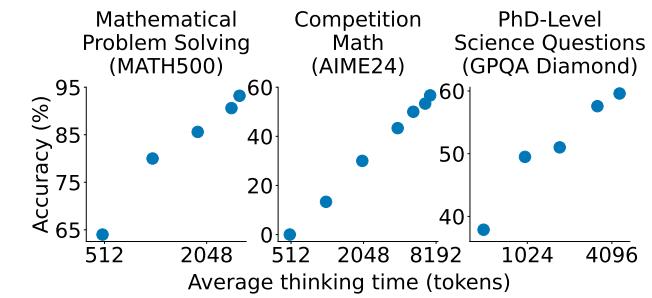


Figure 1. 测试时使用 **s1-32B** 的计算扩展。我们在推理密集型任务上对 **s1-32B** 进行基准测试, 并改变测试时的计算资源。

们的模型 **s1-32B** 在竞赛数学问题上超过了 o1-preview, 最高提升了 27% (MATH 和 AIME24)。此外, 通过预算强制扩展 **s1-32B** 可以使其性能超出没有测试时干预的情况: 从 AIME24 的 50% 提高到 57%。我们的模型、数据和代码在 <https://github.com/simplescaling/s1> 开源。

### 1. Introduction

近年来, 语言模型 (LMs) 性能的提升在很大程度上依赖于使用大规模自监督预训练来增加训练时的计算量 (Kaplan et al., 2020; Hoffmann et al., 2022)。这些强大模型的创建为新的扩展范式奠定了基础: 测试时扩展。这种方法的目的是在测试时增加计算量以获得更好的结果。许多研究探索了这一想法 (Snell et al., 2024; Welleck et al., 2024), 并且这一范式的可行性最近被OpenAI的o1模型所验证 (OpenAI, 2024)。o1展示了强大的推理性能, 并且从测试时计算量的扩展中获得了持续的收益。OpenAI将其方法

cate their models relying on techniques like Monte Carlo Tree Search (Gao et al., 2024b; Zhang et al., 2024a), multi-agent approaches (Qin et al., 2024), and others (Wang et al., 2024a; Huang et al., 2024b; 2025). Among these approaches, DeepSeek R1 (DeepSeek-AI et al., 2025) has successfully replicated o1-level performance, also employing reinforcement learning via millions of samples and multiple training stages. However, despite the large number of o1 replication attempts, none have openly replicated a clear test-time scaling behavior. Thus, we ask: what is the simplest approach to achieve both test-time scaling and strong reasoning performance?

We show that training on only 1,000 samples with next-token prediction and controlling thinking duration via a simple test-time technique we refer to as *budget forcing* leads to a strong reasoning model that scales in performance with more test-time compute. Specifically, we construct **s1K**, which consists of 1,000 carefully curated questions paired with reasoning traces and answers distilled from Gemini Thinking Experimental (Google, 2024). We perform supervised fine-tuning (SFT) of an off-the-shelf pretrained model on our small dataset requiring just 26 minutes of training on 16 H100 GPUs. After training, we control the amount of test-time compute our model spends using *budget forcing*: (I) If the model generates more thinking tokens than a desired limit, we forcefully end the thinking process by appending an end-of-thinking token delimiter. Ending the thinking this way makes the model transition to generating its answer. (II) If we want the model to spend more test-time compute on a problem, we suppress the generation of the end-of-thinking token delimiter and instead append “Wait” to the model’s current reasoning trace to encourage more exploration. Equipped with this simple recipe – SFT on 1,000 samples and test-time budget forcing – our model **s1-32B** exhibits test-time scaling (Figure 1). Further, **s1-32B** is the most sample-efficient reasoning model and outperforms closed-source models like OpenAI’s o1-preview (Figure 2).

We conduct extensive ablation experiments targeting (a) our selection of 1,000 (1K) reasoning samples and (b) our test-time scaling. For (a), we find that jointly incorporating difficulty, diversity, and quality measures into our selection algorithm is important. Random selection, selecting samples with the longest reasoning traces, or only selecting maximally diverse samples all lead to significantly worse performance (around -30% on AIME24 on average). Training on our full data pool of 59K examples, a superset of **s1K**, does not offer substantial gains over our 1K selection. This highlights the importance of careful data selection and echoes prior findings for instruction tuning (Zhou et al., 2023). For (b), we define desiderata for test-time scaling methods to compare different approaches. Budget forcing leads to the best scaling as it has perfect controllability with a clear positive slope leading to strong performance.

In summary, our contributions are: We develop simple methods for creating a sample-efficient reasoning dataset (§2) and test-time scaling (§3); Based on these we build **s1-32B** which is competitive with o1-preview (§4); We ablate subtleties of data (§5.1) and test-time scaling (§5.2). We end with a discussion to motivate future work on simple reasoning (§6). Our code, model, and data are open-source at <https://github.com/simplescaling/s1>.

## 2. Reasoning data curation to create s1K

In this section, we describe our process for creating a large dataset first in §2.1 and then filtering it down to **s1K** in §2.2.

### 2.1. Initial collection of 59K samples

We collect an initial 59,029 questions from 16 diverse sources following three guiding principles. **Quality**: Datasets should be of high quality; we always inspect samples and ignore datasets with, e.g., poor formatting; **Difficulty**: Datasets should be challenging and require

描述为使用大规模强化学习 (RL)，这意味着使用了大量数据 (OpenAI, 2024)。这导致了各种尝试复制其模型的努力，这些努力依赖于诸如蒙特卡洛树搜索 (Gao et al., 2024b; Zhang et al., 2024a)、多代理方法 (Qin et al., 2024) 以及其他方法 (Wang et al., 2024a; Huang et al., 2024b; 2025)。在这些方法中，DeepSeek R1 (DeepSeek-AI et al., 2025) 成功复制了 o1 级别的性能，同样使用了数百万样本和多个训练阶段的强化学习。然而，尽管有大量尝试复制 o1 模型，但没有一个公开复制了明确的测试时扩展行为。因此，我们问：实现测试时扩展和强大推理性能的最简单方法是什么？

我们展示了一个仅在 1,000 个样本上进行训练，使用下一个标记预测，并通过我们称为预算强制的简单测试时技术控制思考时间的方法，可以生成一个性能随测试时计算量增加而扩展的强大推理模型。具体来说，我们构建了 **s1K**，它包含 1,000 个精心策划的问题，这些问题附有从 Gemini Thinking Experimental (Google, 2024) 提炼出的推理轨迹和答案。我们在小型数据集上对现成的预训练模型进行监督微调 (SFT)，仅需在 16 个 H100 GPU 上训练 26 分钟。训练后，我们使用预算强制来控制模型在测试时的计算量：(I) 如果模型生成的思考标记超过预限制，我们通过附加一个思考结束标记分隔符来强制结束思考过程。这样结束思考会使模型转换为生成答案。(II) 如果我们希望模型在某个问题上花费更多测试时计算量，我们抑制思考结束标记分隔符的生成，并在模型当前的推理轨迹后附加“等待”以鼓励更多探索。通过这一简单的配方——在 1,000 个样本上进行 SFT 和测试时预算强制——我们的模型 **s1-32B** 表现出测试时扩展（见图 Figure 1）。此外，**s1-32B** 是最样本高效的推理模型，优于 OpenAI 的 o1-preview 等闭源模型（见图 Figure 2）。

我们进行了广泛的消融实验，针对 (a) 我们选择的 1,000 (1K) 个推理样本和 (b) 我们的测试时扩展。对于 (a)，我们发现将难度、多样性和质量度量共同纳入选择算法中是重要的。随机选择、选择具有最长推理轨迹的样本或仅选择最多样化的样本都会导致显著更差的性能（平均在 AIME24 上降低

约 30%）。在我们的 59K 个样本的完整数据池上进行训练，即 **s1K** 的超集，并没有比我们的 1K 选择提供实质性的收益。这突显了精心选择数据的重要性，并呼应了先前关于指令调优的发现 (Zhou et al., 2023)。对于 (b)，我们定义了测试时扩展方法的期望标准，以比较不同的方法。预算强制导致最佳的扩展，因为它具有完美的可控性，并且具有明显的正斜率，从而实现强大的性能。总之，我们的贡献包括：我们开发了创建样本高效推理数据集 (§2) 和测试时缩放 (§3) 的简单方法；基于这些方法，我们构建了 **s1-32B**，其性能与 o1-preview 相当 (§4)；我们对数据 (§5.1) 和测试时缩放 (§5.2) 的细微之处进行了消融研究。最后，我们通过讨论来激励未来在简单推理方面的研究 (§6)。我们的代码、模型和数据在 <https://github.com/simplescaling/s1> 开源。

## 2. Reasoning data curation to create s1K

在本节中，我们首先描述了创建大型数据集的过程（见 §2.1），然后介绍了筛选出 **s1K** 的标准（见 §2.2）。

### 2.1. Initial collection of 59K samples

我们从 16 个不同的来源收集了初始的 59,029 个问题，遵循三个指导原则。\*\*质量\*\*：数据集应该是高质量的；我们始终检查样本并忽略格式不佳的数据集；\*\*难度\*\*：数据集应该是具有挑战性的，需要大量的推理努力；\*\*多样性\*\*：数据集应来自不同的领域，以涵盖不同的推理任务。我们收集了两类数据集：

**现有数据集的整理** 我们最大的来源是 NuminaMATH (Li et al., 2024)，其中包含来自在线网站的 30,660 个数学问题。我们还包含了 1983 年至 2021 年的历史 AIME 问题。为了增强多样性，我们加入了 OlympicArena (Huang et al., 2024a)，其中包含来自各种奥林匹克竞赛的 4,250 个问题，涵盖天文学、生物学、化学、计算机科学、地理学、数学和物理学。OmniMath (Gao et al.,

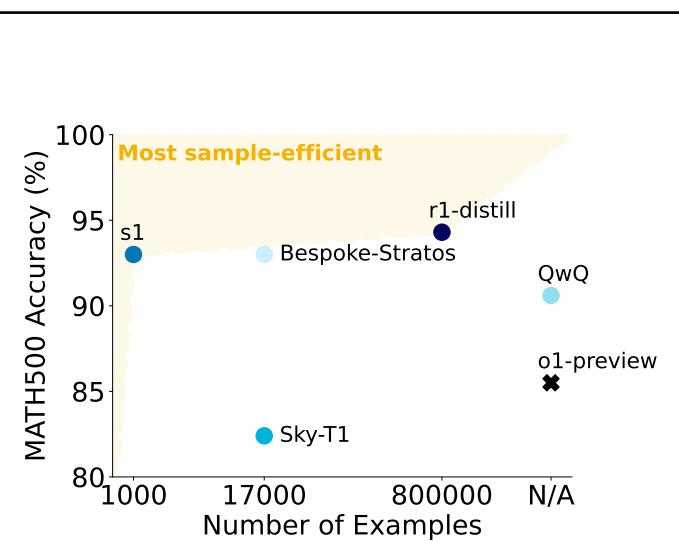


Figure 2. **s1K and s1-32B.** (left) **s1K** is a dataset of 1,000 high-quality, diverse, and difficult questions with reasoning traces. (right) **s1-32B**, a 32B parameter model finetuned on **s1K** is on the sample-efficiency frontier. See Table 1 for details on other models.

significant reasoning effort; **Diversity**: Datasets should stem from different fields to cover different reasoning tasks. We collect datasets of two categories:

**Curation of existing datasets** Our largest source is NuminaMATH (Li et al., 2024) with 30,660 mathematical problems from online websites. We also include historical AIME problems (1983–2021). To enhance diversity, we add OlympicArena (Huang et al., 2024a) with 4,250 questions spanning Astronomy, Biology, Chemistry, Computer Science, Geography, Mathematics, and Physics from various Olympiads. OmniMath (Gao et al., 2024a) adds 4,238 competition-level mathematics problems. We also include 2,385 problems from AGIEval (Zhong et al., 2023), which features questions from standardized tests like SAT and LSAT, covering English, Law, and Logic. We refer to Table 6 in §B for our other sources.

**New datasets in quantitative reasoning** To complement these existing datasets, we create two original datasets. **s1-prob** consists of 182 questions from the probability section of Stanford University’s Statistics Department’s PhD Qualifying Exams (<https://statistics.stanford.edu>), accompanied by handwritten solutions that cover difficult proofs. The

probability qualifying exam is held yearly and requires professional-level mathematical problem-solving. **s1-teasers** comprises 23 challenging brain-teasers commonly used in interview questions for quantitative trading positions. Each sample consists of a problem and solution taken from PuzzledQuant (<https://www.puzzledquant.com/>). We only take examples with the highest difficulty level (“Hard”).

For each question, we generate a reasoning trace and solution using the Google Gemini Flash Thinking API (Google, 2024) extracting its reasoning trace and response. This yields 59K triplets of a question, generated reasoning trace, and generated solution. Examples from our dataset are in §C.2. We decontaminate all samples against our evaluation questions (MATH500, GPQA Diamond, AIME24; §B.5) using 8-grams and deduplicate the data.

## 2.2. Final selection of 1K samples

We could directly train on our pool of 59K questions, however, our goal is to find the *simplest* approach with minimal resources. Thus, we go through three stages of filtering to arrive at a minimal set of 1,000 samples relying on our three guiding data principles: Quality,

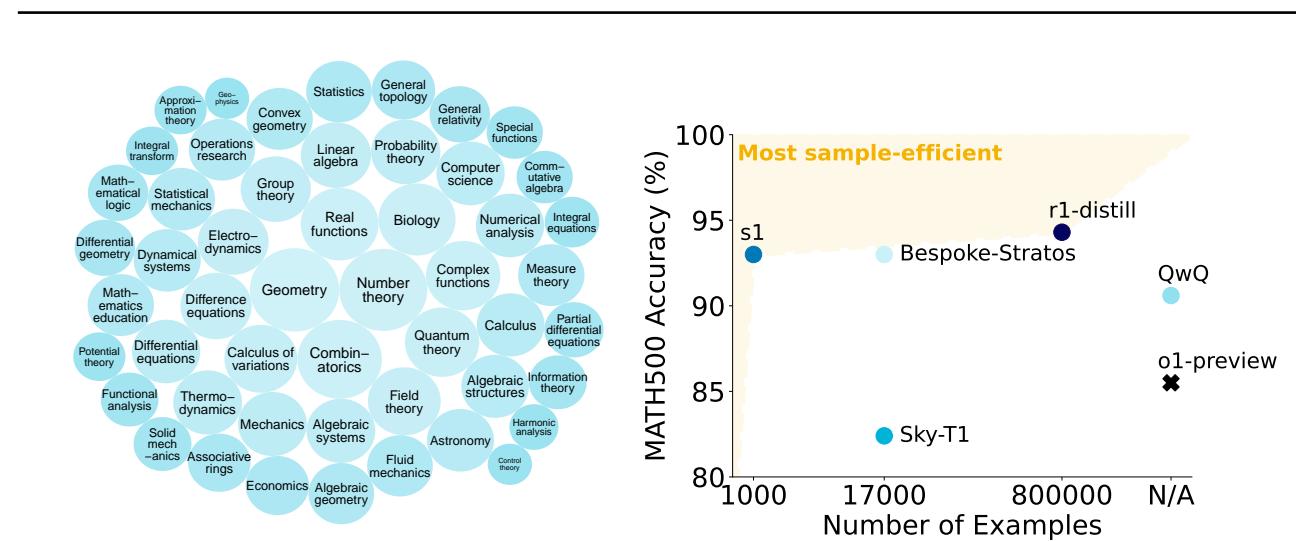


Figure 2. **s1K 和 s1-32B.** (左) **s1K** 是一个包含1,000个高质量、多样化且具有挑战性问题的数据集，每个问题都附有推理轨迹。 (右) **s1-32B** 是一个在 **s1K** 上微调的32B参数模型，位于样本效率前沿。有关其他模型的详细信息，请参见 Table 1。

2024a) 增加了4,238个竞赛级别的数学问题。我们还包含了来自AGIEval (Zhong et al., 2023) 的2,385个问题，这些问题来自SAT和LSAT等标准化测试，涵盖英语、法律和逻辑。我们参考 Table 6 在 §B 中的其他来源。

**新的定量推理数据集** 为了补充这些现有的数据集，我们创建了两个原创数据集。**s1-prob** 包含182个来自斯坦福大学统计系博士资格考试 (<https://statistics.stanford.edu>) 的概率部分的问题，附有手写解决方案，涵盖了复杂的证明。概率资格考试每年举行一次，需要专业级别的数学问题解决能力。**s1-teasers** 包含23个具有挑战性的谜题，这些谜题通常用于定量交易职位的面试问题。每个样本包括一个问题和解决方案，来自PuzzledQuant (<https://www.puzzledquant.com/>)。我们只选取难度最高的示例 (“Hard”)。

对于每个问题，我们使用Google Gemini Flash Thinking API (Google, 2024) 生成推理轨迹和解决方案，提取其推理轨迹和响应。这产生了59,000个问题、生成的推理轨迹和生成的解决方案的三元组。我们数据集中的示例见 §C.2。我们使用8-grams对所有样本进行去污染处理，并对数据进行去重，以防止与我们的评估问题 (MATH500, GPQA Diamond, AIME24; §B.5) 重复。

## 2.2. Final selection of 1K samples

我们可以直接在59K个问题的池中进行训练，然而，我们的目标是找到一种资源最少的最简单方法。因此，我们经过三个阶段的筛选，最终得到1,000个样本，这些样本基于我们的三个指导原则：质量、难度和多样性。

**质量** 首先，我们移除所有遇到API错误的问题，将数据集减少到54,116个样本。接下来，我们通过检查问题中是否存在格式问题的字符串模式（如ASCII艺术图、不存在的图像引用或不一致的问题编号）来过滤低质量的样本，将数据集减少到51,581个样本。从这个池中，我们从我们认为高质量且不需要进一步筛选的数据集中识别出384个样本，作为最终1,000个样本的一部分（详细信息见§B.4）。

**难度** 对于难度，我们使用两个指标：模型性能和推理轨迹长度。我们评估两个模型在每个问题上的表现：Qwen2.5-7B-Instruct和Qwen2.5-32B-Instruct (Qwen et al., 2024)，正确性由Claude 3.5 Sonnet评估，通过将每次尝试与参考答案进行比较（评分协议见§B.3）。我们使用Qwen2.5分词器测量每个推

Difficulty, and Diversity.

**Quality** We first remove any questions where we ran into any API errors reducing our dataset to 54,116 samples. Next, we filter out low-quality examples by checking if they contain any string patterns with formatting issues, such as ASCII art diagrams, non-existent image references, or inconsistent question numbering reducing our dataset to 51,581 examples. From this pool, we identify 384 samples for our final 1,000 samples from datasets that we perceive as high-quality and not in need of further filtering (see §B.4 for details).

**Difficulty** For difficulty, we use two indicators: model performance and reasoning trace length. We evaluate two models on each question: Qwen2.5-7B-Instruct and Qwen2.5-32B-Instruct (Qwen et al., 2024), with correctness assessed by Claude 3.5 Sonnet comparing each attempt against the reference solution (see §B.3 for the grading protocol). We measure the token length of each reasoning trace to indicate problem difficulty using the Qwen2.5 tokenizer. This relies on the assumption that more difficult problems require more thinking tokens. Based on the grading, we remove questions that either Qwen2.5-7B-Instruct or Qwen2.5-32B-Instruct can solve correctly and thus may be too easy. By using two models we reduce the likelihood of an easy sample slipping through our filtering due to a rare mistake on an easy question of one of the models. This brings our total samples down to 24,496, setting the stage for the next round of subsampling based on diversity. While filtering with these two models may be optimized for our setup as we will also use Qwen2.5-32B-Instruct as our model to finetune, the idea of model-based filtering generalizes to other setups.

**Diversity** To quantify diversity we classify each question into specific domains using Claude 3.5 Sonnet based on the Mathematics Subject Classification (MSC) system (e.g., geometry, dynamic systems, real analysis, etc.)

from the American Mathematical Society.<sup>1</sup> The taxonomy focuses on topics in mathematics but also includes other sciences such as biology, physics, and economics. To select our final examples from the pool of 24,496 questions, we first choose one domain uniformly at random. Then, we sample one problem from this domain according to a distribution that favors longer reasoning traces (see §B.4 for details) as motivated in *Difficulty*. We repeat this process until we have 1,000 total samples.

This three-stage process yields a dataset spanning 50 different domains (see Table 5). In §5.1, we will show that using our three criteria in combination is important, as only relying on quality, diversity, or difficulty in isolation leads to worse datasets. Examples from our dataset are in §C.2.

### 3. Test-time scaling

#### 3.1. Method

We classify test-time scaling methods into 1) **Sequential**, where later computations depend on earlier ones (e.g., a long reasoning trace), and 2) **Parallel**, where computations run independently (e.g., majority voting) (Snell et al., 2024; Brown et al., 2024). We focus on sequential scaling as intuitively we believe it should scale better, since later computations can build on intermediate results, allowing for deeper reasoning and iterative refinement. We propose new sequential scaling methods and ways to benchmark them.

**Budget forcing** We propose a simple decoding-time intervention by forcing a maximum and/or minimum number of thinking tokens at test time. Specifically, we enforce a maximum token count by simply appending the end-of-thinking token delimiter and “Final Answer:” to early exit the thinking stage and make the model provide its current best answer. To enforce a minimum, we suppress the generation of the end-of-

<sup>1</sup><https://mathscinet.ams.org/mathscinet/msc/msc2020.html>

理轨迹的标记长度，以表示问题的难度，这基于一个假设，即更难的问题需要更多的思考标记。根据评分，我们移除Qwen2.5-7B-Instruct或Qwen2.5-32B-Instruct能够正确解决的问题，因为这些可能太简单。通过使用两个模型，我们减少了由于模型在简单问题上的罕见错误而导致简单样本漏过筛选的可能性。这使我们的总样本数减少到24,496，为基于多样性的下一轮子采样奠定了基础。虽然使用这两个模型的筛选可能针对我们的设置进行了优化，因为我们还将使用Qwen2.5-32B-Instruct作为微调模型，但基于模型的筛选方法可以推广到其他设置。

**多样性** 为了量化多样性，我们使用Claude 3.5 Sonnet根据美国数学学会的数学主题分类（MSC）系统（例如，几何、动态系统、实分析等）将每个问题分类到特定领域。<sup>1</sup> 该分类法侧重于数学主题，但也包括生物学、物理学和经济学等其他科学。为了从24,496个问题的池中选择最终的样本，我们首先随机均匀地选择一个领域。然后，根据一个偏向于较长推理轨迹的分布从该领域中选择一个问题（详细信息见§B.4），这受到难度部分的启发。我们重复这个过程，直到总共有1,000个样本。

这个三阶段的过程产生了一个涵盖50个不同领域的数据集（见Table 5）。在§5.1中，我们将展示结合使用我们的三个标准是重要的，因为仅依赖质量、多样性和难度中的任何一个标准会导致更差的数据集。我们数据集中的示例见§C.2。

### 3. Test-time scaling

#### 3.1. Method

我们将测试时扩展方法分类为 1) **顺序**，其中后续计算依赖于早期计算（例如，长推理链），和 2) **并行**，其中计算独立运行（例如，多数投票）(Snell et al., 2024; Brown et al., 2024)。我们专注于顺序扩展，因为我们直观地认为它应该扩展得更好，因为后续计算可以建立在中间结果上，允许进行更深入的推理和迭代改进。我们提出新的顺序扩展方法及其评估

<sup>1</sup><https://mathscinet.ams.org/mathscinet/msc/msc2020.html>

How many r in raspberry?

Question

Let's break down the process of counting the letter 'r' in the word "raspberry" ...

\* First letter: 'r' - This is an 'r', count = 1.

\* Second letter: 'a' - Not an 'r', count remains 1 ...

\* Sixth letter: 'e' - Not an 'r', count remains 1.

\* Seventh letter: 'r' - This is an 'r', count = 2.

\* Eighth letter: 'y' - Not an 'r', count remains 2 ...

The number of 'r's in "raspberry" is 2.

Wait, let's re-read the question carefully. It asks "How many r in raspberry?" ... \* r - a - s - p - b - e - r - r - y ... \* First 'r' ... \* Second 'r' ... \* Third 'r' ... Count = 3 ...

Reasoning trace

My initial answer of 2 was incorrect due to a quick reading of the word. Final Answer: The final answer is 3

Response

Figure 3. 使用s1-32B进行预算强制。模型试图在“...是2。”之后停止，但我们抑制了思维结束的令牌分隔符，而是附加了“等待”，这使得s1-32B自我纠正其答案。

方法。

**预算强制** 我们提出了一种简单的解码时干预方法，通过在测试时强制执行最大和/或最小思考令牌数。具体来说，我们通过简单地附加思考结束令牌分隔符和“Final Answer:”来强制执行最大令牌数，以提前退出思考阶段并使模型提供其当前的最佳答案。为了强制执行最小值，我们抑制思考结束令牌分隔符的生成，并可选地将字符串“Wait”附加到模型当前的推理链中，以鼓励模型反思其当前的生成。Figure 3 包含了这种简单方法如何引导模型得出更好答案的示例。

**基线** 我们使用以下方法对预算强制进行基准测试：(I) 条件长度控制方法，这些方法依赖于在提示中告诉模型应生成多长时间。我们按粒度将它们分为 (a) 令牌条件控制：我们在提示中指定思考令牌的上限；(b) 步骤条件控制：我们指定思考步骤的上限，每步大约100个令牌；(c) 类别条件控制：我们编写两个通用提示，告诉模型思考时间应短或长（详见 §D.1）。(II) 拒绝采样，这会一直采样直到生成的内容符合预定的计算预算。这个先验捕获了基于其长度的响应后验。

How many r in raspberry? **Question**

Let's break down the process of counting the letter 'r' in the word "raspberry" ...

- \* First letter: 'r' - This is an 'r', count = 1.
- \* Second letter: 'a' - Not an 'r', count remains 1 ...
- \* Sixth letter: 'e' - Not an 'r', count remains 1.
- \* Seventh letter: 'r' - This is an 'r', count = 2.
- \* Eighth letter: 'y' - Not an 'r', count remains 2 ...

The number of 'r's in "raspberry" is 2.

**Wait**, let's re-read the question carefully. It asks "How many r in raspberry?" ... \* r - a - s - p - b - e - r - r - y ... \* First 'r' ... \* Second 'r' ... \* Third 'r' ... Count = 3 ... **Reasoning trace**

My initial answer of 2 was incorrect due to a quick reading of the word. **Final Answer:** The final answer is 3 **Response**

**Figure 3. Budget forcing with s1-32B.** The model tries to stop after "...is 2.", but we suppress the end-of-thinking token delimiter instead appending "Wait" leading s1-32B to self-correct its answer.

thinking token delimiter and optionally append the string "Wait" to the model's current reasoning trace to encourage the model to reflect on its current generation. Figure 3 contains an example of how this simple approach can lead the model to arrive at a better answer.

**Baselines** We benchmark budget forcing with: **(I) Conditional length-control methods**, which rely on telling the model in the prompt how long it should generate for. We group them by granularity into (a) Token-conditional control: We specify an upper bound of thinking tokens in the prompt; (b) Step-conditional control: We specify an upper bound of thinking steps, where each step is around 100 tokens; (c) Class-conditional control: We write two generic prompts that tell the model to either think for a short or long amount of time (see §D.1 for details). **(II) Rejection sampling**, which samples until a generation fits a predetermined compute budget. This oracle captures the posterior over responses conditioned on its length.

### 3.2. Metrics

We establish a set of desiderata as evaluation metrics to measure test-time scaling across methods. Importantly, we do not only care about the accuracy a method can

achieve but also its controllability and test-time scaling slope. For each method we consider, we run a set of evaluations  $a \in \mathcal{A}$  varying test-time compute on a fixed benchmark, e.g. AIME24. This produces a piece-wise linear function  $f$  with compute as the x-axis measured in thinking tokens and accuracy as the y-axis (see Figure 1, where the rightmost dot for AIME24 corresponds to  $f(7320) = 57\%$ ). We measure three metrics:

$$\text{Control} = \frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} \mathbb{I}(a_{\min} \leq a \leq a_{\max}) \quad (1)$$

where  $a_{\min}, a_{\max}$  refer to a pre-specified minimum and maximum amount of test-time compute; in our case thinking tokens. We usually only constrain  $a_{\max}$ . As tokens generated correspond to the amount of test-time compute spent, this metric measures the extent to which a method allows controllability over the use of that test-time compute. We report it as a percentage with 100% being perfect control.

$$\text{Scaling} = \frac{1}{\binom{|\mathcal{A}|}{2}} \sum_{\substack{a,b \in \mathcal{A} \\ b > a}} \frac{f(b) - f(a)}{b - a} \quad (2)$$

Scaling is the average slope of the piece-wise linear function. It must be positive for useful methods and larger is better.

$$\text{Performance} = \max_{a \in \mathcal{A}} f(a) \quad (3)$$

Performance is simply the maximum performance the method achieves on the benchmark. A method with monotonically increasing scaling achieves 100% performance on any benchmark in the limit. However, the methods we investigate eventually flatten out or further scaling fails due to control or context window limitations.

## 4. Results

### 4.1. Setup

**Training** We perform supervised finetuning on Qwen2.5-32B-Instruct using s1K to obtain our model

### 3.2. Metrics

我们建立了一套评估标准作为评价指标，用于衡量不同方法在测试时间扩展上的表现。重要的是，我们不仅关注方法可以达到的准确性，还关注其可控性和测试时间扩展的斜率。对于每种方法，我们在固定的基准测试（例如 AIME24）上运行一组评估  $a \in \mathcal{A}$ ，改变测试时间的计算量。这产生了一个分段线性函数  $f$ ，其中 x 轴表示计算量（以思考令牌为单位），y 轴表示准确性（参见 Figure 1，其中 AIME24 的最右侧点对应于  $f(7320) = 57\%$ ）。我们测量了三个指标：

$$\text{Control} = \frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} \mathbb{I}(a_{\min} \leq a \leq a_{\max}) \quad (1)$$

其中  $a_{\min}, a_{\max}$  指的是预设的测试时计算量的最小值和最大值；在我们的例子中是指思考 token。我们通常只限制  $a_{\max}$ 。由于生成的 token 数量对应于测试时所花费的计算量，这一指标衡量了一种方法在控制测试时计算量使用方面的程度。我们以百分比的形式报告这一指标，100% 表示完全控制。

$$\text{Scaling} = \frac{1}{\binom{|\mathcal{A}|}{2}} \sum_{\substack{a,b \in \mathcal{A} \\ b > a}} \frac{f(b) - f(a)}{b - a} \quad (2)$$

Scaling 是分段线性函数的平均斜率。对于有用的方法，它必须是正数，并且越大越好。

$$\text{Performance} = \max_{a \in \mathcal{A}} f(a) \quad (3)$$

性能仅仅是该方法在基准测试中达到的最大性能。一个具有单调递增扩展性的方法在任何基准测试中最终都能达到100%的性能。然而，我们研究的方法最终会趋于平稳，或者由于控制或上下文窗口的限制，进一步的扩展会失败。

## 4. Results

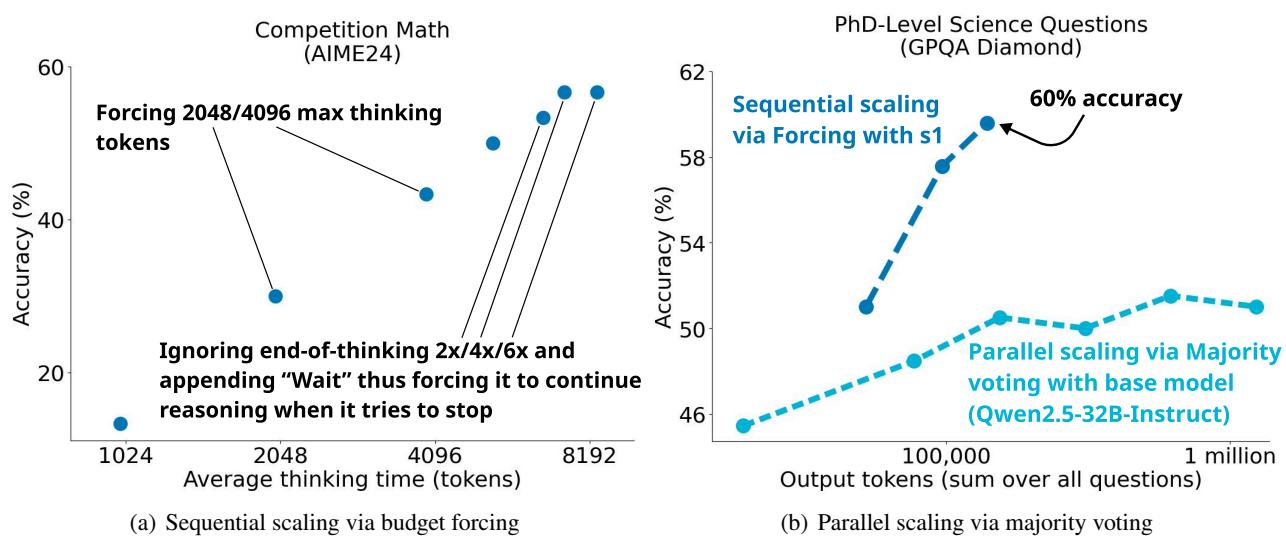
### 4.1. Setup

**训练** 我们在 Qwen2.5-32B-Instruct 上使用 s1K 进行监督微调，以获得我们的模型 s1-32B，基本超参

数如 §C 所述。微调在 16 个 NVIDIA H100 GPU 上使用 PyTorch FSDP 花费了 26 分钟。

**评估** 我们选择了三个在该领域广泛使用的代表性推理基准：AIME24 (of America, 2024) 包含了 2024 年 1 月 31 日至 2 月 1 日举行的 2024 年美国数学邀请赛 (AIME) 中的 30 道题目。AIME 测试包括算术、代数、计数、几何、数论、概率和其他中学数学主题的数学问题解决能力。测试中得分较高的高中生将被邀请参加美国数学奥林匹克竞赛 (USAMO)。所有 AIME 答案都是从 000 到 999 (包括两端) 的整数。一些 AIME 问题依赖于我们使用矢量图形语言 Asymptote 提供给模型的图形，因为它不能接受图像输入。MATH500 (Hendrycks et al., 2021) 是一个包含不同难度的竞赛数学问题的基准。我们评估了 OpenAI 在先前工作中 (Lightman et al., 2023) 选择的相同 500 个样本。GPQA Diamond (Rein et al., 2023) 包含了 198 个来自生物学、化学和物理学的博士水平科学问题。相应领域的博士专家在 GPQA Diamond 上仅达到了 69.7% 的准确率 (OpenAI, 2024)。在本工作的评估上下文中，当我们提到“GPQA”时，始终是指 Diamond 子集。我们基于“lm-evaluation-harness”框架 (Gao et al., 2021; Biderman et al., 2024) 进行构建。

**其他模型** 我们将 s1-32B 与以下模型进行基准测试：OpenAI o1 系列 (OpenAI, 2024)，这些是封闭源代码模型，普及了测试时扩展的概念；DeepSeek r1 系列 (Team, 2024a)，这些是开放权重推理模型，性能可达 o1 水平，与我们的模型同时发布；Qwen 的 QwQ-32B-preview (Team, 2024b)，一个 32B 的开放权重推理模型，未公开方法；Sky-T1-32B-Preview (Team, 2025) 和 Bespoke-32B (Labs, 2025)，这些都是开放模型，从 QwQ-32B-preview 和 r1 中提取的开放推理数据；Google Gemini 2.0 Flash Thinking Experimental (Google, 2024)，我们从中提取的 API。由于没有官方评估分数，我们使用 Gemini API 自行进行基准测试。然而，Gemini API 的“背诵错误”



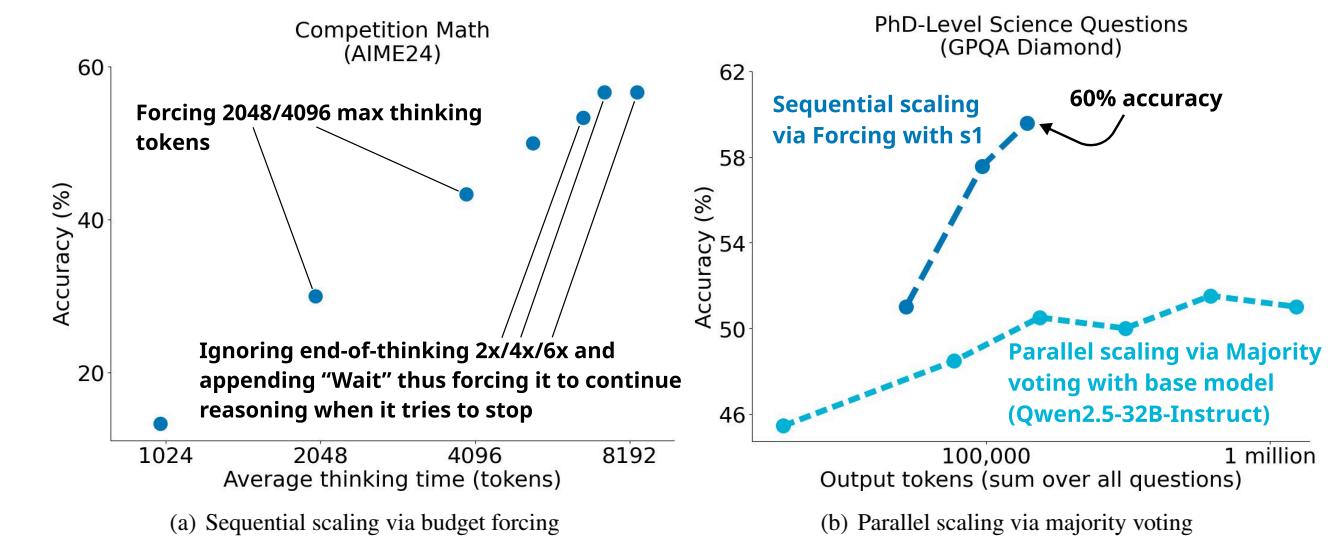
**Figure 4. Sequential and parallel test-time scaling.** (a): Budget forcing shows clear scaling trends and extrapolates to some extent. For the three rightmost dots, we prevent the model from stopping its thinking 2/4/6 times, each time appending “Wait” to its current reasoning trace. (b): For Qwen2.5-32B-Instruct we perform 64 evaluations for each sample with a temperature of 1 and visualize the performance when majority voting across 2, 4, 8, 16, 32, and 64 of these.

**s1-32B** using basic hyperparameters outlined in §C. Finetuning took 26 minutes on 16 NVIDIA H100 GPUs with PyTorch FSDP.

**Evaluation** We select three representative reasoning benchmarks widely used in the field: **AIME24** (of America, 2024) consists of 30 problems that were used in the 2024 American Invitational Mathematics Examination (AIME) held from Wednesday, January 31 – Thursday, February 1, 2024. AIME tests mathematical problem-solving with arithmetic, algebra, counting, geometry, number theory, probability, and other secondary school math topics. High-scoring high school students in the test are invited to participate in the United States of America Mathematics Olympiad (USAMO). All AIME answers are integers ranging from 000 to 999, inclusive. Some AIME problems rely on figures that we provide to our model using the vector graphics language Asymptote as it cannot take image inputs. **MATH500** (Hendrycks et al., 2021) is a benchmark of competition math problems of varying difficulty. We evaluate on the same 500 samples selected by OpenAI in prior work (Lightman et al.,

2023). **GPQA Diamond** (Rein et al., 2023) consists of 198 PhD-level science questions from Biology, Chemistry and Physics. Experts with PhDs in the corresponding domains only achieved 69.7% on GPQA Diamond (OpenAI, 2024). When we write “GPQA” in the context of evaluation in this work, we always refer to the Diamond subset. We build on the “lm-evaluation-harness” framework (Gao et al., 2021; Biderman et al., 2024).

**Other models** We benchmark **s1-32B** against: **OpenAI o1 series** (OpenAI, 2024), which are closed-source models that popularized the idea of test-time scaling; **DeepSeek r1 series** (Team, 2024a), which are open-weight reasoning models with up to o1-level performance, concurrently released to ours; Qwen’s **QwQ-32B-preview** (Team, 2024b), a 32B open-weight reasoning model without disclosed methodology; **Sky-T1-32B-Preview** (Team, 2025) and **Bespoke-32B** (Labs, 2025), which are open models with open reasoning data distilled from QwQ-32B-preview and r1; **Google Gemini 2.0 Flash Thinking Experimental** (Google, 2024), the API that we distill from. As it has no official evaluation



**Figure 4. 顺序和并行测试时扩展。** (a): 预算强制显示了明显的扩展趋势，并在一定程度上进行了外推。对于最右侧的三个点，我们分别阻止模型停止思考2/4/6次，每次在其当前推理轨迹后附加“Wait”。(b): 对于Qwen2.5-32B-Instruct，我们对每个样本执行64次评估，温度为1，并可视化在2、4、8、16、32和64次评估中的多数投票性能。

使得评估具有挑战性。<sup>2</sup>为了规避这一问题，我们在其网页界面中手动插入了所有30个AIME24问题，因为该错误不会在此处出现。然而，我们省略了MATH500（500个问题）和GPQA Diamond（198个问题），因此在Table 1中它们是N.A.。我们的模型**s1-32B**完全开放，包括权重、推理数据和代码。

#### 4.2. Performance

**测试时扩展** Figure 1 显示了**s1-32B**在预算强制下的性能随着更多测试时计算资源的变化。在Figure 4（左）中，我们扩展了Figure 1（中）的图表，表明虽然我们可以使用我们的预算强制技术（§3）和更多的测试时计算资源来提高AIME24的性能，但最终在六倍时趋于平稳。过于频繁地抑制思考结束标记分隔符可能会导致模型陷入重复循环而不是继续推理。在Figure 4（右）中，我们展示了在我们的1,000个样本上训练Qwen2.5-32B-Instruct以生成**s1-32B**并配备简单的预算强制技术后，它在不同的扩展范式中运行。通过多数投票在基础模型上扩展测试时计算资源无法赶上**s1-32B**的性能，这验证了我们在§3中的直觉，即顺序扩展比并行扩展更有

<sup>2</sup><https://github.com/google/generative-ai-docs/issues/257>

**Table 1.** **s1-32B**是一个开放且样本高效的推理模型。我们评估了**s1-32B**、Qwen 和 Gemini（某些条目未知（N.A.），见§4）。其他结果来自各自的报告（Qwen et al., 2024; Team, 2024b; OpenAI, 2024; DeepSeek-AI et al., 2025; Labs, 2025; Team, 2025）。# ex. = 用于推理微调的示例数量；BF = 预算强制。

Model	# ex.	AIME 2024	MATH 500	GPQA Diamond
<b>API only</b>				
o1-preview	N.A.	44.6	85.5	73.3
o1-mini	N.A.	70.0	90.0	60.0
o1	N.A.	<b>74.4</b>	<b>94.8</b>	<b>77.3</b>
Gemini 2.0	N.A.	60.0	N.A.	N.A.
Flash Think.				
<b>Open Weights</b>				
Qwen2.5-32B-Instruct	N.A.	26.7	84.0	49.0
QwQ-32B	N.A.	50.0	90.6	65.2
r1	>>800K	<b>79.8</b>	<b>97.3</b>	<b>71.5</b>
r1-distill	800K	72.6	94.3	62.1
<b>Open Weights and Open Data</b>				
Sky-T1	17K	43.3	82.4	56.8
Bespoke-32B	17K	<b>63.3</b>	<b>93.0</b>	58.1
s1 w/o BF	<b>1K</b>	50.0	92.6	56.6
<b>s1-32B</b>	<b>1K</b>	56.7	<b>93.0</b>	<b>59.6</b>

scores, we use the Gemini API to benchmark it ourselves. However, the “recitation error” of the Gemini API makes evaluation challenging.<sup>2</sup> We circumvent this, by manually inserting all 30 AIME24 questions in its web interface where the error does not appear. However, we leave out MATH500 (500 questions) and GPQA Diamond (198 questions), thus they are N.A. in Table 1. Our model, **s1-32B**, is fully open including weights, reasoning data, and code.

**Table 1.** **s1-32B** is an open and sample-efficient reasoning model. We evaluate **s1-32B**, Qwen, and Gemini (some entries are unknown (N.A.), see §4). Other results are from the respective reports (Qwen et al., 2024; Team, 2024b; OpenAI, 2024; DeepSeek-AI et al., 2025; Labs, 2025; Team, 2025). # ex. = number examples used for reasoning finetuning; BF = budget forcing.

Model	# ex.	AIME 2024	MATH 500	GPQA Diamond
<b>API only</b>				
o1-preview	N.A.	44.6	85.5	73.3
o1-mini	N.A.	70.0	90.0	60.0
o1	N.A.	<b>74.4</b>	<b>94.8</b>	<b>77.3</b>
Gemini 2.0	N.A.	60.0	N.A.	N.A.
Flash Think.				
<b>Open Weights</b>				
Qwen2.5-32B-Instruct	N.A.	26.7	84.0	49.0
QwQ-32B	N.A.	50.0	90.6	65.2
r1	≥800K	<b>79.8</b>	<b>97.3</b>	<b>71.5</b>
r1-distill	800K	72.6	94.3	62.1
<b>Open Weights and Open Data</b>				
Sky-T1	17K	43.3	82.4	56.8
Bespoke-32B	17K	<b>63.3</b>	<b>93.0</b>	58.1
s1 w/o BF	<b>1K</b>	50.0	92.6	56.6
<b>s1-32B</b>	<b>1K</b>	56.7	<b>93.0</b>	<b>59.6</b>

## 4.2. Performance

**Test-time scaling** Figure 1 shows the performance of **s1-32B** with budget forcing scales with more test-time compute. In Figure 4 (left), we expand the plot from

<sup>2</sup><https://github.com/google/generative-ai-docs/issues/257>

Figure 1 (middle) showing that while we can improve AIME24 performance using our budget forcing technique (§3) and more test-time compute it does eventually flatten out at six times. Suppressing the end-of-thinking token delimiter too often can lead the model into repetitive loops instead of continued reasoning. In Figure 4 (right), we show that after training Qwen2.5-32B-Instruct on our 1,000 samples to produce **s1-32B** and equipping it with the simple budget forcing technique, it operates in a different scaling paradigm. Scaling test-time compute on the base model via majority voting cannot catch up with the performance of **s1-32B** which validates our intuition from §3 that sequential scaling is more effective than parallel. We provide example generations of **s1-32B** in Figure 5.

**Sample-efficiency** In Figure 2 (right) and Table 1 we compare **s1-32B** with other models. We find that **s1-32B** is the most sample-efficient open data reasoning model. It performs significantly better than our base model (Qwen2.5-32B-Instruct) despite just training it on an additional 1,000 samples. The concurrently released r1-32B shows stronger performance than **s1-32B** while also only using SFT (DeepSeek-AI et al., 2025). However, it is trained on  $800 \times$  more reasoning samples. It is an open question whether one can achieve their performance with just 1,000 samples. Finally, our model nearly matches Gemini 2.0 Thinking on AIME24. As **s1-32B** is distilled from Gemini 2.0, this shows our distillation procedure was likely effective.

## 5. Ablations

### 5.1. Data Quantity, Diversity, and Difficulty

In §2 we outlined our three guiding principles in curating **s1K**: Quality, Difficulty, and Diversity. Here we test the importance of combining them and the overall efficacy of our selection. **Only Quality (1K-random)**: After obtaining our high-quality reasoning chains from Gemini, we select 1,000 samples at random; not relying on our

效。我们在 Figure 5 中提供了 **s1-32B** 的生成示例。

**样本效率** 在 Figure 2 (右) 和 Table 1 中, 我们将 **s1-32B** 与其他模型进行了比较。我们发现 **s1-32B** 是最高效的开放数据推理模型。尽管仅在额外的 1,000 个样本上进行了训练, 它的性能显著优于我们的基础模型 (Qwen2.5-32B-Instruct)。同时发布的 r1-32B 表现优于 **s1-32B**, 尽管它也仅使用了 SFT (DeepSeek-AI et al., 2025)。然而, 它是在 800 倍更多的推理样本上训练的。是否可以用仅 1,000 个样本达到其性能仍然是一个开放的问题。最后, 我们的模型在 AIME24 上几乎与 Gemini 2.0 Thinking 匹配。由于 **s1-32B** 是从 Gemini 2.0 蒸馏而来的, 这表明我们的蒸馏过程可能非常有效。

## 5. Ablations

### 5.1. Data Quantity, Diversity, and Difficulty

**Table 2. s1K 数据消融。** 我们为表中的所有分数设置了大约 30,000 个思考令牌的预算力 (BF)。这比没有 BF 的分数 (见表 Table 1) 略好, 因为它允许模型在陷入无限循环时以最佳猜测结束。我们使用 10,000 个引导样本报告相对于 **s1K** 模型的差异的 95% 配对引导置信区间。例如, 区间 [-13%, 20%] 表示, 以 95% 的置信度, 59K-full 和 **s1K** 之间的真正差异在 -13% 到 +20% 之间。如果整个区间为负, 例如 [-27%, -3%], 我们可以自信地说性能比 **s1K** 差。

Method	AIME 2024	MATH 500	GPQA Diamond	A
<b>BF</b>	<b>100%</b>	15	<b>56.7</b>	5
TCC	40%	-24	40.0	5
TCC + BF	<b>100%</b>	13	40.0	5
SCC	60%	3	36.7	5
SCC + BF	<b>100%</b>	6	36.7	5
CCC	50%	<b>25</b>	36.7	2
RS	<b>100%</b>	-35	40.0	5
<b>s1K</b>	50.0	93.0	57.6	

在 §2 中, 我们概述了在策划 **s1K** 时遵循的三个原则: 质量、难度和多样性。在这里, 我们测试了将它们结合起来的重要性以及我们选择的整体有效性。**仅质量 (1K-随机)**: 在从 Gemini 获得高质量的推理链后, 我们随机选择了 1,000 个样本; 完全不依

赖于我们的难度和多样性筛选。Table 2 显示, 这种方法在所有基准测试中的表现远不如 **s1K**。仅多样性 (**1K-多样**): 对于这个数据集, 我们在各个领域均匀采样以最大化多样性, 而不考虑任何难度概念。这种方法也导致了与 **1K-随机** 相似的较差表现。仅难度 (**1K-最长**): 在这里, 我们依赖于在 §2 中介绍的一个难度指标, 选择 1,000 个推理轨迹最长的样本。这种方法显著提高了 GPQA 的表现, 但总体上仍不如使用 **s1K**。最大化数量: 最后, 我们将仅使用我们所有的 59K 样本进行训练进行比较, 这是所有 **1K** 样本版本的超集。这导致了一个强大的模型, 但使用了更多的资源。为了微调 59K 样本, 我们使用了 394 个 H100 GPU 小时, 而 **s1-32B** 仅需要 7 个 H100 GPU 小时。此外, 仅依赖 **s1K** 非常具有竞争力, 如 §2 所示。总体而言, 通过我们在 §2 中提出的方法, 结合所有三个标准——质量、难度、多样性——对于样本高效的推理训练至关重要。

### 5.2. Test-time scaling methods

**Table 3. 在 AIME24 上对测试时计算扩展方法的消融研究。**  $|\mathcal{A}|$  表示用于估计属性的评估运行次数; 因此更高的值表示更稳健。加粗表示我们选择的方法和最佳值。BF = 预算强制, TCC/SCC/CCC = 令牌/步骤/类别条件控制, RS = 拒绝采样。

Method	Control	Scaling	Performance	$ \mathcal{A} $
<b>BF</b>	<b>100%</b>	15	<b>56.7</b>	5
TCC	40%	-24	40.0	5
TCC + BF	<b>100%</b>	13	40.0	5
SCC	60%	3	36.7	5
SCC + BF	<b>100%</b>	6	36.7	5
CCC	50%	<b>25</b>	36.7	2
RS	<b>100%</b>	-35	40.0	5
<b>s1K</b>	50.0	93.0	57.6	

**预算强制** 在 Table 3 中, 我们比较了在 §3 中介绍的测试时缩放方法。总体而言, 我们发现 预算强制 提供了完美的控制, 良好的缩放, 并且导致了我们最好的 AIME24 分数。因此, 这是我们在 Figure 1 和 §4 中使用的方法。在 Table 4 中, 我们比较了不同字符串的性能外推。我们发现“Wait”通常能提供最佳性能。

[AIME24]

[MATH500]

[GPQA]

Alice and Bob play the following game. A stack of  $n$  tokens lies before them. The players take turns with Alice going first. On each turn, the player removes either 1 token or 4 tokens from the stack. Whoever removes the last token wins. Find the number of positive integers  $n$  less than or equal to 2024 for which there exists a strategy for Bob that guarantees that Bob will win the game regardless of Alice's play.

The problem asks for the number of positive integers  $n \leq 2024$  such that Bob has a winning strategy. This is a typical impartial game that can be analyzed using the Sprague-Grundy theorem or by directly finding the winning and losing positions. Let  $W$  be the set of winning positions (number of tokens  $n$  such that the player whose turn it is can force a win) and  $L$  be the set of losing positions. [...]

Let  $P(n)$  be the statement "the player whose turn it is when there are  $n$  tokens has a winning strategy".  $P(0)$  is false (the player whose turn it is and faces 0 tokens has already lost). Consider the small values of  $n$ :  
 $n = 1$ : The player whose turn it is can take 1 token and win. So  $P(1)$  is true.  $1 \in W$ .

$n = 2$ : The player whose turn it is can take 1 token, leaving 1 token for the opponent. Since  $1 \in W$ , the opponent wins. So  $P(2)$  is false.  $2 \in L$ .

$n = 3$ : The player whose turn it is can take 1 token, leaving 2 tokens for the opponent. Since  $2 \in L$ , the player whose turn it is wins. So  $P(3)$  is true.  $3 \in W$ . [...]

We are looking for the number of positive integers  $n \leq 2024$  for which Bob has a winning strategy. This means that if the game starts with  $n$  tokens, and Alice goes first, Bob can win. This happens if and only if  $n$  is a losing position (because if  $n \in L$ , then Alice's first move must lead to a winning position for Bob).

[...] Let's try to find a pattern in the losing positions  $L = \{0, 2, 5, 7, 10, 12, 15, 17, 20, 22, \dots\}$ . Differences between consecutive terms: 2, 3, 2, 3, 2, 3, 2, ... The pattern of differences is alternating 2, 3. [...]

The losing positions are of the form  $5m$  or  $5m + 2$ , where  $m \geq 0$ . We need to find the number of positive integers  $n \leq 2024$  that are in the set of losing positions  $L$ . [...]

Case 1:  $n = 5m$ , where  $m \geq 1$ .  $1 \leq 5m \leq 2024$   $1/5 \leq m \leq 2024/5 = 404.8$  Since  $m$  is an integer,  $1 \leq m \leq 404$ . The number of values of  $m$  is 404.

Case 2:  $n = 5m + 2$ , where  $m \geq 0$ .  $5m + 2 \leq 2024$   $5m \leq 2022$   $m \leq 2022/5 = 404.4$  Since  $m$  is an integer,  $0 \leq m \leq 404$ . The number of values of  $m$  is  $404 - 0 + 1 = 405$ .

The total number of positive integers  $n \leq 2024$  that are losing positions is  $404 + 405 = \boxed{809}$ .

Find the projection of  $a$  onto  $b = \begin{pmatrix} 2 \\ 6 \\ 3 \end{pmatrix}$  if  $a \cdot b = 8$ .

The projection of vector  $a$  onto vector  $b$  is given by the formula:

$$\text{proj}_b a = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{b}\|^2} \mathbf{b}$$

We are given  $\mathbf{b} = \begin{pmatrix} 2 \\ 6 \\ 3 \end{pmatrix}$  and  $\mathbf{a} \cdot \mathbf{b} = 8$ .

First, we need to find the magnitude squared of  $\mathbf{b}$ , which is  $\|\mathbf{b}\|^2$ :

$$\|\mathbf{b}\|^2 = \mathbf{b} \cdot \mathbf{b} = 2^2 + 6^2 + 3^2 = 4 + 36 + 9 = 49.$$

Now we can plug the values into the projection formula:

$$\text{proj}_b a = \frac{8}{49} \mathbf{b}$$

Substitute the vector  $\mathbf{b}$ :

$$\text{proj}_b a = \frac{8}{49} \begin{pmatrix} 2 \\ 6 \\ 3 \end{pmatrix}$$

Multiply the scalar by each component of the vector:

$$\text{proj}_b a = \begin{pmatrix} \frac{8}{49} \times 2 \\ \frac{8}{49} \times 6 \\ \frac{8}{49} \times 3 \end{pmatrix} = \begin{pmatrix} \frac{16}{49} \\ \frac{48}{49} \\ \frac{24}{49} \end{pmatrix}$$

Final Answer: The final answer is

$$\begin{pmatrix} \frac{16}{49} \\ \frac{48}{49} \\ \frac{24}{49} \end{pmatrix}$$

A quantum mechanical particle of mass  $m$  moves in two dimensions in the following potential, as a function of the polar coordinates  $(r, \theta)$ :

$$V(r, \theta) = \frac{1}{2}kr^2 + \frac{3}{2}kr^2 \cos^2(\theta)$$

Find the energy spectrum. Hint: Write the potential in Cartesian coordinates.

The problem asks for the energy spectrum of a particle in a given 2D potential. The potential is given in polar coordinates  $(r, \theta)$ :

$$V(r, \theta) = \frac{1}{2}kr^2 + \frac{3}{2}kr^2 \cos^2(\theta).$$

We can rewrite the potential as:

$$V(r, \theta) = \frac{1}{2}kr^2(1 + 3 \cos^2(\theta)).$$

The Hamiltonian for a particle in 2D is given by:

$$H = -\frac{\hbar^2}{2m} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) + V(x, y).$$

In polar coordinates  $(r, \theta)$ , the Laplacian operator is:

$$\nabla^2 = \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2}.$$

So the Hamiltonian in polar coordinates is:

$$H = -\frac{\hbar^2}{2m} \left[ \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} \right] + V(r, \theta).$$

The time-independent Schrödinger equation is  $H\psi = E\psi$ . We can try to separate variables by assuming a solution of the form  $\psi(r, \theta) = R(r)\Theta(\theta)$ . [...]

This confirms the energy spectrum is:

$$E = \hbar \sqrt{\frac{k}{m}} \left( 2n_x + n_y + \frac{3}{2} \right).$$

Final Answer:  $E = \hbar \sqrt{\frac{k}{m}} \left( 2n_x + n_y + \frac{3}{2} \right)$ .

[AIME24]

[MATH500]

[GPQA]

Alice and Bob play the following game. A stack of  $n$  tokens lies before them. The players take turns with Alice going first. On each turn, the player removes either 1 token or 4 tokens from the stack. Whoever removes the last token wins. Find the number of positive integers  $n$  less than or equal to 2024 for which there exists a strategy for Bob that guarantees that Bob will win the game regardless of Alice's play.

The problem asks for the number of positive integers  $n \leq 2024$  such that Bob has a winning strategy. This is a typical impartial game that can be analyzed using the Sprague-Grundy theorem or by directly finding the winning and losing positions. Let  $W$  be the set of winning positions (number of tokens  $n$  such that the player whose turn it is can force a win) and  $L$  be the set of losing positions. [...]

Let  $P(n)$  be the statement "the player whose turn it is when there are  $n$  tokens has a winning strategy".  $P(0)$  is false (the player whose turn it is and faces 0 tokens has already lost). Consider the small values of  $n$ :  
 $n = 1$ : The player whose turn it is can take 1 token and win. So  $P(1)$  is true.  $1 \in W$ .

$n = 2$ : The player whose turn it is can take 1 token, leaving 1 token for the opponent. Since  $1 \in W$ , the opponent wins. So  $P(2)$  is false.  $2 \in L$ .

$n = 3$ : The player whose turn it is can take 1 token, leaving 2 tokens for the opponent. Since  $2 \in L$ , the player whose turn it is wins. So  $P(3)$  is true.  $3 \in W$ . [...]

We are looking for the number of positive integers  $n \leq 2024$  for which Bob has a winning strategy. This means that if the game starts with  $n$  tokens, and Alice goes first, Bob can win. This happens if and only if  $n$  is a losing position (because if  $n \in L$ , then Alice's first move must lead to a winning position for Bob).

[...] Let's try to find a pattern in the losing positions  $L = \{0, 2, 5, 7, 10, 12, 15, 17, 20, 22, \dots\}$ . Differences between consecutive terms: 2, 3, 2, 3, 2, 3, 2, ... The pattern of differences is alternating 2, 3. [...]

The losing positions are of the form  $5m$  or  $5m + 2$ , where  $m \geq 0$ . We need to find the number of positive integers  $n \leq 2024$  that are in the set of losing positions  $L$ . [...]

Case 1:  $n = 5m$ , where  $m \geq 1$ .  $1 \leq 5m \leq 2024$   $1/5 \leq m \leq 2024/5 = 404.8$  Since  $m$  is an integer,  $1 \leq m \leq 404$ . The number of values of  $m$  is 404.

Case 2:  $n = 5m + 2$ , where  $m \geq 0$ .  $5m + 2 \leq 2024$   $5m \leq 2022$   $m \leq 2022/5 = 404.4$  Since  $m$  is an integer,  $0 \leq m \leq 404$ . The number of values of  $m$  is  $404 - 0 + 1 = 405$ .

The total number of positive integers  $n \leq 2024$  that are losing positions is  $404 + 405 = \boxed{809}$ .

Find the projection of  $a$  onto  $b = \begin{pmatrix} 2 \\ 6 \\ 3 \end{pmatrix}$  if  $a \cdot b = 8$ .

The projection of vector  $a$  onto vector  $b$  is given by the formula:

$$\text{proj}_b a = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{b}\|^2} \mathbf{b}$$

We are given  $\mathbf{b} = \begin{pmatrix} 2 \\ 6 \\ 3 \end{pmatrix}$  and  $\mathbf{a} \cdot \mathbf{b} = 8$ .

First, we need to find the magnitude squared of  $\mathbf{b}$ , which is  $\|\mathbf{b}\|^2$ :

$$\|\mathbf{b}\|^2 = \mathbf{b} \cdot \mathbf{b} = 2^2 + 6^2 + 3^2 = 4 + 36 + 9 = 49.$$

Now we can plug the values into the projection formula:

$$\text{proj}_b a = \frac{8}{49} \mathbf{b}$$

Substitute the vector  $\mathbf{b}$ :

$$\text{proj}_b a = \frac{8}{49} \begin{pmatrix} 2 \\ 6 \\ 3 \end{pmatrix}$$

Multiply the scalar by each component of the vector:

$$\text{proj}_b a = \begin{pmatrix} \frac{8}{49} \times 2 \\ \frac{8}{49} \times 6 \\ \frac{8}{49} \times 3 \end{pmatrix} = \begin{pmatrix} \frac{16}{49} \\ \frac{48}{49} \\ \frac{24}{49} \end{pmatrix}$$

Final Answer: The final answer is

$$\begin{pmatrix} \frac{16}{49} \\ \frac{48}{49} \\ \frac{24}{49} \end{pmatrix}$$

Final Answer:  $E = \hbar \sqrt{\frac{k}{m}} \left( 2n_x + n_y + \frac{3}{2} \right)$ .

A quantum mechanical particle of mass  $m$  moves in two dimensions in the following potential, as a function of the polar coordinates  $(r, \theta)$ :

$$V(r, \theta) = \frac{1}{2}kr^2 + \frac{3}{2}kr^2 \cos^2(\theta)$$

Find the energy spectrum. Hint: Write the potential in Cartesian coordinates.

The problem asks for the energy spectrum of a particle in a given 2D potential. The potential is given in polar coordinates  $(r, \theta)$ :

$$V(r, \theta) = \frac{1}{2}kr^2(1 + 3 \cos^2(\theta)).$$

We can rewrite the potential as:

$$V(r, \theta) = \frac{1}{2}kr^2(1 + 3 \cos^2(\theta)).$$

The Hamiltonian for a particle in 2D is given by:

$$H = -\frac{\hbar^2}{2m} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) + V(x, y).$$

In polar coordinates  $(r, \theta)$ , the Laplacian operator is:

$$\nabla^2 = \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2}.$$

So the Hamiltonian in polar coordinates is:

$$H = -\frac{\hbar^2}{2m} \left[ \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} \right] + V(r, \theta).$$

The time-independent Schrödinger equation is  $H\psi = E\psi$ . We can try to separate variables by assuming a solution of the form  $\psi(r, \theta) = R(r)\Theta(\theta)$ . [...]

This confirms the energy spectrum is:

$$E = \hbar \sqrt{\frac{k}{m}} \left( 2n_x + n_y + \frac{3}{2} \right).$$

Final Answer:  $E = \hbar \sqrt{\frac{k}{m}} \left( 2n_x + n_y + \frac{3}{2} \right)$ .

Figure 5. Example model outputs. We pick a question from AIME24 (left), MATH500 (middle), and GPQA (right), where our model generates the correct answer. The black text is the prompt, the light blue text is the reasoning trace, and the blue text is the answer of s1-32B. The gray ellipsis [...] indicates that the text was trimmed to fit this page, but the generated text is actually longer.

Figure 5. 例示模型输出。我们从 AIME24 (左)、MATH500 (中) 和 GPQA (右) 中挑选了一个问题，其中我们的模型生成了正确答案。黑色文本是提示，浅蓝色文本是推理过程，蓝色文本是 s1-32B 的答案。灰色省略号 [...] 表示文本被裁剪以适应此页面，但生成的文本实际上更长。

**Table 2. s1K data ablations.** We budget force (BF) a maximum of around 30,000 thinking tokens for all scores in this table. This performs slightly better than the scores without BF (Table 1) as it allows the model to finish with a best guess when stuck in an infinite loop. We report 95% paired bootstrap confidence intervals for differences relative to the **s1K** model using 10,000 bootstrap samples. E.g., the interval [-13%, 20%] means that, with 95% confidence, the true difference between 59K-full and **s1K** is between -13% and +20%. If the entire interval is negative, e.g. [-27%, -3%], we can confidently say that the performance is worse than **s1K**.

Model	AIME 2024	MATH 500	GPQA Diamond	
<b>BF</b>	<b>100%</b>	15	<b>56.7</b>	5
1K-random	36.7	90.6	52.0	
	[-26.7%, -3.3%]	[-4.8%, 0.0%]	[-12.6%, 2.5%]	
1K-diverse	26.7	91.2	54.6	
	[-40.0%, -10.0%]	[-4.0%, 0.2%]	[-10.1%, 5.1%]	
1K-longest	33.3	90.4	59.6	
	[-36.7%, 0.0%]	[-5.0%, -0.2%]	[-5.1%, 10.1%]	
59K-full	53.3	92.8	58.1	
	[-13.3%, 20.0%]	[-2.6%, 2.2%]	[-6.6%, 8.6%]	
<b>s1K</b>	50.0	93.0	57.6	

difficulty and diversity filtering at all. Table 2 shows this approach performs much worse than **s1K** across all benchmarks. **Only Diversity (1K-diverse)**: For this dataset, we sample uniformly across domains to maximize diversity disregarding any notion of difficulty. This approach also leads to poor performance similar to 1K-random. **Only Difficulty (1K-longest)**: Here we rely on one of our difficulty indicators introduced in §2 by selecting the 1,000 samples with the longest reasoning traces. This approach significantly boosts GPQA performance but overall still falls short of using **s1K**. **Maximize Quantity**: Finally, we compare with just training on all of our 59K samples, a superset of all the 1K-sample versions. This leads to a strong model but uses much more resources. To finetune on 59K samples, we use 394 H100 GPU hours while **s1-32B** only required 7 H100 GPU hours. Moreover, relying only on **s1K** is extremely competitive as shown in §2. Overall, combining all three criteria – *Quality*, *Difficulty*, *Diversity* – via our methodology in §2 is key for sample-efficient reasoning training.

## 5.2. Test-time scaling methods

**Table 3. Ablations on methods to scale test-time compute on AIME24.**  $|\mathcal{A}|$  refers to the number of evaluation runs used to estimate the properties; thus a higher value indicates more robustness. **Bold** indicates our chosen method and the best values. BF = budget forcing, TCC/SCC/CCC = token/step/class-conditional control, RS = rejection sampling.

Method	Control	Scaling	Performance	$ \mathcal{A} $
<b>BF</b>	<b>100%</b>	15	<b>56.7</b>	5
TCC	40%	-24	40.0	5
TCC + BF	<b>100%</b>	13	40.0	5
SCC	60%	3	36.7	5
SCC + BF	<b>100%</b>	6	36.7	5
CCC	50%	<b>25</b>	36.7	2
RS	<b>100%</b>	-35	40.0	5

**Budget forcing** In Table 3 we compare the test-time scaling methods we have introduced in §3. Overall, we find that *budget forcing* provides perfect control, good scaling, and leads to our best AIME24 score. Thus, this is the method we use for **s1-32B** in Figure 1 and in §4. In Table 4, we compare different strings for extrapolating performance. We find that “Wait” generally gives the best performance.

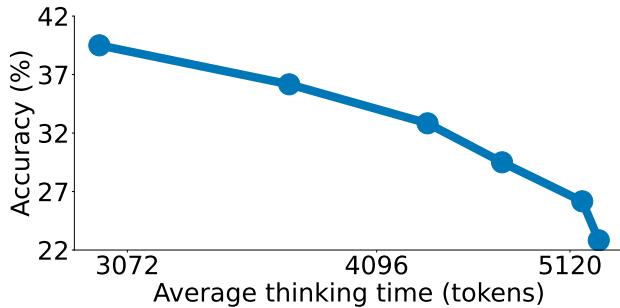
**Class-conditional control** We provide benchmark scores for this method in §D.1 and summarize three findings here: (1) Token-conditional control fails without budget forcing, as our model cannot reliably count tokens - even when trained to do so. (2) Under step-conditional control, the model generates a similar total number of tokens when given different step targets, as the model goes from few steps with many tokens per step, to many steps with few tokens in each step. Thus, the model learns to hack its way around the compute constraint making the controllability of this method mediocre. (3) Class-conditional control can work - telling a model to simply think longer can increase its test-time compute and performance, which leads good scaling in Table 3.

**类别条件控制** 我们在 §D.1 中提供了此方法的基准分数，并在此总结了三个发现：(1) 令牌条件控制在没有预算强制的情况下会失败，因为我们的模型无法可靠地计数令牌——即使经过训练也是如此。 (2) 在步骤条件控制下，当给定不同的步骤目标时，模型生成的总令牌数相似，因为模型从少量步骤每步多个令牌变为多步骤每步少量令牌。因此，模型学会了绕过计算约束的方法，使得这种方法的可控性较差。 (3) 类别条件控制可以工作——告诉模型简单地思考更长时间可以增加其测试时的计算量和性能，这在 Table 3 中导致了良好的缩放。

**Table 4. 预算强制外推消融。** 我们比较了两次忽略思维结束分隔符以及附加无字符串或各种字符串的情况。

Model	AIME 2024	MATH 500	GPQA Diamond
No extrapolation	50.0	<b>93.0</b>	57.6
2x without string	50.0	90.2	55.1
2x “Alternatively”	50.0	92.2	<b>59.6</b>
2x “Hmm”	50.0	<b>93.0</b>	<b>59.6</b>
2x “Wait”	<b>53.3</b>	<b>93.0</b>	<b>59.6</b>

**拒绝采样** 令人惊讶的是，我们发现简单地采样直到生成内容符合特定长度会导致如图Figure 6所示的逆向缩放趋势。在 §D.2 中，我们检查了一个问题，该问题在拒绝采样为  $\leq 4000$  时被模型正确回答，但在  $\leq 8000$  token 设置下却没有。在  $\leq 4000$  设置下，模型直接跳到了正确的方法，而在  $\leq 8000$  设置下，模



**Figure 6. 在AIME24上使用s1-32B进行拒绝采样。** 我们以1的温度进行采样，直到所有生成的思考令牌数量少于（从左到右）3500、4000、5000、8000和16000，每个样本平均需要655、97、8、3、2和1次尝试。

型却回溯了很多。我们假设存在一种相关性，即较短的生成内容往往是模型从一开始就正确的，而较长的生成内容往往是模型犯错并因此回溯或自我质疑的情况。这导致在拒绝采样时，较长的样本往往不正确，从而导致逆向缩放趋势。

## 6. Discussion and related work

### 6.1. Sample-efficient reasoning

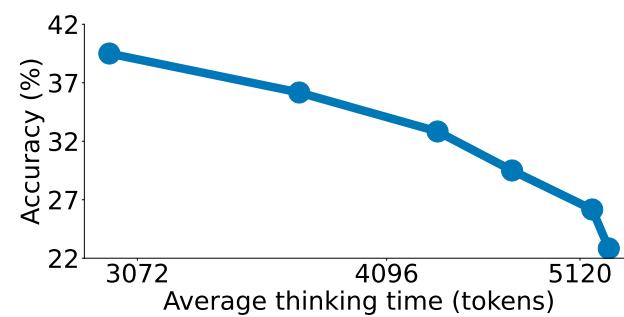
**模型** 有许多并行的努力旨在构建能够复制 o1 (OpenAI, 2024) 性能的模型。例如，DeepSeek-r1 和 k1.5 (DeepSeek-AI et al., 2025; Team et al., 2025) 是使用强化学习方法构建的，而其他模型则依赖于使用数万个蒸馏示例的 SFT (Team, 2025; Xu et al., 2025; Labs, 2025)。我们展示了仅使用 1,000 个示例进行 SFT 就足以构建一个与 o1-preview 匹配的具有竞争力的推理模型，并且该模型位于帕累托前沿 (Figure 2)。此外，我们引入了预算强制，结合我们的推理模型，首次重现了 OpenAI 的测试时扩展曲线 (OpenAI, 2024)。为什么仅在 1,000 个样本上进行监督微调就能带来如此大的性能提升？我们假设模型在预训练期间已经接触到了大量的推理数据，这些数据涵盖了数万亿个标记。因此，推理能力已经存在于我们的模型中。我们样本高效的微调阶段只是激活了这种能力，并且我们在测试时通过预算强制进一步扩展了它。这类似于 LIMA (Zhou et al., 2023) 中提出的“表面一致性假设”，作者发现 1,000 个示例足以使模型符合用户偏好。

**基准和方法** 为了评估和推动这些模型的极限，已经引入了越来越具有挑战性的基准，例如奥林匹克级别的科学竞赛 (He et al., 2024; Jain et al., 2024; Zhong et al., 2023) 和其他基准 (Srivastava et al., 2023; Glazer et al., 2024; Su et al., 2024; Kim et al., 2024; Phan et al., 2025)。为了提高模型在推理相关任务上的性能，研究人员探索了多种策略：先前的工作已经探索了在与数学和科学相关的专业语料库上继续训练语言模型 (Azerbaiyev et al., 2023; Yang et al., 2024)，有时甚至是合成生成的数据 (Yu et al., 2024)。其他人则开发了专门针对推理性能的训练方法 (Zelikman

**Table 4. Budget forcing extrapolation ablations.** We compare ignoring the end-of-thinking delimiter twice and appending none or various strings.

Model	AIME 2024	MATH 500	GPQA Diamond
No extrapolation	50.0	<b>93.0</b>	57.6
2x without string	50.0	90.2	55.1
2x “Alternatively”	50.0	92.2	<b>59.6</b>
2x “Hmm”	50.0	<b>93.0</b>	<b>59.6</b>
2x “Wait”	<b>53.3</b>	<b>93.0</b>	<b>59.6</b>

**Rejection sampling** Surprisingly, we find that simply sampling until the generation fits a specific length leads to an inverse scaling trend as depicted in Figure 6. In §D.2 we inspect a question, which was answered correctly by the model when rejection sampling for  $\leq 4000$ , but not for the  $\leq 8000$  token setting. In the  $\leq 4000$  setting the model directly jumps to the correct approach, while for the  $\leq 8000$  setting it backtracks a lot. We hypothesize that there is a correlation such that shorter generations tend to be the ones where the model was on the right track from the start, whereas longer ones tend to be ones where the model made mistakes and thus backtracks or questions itself. This leads to longer samples often being wrong when rejection sampling and thus the inverse scaling trend.



**Figure 6. Rejection sampling on AIME24 with s1-32B.** We sample with a temperature of 1 until all generations have less than (from left to right) 3500, 4000, 5000, 8000, and 16000 thinking tokens requiring an average of 655, 97, 8, 3, 2, and 1 tries per sample.

## 6. Discussion and related work

### 6.1. Sample-efficient reasoning

**Models** There are a number of concurrent efforts to build models that replicate the performance of o1 (OpenAI, 2024). For example, DeepSeek-r1 and k1.5 (DeepSeek-AI et al., 2025; Team et al., 2025) are built with reinforcement learning methods, while others rely on SFT using tens of thousands of distilled examples (Team, 2025; Xu et al., 2025; Labs, 2025). We show that SFT on only 1,000 examples suffices to build a competitive reasoning model matching o1-preview and produces a model that lies on the pareto frontier (Figure 2). Further, we introduce budget forcing which combined with our reasoning model leads to the first reproduction of OpenAI’s test-time scaling curves (OpenAI, 2024). Why does supervised finetuning on just 1,000 samples lead to such performance gains? We hypothesize that the model is already exposed to large amounts of reasoning data during pretraining which spans trillions of tokens. Thus, the ability to perform reasoning is already present in our model. Our sample-efficient finetuning stage just activates it and we scale it further at test time with budget forcing. This is similar to the "Superficial Alignment Hypothesis" presented in LIMA (Zhou et al., 2023), where the authors find that 1,000 examples can be sufficient to align a model to adhere to user preferences.

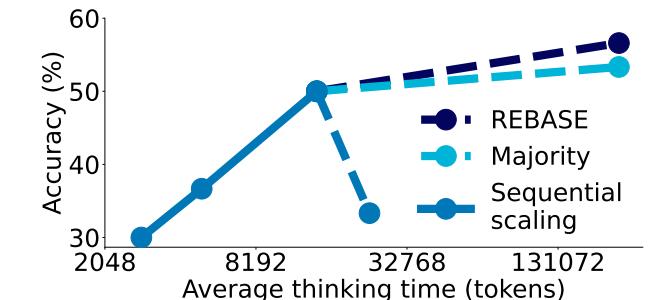
**Benchmarks and methods** To evaluate and push the limits of these models, increasingly challenging benchmarks have been introduced, such as Olympiad-level science competitions (He et al., 2024; Jain et al., 2024; Zhong et al., 2023) and others (Srivastava et al., 2023; Glazer et al., 2024; Su et al., 2024; Kim et al., 2024; Phan et al., 2025). To enhance models’ performance on reasoning-related tasks, researchers have pursued several strategies: Prior works have explored continuing training language models on specialized corpora related to mathematics and science (Azerbayev et al., 2023; Yang et al., 2024), sometimes even synthetically generated

et al., 2022; 2024; Luo et al., 2025; Yuan et al., 2025; Wu et al., 2024a)。另一条重要的研究路线集中在基于提示的方法，以激发和提高推理能力，包括链式思维提示方法 (Wei et al., 2023; Yao et al., 2023a;b; Bi et al., 2023; Fu et al., 2023; Zhang et al., 2024b; Xiang et al., 2025; Hu et al., 2024)。这些综合努力旨在提高语言模型的推理能力，使它们能够更有效地处理更复杂和抽象的任务。

### 6.2. Test-time scaling

**方法** 如我们在§3中介绍的，我们区分了两种方法来扩展测试时计算：并行和顺序。前者依赖于并生成的多个解决方案尝试，并通过特定标准选择最佳结果。这些标准包括选择最频繁的响应进行多数投票，或基于外部奖励选择最佳响应的Best-of-N (Brown et al., 2024; Irvine et al., 2023; Snell et al., 2024)。与重复采样不同，以前的顺序扩展方法让模型根据先前的尝试顺序生成解决方案尝试，使其能够根据先前的结果改进每次尝试(Snell et al., 2024; Hou et al., 2025; Lee et al., 2025)。基于树的搜索方法(Gandhi et al., 2024; Wu et al., 2024b)提供了一种介于顺序和并行扩展之间的混合方法，例如蒙特卡洛树搜索 (MCTS) (Liu et al., 2024; Zhang et al., 2023; Zhou et al., 2024; Choi et al., 2023)和引导波束搜索(Xie et al., 2023)。REBASE (Wu et al., 2024b)采用过程奖励模型在树搜索期间平衡利用和剪枝。实证研究表明，REBASE在性能上优于基于采样的方法和MCTS (Wu et al., 2024b)。奖励模型(Lightman et al., 2023; Wang et al., 2024b;c)在这些方法中起着关键作用。它们有两种变体：结果奖励模型和过程奖励模型。结果奖励模型(Xin et al., 2024; Ankner et al., 2024)为完整解决方案分配分数，特别适用于Best-of-N选择，而过程奖励模型(Lightman et al., 2023; Wang et al., 2024b; Wu et al., 2024b)评估各个推理步骤，在引导基于树的搜索方法中非常有效。

**进一步测试时扩展的限制** 我们在§4中展示了预算强制允许外推测试时计算，例如，将AIME24的性能从50%提高到57%。然而，进一步扩展时存在两



**Figure 7. 使用并行扩展方法进一步扩展。** 所有指标均为 AIME24 中 30 个问题的平均值。对于 REBASE，平均思考 token 数不包括来自奖励模型的额外计算。对于顺序扩展，我们提示模型使用最多（从左到右）32、64、256 和 512 步。对于 REBASE 和多数投票，我们生成 16 个并行轨迹进行聚合。

个关键限制：它最终会趋于平缓 (Figure 4)，并且基础语言模型的上下文窗口限制了它。尽管存在这些限制，我们的工作显示了在广泛准确度范围内进行测试时扩展 (Figure 1)，部分原因是减少测试时计算的行为是可预测的，并且不会受到这些限制的影响。

继续进行测试时扩展将需要能够进一步外推测试时计算的方法。我们如何实现这种外推？预算强制可能有改进，例如循环使用不同的字符串，而不仅仅是“等待”，或者将其与频率惩罚或更高的温度结合以避免重复循环。未来研究的一个令人兴奋的方向是研究将预算强制应用于通过强化学习训练的推理模型是否能产生更好的外推；或者RL是否允许新的测试时扩展方法超越预算强制。我们的工作定义了正确的指标 (§3.2)——控制、扩展和性能——以促进未来在测试时计算外推方面的研究和进展。

**并行扩展作为解决方案** 并行扩展为顺序扩展的限制提供了一个解决方案，因此我们通过两种方法增强我们的顺序扩展模型：(I) 多数投票：生成  $k$  个解决方案后，最终的解决方案是出现频率最高的那个；(II) 通过 REBASE 进行树搜索：我们使用 REBASE 过程奖励模型，该模型从 LLaMA-34B 初始，而在合成过程奖励建模数据集上进一步微调 (Wu et al., 2024b)。然后我们通过多数投票聚合由 REBASE 生成的解决方案。如 Figure 7 所示，与多

data (Yu et al., 2024). Others have developed training methodologies specifically aimed at reasoning performance (Zelikman et al., 2022; 2024; Luo et al., 2025; Yuan et al., 2025; Wu et al., 2024a). Another significant line of work focuses on prompting-based methods to elicit and improve reasoning abilities, including methods like Chain-of-Thought prompting (Wei et al., 2023; Yao et al., 2023a;b; Bi et al., 2023; Fu et al., 2023; Zhang et al., 2024b; Xiang et al., 2025; Hu et al., 2024). These combined efforts aim to advance the reasoning ability of language models, enabling them to handle more complex and abstract tasks effectively.

## 6.2. Test-time scaling

**Methods** As we introduce in §3, we differentiate two methods to scale test-time compute: **parallel** and **sequential**. The former relies on multiple solution attempts generated in parallel and selecting the best outcome via specific criteria. These criteria include choosing the most frequent response for majority voting or the best response based on an external reward for Best-of-N (Brown et al., 2024; Irvine et al., 2023; Snell et al., 2024). Unlike repeated sampling, previous sequential scaling methods let the model generate solution attempts sequentially based on previous attempts, allowing it to refine each attempt based on previous outcomes (Snell et al., 2024; Hou et al., 2025; Lee et al., 2025). Tree-based search methods (Gandhi et al., 2024; Wu et al., 2024b) offer a hybrid approach between sequential and parallel scaling, such as Monte-Carlo Tree Search (MCTS) (Liu et al., 2024; Zhang et al., 2023; Zhou et al., 2024; Choi et al., 2023) and guided beam search (Xie et al., 2023). REBASE (Wu et al., 2024b) employs a process reward model to balance exploitation and pruning during tree search. Empirically, REBASE has been shown to outperform sampling-based methods and MCTS (Wu et al., 2024b). Reward models (Lightman et al., 2023; Wang et al., 2024b;c) play a key role in these methods. They come in two variants: outcome reward models and process reward models.

Outcome reward models (Xin et al., 2024; Ankner et al., 2024) assign a score to complete solutions and are particularly useful in Best-of-N selection, while process reward models (Lightman et al., 2023; Wang et al., 2024b; Wu et al., 2024b) assess individual reasoning steps and are effective in guiding tree-based search methods.

**Limits to further test-time scaling** We have shown that budget forcing allows extrapolating test-time compute in §4, e.g., improving AIME24 performance from 50% to 57%. However, it has two key limitations when scaling further: it eventually **flattens out** (Figure 4), and the **context window** of the underlying language model constrains it. Despite these constraints, our work shows test-time scaling across a wide range of accuracies (Figure 1), partly because scaling down test-time compute behaves predictably and does not suffer from these constraints.

Continuing test-time scaling will require approaches that can further extrapolate test-time compute. How can we get such extrapolation? There may be improvements to budget forcing such as rotating through different strings, not only “Wait”, or combining it with frequency penalties or higher temperature to avoid repetitive loops. An exciting direction for future work is also researching whether applying budget forcing to a reasoning model trained with reinforcement learning yields better extrapolation; or if RL allows for new ways of test-time scaling beyond budget forcing. Our work defines the right metrics (§3.2) – Control, Scaling, and Performance – to enable future research and progress on extrapolating test-time compute.

**Parallel scaling as a solution** Parallel scaling offers one solution to the limits of sequential scaling, thus we augment our sequentially scaled model with two methods: **(I) Majority voting:** After generating  $k$  solutions, the final solution is the most frequent one across generations; **(II) Tree search via REBASE:** We use the REBASE process reward model, which is initialized from LLaMA-34B and further finetuned on a synthetic process reward

数投票和顺序扩展相比，使用 REBASE 增强的模型扩展效果更好。然而，REBASE 在每一步都需要额外的前向传递以计算奖励模型，这增加了一些计算开销。对于顺序扩展，当提示使用多达 512 步时，在 30 个评估问题中的 12 个问题中，模型生成的响应超过了上下文窗口，导致性能大幅下降。总体而言，我们发现这些并行扩展方法补充了顺序扩展，因此它们为超越固定上下文窗口的测试时计算扩展提供了途径。

## Impact Statement

具有强大推理能力的语言模型有潜力大大提高人类的生产力，从协助复杂的决策到推动科学突破。然而，最近的推理进展，如OpenAI的o1和DeepSeek的r1，缺乏透明度，限制了更广泛的研究进展。我们的工作旨在以完全开放的方式推动推理的前沿，促进创新和合作，加速最终惠及社会的进展。

## Acknowledgements

本研究部分使用了斯坦福大学的Marlowe GPU集群(Kapfer et al., 2025)，该集群的建立得到了斯坦福大学的财政支持。我们感谢Alexander M. Rush、Andrew Ilyas、Banghua Zhu、Chenglei Si、Chunting Zhou、John Yang、Ludwig Schmidt、Samy Jelassi、Tengyu Ma、Xuechen Li、Yu Sun 和 Yue Zhang 提供的非常建设性的讨论。

## 参考文献

Ankner, Z., Paul, M., Cui, B., Chang, J. D., and Ammanabrolu, P. Critique-out-loud reward models, 2024. URL <https://arxiv.org/abs/2408.11791>.

Arora, D., Singh, H. G., and Mausam. Have llms advanced enough? a challenging problem solving benchmark for large language models, 2023. URL <https://arxiv.org/abs/2305.15074>.

Azerbayev, Z., Schoelkopf, H., Paster, K., Santos, M. D., McAleer, S., Jiang, A. Q., Deng, J., Biderman, S., and Welleck, S. Llemma: An open language model for mathematics, 2023.

Bi, Z., Zhang, N., Jiang, Y., Deng, S., Zheng, G., and Chen, H. When do program-of-thoughts work for reasoning?, 2023. URL <https://arxiv.org/abs/2308.15452>.

Biderman, S., Schoelkopf, H., Sutawika, L., Gao, L., Tow, J., Abbasi, B., Aji, A. F., Ammanamanchi, P. S., Black, S., Clive, J., DiPofi, A., Etxaniz, J., Fattori, B., Forde, J. Z., Foster, C., Hsu, J., Jaiswal, M., Lee, W. Y., Li, H., Lovering, C., Muennighoff, N., Pavlick, E., Phang, J., Skowron, A., Tan, S., Tang, X., Wang, K. A., Winata, G. I., Yvon, F., and Zou, A. Lessons from the trenches on reproducible evaluation of language models, 2024.

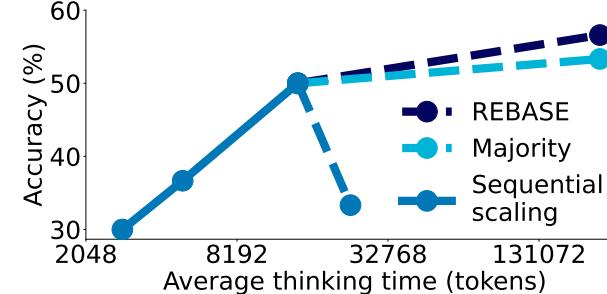
Brown, B., Juravsky, J., Ehrlich, R., Clark, R., Le, Q. V., Ré, C., and Mirhoseini, A. Large language monkeys: Scaling inference compute with repeated sampling, 2024. URL <https://arxiv.org/abs/2407.21787>.

Cesista, F. L. Multimodal structured generation: Cvpr’s 2nd mmfm challenge technical report, 2024. URL <https://arxiv.org/abs/2406.11403>.

Chen, W., Yin, M., Ku, M., Lu, P., Wan, Y., Ma, X., Xu, J., Wang, X., and Xia, T. Theoremqa: A theorem-driven question answering dataset, 2023. URL <https://arxiv.org/abs/2305.12524>.

Choi, S., Fang, T., Wang, Z., and Song, Y. Kcts: Knowledge-constrained tree search decoding with token-level hallucination detection, 2023. URL <https://arxiv.org/abs/2310.09044>.

DeepSeek-AI, Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., Zhu, Q., Ma, S., Wang, P., Bi, X., Zhang, X., Yu, X., Wu, Y., Wu, Z. F., Gou, Z., Shao,



**Figure 7. Scaling further with parallel scaling methods.** All metrics averaged over the 30 questions in AIME24. Average thinking tokens for REBASE do not account for the additional compute from the reward model. For sequential scaling, we prompt the model to use up to (from left to right) 32, 64, 256, and 512 steps. For REBASE and majority voting we generate 16 parallel trajectories to aggregate across.

modeling dataset (Wu et al., 2024b). We then aggregate the solutions generated by REBASE via majority voting. As shown in Figure 7, augmenting our model with REBASE scales better than majority voting, and even sequential scaling in this scenario. However, REBASE requires an additional forward pass at each step for the reward model adding some computation overhead. For sequential scaling, when prompted to use up to 512 steps, for 12 out of the 30 evaluation questions the model generates a response that exceeds the context window leading to a large performance drop. Overall, we find that these parallel scaling methods complement sequential scaling thus they offer an avenue for scaling test-time compute even further; beyond fixed context windows.

## Impact Statement

Language models with strong reasoning capabilities have the potential to greatly enhance human productivity, from assisting in complex decision-making to driving scientific breakthroughs. However, recent advances in reasoning, such as OpenAI’s o1 and DeepSeek’s r1, lack transparency, limiting broader research progress. Our work aims to push the frontier of reasoning in a fully open manner, fostering innovation and collaboration to accelerate advancements that ultimately benefit society.

## Acknowledgements

This work was partly conducted using the Stanford Marlowe GPU cluster (Kapfer et al., 2025) made possible by financial support from Stanford University. We thank Alexander M. Rush, Andrew Ilyas, Banghua Zhu, Chenglei Si, Chunting Zhou, John Yang, Ludwig Schmidt, Samy Jelassi, Tengyu Ma, Xuechen Li, Yu Sun and Yue Zhang for very constructive discussions.

## 参考文献

- Ankner, Z., Paul, M., Cui, B., Chang, J. D., and Ammanabrolu, P. Critique-out-loud reward models, 2024. URL <https://arxiv.org/abs/2408.11791>.
- Arora, D., Singh, H. G., and Mausam. Have llms advanced enough? a challenging problem solving benchmark for large language models, 2023. URL <https://arxiv.org/abs/2305.15074>.
- Azerbayev, Z., Schoelkopf, H., Paster, K., Santos, M. D., McAleer, S., Jiang, A. Q., Deng, J., Biderman, S., and Welleck, S. Llemma: An open language model for mathematics, 2023.
- Bi, Z., Zhang, N., Jiang, Y., Deng, S., Zheng, G., and Chen, H. When do program-of-thoughts work for reasoning?, 2023. URL <https://arxiv.org/abs/2308.15452>.
- Biderman, S., Schoelkopf, H., Sutawika, L., Gao, L., Tow, J., Abbasi, B., Aji, A. F., Ammanamanchi, P. S., Black, S., Clive, J., DiPofi, A., Etxaniz, J., Fattori, B., Forde, J. Z., Foster, C., Hsu, J., Jaiswal, M., Lee, W. Y., Li, H., Lovering, C., Muennighoff, N., Pavlick, E., Phang, J., Skowron, A., Tan, S., Tang, X., Wang, K. A., Winata, G. I., Yvon, F., and Zou, A. Lessons from the trenches on reproducible evaluation of language models, 2024.
- Brown, B., Juravsky, J., Ehrlich, R., Clark, R., Le, Q. V., Ré, C., and Mirhoseini, A. Large language monkeys: Z., Li, Z., Gao, Z., Liu, A., Xue, B., Wang, B., Wu, B., Feng, B., Lu, C., Zhao, C., Deng, C., Zhang, C., Ruan, C., Dai, D., Chen, D., Ji, D., Li, E., Lin, F., Dai, F., Luo, F., Hao, G., Chen, G., Li, G., Zhang, H., Bao, H., Xu, H., Wang, H., Ding, H., Xin, H., Gao, H., Qu, H., Li, H., Guo, J., Li, J., Wang, J., Chen, J., Yuan, J., Qiu, J., Li, J., Cai, J. L., Ni, J., Liang, J., Chen, J., Dong, K., Hu, K., Gao, K., Guan, K., Huang, K., Yu, K., Wang, L., Zhang, L., Zhao, L., Wang, L., Zhang, L., Xu, L., Xia, L., Zhang, M., Zhang, M., Tang, M., Li, M., Wang, M., Li, M., Tian, N., Huang, P., Zhang, P., Wang, Q., Chen, Q., Du, Q., Ge, R., Zhang, R., Pan, R., Wang, R., Chen, R. J., Jin, R. L., Chen, R., Lu, S., Zhou, S., Chen, S., Ye, S., Wang, S., Yu, S., Zhou, S., Pan, S., Li, S. S., Zhou, S., Wu, S., Ye, S., Yun, T., Pei, T., Sun, T., Wang, T., Zeng, W., Zhao, W., Liu, W., Liang, W., Gao, W., Yu, W., Zhang, W., Xiao, W. L., An, W., Liu, X., Wang, X., Chen, X., Nie, X., Cheng, X., Liu, X., Xie, X., Liu, X., Yang, X., Li, X., Su, X., Lin, X., Li, X. Q., Jin, X., Shen, X., Chen, X., Sun, X., Wang, X., Song, X., Zhou, X., Wang, X., Shan, X., Li, Y. K., Wang, Y. Q., Wei, Y. X., Zhang, Y., Xu, Y., Li, Y., Zhao, Y., Sun, Y., Wang, Y., Yu, Y., Zhang, Y., Shi, Y., Xiong, Y., He, Y., Piao, Y., Wang, Y., Tan, Y., Ma, Y., Liu, Y., Guo, Y., Ou, Y., Wang, Y., Gong, Y., Zou, Y., He, Y., Xiong, Y., Luo, Y., You, Y., Liu, Y., Zhou, Y., Zhu, Y. X., Xu, Y., Huang, Y., Li, Y., Zheng, Y., Zhu, Y., Ma, Y., Tang, Y., Zha, Y., Yan, Y., Ren, Z. Z., Ren, Z., Sha, Z., Fu, Z., Xu, Z., Xie, Z., Zhang, Z., Hao, Z., Ma, Z., Yan, Z., Wu, Z., Gu, Z., Zhu, Z., Liu, Z., Li, Z., Xie, Z., Song, Z., Pan, Z., Huang, Z., Xu, Z., Zhang, Z., and Zhang, Z. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, 2025. URL <https://arxiv.org/abs/2501.12948>.
- Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Yang, A., Fan, A., Goyal, A., Hartshorn, A., Yang, A., Mitra, A., Sravankumar, A., Korenev, A., Hinsvark, A., Rao, A., Zhang, A., Rodriguez, A., Gregerson, A., et al. The llama 3 herd of models, 2024. URL <https://arxiv.org/abs/2407.21783>.
- Fu, Y., Peng, H., Sabharwal, A., Clark, P., and Khot, T. Complexity-based prompting for multi-step reasoning, 2023. URL <https://arxiv.org/abs/2210.00720>.
- Gandhi, K., Lee, D., Grand, G., Liu, M., Cheng, W., Sharma, A., and Goodman, N. D. Stream of search (sos): Learning to search in language, 2024. URL <https://arxiv.org/abs/2404.03683>.
- Gao, B., Song, F., Yang, Z., Cai, Z., Miao, Y., Dong, Q., Li, L., Ma, C., Chen, L., Xu, R., Tang, Z., Wang, B., Zan, D., Quan, S., Zhang, G., Sha, L., Zhang, Y., Ren, X., Liu, T., and Chang, B. Omni-math: A universal olympiad level mathematic benchmark for large language models, 2024a. URL <https://arxiv.org/abs/2410.07985>.
- Gao, L., Tow, J., Biderman, S., Black, S., DiPofi, A., Foster, C., Golding, L., Hsu, J., McDonell, K., Muennighoff, N., Phang, J., Reynolds, L., Tang, E., Thite, A., Wang, B., Wang, K., and Zou, A. A framework for few-shot language model evaluation, September 2021. URL <https://doi.org/10.5281/zenodo.5371628>.
- Gao, Z., Niu, B., He, X., Xu, H., Liu, H., Liu, A., Hu, X., and Wen, L. Interpretable contrastive monte carlo tree search reasoning, 2024b. URL <https://arxiv.org/abs/2410.01707>.
- Glazer, E., Erdil, E., Besiroglu, T., Chicharro, D., Chen, E., Gunning, A., Olsson, C. F., Denain, J.-S., Ho, A., de Oliveira Santos, E., Järvinemi, O., Barnett, M., Sandler, R., Vrzala, M., Sevilla, J., Ren, Q., Pratt, E., Levine, L., Barkley, G., Stewart, N., Grechuk, B., Grechuk, T., Enugandla, S. V., and Wildon, M. Frontiermath: A benchmark for evaluating advanced

- 
- Scaling inference compute with repeated sampling, 2024. URL <https://arxiv.org/abs/2407.21787>.
- Cesista, F. L. Multimodal structured generation: Cvpr’s 2nd mmfm challenge technical report, 2024. URL <https://arxiv.org/abs/2406.11403>.
- Chen, W., Yin, M., Ku, M., Lu, P., Wan, Y., Ma, X., Xu, J., Wang, X., and Xia, T. Theoremqa: A theorem-driven question answering dataset, 2023. URL <https://arxiv.org/abs/2305.12524>.
- Choi, S., Fang, T., Wang, Z., and Song, Y. Kcts: Knowledge-constrained tree search decoding with token-level hallucination detection, 2023. URL <https://arxiv.org/abs/2310.09044>.
- DeepSeek-AI, Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., Zhu, Q., Ma, S., Wang, P., Bi, X., Zhang, X., Yu, X., Wu, Y., Wu, Z. F., Gou, Z., Shao, Z., Li, Z., Gao, Z., Liu, A., Xue, B., Wang, B., Wu, B., Feng, B., Lu, C., Zhao, C., Deng, C., Zhang, C., Ruan, C., Dai, D., Chen, D., Ji, D., Li, E., Lin, F., Dai, F., Luo, F., Hao, G., Chen, G., Li, G., Zhang, H., Bao, H., Xu, H., Wang, H., Ding, H., Xin, H., Gao, H., Qu, H., Li, H., Guo, J., Li, J., Wang, J., Chen, J., Yuan, J., Qiu, J., Li, J., Cai, J. L., Ni, J., Liang, J., Chen, J., Dong, K., Hu, K., Gao, K., Guan, K., Huang, K., Yu, K., Wang, L., Zhang, L., Zhao, L., Wang, L., Zhang, L., Xu, L., Xia, L., Zhang, M., Zhang, M., Tang, M., Li, M., Wang, M., Li, M., Tian, N., Huang, P., Zhang, P., Wang, Q., Chen, Q., Du, Q., Ge, R., Zhang, R., Pan, R., Wang, R., Chen, R. J., Jin, R. L., Chen, R., Lu, S., Zhou, S., Chen, S., Ye, S., Wang, S., Yu, S., Zhou, S., Pan, S., Li, S. S., Zhou, S., Wu, S., Ye, S., Yun, T., Pei, T., Sun, T., Wang, T., Zeng, W., Zhao, W., Liu, W., Liang, W., Gao, W., Yu, W., Zhang, W., Xiao, W. L., An, W., Liu, X., Wang, X., Chen, X., Nie, X., Cheng, X., Liu, X., Xie, X., Liu, X., Yang, X., Li, X., Su, X., Lin, X., Li, X. Q., Jin, X., Shen, X., Chen, X., Sun, X., Wang, X., Song, X., Zhou, X., Wang, X., Shan, X., Li, Y. K., Wang, Y. Q., Wei, Y. X., Zhang, Y., Xu, Y., Li, Y., Zhao, Y., Sun, Y., Wang, Y., Yu, Y., Zhang, Y., Shi, Y., Xiong, Y., He, Y., Piao, Y., Wang, Y., Tan, Y., Ma, Y., Liu, Y., Guo, Y., Ou, Y., Wang, Y., Gong, Y., Zou, Y., He, Y., Xiong, Y., Luo, Y., You, Y., Liu, Y., Zhou, Y., Zhu, Y. X., Xu, Y., Huang, Y., Li, Y., Zheng, Y., Zhu, Y., Ma, Y., Tang, Y., Zha, Y., Yan, Y., Ren, Z. Z., Ren, Z., Sha, Z., Fu, Z., Xu, Z., Xie, Z., Zhang, Z., Hao, Z., Ma, Z., Yan, Z., Wu, Z., Gu, Z., Zhu, Z., Liu, Z., Li, Z., Xie, Z., Song, Z., Pan, Z., Huang, Z., Xu, Z., Zhang, Z., and Zhang, Z. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, 2025. URL <https://arxiv.org/abs/2501.12948>.
- Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Yang, A., Fan, A., Goyal, A., Hartshorn, A., Yang, A., Mitra, A., Sravankumar, A., Korenev, A., Hinsvark, A., Rao, A., Zhang, A., Rodriguez, A., Gregerson, A., et al. The llama 3 herd of models, 2024. URL <https://arxiv.org/abs/2407.21783>.
- Fu, Y., Peng, H., Sabharwal, A., Clark, P., and Khot, T. Complexity-based prompting for multi-step reasoning, 2023. URL <https://arxiv.org/abs/2210.00720>.
- Gandhi, K., Lee, D., Grand, G., Liu, M., Cheng, W., Sharma, A., and Goodman, N. D. Stream of search (sos): Learning to search in language, 2024. URL <https://arxiv.org/abs/2404.03683>.
- Gao, B., Song, F., Yang, Z., Cai, Z., Miao, Y., Dong, Q., Li, L., Ma, C., Chen, L., Xu, R., Tang, Z., Wang, B., Zan, D., Quan, S., Zhang, G., Sha, L., Zhang, Y., Ren, X., Liu, T., and Chang, B. Omni-math: A universal olympiad level mathematic benchmark for large language models, 2024a. URL <https://arxiv.org/abs/2410.07985>.
- Gao, L., Tow, J., Biderman, S., Black, S., DiPofi, A., Foster, C., Golding, L., Hsu, J., McDonell, K., Muennighoff, N., Naik, A., Nam, C., Peters, M. E., Pyatkin, V., Ravichander, A., Schwenk, D., Shah, S., Smith, W., Strubell, E., Subramani, N., Wortsman, M., Dasigi, P., Lambert, N., Richardson, K., Zettlemoyer, L., Dodge, J., Lo, K., Soldaini, L., Smith, N. A., and Hajishirzi, H. Olmo: Accelerating the science of language models, 2024.
- He, C., Luo, R., Bai, Y., Hu, S., Thai, Z. L., Shen, J., Hu, J., Han, X., Huang, Y., Zhang, Y., Liu, J., Qi, L., Liu, Z., and Sun, M. Olympiadbench: A challenging benchmark for promoting agi with olympiad-level bilingual multimodal scientific problems, 2024. URL <https://arxiv.org/abs/2402.14008>.
- Hendrycks, D., Burns, C., Kadavath, S., Arora, A., Basart, S., Tang, E., Song, D., and Steinhardt, J. Measuring mathematical problem solving with the math dataset, 2021. URL <https://arxiv.org/abs/2103.03874>.
- Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., de Las Casas, D., Hendricks, L. A., Welbl, J., Clark, A., Hennigan, T., Noland, E., Millican, K., van den Driessche, G., Damoc, B., Guy, A., Osindero, S., Simonyan, K., Elsen, E., Rae, J. W., Vinyals, O., and Sifre, L. Training compute-optimal large language models, 2022. URL <https://arxiv.org/abs/2203.15556>.
- Hou, Z., Lv, X., Lu, R., Zhang, J., Li, Y., Yao, Z., Li, J., Tang, J., and Dong, Y. Advancing language model reasoning through reinforcement learning and inference scaling, 2025. URL <https://arxiv.org/abs/2501.11651>.
- Hu, Y., Shi, W., Fu, X., Roth, D., Ostendorf, M., Zettlemoyer, L., Smith, N. A., and Krishna, R. Visual sketchpad: Sketching as a visual chain of thought for multimodal language models, 2024. URL <https://arxiv.org/abs/2406.09403>.
- Huang, Z., Wang, Z., Xia, S., Li, X., Zou, H., Xu, R., Fan, R.-Z., Ye, L., Chern, E., Ye, Y., Zhang, Y., Yang, Y., Wu, T., Wang, B., Sun, S., Xiao, Y., Li, Y., Zhou, F., Chern, S., Qin, Y., Ma, Y., Su, J., Liu, Y., Zheng, Y., Zhang, S., Lin, D., Qiao, Y., and Liu, P. Olympicarena: Benchmarking multi-discipline cognitive reasoning for superintelligent ai, 2024a. URL <https://arxiv.org/abs/2406.12753>.
- Huang, Z., Zou, H., Li, X., Liu, Y., Zheng, Y., Chern, E., Xia, S., Qin, Y., Yuan, W., and Liu, P. O1 replication journey – part 2: Surpassing o1-preview through simple distillation, big progress or bitter lesson?, 2024b. URL <https://arxiv.org/abs/2411.16489>.
- Huang, Z., Geng, G., Hua, S., Huang, Z., Zou, H., Zhang, S., Liu, P., and Zhang, X. O1 replication journey – part 3: Inference-time scaling for medical reasoning, 2025. URL <https://arxiv.org/abs/2501.06458>.
- Irvine, R., Boubert, D., Raina, V., Liusie, A., Zhu, Z., Mudupalli, V., Korshuk, A., Liu, Z., Cremer, F., Assassi, V., Beauchamp, C.-C., Lu, X., Rialan, T., and Beauchamp, W. Rewarding chatbots for real-world engagement with millions of users, 2023. URL <https://arxiv.org/abs/2303.06135>.
- Jain, N., Han, K., Gu, A., Li, W.-D., Yan, F., Zhang, T., Wang, S., Solar-Lezama, A., Sen, K., and Stoica, I.

- 
- nighoff, N., Phang, J., Reynolds, L., Tang, E., Thite, A., Wang, B., Wang, K., and Zou, A. A framework for few-shot language model evaluation, September 2021. URL <https://doi.org/10.5281/zenodo.5371628>.
- Gao, Z., Niu, B., He, X., Xu, H., Liu, H., Liu, A., Hu, X., and Wen, L. Interpretable contrastive monte carlo tree search reasoning, 2024b. URL <https://arxiv.org/abs/2410.01707>.
- Glazer, E., Erdil, E., Besiroglu, T., Chicharro, D., Chen, E., Gunning, A., Olsson, C. F., Denain, J.-S., Ho, A., de Oliveira Santos, E., Järvinen, O., Barnett, M., Sandler, R., Vrzala, M., Sevilla, J., Ren, Q., Pratt, E., Levine, L., Barkley, G., Stewart, N., Grechuk, B., Grechuk, T., Enugandla, S. V., and Wildon, M. Frontiermath: A benchmark for evaluating advanced mathematical reasoning in ai, 2024. URL <https://arxiv.org/abs/2411.04872>.
- Google. Gemini 2.0 flash thinking mode (gemini-2.0-flash-thinking-exp-1219), December 2024. URL <https://cloud.google.com/vertex-ai/generative-ai/docs/thinking-mode>.
- Groeneveld, D., Beltagy, I., Walsh, P., Bhagia, A., Kinney, R., Tafjord, O., Jha, A. H., Ivison, H., Magnusson, I., Wang, Y., Arora, S., Atkinson, D., Author, R., Chandu, K. R., Cohan, A., Dumas, J., Elazar, Y., Gu, Y., Hessel, J., Khot, T., Merrill, W., Morrison, J., Muennighoff, N., Naik, A., Nam, C., Peters, M. E., Pyatkin, V., Ravichander, A., Schwenk, D., Shah, S., Smith, W., Strubell, E., Subramani, N., Wortsman, M., Dasigi, P., Lambert, N., Richardson, K., Zettlemoyer, L., Dodge, J., Lo, K., Soldaini, L., Smith, N. A., and Hajishirzi, H. Olmo: Accelerating the science of language models, 2024.
- He, C., Luo, R., Bai, Y., Hu, S., Thai, Z. L., Shen, J., Hu, J., Han, X., Huang, Y., Zhang, Y., Liu, J., Qi, L., Liu, Z., and Sun, M. Olympiadbench: A challenging benchmark for promoting agi with olympiad-level bilingual multimodal scientific problems, 2024. URL <https://arxiv.org/abs/2402.14008>.
- Hendrycks, D., Burns, C., Kadavath, S., Arora, A., Basart, S., Tang, E., Song, D., and Steinhardt, J. Measuring mathematical problem solving with the math dataset, 2021. URL <https://arxiv.org/abs/2103.03874>.
- Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., de Las Casas, D., Hendricks, L. A., Welbl, J., Clark, A., Hennigan, T., Noland, E., Millican, K., van den Driessche, G., Damoc, B., Guy, A., Osindero, S., Simonyan, K., Elsen, E., Rae, J. W., Vinyals, O., and Sifre, L. Training compute-optimal large language models, 2022. URL <https://arxiv.org/abs/2203.15556>.
- Hou, Z., Lv, X., Lu, R., Zhang, J., Li, Y., Yao, Z., Li, J., Tang, J., and Dong, Y. Advancing language model reasoning through reinforcement learning and inference scaling, 2025. URL <https://arxiv.org/abs/2501.11651>.
- Hu, Y., Shi, W., Fu, X., Roth, D., Ostendorf, M., Zettlemoyer, L., Smith, N. A., and Krishna, R. Visual sketchpad: Sketching as a visual chain of thought for multimodal language models, 2024. URL <https://arxiv.org/abs/2406.09403>.
- Huang, Z., Wang, Z., Xia, S., Li, X., Zou, H., Xu, R., Fan, R.-Z., Ye, L., Chern, E., Ye, Y., Zhang, Y., Yang, Y., Wu, T., Wang, B., Sun, S., Xiao, Y., Li, Y., Zhou, F., Chern, S., Qin, Y., Ma, Y., Su, J., Liu, Y., Zheng, Y., Zhang, S., Lin, D., Qiao, Y., and Liu, P. Olympicarena: Benchmarking multi-discipline cognitive reasoning for superintelligent ai, 2024a. URL <https://arxiv.org/abs/2406.12753>.
- Huang, Z., Zou, H., Li, X., Liu, Y., Zheng, Y., Chern, E., Xia, S., Qin, Y., Yuan, W., and Liu, P. O1 replication journey – part 2: Surpassing o1-preview
- Livecodebench: Holistic and contamination free evaluation of large language models for code, 2024. URL <https://arxiv.org/abs/2403.07974>.
- Kapfer, C., Stine, K., Narasimhan, B., Mentzel, C., and Candes, E. Marlowe: Stanford’s gpu-based computational instrument, January 2025. URL <https://doi.org/10.5281/zenodo.14751899>.
- Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., and Amodei, D. Scaling laws for neural language models, 2020. URL <https://arxiv.org/abs/2001.08361>.
- Kim, E., Suk, J., Kim, S., Muennighoff, N., Kim, D., and Oh, A. Llm-as-an-interviewer: Beyond static testing through dynamic llm evaluation, 2024. URL <https://arxiv.org/abs/2412.10424>.
- Kwon, W., Li, Z., Zhuang, S., Sheng, Y., Zheng, L., Yu, C. H., Gonzalez, J. E., Zhang, H., and Stoica, I. Efficient memory management for large language model serving with pagedattention, 2023. URL <https://arxiv.org/abs/2309.06180>.
- Labs, B. Bespoke-stratos: The unreasonable effectiveness of reasoning distillation, 2025. URL <https://hf.co/bespokelabs/Bespoke-Stratos-32B>. Accessed: 2025-01-22.
- Lee, K.-H., Fischer, I., Wu, Y.-H., Marwood, D., Baluja, S., Schuurmans, D., and Chen, X. Evolving deeper llm thinking, 2025. URL <https://arxiv.org/abs/2501.09891>.
- LI, J., Beeching, E., Tunstall, L., Lipkin, B., Solotskyi, R., Huang, S. C., Rasul, K., Yu, L., Jiang, A., Shen, Z., Qin, Z., Dong, B., Zhou, L., Fleureau, Y., Lample, G., and Polu, S. Numinamath, 2024. URL [https://github.com/project-numina/aimo-progress-prize/blob/main/report/numina\\_dataset.pdf](https://github.com/project-numina/aimo-progress-prize/blob/main/report/numina_dataset.pdf).
- Lightman, H., Kosaraju, V., Burda, Y., Edwards, H., Baker, B., Lee, T., Leike, J., Schulman, J., Sutskever, I., and Cobbe, K. Let’s verify step by step, 2023. URL <https://arxiv.org/abs/2305.20050>.
- Ling, W., Yogatama, D., Dyer, C., and Blunsom, P. Program induction by rationale generation : Learning to solve and explain algebraic word problems, 2017. URL <https://arxiv.org/abs/1705.04146>.
- Liu, J., Cui, L., Liu, H., Huang, D., Wang, Y., and Zhang, Y. Logiqa: A challenge dataset for machine reading comprehension with logical reasoning, 2020. URL <https://arxiv.org/abs/2007.08124>.
- Liu, J., Cohen, A., Pasunuru, R., Choi, Y., Hajishirzi, H., and Celikyilmaz, A. Don’t throw away your value model! generating more preferable text with value-guided monte-carlo tree search decoding, 2024. URL <https://arxiv.org/abs/2309.15028>.
- Loshchilov, I. and Hutter, F. Decoupled weight decay regularization, 2019.
- Luo, H., Sun, Q., Xu, C., Zhao, P., Lou, J., Tao, C., Geng, X., Lin, Q., Chen, S., Tang, Y., and Zhang, D. Wizardmath: Empowering mathematical reasoning for large language models via reinforced evol-instruct, 2025. URL <https://arxiv.org/abs/2308.09583>.
- Muennighoff, N., Soldaini, L., Groeneveld, D., Lo, K., Morrison, J., Min, S., Shi, W., Walsh, P., Tafjord, O., Lambert, N., Gu, Y., Arora, S., Bhagia, A., Schwenk, D., Wadden, D., Wettig, A., Hui, B., Dettmers, T., Kiela, D., Farhadi, A., Smith, N. A., Koh, P. W., Singh, A., and Hajishirzi, H. Olmoe: Open mixture-of-experts language models, 2024. URL <https://arxiv.org/abs/2409.02060>.
- of America, M. A. Aime, February 2024. URL [https://artofproblemsolving.com/wiki/index.php/AIME\\_Problems\\_and\\_Solutions/](https://artofproblemsolving.com/wiki/index.php/AIME_Problems_and_Solutions/).

- 
- through simple distillation, big progress or bitter lesson?, 2024b. URL <https://arxiv.org/abs/2411.16489>.
- Huang, Z., Geng, G., Hua, S., Huang, Z., Zou, H., Zhang, S., Liu, P., and Zhang, X. O1 replication journey – part 3: Inference-time scaling for medical reasoning, 2025. URL <https://arxiv.org/abs/2501.06458>.
- Irvine, R., Boubert, D., Raina, V., Liusie, A., Zhu, Z., Mudupalli, V., Korshuk, A., Liu, Z., Cremer, F., Assassi, V., Beauchamp, C.-C., Lu, X., Rialan, T., and Beauchamp, W. Rewarding chatbots for real-world engagement with millions of users, 2023. URL <https://arxiv.org/abs/2303.06135>.
- Jain, N., Han, K., Gu, A., Li, W.-D., Yan, F., Zhang, T., Wang, S., Solar-Lezama, A., Sen, K., and Stoica, I. Livecodebench: Holistic and contamination free evaluation of large language models for code, 2024. URL <https://arxiv.org/abs/2403.07974>.
- Kapfer, C., Stine, K., Narasimhan, B., Mentzel, C., and Candes, E. Marlowe: Stanford’s gpu-based computational instrument, January 2025. URL <https://doi.org/10.5281/zenodo.14751899>.
- Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., and Amodei, D. Scaling laws for neural language models, 2020. URL <https://arxiv.org/abs/2001.08361>.
- Kim, E., Suk, J., Kim, S., Muennighoff, N., Kim, D., and Oh, A. Llm-as-an-interviewer: Beyond static testing through dynamic llm evaluation, 2024. URL <https://arxiv.org/abs/2412.10424>.
- Kwon, W., Li, Z., Zhuang, S., Sheng, Y., Zheng, L., Yu, C. H., Gonzalez, J. E., Zhang, H., and Stoica, I. Efficient memory management for large language model serving with pagedattention, 2023. URL <https://arxiv.org/abs/2309.06180>.
- Labs, B. Bespoke-stratos: The unreasonable effectiveness of reasoning distillation, 2025. URL <https://hf.co/bespokelabs/Bespoke-Stratos-32B>. Accessed: 2025-01-22.
- Lee, K.-H., Fischer, I., Wu, Y.-H., Marwood, D., Baluja, S., Schuurmans, D., and Chen, X. Evolving deeper llm thinking, 2025. URL <https://arxiv.org/abs/2501.09891>.
- LI, J., Beeching, E., Tunstall, L., Lipkin, B., Solotskyi, R., Huang, S. C., Rasul, K., Yu, L., Jiang, A., Shen, Z., Qin, Z., Dong, B., Zhou, L., Fleureau, Y., Lample, G., and Polu, S. Numinamath, 2024. URL [https://github.com/project-numina/aimo-progress-prize/blob/main/report/numina\\_dataset.pdf](https://github.com/project-numina/aimo-progress-prize/blob/main/report/numina_dataset.pdf).
- Lightman, H., Kosaraju, V., Burda, Y., Edwards, H., Baker, B., Lee, T., Leike, J., Schulman, J., Sutskever, I., and Cobbe, K. Let’s verify step by step, 2023. URL <https://arxiv.org/abs/2305.20050>.
- Ling, W., Yogatama, D., Dyer, C., and Blunsom, P. Program induction by rationale generation : Learning to solve and explain algebraic word problems, 2017. URL <https://arxiv.org/abs/1705.04146>.
- Liu, J., Cui, L., Liu, H., Huang, D., Wang, Y., and Zhang, Y. Logiqqa: A challenge dataset for machine reading comprehension with logical reasoning, 2020. URL <https://arxiv.org/abs/2007.08124>.
- Liu, J., Cohen, A., Pasunuru, R., Choi, Y., Hajishirzi, H., and Celikyilmaz, A. Don’t throw away your value model! generating more preferable text with value-guided monte-carlo tree search decoding, 2024. URL <https://arxiv.org/abs/2309.15028>.
- Loshchilov, I. and Hutter, F. Decoupled weight decay regularization, 2019.
- OpenAI. Learning to reason with llms, September 2024. URL <https://openai.com/index/learning-to-reason-with-llms/>.
- Phan, L., Gatti, A., Han, Z., Li, N., Hu, J., Zhang, H., Shi, S., Choi, M., Agrawal, A., Chopra, A., et al. Humanity’s last exam, 2025. URL <https://arxiv.org/abs/2501.14249>.
- Qin, Y., Li, X., Zou, H., Liu, Y., Xia, S., Huang, Z., Ye, Y., Yuan, W., Liu, H., Li, Y., and Liu, P. O1 replication journey: A strategic progress report – part 1, 2024. URL <https://arxiv.org/abs/2410.18982>.
- Qwen, :, Yang, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Li, C., Liu, D., Huang, F., Wei, H., Lin, H., Yang, J., Tu, J., Zhang, J., Yang, J., Yang, J., Zhou, J., Lin, J., Dang, K., Lu, K., Bao, K., Yang, K., Yu, L., Li, M., Xue, M., Zhang, P., Zhu, Q., Men, R., Lin, R., Li, T., Xia, T., Ren, X., Ren, X., Fan, Y., Su, Y., Zhang, Y., Wan, Y., Liu, Y., Cui, Z., Zhang, Z., and Qiu, Z. Qwen2.5 technical report, 2024. URL <https://arxiv.org/abs/2412.15115>.
- Rein, D., Hou, B. L., Stickland, A. C., Petty, J., Pang, R. Y., Dirani, J., Michael, J., and Bowman, S. R. Gpqa: A graduate-level google-proof q&a benchmark, 2023. URL <https://arxiv.org/abs/2311.12022>.
- Shi, Q., Tang, M., Narasimhan, K., and Yao, S. Can language models solve olympiad programming?, 2024. URL <https://arxiv.org/abs/2404.10952>.
- Snell, C., Lee, J., Xu, K., and Kumar, A. Scaling llm test-time compute optimally can be more effective than scaling model parameters, 2024. URL <https://arxiv.org/abs/2408.03314>.
- Srivastava, A., Rastogi, A., Rao, A., Shoeb, A. A. M., Abid, A., Fisch, A., Brown, A. R., Santoro, A., Gupta, A., Garriga-Alonso, A., et al. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models, 2023.
- Su, H., Yen, H., Xia, M., Shi, W., Muennighoff, N., yu Wang, H., Liu, H., Shi, Q., Siegel, Z. S., Tang, M., Sun, R., Yoon, J., Arik, S. O., Chen, D., and Yu, T. Bright: A realistic and challenging benchmark for reasoning-intensive retrieval, 2024. URL <https://arxiv.org/abs/2407.12883>.
- Sun, L., Han, Y., Zhao, Z., Ma, D., Shen, Z., Chen, B., Chen, L., and Yu, K. Scieval: A multi-level large language model evaluation benchmark for scientific research, 2024. URL <https://arxiv.org/abs/2308.13149>.
- Team, D. Deepseek r1, November 2024a. URL [https://x.com/deepseek\\_ai/status/185920014135536422](https://x.com/deepseek_ai/status/185920014135536422).
- Team, K., Du, A., Gao, B., Xing, B., Jiang, C., Chen, C., Li, C., Xiao, C., Du, C., Liao, C., Tang, C., Wang, C., Zhang, D., Yuan, E., Lu, E., Tang, F., Sung, F., Wei, G., Lai, G., Guo, H., Zhu, H., Ding, H., Hu, H., Yang, H., Zhang, H., Yao, H., Zhao, H., Lu, H., Li, H., Yu, H., Gao, H., Zheng, H., Yuan, H., Chen, J., Guo, J., Su, J., Wang, J., Zhao, J., Zhang, J., Liu, J., Yan, J., Wu, J., Shi, L., Ye, L., Yu, L., Dong, M., Zhang, N., Ma, N., Pan, Q., Gong, Q., Liu, S., Ma, S., Wei, S., Cao, S., Huang, S., Jiang, T., Gao, W., Xiong, W., He, W., Huang, W., Wu, W., He, W., Wei, X., Jia, X., Wu, X., Xu, X., Zu, X., Zhou, X., Pan, X., Charles, Y., Li, Y., Hu, Y., Liu, Y., Chen, Y., Wang, Y., Liu, Y., Qin, Y., Liu, Y., Yang, Y., Bao, Y., Du, Y., Wu, Y., Wang, Y., Zhou, Z., Wang, Z., Li, Z., Zhu, Z., Zhang, Z., Wang, Z., Yang, Z., Huang, Z., Hu, Z., Huang, Z., Xu, Z., and Yang, Z. Kimi k1.5: Scaling reinforcement learning with llms, 2025. URL <https://arxiv.org/abs/2501.12599>.
- Team, N. Sky-t1: Fully open-source reasoning model with o1-preview performance in \$450 budget,

- Luo, H., Sun, Q., Xu, C., Zhao, P., Lou, J., Tao, C., Geng, X., Lin, Q., Chen, S., Tang, Y., and Zhang, D. Wizardmath: Empowering mathematical reasoning for large language models via reinforced evol-instruct, 2025. URL <https://arxiv.org/abs/2308.09583>.
- Muennighoff, N., Soldaini, L., Groeneveld, D., Lo, K., Morrison, J., Min, S., Shi, W., Walsh, P., Tafjord, O., Lambert, N., Gu, Y., Arora, S., Bhagia, A., Schwenk, D., Wadden, D., Wettig, A., Hui, B., Dettmers, T., Kiela, D., Farhadi, A., Smith, N. A., Koh, P. W., Singh, A., and Hajishirzi, H. Olmoe: Open mixture-of-experts language models, 2024. URL <https://arxiv.org/abs/2409.02060>.
- of America, M. A. Aime, February 2024. URL [https://artofproblemsolving.com/wiki/index.php/AIME\\_Problems\\_and\\_Solutions/](https://artofproblemsolving.com/wiki/index.php/AIME_Problems_and_Solutions/).
- OpenAI. Learning to reason with llms, September 2024. URL <https://openai.com/index/learning-to-reason-with-llms/>.
- Phan, L., Gatti, A., Han, Z., Li, N., Hu, J., Zhang, H., Shi, S., Choi, M., Agrawal, A., Chopra, A., et al. Humanity’s last exam, 2025. URL <https://arxiv.org/abs/2501.14249>.
- Qin, Y., Li, X., Zou, H., Liu, Y., Xia, S., Huang, Z., Ye, Y., Yuan, W., Liu, H., Li, Y., and Liu, P. O1 replication journey: A strategic progress report – part 1, 2024. URL <https://arxiv.org/abs/2410.18982>.
- Qwen, ;, Yang, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Li, C., Liu, D., Huang, F., Wei, H., Lin, H., Yang, J., Tu, J., Zhang, J., Yang, J., Yang, J., Zhou, J., Lin, J., Dang, K., Lu, K., Bao, K., Yang, K., Yu, L., Li, M., Xue, M., Zhang, P., Zhu, Q., Men, R., Lin, R., Li, T., Xia, T., Ren, X., Ren, X., Fan, Y., Su, Y., Zhang, Y., Wan, Y., Liu, Y., Cui, Z., Zhang, Z., and Qiu, Z. Qwen2.5 technical report, 2024. URL <https://arxiv.org/abs/2412.15115>.
- Rein, D., Hou, B. L., Stickland, A. C., Petty, J., Pang, R. Y., Dirani, J., Michael, J., and Bowman, S. R. Gpqa: A graduate-level google-proof q&a benchmark, 2023. URL <https://arxiv.org/abs/2311.12022>.
- Shi, Q., Tang, M., Narasimhan, K., and Yao, S. Can language models solve olympiad programming?, 2024. URL <https://arxiv.org/abs/2404.10952>.
- Snell, C., Lee, J., Xu, K., and Kumar, A. Scaling llm test-time compute optimally can be more effective than scaling model parameters, 2024. URL <https://arxiv.org/abs/2408.03314>.
- Srivastava, A., Rastogi, A., Rao, A., Shoeb, A. A. M., Abid, A., Fisch, A., Brown, A. R., Santoro, A., Gupta, A., Garriga-Alonso, A., et al. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models, 2023.
- Su, H., Yen, H., Xia, M., Shi, W., Muennighoff, N., yu Wang, H., Liu, H., Shi, Q., Siegel, Z. S., Tang, M., Sun, R., Yoon, J., Arik, S. O., Chen, D., and Yu, T. Bright: A realistic and challenging benchmark for reasoning-intensive retrieval, 2024. URL <https://arxiv.org/abs/2407.12883>.
- Sun, L., Han, Y., Zhao, Z., Ma, D., Shen, Z., Chen, B., Chen, L., and Yu, K. Scieval: A multi-level large language model evaluation benchmark for scientific research, 2024. URL <https://arxiv.org/abs/2308.13149>.
- Team, D. Deepseek r1, November 2024a. URL [https://x.com/deepseek\\_ai/status/1859200141355536422](https://x.com/deepseek_ai/status/1859200141355536422).
- Team, K., Du, A., Gao, B., Xing, B., Jiang, C., Chen, C., Li, C., Xiao, C., Du, C., Liao, C., Tang, C., Wang, C., Wu, T., Lan, J., Yuan, W., Jiao, J., Weston, J., and Sukhbaatar, S. Thinking llms: General instruction following with thought generation, 2024a. URL <https://arxiv.org/abs/2410.10630>.
- Wu, Y., Sun, Z., Li, S., Welleck, S., and Yang, Y. Inference scaling laws: An empirical analysis of compute-optimal inference for problem-solving with language models, 2024b. URL <https://arxiv.org/abs/2408.00724>.
- Xiang, V., Snell, C., Gandhi, K., Albalak, A., Singh, A., Blagden, C., Phung, D., Rafailov, R., Lile, N., Mahan, D., Castricato, L., Franken, J.-P., Haber, N., and Finn, C. Towards system 2 reasoning in llms: Learning how to think with meta chain-of-thought, 2025. URL <https://arxiv.org/abs/2501.04682>.
- Xie, Y., Kawaguchi, K., Zhao, Y., Zhao, X., Kan, M.-Y., He, J., and Xie, Q. Self-evaluation guided beam search for reasoning, 2023. URL <https://arxiv.org/abs/2305.00633>.
- Xin, H., Guo, D., Shao, Z., Ren, Z., Zhu, Q., Liu, B., Ruan, C., Li, W., and Liang, X. Deepseek-prover: Advancing theorem proving in llms through large-scale synthetic data, 2024. URL <https://arxiv.org/abs/2405.14333>.
- Xu, H., Wu, X., Wang, W., Li, Z., Zheng, D., Chen, B., Hu, Y., Kang, S., Ji, J., Zhang, Y., Guo, Z., Yang, Y., Zhang, M., and Zhang, D. Redstar: Does scaling long-cot data unlock better slow-reasoning systems?, 2025. URL <https://arxiv.org/abs/2501.11284>.
- Yang, Z., Band, N., Li, S., Candès, E., and Hashimoto, T. Synthetic continued pretraining, 2024. URL <https://arxiv.org/abs/2409.07431>.
- Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao, Y., and Narasimhan, K. Tree of thoughts: Deliberate problem solving with large language models, 2023a. URL <https://arxiv.org/abs/2305.10601>.

- 
- C., Zhang, D., Yuan, E., Lu, E., Tang, F., Sung, F., Wei, G., Lai, G., Guo, H., Zhu, H., Ding, H., Hu, H., Yang, H., Zhang, H., Yao, H., Zhao, H., Lu, H., Li, H., Yu, H., Gao, H., Zheng, H., Yuan, H., Chen, J., Guo, J., Su, J., Wang, J., Zhao, J., Zhang, J., Liu, J., Yan, J., Wu, J., Shi, L., Ye, L., Yu, L., Dong, M., Zhang, N., Ma, N., Pan, Q., Gong, Q., Liu, S., Ma, S., Wei, S., Cao, S., Huang, S., Jiang, T., Gao, W., Xiong, W., He, W., Huang, W., Wu, W., He, W., Wei, X., Jia, X., Wu, X., Xu, X., Zu, X., Zhou, X., Pan, X., Charles, Y., Li, Y., Hu, Y., Liu, Y., Chen, Y., Wang, Y., Liu, Y., Qin, Y., Liu, Y., Yang, Y., Bao, Y., Du, Y., Wu, Y., Wang, Y., Zhou, Z., Wang, Z., Li, Z., Zhu, Z., Zhang, Z., Wang, Z., Yang, Z., Huang, Z., Huang, Z., Xu, Z., and Yang, Z. Kimi k1.5: Scaling reinforcement learning with llms, 2025. URL <https://arxiv.org/abs/2501.12599>.
- Team, N. Sky-t1: Fully open-source reasoning model with o1-preview performance in \$450 budget, 2025. URL <https://novasky-ai.github.io/posts/sky-t1>. Accessed: 2025-01-09.
- Team, Q. Qwq: Reflect deeply on the boundaries of the unknown, November 2024b. URL <https://qwenlm.github.io/blog/qwq-32b-preview/>.
- Wang, J., Meng, F., Liang, Y., and Zhou, J. Drt-o1: Optimized deep reasoning translation via long chain-of-thought, 2024a. URL <https://arxiv.org/abs/2412.17498>.
- Wang, P., Li, L., Shao, Z., Xu, R. X., Dai, D., Li, Y., Chen, D., Wu, Y., and Sui, Z. Math-shepherd: Verify and reinforce llms step-by-step without human annotations, 2024b. URL <https://arxiv.org/abs/2312.08935>.
- Wang, S., Liu, Z., Zhong, W., Zhou, M., Wei, Z., Chen, Z., and Duan, N. From lsat: The progress and challenges of complex reasoning, 2021. URL <https://arxiv.org/abs/2108.00648>.
- Wang, Z., Dong, Y., Delalleau, O., Zeng, J., Shen, G., Egert, D., Zhang, J. J., Sreedhar, M. N., and Kuchaiiev, O. Helpsteer2: Open-source dataset for training top-performing reward models, 2024c. URL <https://arxiv.org/abs/2406.08673>.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., and Zhou, D. Chain-of-thought prompting elicits reasoning in large language models, 2023. URL <https://arxiv.org/abs/2201.11903>.
- Welleck, S., Bertsch, A., Finlayson, M., Schoelkopf, H., Xie, A., Neubig, G., Kulikov, I., and Harchaoui, Z. From decoding to meta-generation: Inference-time algorithms for large language models, 2024. URL <https://arxiv.org/abs/2406.16838>.
- Wu, T., Lan, J., Yuan, W., Jiao, J., Weston, J., and Sukhbaatar, S. Thinking llms: General instruction following with thought generation, 2024a. URL <https://arxiv.org/abs/2410.10630>.
- Wu, Y., Sun, Z., Li, S., Welleck, S., and Yang, Y. Inference scaling laws: An empirical analysis of compute-optimal inference for problem-solving with language models, 2024b. URL <https://arxiv.org/abs/2408.00724>.
- Xiang, V., Snell, C., Gandhi, K., Albalak, A., Singh, A., Blagden, C., Phung, D., Rafailov, R., Lile, N., Mahan, D., Castricato, L., Franken, J.-P., Haber, N., and Finn, C. Towards system 2 reasoning in llms: Learning how to think with meta chain-of-thought, 2025. URL <https://arxiv.org/abs/2501.04682>.
- Xie, Y., Kawaguchi, K., Zhao, Y., Zhao, X., Kan, M.-Y., He, J., and Xie, Q. Self-evaluation guided beam search for reasoning, 2023. URL <https://arxiv.org/abs/2305.00633>.
- Xin, H., Guo, D., Shao, Z., Ren, Z., Zhu, Q., Liu, B., Ruan, C., Li, W., and Liang, X. Deepseek-prover:
- Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., and Cao, Y. React: Synergizing reasoning and acting in language models, 2023b. URL <https://arxiv.org/abs/2210.03629>.
- Yu, L., Jiang, W., Shi, H., Yu, J., Liu, Z., Zhang, Y., Kwok, J. T., Li, Z., Weller, A., and Liu, W. Metamath: Bootstrap your own mathematical questions for large language models, 2024. URL <https://arxiv.org/abs/2309.12284>.
- Yuan, S., Chen, Z., Xi, Z., Ye, J., Du, Z., and Chen, J. Agent-r: Training language model agents to reflect via iterative self-training, 2025. URL <https://arxiv.org/abs/2501.11425>.
- Zelikman, E., Wu, Y., Mu, J., and Goodman, N. D. Star: Bootstrapping reasoning with reasoning, 2022. URL <https://arxiv.org/abs/2203.14465>.
- Zelikman, E., Harik, G., Shao, Y., Jayasiri, V., Haber, N., and Goodman, N. D. Quiet-star: Language models can teach themselves to think before speaking, 2024. URL <https://arxiv.org/abs/2403.09629>.
- Zhang, H. and Chen, C. Test-time compute scaling laws, 2024. URL [https://github.com/hughbzhang/o1\\_inference\\_scaling\\_laws](https://github.com/hughbzhang/o1_inference_scaling_laws).
- Zhang, S., Chen, Z., Shen, Y., Ding, M., Tenenbaum, J., B., and Gan, C. Planning with large language models for code generation, 2023. URL <https://arxiv.org/abs/2303.05510>.
- Zhang, Y., Wu, S., Yang, Y., Shu, J., Xiao, J., Kong, C., and Sang, J. o1-coder: an o1 replication for coding, 2024a. URL <https://arxiv.org/abs/2412.00154>.
- Zhang, Y., Yang, J., Yuan, Y., and Yao, A. C.-C. Cumulative reasoning with large language models, 2024b. URL <https://arxiv.org/abs/2308.04371>.
- Zhong, H., Xiao, C., Tu, C., Zhang, T., Liu, Z., and Sun, M. Jec-qa: A legal-domain question answering dataset, 2019. URL <https://arxiv.org/abs/1911.12011>.
- Zhong, W., Cui, R., Guo, Y., Liang, Y., Lu, S., Wang, Y., Saied, A., Chen, W., and Duan, N. Agieval: A human-centric benchmark for evaluating foundation models, 2023. URL <https://arxiv.org/abs/2304.06364>.
- Zhou, A., Yan, K., Shlapentokh-Rothman, M., Wang, H., and Wang, Y.-X. Language agent tree search unifies reasoning acting and planning in language models, 2024. URL <https://arxiv.org/abs/2310.04406>.
- Zhou, C., Liu, P., Xu, P., Iyer, S., Sun, J., Mao, Y., Ma, X., Efrat, A., Yu, P., Yu, L., Zhang, S., Ghosh, G., Lewis, M., Zettlemoyer, L., and Levy, O. Lima: Less is more for alignment, 2023. URL <https://arxiv.org/abs/2305.11206>.

Advancing theorem proving in llms through large-scale synthetic data, 2024. URL <a href="https://arxiv.org/abs/2405.14333">https://arxiv.org/abs/2405.14333</a> .	teach themselves to think before speaking, 2024. URL <a href="https://arxiv.org/abs/2403.09629">https://arxiv.org/abs/2403.09629</a> .
Xu, H., Wu, X., Wang, W., Li, Z., Zheng, D., Chen, B., Hu, Y., Kang, S., Ji, J., Zhang, Y., Guo, Z., Yang, Y., Zhang, M., and Zhang, D. Redstar: Does scaling long-cot data unlock better slow-reasoning systems?, 2025. URL <a href="https://arxiv.org/abs/2501.11284">https://arxiv.org/abs/2501.11284</a> .	Zhang, H. and Chen, C. Test-time compute scaling laws, 2024. URL <a href="https://github.com/hughbzhang/o1_inference_scaling_laws">https://github.com/hughbzhang/o1_inference_scaling_laws</a> .
Yang, Z., Band, N., Li, S., Candès, E., and Hashimoto, T. Synthetic continued pretraining, 2024. URL <a href="https://arxiv.org/abs/2409.07431">https://arxiv.org/abs/2409.07431</a> .	Zhang, S., Chen, Z., Shen, Y., Ding, M., Tenenbaum, J. B., and Gan, C. Planning with large language models for code generation, 2023. URL <a href="https://arxiv.org/abs/2303.05510">https://arxiv.org/abs/2303.05510</a> .
Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao, Y., and Narasimhan, K. Tree of thoughts: Deliberate problem solving with large language models, 2023a. URL <a href="https://arxiv.org/abs/2305.10601">https://arxiv.org/abs/2305.10601</a> .	Zhang, Y., Wu, S., Yang, Y., Shu, J., Xiao, J., Kong, C., and Sang, J. o1-coder: an o1 replication for coding, 2024a. URL <a href="https://arxiv.org/abs/2412.00154">https://arxiv.org/abs/2412.00154</a> .
Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., and Cao, Y. React: Synergizing reasoning and acting in language models, 2023b. URL <a href="https://arxiv.org/abs/2210.03629">https://arxiv.org/abs/2210.03629</a> .	Zhang, Y., Yang, J., Yuan, Y., and Yao, A. C.-C. Cumulative reasoning with large language models, 2024b. URL <a href="https://arxiv.org/abs/2308.04371">https://arxiv.org/abs/2308.04371</a> .
Yu, L., Jiang, W., Shi, H., Yu, J., Liu, Z., Zhang, Y., Kwok, J. T., Li, Z., Weller, A., and Liu, W. Metamath: Bootstrap your own mathematical questions for large language models, 2024. URL <a href="https://arxiv.org/abs/2309.12284">https://arxiv.org/abs/2309.12284</a> .	Zhong, H., Xiao, C., Tu, C., Zhang, T., Liu, Z., and Sun, M. Jec-qa: A legal-domain question answering dataset, 2019. URL <a href="https://arxiv.org/abs/1911.12011">https://arxiv.org/abs/1911.12011</a> .
Yuan, S., Chen, Z., Xi, Z., Ye, J., Du, Z., and Chen, J. Agent-r: Training language model agents to reflect via iterative self-training, 2025. URL <a href="https://arxiv.org/abs/2501.11425">https://arxiv.org/abs/2501.11425</a> .	Zhong, W., Cui, R., Guo, Y., Liang, Y., Lu, S., Wang, Y., Saeid, A., Chen, W., and Duan, N. Agieval: A human-centric benchmark for evaluating foundation models, 2023. URL <a href="https://arxiv.org/abs/2304.06364">https://arxiv.org/abs/2304.06364</a> .
Zelikman, E., Wu, Y., Mu, J., and Goodman, N. D. Star: Bootstrapping reasoning with reasoning, 2022. URL <a href="https://arxiv.org/abs/2203.14465">https://arxiv.org/abs/2203.14465</a> .	Zhou, A., Yan, K., Shlapentokh-Rothman, M., Wang, H., and Wang, Y.-X. Language agent tree search unifies reasoning acting and planning in language models, 2024. URL <a href="https://arxiv.org/abs/2310.04406">https://arxiv.org/abs/2310.04406</a> .
Zelikman, E., Harik, G., Shao, Y., Jayasiri, V., Haber, N., and Goodman, N. D. Quiet-star: Language models can	Zhou, C., Liu, P., Xu, P., Iyer, S., Sun, J., Mao, Y., Ma, X., Efrat, A., Yu, P., Yu, L., Zhang, S., Ghosh, G., Lewis, M., Zettlemoyer, L., and Levy, O. Lima: Less is more for alignment, 2023. URL <a href="https://arxiv.org/abs/2305.11206">https://arxiv.org/abs/2305.11206</a> .

## 目录

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Reasoning data curation to create s1K</b>	<b>2</b>
2.1	Initial collection of 59K samples	2
2.2	Final selection of 1K samples	3
<b>3</b>	<b>Test-time scaling</b>	<b>4</b>
3.1	Method	4
3.2	Metrics	5
<b>4</b>	<b>Results</b>	<b>5</b>
4.1	Setup	5
4.2	Performance	6
<b>5</b>	<b>Ablations</b>	<b>7</b>
5.1	Data Quantity, Diversity, and Difficulty	7
5.2	Test-time scaling methods	7
<b>6</b>	<b>Discussion and related work</b>	<b>9</b>
6.1	Sample-efficient reasoning	9
6.2	Test-time scaling	10
<b>A</b>	<b>Evaluation determinism</b>	<b>19</b>
<b>B</b>	<b>s1K details</b>	<b>19</b>
B.1	s1K summary	19
B.2	Dataset composition for full 59K questions	19
B.3	s1K grading prompt	22
B.4	s1K diversity selection	22
B.5	Decontamination	22
<b>C</b>	<b>Training details</b>	<b>24</b>
C.1	Training Ablations: Sequence length	24
C.2	Training Samples	24
<b>D</b>	<b>Test-time scaling details</b>	<b>42</b>
D.1	Sequential scaling ablations	42
D.2	Examples for rejection sampling ablation	44

---

## 目录

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Reasoning data curation to create s1K</b>	<b>2</b>
2.1	Initial collection of 59K samples . . . . .	2
2.2	Final selection of 1K samples . . . . .	3
<b>3</b>	<b>Test-time scaling</b>	<b>4</b>
3.1	Method . . . . .	4
3.2	Metrics . . . . .	5
<b>4</b>	<b>Results</b>	<b>5</b>
4.1	Setup . . . . .	5
4.2	Performance . . . . .	7
<b>5</b>	<b>Ablations</b>	<b>7</b>
5.1	Data Quantity, Diversity, and Difficulty . . . . .	7
5.2	Test-time scaling methods . . . . .	9
<b>6</b>	<b>Discussion and related work</b>	<b>10</b>
6.1	Sample-efficient reasoning . . . . .	10
6.2	Test-time scaling . . . . .	11
<b>A</b>	<b>Evaluation determinism</b>	<b>20</b>
<b>B</b>	<b>s1K details</b>	<b>20</b>
B.1	s1K summary . . . . .	20
B.2	Dataset composition for full 59K questions . . . . .	20
B.3	s1K grading prompt . . . . .	23
B.4	s1K diversity selection . . . . .	23
B.5	Decontamination . . . . .	23
<b>C</b>	<b>Training details</b>	<b>25</b>
C.1	Training Ablations: Sequence length . . . . .	25
C.2	Training Samples . . . . .	26
<b>D</b>	<b>Test-time scaling details</b>	<b>44</b>
D.1	Sequential scaling ablations . . . . .	44
D.2	Examples for rejection sampling ablation . . . . .	47

---

## A. Evaluation determinism

我们使用vLLM (Kwon et al., 2023)进行评估，因为它比我们尝试的其他选项更快。然而，我们发现即使使用相同的随机种子和贪婪采样，评估分数在不同运行中仍会有显著变化：

- 不同的批量大小导致不同的结果，见 <https://github.com/vllm-project/vllm/issues/5898>
- 持续的代际导致不同的结果，见 <https://github.com/vllm-project/vllm/issues/11783>
- 张量并行性的变化导致不同的结果

由于我们的模型在生成答案之前会产生较长的推理轨迹，因此小的数值变化可能会滚雪球般地导致巨大的差异。我们遇到了许多生成结果，它们在数千个标记上完全相同，然后突然在一个标记上有所不同，最终导致完全不同的答案。为了部分地应对这一问题，除非另有说明，我们通常在最终评估中使用全精度。

## B. s1K details

### B.1. s1K summary

### B.2. Dataset composition for full 59K questions

## A. Evaluation determinism

We run our evaluations using vLLM (Kwon et al., 2023) as it is faster than the alternatives we tried. However, we find that even when using the same random seeds and greedy sampling, evaluation scores can change significantly across runs:

- Different batch sizes causing different results see <https://github.com/vllm-project/vllm/issues/5898>
- Continuing generations causing different results see <https://github.com/vllm-project/vllm/issues/11783>
- Changes in tensor parallelism causing different results

As our model generates long reasoning traces prior to its answer, small numeric changes can snowball into large differences. We encounter many generations that are exactly the same for thousands of tokens and then suddenly differ in one token eventually ending up with an entirely different answer. To partly counter this issue we generally run our final evaluations using full precision unless otherwise indicated.

## B. s1K details

### B.1. s1K summary

### B.2. Dataset composition for full 59K questions

Table 5. 我们的数据集 s1K 概述。使用 Qwen-2.5 分词器测量的 token 数量。我们提示 Claude 根据领域内的几个问题生成关键词。

Domain	#questions	Total token count	Keywords
Geometry	109	560.2K	Area, Triangle, Distance
Number theory	98	522.5K	Sequences, Divisibility
Combinatorics	75	384.7K	Permutations, Counting
Real functions	43	234.8K	Trigonometry, Calculus
Biology	41	120.9K	Organic reactions
Complex functions	32	170.2K	Complex roots
Quantum theory	32	127.9K	Particles, Wave functions
Field theory	28	150.1K	Polynomials, Roots
Calculus of variations	28	155.5K	Optimization, Control
Difference equations	24	132.5K	Recurrence, Recursion
Electromagnetic theory	23	95.8K	Optics, Waves, Diffraction
Group theory	22	100.0K	Groups, Automorphisms
Linear algebra	22	128.3K	Matrices, Determinants
Probability theory	20	114.6K	Random walk, Expectation
Algebraic systems	19	109.9K	Functional equations
Mechanics	19	103.6K	Forces, Motion, Energy
Thermodynamics	19	74.2K	Heat engines, Entropy
Differential equations	18	89.6K	Substitution, Existence
Computer science	18	34.2K	Complexity theory, Algorithms
Numerical analysis	18	76.5K	Error analysis, Stability
Calculus	17	96.3K	Convergence, Summation
Algebraic structures	17	90.4K	Inequalities, Sets
Astronomy	16	37.7K	Stellar populations, Orbits
Remaining 27 domains	242	982.2K	Domains with $\leq 16$ questions
All domains (51)	1000	4.7M	s1K

**Table 5. Summary of our dataset s1K.** Token count measured by the Qwen-2.5 tokenizer. We prompt Claude to produce keywords given several questions from the domain.

Domain	#questions	Total token count	Keywords
Geometry	109	560.2K	Area, Triangle, Distance
Number theory	98	522.5K	Sequences, Divisibility
Combinatorics	75	384.7K	Permutations, Counting
Real functions	43	234.8K	Trigonometry, Calculus
Biology	41	120.9K	Organic reactions
Complex functions	32	170.2K	Complex roots
Quantum theory	32	127.9K	Particles, Wave functions
Field theory	28	150.1K	Polynomials, Roots
Calculus of variations	28	155.5K	Optimization, Control
Difference equations	24	132.5K	Recurrence, Recursion
Electromagnetic theory	23	95.8K	Optics, Waves, Diffraction
Group theory	22	100.0K	Groups, Automorphisms
Linear algebra	22	128.3K	Matrices, Determinants
Probability theory	20	114.6K	Random walk, Expectation
Algebraic systems	19	109.9K	Functional equations
Mechanics	19	103.6K	Forces, Motion, Energy
Thermodynamics	19	74.2K	Heat engines, Entropy
Differential equations	18	89.6K	Substitution, Existence
Computer science	18	34.2K	Complexity theory, Algorithms
Numerical analysis	18	76.5K	Error analysis, Stability
Calculus	17	96.3K	Convergence, Summation
Algebraic structures	17	90.4K	Inequalities, Sets
Astronomy	16	37.7K	Stellar populations, Orbits
Remaining 27 domains	242	982.2K	Domains with $\leq 16$ questions
All domains (51)	1000	4.7M	<b>s1K</b>

**Table 6. 59K 问题的组成。** 思考和回答的长度使用 Qwen2.5-32B-Instruct 分词器 (Qwen et al., 2024) 以 token 为单位进行测量。除了排除我们的评估基准 AIME24 之外，我们还排除了 2022-2023 年的 AIME 问题，因为我们在 **s1-32B** 的开发阶段使用了这 90 个问题。

Source	Description	#Samples	Avg. thinking length
NuminaMATH (Li et al., 2024)	Math problems from online websites	30660	4.1K
MATH (Hendrycks et al., 2021)	Math problems from competitions	11999	2.9K
OlympicArena (Huang et al., 2024a)	Astronomy, Biology, Chemistry, Computer Science, Geography, Math, and Physics olympiad questions	4250	3.2K
OmniMath (Gao et al., 2024a)	Math problems from competitions	4238	4.4K
AGIEval (Zhong et al., 2023; Ling et al., 2017; Hendrycks et al., 2021; Liu et al., 2020; Zhong et al., 2019; Wang et al., 2021)	English, Law, Logic and Math problems from the SAT, LSAT and other exams	2385	1.2K
xword	Crossword puzzles	999	0.7K
OlympiadBench (He et al., 2024)	Math and Physics olympiad questions	896	3.9K
AIME (1983-2021)	American Invitational Mathematics Examination	890	4.7K
TheoremQA (Chen et al., 2023)	Computer Science, Finance, Math, and Physics university-level questions relating to theorems	747	2.1K
USACO (Shi et al., 2024)	Code problems from the USA Computing Olympiad	519	3.6K
JEEBench (Arora et al., 2023)	Chemistry, Math, and Physics problems used in the university entrance examination of the Indian Institute of Technology	515	2.9K
GPQA (Rein et al., 2023)	PhD-Level Science Questions	348	2.9K
SciEval (Sun et al., 2024)	Biology, Chemistry, and Physics problems from various sources	227	0.7K
s1-prob	Stanford statistics qualifying exams	182	4.0K
LiveCodeBench (Jain et al., 2024)	Code problems from coding websites (LeetCode, AtCoder, and CodeForces)	151	3.5K
s1-teasers	Math brain-teasers crawled from the Internet	23	4.1K
<b>All 59K questions</b>	Composite of the above datasets with reasoning traces and solutions	59029	3.6K

**Table 6. Composition of full 59K questions.** Thinking and response lengths are measured in tokens using the Qwen2.5-32B-Instruct tokenizer (Qwen et al., 2024). In addition to excluding our evaluation benchmark, AIME24, we also exclude AIME questions from 2022-2023 as we use these 90 questions during our development stage of s1-32B.

Source	Description	#Samples	Avg. think- ing length
NuminaMATH (LI et al., 2024)	Math problems from online websites	30660	4.1K
MATH (Hendrycks et al., 2021)	Math problems from competitions	11999	2.9K
OlympicArena (Huang et al., 2024a)	Astronomy, Biology, Chemistry, Computer Science, Geography, Math, and Physics olympiad questions	4250	3.2K
OmniMath (Gao et al., 2024a)	Math problems from competitions	4238	4.4K
AGIEval (Zhong et al., 2023; Ling et al., 2017; Hendrycks et al., 2021; Liu et al., 2020; Zhong et al., 2019; Wang et al., 2021)	English, Law, Logic and Math problems from the SAT, LSAT and other exams	2385	1.2K
xword	Crossword puzzles	999	0.7K
OlympiadBench (He et al., 2024)	Math and Physics olympiad questions	896	3.9K
AIME (1983-2021)	American Invitational Mathematics Examination	890	4.7K
TheoremQA (Chen et al., 2023)	Computer Science, Finance, Math, and Physics university-level questions relating to theorems	747	2.1K
USACO (Shi et al., 2024)	Code problems from the USA Computing Olympiad	519	3.6K
JEEBench (Arora et al., 2023)	Chemistry, Math, and Physics problems used in the university entrance examination of the Indian Institute of Technology	515	2.9K
GPQA (Rein et al., 2023)	PhD-Level Science Questions	348	2.9K
SciEval (Sun et al., 2024)	Biology, Chemistry, and Physics problems from various sources	227	0.7K
s1-prob	Stanford statistics qualifying exams	182	4.0K
LiveCodeBench (Jain et al., 2024)	Code problems from coding websites (LeetCode, AtCoder, and CodeForces)	151	3.5K
s1-teasers	Math brain-teasers crawled from the Internet	23	4.1K
<b>All 59K questions</b>	Composite of the above datasets with reasoning traces and solutions	59029	3.6K

### B.3. s1K grading prompt

为了评估某个示例是否适合我们数据集的选择（见§2），我们使用了Figure 8中的提示。

You are an AI assistant for grading a science problem. The user will provide you with the question itself, an attempt made by a student and the correct answer to the problem. Your job is to judge whether the attempt is correct by comparing it with the correct answer. If the expected solution concludes with a number or choice, there should be no ambiguity. If the expected solution involves going through the entire reasoning process, you should judge the attempt based on whether the reasoning process is correct with correct answer if helpful.

The user will provide the attempt and the correct answer in the following format:

```
# Problem
{problem}

## Attempt
{attempt}

## Correct answer
{solution}
```

Explain your reasoning, and end your response on a new line with only "Yes" or "No" (without quotes).

Figure 8. Grading prompt.

### B.4. s1K diversity selection

Algorithm 1 提供了我们在多样性选择阶段选择数据的算法。如 §2 所述，我们还纳入了一些我们认为高质量的特定基准样本。这些样本与我们的最终评估没有重叠。

### B.5. Decontamination

我们通过检查选定示例与评估基准（MATH500、GPTQA Diamond 和 AIME24）之间的8-gram重叠来过滤所有样本。我们排除了超过8-gram重叠的问题。

### B.3. s1K grading prompt

To grade whether an example is correct for our dataset selection in §2, we use the prompt in Figure 8.

---

You are an AI assistant for grading a science problem. The user will provide you with the question itself, an attempt made by a student and the correct answer to the problem. Your job is to judge whether the attempt is correct by comparing it with the correct answer. If the expected solution concludes with a number or choice, there should be no ambiguity. If the expected solution involves going through the entire reasoning process, you should judge the attempt based on whether the reasoning process is correct with correct answer if helpful.

The user will provide the attempt and the correct answer in the following format:

```
# Problem
{problem}

## Attempt
{attempt}

## Correct answer
{solution}
```

Explain your reasoning, and end your response on a new line with only "Yes" or "No" (without quotes).

---

Figure 8. Grading prompt.

### B.4. s1K diversity selection

Algorithm 1 provides our algorithm for selecting data in our diversity selection stage. As mentioned in §2, we also include samples from some specific benchmarks we perceive as high-quality. None of the samples overlap with our final evaluation.

### B.5. Decontamination

We filter all samples by checking for an 8-gram overlap between the selected examples and the evaluation benchmarks: MATH500, GPTQA Diamond, and AIME24. We exclude questions with more than an 8-gram overlap.

---

**Algorithm 1** Two-stage sampling for s1K

```

1: Input:  $\mathcal{Q} :=$  Set of 24,496 questions with features
2: Output:  $\mathcal{S} :=$  Set of 1,000 selected questions
3:  $\mathcal{S} \leftarrow \emptyset$  Initialize the output set (only tracks unique elements)
4: for  $q \in \mathcal{Q}$  do
5:   if IsGeminiCorrect( $q$ ) and (IsAIME( $q$ ) or IsGPQA( $q$ )) then
6:      $\mathcal{S} \leftarrow \mathcal{S} \cup \{q\}$ 
7:   else if IsGeminiCorrect( $q$ ) and IsMATH( $q$ ) and ThinkingLength( $q$ ) > 5600 then
8:      $\mathcal{S} \leftarrow \mathcal{S} \cup \{q\}$ 
9:   end if
10:  end for
11:   $\mathcal{D} \leftarrow$  All available domains
12:  while  $|\mathcal{S}| < 1000$  do Initialize domain pool
13:     $d \leftarrow$  RandomChoice( $\mathcal{D}$ )
14:     $\mathcal{Q}_d \leftarrow$  Questions in domain  $d$  Randomly select a domain
15:     $\text{ranks} \leftarrow$  RankByThinkingLength( $\mathcal{Q}_d$ )
16:     $\text{weights} \leftarrow 2^{-\text{ranks}}$  Get questions from this domain
17:     $q \leftarrow$  WeightedSample( $\mathcal{Q}_d$ , weights) Rank by thinking length
18:     $\mathcal{S} \leftarrow \mathcal{S} \cup \{q\}$  Apply power-law weighting
19:     $\mathcal{Q}_d \leftarrow \mathcal{Q}_d \setminus \{q\}$  Sample favoring longer chains
20:    if  $\mathcal{Q}_d = \emptyset$  then
21:       $\mathcal{D} \leftarrow \mathcal{D} \setminus \{d\}$  Add selected question
22:    end if
23:  end while Remove exhausted domains
```

---

---

**Algorithm 1** Two-stage sampling for s1K

```

1: Input:  $\mathcal{Q}$  := Set of 24,496 questions with features
2: Output:  $\mathcal{S}$  := Set of 1,000 selected questions
3:  $\mathcal{S} \leftarrow \emptyset$  Initialize the output set (only tracks unique elements)
4: for  $q \in \mathcal{Q}$  do
5:   if IsGeminiCorrect( $q$ ) and (IsAIME( $q$ ) or IsGPQA( $q$ )) then
6:      $\mathcal{S} \leftarrow \mathcal{S} \cup \{q\}$  Select all correct AIME/GPQA solutions
7:   else if IsGeminiCorrect( $q$ ) and IsMATH( $q$ ) and ThinkingLength( $q$ ) > 5600 then
8:      $\mathcal{S} \leftarrow \mathcal{S} \cup \{q\}$  Select correct MATH500 solutions with long chains
9:   end if
10:  end for
11:  $\mathcal{D} \leftarrow$  All available domains
12: while  $|\mathcal{S}| < 1000$  do
13:    $d \leftarrow$  RandomChoice( $\mathcal{D}$ ) Initialize domain pool
14:    $Q_d \leftarrow$  Questions in domain  $d$  Randomly select a domain
15:   ranks  $\leftarrow$  RankByThinkingLength( $Q_d$ ) Get questions from this domain
16:   weights  $\leftarrow 2^{-\text{ranks}}$  Rank by thinking length
17:    $q \leftarrow$  WeightedSample( $Q_d$ , weights) Apply power-law weighting
18:    $\mathcal{S} \leftarrow \mathcal{S} \cup \{q\}$  Sample favoring longer chains
19:   if  $Q_d = \emptyset$  then Add selected question
20:      $Q_d \leftarrow Q_d \setminus \{q\}$ 
21:     if  $Q_d = \emptyset$  then Remove exhausted domains
22:        $\mathcal{D} \leftarrow \mathcal{D} \setminus \{d\}$ 
23:     end if
24:   end if
25: end while

```

---

## C. Training details

我们采用一个已经预训练和指令调优的模型，并进一步对其进行推理微调。具体来说，我们使用 Qwen2.5-32B-Instruct (Qwen et al., 2024)，该模型在数学任务上通常与更大的 Qwen2.5-72B-Instruct (Qwen et al., 2024) 或其他开源模型 (Dubey et al., 2024; Groeneveld et al., 2024; Muennighoff et al., 2024) 相当或更优。我们使用标记分隔思考阶段和回答阶段。思考阶段用

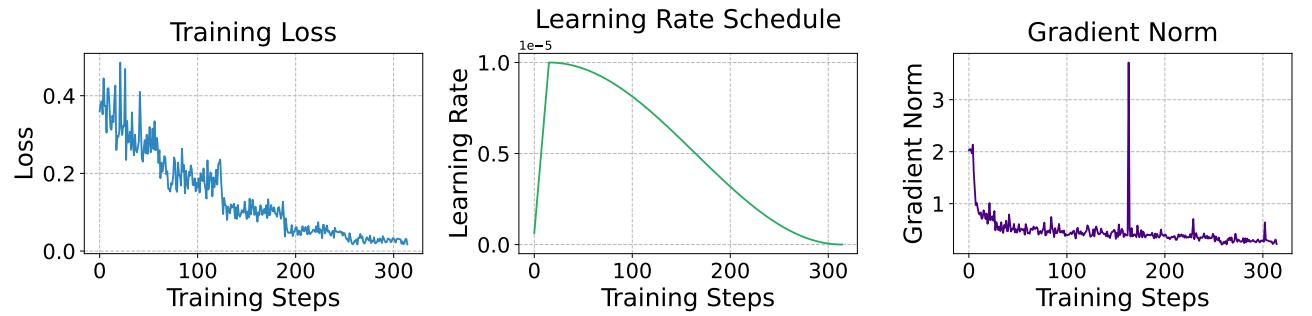


Figure 9. s1-32B 在 s1K 上的训练动态。

### C.1. Training Ablations: Sequence length

Table 7. 训练序列长度消融。我们报告“准确率 / 每样本平均思考标记数”；准确率越高且思考标记数（推理成本）越少越好。

	Model A	Model B
Training sequence length	4096	32768
% training samples cutoff	74%	0%
AIME24	30.0% / 20721	50.0% / 6984
MATH500	90.0% / 5324	91.0% / 3268
GPQA	52.5% / 6841	53.0% / 3568

除了我们在§5.2中的缩放消融实验外，主要的训练超参数消融是我们训练时使用的序列长度。我们发现，较短的训练序列长度会导致测试时的推理轨迹更长。这是因为当使用较短的序列长度进行训练时，训练样本的答案部分更常被截断。相反，当训练序列长度较长时，更多的样本会完整地出现，包括模型作答的部分。因此，模型在学习生成答案的链之后会接收到更多的梯度更新。这进而导致在生成过程中的任何时刻，答案部分的对数概率更高，从而在测试时产生更短的推理轨迹。在性能方面，我们还发现，使用较长序列长度训练的模型表现更好。因此，我们选择最长的训练序列长度，因为它不仅提高了性能，还通过产生更短的推理轨迹使推理更加高效。

### C.2. Training Samples

Table 8, Table 9, Table 10 包含来自 s1K 的训练样本。

### C. Training details

We take a model that has already been pretrained and instruction tuned and further finetune it for reasoning. Specifically, we use Qwen2.5-32B-Instruct (Qwen et al., 2024), which on math tasks generally matches or outperforms the larger Qwen2.5-72B-Instruct (Qwen et al., 2024) or other open models (Dubey et al., 2024; Groeneveld et al., 2024; Muennighoff et al., 2024). We use token delimiters to separate the thinking stage from the answering stage. We enclose the thinking stage with `<|im_start|>`think and `<|im_start|>`answer; both preceded and followed by a newline. Samples from our dataset are in §C.2. We use basic fine-tuning hyperparameters: we train for 5 epochs with a batch size of 16 for a total of 315 gradient steps. We train in bfloat16 precision with a learning rate of  $1e-5$  warmed up linearly for 5% (16 steps) and then decayed to 0 over the rest of training (299 steps) following a cosine schedule. We use the AdamW optimizer (Loshchilov & Hutter, 2019) with  $\beta_1 = 0.9$ ,  $\beta_2 = 0.95$  and weight decay of  $1e-4$ . We do not compute loss on questions, only on reasoning traces and solutions. We ensure the sequence length is large enough to avoid cutting off any samples; a setting we ablate in §C.1. The training takes just 26 minutes on 16 NVIDIA H100 GPUs.

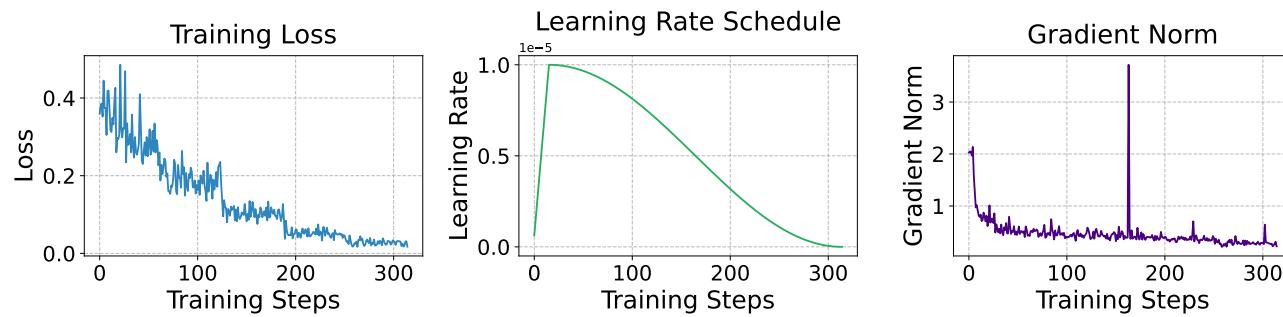


Figure 9. Training dynamics of s1-32B on s1K.

### C.1. Training Ablations: Sequence length

Table 7. Training sequence length ablation. We report “accuracy / average thinking tokens per sample”; the higher the accuracy and the fewer the thinking tokens (inference cost) the better.

	Model A	Model B
Training sequence length	4096	32768
% training samples cutoff	74%	0%
AIME24	30.0% / 20721	50.0% / 6984
MATH500	90.0% / 5324	91.0% / 3268
GPQA	52.5% / 6841	53.0% / 3568

Besides our scaling ablations in §5.2, the main training hyperparameter we ablate is the sequence length used during training. We find that a **shorter** training sequence length leads to **longer** reasoning traces at test time. This is because when training with a shorter sequence length the answer section of the training sample is more commonly cut off. Inversely, when the training sequence length is longer, more samples appear in their entirety with the section where the model answers. Thus the model receives more gradient updates where it learns to generate an answer following its

Table 8. s1K 从 s1-prob 生成的样本。黑色文本是问题，黄色文本是解决方案。仅根据问题，我们生成 浅蓝色 推理轨迹和 蓝色 回答，我们使用这些来训练 Gemini。

给定在同一个概率空间  $(\Omega, \mathcal{F}, \mathbb{P})$  上的可积随机变量  $X, Y_0$  和  $Z_0$ ，以及两个  $\sigma$ -代数  $\mathcal{A} \subset \mathcal{F}, \mathcal{B} \subset \mathcal{F}$ . 对于  $k = 1, 2, \dots$ ，令

$$Y_k := \mathbb{E}[X|\sigma(\mathcal{A}, Z_0, \dots, Z_{k-1})], \quad Z_k := \mathbb{E}[X|\sigma(\mathcal{B}, Y_0, \dots, Y_{k-1})].$$

(a) 证明存在可积随机变量  $Y_\infty$  和  $Z_\infty$ ，使得当  $n \rightarrow \infty$  时， $\mathbb{E}[|Y_n - Y_\infty|] \rightarrow 0$  和  $\mathbb{E}[|Z_n - Z_\infty|] \rightarrow 0$ . (b) Prove that almost surely  $Y_\infty = Z_\infty$ .

1. 引入记号， $\mathcal{F}_n := \sigma(\mathcal{A}, Z_0, \dots, Z_{n-1})$  和  $\mathcal{G}_n := \sigma(\mathcal{B}, Y_0, \dots, Y_{n-1})$ ，对所有  $n \geq 1$ ，并且  $\mathcal{F}_0 := \mathcal{A}$ ,  $\mathcal{G}_0 := \mathcal{B}$ 。显然，两者都是滤波，并且根据定义， $Y_n = \mathbb{E}[X|\mathcal{F}_n]$  和  $Z_n = \mathbb{E}[X|\mathcal{G}_n]$ ，对所有  $n \geq 0$ 。因此  $\{Y_n\}_{n \geq 0}$  (或  $\{Z_n\}_{n \geq 0}$ ) 是关于滤波  $\{\mathcal{F}_n\}_{n \geq 0}$  (或  $\{\mathcal{G}_n\}_{n \geq 0}$ ) 的 Doob's Martingale，因此是U.I.，因为  $X$  是可积的。现在应用一个鞅是U.I.当且仅当它在  $L^1$  中收敛的事实，并得出结论。

2. We provide two different proofs.

1. 观察以下恒等式，它源自 Tower Property。

$$\mathbb{E}[Y_n|\mathcal{G}_n] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_n]|\mathcal{G}_n] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}_n]|\mathcal{F}_n] = \mathbb{E}[Z_n|\mathcal{F}_n]. \quad (4)$$

请注意  $\mathbb{E}[\mathbb{E}[Y_n|\mathcal{G}_n]] - \mathbb{E}[Y_\infty|\mathcal{G}_n] \leq \mathbb{E}|Y_n - Y_\infty| = o(1)$ ，并且根据 Levy's Upward Theorem， $\mathbb{E}[Y_\infty|\mathcal{G}_n] \xrightarrow{L^1} \mathbb{E}[Y_\infty|\mathcal{G}_\infty]$ ，其中  $\mathcal{G}_\infty = \sigma(\cup_{n \geq 0} \mathcal{G}_n) = \sigma(\mathcal{B}, Y_0, Y_1, \dots)$ 。显然， $Y_\infty \in m\mathcal{G}_\infty$  (因为它是  $Y_n \in m\mathcal{G}_n$  的几乎必然极限)，因此  $\mathbb{E}[Y_\infty|\mathcal{G}_\infty] = Y_\infty$  几乎必然。结合上述观察，我们得出结论  $\mathbb{E}[Y_n|\mathcal{G}_n] \xrightarrow{L^1} Y_\infty$ 。类似地， $\mathbb{E}[Z_n|\mathcal{F}_n] \xrightarrow{L^1} Z_\infty$ 。现在引用方程 (4)，得出  $Y_\infty = Z_\infty$  几乎必然。

2. 由于  $X$  是可积的，根据 Levy's Upward Theorem，我们有  $Y_\infty = \lim_{n \rightarrow \infty} Y_n = \lim_{n \rightarrow \infty} \mathbb{E}[X|\mathcal{F}_n] = \mathbb{E}[X|\mathcal{F}_\infty]$ 。同样地， $Z_\infty = \mathbb{E}[X|\mathcal{G}_\infty]$ 。另一方面，在第一次证明中我们观察到  $Y_\infty \in m\mathcal{G}_\infty$ ，因此  $Y_\infty \in m\mathcal{G}_\infty \cap m\mathcal{F}_\infty = m(\mathcal{G}_\infty \cap \mathcal{F}_\infty)$ 。这导致了以下结论。

$$Y_\infty = \mathbb{E}[Y_\infty|\mathcal{G}_\infty \cap \mathcal{F}_\infty] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_\infty]|\mathcal{G}_\infty \cap \mathcal{F}_\infty] = \mathbb{E}[X|\mathcal{G}_\infty \cap \mathcal{F}_\infty].$$

同样地， $Z_\infty = \mathbb{E}[X|\mathcal{G}_\infty \cap \mathcal{F}_\infty]$ 。这证明了  $Y_\infty = Z_\infty$  几乎必然成立。

---

chain. This in turn leads to a higher log probability of the answer section at any point during the generation and thus shorter reasoning traces at test time. Performance-wise, we also find that the model trained with a longer sequence length performs better. Thus we opt for the longest training sequence length as it leads to better performance and makes inference more efficient by leading to shorter reasoning traces.

## C.2. Training Samples

Table 8, Table 9, Table 10 contain training samples from **s1K**.

让我们分析这个问题。我们有两个由条件期望定义的随机变量序列  $Y_k$  和  $Z_k$ 。我们需要证明这些序列在  $L^1$  中收敛，并且它们的极限几乎处处相等。

部分 (a): 在  $L^1$  中的收敛性。

涉及的  $\sigma$ -代数是递增的:

$$\mathcal{F}_k = \sigma(\mathcal{A}, Z_0, \dots, Z_{k-1})$$

$$\mathcal{G}_k = \sigma(B, Y_0, \dots, Y_{k-1})$$

对于  $Y_k$  序列:

$$Y_1 = \mathbb{E}[X | \sigma(\mathcal{A})]$$

$$Y_2 = \mathbb{E}[X | \sigma(\mathcal{A}, Z_0)]$$

$$Y_3 = \mathbb{E}[X | \sigma(\mathcal{A}, Z_0, Z_1)]$$

...

$\sigma$ -代数序列  $\sigma(\mathcal{A}, Z_0, \dots, Z_{k-1})$  是递增的。根据鞅收敛定理（对于条件期望），由于  $X$  是可积的，序列  $Y_k$  几乎处处和在  $L^1$  中收敛到  $\mathbb{E}[X | \sigma(\mathcal{A}, Z_0, Z_1, \dots)]$ 。设  $Y_\infty = \mathbb{E}[X | \sigma(\mathcal{A}, Z_0, Z_1, \dots)]$ 。

Table 8. **s1K sample stemming from s1-prob.** The black text is the question and the yellow text is the solution. Based solely on the question we generate the light blue reasoning trace and blue response with Gemini, which we use for training.

You are given integrable random variables  $X, Y_0$  and  $Z_0$  on the same probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , and two  $\sigma$ -algebras  $\mathcal{A} \subset \mathcal{F}, \mathcal{B} \subset \mathcal{F}$ . For  $k = 1, 2, \dots$ , let

$$Y_k := \mathbb{E}[X|\sigma(\mathcal{A}, Z_0, \dots, Z_{k-1})], \quad Z_k := \mathbb{E}[X|\sigma(\mathcal{B}, Y_0, \dots, Y_{k-1})].$$

- (a) Show that there exist integrable random variables  $Y_\infty$  and  $Z_\infty$  such that as  $n \rightarrow \infty$  both  $\mathbb{E}[|Y_n - Y_\infty|] \rightarrow 0$  and  $\mathbb{E}[|Z_n - Z_\infty|] \rightarrow 0$ . (b) Prove that almost surely  $Y_\infty = Z_\infty$ .

1. Introduce the notations,  $\mathcal{F}_n := \sigma(\mathcal{A}, Z_0, \dots, Z_{n-1})$  and  $\mathcal{G}_n := \sigma(\mathcal{B}, Y_0, \dots, Y_{n-1})$ , for all  $n \geq 1$  and  $\mathcal{F}_0 := \mathcal{A}$ ,  $\mathcal{G}_0 := \mathcal{B}$ . Clearly, both are filtrations and by definition,  $Y_n = \mathbb{E}[X|\mathcal{F}_n]$  and  $Z_n = \mathbb{E}[X|\mathcal{G}_n]$ , for all  $n \geq 0$ . Thus  $\{Y_n\}_{n \geq 0}$  (or  $\{Z_n\}_{n \geq 0}$ ) is a *Doob's Martingale* with respect to filtration  $\{\mathcal{F}_n\}_{n \geq 0}$  (or  $\{\mathcal{G}_n\}_{n \geq 0}$ ) and hence is U.I.,  $X$  being integrable. Now apply the fact that a Martingale is U.I. if and only if it converges in  $L^1$  and conclude.

2. We provide two different proofs.

1. Observe the following identity which follows from *Tower Property*.

$$\mathbb{E}[Y_n|\mathcal{G}_n] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_n]|\mathcal{G}_n] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}_n]|\mathcal{F}_n] = \mathbb{E}[Z_n|\mathcal{F}_n]. \quad (4)$$

Note that  $\mathbb{E}|\mathbb{E}[Y_n|\mathcal{G}_n] - \mathbb{E}[Y_\infty|\mathcal{G}_n]| \leq \mathbb{E}|Y_n - Y_\infty| = o(1)$  and by *Levy's Upward Theorem*,  $\mathbb{E}[Y_\infty|\mathcal{G}_n] \xrightarrow{L^1} \mathbb{E}[Y_\infty|\mathcal{G}_\infty]$ , where  $\mathcal{G}_\infty = \sigma(\cup_{n \geq 0} \mathcal{G}_n) = \sigma(\mathcal{B}, Y_0, Y_1, \dots)$ . It is evident that  $Y_\infty \in m\mathcal{G}_\infty$  (since it is almost sure limit of  $Y_n \in m\mathcal{G}_n$ ) and hence  $\mathbb{E}[Y_\infty|\mathcal{G}_\infty] = Y_\infty$  almost surely. Combining the above observations we conclude that  $\mathbb{E}[Y_n|\mathcal{G}_n] \xrightarrow{L^1} Y_\infty$ . Similarly,  $\mathbb{E}[Z_n|\mathcal{F}_n] \xrightarrow{L^1} Z_\infty$ . Now invoke Equation (4), to conclude that  $Y_\infty = Z_\infty$  almost surely.

2. Since  $X$  is integrable, we have by *Levy's Upward Theorem*,  $Y_\infty = \lim_{n \rightarrow \infty} Y_n = \lim_{n \rightarrow \infty} \mathbb{E}[X|\mathcal{F}_n] = \mathbb{E}[X|\mathcal{F}_\infty]$ .

Similarly,  $Z_\infty = \mathbb{E}[X|\mathcal{G}_\infty]$ . On the otherhand, in first proof we have observed that  $Y_\infty \in m\mathcal{G}_\infty$  and hence  $Y_\infty \in m\mathcal{G}_\infty \cap m\mathcal{F}_\infty = m(\mathcal{G}_\infty \cap \mathcal{F}_\infty)$ . This yields the following.

$$Y_\infty = \mathbb{E}[Y_\infty|\mathcal{G}_\infty \cap \mathcal{F}_\infty] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_\infty]|\mathcal{G}_\infty \cap \mathcal{F}_\infty] = \mathbb{E}[X|\mathcal{G}_\infty \cap \mathcal{F}_\infty].$$

Similarly,  $Z_\infty = \mathbb{E}[X|\mathcal{G}_\infty \cap \mathcal{F}_\infty]$ . This proves that  $Y_\infty = Z_\infty$  almost surely.

同样地, 对于  $Z_k$  序列:

$$Z_1 = \mathbb{E}[X|\sigma(\mathcal{B})]$$

$$Z_2 = \mathbb{E}[X|\sigma(\mathcal{B}, Y_0)]$$

$$Z_3 = \mathbb{E}[X|\sigma(\mathcal{B}, Y_0, Y_1)]$$

...

$\sigma$ -代数序列  $\sigma(\mathcal{B}, Y_0, \dots, Y_{k-1})$  是递增的。根据鞅收敛定理, 序列  $Z_k$  几乎处处和在  $L^1$  中收敛到  $\mathbb{E}[X|\sigma(\mathcal{B}, Y_0, Y_1, \dots)]$ 。设  $Z_\infty = \mathbb{E}[X|\sigma(\mathcal{B}, Y_0, Y_1, \dots)]$ 。

部分 (b): 证明几乎处处  $Y_\infty = Z_\infty$ 。

我们有  $Y_\infty = \mathbb{E}[X|\sigma(\mathcal{A}, Z_0, Z_1, \dots)]$  和  $Z_\infty = \mathbb{E}[X|\sigma(\mathcal{B}, Y_0, Y_1, \dots)]$ 。

设  $\mathcal{Y}_\infty = \sigma(\mathcal{A}, Z_0, Z_1, \dots)$  和  $\mathcal{Z}_\infty = \sigma(\mathcal{B}, Y_0, Y_1, \dots)$ 。然后  $Y_\infty = \mathbb{E}[X|\mathcal{Y}_\infty]$  和  $Z_\infty = \mathbb{E}[X|\mathcal{Z}_\infty]$ 。

考虑  $Y_1 = \mathbb{E}[X|\mathcal{A}]$ 。  $Z_1 = \mathbb{E}[X|\mathcal{B}]$ 。

$Y_2 = \mathbb{E}[X|\sigma(\mathcal{A}, Z_0)]$ 。由于  $Z_0 = \mathbb{E}[X|\mathcal{B}]$ ,  $\sigma(\mathcal{A}, Z_0) = \sigma(\mathcal{A}, \mathbb{E}[X|\mathcal{B}])$ 。

$Z_2 = \mathbb{E}[X|\sigma(\mathcal{B}, Y_0)]$ 。由于  $Y_0$  未定义, 假设  $Y_0$  和  $Z_0$  是给定的。

让我们考虑由所有  $Y_i$  和  $Z_i$  生成的  $\sigma$ -代数。

令  $\mathcal{G} = \sigma(Y_0, Z_0, Y_1, Z_1, \dots)$ 。

$Y_k$  和  $Z_k$  都相对于  $\mathcal{G}$  可测。因此,  $Y_\infty$  和  $Z_\infty$  也相对于  $\mathcal{G}$  可测。

考虑  $\mathbb{E}[Y_{n+1}|\mathcal{A}, Z_0, \dots, Z_{n-1}] = \mathbb{E}[\mathbb{E}[X|\sigma(\mathcal{A}, Z_0, \dots, Z_n)]|\mathcal{A}, Z_0, \dots, Z_{n-1}] = Y_n$ 。

考虑  $\mathbb{E}[Z_{n+1}|\mathcal{B}, Y_0, \dots, Y_{n-1}] = \mathbb{E}[\mathbb{E}[X|\sigma(\mathcal{B}, Y_0, \dots, Y_n)]|\mathcal{B}, Y_0, \dots, Y_{n-1}] = Z_n$ 。

令  $\mathcal{H} = \sigma(\mathcal{A} \cup \mathcal{B} \cup \sigma(Y_0, Z_0, Y_1, Z_1, \dots))$ 。

考虑  $\mathbb{E}[X|\mathcal{A}, Z_0, \dots, Z_{k-1}] = Y_k$ 。

考虑  $\mathbb{E}[X|\mathcal{B}, Y_0, \dots, Y_{k-1}] = Z_k$ 。

让我们考虑极限  $\sigma$ -代数  $\mathcal{Y}_\infty = \sigma(\mathcal{A}, Z_0, Z_1, \dots)$  和  $\mathcal{Z}_\infty = \sigma(\mathcal{B}, Y_0, Y_1, \dots)$ 。

我们知道  $Y_\infty$  是  $\mathcal{Y}_\infty$ -可测的,  $Z_\infty$  是  $\mathcal{Z}_\infty$ -可测的。

令  $\mathcal{C} = \sigma(\mathcal{A} \cup \mathcal{B})$ 。

$Y_1 = \mathbb{E}[X|\mathcal{A}]$ 。  $Z_1 = \mathbb{E}[X|\mathcal{B}]$ 。

$Y_2 = \mathbb{E}[X|\sigma(\mathcal{A}, Z_0)] = \mathbb{E}[X|\sigma(\mathcal{A}, \mathbb{E}[X|\mathcal{B}])]$ 。

$Z_2 = \mathbb{E}[X|\sigma(\mathcal{B}, Y_0)] = \mathbb{E}[X|\sigma(\mathcal{B}, \mathbb{E}[X|\mathcal{A}])]$ 。

考虑当  $\mathcal{A}$  和  $\mathcal{B}$  独立时的情况。那么  $Y_1 = \mathbb{E}[X|\mathcal{A}]$  和  $Z_1 = \mathbb{E}[X|\mathcal{B}]$ 。  $Y_2 = \mathbb{E}[X|\sigma(\mathcal{A}, \mathbb{E}[X|\mathcal{B}])]$ 。

让我们考虑由所有信息生成的  $\sigma$ -代数。设  $\mathcal{F}_\infty = \sigma(\mathcal{A}, \mathcal{B}, Y_0, Z_0, Y_1, Z_1, \dots)$ 。

那么  $Y_\infty$  和  $Z_\infty$  都是  $\mathcal{F}_\infty$ -可测的。

我们有  $Y_\infty = \lim_{n \rightarrow \infty} Y_n$  和  $Z_\infty = \lim_{n \rightarrow \infty} Z_n$  几乎处处成立且在  $L^1$  中成立。

考虑  $\mathbb{E}[X|\mathcal{A}, Z_0, Z_1, \dots] = Y_\infty$  和  $\mathbb{E}[X|\mathcal{B}, Y_0, Y_1, \dots] = Z_\infty$ 。

设  $\mathcal{H}_\infty = \sigma(\mathcal{A} \cup \mathcal{B})$ 。

考虑  $\mathbb{E}[X|\mathcal{H}_\infty]$ 。

让我们尝试证明  $Y_\infty$  和  $Z_\infty$  都等于  $\mathbb{E}[X|\sigma(\mathcal{A} \cup \mathcal{B})]$ 。这似乎太简单了。

让我们考虑条件期望的定义。对于任何集合  $A \in \mathcal{Y}_\infty$ , 有  $\mathbb{E}[Y_\infty 1_A] = \mathbb{E}[X 1_A]$ 。

对于任何集合  $B \in \mathcal{Z}_\infty$ , 有  $\mathbb{E}[Z_\infty 1_B] = \mathbb{E}[X 1_B]$ 。

考虑  $\mathbb{E}[Y_{k+1}|\mathcal{F}_k] = Y_k$ , 其中  $\mathcal{F}_k = \sigma(\mathcal{A}, Z_0, \dots, Z_{k-1})$ 。

考虑  $\mathbb{E}[Z_{k+1}|\mathcal{G}_k] = Z_k$ , 其中  $\mathcal{G}_k = \sigma(\mathcal{B}, Y_0, \dots, Y_{k-1})$ 。

考虑  $\sigma$ -代数  $\sigma(\mathcal{A} \cup \mathcal{B})$ 。

设  $W = \mathbb{E}[X|\sigma(\mathcal{A} \cup \mathcal{B})]$ 。

我们有  $Y_\infty = \mathbb{E}[X|\sigma(\mathcal{A}, Z_0, Z_1, \dots)]$ 。

由于  $Z_k$  依赖于  $Y_{k-1}$ , 而  $Y_{k-1}$  依赖于  $Z_{k-2}$  等等, 因此这是复杂的。

让我们考虑  $\mathbb{E}[X|\sigma(\mathcal{A} \cup \mathcal{B})]$  的情况。

Let's analyze the problem. We are given two sequences of random variables  $Y_k$  and  $Z_k$  defined by conditional expectations. We need to show that these sequences converge in  $L^1$  and that their limits are equal almost surely.

Part (a): Convergence in  $L^1$ .

The sigma-algebras involved are increasing:

$$\mathcal{F}_k = \sigma(\mathcal{A}, Z_0, \dots, Z_{k-1})$$

$$\mathcal{G}_k = \sigma(\mathcal{B}, Y_0, \dots, Y_{k-1})$$

For the  $Y_k$  sequence:

$$Y_1 = \mathbb{E}[X|\sigma(\mathcal{A})]$$

$$Y_2 = \mathbb{E}[X|\sigma(\mathcal{A}, Z_0)]$$

$$Y_3 = \mathbb{E}[X|\sigma(\mathcal{A}, Z_0, Z_1)]$$

...

The sequence of sigma-algebras  $\sigma(\mathcal{A}, Z_0, \dots, Z_{k-1})$  is increasing. By the Martingale Convergence Theorem (for conditional expectations), since  $X$  is integrable, the sequence  $Y_k$  converges almost surely and in  $L^1$  to  $\mathbb{E}[X|\sigma(\mathcal{A}, Z_0, Z_1, \dots)]$ .

Let  $Y_\infty = \mathbb{E}[X|\sigma(\mathcal{A}, Z_0, Z_1, \dots)]$ .

考虑  $\mathbb{E}[X|\mathcal{A}, Z_0, Z_1, \dots] = Y_\infty$ 。这意味着对于任何  $A \in \sigma(\mathcal{A}, Z_0, Z_1, \dots)$ , 有  $\mathbb{E}[X1_A] = \mathbb{E}[Y_\infty 1_A]$ 。

考虑  $\mathbb{E}[X|\mathcal{B}, Y_0, Y_1, \dots] = Z_\infty$ 。这意味着对于任何  $B \in \sigma(\mathcal{B}, Y_0, Y_1, \dots)$ , 有  $\mathbb{E}[X1_B] = \mathbb{E}[Z_\infty 1_B]$ 。

令  $\mathcal{H} = \sigma(\mathcal{A}, \mathcal{B})$ 。

考虑  $\mathbb{E}[X|\mathcal{H}]$ 。

考虑  $\mathbb{E}[Y_\infty|\mathcal{B}, Y_0, Y_1, \dots] = \mathbb{E}[\mathbb{E}[X|\mathcal{A}, Z_0, Z_1, \dots]]|\mathcal{B}, Y_0, Y_1, \dots]$ 。

利用  $Y_\infty$  是  $\sigma(\mathcal{A}, Z_0, Z_1, \dots)$ -可测的, 而  $Z_\infty$  是  $\sigma(\mathcal{B}, Y_0, Y_1, \dots)$ -可测的这一事实。

考虑包含  $\sigma(\mathcal{A}, Z_0, Z_1, \dots)$  和  $\sigma(\mathcal{B}, Y_0, Y_1, \dots)$  的最小  $\sigma$ -代数。

这是  $\sigma(\mathcal{A} \cup \mathcal{B} \cup \{Y_i\}_{i \geq 0} \cup \{Z_i\}_{i \geq 0})$ 。

让我们尝试证明  $\mathbb{E}[(Y_\infty - Z_\infty)^2] = 0$ 。

$$\mathbb{E}[(Y_\infty - Z_\infty)^2] = \mathbb{E}[Y_\infty^2] + \mathbb{E}[Z_\infty^2] - 2\mathbb{E}[Y_\infty Z_\infty].$$

考虑  $Y_k$  和  $Z_k$  的定义。

$$\mathbb{E}[Y_{k+1}|\mathcal{A}, Z_0, \dots, Z_{k-1}] = Y_k. \quad \mathbb{E}[Z_{k+1}|\mathcal{B}, Y_0, \dots, Y_{k-1}] = Z_k.$$

考虑  $\mathbb{E}[Y_\infty|\mathcal{B}, Y_0, Y_1, \dots]$ 。

由于  $Y_\infty$  关于  $\sigma(\mathcal{A}, Z_0, Z_1, \dots)$  可测, 这并不容易简化。

假设  $Y_\infty \neq Z_\infty$  在一个正测度的集合上。

考虑  $\mathcal{A}$  和  $\mathcal{B}$  是平凡的情况。

则  $Y_k = \mathbb{E}[X|\sigma(Z_0, \dots, Z_{k-1})]$  和  $Z_k = \mathbb{E}[X|\sigma(Y_0, \dots, Y_{k-1})]$ 。

令  $\mathcal{H}_\infty = \sigma(\mathcal{A} \cup \mathcal{B} \cup \{Y_i, Z_i\}_{i \geq 0})$ 。

考虑  $\mathbb{E}[X|\mathcal{H}_\infty]$ 。

我们有  $Y_\infty = \mathbb{E}[X|\sigma(\mathcal{A}, Z_0, Z_1, \dots)]$ 。

根据条件期望的性质,  $\mathbb{E}[Y_\infty|\mathcal{A}] = \mathbb{E}[\mathbb{E}[X|\mathcal{A}, Z_0, Z_1, \dots]]|\mathcal{A}]$ 。

考虑由所有涉及的随机变量生成的sigma代数  $\mathcal{D}$ 。

$\mathcal{D} = \sigma(\mathcal{A} \cup \mathcal{B} \cup \{Y_i\}_{i \geq 0} \cup \{Z_i\}_{i \geq 0})$ 。

令  $M_\infty = \mathbb{E}[X|\mathcal{A} \vee \mathcal{B}]$ 。

考虑极限。当  $n \rightarrow \infty$  时,  $\sigma(\mathcal{A}, Z_0, \dots, Z_{n-1}) \uparrow \sigma(\mathcal{A}, Z_0, Z_1, \dots)$ 。

当  $n \rightarrow \infty$  时,  $\sigma(\mathcal{B}, Y_0, \dots, Y_{n-1}) \uparrow \sigma(\mathcal{B}, Y_0, Y_1, \dots)$ 。

令  $\mathcal{Y}_\infty = \sigma(\mathcal{A}, Z_0, Z_1, \dots)$  和  $\mathcal{Z}_\infty = \sigma(\mathcal{B}, Y_0, Y_1, \dots)$ 。

$$\begin{aligned} \text{我们有 } Y_\infty &= \mathbb{E}[X|\mathcal{Y}_\infty] \text{ 和 } Z_\infty = \mathbb{E}[X|\mathcal{Z}_\infty]. \text{ 考虑 } \mathbb{E}[Y_\infty - Z_\infty]^2 = \mathbb{E}[(Y_\infty - X) + (X - Z_\infty)]^2 \\ &= \mathbb{E}[(Y_\infty - X)^2] + \mathbb{E}[(X - Z_\infty)^2] + 2\mathbb{E}[(Y_\infty - X)(X - Z_\infty)] \\ &= \mathbb{E}[(X - \mathbb{E}[X|\mathcal{Y}_\infty])^2] + \mathbb{E}[(X - \mathbb{E}[X|\mathcal{Z}_\infty])^2] + 2\mathbb{E}[Y_\infty X - Y_\infty Z_\infty - X^2 + XZ_\infty] \\ &= \mathbb{E}[X^2] - \mathbb{E}[Y_\infty^2] + \mathbb{E}[X^2] - \mathbb{E}[Z_\infty^2] + 2\mathbb{E}[Y_\infty X] - 2\mathbb{E}[Y_\infty Z_\infty] - 2\mathbb{E}[X^2] + 2\mathbb{E}[XZ_\infty] \\ &= -\mathbb{E}[Y_\infty^2] - \mathbb{E}[Z_\infty^2] + 2\mathbb{E}[Y_\infty^2] - 2\mathbb{E}[Y_\infty Z_\infty] + 2\mathbb{E}[Z_\infty^2] \\ &= \mathbb{E}[Y_\infty^2] + \mathbb{E}[Z_\infty^2] - 2\mathbb{E}[Y_\infty Z_\infty] \end{aligned}$$

考虑sigma代数  $\mathcal{C} = \sigma(\mathcal{A} \cup \mathcal{B})$ 。

令  $W = \mathbb{E}[X|\mathcal{C}]$ 。

考虑由所有信息生成的极限sigma代数。

令  $\mathcal{F}_\infty = \sigma(\mathcal{A}, \mathcal{B}, Y_0, Z_0, Y_1, Z_1, \dots)$ 。

则  $\mathcal{Y}_\infty \subseteq \mathcal{F}_\infty$  且  $\mathcal{Z}_\infty \subseteq \mathcal{F}_\infty$ 。

我们有  $\mathbb{E}[X|\mathcal{Y}_\infty] = Y_\infty$  和  $\mathbb{E}[X|\mathcal{Z}_\infty] = Z_\infty$ 。

考虑  $\mathbb{E}[Y_\infty|\mathcal{Z}_\infty]$  和  $\mathbb{E}[Z_\infty|\mathcal{Y}_\infty]$ 。

令  $\mathcal{H} = \sigma(\mathcal{A} \cup \mathcal{B})$ 。

考虑  $\mathbb{E}[X|\mathcal{H}]$ 。

考虑序列对  $(Y_k, Z_k)$ 。

Similarly, for the  $Z_k$  sequence:

$$Z_1 = \mathbb{E}[X|\sigma(\mathcal{B})]$$

$$Z_2 = \mathbb{E}[X|\sigma(\mathcal{B}, Y_0)]$$

$$Z_3 = \mathbb{E}[X|\sigma(\mathcal{B}, Y_0, Y_1)]$$

...

The sequence of sigma-algebras  $\sigma(\mathcal{B}, Y_0, \dots, Y_{k-1})$  is increasing. By the Martingale Convergence Theorem, the sequence  $Z_k$  converges almost surely and in  $L^1$  to  $\mathbb{E}[X|\sigma(\mathcal{B}, Y_0, Y_1, \dots)]$ . Let  $Z_\infty = \mathbb{E}[X|\sigma(\mathcal{B}, Y_0, Y_1, \dots)]$ .

Part (b): Prove that almost surely  $Y_\infty = Z_\infty$ .

We have  $Y_\infty = \mathbb{E}[X|\sigma(\mathcal{A}, Z_0, Z_1, \dots)]$  and  $Z_\infty = \mathbb{E}[X|\sigma(\mathcal{B}, Y_0, Y_1, \dots)]$ .

Let  $\mathcal{Y}_\infty = \sigma(\mathcal{A}, Z_0, Z_1, \dots)$  and  $\mathcal{Z}_\infty = \sigma(\mathcal{B}, Y_0, Y_1, \dots)$ .

Then  $Y_\infty = \mathbb{E}[X|\mathcal{Y}_\infty]$  and  $Z_\infty = \mathbb{E}[X|\mathcal{Z}_\infty]$ .

Consider  $Y_1 = \mathbb{E}[X|\mathcal{A}]$ .  $Z_1 = \mathbb{E}[X|\mathcal{B}]$ .

$Y_2 = \mathbb{E}[X|\sigma(\mathcal{A}, Z_0)]$ . Since  $Z_0 = \mathbb{E}[X|\mathcal{B}]$ ,  $\sigma(\mathcal{A}, Z_0) = \sigma(\mathcal{A}, \mathbb{E}[X|\mathcal{B}])$ .

$Z_2 = \mathbb{E}[X|\sigma(\mathcal{B}, Y_0)]$ . Since  $Y_0$  is not defined, let's assume  $Y_0$  and  $Z_0$  are given.

Let's consider the sigma-algebra generated by all the  $Y_i$  and  $Z_i$ .

Let  $\mathcal{G} = \sigma(Y_0, Z_0, Y_1, Z_1, \dots)$ .

Both  $Y_k$  and  $Z_k$  are measurable with respect to  $\mathcal{G}$ . Thus,  $Y_\infty$  and  $Z_\infty$  are measurable with respect to  $\mathcal{G}$ .

Consider  $\mathbb{E}[Y_{n+1}|\mathcal{A}, Z_0, \dots, Z_{n-1}] = \mathbb{E}[\mathbb{E}[X|\sigma(\mathcal{A}, Z_0, \dots, Z_n)]|\mathcal{A}, Z_0, \dots, Z_{n-1}] = Y_n$ .

Consider  $\mathbb{E}[Z_{n+1}|\mathcal{B}, Y_0, \dots, Y_{n-1}] = \mathbb{E}[\mathbb{E}[X|\sigma(\mathcal{B}, Y_0, \dots, Y_n)]|\mathcal{B}, Y_0, \dots, Y_{n-1}] = Z_n$ .

Let  $\mathcal{H} = \sigma(\mathcal{A} \cup \mathcal{B} \cup \sigma(Y_0, Z_0, Y_1, Z_1, \dots))$ .

Consider  $\mathbb{E}[X|\mathcal{A}, Z_0, \dots, Z_{k-1}] = Y_k$ .

Consider  $\mathbb{E}[X|\mathcal{B}, Y_0, \dots, Y_{k-1}] = Z_k$ .

Let's consider the limit sigma-algebras  $\mathcal{Y}_\infty = \sigma(\mathcal{A}, Z_0, Z_1, \dots)$  and  $\mathcal{Z}_\infty = \sigma(\mathcal{B}, Y_0, Y_1, \dots)$ .

We know that  $Y_\infty$  is  $\mathcal{Y}_\infty$ -measurable and  $Z_\infty$  is  $\mathcal{Z}_\infty$ -measurable.

Let  $\mathcal{C} = \sigma(\mathcal{A} \cup \mathcal{B})$ .

$Y_1 = \mathbb{E}[X|\mathcal{A}]$ .  $Z_1 = \mathbb{E}[X|\mathcal{B}]$ .

$Y_2 = \mathbb{E}[X|\sigma(\mathcal{A}, Z_0)] = \mathbb{E}[X|\sigma(\mathcal{A}, \mathbb{E}[X|\mathcal{B}])]$ .

$Z_2 = \mathbb{E}[X|\sigma(\mathcal{B}, Y_0)] = \mathbb{E}[X|\sigma(\mathcal{B}, \mathbb{E}[X|\mathcal{A}])]$ .

Consider the case when  $\mathcal{A}$  and  $\mathcal{B}$  are independent. Then  $Y_1 = \mathbb{E}[X|\mathcal{A}]$  and  $Z_1 = \mathbb{E}[X|\mathcal{B}]$ .

$Y_2 = \mathbb{E}[X|\sigma(\mathcal{A}, \mathbb{E}[X|\mathcal{B}])]$ .

Let's consider the sigma-algebra generated by all the information. Let  $\mathcal{F}_\infty = \sigma(\mathcal{A}, \mathcal{B}, Y_0, Z_0, Y_1, Z_1, \dots)$ .

Then both  $Y_\infty$  and  $Z_\infty$  are  $\mathcal{F}_\infty$ -measurable.

We have  $Y_\infty = \lim_{n \rightarrow \infty} Y_n$  and  $Z_\infty = \lim_{n \rightarrow \infty} Z_n$  a.s. and in  $L^1$ .

Consider  $\mathbb{E}[X|\mathcal{A}, Z_0, Z_1, \dots] = Y_\infty$  and  $\mathbb{E}[X|\mathcal{B}, Y_0, Y_1, \dots] = Z_\infty$ .

Let  $\mathcal{H}_\infty = \sigma(\mathcal{A} \cup \mathcal{B})$ .

Consider  $\mathbb{E}[X|\mathcal{H}_\infty]$ .

Let's try to show that both  $Y_\infty$  and  $Z_\infty$  are equal to  $\mathbb{E}[X|\sigma(\mathcal{A} \cup \mathcal{B})]$ . This seems too simple.

Let's consider the definition of conditional expectation. For any set  $A \in \mathcal{Y}_\infty$ ,  $\mathbb{E}[Y_\infty 1_A] = \mathbb{E}[X 1_A]$ .

For any set  $B \in \mathcal{Z}_\infty$ ,  $\mathbb{E}[Z_\infty 1_B] = \mathbb{E}[X 1_B]$ .

Consider  $\mathbb{E}[Y_{k+1}|\mathcal{F}_k] = Y_k$  where  $\mathcal{F}_k = \sigma(\mathcal{A}, Z_0, \dots, Z_{k-1})$ .

Consider  $\mathbb{E}[Z_{k+1}|\mathcal{G}_k] = Z_k$  where  $\mathcal{G}_k = \sigma(\mathcal{B}, Y_0, \dots, Y_{k-1})$ .

Consider the sigma-algebra  $\sigma(\mathcal{A} \cup \mathcal{B})$ .

Let  $W = \mathbb{E}[X|\sigma(\mathcal{A} \cup \mathcal{B})]$ .

考虑极限sigma代数  $\mathcal{L} = \sigma(\mathcal{A} \cup \mathcal{B} \cup \sigma(Y_0, Z_0, \dots))$ 。

我们有  $Y_\infty = \mathbb{E}[X|\sigma(\mathcal{A}, Z_0, Z_1, \dots)]$  和  $Z_\infty = \mathbb{E}[X|\sigma(\mathcal{B}, Y_0, Y_1, \dots)]$ 。

设  $\mathcal{M} = \sigma(\mathcal{A}, Z_0, Z_1, \dots) \vee \sigma(\mathcal{B}, Y_0, Y_1, \dots)$ 。

考虑  $\mathbb{E}[X|\mathcal{M}]$ 。

考虑随机变量  $X$ 。则  $Y_\infty = \lim_n \mathbb{E}[X|\mathcal{F}_n]$  和  $Z_\infty = \lim_n \mathbb{E}[X|\mathcal{G}_n]$ 。

设  $\mathcal{H}_\infty = \sigma(\mathcal{A} \cup \mathcal{B} \cup \sigma(Y_0, Z_0, \dots))$ 。

考虑  $E = \{\omega : Y_\infty(\omega) \neq Z_\infty(\omega)\}$ 。假设  $P(E) > 0$ 。

考虑关于sigma代数并的条件期望。

设  $\mathcal{C}_\infty = \sigma(\mathcal{A}, \mathcal{B}, Y_0, Z_0, \dots)$ 。考虑  $\mathbb{E}[X|\mathcal{C}_\infty]$ 。

那么  $\mathbb{E}[\mathbb{E}[X|\mathcal{C}_\infty]]|\mathcal{A}, Z_0, \dots] = \mathbb{E}[X|\mathcal{A}, Z_0, \dots] = Y_{n+1}$ 。

并且  $\mathbb{E}[\mathbb{E}[X|\mathcal{C}_\infty]]|\mathcal{B}, Y_0, \dots] = \mathbb{E}[X|\mathcal{B}, Y_0, \dots] = Z_{n+1}$ 。

所以  $Y_\infty = \mathbb{E}[X|\sigma(\mathcal{A}, Z_0, Z_1, \dots)]$  和  $Z_\infty = \mathbb{E}[X|\sigma(\mathcal{B}, Y_0, Y_1, \dots)]$ 。

设  $\mathcal{F}_\infty = \sigma(\mathcal{A}, Z_0, Z_1, \dots)$  和  $\mathcal{G}_\infty = \sigma(\mathcal{B}, Y_0, Y_1, \dots)$ 。

考虑  $\mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty]$ 。

那么  $\mathbb{E}[\mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty]]|\mathcal{F}_\infty] = \mathbb{E}[X|\mathcal{F}_\infty] = Y_\infty$ 。

并且  $\mathbb{E}[\mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty]]|\mathcal{G}_\infty] = \mathbb{E}[X|\mathcal{G}_\infty] = Z_\infty$ 。

这意味着  $Y_\infty = \mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty]$  几乎处处成立，且  $Z_\infty = \mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty]$  几乎处处成立。

因此， $Y_\infty = Z_\infty$  几乎处处成立。

解：

(a) 对于序列  $Y_k$ ,  $\sigma$ -代数  $\mathcal{F}_k = \sigma(\mathcal{A}, Z_0, \dots, Z_{k-1})$  随着  $k$  的增加而增加。由于  $X$  是可积的, 根据条件期望的鞅收敛定理, 序列  $Y_k = \mathbb{E}[X|\mathcal{F}_k]$  几乎处处和在  $L^1$  中收敛到  $Y_\infty = \mathbb{E}[X|\sigma(\mathcal{A}, Z_0, Z_1, \dots)]$ 。同样, 对于序列  $Z_k$ ,  $\sigma$ -代数  $\mathcal{G}_k = \sigma(\mathcal{B}, Y_0, \dots, Y_{k-1})$  随着  $k$  的增加而增加。由于  $X$  是可积的, 根据鞅收敛定理, 序列  $Z_k = \mathbb{E}[X|\mathcal{G}_k]$  几乎处处和在  $L^1$  中收敛到  $Z_\infty = \mathbb{E}[X|\sigma(\mathcal{B}, Y_0, Y_1, \dots)]$ 。(b) 令  $\mathcal{F}_\infty = \sigma(\mathcal{A}, Z_0, Z_1, \dots)$  和  $\mathcal{G}_\infty = \sigma(\mathcal{B}, Y_0, Y_1, \dots)$ 。我们有  $Y_\infty = \mathbb{E}[X|\mathcal{F}_\infty]$  和  $Z_\infty = \mathbb{E}[X|\mathcal{G}_\infty]$ 。考虑  $\sigma$ -代数  $\mathcal{H} = \mathcal{F}_\infty \vee \mathcal{G}_\infty = \sigma(\mathcal{A}, \mathcal{B}, Y_0, Z_0, Y_1, Z_1, \dots)$ 。考虑  $W = \mathbb{E}[X|\mathcal{H}]$ 。

我们有  $\mathcal{F}_\infty \subseteq \mathcal{H}$ , 所以  $\mathbb{E}[W|\mathcal{F}_\infty] = \mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{F}_\infty] = \mathbb{E}[X|\mathcal{F}_\infty] = Y_\infty$ 。

由于  $W$  是  $\mathcal{H}$ -可测的,  $\mathbb{E}[W|\mathcal{F}_\infty] = Y_\infty$  意味着  $Y_\infty$  是  $W$  关于  $\mathcal{F}_\infty$  的条件期望。但由于  $Y_\infty$  已经是  $\mathcal{F}_\infty$ -可测的, 我们有  $\mathbb{E}[W|\mathcal{F}_\infty] = W$ , 如果  $W$  是  $\mathcal{F}_\infty$ -可测的, 这不一定是真的。

然而, 我们知道  $\mathbb{E}[Y_\infty|\mathcal{F}_\infty] = Y_\infty$  和  $\mathbb{E}[Z_\infty|\mathcal{G}_\infty] = Z_\infty$ 。

考虑  $\mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty]$ 。

那么  $\mathbb{E}[\mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty]]|\mathcal{F}_\infty] = \mathbb{E}[X|\mathcal{F}_\infty] = Y_\infty$ 。

这表明  $\mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty] = Y_\infty$  几乎处处成立。

同样地,  $\mathbb{E}[\mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty]]|\mathcal{G}_\infty] = \mathbb{E}[X|\mathcal{G}_\infty] = Z_\infty$ 。

这表明  $\mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty] = Z_\infty$  几乎处处成立。

因此,  $Y_\infty = Z_\infty$  几乎处处成立。

最终答案: 最终答案是  $Y_\infty = Z_\infty$

Consider  $\mathbb{E}[X|\mathcal{A}, Z_0, Z_1, \dots] = Y_\infty$ . This means for any  $A \in \sigma(\mathcal{A}, Z_0, Z_1, \dots)$ ,  $\mathbb{E}[X1_A] = \mathbb{E}[Y_\infty 1_A]$ .

Consider  $\mathbb{E}[X|\mathcal{B}, Y_0, Y_1, \dots] = Z_\infty$ . This means for any  $B \in \sigma(\mathcal{B}, Y_0, Y_1, \dots)$ ,  $\mathbb{E}[X1_B] = \mathbb{E}[Z_\infty 1_B]$ .

Let  $\mathcal{H} = \sigma(\mathcal{A}, \mathcal{B})$ .

Consider  $\mathbb{E}[X|\mathcal{H}]$ .

Consider  $\mathbb{E}[Y_\infty|\mathcal{B}, Y_0, Y_1, \dots] = \mathbb{E}[\mathbb{E}[X|\mathcal{A}, Z_0, Z_1, \dots]|\mathcal{B}, Y_0, Y_1, \dots]$ .

Let's use the fact that  $Y_\infty$  is  $\sigma(\mathcal{A}, Z_0, Z_1, \dots)$ -measurable and  $Z_\infty$  is  $\sigma(\mathcal{B}, Y_0, Y_1, \dots)$ -measurable.

Consider the smallest sigma-algebra containing both  $\sigma(\mathcal{A}, Z_0, Z_1, \dots)$  and  $\sigma(\mathcal{B}, Y_0, Y_1, \dots)$ .

This is  $\sigma(\mathcal{A} \cup \mathcal{B} \cup \{Y_i\}_{i \geq 0} \cup \{Z_i\}_{i \geq 0})$ .

Let's try to show that  $\mathbb{E}[(Y_\infty - Z_\infty)^2] = 0$ .

$$\mathbb{E}[(Y_\infty - Z_\infty)^2] = \mathbb{E}[Y_\infty^2] + \mathbb{E}[Z_\infty^2] - 2\mathbb{E}[Y_\infty Z_\infty].$$

Consider the definition of  $Y_k$  and  $Z_k$ .

$$\mathbb{E}[Y_{k+1}|\mathcal{A}, Z_0, \dots, Z_{k-1}] = Y_k.$$

$$\mathbb{E}[Z_{k+1}|\mathcal{B}, Y_0, \dots, Y_{k-1}] = Z_k.$$

Consider  $\mathbb{E}[Y_\infty|\mathcal{B}, Y_0, Y_1, \dots]$ .

Since  $Y_\infty$  is measurable with respect to  $\sigma(\mathcal{A}, Z_0, Z_1, \dots)$ , this does not simplify easily.

Let's assume  $Y_\infty \neq Z_\infty$  on a set of positive measure.

Consider the case where  $\mathcal{A}$  and  $\mathcal{B}$  are trivial.

Then  $Y_k = \mathbb{E}[X|\sigma(Z_0, \dots, Z_{k-1})]$  and  $Z_k = \mathbb{E}[X|\sigma(Y_0, \dots, Y_{k-1})]$ .

Let  $\mathcal{H}_\infty = \sigma(\mathcal{A} \cup \mathcal{B} \cup \{Y_i, Z_i\}_{i \geq 0})$ .

Consider  $\mathbb{E}[X|\mathcal{H}_\infty]$ .

We have  $Y_\infty = \mathbb{E}[X|\sigma(\mathcal{A}, Z_0, Z_1, \dots)]$ .

By the properties of conditional expectation,  $\mathbb{E}[Y_\infty|\mathcal{A}] = \mathbb{E}[\mathbb{E}[X|\mathcal{A}, Z_0, Z_1, \dots]|\mathcal{A}]$ .

Consider the sigma-algebra  $\mathcal{D}$  generated by all the random variables involved.

$\mathcal{D} = \sigma(\mathcal{A} \cup \mathcal{B} \cup \{Y_i\}_{i \geq 0} \cup \{Z_i\}_{i \geq 0})$ .

Let  $M_\infty = \mathbb{E}[X|\mathcal{A} \vee \mathcal{B}]$ .

Consider the limits. As  $n \rightarrow \infty$ ,  $\sigma(\mathcal{A}, Z_0, \dots, Z_{n-1}) \uparrow \sigma(\mathcal{A}, Z_0, Z_1, \dots)$ .

As  $n \rightarrow \infty$ ,  $\sigma(\mathcal{B}, Y_0, \dots, Y_{n-1}) \uparrow \sigma(\mathcal{B}, Y_0, Y_1, \dots)$ .

Let  $\mathcal{Y}_\infty = \sigma(\mathcal{A}, Z_0, Z_1, \dots)$  and  $\mathcal{Z}_\infty = \sigma(\mathcal{B}, Y_0, Y_1, \dots)$ .

We have  $Y_\infty = \mathbb{E}[X|\mathcal{Y}_\infty]$  and  $Z_\infty = \mathbb{E}[X|\mathcal{Z}_\infty]$ .

Consider  $\mathbb{E}[Y_\infty - Z_\infty]^2 = \mathbb{E}[(Y_\infty - X) + (X - Z_\infty)]^2$

$$= \mathbb{E}[(Y_\infty - X)^2] + \mathbb{E}[(X - Z_\infty)^2] + 2\mathbb{E}[(Y_\infty - X)(X - Z_\infty)]$$

$$= \mathbb{E}[(X - \mathbb{E}[X|\mathcal{Y}_\infty])^2] + \mathbb{E}[(X - \mathbb{E}[X|\mathcal{Z}_\infty])^2] + 2\mathbb{E}[Y_\infty X - Y_\infty Z_\infty - X^2 + X Z_\infty]$$

$$= \mathbb{E}[X^2] - \mathbb{E}[Y_\infty^2] + \mathbb{E}[X^2] - \mathbb{E}[Z_\infty^2] + 2\mathbb{E}[Y_\infty X] - 2\mathbb{E}[Y_\infty Z_\infty] - 2\mathbb{E}[X^2] + 2\mathbb{E}[X Z_\infty]$$

$$= -\mathbb{E}[Y_\infty^2] - \mathbb{E}[Z_\infty^2] + 2\mathbb{E}[Y_\infty^2] - 2\mathbb{E}[Y_\infty Z_\infty] + 2\mathbb{E}[Z_\infty^2]$$

$$= \mathbb{E}[Y_\infty^2] + \mathbb{E}[Z_\infty^2] - 2\mathbb{E}[Y_\infty Z_\infty]$$

Consider the sigma-algebra  $\mathcal{C} = \sigma(\mathcal{A} \cup \mathcal{B})$ .

Let  $W = \mathbb{E}[X|\mathcal{C}]$ .

Consider the limit sigma-algebra generated by all the information.

Let  $\mathcal{F}_\infty = \sigma(\mathcal{A}, \mathcal{B}, Y_0, Y_1, Z_0, Z_1, \dots)$ .

Then  $\mathcal{Y}_\infty \subseteq \mathcal{F}_\infty$  and  $\mathcal{Z}_\infty \subseteq \mathcal{F}_\infty$ .

We have  $\mathbb{E}[X|\mathcal{Y}_\infty] = Y_\infty$  and  $\mathbb{E}[X|\mathcal{Z}_\infty] = Z_\infty$ .

Consider  $\mathbb{E}[Y_\infty|\mathcal{Z}_\infty]$  and  $\mathbb{E}[Z_\infty|\mathcal{Y}_\infty]$ .

Let  $\mathcal{H} = \sigma(\mathcal{A} \cup \mathcal{B})$ .

Table 9. s1K 从 s1-teasers 中提取的样本。黑色文本是问题，黄色文本是答案。仅根据问题，我们生成 浅蓝色 推理轨迹和 蓝色 回答，这些用于训练 Gemini。

爱丽丝和鲍勃各自有一枚硬币，他们一直翻转硬币直到出现正面。如果鲍勃翻转硬币的次数比爱丽丝多，那么爱丽丝翻转硬币的期望次数是多少？

4/3

Consider the limit sigma-algebra  $\mathcal{L} = \sigma(\mathcal{A} \cup \mathcal{B} \cup \sigma(Y_0, Z_0, \dots))$ .

We have  $Y_\infty = \mathbb{E}[X|\sigma(\mathcal{A}, Z_0, Z_1, \dots)]$  and  $Z_\infty = \mathbb{E}[X|\sigma(\mathcal{B}, Y_0, Y_1, \dots)]$ .

Let  $\mathcal{M} = \sigma(\mathcal{A}, Z_0, Z_1, \dots) \vee \sigma(\mathcal{B}, Y_0, Y_1, \dots)$ .

Consider  $\mathbb{E}[X|\mathcal{M}]$ .

Consider the random variable  $X$ . Then  $Y_\infty = \lim_n \mathbb{E}[X|\mathcal{F}_n]$  and  $Z_\infty = \lim_n \mathbb{E}[X|\mathcal{G}_n]$ .

Let  $\mathcal{H}_\infty = \sigma(\mathcal{A} \cup \mathcal{B} \cup \sigma(Y_0, Z_0, \dots))$ .

Consider  $E = \{\omega : Y_\infty(\omega) \neq Z_\infty(\omega)\}$ . Suppose  $P(E) > 0$ .

Consider the conditional expectation with respect to the join of the sigma-algebras.

Let  $\mathcal{C}_\infty = \sigma(\mathcal{A}, \mathcal{B}, Y_0, Z_0, \dots)$ .

Consider  $\mathbb{E}[X|\mathcal{C}_\infty]$ .

Then  $\mathbb{E}[\mathbb{E}[X|\mathcal{C}_\infty]|\mathcal{A}, Z_0, \dots] = \mathbb{E}[X|\mathcal{A}, Z_0, \dots] = Y_{n+1}$ .

And  $\mathbb{E}[\mathbb{E}[X|\mathcal{C}_\infty]|\mathcal{B}, Y_0, \dots] = \mathbb{E}[X|\mathcal{B}, Y_0, \dots] = Z_{n+1}$ .

So  $Y_\infty = \mathbb{E}[X|\sigma(\mathcal{A}, Z_0, Z_1, \dots)]$  and  $Z_\infty = \mathbb{E}[X|\sigma(\mathcal{B}, Y_0, Y_1, \dots)]$ .

Let  $\mathcal{F}_\infty = \sigma(\mathcal{A}, Z_0, Z_1, \dots)$  and  $\mathcal{G}_\infty = \sigma(\mathcal{B}, Y_0, Y_1, \dots)$ .

Consider  $\mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty]$ .

Then  $\mathbb{E}[\mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty]|\mathcal{F}_\infty] = \mathbb{E}[X|\mathcal{F}_\infty] = Y_\infty$ .

And  $\mathbb{E}[\mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty]|\mathcal{G}_\infty] = \mathbb{E}[X|\mathcal{G}_\infty] = Z_\infty$ .

This means  $Y_\infty = \mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty]$  a.s. and  $Z_\infty = \mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty]$  a.s.

Therefore,  $Y_\infty = Z_\infty$  a.s.

Solution:

(a) For the sequence  $Y_k$ , the  $\sigma$ -algebras  $\mathcal{F}_k = \sigma(\mathcal{A}, Z_0, \dots, Z_{k-1})$  are increasing as  $k$  increases. Since  $X$  is integrable, by the Martingale Convergence Theorem for conditional expectations, the sequence  $Y_k = \mathbb{E}[X|\mathcal{F}_k]$  converges almost surely and in  $L^1$  to  $Y_\infty = \mathbb{E}[X|\sigma(\mathcal{A}, Z_0, Z_1, \dots)]$ . Similarly, for the sequence  $Z_k$ , the  $\sigma$ -algebras  $\mathcal{G}_k = \sigma(\mathcal{B}, Y_0, \dots, Y_{k-1})$  are increasing. Since  $X$  is integrable, by the Martingale Convergence Theorem, the sequence  $Z_k = \mathbb{E}[X|\mathcal{G}_k]$  converges almost surely and in  $L^1$  to  $Z_\infty = \mathbb{E}[X|\sigma(\mathcal{B}, Y_0, Y_1, \dots)]$ .

(b) Let  $\mathcal{F}_\infty = \sigma(\mathcal{A}, Z_0, Z_1, \dots)$  and  $\mathcal{G}_\infty = \sigma(\mathcal{B}, Y_0, Y_1, \dots)$ . We have  $Y_\infty = \mathbb{E}[X|\mathcal{F}_\infty]$  and  $Z_\infty = \mathbb{E}[X|\mathcal{G}_\infty]$ . Consider the  $\sigma$ -algebra  $\mathcal{H} = \mathcal{F}_\infty \vee \mathcal{G}_\infty = \sigma(\mathcal{A}, \mathcal{B}, Y_0, Z_0, Y_1, Z_1, \dots)$ . Consider  $W = \mathbb{E}[X|\mathcal{H}]$ .

We have  $\mathcal{F}_\infty \subseteq \mathcal{H}$ , so  $\mathbb{E}[W|\mathcal{F}_\infty] = \mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{F}_\infty] = \mathbb{E}[X|\mathcal{F}_\infty] = Y_\infty$ .

Since  $W$  is  $\mathcal{H}$ -measurable,  $\mathbb{E}[W|\mathcal{F}_\infty] = Y_\infty$  implies that  $Y_\infty$  is the conditional expectation of  $W$  given  $\mathcal{F}_\infty$ . But since  $Y_\infty$  is already  $\mathcal{F}_\infty$ -measurable, we have  $\mathbb{E}[W|\mathcal{F}_\infty] = W$  if  $W$  is  $\mathcal{F}_\infty$ -measurable, which is not necessarily true.

However, we know that  $\mathbb{E}[Y_\infty|\mathcal{F}_\infty] = Y_\infty$  and  $\mathbb{E}[Z_\infty|\mathcal{G}_\infty] = Z_\infty$ .

Consider  $\mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty]$ .

Then  $\mathbb{E}[\mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty]|\mathcal{F}_\infty] = \mathbb{E}[X|\mathcal{F}_\infty] = Y_\infty$ .

This implies that  $\mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty] = Y_\infty$  almost surely.

Similarly,  $\mathbb{E}[\mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty]|\mathcal{G}_\infty] = \mathbb{E}[X|\mathcal{G}_\infty] = Z_\infty$ .

This implies that  $\mathbb{E}[X|\mathcal{F}_\infty \vee \mathcal{G}_\infty] = Z_\infty$  almost surely.

Therefore,  $Y_\infty = Z_\infty$  almost surely.

Final Answer: The final answer is  $Y_\infty = Z_\infty$

这个问题要求在鲍勃翻转硬币次数多于爱丽丝的情况下，爱丽丝翻转硬币的期望次数。

设  $X$  为爱丽丝翻转硬币直到出现正面的次数。

设  $Y$  为鲍勃翻转硬币直到出现正面的次数。

$X$  和  $Y$  是独立的随机变量，遵循参数为  $p = 0.5$ （出现正面的概率）的几何分布。

几何分布的概率质量函数为  $P(K = k) = (1-p)^{k-1}p$ , 对于  $k = 1, 2, \dots$ 。

在这种情况下,  $p = 0.5$ , 所以  $P(K = k) = (0.5)^{k-1}(0.5) = (0.5)^k = \frac{1}{2^k}$ 。

我们已知的条件是鲍勃翻转硬币的次数多于爱丽丝, 即  $Y > X$ 。

我们想要找到给定  $Y > X$  时  $X$  的期望值, 即  $E[X|Y > X]$ 。

条件期望的公式为  $E[X|A] = \sum_x xP(X = x|A)$ 。

同时,  $P(X = x|A) = \frac{P(X=x \cap A)}{P(A)}$ 。

在我们的情况下,  $A$  是事件  $Y > X$ 。

我们需要找到  $P(Y > X)$ 。

$P(Y > X) = \sum_{x=1}^{\infty} \sum_{y=x+1}^{\infty} P(X = x, Y = y)$

由于  $X$  和  $Y$  是独立的,  $P(X = x, Y = y) = P(X = x)P(Y = y) = \frac{1}{2^x} \frac{1}{2^y} = \frac{1}{2^{x+y}}$ 。

$P(Y > X) = \sum_{x=1}^{\infty} \sum_{y=x+1}^{\infty} \frac{1}{2^{x+y}} = \sum_{x=1}^{\infty} \frac{1}{2^x} \sum_{y=x+1}^{\infty} \frac{1}{2^y}$

内层求和是一个几何级数:  $\sum_{y=x+1}^{\infty} (\frac{1}{2})^y = (\frac{1}{2})^{x+1} + (\frac{1}{2})^{x+2} + \dots$

这是一个首项  $a = (\frac{1}{2})^{x+1}$  和公比  $r = \frac{1}{2}$  的几何级数。

其和为  $\frac{a}{1-r} = \frac{(\frac{1}{2})^{x+1}}{1-\frac{1}{2}} = \frac{(\frac{1}{2})^{x+1}}{\frac{1}{2}} = (\frac{1}{2})^x$ 。

因此,  $P(Y > X) = \sum_{x=1}^{\infty} \frac{1}{2^x} (\frac{1}{2})^x = \sum_{x=1}^{\infty} (\frac{1}{4})^x$ 。

这又是一个几何级数, 首项  $a = \frac{1}{4}$  和公比  $r = \frac{1}{4}$ 。

其和为  $\frac{a}{1-r} = \frac{\frac{1}{4}}{1-\frac{1}{4}} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$ 。

现在, 我们需要找到  $P(X = x \cap Y > X) = P(X = x, Y > x)$ 。

$P(X = x, Y > x) = P(X = x)P(Y > x)$ , 由于独立性。

$P(Y > x) = \sum_{y=x+1}^{\infty} P(Y = y) = \sum_{y=x+1}^{\infty} \frac{1}{2^y} = (\frac{1}{2})^{x+1} + (\frac{1}{2})^{x+2} + \dots = (\frac{1}{2})^x$ 。所以,  $P(X = x, Y > x) = \frac{1}{2^x} \times (\frac{1}{2})^x = (\frac{1}{4})^x$ 。

现在, 我们可以找到  $P(X = x|Y > X) = \frac{P(X=x \cap Y > X)}{P(Y > X)} = \frac{(\frac{1}{4})^x}{\frac{1}{3}} = 3(\frac{1}{4})^x$ 。

最后, 我们可以计算条件期望:

$E[X|Y > X] = \sum_{x=1}^{\infty} xP(X = x|Y > X) = \sum_{x=1}^{\infty} x \cdot 3(\frac{1}{4})^x = 3 \sum_{x=1}^{\infty} x(\frac{1}{4})^x$ 。

考虑级数  $S = \sum_{x=1}^{\infty} xr^x$ , 其中  $|r| < 1$ 。

$S = r + 2r^2 + 3r^3 + 4r^4 + \dots$

$\frac{S}{r} = 1 + 2r + 3r^2 + 4r^3 + \dots$

考虑几何级数  $G = 1 + r + r^2 + r^3 + \dots = \frac{1}{1-r}$ 。

对  $r$  求导:

$\frac{dG}{dr} = 1 + 2r + 3r^2 + 4r^3 + \dots = \frac{d}{dr}(\frac{1}{1-r}) = \frac{1}{(1-r)^2}$ 。

所以,  $\frac{S}{r} = \frac{1}{(1-r)^2}$ , 这意味着  $S = \frac{r}{(1-r)^2}$ 。

在我们的情况下,  $r = \frac{1}{4}$ 。

Table 9. s1K sample stemming from s1-teasers. The black text is the question and the yellow text is the solution. Based solely on the question we generate the light blue reasoning trace and blue response with Gemini, which we use for training.

Alice and Bob each have a coin and flip it until they get a heads. If Bob flipped his coin more times than Alice, what is the expected number of times Alice flipped her coin?

4/3

$$\sum_{x=1}^{\infty} x \left(\frac{1}{4}\right)^x = \frac{\frac{1}{4}}{(1-\frac{1}{4})^2} = \frac{\frac{1}{4}}{(\frac{3}{4})^2} = \frac{\frac{1}{4}}{\frac{9}{16}} = \frac{1}{4} \times \frac{16}{9} = \frac{4}{9}.$$

$$\text{因此, } E[X|Y > X] = 3 \times \frac{4}{9} = \frac{12}{9} = \frac{4}{3}.$$

另一种方法：对称性。

考虑事件  $Y > X$ ,  $X > Y$  和  $X = Y$ 。

$P(Y > X) = P(X > Y)$  由于对称性。

$$P(X = Y) = \sum_{k=1}^{\infty} P(X = k, Y = k) = \sum_{k=1}^{\infty} P(X = k)P(Y = k) = \sum_{k=1}^{\infty} (\frac{1}{2^k})^2 = \sum_{k=1}^{\infty} (\frac{1}{4})^k = \frac{\frac{1}{4}}{1-\frac{1}{4}} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

我们有  $P(Y > X) + P(X > Y) + P(X = Y) = 1$ 。

$$P(Y > X) + P(Y > X) + \frac{1}{3} = 1.$$

$$2P(Y > X) = 1 - \frac{1}{3} = \frac{2}{3}.$$

$P(Y > X) = \frac{1}{3}$ 。这证实了我们之前的计算。

设  $E[X|Y > X] = \mu$ 。

我们正在寻找给定 Bob 翻转次数更多时 Alice 翻转的期望次数。

考虑两人都第一次翻转为反面的情况。那么过程重新开始，条件保持不变。

设  $A_1$  为 Alice 的第一次翻转,  $B_1$  为 Bob 的第一次翻转。

$$P(A_1 = H) = 1/2, P(A_1 = T) = 1/2.$$

$$P(B_1 = H) = 1/2, P(B_1 = T) = 1/2。案例1: A_1 = H, B_1 = H. X = 1, Y = 1, 所以 Y > X 为假。$$

案例2:  $A_1 = H, B_1 = T$ . Alice 在1停止, Bob 继续。 $Y > 1$ 。

案例3:  $A_1 = T, B_1 = H$ . Bob 在1停止, Alice 继续。

案例4:  $A_1 = T, B_1 = T$ . 两者都继续。

考虑给定  $Y > X$  的条件概率。

$$\text{设 } E = E[X|Y > X].$$

如果第一次翻转是 HH, 条件  $Y > X$  不满足。

如果第一次翻转是 HT, Alice 的翻转次数为1。

如果第一次翻转是 TH, Bob 的翻转次数为1, Alice 的至少为2, 所以  $Y < X$ 。

如果第一次翻转是 TT, 问题重新开始, Alice 的额外预期翻转次数仍然是  $E$ 。

设  $N_A$  为 Alice 的翻转次数,  $N_B$  为 Bob 的翻转次数。

已知  $N_B > N_A$ 。

考虑第一次翻转:

HH (概率 1/4): 条件不满足。

HT (概率 1/4): Alice 翻转1次。

TH (概率 1/4): Bob 翻转1次, Alice > 1, 所以  $N_A > N_B$ , 条件不满足。

TT (概率 1/4): 两者都继续。

设  $E$  为给定 Bob 翻转次数更多时 Alice 的预期翻转次数。

如果第一次翻转是 HT, Alice 翻转1次, 条件满足。

如果第一次翻转是 TT, 情况与开始相同, 所以 Alice 的额外预期翻转次数为  $E$ 。

考虑事件  $Y > X$ 。

基于第一次翻转的条件:

$$E[X|Y > X] = E[X|Y > X, HH]P(HH|Y > X) + E[X|Y > X, HT]P(HT|Y > X) + E[X|Y > X, TH]P(TH|Y > X) + E[X|Y > X, TT]P(TT|Y > X)$$

$$P(Y > X|HH) = 0.$$

$$P(Y > X|HT) = 1.$$

$$P(Y > X|TH) = 0.$$

$P(Y > X|TT)$  取决于未来的翻转。

让我们直接使用条件期望的定义。

The problem asks for the expected number of times Alice flipped her coin, given that Bob flipped his coin more times than Alice.

Let  $X$  be the number of flips Alice makes until she gets a head.

Let  $Y$  be the number of flips Bob makes until he gets a head.

$X$  and  $Y$  are independent random variables following a geometric distribution with parameter  $p = 0.5$  (probability of getting a head).

The probability mass function for a geometric distribution is  $P(K = k) = (1 - p)^{k-1} p$ , for  $k = 1, 2, \dots$ .

In this case,  $p = 0.5$ , so  $P(K = k) = (0.5)^{k-1} (0.5) = (0.5)^k = \frac{1}{2^k}$ .

We are given the condition that Bob flipped his coin more times than Alice, which means  $Y > X$ .

We want to find the expected value of  $X$  given that  $Y > X$ , i.e.,  $E[X|Y > X]$ .

The formula for conditional expectation is  $E[X|A] = \sum_x x P(X = x|A)$ .

Also,  $P(X = x|A) = \frac{P(X=x \cap A)}{P(A)}$ .

In our case,  $A$  is the event  $Y > X$ .

We need to find  $P(Y > X)$ .

$$P(Y > X) = \sum_{x=1}^{\infty} \sum_{y=x+1}^{\infty} P(X = x, Y = y)$$

Since  $X$  and  $Y$  are independent,  $P(X = x, Y = y) = P(X = x)P(Y = y) = \frac{1}{2^x} \frac{1}{2^y} = \frac{1}{2^{x+y}}$ .

$$P(Y > X) = \sum_{x=1}^{\infty} \sum_{y=x+1}^{\infty} \frac{1}{2^{x+y}} = \sum_{x=1}^{\infty} \frac{1}{2^x} \sum_{y=x+1}^{\infty} \frac{1}{2^y}$$

The inner sum is a geometric series:  $\sum_{y=x+1}^{\infty} (\frac{1}{2})^y = (\frac{1}{2})^{x+1} + (\frac{1}{2})^{x+2} + \dots$

This is a geometric series with first term  $a = (\frac{1}{2})^{x+1}$  and common ratio  $r = \frac{1}{2}$ .

$$\text{The sum is } \frac{a}{1-r} = \frac{(\frac{1}{2})^{x+1}}{1-\frac{1}{2}} = \frac{(\frac{1}{2})^{x+1}}{\frac{1}{2}} = (\frac{1}{2})^x.$$

$$\text{So, } P(Y > X) = \sum_{x=1}^{\infty} \frac{1}{2^x} (\frac{1}{2})^x = \sum_{x=1}^{\infty} (\frac{1}{4})^x.$$

This is another geometric series with first term  $a = \frac{1}{4}$  and common ratio  $r = \frac{1}{4}$ .

$$\text{The sum is } \frac{a}{1-r} = \frac{\frac{1}{4}}{1-\frac{1}{4}} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

Now, we need to find  $P(X = x \cap Y > X) = P(X = x, Y > x)$ .

$P(X = x, Y > x) = P(X = x)P(Y > x)$ , due to independence.

$$P(Y > x) = \sum_{y=x+1}^{\infty} P(Y = y) = \sum_{y=x+1}^{\infty} \frac{1}{2^y} = (\frac{1}{2})^{x+1} + (\frac{1}{2})^{x+2} + \dots = (\frac{1}{2})^x.$$

$$\text{So, } P(X = x, Y > x) = \frac{1}{2^x} \times (\frac{1}{2})^x = (\frac{1}{4})^x.$$

$$\text{Now, we can find } P(X = x|Y > X) = \frac{P(X=x \cap Y > X)}{P(Y > X)} = \frac{(\frac{1}{4})^x}{\frac{1}{3}} = 3(\frac{1}{4})^x.$$

Finally, we can calculate the conditional expectation:

$$E[X|Y > X] = \sum_{x=1}^{\infty} x P(X = x|Y > X) = \sum_{x=1}^{\infty} x \cdot 3(\frac{1}{4})^x = 3 \sum_{x=1}^{\infty} x(\frac{1}{4})^x.$$

Consider the series  $S = \sum_{x=1}^{\infty} x r^x$ , where  $|r| < 1$ .

$$S = r + 2r^2 + 3r^3 + 4r^4 + \dots$$

$$\frac{S}{r} = 1 + 2r + 3r^2 + 4r^3 + \dots$$

$$\text{Consider the geometric series } G = 1 + r + r^2 + r^3 + \dots = \frac{1}{1-r}.$$

Differentiating with respect to  $r$ :

$$\frac{dG}{dr} = 1 + 2r + 3r^2 + 4r^3 + \dots = \frac{d}{dr}(\frac{1}{1-r}) = \frac{1}{(1-r)^2}.$$

$$\text{So, } \frac{S}{r} = \frac{1}{(1-r)^2}, \text{ which means } S = \frac{r}{(1-r)^2}.$$

In our case,  $r = \frac{1}{4}$ .

$$P(TH|Y > X) = \frac{P(Y > X|TH)P(TH)}{P(Y > X)} = \frac{0 \cdot \frac{1}{4}}{\frac{1}{3}} = 0.$$

$$P(TT|Y > X) = \frac{P(Y > X|TT)P(TT)}{P(Y > X)}$$

如果第一次翻转是 TT, 未来的行为是相同的, 条件  $Y > X$  成立当且仅当 Bob 在后续试验中比 Alice 多翻转次数。

$P(Y > X|TT) = P(\text{Bob 从头开始需要比 Alice 多翻转次数}) = P(Y > X) = 1/3$ .

$$P(TT|Y > X) = \frac{(1/3) \cdot (1/4)}{1/3} = \frac{1}{4}.$$

条件概率之和:  $0 + 3/4 + 0 + 1/4 = 1$ .

现在考虑  $X$  的条件期望:

$E[X|Y > X, HH]$ : 不相关, 因为  $P(HH|Y > X) = 0$ .

$E[X|Y > X, HT]$ : 给定 HT 和  $Y > X$ , Alice 翻转了 1 次。 $E = 1$ .

$E[X|Y > X, TH]$ : 不相关。

$E[X|Y > X, TT]$ : 给定 TT 和  $Y > X$ , Alice 已经翻转了一次, 额外的期望翻转次数是  $E$ 。因此, 总的翻转次数是  $1 + E$ .

$$E = 0 \cdot 0 + 1 \cdot (3/4) + \text{未定义} \cdot 0 + (1 + E) \cdot (1/4)$$

$$E = 0 + 3/4 + 0 + 1/4 + E/4$$

$$E = 1 + E/4$$

$$E - E/4 = 1$$

$$3E/4 = 1$$

$$E = 4/3.$$

这证实了前面的结果。

让我们考虑一下直觉。如果条件是  $Y > X$ , 这意味着 Alice “幸运” 地早期得到了正面, 或者 Bob “不幸”。

考虑随机变量  $Z = Y - X$ 。我们已知  $Z > 0$ .

我们想要  $E[X|Z > 0]$ .

设  $p = P(\text{正面}) = 0.5$ .

第一次出现正面之前的反面次数服从参数为  $p$  的几何分布。

$X - 1 \sim Geom(p)$ ,  $Y - 1 \sim Geom(p)$ .

$$P(X - 1 = k) = (1 - p)^k p.$$

$P(X = x) = (1 - p)^{x-1} p$ 。考虑  $X$  和  $Y$  的联合分布。

$$P(X = x, Y = y) = p^2 (1 - p)^{x-1} (1 - p)^{y-1}$$

考虑事件  $Y = X + k$ , 其中  $k \geq 1$ .

$$P(Y = X + k) = \sum_{x=1}^{\infty} P(X = x, Y = x + k) = \sum_{x=1}^{\infty} p^2 (1 - p)^{x-1} (1 - p)^{x+k-1}$$

$$= p^2 (1 - p)^{k-1} \sum_{x=1}^{\infty} ((1 - p)^2)^{x-1} = p^2 (1 - p)^{k-1} \frac{1}{1 - (1 - p)^2}$$

$$= p^2 (1 - p)^{k-1} \frac{1}{1 - (1 - 2p + p^2)} = p^2 (1 - p)^{k-1} \frac{1}{2p - p^2} = p^2 (1 - p)^{k-1} \frac{1}{p(2 - p)}$$

$$= \frac{p(1 - p)^{k-1}}{2 - p}.$$

如果  $p = 0.5$ , 则  $1 - p = 0.5$ ,  $2 - p = 1.5$ .

$$P(Y = X + k) = \frac{0.5(0.5)^{k-1}}{1.5} = \frac{(0.5)^k}{1.5} = \frac{(1/2)^k}{3/2} = \frac{1}{2^k} \cdot \frac{2}{3} = \frac{1}{3 \cdot 2^{k-1}}.$$

检查  $\sum_{k=1}^{\infty} P(Y = X + k) = P(Y > X) = 1/3$ .

$$\sum_{k=1}^{\infty} \frac{1}{3 \cdot 2^{k-1}} = \frac{1}{3} \sum_{k=1}^{\infty} (\frac{1}{2})^{k-1} = \frac{1}{3} (1 + 1/2 + 1/4 + \dots) = \frac{1}{3} \frac{1}{1 - 1/2} = \frac{1}{3} \cdot 2 = \frac{2}{3}.$$

这里有问题。

让我们重新计算  $P(Y > X)$ .

$$P(Y > X) = \sum_{x=1}^{\infty} \sum_{y=x+1}^{\infty} (1 - p)^{x-1} p (1 - p)^{y-1} p$$

$$= p^2 \sum_{x=1}^{\infty} (1 - p)^{x-1} \sum_{y=x+1}^{\infty} (1 - p)^{y-1}$$

$$\text{内层求和: } (1 - p)^x + (1 - p)^{x+1} + \dots = \frac{(1 - p)^x}{1 - (1 - p)} = \frac{(1 - p)^x}{p}.$$

$$\sum_{x=1}^{\infty} x \left(\frac{1}{4}\right)^x = \frac{\frac{1}{4}}{(1-\frac{1}{4})^2} = \frac{\frac{1}{4}}{(\frac{3}{4})^2} = \frac{\frac{1}{4}}{\frac{9}{16}} = \frac{1}{4} \times \frac{16}{9} = \frac{4}{9}.$$

Therefore,  $E[X|Y > X] = 3 \times \frac{4}{9} = \frac{12}{9} = \frac{4}{3}$ .

Alternative approach: Symmetry.

Consider the events  $Y > X$ ,  $X > Y$ , and  $X = Y$ .

$P(Y > X) = P(X > Y)$  due to symmetry.

$$P(X = Y) = \sum_{k=1}^{\infty} P(X = k, Y = k) = \sum_{k=1}^{\infty} P(X = k)P(Y = k) = \sum_{k=1}^{\infty} (\frac{1}{2^k})^2 = \sum_{k=1}^{\infty} (\frac{1}{4})^k = \frac{\frac{1}{4}}{1-\frac{1}{4}} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

We have  $P(Y > X) + P(X > Y) + P(X = Y) = 1$ .

$$P(Y > X) + P(Y > X) + \frac{1}{3} = 1.$$

$$2P(Y > X) = 1 - \frac{1}{3} = \frac{2}{3}.$$

$P(Y > X) = \frac{1}{3}$ . This confirms our previous calculation.

Let  $E[X|Y > X] = \mu$ .

We are looking for the expected number of flips Alice made, given Bob flipped more times.

Consider the case where the first flip for both is tails. Then the process restarts, and the condition remains.

Let  $A_1$  be Alice's first flip,  $B_1$  be Bob's first flip.

$$P(A_1 = H) = 1/2, P(A_1 = T) = 1/2.$$

$$P(B_1 = H) = 1/2, P(B_1 = T) = 1/2.$$

Case 1:  $A_1 = H, B_1 = H$ .  $X = 1, Y = 1$ , so  $Y > X$  is false.

Case 2:  $A_1 = H, B_1 = T$ . Alice stops at 1, Bob continues.  $Y > 1$ .

Case 3:  $A_1 = T, B_1 = H$ . Bob stops at 1, Alice continues.

Case 4:  $A_1 = T, B_1 = T$ . Both continue.

Consider the conditional probabilities given  $Y > X$ .

Let  $E = E[X|Y > X]$ .

If the first flips are HH, the condition  $Y > X$  is not met.

If the first flips are HT, Alice's number of flips is 1.

If the first flips are TH, Bob's number of flips is 1, Alice's is at least 2, so  $Y < X$ .

If the first flips are TT, the problem restarts, and the expected number of additional flips for Alice is still  $E$ .

Let  $N_A$  be the number of flips Alice makes,  $N_B$  be the number of flips Bob makes.

We are given  $N_B > N_A$ .

Consider the first flips:

HH (prob 1/4): Condition not met.

HT (prob 1/4): Alice flips 1 time.

TH (prob 1/4): Bob flips 1 time, Alice > 1, so  $N_A > N_B$ , condition not met.

TT (prob 1/4): Both continue.

Let  $E$  be the expected number of Alice's flips given Bob flipped more.

If the first flips are HT, Alice flipped 1 time, and the condition is met.

If the first flips are TT, the situation is the same as the beginning, so the expected number of additional flips for Alice is  $E$ .

Consider the event  $Y > X$ .

Conditioning on the first flips:

$$E[X|Y > X] = E[X|Y > X, HH]P(HH|Y > X) + E[X|Y > X, HT]P(HT|Y > X) + E[X|Y >$$

$$X, TH]P(TH|Y > X) + E[X|Y > X, TT]P(TT|Y > X)$$

$$P(Y > X|HH) = 0.$$

$$P(Y > X|HT) = 1.$$

$$P(Y > X|TH) = 0.$$

现在, 让我们找到  $E[X|Y > X]$ 。  $E[X|Y > X] = \sum_{x=1}^{\infty} xP(X = x|Y > X) = \sum_{x=1}^{\infty} x \frac{P(X=x, Y>x)}{P(Y>X)}$ 。

$$P(X = x, Y > x) = P(X = x)P(Y > x) \text{ (独立性)}$$

$$P(X = x) = (1-p)^{x-1}p.$$

$$P(Y > x) = \sum_{y=x+1}^{\infty} (1-p)^{y-1}p = p[(1-p)^x + (1-p)^{x+1} + \dots] = p \frac{(1-p)^x}{1-(1-p)} = (1-p)^x.$$

$$P(X = x, Y > x) = (1-p)^{x-1}p(1-p)^x = p(1-p)^{2x-1}.$$

$$E[X|Y > X] = \frac{1}{P(Y>X)} \sum_{x=1}^{\infty} xp(1-p)^{2x-1}$$

$$= \frac{2-p}{1-p} p(1-p)^{-1} \sum_{x=1}^{\infty} x((1-p)^2)^x \text{ (幂次错误)}$$

$$P(Y > x) = \frac{(1-p)^x}{1-(1-p)} p = (1-p)^x.$$

$$E[X|Y > X] = \frac{1}{P(Y>X)} \sum_{x=1}^{\infty} xP(X = x)P(Y > x)$$

$$= \frac{1}{P(Y>X)} \sum_{x=1}^{\infty} x(1-p)^{x-1}p(1-p)^x$$

$$= \frac{p}{P(Y>X)} \sum_{x=1}^{\infty} x(1-p)^{2x-1}$$

$$= \frac{p}{(1-p)/(2-p)} (1-p)^{-1} \sum_{x=1}^{\infty} x((1-p)^2)^x$$

$$= p \frac{2-p}{(1-p)^2} \sum_{x=1}^{\infty} x((1-p)^2)^x.$$

$$\text{令 } r = (1-p)^2. \sum_{x=1}^{\infty} xr^x = \frac{r}{(1-r)^2}.$$

$$1-r = 1 - (1-p)^2 = 1 - (1-2p+p^2) = 2p - p^2 = p(2-p).$$

$$(1-r)^2 = p^2(2-p)^2.$$

$$\sum_{x=1}^{\infty} x((1-p)^2)^x = \frac{(1-p)^2}{p^2(2-p)^2}.$$

$$E[X|Y > X] = p \frac{2-p}{(1-p)^2} \frac{(1-p)^2}{p^2(2-p)^2} = \frac{1}{p(2-p)}.$$

$$\text{如果 } p = 0.5, E[X|Y > X] = \frac{1}{0.5 \times 1.5} = \frac{1}{0.75} = \frac{1}{3/4} = 4/3.$$

解:

设  $X$  是爱丽丝掷硬币直到出现正面的次数,  $Y$  是鲍勃掷硬币直到出现正面的次数。 $X$  和  $Y$  是独立的几何随机变量, 参数为  $p = 0.5$ 。概率质量函数为  $P(K = k) = (1-p)^{k-1}p$ 。

我们感兴趣的是, 在鲍勃掷硬币次数多于爱丽丝的条件下, 爱丽丝掷硬币的期望次数, 即  $E[X|Y > X]$ 。

使用条件期望的定义,  $E[X|Y > X] = \sum_{x=1}^{\infty} xP(X = x|Y > X)$ 。

$$P(X = x|Y > X) = \frac{P(X=x, Y>x)}{P(Y>X)}.$$

$$P(Y > X) = \sum_{x=1}^{\infty} \sum_{y=x+1}^{\infty} P(X = x, Y = y) = \sum_{x=1}^{\infty} P(X = x)P(Y > x).$$

$$P(X = x) = (0.5)^{x-1}(0.5) = (0.5)^x = \frac{1}{2^x}.$$

$$P(Y > x) = \sum_{y=x+1}^{\infty} (0.5)^{y-1}(0.5) = (0.5)^x.$$

$$P(Y > X) = \sum_{x=1}^{\infty} \frac{1}{2^x} (\frac{1}{2})^x = \sum_{x=1}^{\infty} (\frac{1}{4})^x = \frac{1/4}{1-1/4} = \frac{1/4}{3/4} = \frac{1}{3}.$$

$$\text{接下来, 计算 } P(X = x, Y > x) = P(X = x)P(Y > x) = \frac{1}{2^x} (\frac{1}{2})^x = (\frac{1}{4})^x.$$

$$\text{现在, 求 } P(X = x|Y > X) = \frac{P(X=x, Y>x)}{P(Y>X)} = \frac{(1/4)^x}{1/3} = 3(\frac{1}{4})^x.$$

最后, 计算条件期望:

$$E[X|Y > X] = \sum_{x=1}^{\infty} xP(X = x|Y > X) = \sum_{x=1}^{\infty} x \cdot 3(\frac{1}{4})^x = 3 \sum_{x=1}^{\infty} x(\frac{1}{4})^x.$$

使用公式  $\sum_{x=1}^{\infty} xr^x = \frac{r}{(1-r)^2}$ , 其中  $r = 1/4$ :

$$\sum_{x=1}^{\infty} x(\frac{1}{4})^x = \frac{1/4}{(1-1/4)^2} = \frac{1/4}{(3/4)^2} = \frac{1/4}{9/16} = \frac{1/4}{4/9} = \frac{1}{4} \times \frac{16}{9} = \frac{4}{9}.$$

$$E[X|Y > X] = 3 \times \frac{4}{9} = \frac{12}{9} = \frac{4}{3}.$$

最终答案: 最终答案是  $\boxed{4/3}$ .

Final Answer: The final answer is  $\boxed{4/3}$

$$P(TH|Y > X) = \frac{P(Y > X|TH)P(TH)}{P(Y > X)} = \frac{0 \cdot 1/4}{1/3} = 0.$$

$$P(TT|Y > X) = \frac{P(Y > X|TT)P(TT)}{P(Y > X)}$$

If the first flips are TT, the future behavior is the same, and the condition  $Y > X$  holds if Bob takes more flips than Alice in the subsequent trials.

$$P(Y > X|TT) = P(\text{Bob needs more flips than Alice starting from scratch}) = P(Y > X) = 1/3.$$

$$P(TT|Y > X) = \frac{(1/3) \cdot (1/4)}{1/3} = \frac{1}{4}.$$

Sum of conditional probabilities:  $0 + 3/4 + 0 + 1/4 = 1$ .

Now consider the conditional expectations of  $X$ :

$E[X|Y > X, HH]$ : Not relevant since  $P(HH|Y > X) = 0$ .

$E[X|Y > X, HT]$ : Given HT and  $Y > X$ , Alice flipped 1 time.  $E = 1$ .

$E[X|Y > X, TH]$ : Not relevant.

$E[X|Y > X, TT]$ : Given TT and  $Y > X$ , Alice has already flipped once, and the expected number of additional flips is  $E$ . So, the total number of flips is  $1 + E$ .

$$E = 0 \cdot 0 + 1 \cdot (3/4) + \text{undefined} \cdot 0 + (1 + E) \cdot (1/4)$$

$$E = 0 + 3/4 + 0 + 1/4 + E/4$$

$$E = 1 + E/4$$

$$E - E/4 = 1$$

$$3E/4 = 1$$

$$E = 4/3.$$

This confirms the previous result.

Let's think about the intuition. If we have the condition  $Y > X$ , it means Alice was "lucky" and got a head early, or Bob was "unlucky".

Consider the random variable  $Z = Y - X$ . We are given  $Z > 0$ .

We want  $E[X|Z > 0]$ .

Let  $p = P(\text{head}) = 0.5$ .

The number of tails before the first head follows a geometric distribution with parameter  $p$ .

$X - 1 \sim \text{Geom}(p)$ ,  $Y - 1 \sim \text{Geom}(p)$ .

$$P(X - 1 = k) = (1 - p)^k p.$$

$$P(X = x) = (1 - p)^{x-1} p.$$

Consider the joint distribution of  $X$  and  $Y$ .

$$P(X = x, Y = y) = p^2(1 - p)^{x-1}(1 - p)^{y-1}$$

Consider the event  $Y = X + k$ , where  $k \geq 1$ .

$$\begin{aligned} P(Y = X + k) &= \sum_{x=1}^{\infty} P(X = x, Y = x + k) = \sum_{x=1}^{\infty} p^2(1 - p)^{x-1}(1 - p)^{x+k-1} \\ &= p^2(1 - p)^{k-1} \sum_{x=1}^{\infty} ((1 - p)^2)^{x-1} = p^2(1 - p)^{k-1} \frac{1}{1 - (1 - p)^2} \\ &= p^2(1 - p)^{k-1} \frac{1}{1 - (2p - p^2)} = p^2(1 - p)^{k-1} \frac{1}{2p - p^2} = p^2(1 - p)^{k-1} \frac{1}{p(2 - p)} \\ &= \frac{p(1 - p)^{k-1}}{2 - p}. \end{aligned}$$

If  $p = 0.5$ , then  $1 - p = 0.5$ ,  $2 - p = 1.5$ .

$$P(Y = X + k) = \frac{0.5(0.5)^{k-1}}{1.5} = \frac{(0.5)^k}{1.5} = \frac{(1/2)^k}{3/2} = \frac{1}{2^k} \cdot \frac{2}{3} = \frac{1}{3 \cdot 2^{k-1}}.$$

Check if  $\sum_{k=1}^{\infty} P(Y = X + k) = P(Y > X) = 1/3$ .

$$\sum_{k=1}^{\infty} \frac{1}{3 \cdot 2^{k-1}} = \frac{1}{3} \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^{k-1} = \frac{1}{3} (1 + 1/2 + 1/4 + \dots) = \frac{1}{3} \frac{1}{1 - 1/2} = \frac{1}{3} \cdot 2 = \frac{2}{3}.$$

Something is wrong here.

35

Let's recalculate  $P(Y > X)$ .

$$P(Y > X) = \sum_{x=1}^{\infty} \sum_{y=x+1}^{\infty} (1 - p)^{x-1} p (1 - p)^{y-1} p$$

$$= p^2 \sum_{x=1}^{\infty} (1 - p)^{x-1} \sum_{y=x+1}^{\infty} (1 - p)^{y-1}$$

35

Now, let's find  $E[X|Y > X]$ .

$$E[X|Y > X] = \sum_{x=1}^{\infty} x P(X = x|Y > X) = \sum_{x=1}^{\infty} x \frac{P(X=x, Y>x)}{P(Y>X)}.$$

$P(X = x, Y > x) = P(X = x)P(Y > x)$  (independence)

$$P(X = x) = (1 - p)^{x-1} p.$$

$$P(Y > x) = \sum_{y=x+1}^{\infty} (1 - p)^{y-1} p = p[(1 - p)^x + (1 - p)^{x+1} + \dots] = p \frac{(1-p)^x}{1-(1-p)} = (1 - p)^x.$$

$$P(X = x, Y > x) = (1 - p)^{x-1} p(1 - p)^x = p(1 - p)^{2x-1}.$$

$$E[X|Y > X] = \frac{1}{P(Y>X)} \sum_{x=1}^{\infty} x p(1 - p)^{2x-1}$$

$$= \frac{2-p}{1-p} p(1 - p)^{-1} \sum_{x=1}^{\infty} x((1 - p)^2)^x \text{ (error in power)}$$

$$P(Y > x) = \frac{(1-p)^x}{1-(1-p)} p = (1 - p)^x.$$

$$E[X|Y > X] = \frac{1}{P(Y>X)} \sum_{x=1}^{\infty} x P(X = x) P(Y > x)$$

$$= \frac{1}{P(Y>X)} \sum_{x=1}^{\infty} x(1 - p)^{x-1} p(1 - p)^x$$

$$= \frac{p}{P(Y>X)} \sum_{x=1}^{\infty} x(1 - p)^{2x-1}$$

$$= \frac{p}{(1-p)/(2-p)} (1 - p)^{-1} \sum_{x=1}^{\infty} x((1 - p)^2)^x$$

$$= p \frac{2-p}{(1-p)^2} \sum_{x=1}^{\infty} x((1 - p)^2)^x.$$

$$\text{Let } r = (1 - p)^2. \sum_{x=1}^{\infty} x r^x = \frac{r}{(1-r)^2}.$$

$$1 - r = 1 - (1 - p)^2 = 1 - (1 - 2p + p^2) = 2p - p^2 = p(2 - p).$$

$$(1 - r)^2 = p^2(2 - p)^2.$$

$$\sum_{x=1}^{\infty} x((1 - p)^2)^x = \frac{(1-p)^2}{p^2(2-p)^2}.$$

$$E[X|Y > X] = p \frac{2-p}{(1-p)^2} \frac{(1-p)^2}{p^2(2-p)^2} = \frac{1}{p(2-p)}.$$

$$\text{If } p = 0.5, E[X|Y > X] = \frac{1}{0.5 \times 1.5} = \frac{1}{0.75} = \frac{1}{3/4} = 4/3.$$

Table 10. s1K 样本来自 OlympicArena 的天文学子集。黑色文本是问题，蓝色文本是答案。仅基于问题，我们生成 浅蓝色 推理轨迹和 蓝色 回答，我们用这些来训练 Gemini。

一个经常被提及的有趣事实是，人类单位体积产生的功率比恒星还要高。如果太阳的大小相同，但其单位体积产生的功率与人类相同，那么它的表面温度会是多少？假设“平均人类”产生的功率为100瓦，体积为66400立方厘米。

A: 3500 K

B: 10000 K

C: 25000 K

D: 40000 K

E: 50000 K

根据问题中的数据，平均人类产生的功率为

$$u = \frac{100}{66400 \times 10^{-6}} = 1506 \text{ W/m}^3$$

太阳的体积是  $\frac{4}{3}\pi R_{\odot}^3$ ，因此其新的功率输出为  $P = \frac{4}{3}\pi R_{\odot}^3 u$ 。为了找到平衡温度，我们使用

$$P = \sigma A T^4 = 4\pi\sigma R_{\odot}^2 T^4$$

Solving for  $T$ , we get

$$T = \sqrt[4]{\frac{R_{\odot} u}{3\sigma}} = \sqrt[4]{\frac{(6.96 \times 10^8)(1506)}{3(5.67 \times 10^{-8})}} = 49823 \approx 50000 \text{ K}.$$

Solution:

Let  $X$  be the number of flips Alice makes until she gets a head, and  $Y$  be the number of flips Bob makes until he gets a head.  $X$  and  $Y$  are independent geometric random variables with parameter  $p = 0.5$ . The probability mass function is  $P(K = k) = (1 - p)^{k-1} p$ .

We are interested in the expected number of times Alice flipped her coin, given that Bob flipped his coin more times than Alice, i.e.,  $E[X|Y > X]$ .

Using the definition of conditional expectation,  $E[X|Y > X] = \sum_{x=1}^{\infty} x P(X = x|Y > X)$ .

$$P(X = x|Y > X) = \frac{P(X=x, Y>x)}{P(Y>X)}$$

First, calculate  $P(Y > X)$ .

$$P(Y > X) = \sum_{x=1}^{\infty} \sum_{y=x+1}^{\infty} P(X = x, Y = y) = \sum_{x=1}^{\infty} P(X = x) P(Y > x)$$

$$P(X = x) = (0.5)^{x-1}(0.5) = (0.5)^x = \frac{1}{2^x}$$

$$P(Y > x) = \sum_{y=x+1}^{\infty} (0.5)^{y-1}(0.5) = (0.5)^x$$

$$P(Y > X) = \sum_{x=1}^{\infty} \frac{1}{2^x} \left(\frac{1}{2}\right)^x = \sum_{x=1}^{\infty} \left(\frac{1}{4}\right)^x = \frac{1/4}{1-1/4} = \frac{1/4}{3/4} = \frac{1}{3}$$

$$\text{Next, calculate } P(X = x, Y > x) = P(X = x)P(Y > x) = \frac{1}{2^x} \left(\frac{1}{2}\right)^x = \left(\frac{1}{4}\right)^x$$

$$\text{Now, find } P(X = x|Y > X) = \frac{P(X=x, Y>x)}{P(Y>X)} = \frac{(1/4)^x}{1/3} = 3\left(\frac{1}{4}\right)^x$$

Finally, calculate the conditional expectation:

$$E[X|Y > X] = \sum_{x=1}^{\infty} x P(X = x|Y > X) = \sum_{x=1}^{\infty} x \cdot 3\left(\frac{1}{4}\right)^x = 3 \sum_{x=1}^{\infty} x\left(\frac{1}{4}\right)^x$$

Using the formula  $\sum_{x=1}^{\infty} x r^x = \frac{r}{(1-r)^2}$  with  $r = 1/4$ :

$$\sum_{x=1}^{\infty} x\left(\frac{1}{4}\right)^x = \frac{1/4}{(1-1/4)^2} = \frac{1/4}{(3/4)^2} = \frac{1/4}{9/16} = \frac{1}{4} \times \frac{16}{9} = \frac{4}{9}$$

$$E[X|Y > X] = 3 \times \frac{4}{9} = \frac{12}{9} = \frac{4}{3}$$

Final Answer: The final answer is 4/3

问题要求计算如果太阳的大小不变, 但每单位体积产生的功率与人类相同, 太阳的表面温度会是多少。

首先, 我们找到人类的功率密度。

人类产生的功率,  $P_{human} = 100 \text{ W}$

人类的体积,  $V_{human} = 66400 \text{ cm}^3$

将体积转换为立方米:

$$1 \text{ m} = 100 \text{ cm}$$

$$1 \text{ m}^3 = (100 \text{ cm})^3 = 10^6 \text{ cm}^3$$

$$V_{human} = 66400 \text{ cm}^3 = 66400 \times 10^{-6} \text{ m}^3 = 0.0664 \text{ m}^3$$

$$\text{人类的功率密度, } \rho_{power,human} = \frac{P_{human}}{V_{human}} = \frac{100 \text{ W}}{0.0664 \text{ m}^3}$$

$$\rho_{power,human} = 1506.024 \text{ W/m}^3$$

现在, 考虑太阳的大小保持不变, 但功率密度与人类相同的情况。

我们需要太阳的半径来计算其体积。太阳的平均半径是  $R_{sun} = 6.95 \times 10^8 \text{ m}$ 。

$$\text{太阳的体积, } V_{sun} = \frac{4}{3}\pi R_{sun}^3 = \frac{4}{3}\pi(6.95 \times 10^8 \text{ m})^3$$

$$V_{sun} = \frac{4}{3}\pi(337.159375 \times 10^{24}) \text{ m}^3$$

$$V_{sun} = 1410.24 \times 10^{24} \times \pi \text{ m}^3$$

$$V_{sun} = 4.4307 \times 10^{27} \text{ m}^3$$

如果太阳的功率密度与人类相同, 这种修改后的太阳产生的总功率为:

$$P_{modified\_sun} = \rho_{power,human} \times V_{sun}$$

$$P_{modified\_sun} = 1506.024 \text{ W/m}^3 \times 4.4307 \times 10^{27} \text{ m}^3$$

$$P_{modified\_sun} = 6.6726 \times 10^{30} \text{ W}$$

Table 10. s1K sample stemming from the astronomy subset of OlympicArena. The black text is the question and the blue text is the solution. Based solely on the question we generate the light blue reasoning trace and blue response with Gemini, which we use for training.

An often-repeated fun fact is that humans produce more power per unit volume than stars. If the sun were the same size, but it produced the same amount of power per unit volume as a human, what would its surface temperature be? Assume the "average human" produces 100 watts of power and has a volume of 66400 cubic centimeters.

- A: 3500 K
- B: 10000 K
- C: 25000 K
- D: 40000 K
- E: 50000 K

Using the numbers from the problem, the average human produces

$$u = \frac{100}{66400 \times 10^{-6}} = 1506 \text{ W/m}^3$$

The volume of the sun is  $\frac{4}{3}\pi R_\odot^3$ , so its new power output would be  $P = \frac{4}{3}\pi R_\odot^3 u$ . To find the equilibrium temperature, we use

$$P = \sigma AT^4 = 4\pi\sigma R_\odot^2 T^4$$

Solving for  $T$ , we get

$$T = \sqrt[4]{\frac{R_\odot u}{3\sigma}} = \sqrt[4]{\frac{(6.96 \times 10^8)(1506)}{3(5.67 \times 10^{-8})}} = 49823 \approx 50000 \text{ K.}$$

恒星的表面温度与其光度（总功率输出）通过斯特藩-玻尔兹曼定律相关：

$$L = 4\pi R^2 \sigma T^4$$

其中  $L$  是光度（总功率输出）， $R$  是恒星的半径， $\sigma$  是斯特藩-玻尔兹曼常数， $T$  是表面温度。

在这种情况下，修改后的太阳的光度为  $P_{modified\_sun}$ 。修改后的太阳的半径与实际太阳的半径相同， $R_{sun}$ 。斯特藩-玻尔兹曼常数， $\sigma = 5.67 \times 10^{-8} \text{ Wm}^{-2}\text{K}^{-4}$ 。

$$P_{modified\_sun} = 4\pi R_{sun}^2 \sigma T^4$$

$$6.6726 \times 10^{30} \text{ W} = 4\pi(6.95 \times 10^8 \text{ m})^2 \times 5.67 \times 10^{-8} \text{ Wm}^{-2}\text{K}^{-4} \times T^4$$

计算项  $4\pi R_{sun}^2 \sigma$ :

$$R_{sun}^2 = (6.95 \times 10^8)^2 = 48.3025 \times 10^{16} \text{ m}^2$$

$$4\pi R_{sun}^2 = 4\pi \times 48.3025 \times 10^{16} = 607.08 \times 10^{16} \text{ m}^2$$

$$4\pi R_{sun}^2 \sigma = 607.08 \times 10^{16} \text{ m}^2 \times 5.67 \times 10^{-8} \text{ Wm}^{-2}\text{K}^{-4}$$

$$4\pi R_{sun}^2 \sigma = 3441.93 \times 10^8 \text{ WK}^{-4} = 3.44193 \times 10^{11} \text{ WK}^{-4}$$

现在，将这个结果代入方程：

$$6.6726 \times 10^{30} = 3.44193 \times 10^{11} \times T^4$$

解出  $T^4$ :

$$T^4 = \frac{6.6726 \times 10^{30}}{3.44193 \times 10^{11}} = 1.9386 \times 10^{19}$$

解出  $T$ :

$$T = (1.9386 \times 10^{19})^{1/4} = (193.86 \times 10^{17})^{1/4}$$

$$T = (1.9386)^{1/4} \times (10^{19})^{1/4} = (1.9386)^{0.25} \times 10^{4.75}$$

计算  $(1.9386)^{0.25}$ :

$$\text{令 } y = 1.9386^{0.25}, \text{ 则 } \ln y = 0.25 \ln(1.9386)$$

$$\ln(1.9386) \approx 0.6619$$

$$\ln y = 0.25 \times 0.6619 = 0.165475$$

$$y = e^{0.165475} \approx 1.1799$$

$$\text{因此, } T = 1.1799 \times 10^{4.75} = 1.1799 \times 10^4 \times 10^{0.75}$$

$$10^{0.75} = 10^{3/4} = (10^3)^{1/4} = (1000)^{1/4}$$

我们知道  $5^4 = 625$  和  $6^4 = 1296$ ，所以  $(1000)^{1/4}$  在 5 和 6 之间。

$$\text{或者, } 10^{0.75} = 10^{0.5} \times 10^{0.25} = \sqrt{10} \times 10^{0.25} \approx 3.16 \times 1.778 = 5.62$$

$$T = 1.1799 \times 10^4 \times 5.62 = 6.63 \times 10^4 \text{ K} = 66300 \text{ K}$$

让我们检查一下计算。

人类的功率密度:  $\rho_{power,human} = 1506.02 \text{ W/m}^3$

太阳的体积:  $V_{sun} = 4.43 \times 10^{27} \text{ m}^3$

修改后的太阳的功率:  $P_{modified\_sun} = 1506.02 \times 4.43 \times 10^{27} = 6671.67 \times 10^{27} = 6.67 \times 10^{30} \text{ W}$

斯特藩-玻尔兹曼定律:  $P = 4\pi R^2 \sigma T^4$

$$T^4 = \frac{P}{4\pi R^2 \sigma}$$

$$4\pi R_{sun}^2 \sigma = 3.44 \times 10^{11}$$

$$T^4 = \frac{6.67 \times 10^{30}}{3.44 \times 10^{11}} = 1.939 \times 10^{19}$$

$$T = (1.939 \times 10^{19})^{1/4} = (193.9 \times 10^{17})^{1/4}$$

$$T = (193.9)^{0.25} \times (10^{17})^{0.25} = (193.9)^{0.25} \times 10^{4.25}$$

计算  $(193.9)^{0.25}$ :

$$\text{令 } y = 193.9^{0.25}, \text{ 则 } \ln y = 0.25 \ln(193.9)$$

$$\ln(193.9) \approx 5.267 \ln y = 0.25 \times 5.267 = 1.31675$$

$$y = e^{1.31675} \approx 3.73$$

$$T = 3.73 \times 10^{4.25} = 3.73 \times 10^4 \times 10^{0.25} = 37300 \times 1.778 = 66329 \text{ K}$$

结果大约是 66000 K。这不在选项中。让我们回顾假设和计算。

假设：太阳像人一样行动。这对于讨论表面温度是合理的。

The problem asks for the surface temperature of the sun if it had the same size but produced power per unit volume as a human.

First, let's find the power density of a human.

Power produced by a human,  $P_{\text{human}} = 100 \text{ W}$

Volume of a human,  $V_{\text{human}} = 66400 \text{ cm}^3$

Convert the volume to cubic meters:

$$1 \text{ m} = 100 \text{ cm}$$

$$1 \text{ m}^3 = (100 \text{ cm})^3 = 10^6 \text{ cm}^3$$

$$V_{\text{human}} = 66400 \text{ cm}^3 = 66400 \times 10^{-6} \text{ m}^3 = 0.0664 \text{ m}^3$$

$$\text{Power density of a human, } \rho_{\text{power,human}} = \frac{P_{\text{human}}}{V_{\text{human}}} = \frac{100 \text{ W}}{0.0664 \text{ m}^3}$$

$$\rho_{\text{power,human}} = 1506.024 \text{ W/m}^3$$

Now, consider the sun with the same size as it currently is, but with a power density equal to that of a human.

We need the radius of the sun to calculate its volume. The average radius of the sun is  $R_{\text{sun}} = 6.95 \times 10^8 \text{ m}$ .

$$\text{Volume of the sun, } V_{\text{sun}} = \frac{4}{3}\pi R_{\text{sun}}^3 = \frac{4}{3}\pi(6.95 \times 10^8 \text{ m})^3$$

$$V_{\text{sun}} = \frac{4}{3}\pi(337.159375 \times 10^{24}) \text{ m}^3$$

$$V_{\text{sun}} = 1410.24 \times 10^{24} \times \pi \text{ m}^3$$

$$V_{\text{sun}} = 4.4307 \times 10^{27} \text{ m}^3$$

If the sun had the same power density as a human, the total power produced by this modified sun would be:

$$P_{\text{modified\_sun}} = \rho_{\text{power,human}} \times V_{\text{sun}}$$

$$P_{\text{modified\_sun}} = 1506.024 \text{ W/m}^3 \times 4.4307 \times 10^{27} \text{ m}^3$$

$$P_{\text{modified\_sun}} = 6.6726 \times 10^{30} \text{ W}$$

考虑计算中的潜在错误。

让我们尝试通过设置比例来采用更简单的方法。

修改后的太阳的单位体积功率等于人类的单位体积功率。

$$\frac{P_{\text{modified\_sun}}}{V_{\text{sun}}} = \frac{P_{\text{human}}}{V_{\text{human}}}$$

$$\text{我们还知道 } P_{\text{modified\_sun}} = 4\pi R_{\text{sun}}^2 \sigma T_{\text{modified\_sun}}^4$$

$$\frac{4\pi R_{\text{sun}}^2 \sigma T_{\text{modified\_sun}}^4}{\frac{4}{3}\pi R_{\text{sun}}^3} = \frac{P_{\text{human}}}{V_{\text{human}}}$$

$$\frac{3\sigma T_{\text{modified\_sun}}^4}{R_{\text{sun}}} = \rho_{\text{power,human}}$$

$$T_{\text{modified\_sun}}^4 = \frac{\rho_{\text{power,human}} R_{\text{sun}}}{3\sigma}$$

代入值:

$$\rho_{\text{power,human}} = 1506.02 \text{ W/m}^3$$

$$R_{\text{sun}} = 6.95 \times 10^8 \text{ m}$$

$$\sigma = 5.67 \times 10^{-8} \text{ Wm}^{-2}\text{K}^{-4}$$

$$T_{\text{modified\_sun}}^4 = \frac{1506.02 \times 6.95 \times 10^8}{3 \times 5.67 \times 10^{-8}}$$

$$T_{\text{modified\_sun}}^4 = \frac{10466.839 \times 10^8}{17.01 \times 10^{-8}} = \frac{1.0466 \times 10^{12}}{1.701 \times 10^{-7}} = 0.6153 \times 10^{19} = 6.153 \times 10^{18}$$

$$T_{\text{modified\_sun}} = (6.153 \times 10^{18})^{1/4} = (6153 \times 10^{15})^{1/4}$$

$$T_{\text{modified\_sun}} = (6153)^{0.25} \times (10^{15})^{0.25} = (6153)^{0.25} \times 10^{3.75}$$

计算  $(6153)^{0.25}$ :

$$\text{设 } y = 6153^{0.25}, \text{ 则 } \ln y = 0.25 \ln(6153)$$

$$\ln(6153) \approx 8.724$$

$$\ln y = 0.25 \times 8.724 = 2.181$$

$$y = e^{2.181} \approx 8.85$$

$$T_{\text{modified\_sun}} = 8.85 \times 10^{3.75} = 8.85 \times 10^3 \times 10^{0.75} = 8850 \times 5.62 = 49737 \text{ K}$$

这接近选项 E: 50000 K。

让我们仔细检查单位和计算。

$$\text{人类的功率密度: } \frac{100 \text{ W}}{0.0664 \text{ m}^3} = 1506.02 \text{ W/m}^3$$

$$\text{推导出的公式是 } T^4 = \frac{\rho_{\text{power,human}} R_{\text{sun}}}{3\sigma}.$$

单位检查:

$$[\rho_{\text{power,human}}] = \text{W/m}^3$$

$$[R_{\text{sun}}] = \text{m}$$

$$[\sigma] = \text{Wm}^{-2}\text{K}^{-4} \frac{(\text{W/m}^3) \times \text{m}}{\text{Wm}^{-2}\text{K}^{-4}} = \frac{\text{W/m}^2}{\text{Wm}^{-2}\text{K}^{-4}} = \frac{\text{W}}{\text{m}^2} \times \frac{\text{m}^2\text{K}^4}{\text{W}} = \text{K}^4$$

单位与  $T^4$  匹配。

数值计算检查:

$$\text{分子: } 1506.02 \times 6.95 \times 10^8 = 10466.839 \times 10^8 = 1.0466 \times 10^{12}$$

$$\text{分母: } 3 \times 5.67 \times 10^{-8} = 17.01 \times 10^{-8}$$

$$\text{比值: } \frac{1.0466 \times 10^{12}}{17.01 \times 10^{-8}} = 0.061528 \times 10^{20} = 6.1528 \times 10^{18}$$

$$T^4 = 6.1528 \times 10^{18}$$

$$T = (6.1528 \times 10^{18})^{1/4} = (6152.8 \times 10^{15})^{1/4}$$

$$T = (6152.8)^{0.25} \times 10^{3.75}$$

让我们使用对数以提高精度。

$$\log_{10}(6152.8) \approx 3.789$$

$$0.25 \times 3.789 = 0.94725$$

$$10^{0.94725} \approx 8.856$$

$$T = 8.856 \times 10^{3.75} = 8.856 \times 10^3 \times 10^{0.75} = 8856 \times 5.623 = 49806 \text{ K}$$

这非常接近 50000 K。

The surface temperature of a star is related to its luminosity (total power output) through the Stefan-Boltzmann law:

$$L = 4\pi R^2 \sigma T^4$$

where  $L$  is the luminosity (total power output),  $R$  is the radius of the star,  $\sigma$  is the Stefan-Boltzmann constant, and  $T$  is the surface temperature.

In this case, the luminosity of the modified sun is  $P_{modified\_sun}$ . The radius of the modified sun is the same as the actual sun,  $R_{sun}$ .

The Stefan-Boltzmann constant,  $\sigma = 5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$ .

$$P_{modified\_sun} = 4\pi R_{sun}^2 \sigma T^4$$

$$6.6726 \times 10^{30} \text{ W} = 4\pi(6.95 \times 10^8 \text{ m})^2 \times 5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4} \times T^4$$

Calculate the term  $4\pi R_{sun}^2 \sigma$ :

$$R_{sun}^2 = (6.95 \times 10^8)^2 = 48.3025 \times 10^{16} \text{ m}^2$$

$$4\pi R_{sun}^2 = 4\pi \times 48.3025 \times 10^{16} = 607.08 \times 10^{16} \text{ m}^2$$

$$4\pi R_{sun}^2 \sigma = 607.08 \times 10^{16} \text{ m}^2 \times 5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$$

$$4\pi R_{sun}^2 \sigma = 3441.93 \times 10^8 \text{ WK}^{-4} = 3.44193 \times 10^{11} \text{ WK}^{-4}$$

Now, substitute this back into the equation:

$$6.6726 \times 10^{30} = 3.44193 \times 10^{11} \times T^4$$

Solve for  $T^4$ :

$$T^4 = \frac{6.6726 \times 10^{30}}{3.44193 \times 10^{11}} = 1.9386 \times 10^{19}$$

Solve for  $T$ :

$$T = (1.9386 \times 10^{19})^{1/4} = (193.86 \times 10^{17})^{1/4}$$

$$T = (1.9386)^{1/4} \times (10^{19})^{1/4} = (1.9386)^{0.25} \times 10^{4.75}$$

Calculate  $(1.9386)^{0.25}$ :

Let  $y = 1.9386^{0.25}$ , then  $\ln y = 0.25 \ln(1.9386)$

$$\ln(1.9386) \approx 0.6619$$

$$\ln y = 0.25 \times 0.6619 = 0.165475$$

$$y = e^{0.165475} \approx 1.1799$$

$$\text{So, } T = 1.1799 \times 10^{4.75} = 1.1799 \times 10^4 \times 10^{0.75}$$

$$10^{0.75} = 10^{3/4} = (10^3)^{1/4} = (1000)^{1/4}$$

We know  $5^4 = 625$  and  $6^4 = 1296$ , so  $(1000)^{1/4}$  is between 5 and 6.

$$\text{Alternatively, } 10^{0.75} = 10^{0.5} \times 10^{0.25} = \sqrt{10} \times 10^{0.25} \approx 3.16 \times 1.778 = 5.62$$

$$T = 1.1799 \times 10^4 \times 5.62 = 6.63 \times 10^4 \text{ K} = 66300 \text{ K}$$

Let's double-check the calculations.

Power density of human:  $\rho_{power,human} = 1506.02 \text{ W/m}^3$

Volume of sun:  $V_{sun} = 4.43 \times 10^{27} \text{ m}^3$

Power of modified sun:  $P_{modified\_sun} = 1506.02 \times 4.43 \times 10^{27} = 6671.67 \times 10^{27} = 6.67 \times 10^{30} \text{ W}$

Stefan-Boltzmann law:  $P = 4\pi R^2 \sigma T^4$

$$T^4 = \frac{P}{4\pi R^2 \sigma}$$

$$4\pi R_{sun}^2 \sigma = 3.44 \times 10^{11}$$

$$T^4 = \frac{6.67 \times 10^{30}}{3.44 \times 10^{11}} = 1.939 \times 10^{19}$$

$$T = (1.939 \times 10^{19})^{1/4} = (193.9 \times 10^{17})^{1/4}$$

$$T = (193.9)^{0.25} \times (10^{17})^{0.25} = (193.9)^{0.25} \times 10^{4.25}$$

Calculate  $(193.9)^{0.25}$ :

Let  $y = 193.9^{0.25}$ , then  $\ln y = 0.25 \ln(193.9)$

$$\ln(193.9) \approx 5.267$$

$$\rho_{power} = \frac{L}{V}$$

$$L = 4\pi R^2 \sigma T^4$$

$$V = \frac{4}{3}\pi R^3$$

$$\rho_{power} = \frac{4\pi R^2 \sigma T^4}{\frac{4}{3}\pi R^3} = \frac{3\sigma T^4}{R}$$

如果修改后的太阳具有人类的功率密度:

$$\rho_{power,human} = \frac{3\sigma T_{modified\_sun}^4}{R_{sun}}$$

$$T_{modified\_sun}^4 = \frac{\rho_{power,human} R_{sun}}{3\sigma}$$

公式的推导似乎正确。

让我们使用近似值来检查数量级。

$$\rho_{power,human} \approx 10^3 \text{ W/m}^3$$

$$R_{sun} \approx 7 \times 10^8 \text{ m}$$

$$\sigma \approx 6 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$$

$$T^4 \approx \frac{10^3 \times 7 \times 10^8}{3 \times 6 \times 10^{-8}} = \frac{7 \times 10^{11}}{18 \times 10^{-8}} = 0.38 \times 10^{19} = 3.8 \times 10^{18}$$

$$T \approx (3.8 \times 10^{18})^{1/4} = (3800 \times 10^{15})^{1/4}$$

$$T \approx (3800)^{0.25} \times 10^{3.75}$$

$(3800)^{0.25} \cdot 7^4 = 2401, 8^4 = 4096$ 。因此，介于7和8之间。

$$\approx 7.8$$

$$T \approx 7.8 \times 10^{3.75} = 7800 \times 5.6 = 43680 \text{ K}$$

数量级与详细计算一致。

问题陈述是直接的，似乎没有隐藏的复杂性。所做的假设对于此类问题来说是标准的。考虑给定值的精度。

人类的功率: 100 W (精确值)

人类的体积: 66400 cm<sup>3</sup> (精确值)

获得的结果非常接近选项E。鉴于选择题的性质，最接近的选项很可能是正确答案。

如果对问题有任何误解，可能是在设定上。然而，短语“如果太阳的大小相同，但每单位体积产生的功率与人类相同”似乎很清楚。

表面温度是使用斯特藩-玻尔兹曼定律计算的，假设为黑体辐射。

再次检查提供的选项:

A: 3500 K

B: 10000 K

C: 25000 K

D: 40000 K

E: 50000 K

我们计算的值约为 49800 K，最接近 50000 K。

如果我们使用更精确的常数值，结果可能会略有不同，但可能不足以转移到另一个选项。

问题中提到的趣闻表明，人类的功率密度高于太阳。让我们验证这一点。

太阳的功率输出:  $3.846 \times 10^{26} \text{ W}$

太阳的体积:  $4.43 \times 10^{27} \text{ m}^3$

太阳的功率密度:  $\frac{3.846 \times 10^{26}}{4.43 \times 10^{27}} = 0.868 \text{ W/m}^3$

人类的功率密度:  $1506 \text{ W/m}^3$

确实，人类的功率密度显著高于太阳。

最后检查计算步骤和逻辑。将修改后的太阳的功率密度等同于人类的功率密度，然后使用斯特藩-玻尔兹曼定律的方法似乎是正确的。

Consider potential errors in calculations.

Let's try a simpler approach by setting up ratios.

The power per unit volume of the modified sun is equal to that of a human.

$$\frac{P_{\text{modified\_sun}}}{V_{\text{sun}}} = \frac{P_{\text{human}}}{V_{\text{human}}}$$

We also know  $P_{\text{modified\_sun}} = 4\pi R_{\text{sun}}^2 \sigma T_{\text{modified\_sun}}^4$ .

$$\frac{4\pi R_{\text{sun}}^2 \sigma T_{\text{modified\_sun}}^4}{\frac{4}{3}\pi R_{\text{sun}}^3} = \frac{P_{\text{human}}}{V_{\text{human}}}$$

$$\frac{3\sigma T_{\text{modified\_sun}}^4}{R_{\text{sun}}} = \rho_{\text{power,human}}$$

$$T_{\text{modified\_sun}}^4 = \frac{\rho_{\text{power,human}} R_{\text{sun}}}{3\sigma}$$

Substitute the values:

$$\rho_{\text{power,human}} = 1506.02 \text{ W/m}^3$$

$$R_{\text{sun}} = 6.95 \times 10^8 \text{ m}$$

$$\sigma = 5.67 \times 10^{-8} \text{ Wm}^{-2}\text{K}^{-4}$$

$$T_{\text{modified\_sun}}^4 = \frac{1506.02 \times 6.95 \times 10^8}{3 \times 5.67 \times 10^{-8}}$$

$$T_{\text{modified\_sun}}^4 = \frac{10466.839 \times 10^8}{17.01 \times 10^{-8}} = \frac{1.0466 \times 10^{12}}{1.701 \times 10^{-7}} = 0.6153 \times 10^{19} = 6.153 \times 10^{18}$$

$$T_{\text{modified\_sun}} = (6.153 \times 10^{18})^{1/4} = (6153 \times 10^{15})^{1/4}$$

$$T_{\text{modified\_sun}} = (6153)^{0.25} \times (10^{15})^{0.25} = (6153)^{0.25} \times 10^{3.75}$$

Calculate  $(6153)^{0.25}$ :

Let  $y = 6153^{0.25}$ , then  $\ln y = 0.25 \ln(6153)$

$$\ln(6153) \approx 8.724$$

$$\ln y = 0.25 \times 8.724 = 2.181$$

$$y = e^{2.181} \approx 8.85$$

$$T_{\text{modified\_sun}} = 8.85 \times 10^{3.75} = 8.85 \times 10^3 \times 10^{0.75} = 8850 \times 5.62 = 49737 \text{ K}$$

This is close to option E: 50000 K.

Let's re-check the units and calculations carefully.

$$\text{Power density of human: } \frac{100 \text{ W}}{0.0664 \text{ m}^3} = 1506.02 \text{ W/m}^3$$

$$\text{The formula derived was } T^4 = \frac{\rho_{\text{power,human}} R_{\text{sun}}}{3\sigma}.$$

Units check:

$$[\rho_{\text{power,human}}] = \text{W/m}^3$$

$$[R_{\text{sun}}] = \text{m}$$

$$[\sigma] = \text{Wm}^{-2}\text{K}^{-4}$$

$$\frac{(\text{W/m}^3) \times \text{m}}{\text{Wm}^{-2}\text{K}^{-4}} = \frac{\text{W}/\text{m}^2}{\text{Wm}^{-2}\text{K}^{-4}} = \frac{\text{W}}{\text{m}^2} \times \frac{\text{m}^2\text{K}^4}{\text{W}} = \text{K}^4$$

The units match for  $T^4$ .

Numerical calculation check:

$$\text{Numerator: } 1506.02 \times 6.95 \times 10^8 = 10466.839 \times 10^8 = 1.0466 \times 10^{12}$$

$$\text{Denominator: } 3 \times 5.67 \times 10^{-8} = 17.01 \times 10^{-8}$$

$$\text{Ratio: } \frac{1.0466 \times 10^{12}}{17.01 \times 10^{-8}} = 0.061528 \times 10^{20} = 6.1528 \times 10^{18}$$

$$T^4 = 6.1528 \times 10^{18}$$

$$T = (6.1528 \times 10^{18})^{1/4} = (6152.8 \times 10^{15})^{1/4}$$

$$T = (6152.8)^{0.25} \times 10^{3.75}$$

Let's use logarithms for precision.

$$\log_{10}(6152.8) \approx 3.789$$

$$0.25 \times 3.789 = 0.94725$$

$$10^{0.94725} \approx 8.856$$

$$T = 8.856 \times 10^{3.75} = 8.856 \times 10^3 \times 10^{0.75} = 8856 \times 5.623 = 49806 \text{ K}$$

解:

一个普通人类产生的功率为  $P_{\text{human}} = 100 \text{ W}$ , 普通人类的体积为  $V_{\text{human}} = 66400 \text{ cm}^3 = 0.0664 \text{ m}^3$ 。

$$\text{人类的功率密度为 } \rho_{\text{power,human}} = \frac{P_{\text{human}}}{V_{\text{human}}} = \frac{100 \text{ W}}{0.0664 \text{ m}^3} \approx 1506.02 \text{ W/m}^3.$$

如果太阳的大小相同, 但每单位体积产生的功率与人类相同, 其功率密度将为  $\rho_{\text{power,modified\_sun}} = \rho_{\text{power,human}}$ 。

$$\text{太阳的半径为 } R_{\text{sun}} = 6.95 \times 10^8 \text{ m}.$$

$$\text{太阳的体积为 } V_{\text{sun}} = \frac{4}{3}\pi R_{\text{sun}}^3.$$

$$\text{修改后的太阳的功率密度由 } \rho_{\text{power,modified\_sun}} = \frac{P_{\text{modified\_sun}}}{V_{\text{sun}}} \text{ 给出。}$$

我们还知道  $P_{\text{modified\_sun}} = 4\pi R_{\text{sun}}^2 \sigma T^4$ , 其中  $T$  是修改后的太阳的表面温度。

等同功率密度:

$$\rho_{\text{power,human}} = \frac{4\pi R_{\text{sun}}^2 \sigma T^4}{\frac{4}{3}\pi R_{\text{sun}}^3} = \frac{3\sigma T^4}{R_{\text{sun}}}$$

解  $T^4$ :

$$T^4 = \frac{\rho_{\text{power,human}} R_{\text{sun}}}{3\sigma}$$

代入值:

$$T^4 = \frac{1506.02 \text{ W/m}^3 \times 6.95 \times 10^8 \text{ m}}{3 \times 5.67 \times 10^{-8} \text{ Wm}^{-2}\text{K}^{-4}}$$

$$T^4 = \frac{1.0466 \times 10^{12}}{1.701 \times 10^{-7}} = 6.1528 \times 10^{18} \text{ K}^4$$

$$\text{解 } T: T = (6.1528 \times 10^{18})^{1/4} = (6152.8)^{0.25} \times 10^{18/4} = (6152.8)^{0.25} \times 10^{4.5}$$

使用对数:

$$\log_{10}(6152.8) \approx 3.789$$

$$0.25 \times 3.789 = 0.94725$$

$$(6152.8)^{0.25} = 10^{0.94725} \approx 8.856$$

$$T = 8.856 \times 10^{4.5} = 8.856 \times 10^4 \times 10^{0.5} = 88560 \times \sqrt{10} \approx 88560 \times 3.16 = 280000 \text{ K}$$

让我们重新计算指数:  $18/4 = 4.5$ 。

$T = 8.856 \times 10^{4.5} = 8.856 \times 10^4 \times 10^{0.5}$  - 前一步骤中有错误。

$$T = (6.1528 \times 10^{18})^{0.25} = 49806 \text{ K}$$

表面温度大约为 49806 K, 最接近 50000 K。

最终答案: 最终答案是 E

$$\rho_{power} = \frac{L}{V}$$

$$L = 4\pi R^2 \sigma T^4$$

$$V = \frac{4}{3}\pi R^3$$

$$\rho_{power} = \frac{4\pi R^2 \sigma T^4}{\frac{4}{3}\pi R^3} = \frac{3\sigma T^4}{R}$$

If the modified sun has the power density of a human:

$$\rho_{power,human} = \frac{3\sigma T^4}{R_{sun}}$$

$$T_{modified\_sun}^4 = \frac{\rho_{power,human} R_{sun}}{3\sigma}$$

The formula derivation seems correct.

Let's use approximate values to check the order of magnitude.

$$\rho_{power,human} \approx 10^3 \text{ W/m}^3$$

$$R_{sun} \approx 7 \times 10^8 \text{ m}$$

$$\sigma \approx 6 \times 10^{-8} \text{ Wm}^{-2}\text{K}^{-4}$$

$$T^4 \approx \frac{10^3 \times 7 \times 10^8}{3 \times 6 \times 10^{-8}} = \frac{7 \times 10^{11}}{18 \times 10^{-8}} = 0.38 \times 10^{19} = 3.8 \times 10^{18}$$

$$T \approx (3.8 \times 10^{18})^{1/4} = (3800 \times 10^{15})^{1/4}$$

$$T \approx (3800)^{0.25} \times 10^{3.75}$$

$(3800)^{0.25}$ :  $7^4 = 2401$ ,  $8^4 = 4096$ . So, between 7 and 8.

$\approx 7.8$

$$T \approx 7.8 \times 10^{3.75} = 7800 \times 5.6 = 43680 \text{ K}$$

The order of magnitude seems consistent with the detailed calculation.

The problem statement is straightforward and doesn't seem to have any hidden complexities. The assumptions made are standard for such problems.

Consider the precision of the given values.

Power of human: 100 W (exact)

Volume of human: 66400 cm<sup>3</sup> (exact)

The result obtained is very close to option E. Given the nature of multiple-choice questions, the closest option is likely the intended answer.

If there were any misinterpretations of the question, it would likely be in the setup. However, the phrase "If the sun were the same size, but it produced the same amount of power per unit volume as a human" seems clear.

The surface temperature is calculated using the Stefan-Boltzmann law, assuming black-body radiation.

Double-check the provided options:

A: 3500 K

B: 10000 K

C: 25000 K

D: 40000 K

E: 50000 K

Our calculated value is approximately 49800 K, which is closest to 50000 K.

If we used more precise values for the constants, the result might be slightly different, but likely not enough to shift to another option.

The fun fact mentioned in the question implies that the power density of a human is higher than that of the sun. Let's verify this.

Power output of the sun:  $3.846 \times 10^{26}$  W

42

Volume of the sun:  $4.43 \times 10^{27}$  m<sup>3</sup>

Power density of the sun:  $\frac{3.846 \times 10^{26}}{4.43 \times 10^{27}} = 0.868 \text{ W/m}^3$

Power density of human: 1506 W/m<sup>3</sup>

## D. Test-time scaling details

### D.1. Sequential scaling ablations

<lim_start>user	<lim_start>user
What is the answer to Life, the Universe and Everything?	What is the answer to Life, the Universe and Everything?
Think for up to 2048 tokens.	Think for up to 64 steps.
<lim_start>assistant	<lim_start>assistant
<lim_start>think	<lim_start>64 steps left
Let me break down this question into the three parts it is asking for: 1) Life 2) Universe 3) Everything	Let me break down this question into the three parts it is asking for: 1) Life 2) Universe 3) Everything
Let me start with life...	Let me start with life...
<lim_start>answer	<lim_start>answer
The answer is...	The answer is...

Figure 10. 用于控制测试时计算的令牌和步骤指令数据格式。我们仅在 推理轨迹 和 答案 上训练我们的模型。

Table 11. 通过令牌条件控制扩展思考时间。所有指标均取自 AIME24 中的 30 个问题的平均值。

Tokens instructed ( $\rightarrow$ )	1024	2048	4096	8192	16384
No intervention at test-time					
Thinking tokens	7939	7158	8263	7108	7500
Answer tokens	689	669	659	722	724
AIME24	26.7	30.0	33.3	33.3	40.0
Forcing end of thinking when token budget is reached					
Thinking tokens	1024	2048	4031	5664	6330
Answer tokens	15	15	142	722	691
AIME24	3.3	30.0	33.3	33.3	40.0

**基于标记的条件控制** 一种通用的方法是在提示中明确告诉模型应该生成多少个标记。理想情况下，模型可以跟踪其标记计数，并调整其生成以在所需的限制内完成。我们通过使用 Figure 10 (左) 中的格式，通过带有标记指令训练模型来实验这种方法。我们将1,000个训练示例中的推理轨迹长度分桶为2的幂（向上取整），并在用户提示中添加相应的指令。例如，如果指令说“思考最多2048个标记”，那么推理轨迹将有1024到2048个标记。在 Table 11 中，我们显示训练后的模型几乎不遵循标记指令。当给定更高的限制时，它有时会生成更多的标记，但经常超过限制。这可能不仅限于我们的模型，因为先前的研究表明，OpenAI o1-mini 也无法遵循标记指令 (Zhang & Chen, 2024)。为了防止超过限制，我们测试了预算强制，即一旦达到限制就结束思考。这导致了完美的控制 (Table 11 (下))。通过预算强制，当给定较小的思考预算时，模型不再超过限制，因此趋势更加明显。这导致 Table 3 中 标记提示 + 预算强制 的测试时间缩放值更好。为了计算 Table 3 中报告的控制值，对于基于标记的条件控制变体，我们将 Table 11 中思考标记少于上限的次数除以总评估次数（无干预为2/5；有干预为5/5）。

**步骤条件控制** 令牌指令失败，因为当前模型无法计数令牌。为了适应这种能力的缺乏，我们尝试使计数更加粗略。我们将推理轨迹划分为步骤，并要求模型思考特定数量的步骤而不是令牌。我们根据手动

Solution:

The power produced by an average human is  $P_{\text{human}} = 100 \text{ W}$ , and the volume of an average human is  $V_{\text{human}} = 66400 \text{ cm}^3 = 0.0664 \text{ m}^3$ .

The power density of a human is  $\rho_{\text{power,human}} = \frac{P_{\text{human}}}{V_{\text{human}}} = \frac{100 \text{ W}}{0.0664 \text{ m}^3} \approx 1506.02 \text{ W/m}^3$ .

If the sun were the same size but produced the same amount of power per unit volume as a human, its power density would be  $\rho_{\text{power,modified_sun}} = \rho_{\text{power,human}}$ .

The radius of the sun is  $R_{\text{sun}} = 6.95 \times 10^8 \text{ m}$ .

The volume of the sun is  $V_{\text{sun}} = \frac{4}{3}\pi R_{\text{sun}}^3$ .

The power density of the modified sun is given by  $\rho_{\text{power,modified_sun}} = \frac{P_{\text{modified_sun}}}{V_{\text{sun}}}$ .

We also know that  $P_{\text{modified_sun}} = 4\pi R_{\text{sun}}^2 \sigma T^4$ , where  $T$  is the surface temperature of the modified sun.

Equating the power density:

$$\rho_{\text{power,human}} = \frac{\frac{4\pi R_{\text{sun}}^2 \sigma T^4}{3\pi R_{\text{sun}}^3}}{R_{\text{sun}}} = \frac{3\sigma T^4}{R_{\text{sun}}}$$

Solving for  $T^4$ :

$$T^4 = \frac{\rho_{\text{power,human}} R_{\text{sun}}}{3\sigma}$$

Substitute the values:

$$T^4 = \frac{1506.02 \text{ W/m}^3 \times 6.95 \times 10^8 \text{ m}}{3 \times 5.67 \times 10^{-8} \text{ Wm}^{-2}\text{K}^{-4}}$$

$$T^4 = \frac{1.0466 \times 10^{12}}{1.701 \times 10^{-7}} = 6.1528 \times 10^{18} \text{ K}^4$$

Solving for  $T$ :

$$T = (6.1528 \times 10^{18})^{1/4} = (6152.8)^{0.25} \times 10^{18/4} = (6152.8)^{0.25} \times 10^{4.5}$$

Using logarithms:

$$\log_{10}(6152.8) \approx 3.789$$

$$0.25 \times 3.789 = 0.94725$$

$$(6152.8)^{0.25} = 10^{0.94725} \approx 8.856$$

$$T = 8.856 \times 10^{4.5} = 8.856 \times 10^4 \times 10^{0.5} = 88560 \times \sqrt{10} \approx 88560 \times 3.16 = 280000 \text{ K}$$

Let's redo the exponent calculation:  $18/4 = 4.5$ .

$$T = 8.856 \times 10^{4.5} = 8.856 \times 10^4 \times 10^{0.5} - \text{error in previous step.}$$

$$T = (6.1528 \times 10^{18})^{0.25} = 49806 \text{ K}$$

The surface temperature would be approximately 49806 K, which is closest to 50000 K.

Final Answer: The final answer is E

Table 12. 通过步骤条件控制扩展思考时间。所有指标均取自 AIME24 中的 30 个样本的平均值。Token 计数不包括思考和步骤分隔符。

Steps instructed ( $\rightarrow$ )	16	32	64	128	256
<i>No intervention at test-time</i>					
Steps used	123	90	80	82	136
Tokens per step	60	70	69	66	56
Thinking tokens	7252	6277	5396	5552	7551
Answer tokens	665	653	735	777	754
AIME24	33.3	23.3	33.3	36.7	33.3
<i>Forcing end of thinking when 0 steps are reached</i>					
Steps used	16	32	59	78	136
Tokens per step	96	94	80	70	56
Thinking tokens	1517	2963	4636	5409	7551
Answer tokens	1111	788	799	794	754
AIME24	23.3	23.3	33.3	36.7	33.3

Table 13. 通过类别条件控制扩展思考时间。我们报告“准确率 / 每样本平均思考标记数”；准确率越高，思考标记数（推理成本）越少越好。

Prompt appended to the question after two newlines	AIME24	MATH500 GPQA
Answer after a short amount of thinking. Do not spend excessive time double-checking your work.	30.0% / 8033	90.4% / 2537
		56.6% / 4177
Answer after a long amount of thinking. If you feel like you are finished early, spend the extra time trying to double-check your work until you are absolutely sure that you have the correct answer.	36.7% / 9651	91.4% / 3875
		51.0% / 4827
Without generic prompt appending	50.0% / 6109	93.0% / 3298
		57.6% / 3510

## D. Test-time scaling details

### D.1. Sequential scaling ablations

<lim_start>user	<lim_start>user
What is the answer to Life, the Universe and Everything?	What is the answer to Life, the Universe and Everything?
Think for up to 2048 tokens.	Think for up to 64 steps.
<lim_start>assistant	<lim_start>assistant
<lim_start>think	<lim_start>64 steps left
Let me break down this question into the three parts it is asking for: 1) Life 2) Universe 3) Everything	Let me break down this question into the three parts it is asking for: 1) Life 2) Universe 3) Everything
Let me start with life...	<lim_start>63 steps left
<lim_start>answer	Let me start with life...
The answer is...	<lim_start>answer
	The answer is...

Figure 10. Token and step instruction data formats for controlling test-time compute. We only train our model on the reasoning trace and the answer.

检查样本发现的直观分隔符，将推理轨迹在双换行处分割为步骤。我们根据步骤数量将训练样本分为2的幂，并添加相应的步骤指令，格式如Figure 10（右）所示。这种格式基于早期实验，我们发现当倒数（“还剩3步...还剩2步”）而不是正数（“第2步...第3步...”）时，模型更有可能遵守步骤限制。这可能是因为如果倒数，最后一步总是1，这将作为模型完成生成的强烈先验。如果正数，答案前的最后一步会变化，因此如果模型不记得原始步骤指令，它可能会停止失败。我们从Table 12的结果中得出以下结论：(1) 模型仍然难以遵守步骤限制。有时模型会继续计数到负数步骤，例如“还剩-1步”。为了解决这个问题，我们一旦达到0步就自动停止思考过程，然后通过附加答案令牌分隔符（§3）强制模型转换到回答模式。这导致了完美的步骤遵守（Table 12的下半部分），但问题仍然存在。(2) 模型通过使每一步更长来补偿较少的步骤。例如，当被迫使用16步与256步时，模型每步生成的平均令牌数分别为96和56。尽管存在这个问题，更多的步骤仍然明显与更多的总思考令牌和更好的性能相关，导致Table 12中的正斜率(3) 步骤指令比其他方法更昂贵。步骤分隔符每个需要大约6个令牌，例如64步总共需要大约380个令牌。当在Table 12中忽略步骤分隔符的令牌计数时，模型仍然需要7551个思考令牌才能在AIME24上达到33.3%的准确率。为了计算Table 3中报告的步骤条件控制变体的控制，我们首先决定每步的上限为100个令牌，然后将这个数字乘以指示的步骤数，以得出一个代理的总令牌限制，例如16步指示的1600。然后我们检查Table 12中的思考令牌是否在每个评估运行中符合相应的限制（没有干预的3/5；有干预的5/5）。对于Figure 7中的模型，我们使用在数据的早期版本上训练的具有步骤条件控制的模型，并使用我们评估代码库的早期版本。

Table 11. Scaling thinking time via tokens-conditional control. All metrics are averaged over the 30 questions in AIME24.

Tokens instructed (→)	1024	2048	4096	8192	16384
<i>No intervention at test-time</i>					
Thinking tokens	7939	7158	8263	7108	7500
Answer tokens	689	669	659	722	724
AIME24	26.7	30.0	33.3	33.3	40.0
<i>Forcing end of thinking when token budget is reached</i>					
Thinking tokens	1024	2048	4031	5664	6330
Answer tokens	15	15	142	722	691
AIME24	3.3	30.0	33.3	33.3	40.0

**Token-conditional control** One general approach is to simply tell a model in the prompt precisely how many tokens it should generate. Ideally, the model can keep track of its token count and adjust its generation to finish within the desired limits. We experiment with this approach by training a model with token instructions using the format in Figure 10 (left). We bucket the lengths of the reasoning traces from our 1,000 training examples into powers of two (rounded upwards) and add a corresponding instruction to the user prompt. For example, if the instruction says “Think for up to 2048 tokens”, then the reasoning trace has anywhere between 1024 and 2048 tokens. In Table 11, we show that after training the model hardly follows the token instruction. It does sometimes generate more tokens when given a higher limit but often overshoots the limit. This may not be unique to our model as prior work suggests that OpenAI o1-mini can also not follow token instructions (Zhang & Chen, 2024). To prevent exceeding the limit, we test budget forcing the thinking to end once the limit is reached. This leads to perfect control (Table 11 (lower)). With budget forcing, the scaling trend is also clearer as the model can no longer overshoot the limit when given a small thinking budget. This leads to better test-time scaling values for *Token Prompting + budget forcing* in Table 3. To compute Control reported in Table 3 for token-conditional control variants we divide the number of times the thinking tokens in

**类别条件控制** OpenAI 通过一个具有三种可能设置的“reasoning\_effort” API参数（低、中、高）向用户开放测试时计算控制。<sup>3</sup> OpenAI 的文档还指出，“减少推理努力可以导致更快的响应和更少的推理令牌。”这表明他们无法保证控制测试时的计算。因此，OpenAI 可能只是根据所需的推理努力调整提示或系统指令。在Table 13中，我们展示了为短思考和长思考分别设置提示可以让我们在一定程度上控制思考时间：提示模型思考更长时间会导致更长的思考。然而，它并不能可靠地提高性能，控制也不精确。当前的控制遵守程度可能在我们只有三个类别时足够，但可能无法扩展到更细粒度的类别。为了计算Table 3中报告的这种方法的控制，我们假设在Table 13中提示模型短时间思考应该比AIME24的默认值生成更少的令牌，而长提示应该生成更多。由于 $8033 > 6109$  和  $9651 > 6109$ ，两个中有一个是符合我们预期的控制，因此控制率为50%。

### D.2. Examples for rejection sampling ablation

<sup>3</sup>[https://github.com/openai/openai-python/blob/44d6210f101abedeb2dd68507fcffcb329df70ea/src/openai/types/chat/completion\\_create\\_params.py#L172](https://github.com/openai/openai-python/blob/44d6210f101abedeb2dd68507fcffcb329df70ea/src/openai/types/chat/completion_create_params.py#L172)

Table 12. Scaling thinking time via step-conditional control. All metrics are averaged over the 30 samples in AIME24. Token counts ignore the thinking and step delimiters.

Steps instructed ( $\rightarrow$ )	16	32	64	128	256
<i>No intervention at test-time</i>					
Steps used	123	90	80	82	136
Tokens per step	60	70	69	66	56
Thinking tokens	7252	6277	5396	5552	7551
Answer tokens	665	653	735	777	754
AIME24	33.3	23.3	33.3	36.7	33.3
<i>Forcing end of thinking when 0 steps are reached</i>					
Steps used	16	32	59	78	136
Tokens per step	96	94	80	70	56
Thinking tokens	1517	2963	4636	5409	7551
Answer tokens	1111	788	799	794	754
AIME24	23.3	23.3	33.3	36.7	33.3

Table 13. Scaling thinking time via class-conditional control. We report “accuracy / average thinking tokens per sample”; the higher the accuracy and the fewer the thinking tokens (inference cost) the better.

Prompt appended to the question after two newlines	AIME24	MATH500	GPQA
Answer after a short amount of thinking. Do not spend excessive time double-checking your work.	30.0% / 8033	90.4% / 2537	56.6% / 4177
Answer after a long amount of thinking. If you feel like you are finished early, spend the extra time trying to double-check your work until you are absolutely sure that you have the correct answer.	36.7% / 9651	91.4% / 3875	51.0% / 4827
Without generic prompt appending	50.0% / 6109	93.0% / 3298	57.6% / 3510

Table 14. 由 s1-32B 的早期版本在 AIME24 上进行拒绝采样时生成的答案。黑色文本是问题，黄色文本是解答。仅基于问题，我们使用温度为 1 的 s1-32B 生成 浅蓝色 推理轨迹和 蓝色 回应。我们生成的推理轨迹在第一个例子中少于 4000 个标记，在第二个例子中少于 8000 个标记。平均每个例子需要 199 (1) 次尝试，直到思考的标记数  $\leq 4000$  ( $\leq 8000$  标记)。

设  $O(0, 0)$ ,  $A(\frac{1}{2}, 0)$ , 和  $B(0, \frac{\sqrt{3}}{2})$  是坐标平面上的点。设  $\mathcal{F}$  是第一象限中长度为单位的线段  $\overline{PQ}$  的集合，其中  $P$  在  $x$  轴上， $Q$  在  $y$  轴上。存在一个唯一的点  $C$  在  $\overline{AB}$  上，不同于  $A$  和  $B$ ，并且不属于  $\mathcal{F}$  中的任何其他线段，除了  $\overline{AB}$ 。那么  $OC^2 = \frac{p}{q}$ ，其中  $p$  和  $q$  是互质的正整数。求  $p + q$ 。

23

[使用 3981 个思考令牌和 1055 个答案令牌正确生成]

问题要求在  $\overline{AB}$  上找到一个点  $C$ ，该点不包含在  $\mathcal{F}$  中的任何其他线段中，除了  $\overline{AB}$  本身。

点  $O, A, B$  在第一象限形成一个三角形。

$$O = (0, 0)$$

$$A = (\frac{1}{2}, 0)$$

$$B = (0, \frac{\sqrt{3}}{2})$$

线段  $\overline{AB}$  连接  $A(\frac{1}{2}, 0)$  和  $B(0, \frac{\sqrt{3}}{2})$ 。

包含  $\overline{AB}$  的直线的方程可以通过两个点来确定：

$$\text{斜率 } m = \frac{\frac{\sqrt{3}}{2} - 0}{0 - \frac{1}{2}} = -\sqrt{3}$$

使用点斜式  $y - y_1 = m(x - x_1)$  与点  $A(\frac{1}{2}, 0)$ :

$$y - 0 = -\sqrt{3}(x - \frac{1}{2})$$

$$y = -\sqrt{3}x + \frac{\sqrt{3}}{2}$$

线段  $\overline{AB}$  位于第一象限，连接  $(\frac{1}{2}, 0)$  到  $(0, \frac{\sqrt{3}}{2})$ 。

集合  $\mathcal{F}$  包含长度为单位的线段  $\overline{PQ}$ ，其中  $P$  在  $x$  轴上， $Q$  在  $y$  轴上。设  $P = (a, 0)$  和  $Q = (0, b)$ 。

线段  $\overline{PQ}$  的长度为  $\sqrt{(a - 0)^2 + (0 - b)^2} = \sqrt{a^2 + b^2} = 1$ 。

因此  $a^2 + b^2 = 1$ 。

由于  $P$  在正  $x$  轴上， $Q$  在正  $y$  轴上，我们有  $a > 0$  和  $b > 0$ 。

线段  $\overline{PQ}$  连接  $(a, 0)$  和  $(0, b)$ 。包含  $\overline{PQ}$  的直线的方程为  $\frac{x}{a} + \frac{y}{b} = 1$ 。

集合  $\mathcal{F}$  覆盖的区域是由这些线段包含的线所界定的区域。端点  $(a, 0)$  和  $(0, b)$  变化，使得  $a^2 + b^2 = 1$ 。

考虑参数化  $a = \cos \theta$ ,  $b = \sin \theta$ , 其中  $0 < \theta < \frac{\pi}{2}$ 。

线段  $\overline{PQ}$  连接  $(\cos \theta, 0)$  和  $(0, \sin \theta)$ 。直线的方程为  $\frac{x}{\cos \theta} + \frac{y}{\sin \theta} = 1$ 。

我们正在寻找一个唯一的点  $C$  在  $\overline{AB}$  上，不同于  $A$  和  $B$ ，并且不属于  $\mathcal{F}$  中的任何其他线段，除了  $\overline{AB}$ 。

设  $C$  是  $\overline{AB}$  上的一个点。我们可以参数化  $\overline{AB}$  上的点为  $(1-t)A + tB$ ，其中  $0 \leq t \leq 1$ 。 $C = (1-t)(\frac{1}{2}, 0) + t(0, \frac{\sqrt{3}}{2}) = (\frac{1}{2} - \frac{1}{2}t, \frac{\sqrt{3}}{2}t)$ 。

由于  $C$  与  $A$  和  $B$  不同，我们有  $0 < t < 1$ 。

假设点  $C$  位于某线段  $\overline{PQ} \in \mathcal{F}$  上，其中  $P = (\cos \theta, 0)$  和  $Q = (0, \sin \theta)$ 。

点  $C$  满足方程  $\frac{x}{\cos \theta} + \frac{y}{\sin \theta} = 1$ 。

Table 11 are less than the upper limit by the total evaluations (2/5 for without intervention; 5/5 for with intervention).

**Step-conditional control** Token instructions fail as current models cannot count tokens. To accommodate this lack of capability, we experiment with making the counting more coarse-grained. We partition the reasoning traces into steps and ask the model to think for a specific number of steps rather than tokens. We split our reasoning traces on double newlines into steps, which we find act as intuitive separators based on manual inspection of samples. We bucket our training samples into powers of 2 depending on their number of steps and add a corresponding step instruction following the format in Figure 10 (right). This format is based on early experiments, where we found the model to be more likely to adhere to the step limit when counting down (“3 steps left...2 steps left”) rather than counting up (“Step2...Step3...”). This is likely because if counting down, the final step is always 1, which will act as a strong prior to the model to finish its generation. If counting up, the final step before the answer varies, thus if the model does not remember the original step instruction, it may fail to stop. We conclude the following from our results in Table 12: (1) The model still struggles to adhere to the step limit. The model sometimes simply continues counting into negative steps, e.g. “-1 steps left”. To solve this issue, we automatically stop the thinking process once 0 steps are reached and then force the model to transition to answering mode by appending the answer token delimiter (§3). This leads to perfect step adherence (lower half of Table 12), yet problems remain. (2) The model compensates for fewer steps by making each step longer. For example, when forced to use up to 16 steps vs 256 steps, the model generates an average of 96 tokens per step vs 56. Despite this issue, more steps still clearly correlate with more total thinking tokens in Table 12 and better performance leading to a positive slope (3) Step instructions are more costly than other methods. The step delimiters require around 6 tokens each which for e.g. 64 steps adds up to a total of around 380 tokens. When ignoring the step delimiters in token counts as in Table 12, the model still requires 7551 thinking tokens on average to achieve only 33.3% on AIME24. To compute Control reported in Table 3 for step-conditional control variants, we first decide that 100 tokens are an upper limit per step and then multiply this number by the steps instructed to arrive at a proxy total token limit, e.g. 1600 for 16 steps instructed. We then check whether the thinking tokens in Table 12 fit within the respective limit for each evaluation run (3/5 for without intervention; 5/5 for with intervention). For the model in Figure 7, we use a model with step-conditional control trained on an earlier version of our data and using an earlier version of our evaluation codebase.

**Class-conditional control** OpenAI exposes test-time compute control to users via a “reasoning\_effort” API parameter with three possible settings: low, medium, and high.<sup>3</sup> The OpenAI documentation also states that “Reducing reasoning effort can result in faster responses and fewer tokens used on reasoning in a response.” suggesting that they are unable to control test-time compute with guarantees. Thus, maybe OpenAI simply adjusts the prompt or system instruction depending on the reasoning effort desired. In Table 13, we show that separate prompts for short and long thinking allow us to control thinking time to some extent: Prompting the model to think for longer leads to longer thinking. However, it does not reliably improve performance and control is not precise. The current adherence to control may suffice when we only have three classes, but it might not scale to finer-grained classes. To compute Control reported in Table 3 for this method, we assume that prompting the model to think for a short time in Table 13 should produce fewer tokens than the default for AIME24, while the long prompt should produce more. As 8033 > 6109 and 9651 > 6109,

<sup>3</sup>[https://github.com/openai/openai-python/blob/44d6210f101abedeb2dd68507fcffcb329df70ea/src/openai/types/chat/completion\\_create\\_params.py#L172](https://github.com/openai/openai-python/blob/44d6210f101abedeb2dd68507fcffcb329df70ea/src/openai/types/chat/completion_create_params.py#L172)

代入  $C$  的坐标:

$$\frac{\frac{1-t}{2}}{\cos \theta} + \frac{\frac{\sqrt{3}t}{2}}{\sin \theta} = 1$$

$$\frac{1-t}{2\cos \theta} + \frac{\sqrt{3}t}{2\sin \theta} = 1$$

乘以 2:

$$\frac{1-t}{\cos \theta} + \frac{\sqrt{3}t}{\sin \theta} = 2$$

$$(1-t)\sin \theta + \sqrt{3}t\cos \theta = 2\sin \theta \cos \theta$$

$$\sin \theta - t\sin \theta + \sqrt{3}t\cos \theta = \sin(2\theta)/2$$

我们正在寻找一个点  $C$  在  $\overline{AB}$  上, 使得包含  $C$  的唯一线段是  $\overline{AB}$  本身。

线段  $\overline{AB}$  是族  $F$  的一部分。

对于  $\overline{AB}$ ,  $P = A = (\frac{1}{2}, 0)$  和  $Q = B = (0, \frac{\sqrt{3}}{2})$ 。

这里  $a = \frac{1}{2}$ ,  $b = \frac{\sqrt{3}}{2}$ 。

$$a^2 + b^2 = (\frac{1}{2})^2 + (\frac{\sqrt{3}}{2})^2 = \frac{1}{4} + \frac{3}{4} = 1.$$

线段  $\overline{AB}$  连接  $(\frac{1}{2}, 0)$  和  $(0, \frac{\sqrt{3}}{2})$ 。

线的方程是  $\frac{x}{1/2} + \frac{y}{\sqrt{3}/2} = 1$ , 简化为  $2x + \frac{2}{\sqrt{3}}y = 1$ , 或  $\sqrt{3}x + y = \frac{\sqrt{3}}{2}$ 。

这与方程  $y = -\sqrt{3}x + \frac{\sqrt{3}}{2}$  一致。

考虑线族  $\frac{x}{\cos \theta} + \frac{y}{\sin \theta} = 1$  的包络线。

对  $\theta$  求导:

$$-\frac{x \sin \theta}{\cos^2 \theta} + \frac{y \cos \theta}{\sin^2 \theta} = 0$$

$$\frac{y \cos \theta}{\sin^2 \theta} = \frac{x \sin \theta}{\cos^2 \theta}$$

$$y \cos^3 \theta = x \sin^3 \theta$$

$$\frac{y}{x} = \tan^3 \theta$$

$$\tan \theta = (\frac{y}{x})^{1/3}$$

从线方程:  $y \sin \theta = \sin \theta \cos \theta - x \sin \theta$

$$y = \cos \theta - x$$

$$\text{代入 } \cos \theta = \frac{1}{\sqrt{1+\tan^2 \theta}} = \frac{1}{\sqrt{1+(y/x)^2/3}}$$

$$\cos \theta = \frac{x^{1/3}}{(x^{2/3}+y^{2/3})^{1/2}} \sin \theta = \tan \theta \cos \theta = (\frac{y}{x})^{1/3} \frac{x^{1/3}}{(x^{2/3}+y^{2/3})^{1/2}} = \frac{y^{1/3}}{(x^{2/3}+y^{2/3})^{1/2}}$$

将上述结果代入直线方程:

$$\frac{x}{\cos \theta} + \frac{y}{\sin \theta} = 1$$

$$x(x^{2/3} + y^{2/3})^{1/2}/x^{1/3} + y(x^{2/3} + y^{2/3})^{1/2}/y^{1/3} = 1$$

$$x^{2/3}(x^{2/3} + y^{2/3})^{1/2} + y^{2/3}(x^{2/3} + y^{2/3})^{1/2} = 1$$

$$(x^{2/3} + y^{2/3})^{3/2} = 1$$

$$x^{2/3} + y^{2/3} = 1$$

族  $F$  的包络线是星形线  $x^{2/3} + y^{2/3} = 1$ 。

族  $F$  的线段与该星形线相切。

我们寻找的点  $C$  在线段  $\overline{AB}$  上, 且  $\overline{AB}$  与星形线相切。

线段  $\overline{AB}$  的方程为  $y = -\sqrt{3}x + \frac{\sqrt{3}}{2}$ 。

代入星形线方程:

$$x^{2/3} + (-\sqrt{3}x + \frac{\sqrt{3}}{2})^{2/3} = 1$$

设  $u = x^{1/3}$  和  $v = y^{1/3}$ . 星形线方程为  $u^2 + v^2 = 1$ 。

点  $C$  既在直线  $\overline{AB}$  上, 也在星形线上。

46

对星形线方程隐式求导:

$$\frac{2}{3}x^{-1/3} + \frac{2}{3}y^{-1/3}y' = 0$$

$$y' = -\frac{y^{1/3}}{x^{1/3}} = -(\frac{y}{x})^{1/3}$$

one out of two follows our expected control thus Control is 50%.

## D.2. Examples for rejection sampling ablation

点  $(x, y)$  处星形线的切线斜率为  $-\sqrt[3]{\frac{y}{x}}$ 。

直线  $\overline{AB}$  的斜率为  $-\sqrt{3}$ 。

因此，在切点处， $-\sqrt[3]{\frac{y}{x}} = -\sqrt{3}$ 。

$$\sqrt[3]{\frac{y}{x}} = \sqrt{3}$$

$$\frac{y}{x} = (\sqrt{3})^3 = 3\sqrt{3}$$

点  $C$  位于直线  $y = -\sqrt{3}x + \frac{\sqrt{3}}{2}$  上，并满足  $\frac{y}{x} = 3\sqrt{3}$ 。

因此  $y = 3\sqrt{3}x$ 。

代入直线方程：

$$3\sqrt{3}x = -\sqrt{3}x + \frac{\sqrt{3}}{2}$$

$$3x = -x + \frac{1}{2}$$

$$4x = \frac{1}{2}$$

$$x = \frac{1}{8}$$

因此  $y = 3\sqrt{3}(\frac{1}{8}) = \frac{3\sqrt{3}}{8}$ 。点  $C$  是  $(\frac{1}{8}, \frac{3\sqrt{3}}{8})$ 。

我们需要检查这个点是否在段  $\overline{AB}$  上。

点  $C$  的形式为  $(\frac{1}{2} - \frac{1}{2}t, \frac{\sqrt{3}}{2}t)$ 。

$$\frac{1}{2} - \frac{1}{2}t = \frac{1}{8} \implies 4 - 4t = 1 \implies 3 = 4t \implies t = \frac{3}{4}$$

$$\frac{\sqrt{3}}{2}t = \frac{\sqrt{3}}{2} \cdot \frac{3}{4} = \frac{3\sqrt{3}}{8}$$

所以  $C$  在  $\overline{AB}$  上。

我们需要确保  $0 < t < 1$ ，对于  $t = \frac{3}{4}$  这是成立的。

现在，我们需要检查这个点  $C$  是否不属于  $\mathcal{F}$  中的任何其他段。

点  $C$  位于星形线上。段  $\overline{AB}$  在  $C$  处与星形线相切。

$\mathcal{F}$  中的任何其他包含  $C$  的段都会将  $C$  作为内部点，而不是切点。

考虑  $\mathcal{F}$  的参数化。

点  $(x, y)$  属于  $\mathcal{F}$  中的某个段当且仅当存在  $\theta$  使得  $\frac{x}{\cos \theta} + \frac{y}{\sin \theta} = 1$ 。

对于点  $C = (\frac{1}{8}, \frac{3\sqrt{3}}{8})$ ，我们有：

$$\frac{1/8}{\cos \theta} + \frac{3\sqrt{3}/8}{\sin \theta} = 1$$

$$\frac{1}{\cos \theta} + \frac{3\sqrt{3}}{\sin \theta} = 8$$

$$\sin \theta + 3\sqrt{3} \cos \theta = 8 \sin \theta \cos \theta$$

$$\sin \theta + 3\sqrt{3} \cos \theta = 4 \sin(2\theta)$$

我们知道  $C$  位于  $\overline{AB}$  上，这对应于  $\mathcal{F}$  中的某个特定段。

问题陈述假设存在且唯一这样的点  $C$ 。

$\overline{AB}$  上的唯一点  $C$ ，不同于  $A$  和  $B$ ，且不属于  $\mathcal{F}$  中的任何其他段。

这意味着如果  $C$  位于另一个段  $\overline{P'Q'} \in \mathcal{F}$  上，那么  $\overline{P'Q'} = \overline{AB}$ 。

点  $C$  是  $\overline{AB}$  与星形线的切点。

现在我们需要找到  $OC^2$ 。

$$OC^2 = (\frac{1}{8})^2 + (\frac{3\sqrt{3}}{8})^2 = \frac{1}{64} + \frac{27}{64} = \frac{28}{64} = \frac{7}{16}$$

这里  $p = 7$  和  $q = 16$ 。

$p$  和  $q$  是互质的。

$p + q = 7 + 16 = 23$ 。最终检查推理：

家族  $\mathcal{F}$  由轴上的端点组成的单位长度线段组成。

这个家族的包络线是星形线  $x^{2/3} + y^{2/3} = 1$ 。

来自  $\mathcal{F}$  的线段与星形线相切。

线段  $\overline{AB}$  也是家族  $\mathcal{F}$  的成员。

Table 14. Answers generated by an earlier version of s1-32B when rejection sampling on AIME24. The black text is the question and the yellow text is the solution. Based solely on the question we generate the light blue reasoning trace and blue response with s1-32B using a temperature of 1. We generate until the reasoning trace has less than 4000 tokens for the first and less than 8000 tokens for the second example. It required on average 199 (1) tries per example until thinking was  $\leq 4000$  ( $\leq 8000$  tokens).

Let  $O(0, 0)$ ,  $A(\frac{1}{2}, 0)$ , and  $B(0, \frac{\sqrt{3}}{2})$  be points in the coordinate plane. Let  $\mathcal{F}$  be the family of segments  $\overline{PQ}$  of unit length lying in the first quadrant with  $P$  on the  $x$ -axis and  $Q$  on the  $y$ -axis. There is a unique point  $C$  on  $\overline{AB}$ , distinct from  $A$  and  $B$ , that does not belong to any segment from  $\mathcal{F}$  other than  $\overline{AB}$ . Then  $OC^2 = \frac{p}{q}$ , where  $p$  and  $q$  are relatively prime positive integers. Find  $p + q$ .

23

如果  $\overline{AB}$  上的点位于星形线上，它属于该点的切线段。由于  $\overline{AB}$  是切线段，这个点不属于来自  $\mathcal{F}$  的任何其他线段。

考虑线段  $\overline{AB}$  与星形线相交于两个点的情况。

如果线段  $\overline{AB}$  不与星形线相切，则有两个交点。

设交点为  $C_1$  和  $C_2$ 。

如果  $C_1$  位于  $\overline{AB}$  上，则星形线在  $C_1$  处的切线是线段  $\overline{AB}$ 。

如果  $C_2$  位于  $\overline{AB}$  上，则星形线在  $C_2$  处的切线是线段  $\overline{AB}$ 。

线段  $\overline{AB}$  的斜率是  $-\sqrt{3}$ 。

我们找到了星形线上切线斜率为  $-\sqrt{3}$  的点。

问题陈述保证了这样的点  $C$  的唯一性。

点  $C$  必须位于  $\overline{AB}$  的内部，因此  $0 < t < 1$ 。

参数  $t = \frac{3}{4}$  在这个范围内。

点  $C$  与  $A$  和  $B$  不同。

如果  $C = A$ ，则  $t = 0$ ， $C = (\frac{1}{2}, 0)$ 。

如果  $C = B$ ，则  $t = 1$ ， $C = (0, \frac{\sqrt{3}}{2})$ 。

切点  $C$  是  $\overline{AB}$  上（不包括端点）唯一位于星形线上的点。

[Correct generation with 3981 thinking tokens and 1055 answer tokens]

The problem asks for a point  $C$  on the segment  $\overline{AB}$  that is not contained in any other segment from the family  $\mathcal{F}$ , except for  $\overline{AB}$  itself.

The points  $O, A, B$  form a triangle in the first quadrant.

$$O = (0, 0)$$

$$A = \left(\frac{1}{2}, 0\right)$$

$$B = \left(0, \frac{\sqrt{3}}{2}\right)$$

The segment  $\overline{AB}$  connects  $A\left(\frac{1}{2}, 0\right)$  and  $B\left(0, \frac{\sqrt{3}}{2}\right)$ .

The equation of the line containing  $\overline{AB}$  can be found using the two points:

$$\text{Slope } m = \frac{\frac{\sqrt{3}}{2} - 0}{0 - \frac{1}{2}} = -\sqrt{3}$$

Using the point-slope form  $y - y_1 = m(x - x_1)$  with point  $A\left(\frac{1}{2}, 0\right)$ :

$$y - 0 = -\sqrt{3}(x - \frac{1}{2})$$

$$y = -\sqrt{3}x + \frac{\sqrt{3}}{2}$$

The segment  $\overline{AB}$  lies in the first quadrant and connects  $\left(\frac{1}{2}, 0\right)$  to  $\left(0, \frac{\sqrt{3}}{2}\right)$ .

The family  $\mathcal{F}$  consists of segments  $\overline{PQ}$  of unit length, where  $P$  is on the  $x$ -axis and  $Q$  is on the  $y$ -axis. Let  $P = (a, 0)$  and  $Q = (0, b)$ .

The length of  $\overline{PQ}$  is  $\sqrt{(a-0)^2 + (0-b)^2} = \sqrt{a^2 + b^2} = 1$ .

$$\text{So } a^2 + b^2 = 1.$$

Since  $P$  is on the positive  $x$ -axis and  $Q$  is on the positive  $y$ -axis, we have  $a > 0$  and  $b > 0$ .

The segment  $\overline{PQ}$  connects  $(a, 0)$  and  $(0, b)$ . The equation of the line containing  $\overline{PQ}$  is  $\frac{x}{a} + \frac{y}{b} = 1$ .

The region covered by the family  $\mathcal{F}$  is the area bounded by the lines containing these segments. The endpoints  $(a, 0)$  and  $(0, b)$  vary such that  $a^2 + b^2 = 1$ .

Consider the parametrization  $a = \cos \theta$ ,  $b = \sin \theta$ , where  $0 < \theta < \frac{\pi}{2}$ .

The segment  $\overline{PQ}$  connects  $(\cos \theta, 0)$  and  $(0, \sin \theta)$ . The equation of the line is  $\frac{x}{\cos \theta} + \frac{y}{\sin \theta} = 1$ .

We are looking for a unique point  $C$  on  $\overline{AB}$ , distinct from  $A$  and  $B$ , that does not belong to any segment from  $\mathcal{F}$  other than  $\overline{AB}$ .

Let  $C$  be a point on  $\overline{AB}$ . We can parametrize the points on  $\overline{AB}$  as  $(1-t)A + tB$  for  $0 \leq t \leq 1$ .

$$C = (1-t)\left(\frac{1}{2}, 0\right) + t\left(0, \frac{\sqrt{3}}{2}\right) = \left(\frac{1}{2} - \frac{1}{2}t, \frac{\sqrt{3}}{2}t\right).$$

Since  $C$  is distinct from  $A$  and  $B$ , we have  $0 < t < 1$ .

Suppose the point  $C$  lies on some segment  $\overline{PQ} \in \mathcal{F}$ , where  $P = (\cos \theta, 0)$  and  $Q = (0, \sin \theta)$ .

The point  $C$  satisfies the equation  $\frac{x}{\cos \theta} + \frac{y}{\sin \theta} = 1$ .

解决方案:

设线段家族  $\mathcal{F}$  由单位长度的线段  $\overline{PQ}$  组成, 其中  $P$  在正  $x$  轴上,  $Q$  在正  $y$  轴上。设  $P = (a, 0)$  和  $Q = (0, b)$ 。长度条件给出  $a^2 + b^2 = 1$ 。线段  $\overline{PQ}$  位于直线  $\frac{x}{a} + \frac{y}{b} = 1$  上。使用参数化  $a = \cos \theta$ ,  $b = \sin \theta$ , 其中  $0 < \theta < \frac{\pi}{2}$ , 直线的方程为  $\frac{x}{\cos \theta} + \frac{y}{\sin \theta} = 1$ 。

这个家族的包络线是星形线  $x^{2/3} + y^{2/3} = 1$ 。

来自  $\mathcal{F}$  的线段与这条星形线相切。

线段  $\overline{AB}$  连接  $A\left(\frac{1}{2}, 0\right)$  和  $B\left(0, \frac{\sqrt{3}}{2}\right)$ 。包含  $\overline{AB}$  的直线的方程是  $y = -\sqrt{3}x + \frac{\sqrt{3}}{2}$ 。这条线段是家族  $\mathcal{F}$  的一部分。

我们正在寻找  $\overline{AB}$  上的一个唯一点  $C$ , 该点不同于  $A$  和  $B$ , 并且不属于  $\mathcal{F}$  中的任何其他线段, 除了  $\overline{AB}$ 。这意味着  $C$  是  $\overline{AB}$  与星形线相切的点。

星形线上点  $(x, y)$  处的切线斜率是  $-\sqrt[3]{\frac{y}{x}}$ 。 $\overline{AB}$  的斜率是  $-\sqrt{3}$ 。

$$\text{将斜率设置为相等: } -\sqrt[3]{\frac{y}{x}} = -\sqrt{3} \Rightarrow \frac{y}{x} = 3\sqrt{3}.$$

点  $C$  位于直线  $y = -\sqrt{3}x + \frac{\sqrt{3}}{2}$  上, 并且满足  $y = 3\sqrt{3}x$ 。

$$\text{代入, } 3\sqrt{3}x = -\sqrt{3}x + \frac{\sqrt{3}}{2} \Rightarrow 4\sqrt{3}x = \frac{\sqrt{3}}{2} \Rightarrow x = \frac{1}{8}.$$

$$\text{然后 } y = 3\sqrt{3}\left(\frac{1}{8}\right) = \frac{3\sqrt{3}}{8}.$$

$$\text{点 } C \text{ 是 } \left(\frac{1}{8}, \frac{3\sqrt{3}}{8}\right).$$

我们需要检查  $C$  是否位于  $\overline{AB}$  上。 $\overline{AB}$  上的点是  $\left(\frac{1}{2} - \frac{1}{2}t, \frac{\sqrt{3}}{2}t\right)$ , 其中  $0 \leq t \leq 1$ 。

$$\text{对于 } x = \frac{1}{8}, \frac{1}{2} - \frac{1}{2}t = \frac{1}{8} \Rightarrow t = \frac{3}{4}.$$

$$\text{对于 } y = \frac{3\sqrt{3}}{8}, \frac{\sqrt{3}}{2}t = \frac{3\sqrt{3}}{8} \Rightarrow t = \frac{3}{4}.$$

由于  $0 < \frac{3}{4} < 1$ ,  $C$  是  $\overline{AB}$  的一个内点。

点  $C$  是  $\overline{AB}$  上的唯一点 (不包括端点), 位于星形线上。星形线在  $C$  处的切线是线段  $\overline{AB}$ 。因此,  $C$  不属于  $\mathcal{F}$  中的任何其他线段。

最后, 我们计算  $OC^2 = \left(\frac{1}{8}\right)^2 + \left(\frac{3\sqrt{3}}{8}\right)^2 = \frac{1}{64} + \frac{27}{64} = \frac{28}{64} = \frac{7}{16}$ 。

我们有  $p = 7$  和  $q = 16$ 。它们是互质的。

$$p + q = 7 + 16 = 23.$$

最终答案: 最终答案是 23

Plugging in the coordinates of  $C$ :

$$\frac{\frac{1-t}{2} + \frac{\sqrt{3}t}{2}}{\cos \theta} = 1$$

$$\frac{\frac{1-t}{2} + \frac{\sqrt{3}t}{2}}{2\cos \theta} = 1$$

Multiply by 2:

$$\frac{1-t}{\cos \theta} + \frac{\sqrt{3}t}{\sin \theta} = 2$$

$$(1-t)\sin \theta + \sqrt{3}t \cos \theta = 2 \sin \theta \cos \theta$$

$$\sin \theta - t \sin \theta + \sqrt{3}t \cos \theta = \sin(2\theta)/2$$

We are looking for a point  $C$  on  $\overline{AB}$  such that the only segment from  $\mathcal{F}$  containing  $C$  is  $\overline{AB}$  itself.

The segment  $\overline{AB}$  is part of the family  $\mathcal{F}$ .

For  $\overline{AB}$ ,  $P = A = (\frac{1}{2}, 0)$  and  $Q = B = (0, \frac{\sqrt{3}}{2})$ .

$$\text{Here } a = \frac{1}{2}, b = \frac{\sqrt{3}}{2}.$$

$$a^2 + b^2 = (\frac{1}{2})^2 + (\frac{\sqrt{3}}{2})^2 = \frac{1}{4} + \frac{3}{4} = 1.$$

The line segment  $\overline{AB}$  connects  $(\frac{1}{2}, 0)$  and  $(0, \frac{\sqrt{3}}{2})$ .

The equation of the line is  $\frac{x}{1/2} + \frac{y}{\sqrt{3}/2} = 1$ , which simplifies to  $2x + \frac{2}{\sqrt{3}}y = 1$ , or  $\sqrt{3}x + y = \frac{\sqrt{3}}{2}$ .

This matches the equation  $y = -\sqrt{3}x + \frac{\sqrt{3}}{2}$ .

Consider the envelope of the family of lines  $\frac{x}{\cos \theta} + \frac{y}{\sin \theta} = 1$ .

Differentiating with respect to  $\theta$ :

$$-\frac{x \sin \theta}{\cos^2 \theta} + \frac{y \cos \theta}{\sin^2 \theta} = 0$$

$$\frac{y \cos \theta}{\sin^2 \theta} = \frac{x \sin \theta}{\cos^2 \theta}$$

$$y \cos^3 \theta = x \sin^3 \theta$$

$$\frac{y}{x} = \tan^3 \theta$$

$$\tan \theta = (\frac{y}{x})^{1/3}$$

From the line equation:  $y \sin \theta = \sin \theta \cos \theta - x \sin \theta$

$$y = \cos \theta - x$$

$$\text{Substitute } \cos \theta = \frac{1}{\sqrt{1+\tan^2 \theta}} = \frac{1}{\sqrt{1+(y/x)^2/3}}$$

$$\cos \theta = \frac{x^{1/3}}{(x^{2/3}+y^{2/3})^{1/2}}$$

$$\sin \theta = \tan \theta \cos \theta = (\frac{y}{x})^{1/3} \frac{x^{1/3}}{(x^{2/3}+y^{2/3})^{1/2}} = \frac{y^{1/3}}{(x^{2/3}+y^{2/3})^{1/2}}$$

Substitute back into the line equation:

$$\frac{x}{\cos \theta} + \frac{y}{\sin \theta} = 1$$

$$x(x^{2/3}+y^{2/3})^{1/2}/x^{1/3} + y(x^{2/3}+y^{2/3})^{1/2}/y^{1/3} = 1$$

$$x^{2/3}(x^{2/3}+y^{2/3})^{1/2} + y^{2/3}(x^{2/3}+y^{2/3})^{1/2} = 1$$

$$(x^{2/3}+y^{2/3})^{3/2} = 1$$

$$x^{2/3} + y^{2/3} = 1$$

The envelope of the family  $\mathcal{F}$  is the astroid  $x^{2/3} + y^{2/3} = 1$ .

The segments from  $\mathcal{F}$  are tangent to this astroid.

The point  $C$  on  $\overline{AB}$  that we are looking for is the point where  $\overline{AB}$  is tangent to the astroid.

The equation of  $\overline{AB}$  is  $y = -\sqrt{3}x + \frac{\sqrt{3}}{2}$ .

Substitute into the astroid equation:

$$x^{2/3} + (-\sqrt{3}x + \frac{\sqrt{3}}{2})^{2/3} = 1$$

Let  $u = x^{1/3}$  and  $v = y^{1/3}$ . The astroid is  $u^2 + v^2 = 1$ . 50

The point  $C$  lies on both the line  $\overline{AB}$  and the astroid.

Differentiating the astroid equation implicitly:

$$\frac{2}{3}x^{-1/3} + \frac{2}{3}y^{-1/3}v' = 0$$

[错误生成,包含5923个思考标记和754个答案标记]

问题要求在  $\overline{AB}$  上找到一个点  $C$ , 使得  $C$  不在  $\mathcal{F}$  中的任何其他线段上, 除了  $\overline{AB}$  本身。 $\mathcal{F}$  由单位长度的线段  $\overline{PQ}$  组成, 其中  $P$  在  $x$  轴上,  $Q$  在  $y$  轴上, 位于第一象限。

设  $P = (x, 0)$  和  $Q = (0, y)$ , 其中  $x > 0$  且  $y > 0$ 。线段  $\overline{PQ}$  的长度为 1, 因此  $P$  和  $Q$  之间的距离为 1。

$$\sqrt{(x-0)^2 + (0-y)^2} = 1$$

$$\sqrt{x^2 + y^2} = 1$$

$$x^2 + y^2 = 1$$

线段  $\overline{PQ}$  位于第一象限, 因此  $x > 0$  且  $y > 0$ 。方程  $x^2 + y^2 = 1$  表示第一象限中的四分之一圆, 端点分别为  $(1, 0)$  和  $(0, 1)$ 。

线段  $\overline{PQ}$  上的点可以参数化为  $(1-t)x + t(0), (1-t)0 + ty) = (x(1-t), yt)$ , 其中  $0 \leq t \leq 1$ 。

线段  $\overline{AB}$  连接  $A(\frac{1}{2}, 0)$  和  $B(0, \frac{\sqrt{3}}{2})$ 。

包含  $\overline{AB}$  的直线方程可以通过这两个点来确定:

$$\text{斜率 } m = \frac{\frac{\sqrt{3}}{2}-0}{0-\frac{1}{2}} = \frac{\frac{\sqrt{3}}{2}}{-\frac{1}{2}} = -\sqrt{3}.$$

使用点斜式方程和点  $A(\frac{1}{2}, 0)$ :

$$y - 0 = -\sqrt{3}(x - \frac{1}{2})$$

$$y = -\sqrt{3}x + \frac{\sqrt{3}}{2}$$

线段  $\overline{AB}$  位于第一象限, 介于  $x = 0$  和  $x = \frac{1}{2}$  之间。

或者, 我们可以参数化  $\overline{AB}$  上的点。

设  $C$  是  $\overline{AB}$  上的一点。则  $C = (1-\lambda)A + \lambda B$ , 其中  $0 \leq \lambda \leq 1$ 。

$$C = (1-\lambda)(\frac{1}{2}, 0) + \lambda(0, \frac{\sqrt{3}}{2}) = (\frac{1}{2} - \frac{1}{2}\lambda, \frac{\sqrt{3}}{2}\lambda).$$

我们正在寻找  $\overline{AB}$  上的一个唯一点  $C$ , 该点不同于  $A$  和  $B$ , 并且不属于  $\mathcal{F}$  中除  $\overline{AB}$  之外的任何线段。

考虑  $\mathcal{F}$  中的一条线段, 端点为  $P = (x, 0)$  和  $Q = (0, y)$ , 其中  $x^2 + y^2 = 1$ 。

该线段上的点形式为  $(x(1-t), yt)$ , 其中  $0 \leq t \leq 1$ 。

我们希望找到  $\overline{AB}$  上的一个点  $C$ ,  $C \neq A, B$ , 使得如果  $C$  位于  $\mathcal{F}$  中的线段  $\overline{PQ}$  上, 则  $\overline{PQ} = \overline{AB}$ 。

设  $C = (\frac{1}{2} - \frac{1}{2}\lambda, \frac{\sqrt{3}}{2}\lambda)$ , 其中  $0 < \lambda < 1$ 。

如果  $C$  位于  $\mathcal{F}$  中的线段  $\overline{PQ}$  上, 则存在  $x, y > 0$  且  $x^2 + y^2 = 1$ , 以及  $t \in [0, 1]$  使得:

$$\frac{1}{2} - \frac{1}{2}\lambda = x(1-t)$$

$$\frac{\sqrt{3}}{2}\lambda = yt$$

从第一个方程,  $1-t = \frac{\frac{1}{2}-\frac{1}{2}\lambda}{x} = \frac{1-\lambda}{2x}$ 。

$$t = \frac{\frac{\sqrt{3}}{2}\lambda}{y} = \frac{\sqrt{3}\lambda}{2y}$$

由于  $t \in [0, 1]$ , 我们有  $1-\lambda \in [0, 1]$ 。

另外,  $x > 0$ , 所以  $\frac{1-\lambda}{2x} > 0$ , 这意味着  $1-\lambda > 0$ , 或  $\lambda < 1$ 。这与  $\overline{AB}$  上的点的条件一致。

我们有  $1-\lambda = \frac{1-\lambda}{2x}$  和  $t = \frac{\sqrt{3}\lambda}{2y}$ 。

将这两个方程相加:  $1 = \frac{1-\lambda}{2x} + \frac{\sqrt{3}\lambda}{2y} = (1-\lambda)y + \sqrt{3}\lambda x$

The slope of the tangent to the astroid at a point  $(x, y)$  is  $-\sqrt[3]{\frac{y}{x}}$ .

The slope of the line  $\overline{AB}$  is  $-\sqrt{3}$ .

So, at the point of tangency,  $-\sqrt[3]{\frac{y}{x}} = -\sqrt{3}$ .

$$\sqrt[3]{\frac{y}{x}} = \sqrt{3}$$

$$\frac{y}{x} = (\sqrt{3})^3 = 3\sqrt{3}$$

The point  $C$  lies on the line  $y = -\sqrt{3}x + \frac{\sqrt{3}}{2}$  and satisfies  $\frac{y}{x} = 3\sqrt{3}$ .

So  $y = 3\sqrt{3}x$ .

Substitute into the line equation:

$$3\sqrt{3}x = -\sqrt{3}x + \frac{\sqrt{3}}{2}$$

$$3x = -x + \frac{1}{2}$$

$$4x = \frac{1}{2}$$

$$x = \frac{1}{8}$$

$$\text{Then } y = 3\sqrt{3}\left(\frac{1}{8}\right) = \frac{3\sqrt{3}}{8}.$$

The point  $C$  is  $(\frac{1}{8}, \frac{3\sqrt{3}}{8})$ .

We need to check if this point lies on the segment  $\overline{AB}$ .

The point  $C$  is of the form  $(\frac{1}{2} - \frac{1}{2}t, \frac{\sqrt{3}}{2}t)$ .

$$\frac{1}{2} - \frac{1}{2}t = \frac{1}{8} \implies 4 - 4t = 1 \implies 3 = 4t \implies t = \frac{3}{4}.$$

$$\frac{\sqrt{3}}{2}t = \frac{\sqrt{3}}{2} \cdot \frac{3}{4} = \frac{3\sqrt{3}}{8}.$$

So  $C$  lies on  $\overline{AB}$ .

We need to make sure  $0 < t < 1$ , which is true for  $t = \frac{3}{4}$ .

Now, we need to check if this point  $C$  does not belong to any other segment from  $\mathcal{F}$ .

The point  $C$  lies on the astroid. The segment  $\overline{AB}$  is tangent to the astroid at  $C$ .

Any other segment from  $\mathcal{F}$  that contains  $C$  would have  $C$  as an interior point, not a tangent point.

Consider the parametrization of the family  $\mathcal{F}$ .

A point  $(x, y)$  belongs to some segment in  $\mathcal{F}$  iff there exists  $\theta$  such that  $\frac{x}{\cos \theta} + \frac{y}{\sin \theta} = 1$ .

For the point  $C = (\frac{1}{8}, \frac{3\sqrt{3}}{8})$ , we have:

$$\frac{1/8}{\cos \theta} + \frac{3\sqrt{3}/8}{\sin \theta} = 1$$

$$\frac{1}{\cos \theta} + \frac{3\sqrt{3}}{\sin \theta} = 8$$

$$\sin \theta + 3\sqrt{3} \cos \theta = 8 \sin \theta \cos \theta$$

$$\sin \theta + 3\sqrt{3} \cos \theta = 4 \sin(2\theta)$$

We know that  $C$  lies on  $\overline{AB}$ , which corresponds to some specific segment in  $\mathcal{F}$ .

The problem statement assumes the existence and uniqueness of such a point  $C$ .

The unique point  $C$  on  $\overline{AB}$ , distinct from  $A$  and  $B$ , that does not belong to any segment from  $\mathcal{F}$  other than  $\overline{AB}$ .

This means that if  $C$  lies on another segment  $\overline{P'Q'} \in \mathcal{F}$ , then  $\overline{P'Q'} = \overline{AB}$ .

The point  $C$  is the point of tangency of  $\overline{AB}$  with the astroid.

Now we need to find  $OC^2$ .

$$OC^2 = \left(\frac{1}{8}\right)^2 + \left(\frac{3\sqrt{3}}{8}\right)^2 = \frac{1}{64} + \frac{27}{64} = \frac{28}{64} = \frac{7}{16}.$$

Here  $p = 7$  and  $q = 16$ .

$p$  and  $q$  are relatively prime.

$$p + q = 7 + 16 = 23.$$

Final check of the reasoning:

The family  $\mathcal{F}$  consists of unit-length segments with endpoints on the axes.

我们知道  $x^2 + y^2 = 1$ 。

如果点  $C$  属于线段  $\overline{AB}$  本身, 那么线段的端点是  $A$  和  $B$ 。

因此, 线段是从  $(\frac{1}{2}, 0)$  到  $(0, \frac{\sqrt{3}}{2})$ 。

这里,  $x = \frac{1}{2}$  和  $y = \frac{\sqrt{3}}{2}$ 。

检查  $x^2 + y^2 = 1$ :  $(\frac{1}{2})^2 + (\frac{\sqrt{3}}{2})^2 = \frac{1}{4} + \frac{3}{4} = 1$ 。所以  $\overline{AB} \in \mathcal{F}$ 。

如果  $C$  属于  $\mathcal{F}$  中的另一个线段  $\overline{PQ}$ , 端点为  $P = (x, 0)$  和  $Q = (0, y)$ , 其中  $(x, y) \neq (\frac{1}{2}, \frac{\sqrt{3}}{2})$ 。

考虑几何解释。集合  $\mathcal{F}$  由单位长度的线段组成, 这些线段的端点位于坐标轴上。这些线段的中点位于圆

$$(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 = \frac{1}{2}$$
 上。

设点  $C$  对应于某个  $\lambda$ 。

如果  $C$  位于  $\overline{PQ}$  上, 则  $C = (x(1-t), yt)$ 。

$$\text{因此, } \frac{1}{2} - \frac{1}{2}\lambda = x(1-t) \text{ 和 } \frac{\sqrt{3}}{2}\lambda = yt.$$

我们需要找到一个  $\lambda \in (0, 1)$ , 使得在  $x^2 + y^2 = 1$ ,  $x, y > 0$  的条件下, 这些方程的唯一解是  $x = \frac{1}{2}$  和  $y = \frac{\sqrt{3}}{2}$ 。

从方程中,  $1 - t = \frac{1-\lambda}{2x}$  和  $t = \frac{\sqrt{3}\lambda}{2y}$ 。

将这些代入  $x^2 + y^2 = 1$ 。

尝试不同的方法。考虑线段族的切线。

线段由点  $(x, 0)$  和  $(0, y)$  形成, 其中  $x^2 + y^2 = 1$ 。

线段  $\overline{PQ}$  所在直线的方程为  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ , 即  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ 。

直线的方程为  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ 。

考虑线段的参数方程:  $(x(1-t), yt)$ 。

如果点  $C$  位于该线段上, 则  $C = (x_c, y_c)$ , 其中  $x_c = x(1-t)$  和  $y_c = yt$ 。

$$\text{因此, } \frac{x_c}{x} = 1 - t \text{ 和 } \frac{y_c}{y} = t.$$

$$\frac{x_c}{x} + \frac{y_c}{y} = 1 - t + t = 1.$$

集合  $\mathcal{F}$  中的线段是直线  $\frac{x}{x_0} + \frac{y}{y_0} = 1$  的一部分, 其中  $x_0^2 + y_0^2 = 1$ ,  $x_0, y_0 > 0$ 。

考虑点  $C = (\frac{1}{2} - \frac{1}{2}\lambda, \frac{\sqrt{3}}{2}\lambda)$ 。

如果  $C$  位于  $\mathcal{F}$  中的线段上, 端点为  $(x, 0)$  和  $(0, y)$ , 则:

$$\frac{\frac{1}{2} - \frac{1}{2}\lambda}{x} + \frac{\frac{\sqrt{3}}{2}\lambda}{y} = 1$$

$\frac{1-\lambda}{2x} + \frac{\sqrt{3}\lambda}{2y} = 1$  我们需要找到一个唯一的  $\lambda \in (0, 1)$ , 使得这个方程的唯一解  $(x, y)$ , 满足  $x^2 + y^2 = 1$ ,  $x, y > 0$ ,

是  $(\frac{1}{2}, \frac{\sqrt{3}}{2})$ 。

将  $x = \frac{1}{2}$  和  $y = \frac{\sqrt{3}}{2}$  代入方程:

$$\frac{1-\lambda}{2(\frac{1}{2})} + \frac{\sqrt{3}\lambda}{2(\frac{\sqrt{3}}{2})} = 1$$

$$1 - \lambda + \lambda = 1$$

$$1 = 1$$

这表明  $\overline{AB}$  上的任何点都位于线段  $\overline{AB}$  本身上。

考虑包含  $\overline{AB}$  的直线:  $y = -\sqrt{3}x + \frac{\sqrt{3}}{2}$ 。

如果  $\overline{AB}$  上的点  $C$  也位于  $\mathcal{F}$  中的另一个线段上, 该线段的端点为  $(x, 0)$  和  $(0, y)$ , 则有:

$$\frac{x_c}{x} + \frac{y_c}{y} = 1.$$

设点  $C$  将  $\overline{AB}$  分成比例  $\mu : (1 - \mu)$ , 其中  $0 < \mu < 1$ 。

$$C = (1 - \mu)A + \mu B = (1 - \mu)(\frac{1}{2}, 0) + \mu(0, \frac{\sqrt{3}}{2}) = (\frac{1}{2} - \frac{1}{2}\mu, \frac{\sqrt{3}}{2}\mu).$$

如果  $C$  位于端点为  $(x, 0)$  和  $(0, y)$  的线段上, 则有:

$$\frac{\frac{1}{2} - \frac{1}{2}\mu}{x} + \frac{\frac{\sqrt{3}}{2}\mu}{y} = 1.$$

我们正在寻找一个点  $C$ , 使得通过  $C$  的直线形式为  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ , 其中  $x_0^2 + y_0^2 = 1$ , 是唯一的。

If a point on  $\overline{AB}$  lies on the astroid, it belongs to the tangent segment at that point. Since  $\overline{AB}$  is the tangent segment, this point does not belong to any other segment from  $\mathcal{F}$ .

Consider the case when the segment  $\overline{AB}$  intersects the astroid at two points.

If the segment  $\overline{AB}$  is not tangent to the astroid, then there are two points of intersection.

Let the points of intersection be  $C_1$  and  $C_2$ .

If  $C_1$  lies on  $\overline{AB}$ , then the tangent to the astroid at  $C_1$  is the segment  $\overline{AB}$ .

If  $C_2$  lies on  $\overline{AB}$ , then the tangent to the astroid at  $C_2$  is the segment  $\overline{AB}$ .

The slope of  $\overline{AB}$  is  $-\sqrt{3}$ .

We found the point on the astroid with tangent slope  $-\sqrt{3}$ .

The problem statement guarantees the uniqueness of such a point  $C$ .

The point  $C$  must be interior to  $\overline{AB}$ , so  $0 < t < 1$ .

The parameter  $t = \frac{3}{4}$  is in this range.

The point  $C$  is distinct from  $A$  and  $B$ .

If  $C = A$ , then  $t = 0$ ,  $C = (\frac{1}{2}, 0)$ .

If  $C = B$ , then  $t = 1$ ,  $C = (0, \frac{\sqrt{3}}{2})$ .

The point of tangency  $C$  is the unique point on  $\overline{AB}$  (excluding endpoints) that lies on the astroid.

考虑以原点为中心的单位圆的切线。形式为  $\frac{x}{x_0} + \frac{y}{y_0} = 1$  的直线是圆  $x^2 + y^2 = r^2$  的切线。

考虑对偶问题。线段族对应于第一象限中单位圆  $x^2 + y^2 = 1$  上的点。

包含线段的直线是  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ 。

这些直线的包络线是其切线为这些直线的曲线。

方程  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ , 约束条件为  $x_0^2 + y_0^2 = 1$ 。

设  $x_0 = \cos \theta$ ,  $y_0 = \sin \theta$ , 其中  $0 < \theta < \frac{\pi}{2}$ 。

直线的方程为  $x \cos \theta + y \sin \theta = 1$ 。

线族  $x \cos \theta + y \sin \theta = 1$  的包络线是圆  $x^2 + y^2 = 1$ 。

$\mathcal{F}$  中的线段是第一象限单位圆的切线的一部分。

$\overline{AB}$  上的点  $C$  不属于  $\mathcal{F}$  中的任何其他线段, 意味着通过  $C$  的单位圆的切线是唯一的。

考虑点  $C = (\frac{1}{2} - \frac{1}{2}\lambda, \frac{\sqrt{3}}{2}\lambda)$ 。

通过  $C$  的单位圆的切线形式为  $xx_C + yy_C = 1$ 。

如果切线是唯一的, 则  $C$  位于单位圆上。

$$(\frac{1}{2} - \frac{1}{2}\lambda)^2 + (\frac{\sqrt{3}}{2}\lambda)^2 = 1(\frac{1}{4} - \frac{1}{2}\lambda + \frac{1}{4}\lambda^2) + \frac{3}{4}\lambda^2 = 1$$

$$\frac{1}{4} - \frac{1}{2}\lambda + \lambda^2 = 1$$

$$4\lambda^2 - 2\lambda + 1 = 4$$

$$4\lambda^2 - 2\lambda - 3 = 0$$

解  $\lambda$ :

$$\lambda = \frac{-(-2) \pm \sqrt{(-2)^2 - 4(4)(-3)}}{2(4)} = \frac{2 \pm \sqrt{4+48}}{8} = \frac{2 \pm \sqrt{52}}{8} = \frac{2 \pm 2\sqrt{13}}{8} = \frac{1 \pm \sqrt{13}}{4}.$$

我们需要  $0 < \lambda < 1$ 。

$$\sqrt{13} \approx 3.6.$$

$$\frac{1+3.6}{4} = \frac{4.6}{4} > 1.$$

$$\frac{1-3.6}{4} = \frac{-2.6}{4} < 0.$$

有些地方出错了。点  $C$  在单位圆内。

让我们重新表述条件。存在唯一的点  $C$  在  $\overline{AB}$  上, 使得如果  $C$  位于  $\mathcal{F}$  中的线段  $\overline{PQ}$  上, 则  $\overline{PQ} = \overline{AB}$ 。

设点  $C$  在  $\overline{AB}$  上的坐标为  $(\frac{1}{2} - \frac{1}{2}\lambda, \frac{\sqrt{3}}{2}\lambda)$ 。

考虑  $\mathcal{F}$  中的一个线段, 其端点为  $(x, 0)$  和  $(0, y)$ ,  $x^2 + y^2 = 1$ 。

该线段的参数形式为  $(x(1-t), yt)$ 。

如果  $C$  在这条线段上, 那么对于某个  $t \in [0, 1]$ :

$$\frac{1}{2} - \frac{1}{2}\lambda = x(1-t)$$

$$\frac{\sqrt{3}}{2}\lambda = yt$$

如果线段是唯一的, 那么  $x$  和  $y$  的值也是唯一的。

唯一解为  $\frac{1-\lambda}{2x} + \frac{\sqrt{3}\lambda}{2y} = 1$ , 其中  $x^2 + y^2 = 1$ ,  $x, y > 0$ , 应该是  $x = \frac{1}{2}$ ,  $y = \frac{\sqrt{3}}{2}$ 。

考虑包含  $\overline{AB}$  的直线与由  $\mathcal{F}$  中线段形成的区域边界相交的情况。

边界是  $x$ -轴、 $y$ -轴和四分之一圆  $x^2 + y^2 = 1$ 。

再次考虑对偶问题。线族  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ , 其中  $x_0^2 + y_0^2 = 1$ ,  $x_0, y_0 > 0$ 。

这些线的包络线是圆  $x^2 + y^2 = 1$ 。

如果一个点位于圆  $x^2 + y^2 = 1$  上, 那么它属于  $\mathcal{F}$  中的唯一线段。

点  $C$  在  $\overline{AB}$  上满足  $C = tA + (1-t)B$ 。

让我们考虑线段为  $\overline{AB}$  的情况。这里  $x = \frac{1}{2}$ ,  $y = \frac{\sqrt{3}}{2}$ 。

点  $C$  可以表示为  $(\frac{1}{2}(1-s), \frac{\sqrt{3}}{2}s)$ , 其中  $0 \leq s \leq 1$ 。我们正在寻找一个点  $C$  在  $\overline{AB}$  上, 但不位于  $\mathcal{F}$  中的任何其他线段上。

Solution:

Let the family of segments  $\mathcal{F}$  be defined by segments  $\overline{PQ}$  of unit length with  $P$  on the positive  $x$ -axis and  $Q$  on the positive  $y$ -axis. Let  $P = (a, 0)$  and  $Q = (0, b)$ . The length condition gives  $a^2 + b^2 = 1$ . The segment  $\overline{PQ}$  lies on the line  $\frac{x}{a} + \frac{y}{b} = 1$ . Using the parametrization  $a = \cos \theta$ ,  $b = \sin \theta$  with  $0 < \theta < \frac{\pi}{2}$ , the equation of the line is  $\frac{x}{\cos \theta} + \frac{y}{\sin \theta} = 1$ .

The envelope of this family of lines is the astroid  $x^{2/3} + y^{2/3} = 1$ .

The segments from  $\mathcal{F}$  are tangent to this astroid.

The segment  $\overline{AB}$  connects  $A(\frac{1}{2}, 0)$  and  $B(0, \frac{\sqrt{3}}{2})$ . The equation of the line containing  $\overline{AB}$  is  $y = -\sqrt{3}x + \frac{\sqrt{3}}{2}$ . This segment is part of the family  $\mathcal{F}$ .

We are looking for a unique point  $C$  on  $\overline{AB}$ , distinct from  $A$  and  $B$ , that does not belong to any segment from  $\mathcal{F}$  other than  $\overline{AB}$ . This means  $C$  is the point where  $\overline{AB}$  is tangent to the astroid.

The slope of the tangent to the astroid at a point  $(x, y)$  is  $-\sqrt[3]{\frac{y}{x}}$ . The slope of  $\overline{AB}$  is  $-\sqrt{3}$ .

Setting the slopes equal:  $-\sqrt[3]{\frac{y}{x}} = -\sqrt{3} \Rightarrow \frac{y}{x} = 3\sqrt{3}$ .

The point  $C$  lies on the line  $y = -\sqrt{3}x + \frac{\sqrt{3}}{2}$  and satisfies  $y = 3\sqrt{3}x$ .

Substituting,  $3\sqrt{3}x = -\sqrt{3}x + \frac{\sqrt{3}}{2} \Rightarrow 4\sqrt{3}x = \frac{\sqrt{3}}{2} \Rightarrow x = \frac{1}{8}$ .

Then  $y = 3\sqrt{3}(\frac{1}{8}) = \frac{3\sqrt{3}}{8}$ .

The point  $C$  is  $(\frac{1}{8}, \frac{3\sqrt{3}}{8})$ .

We need to check if  $C$  lies on  $\overline{AB}$ . The points on  $\overline{AB}$  are  $(\frac{1}{2} - \frac{1}{2}t, \frac{\sqrt{3}}{2}t)$  for  $0 \leq t \leq 1$ .

For  $x = \frac{1}{8}, \frac{1}{2} - \frac{1}{2}t = \frac{1}{8} \Rightarrow t = \frac{3}{4}$ .

For  $y = \frac{3\sqrt{3}}{8}, \frac{\sqrt{3}}{2}t = \frac{3\sqrt{3}}{8} \Rightarrow t = \frac{3}{4}$ .

Since  $0 < \frac{3}{4} < 1$ ,  $C$  is an interior point of  $\overline{AB}$ .

The point  $C$  is the unique point on  $\overline{AB}$  (excluding endpoints) that lies on the astroid. The tangent to the astroid at  $C$  is the segment  $\overline{AB}$ . Thus,  $C$  does not belong to any other segment from  $\mathcal{F}$ .

Finally, we calculate  $OC^2 = (\frac{1}{8})^2 + (\frac{3\sqrt{3}}{8})^2 = \frac{1}{64} + \frac{27}{64} = \frac{28}{64} = \frac{7}{16}$ .

We have  $p = 7$  and  $q = 16$ . These are relatively prime.

$p + q = 7 + 16 = 23$ .

Final Answer: The final answer is 23

考虑几何。家族  $\mathcal{F}$  形成一个由坐标轴和四分之一圆边界限定的区域。

线段是圆的弦，这些圆的圆心位于直线  $y = x$  上，且与坐标轴相切。

设线段的端点为  $(\cos \theta, 0)$  和  $(0, \sin \theta)$ 。

线段上的一个点为  $(\cos \theta(1-t), \sin \theta t)$ 。

点  $C = (\frac{1}{2} - \frac{1}{2}\lambda, \frac{\sqrt{3}}{2}\lambda)$ 。

如果  $C$  位于这样的线段上，则有  $\frac{\frac{1}{2}-\frac{1}{2}\lambda}{\cos \theta} = 1-t$  和  $\frac{\frac{\sqrt{3}}{2}\lambda}{\sin \theta} = t$ 。

$$\frac{1-\lambda}{2\cos \theta} + \frac{\sqrt{3}\lambda}{2\sin \theta} = 1$$

我们需要一个唯一的  $\lambda \in (0, 1)$ ，使得唯一的解  $(\cos \theta, \sin \theta)$  是  $(\frac{1}{2}, \frac{\sqrt{3}}{2})$ ，这对应于  $\theta = \frac{\pi}{3}$ 。

考虑函数  $f(\theta) = \frac{1}{2\cos \theta} - \frac{\lambda}{2\cos \theta} + \frac{\sqrt{3}\lambda}{2\sin \theta}$ 。

我们需要  $f(\theta) = 1$ 。

如果  $\lambda = \frac{1}{2}$ ，则  $C = (\frac{1}{4}, \frac{\sqrt{3}}{4})$ 。

$$\frac{1-1/2}{2\cos \theta} + \frac{\sqrt{3}(1/2)}{2\sin \theta} = 1$$

$$\frac{1}{4\cos \theta} + \frac{\sqrt{3}}{4\sin \theta} = 1$$

$$\frac{1}{\cos \theta} + \frac{\sqrt{3}}{\sin \theta} = 4$$

$$\sin \theta + \sqrt{3} \cos \theta = 4 \sin \theta \cos \theta$$

$$\sin \theta + \sqrt{3} \cos \theta = 2 \sin(2\theta)$$

$$\text{除以2: } \frac{1}{2} \sin \theta + \frac{\sqrt{3}}{2} \cos \theta = \sin(2\theta)$$

$$\sin(\theta + \frac{\pi}{3}) = \sin(2\theta)$$

$$\theta + \frac{\pi}{3} = 2\theta + 2\pi k, \text{ 或 } \theta + \frac{\pi}{3} = \pi - 2\theta + 2\pi k。$$

$$\theta = \frac{\pi}{3} - 2\pi k, \text{ 所以 } \theta = \frac{\pi}{3}$$

$$3\theta = \pi - \frac{\pi}{3} + 2\pi k = \frac{2\pi}{3} + 2\pi k$$

$$\theta = \frac{2\pi}{9} + \frac{2\pi k}{3}$$

$$\text{如果 } k = 0, \theta = \frac{2\pi}{9}$$

$$\text{如果 } k = 1, \theta = \frac{2\pi}{9} + \frac{2\pi}{3} = \frac{2\pi+6\pi}{9} = \frac{8\pi}{9}$$

$$\text{如果 } \lambda = \frac{1}{2}, \text{ 该点为 } (\frac{1}{4}, \frac{\sqrt{3}}{4})$$

这个点位于对于  $\theta = \frac{\pi}{3}$  和  $\theta = \frac{8\pi}{9}$  的线段上。

唯一的点  $C$  意味着对于其他  $\lambda$  值，存在多个  $(x, y)$  解。

考虑通过  $C$  的单位圆的切线。

如果切线是唯一的， $C$  就在单位圆上。这导致  $\lambda = \frac{1 \pm \sqrt{13}}{4}$ ，这两个值都不在  $(0, 1)$  范围内。

考虑  $\mathcal{F}$  家族的几何解释。线段连接轴上的点，使得距离为 1。

这些线段中心的轨迹是第一象限中的圆  $(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 = \frac{1}{2}$ 。

如果点  $C$  位于  $\mathcal{F}$  中的某一线段上，那么  $C$  位于连接  $(x, 0)$  和  $(0, y)$  的线段上，且  $x^2 + y^2 = 1$ 。

包含该线段的直线为  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ 。

如果  $C$  属于唯一的一条线段，那么通过  $C$  的这种形式的直线是唯一的。

点  $C$  在  $\overline{AB}$  上的坐标为  $(\frac{1}{2} - \frac{1}{2}\lambda, \frac{\sqrt{3}}{2}\lambda)$ 。

我们需要一个唯一的  $(x_0, y_0)$ ，满足  $x_0^2 + y_0^2 = 1$ ,  $x_0, y_0 > 0$ ，使得  $\frac{\frac{1}{2}-\frac{1}{2}\lambda}{x_0} + \frac{\frac{\sqrt{3}}{2}\lambda}{y_0} = 1$ 。

设点  $C$  使得通过  $C$  的形式为  $\frac{x}{x_0} + \frac{y}{y_0} = 1$  的直线是唯一的，且  $x_0^2 + y_0^2 = 1$ 。

这条直线与单位圆相切。

点  $C$  位于包含  $\overline{AB}$  的直线与单位圆的交点上。

$$(\frac{1}{2} - \frac{1}{2}\lambda)^2 + (\frac{\sqrt{3}}{2}\lambda)^2 = 1$$

我们没有找到  $\lambda \in (0, 1)$  的解。

**[Incorrect generation with 5923 thinking tokens and 754 answer tokens]**

The problem asks for a point  $C$  on the segment  $\overline{AB}$  such that  $C$  is not on any other segment from the family  $\mathcal{F}$ , except  $\overline{AB}$  itself. The family  $\mathcal{F}$  consists of segments  $\overline{PQ}$  of unit length, with  $P$  on the  $x$ -axis and  $Q$  on the  $y$ -axis in the first quadrant.

Let  $P = (x, 0)$  and  $Q = (0, y)$ , where  $x > 0$  and  $y > 0$ . The length of the segment  $\overline{PQ}$  is 1, so the distance between  $P$  and  $Q$  is 1.

$$\sqrt{(x-0)^2 + (0-y)^2} = 1$$

$$\sqrt{x^2 + y^2} = 1$$

$$x^2 + y^2 = 1$$

The segment  $\overline{PQ}$  lies in the first quadrant, so  $x > 0$  and  $y > 0$ .

The equation  $x^2 + y^2 = 1$  represents the quarter circle in the first quadrant, with endpoints at  $(1, 0)$  and  $(0, 1)$ .

The points on the segment  $\overline{PQ}$  can be parameterized as  $(1-t)x + t(0), (1-t)0 + ty = (x(1-t), yt)$ , where  $0 \leq t \leq 1$ .

The segment  $\overline{AB}$  connects  $A(\frac{1}{2}, 0)$  and  $B(0, \frac{\sqrt{3}}{2})$ .

The equation of the line containing  $\overline{AB}$  can be found using the two points:

$$\text{Slope } m = \frac{\frac{\sqrt{3}}{2} - 0}{0 - \frac{1}{2}} = \frac{\frac{\sqrt{3}}{2}}{-\frac{1}{2}} = -\sqrt{3}.$$

Using point-slope form with point  $A(\frac{1}{2}, 0)$ :

$$y - 0 = -\sqrt{3}(x - \frac{1}{2})$$

$$y = -\sqrt{3}x + \frac{\sqrt{3}}{2}$$

The segment  $\overline{AB}$  lies in the first quadrant, between  $x = 0$  and  $x = \frac{1}{2}$ .

Alternatively, we can parameterize the points on  $\overline{AB}$ .

Let  $C$  be a point on  $\overline{AB}$ . Then  $C = (1-\lambda)A + \lambda B$  for some  $0 \leq \lambda \leq 1$ .

$$C = (1-\lambda)(\frac{1}{2}, 0) + \lambda(0, \frac{\sqrt{3}}{2}) = (\frac{1}{2} - \frac{1}{2}\lambda, \frac{\sqrt{3}}{2}\lambda).$$

We are looking for a unique point  $C$  on  $\overline{AB}$ , distinct from  $A$  and  $B$ , that does not belong to any segment from  $\mathcal{F}$  other than  $\overline{AB}$ .

Consider a segment from  $\mathcal{F}$ , with endpoints  $P = (x, 0)$  and  $Q = (0, y)$ , where  $x^2 + y^2 = 1$ .

A point on this segment is of the form  $(x(1-t), yt)$ , where  $0 \leq t \leq 1$ .

We want to find a point  $C$  on  $\overline{AB}$ ,  $C \neq A, B$ , such that if  $C$  lies on a segment  $\overline{PQ} \in \mathcal{F}$ , then  $\overline{PQ} = \overline{AB}$ .

$$\text{Let } C = (\frac{1}{2} - \frac{1}{2}\lambda, \frac{\sqrt{3}}{2}\lambda), \text{ where } 0 < \lambda < 1.$$

If  $C$  lies on a segment  $\overline{PQ}$  from  $\mathcal{F}$ , then there exist  $x, y > 0$  with  $x^2 + y^2 = 1$ , and  $t \in [0, 1]$  such that:

$$\frac{1}{2} - \frac{1}{2}\lambda = x(1-t)$$

$$\frac{\sqrt{3}}{2}\lambda = yt$$

$$\text{From the first equation, } 1-t = \frac{\frac{1}{2} - \frac{1}{2}\lambda}{x} = \frac{1-\lambda}{2x}.$$

$$\text{From the second equation, } t = \frac{\frac{\sqrt{3}}{2}\lambda}{y} = \frac{\sqrt{3}\lambda}{2y}.$$

Since  $t \in [0, 1]$ , we have  $1-t \in [0, 1]$ .

Also,  $x > 0$ , so  $\frac{1-\lambda}{2x} > 0$ , which implies  $1-\lambda > 0$ , or  $\lambda < 1$ . This is consistent with the condition for point on  $\overline{AB}$ .

$$\text{We have } 1-t = \frac{1-\lambda}{2x} \text{ and } t = \frac{\sqrt{3}\lambda}{2y}.$$

$$\text{Adding these equations: } 1 = \frac{1-\lambda}{2x} + \frac{\sqrt{3}\lambda}{2y}$$

$$2xy = (1-\lambda)y + \sqrt{3}\lambda x$$

唯一的点  $C$  意味着如果有任何其他线段通过  $C$ , 它必须是  $\overline{AB}$ 。

考虑连接  $(x, 0)$  和  $(0, y)$  的线段所在的直线的斜率。斜率为  $-\frac{y}{x}$ 。

$$\text{方程为 } y = -\frac{y}{x}x + y.$$

$$\text{如果 } C \text{ 位于线段上, 那么 } \frac{1-\lambda}{2x} + \frac{\sqrt{3}\lambda}{2y} = 1.$$

考虑几何平均。

设点  $C$  的坐标为  $(a, b)$ 。连接  $(x, 0)$  和  $(0, y)$  的线段通过  $(a, b)$ 。

$$\frac{a}{x} + \frac{b}{y} = 1.$$

$$\text{如果 } C \text{ 在 } \overline{AB} \text{ 上, 那么 } b = -\sqrt{3}a + \frac{\sqrt{3}}{2}.$$

$$a = \frac{1}{2} - \frac{1}{2}\lambda, \quad b = \frac{\sqrt{3}}{2}\lambda.$$

点  $C$  与线段家族的包络线相切。

We know  $x^2 + y^2 = 1$ .

If the point  $C$  belongs to the segment  $\overline{AB}$  itself, then the endpoints of the segment are  $A$  and  $B$ .

So, the segment is from  $(\frac{1}{2}, 0)$  to  $(0, \frac{\sqrt{3}}{2})$ .

Here,  $x = \frac{1}{2}$  and  $y = \frac{\sqrt{3}}{2}$ .

Check if  $x^2 + y^2 = 1$ :  $(\frac{1}{2})^2 + (\frac{\sqrt{3}}{2})^2 = \frac{1}{4} + \frac{3}{4} = 1$ . So  $\overline{AB} \in \mathcal{F}$ .

If  $C$  belongs to another segment  $\overline{PQ}$  from  $\mathcal{F}$ , with endpoints  $P = (x, 0)$  and  $Q = (0, y)$ , where  $(x, y) \neq (\frac{1}{2}, \frac{\sqrt{3}}{2})$ .

Consider the geometric interpretation. The family  $\mathcal{F}$  consists of segments of unit length with endpoints on the axes.

The midpoints of these segments lie on the circle  $(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 = \frac{1}{2}$ .

Let the point  $C$  correspond to some  $\lambda$ .

If  $C$  lies on  $\overline{PQ}$ , then  $C = (x(1-t), yt)$ .

So,  $\frac{1}{2} - \frac{1}{2}\lambda = x(1-t)$  and  $\frac{\sqrt{3}}{2}\lambda = yt$ .

We need to find a  $\lambda \in (0, 1)$  such that the only solution to these equations, with  $x^2 + y^2 = 1$ ,  $x, y > 0$ , is  $x = \frac{1}{2}$  and  $y = \frac{\sqrt{3}}{2}$ .

From the equations,  $1-t = \frac{1-\lambda}{2x}$  and  $t = \frac{\sqrt{3}\lambda}{2y}$ .

Substitute these into  $x^2 + y^2 = 1$ .

Try a different approach. Consider the tangents to the family of segments.

The segments are formed by points  $(x, 0)$  and  $(0, y)$  with  $x^2 + y^2 = 1$ .

The equation of the line containing the segment  $\overline{PQ}$  is  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ , or  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ .

The equation of the line is  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ .

Consider the equation of the segment in parametric form:  $(x(1-t), yt)$ .

If a point  $C$  lies on this segment, then  $C = (x_c, y_c)$ , where  $x_c = x(1-t)$  and  $y_c = yt$ .

So,  $\frac{x_c}{x_0} = 1-t$  and  $\frac{y_c}{y_0} = t$ .

$\frac{x_c}{x_0} + \frac{y_c}{y_0} = 1-t+t = 1$ .

The segments from  $\mathcal{F}$  are parts of the lines  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ , where  $x_0^2 + y_0^2 = 1$ ,  $x_0, y_0 > 0$ .

Consider the point  $C = (\frac{1}{2} - \frac{1}{2}\lambda, \frac{\sqrt{3}}{2}\lambda)$ .

If  $C$  lies on a segment from  $\mathcal{F}$  with endpoints  $(x, 0)$  and  $(0, y)$ , then:

$$\frac{\frac{1}{2}-\frac{1}{2}\lambda}{x} + \frac{\frac{\sqrt{3}}{2}\lambda}{y} = 1$$

$$\frac{1-\lambda}{2x} + \frac{\sqrt{3}\lambda}{2y} = 1$$

We need to find a unique  $\lambda \in (0, 1)$  such that the only solution  $(x, y)$  to this equation, with  $x^2 + y^2 = 1$ ,  $x, y > 0$ , is

$$(\frac{1}{2}, \frac{\sqrt{3}}{2}).$$

Substitute  $x = \frac{1}{2}$  and  $y = \frac{\sqrt{3}}{2}$  into the equation:

$$\frac{\frac{1-\lambda}{2}}{2(\frac{1}{2})} + \frac{\frac{\sqrt{3}\lambda}{2}}{2(\frac{\sqrt{3}}{2})} = 1$$

$$1 - \lambda + \lambda = 1$$

$$1 = 1$$

This shows that any point on  $\overline{AB}$  lies on the segment  $\overline{AB}$  itself.

Consider the line containing  $\overline{AB}$ :  $y = -\sqrt{3}x + \frac{\sqrt{3}}{2}$ .

If a point  $C$  on  $\overline{AB}$  also lies on another segment from  $\mathcal{F}$  with endpoints  $(x, 0)$  and  $(0, y)$ , then:

$$\frac{x_c}{x} + \frac{y_c}{y} = 1$$

Let the point  $C$  divide  $\overline{AB}$  in the ratio  $\mu : (1-\mu)$ , where  $0 < \mu < 1$ .

$$C = (1-\mu)A + \mu B = (1-\mu)(\frac{1}{2}, 0) + \mu(0, \frac{\sqrt{3}}{2}) = (\frac{1}{2} - \frac{1}{2}\mu, \frac{\sqrt{3}}{2}\mu).$$

If  $C$  lies on a segment with endpoints  $(x, 0)$  and  $(0, y)$ , then:

$$\frac{\frac{1}{2}-\frac{1}{2}\mu}{x} + \frac{\frac{\sqrt{3}}{2}\mu}{y} = 1$$

解:

设线段家族为  $\mathcal{F}$ 。线段的端点为  $P = (x, 0)$  和  $Q = (0, y)$ , 且  $x^2 + y^2 = 1$ ,  $x, y > 0$ 。线段  $\overline{PQ}$  上的点满足  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ 。线段位于直线  $\frac{x}{x_0} + \frac{y}{y_0} = 1$  上, 其中  $x_0^2 + y_0^2 = 1$ ,  $x_0, y_0 > 0$ 。线段  $\overline{AB}$  的端点为  $A(\frac{1}{2}, 0)$  和  $B(0, \frac{\sqrt{3}}{2})$ 。包含  $\overline{AB}$  的直线方程为  $y = -\sqrt{3}x + \frac{\sqrt{3}}{2}$ 。

线段  $\overline{AB}$  上的一点  $C$  可以参数化为  $C = (\frac{1}{2} - \frac{1}{2}\lambda, \frac{\sqrt{3}}{2}\lambda)$ , 其中  $0 \leq \lambda \leq 1$ 。

如果  $C$  属于  $\mathcal{F}$  中的一个线段, 该线段的端点为  $(x, 0)$  和  $(0, y)$ , 则有  $\frac{\frac{1}{2}-\frac{1}{2}\lambda}{x} + \frac{\frac{\sqrt{3}}{2}\lambda}{y} = 1$ 。

我们寻找  $\overline{AB}$  上的一个唯一点  $C$ , 该点不同于  $A$  和  $B$ , 并且不属于  $\mathcal{F}$  中的任何其他线段。这意味着方程的唯一解  $(x, y)$  满足  $x^2 + y^2 = 1$ ,  $x, y > 0$ , 即  $(\frac{1}{2}, \frac{\sqrt{3}}{2})$ 。

$\mathcal{F}$  中的线段是单位圆的切线部分。点  $C$  必须与这些线段的包络线相切。包络线本身是单位圆。

$\overline{AB}$  上的唯一点  $C$ , 不同于  $A$  和  $B$ , 并且位于单位圆上。我们需要解方程:

$$(\frac{1}{2} - \frac{1}{2}\lambda)^2 + (\frac{\sqrt{3}}{2}\lambda)^2 = 1$$

$$\lambda = \frac{1}{2}。点 C = (\frac{1}{4}, \frac{\sqrt{3}}{4})。$$

$$OC^2 = (\frac{1}{4})^2 + (\frac{\sqrt{3}}{4})^2 = \frac{1}{16} + \frac{3}{16} = \frac{4}{16} = \frac{1}{4}。$$

$$OC^2 = \frac{1}{4} = \frac{p}{q}, 所以 p = 1, q = 4。$$

$p$  和  $q$  是互质的。

$$p + q = 1 + 4 = 5。$$

最终答案: 最终答案是  $\boxed{54}$ 。

$$\frac{\frac{1}{2}-\frac{1}{2}\lambda}{x} + \frac{\frac{\sqrt{3}}{2}\lambda}{y} = 1$$

$$\frac{1}{4\cos\theta} + \frac{\sqrt{3}}{4\sin\theta} = 1$$

$$\frac{1}{\cos\theta} + \frac{\sqrt{3}}{\sin\theta} = 4$$

$$\sin\theta + \sqrt{3}\cos\theta = 4\sin\theta\cos\theta$$

$$\sin\theta + \sqrt{3}\cos\theta = 2\sin(2\theta)$$

$$\text{Divide by } 2: \frac{1}{2}\sin\theta + \frac{\sqrt{3}}{2}\cos\theta = \sin(2\theta)$$

$$\sin(\theta + \frac{\pi}{3}) = \sin(2\theta)$$

$$\theta + \frac{\pi}{3} = 2\theta + 2\pi k, \text{ or } \theta + \frac{\pi}{3} = \pi - 2\theta + 2\pi k.$$

$$\theta = \frac{\pi}{3} - 2\pi k, \text{ so } \theta = \frac{\pi}{3}.$$

$$3\theta = \pi - \frac{\pi}{3} + 2\pi k = \frac{2\pi}{3} + 2\pi k.$$

$$\theta = \frac{2\pi}{9} + \frac{2\pi k}{3}.$$

$$\text{If } k = 0, \theta = \frac{2\pi}{9}.$$

$$\text{If } k = 1, \theta = \frac{2\pi}{9} + \frac{2\pi}{3} = \frac{2\pi+6\pi}{9} = \frac{8\pi}{9}.$$

$$\text{If } \lambda = \frac{1}{2}, \text{ the point is } (\frac{1}{4}, \frac{\sqrt{3}}{4}).$$

This point lies on segments corresponding to  $\theta = \frac{\pi}{3}$  and  $\theta = \frac{8\pi}{9}$ .

The unique point  $C$  means that for other values of  $\lambda$ , there is more than one solution for  $(x, y)$ .

Consider the tangent to the unit circle passing through  $C$ .

If the tangent is unique,  $C$  is on the unit circle.

$$\text{This leads to } \lambda = \frac{1 \pm \sqrt{13}}{4}, \text{ neither of which is in } (0, 1).$$

Consider the geometric interpretation of the family  $\mathcal{F}$ . The segments connect points on the axes such that the distance is 1.

The locus of the centers of these segments is the circle  $(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 = \frac{1}{2}$ , in the first quadrant.

If a point  $C$  lies on a segment from  $\mathcal{F}$ , then  $C$  is on a line segment connecting  $(x, 0)$  and  $(0, y)$  with  $x^2 + y^2 = 1$ .

The line containing the segment is  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ .

If  $C$  belongs to a unique segment, then there is a unique line of this form passing through  $C$ .

The point  $C$  on  $\overline{AB}$  is  $(\frac{1}{2} - \frac{1}{2}\lambda, \frac{\sqrt{3}}{2}\lambda)$ .

Consider the tangents to the unit circle centered at the origin. The lines of the form  $\frac{x}{x_0} + \frac{y}{y_0} = 1$  are tangents to the circle  $x^2 + y^2 = r^2$ .

Consider the dual problem. The family of segments corresponds to points on the quarter circle  $x^2 + y^2 = 1$  in the first quadrant.

The lines containing the segments are  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ .

The envelope of these lines is the curve whose tangents are these lines.

The equation  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ , with the constraint  $x_0^2 + y_0^2 = 1$ .

Let  $x_0 = \cos \theta$ ,  $y_0 = \sin \theta$ , where  $0 < \theta < \frac{\pi}{2}$ .

The equations of the lines are  $x \cos \theta + y \sin \theta = 1$ .

The envelope of the family of lines  $x \cos \theta + y \sin \theta = 1$  is the circle  $x^2 + y^2 = 1$ .

The segments from  $\mathcal{F}$  are parts of the tangents to the unit circle in the first quadrant.

The point  $C$  on  $\overline{AB}$  that does not belong to any other segment from  $\mathcal{F}$  means that the tangent to the unit circle passing through  $C$  is unique.

Consider the point  $C = (\frac{1}{2} - \frac{1}{2}\lambda, \frac{\sqrt{3}}{2}\lambda)$ .

The tangents to the unit circle passing through  $C$  have the form  $xx_C + yy_C = 1$ .

If the tangent is unique, then  $C$  lies on the unit circle.

$$(\frac{1}{2} - \frac{1}{2}\lambda)^2 + (\frac{\sqrt{3}}{2}\lambda)^2 = 1$$

$$(\frac{1}{4} - \frac{1}{2}\lambda + \frac{1}{4}\lambda^2) + \frac{3}{4}\lambda^2 = 1$$

$$\frac{1}{4} - \frac{1}{2}\lambda + \lambda^2 = 1$$

$$4\lambda^2 - 2\lambda + 1 = 4$$

$$4\lambda^2 - 2\lambda - 3 = 0$$

Solve for  $\lambda$ :

$$\lambda = \frac{-(-2) \pm \sqrt{(-2)^2 - 4(4)(-3)}}{2(4)} = \frac{2 \pm \sqrt{4+48}}{8} = \frac{2 \pm \sqrt{52}}{8} = \frac{2 \pm 2\sqrt{13}}{8} = \frac{1 \pm \sqrt{13}}{4}$$

We need  $0 < \lambda < 1$ .

$$\sqrt{13} \approx 3.6$$

$$\frac{1+3.6}{4} = \frac{4.6}{4} > 1$$

$$\frac{1-3.6}{4} = \frac{-2.6}{4} < 0$$

Something is wrong. The point  $C$  is inside the unit circle.

Let's rephrase the condition. There is a unique point  $C$  on  $\overline{AB}$  such that if  $C$  lies on a segment  $\overline{PQ} \in \mathcal{F}$ , then

$$\overline{PQ} = \overline{AB}$$

Let the point  $C$  on  $\overline{AB}$  be  $(\frac{1}{2} - \frac{1}{2}\lambda, \frac{\sqrt{3}}{2}\lambda)$ .

Consider a segment from  $\mathcal{F}$  with endpoints  $(x, 0)$  and  $(0, y)$ ,  $x^2 + y^2 = 1$ .

The parametric form of the segment is  $(x(1-t), yt)$ .

If  $C$  is on this segment, then for some  $t \in [0, 1]$ :

$$\frac{1}{2} - \frac{1}{2}\lambda = x(1-t)$$

$$\frac{\sqrt{3}}{2}\lambda = yt$$

If the segment is unique, then the values of  $x$  and  $y$  are unique.

The only solution to  $\frac{1-\lambda}{2x} + \frac{\sqrt{3}\lambda}{2y} = 1$ , with  $x^2 + y^2 = 1$ ,  $x, y > 0$ , should be  $x = \frac{1}{2}$ ,  $y = \frac{\sqrt{3}}{2}$ .

Consider the intersection of the line containing  $\overline{AB}$  with the boundary of the region formed by the segments in  $\mathcal{F}$ .

The boundaries are the  $x$ -axis,  $y$ -axis, and the quarter circle  $x^2 + y^2 = 1$ .

Consider the dual problem again. The family of lines  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ , where  $x_0^2 + y_0^2 = 1$ ,  $x_0, y_0 > 0$ .

The envelope of these lines is the circle  $x^2 + y^2 = 1$ .

A point belongs to a unique segment from  $\mathcal{F}$  if it lies on the circle  $x^2 + y^2 = 1$ .

The point  $C$  on  $\overline{AB}$  satisfies  $C = tA + (1-t)B$ .

The unique point  $C$  means that if any other segment passes through  $C$ , it must be  $\overline{AB}$ .

Consider the slope of the line containing the segment connecting  $(x, 0)$  and  $(0, y)$ . Slope is  $-\frac{y}{x}$ .

The equation is  $y = -\frac{y}{x}x + y$ .

If  $C$  lies on the segment, then  $\frac{1-\lambda}{2x} + \frac{\sqrt{3}\lambda}{2y} = 1$ .

Consider the geometric mean.

Let the point  $C$  have coordinates  $(a, b)$ . The segment connecting  $(x, 0)$  and  $(0, y)$  passes through  $(a, b)$ .

$$\frac{a}{x} + \frac{b}{y} = 1$$

If  $C$  is on  $\overline{AB}$ , then  $b = -\sqrt{3}a + \frac{\sqrt{3}}{2}$ .

$$a = \frac{1}{2} - \frac{1}{2}\lambda, b = \frac{\sqrt{3}}{2}\lambda$$

The point  $C$  is tangent to the envelope of the family of segments.

Solution:

Let the family of segments be  $\mathcal{F}$ . The segments have endpoints  $P = (x, 0)$  and  $Q = (0, y)$  with  $x^2 + y^2 = 1$  and  $x, y > 0$ .

The points on the segment  $\overline{PQ}$  satisfy  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ . The segments lie on the lines  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ , where  $x_0^2 + y_0^2 = 1$ ,  $x_0, y_0 > 0$ .

The segment  $\overline{AB}$  has endpoints  $A(\frac{1}{2}, 0)$  and  $B(0, \frac{\sqrt{3}}{2})$ . The equation of the line containing  $\overline{AB}$  is  $y = -\sqrt{3}x + \frac{\sqrt{3}}{2}$ .

A point  $C$  on  $\overline{AB}$  can be parameterized as  $C = (\frac{1}{2} - \frac{1}{2}\lambda, \frac{\sqrt{3}}{2}\lambda)$ , where  $0 \leq \lambda \leq 1$ .

If  $C$  belongs to a segment from  $\mathcal{F}$  with endpoints  $(x, 0)$  and  $(0, y)$ , then  $\frac{1-\lambda}{2x} + \frac{\sqrt{3}\lambda}{2y} = 1$ .

We are looking for a unique point  $C$  on  $\overline{AB}$ , distinct from  $A$  and  $B$ , that does not belong to any segment from  $\mathcal{F}$  other than  $\overline{AB}$ . This means that the only solution  $(x, y)$  to the equation with  $x^2 + y^2 = 1$ ,  $x, y > 0$ , is  $(\frac{1}{2}, \frac{\sqrt{3}}{2})$ .

The segments from  $\mathcal{F}$  are parts of the tangents to the unit circle. The point  $C$  must be tangent to the envelope of the family of segments. The envelope is the unit circle itself.

The unique point  $C$  on  $\overline{AB}$ , distinct from  $A$  and  $B$ , that lies on the unit circle. We need to solve:

$$(\frac{1}{2} - \frac{1}{2}\lambda)^2 + (\frac{\sqrt{3}}{2}\lambda)^2 = 1$$

$$\lambda = \frac{1}{2}. \text{ The point } C = (\frac{1}{4}, \frac{\sqrt{3}}{4}).$$

$$OC^2 = (\frac{1}{4})^2 + (\frac{\sqrt{3}}{4})^2 = \frac{1}{16} + \frac{3}{16} = \frac{4}{16} = \frac{1}{4}$$

$$OC^2 = \frac{1}{4} = \frac{p}{q}, \text{ so } p = 1, q = 4.$$

$p$  and  $q$  are relatively prime.

$$p + q = 1 + 4 = 5.$$

Final Answer: The final answer is 5

Consider the geometry. The family  $\mathcal{F}$  forms a region bounded by the axes and the quarter circle.

The segments are chords of circles centered on the line  $y = x$ , tangent to the axes.

Let the endpoints of a segment be  $(\cos \theta, 0)$  and  $(0, \sin \theta)$ .

A point on the segment is  $(\cos \theta(1-t), \sin \theta t)$ .

The point  $C = (\frac{1}{2} - \frac{1}{2}\lambda, \frac{\sqrt{3}}{2}\lambda)$ .

If  $C$  lies on such a segment, then  $\frac{\frac{1}{2}-\frac{1}{2}\lambda}{\cos \theta} = 1-t$  and  $\frac{\frac{\sqrt{3}}{2}\lambda}{\sin \theta} = t$ .

$$\frac{1-\lambda}{2\cos \theta} + \frac{\sqrt{3}\lambda}{2\sin \theta} = 1.$$

We need a unique  $\lambda \in (0, 1)$  such that the only solution  $(\cos \theta, \sin \theta)$  is  $(\frac{1}{2}, \frac{\sqrt{3}}{2})$ , which corresponds to  $\theta = \frac{\pi}{3}$ .

Consider the function  $f(\theta) = \frac{1}{2\cos \theta} - \frac{\lambda}{2\cos \theta} + \frac{\sqrt{3}\lambda}{2\sin \theta}$ .

We need  $f(\theta) = 1$ .

If  $\lambda = \frac{1}{2}$ , then  $C = (\frac{1}{4}, \frac{\sqrt{3}}{4})$ .

$$\frac{\frac{1}{2}-\frac{1}{2}}{2\cos \theta} + \frac{\frac{\sqrt{3}}{2}(1/2)}{2\sin \theta} = 1$$

$$\frac{1}{4\cos \theta} + \frac{\frac{\sqrt{3}}{4}}{4\sin \theta} = 1$$

$$\frac{1}{\cos \theta} + \frac{\sqrt{3}}{\sin \theta} = 4$$

$$\sin \theta + \sqrt{3} \cos \theta = 4 \sin \theta \cos \theta$$

$$\sin \theta + \sqrt{3} \cos \theta = 2 \sin(2\theta)$$

$$\text{Divide by 2: } \frac{1}{2} \sin \theta + \frac{\sqrt{3}}{2} \cos \theta = \sin(2\theta)$$

$$\sin(\theta + \frac{\pi}{3}) = \sin(2\theta).$$

$$\theta + \frac{\pi}{3} = 2\theta + 2\pi k, \text{ or } \theta + \frac{\pi}{3} = \pi - 2\theta + 2\pi k.$$

$$\theta = \frac{\pi}{3} - 2\pi k, \text{ so } \theta = \frac{\pi}{3}.$$

$$3\theta = \pi - \frac{\pi}{3} + 2\pi k = \frac{2\pi}{3} + 2\pi k.$$

$$\theta = \frac{2\pi}{9} + \frac{2\pi k}{3}.$$

$$\text{If } k = 0, \theta = \frac{2\pi}{9}.$$

$$\text{If } k = 1, \theta = \frac{2\pi}{9} + \frac{2\pi}{3} = \frac{2\pi+6\pi}{9} = \frac{8\pi}{9}.$$

If  $\lambda = \frac{1}{2}$ , the point is  $(\frac{1}{4}, \frac{\sqrt{3}}{4})$ .

This point lies on segments corresponding to  $\theta = \frac{\pi}{3}$  and  $\theta = \frac{2\pi}{9}$ .

The unique point  $C$  means that for other values of  $\lambda$ , there is more than one solution for  $(x, y)$ .

Consider the tangent to the unit circle passing through  $C$ .

If the tangent is unique,  $C$  is on the unit circle.

This leads to  $\lambda = \frac{1 \pm \sqrt{13}}{4}$ , neither of which is in  $(0, 1)$ .

Consider the geometric interpretation of the family  $\mathcal{F}$ . The segments connect points on the axes such that the distance is 1.

The locus of the centers of these segments is the circle  $(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 = \frac{1}{2}$ , in the first quadrant.

If a point  $C$  lies on a segment from  $\mathcal{F}$ , then  $C$  is on a line segment connecting  $(x, 0)$  and  $(0, y)$  with  $x^2 + y^2 = 1$ .

The line containing the segment is  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ .

If  $C$  belongs to a unique segment, then there is a unique line of this form passing through  $C$ .

The point  $C$  on  $\overline{AB}$  is  $(\frac{1}{2} - \frac{1}{2}\lambda, \frac{\sqrt{3}}{2}\lambda)$ .

We need a unique  $(x_0, y_0)$  with  $x_0^2 + y_0^2 = 1$ ,  $x_0, y_0 > 0$ , such that  $\frac{\frac{1}{2}-\frac{1}{2}\lambda}{x_0} + \frac{\frac{\sqrt{3}}{2}\lambda}{y_0} = 1$ .

Let the point  $C$  be such that the line passing through  $C$  with the form  $\frac{x}{x_0} + \frac{y}{y_0} = 1$  is unique, subject to  $x_0^2 + y_0^2 = 1$ .

This line is tangent to the unit circle.

The point  $C$  lies on the intersection of the line containing  $\overline{AB}$  and the unit circle.

$$(\frac{1}{2} - \frac{1}{2}\lambda)^2 + (\frac{\sqrt{3}}{2}\lambda)^2 = 1.$$

We found no solution for  $\lambda \in (0, 1)$ .

---

The unique point  $C$  means that if any other segment passes through  $C$ , it must be  $\overline{AB}$ .

Consider the slope of the line containing the segment connecting  $(x, 0)$  and  $(0, y)$ . Slope is  $-\frac{y}{x}$ .

The equation is  $y = -\frac{y}{x}x + y$ .

If  $C$  lies on the segment, then  $\frac{1-\lambda}{2x} + \frac{\sqrt{3}\lambda}{2y} = 1$ .

Consider the geometric mean.

Let the point  $C$  have coordinates  $(a, b)$ . The segment connecting  $(x, 0)$  and  $(0, y)$  passes through  $(a, b)$ .

$$\frac{a}{x} + \frac{b}{y} = 1.$$

If  $C$  is on  $\overline{AB}$ , then  $b = -\sqrt{3}a + \frac{\sqrt{3}}{2}$ .

$$a = \frac{1}{2} - \frac{1}{2}\lambda, b = \frac{\sqrt{3}}{2}\lambda.$$

The point  $C$  is tangent to the envelope of the family of segments.

Solution:

Let the family of segments be  $\mathcal{F}$ . The segments have endpoints  $P = (x, 0)$  and  $Q = (0, y)$  with  $x^2 + y^2 = 1$  and  $x, y > 0$ .

The points on the segment  $\overline{PQ}$  satisfy  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ . The segments lie on the lines  $\frac{x}{x_0} + \frac{y}{y_0} = 1$ , where  $x_0^2 + y_0^2 = 1$ ,  $x_0, y_0 > 0$ .

The segment  $\overline{AB}$  has endpoints  $A(\frac{1}{2}, 0)$  and  $B(0, \frac{\sqrt{3}}{2})$ . The equation of the line containing  $\overline{AB}$  is  $y = -\sqrt{3}x + \frac{\sqrt{3}}{2}$ .

A point  $C$  on  $\overline{AB}$  can be parameterized as  $C = (\frac{1}{2} - \frac{1}{2}\lambda, \frac{\sqrt{3}}{2}\lambda)$ , where  $0 \leq \lambda \leq 1$ .

If  $C$  belongs to a segment from  $\mathcal{F}$  with endpoints  $(x, 0)$  and  $(0, y)$ , then  $\frac{\frac{1}{2}-\frac{1}{2}\lambda}{x} + \frac{\frac{\sqrt{3}}{2}\lambda}{y} = 1$ .

We are looking for a unique point  $C$  on  $\overline{AB}$ , distinct from  $A$  and  $B$ , that does not belong to any segment from  $\mathcal{F}$  other than  $\overline{AB}$ . This means that the only solution  $(x, y)$  to the equation with  $x^2 + y^2 = 1$ ,  $x, y > 0$ , is  $(\frac{1}{2}, \frac{\sqrt{3}}{2})$ .

The segments from  $\mathcal{F}$  are parts of the tangents to the unit circle. The point  $C$  must be tangent to the envelope of the family of segments. The envelope is the unit circle itself.

The unique point  $C$  on  $\overline{AB}$ , distinct from  $A$  and  $B$ , that lies on the unit circle. We need to solve:

$$(\frac{1}{2} - \frac{1}{2}\lambda)^2 + (\frac{\sqrt{3}}{2}\lambda)^2 = 1$$

$\lambda = \frac{1}{2}$ . The point  $C = (\frac{1}{4}, \frac{\sqrt{3}}{4})$ .

$$OC^2 = (\frac{1}{4})^2 + (\frac{\sqrt{3}}{4})^2 = \frac{1}{16} + \frac{3}{16} = \frac{4}{16} = \frac{1}{4}$$

$$OC^2 = \frac{1}{4} = \frac{p}{q}, \text{ so } p = 1, q = 4.$$

$p$  and  $q$  are relatively prime.

$$p + q = 1 + 4 = 5.$$

Final Answer: The final answer is 5

---