Listing 4 | An example prompt of using DeepSeek-V3 as a judge.

```
As an advanced reasoning problem evaluation assistant, your primary responsibility is
    to assess the accuracy of provided answers. You will be presented with a reasoning-
    related question, its corresponding reference answer, and an answer requiring
    evaluation.

## Answer Quality Classification
You have to carefully analyze and classify the answer into one of the following two
    levels:
1. **correct**: The answer fully aligns with the reference answer in both reasoning
    process and final conclusion, and address the question without any errors or
    omissions.
2. **incorrect**: The answer contains major errors in key reasoning steps or the final
    conclusion, or completely deviates from the core of the question. This indicates a
    fundamental misunderstanding or error in comprehending the question.

## Question
{question}

## Reference Answer
{reference}

## Answer to be Evaluated
{answer}

## Output Format
You need to combine the question and reference answer, first provide a detailed
    explanation of your analysis of the answer to be evaluated, then conclude with the
    final answer quality classification.
Output the following content in **JSON** format, including two key:
1. 'analysis': analysis of the answer's correctness;
2. 'correctness': correct/incorrect
```

B.3.3. 800K Supervised Data

Reasoning Data We curate a large set of reasoning prompts and generate reasoning trajectories by performing rejection sampling from the checkpoint of the first-stage RL training. In the previous stage, we only included data that could be evaluated using rule-based rewards. However, in this stage, we expand the dataset by incorporating additional data, some of which uses a generative reward model by feeding the ground-truth and model predictions into DeepSeek-V3 for judgment, an example prompt is provided in Listing 4. Additionally, because the model output is sometimes chaotic and difficult to read, we have filtered out chain-of-thought with mixed languages, long paragraphs, and code blocks. For each prompt, we sample multiple responses and retain only the correct ones. In total, we collect about 600k reasoning-related training samples.

Non-Reasoning Data For non-reasoning data, such as writing, factual QA, self-cognition, and translation, we adopt the DeepSeek-V3 pipeline and reuse portions of the SFT dataset of DeepSeek-V3. We also incorporate software engineering-focused data, including program repair and front-end web development, to enhance the model's ability to solve real-world problems. For certain non-reasoning tasks, we call DeepSeek-V3 to generate a potential chain-of-thought before answering the question by prompting. However, for simpler queries, such as "hello" we do not provide a CoT in response. In the end, we collected a total of approximately 200k training samples that are unrelated to reasoning.

When designing our thinking process style, we ask the model to follow key principles: First, keep each paragraph concise and digestible. Short paragraphs make ideas clearer and easier to follow. Second, adopt a conversational tone that feels natural and engaging. We avoid technical formatting like markdown to maintain a smooth reading experience. Third, and most importantly, the thinking process begins by understanding the complete user context. This means analyzing who our users are, what situations they're dealing with, and what they truly need - including those unstated needs that may lie beneath the surface of their initial request.

After eliciting these thinking processes from the model, human annotators meticulously verify the accuracy of the outputs. Our findings indicate that these artificial reasoning traces enhance the model's precision in interpreting user queries. Specifically, they effectively highlight format constraints, clarify user intentions, and elucidate the requisite structure of outputs. This methodological approach facilitates more accurate and responsive interactions between the model and users.

Table 5 | Data Statistics of SFT Data.

Domain	Num Samples	Avg Rounds	Avg Tokens
Math	395285	1.0	6094.2
Code	211129	1.1	7435.7
STEM	10124	1.0	4928.8
Logic	10395	1.0	2739.0
General	177812	1.1	1419.8
Total	804745	1.0	5355.3

SFT Data Statistics Table 5 summarizes the data statistics across various domains, based on approximately 800,000 supervised samples. It is worth noting that the majority of the data consists of single-turn interactions, which may limit the multi-turn conversational capabilities of DeepSeek-R1. We leave the expansion to multi-turn dialogue data as future work. The math-related data are primarily in Chinese and English, spanning a wide range of topics and difficulty levels. These questions are verifiable, either through deterministic rules or by reference to specific ground-truth answers. The code dataset encompasses not only competitive programming problems but also debugging tasks and project-oriented coding queries. STEM and logic-related questions, although smaller in volume compared to mathematics and code, are sourced from publicly available textbooks and online repositories. The general-domain queries are diverse, predominantly in Chinese and English, and include tasks such as creative writing, open-ended question answering, role-playing scenarios, and more.

Examples of SFT Trajectories We provide examples of SFT trajectories related to math, code, and non-reasoning in Listings 5, 6, and 7.

Listing 5 | An example SFT trajectory from reasoning data related to mathematics.

```
## Question
For positive integer $n$, define $S_n$ to be the minimum value of the sum \[ \sum_{k
    =1}^n \sqrt{(2k-1)^2+a_k^2}, \] where $a_1,a_2,\ldots,a_n$ are positive real
    numbers whose sum is 17. There is a unique positive integer $n$ for which $S_n$ is
    also an integer. Find this $n$.

## Response
```

```
<think>
Alright, so I'm trying to solve this problem where I need to find a positive integer \(
     n \) such that \( S_n \), the minimum value of the sum \( \sum_{k=1}^n \sqrt{(2k
    -1)^2 + a_k^2} \), is an integer. Here, \( a_1, a_2, \ldots, a_n \) are positive
    real numbers that add up to 17.

Hmm, okay, let me break this down. I know that when you have a sum of square roots like
     this, it often relates to geometry, maybe minimizing distances. It reminds me of
    the problem where you have to break up a path into segments with certain slopes to
    minimize total distance. Maybe this is similar to finding the shortest path in a
    staircase.

Let me visualize it. Each term \( \sqrt{(2k - 1)^2 + a_k^2} \) looks like the
    hypotenuse of a right-angled triangle with legs \( (2k - 1) \) and \( a_k \). So,
    if I imagine a grid where I move horizontally by \( a_k \) and vertically by \( (2k
     - 1) \) for each step \( k \), then \( S_n \) would be the total length of this
    path.

Now, the sum of all \( a_k \) is 17, which would be the total horizontal distance. I
    need to figure out the vertical distances, which are fixed based on \( k \). So the
     total vertical distance would be \( \sum_{k=1}^n (2k - 1) \). Let me compute that:

The sum \( \sum_{k=1}^n (2k - 1) \) is equal to \( 2 \sum_{k=1}^n k - \sum_{k=1}^n 1 =
    2 \cdot \frac{n(n + 1)}{2} - n = n(n + 1) - n = n^2 \). So the total vertical
    distance is \( n^2 \).

Wait a second, so this entire setup is like moving from the origin to the point \( (17,
     n^2) \) with \( n \) segments. Each segment has a horizontal component \( a_k \)
    and a vertical component \( (2k - 1) \). But in reality, this entire path from \(
    (0, 0) \) to \( (17, n^2) \) can be considered as a single straight line if we
    could somehow combine all these segments.

But we can't change the number of segments \( n \); each segment's vertical component
    is fixed. However, the horizontal components \( a_k \) can be adjusted as long as
    their sum is 17. So, to find the minimal total distance, maybe we can think of it
    as if each segment is contributing to the total path optimally. If we scale each
    horizontal component \( a_k \) proportionally, maybe this also corresponds to
    scaling the vertical components.

Alternatively, I can think of calculus optimization techniques. If I need to minimize
    the sum \( \sum_{k=1}^n \sqrt{(2k - 1)^2 + a_k^2} \) subject to \( \sum_{k=1}^n a_k
     = 17 \), I can use the method of Lagrange multipliers.

Let me set up the Lagrangian. Let \( L = \sum_{k=1}^n \sqrt{(2k - 1)^2 + a_k^2} + \
    lambda \left(17 - \sum_{k=1}^n a_k\right) \).

Taking the derivative of \( L \) with respect to each \( a_k \):

\( \frac{\partial L}{\partial a_k} = \frac{a_k}{\sqrt{(2k - 1)^2 + a_k^2}} - \lambda =
    0 \).

So, for each \( k \), we have \( \frac{a_k}{\sqrt{(2k - 1)^2 + a_k^2}} = \lambda \).
    Let me solve for \( a_k \):

Squaring both sides gives \( \frac{a_k^2}{(2k - 1)^2 + a_k^2} = \lambda^2 \).

Rearranging terms: \( a_k^2 = \lambda^2 \left( (2k - 1)^2 + a_k^2 \right) \).

Divide both sides by \( a_k^2 \) (assuming \( a_k \neq 0 \)): \( 1 = \lambda^2 \left( \
    frac{(2k - 1)^2}{a_k^2} + 1 \right) \).
```

```
Simplify: \( 1 - \lambda^2 = \lambda^2 \cdot \frac{(2k - 1)^2}{a_k^2} \).

Then, \( \frac{(2k - 1)^2}{a_k^2} = \frac{1 - \lambda^2}{\lambda^2} \).

Therefore, \( \frac{2k - 1}{a_k} = \sqrt{\frac{1 - \lambda^2}{\lambda^2}} = \frac{\sqrt{1 - \lambda^2}}{\lambda} \).

Let me denote \( c = \frac{\sqrt{1 - \lambda^2}}{\lambda} \). Then, \( a_k = \frac{2k - 1}{c} \).

So, each \( a_k \) is proportional to \( (2k - 1) \). That is, \( a_k = \frac{2k - 1}{c} \).

Since the sum of all \( a_k \) is 17, we can write:

\( \sum_{k=1}^n a_k = \sum_{k=1}^n \frac{2k - 1}{c} = \frac{1}{c} \sum_{k=1}^n (2k - 1) \).

Earlier, I found that \( \sum_{k=1}^n (2k - 1) = n^2 \). Therefore,

\( \frac{n^2}{c} = 17 \) => \( c = \frac{n^2}{17} \).

Thus, \( a_k = \frac{2k - 1}{c} = \frac{(2k - 1) \cdot 17}{n^2} \).

Now, going back to the original expression for \( S_n \):

\( S_n = \sum_{k=1}^n \sqrt{(2k - 1)^2 + a_k^2} \).

Substituting \( a_k = \frac{17(2k - 1)}{n^2} \), we get:

\( \sqrt{(2k - 1)^2 + \left( \frac{17(2k - 1)}{n^2} \right)^2 } = (2k - 1) \sqrt{1 + \left( \frac{17}{n^2} \right)^2 } \).

So, factor that square root out of the sum:

\( S_n = \sqrt{1 + \left( \frac{17}{n^2} \right)^2 } \cdot \sum_{k=1}^n (2k - 1) \).

We already know that \( \sum_{k=1}^n (2k - 1) = n^2 \), so this simplifies to:

\( S_n = \sqrt{1 + \left( \frac{17}{n^2} \right)^2 } \cdot n^2 = \sqrt{n^4 + 17^2} = \sqrt{n^4 + 289} \).

Wait, that seems too simplified. Let me check my steps:

1. I set up the Lagrangian correctly.
2. Took the derivative and found \( a_k = \frac{2k - 1}{c} \).
3. Summed them up to get \( c = \frac{n^2}{17} \).
4. Plugged back into \( S_n \) and factored out \( (2k -1) \), so each term becomes \( (2k -1) \sqrt{1 + \left( \frac{17}{n^2} \right)^2 } \).
5. Since \( \sum (2k -1) = n^2 \), so total sum is \( n^2 \times \sqrt{1 + \left( \frac{17}{n^2} \right)^2 } = \sqrt{n^4 + 17^2} \).

Yeah, that seems to check out. So, \( S_n = \sqrt{n^4 + 289} \). Wait, \( 17^2 is 289.

Therefore, \( S_n = \sqrt{n^4 + 289} \). We need this to be an integer.

So, I need to find an integer \( n \) such that \( n^4 + 289 \) is a perfect square.

Let me denote \( S_n = m \), where \( m \) is an integer, so:

\( n^4 + 289 = m^2 \).
```

```
This implies \( m^2 - n^4 = 289 \).

Hmm, 289 is 17^2. So, \( m^2 - n^4 = 17^2 \). That can be written as:

\( m^2 - n^4 = (17)^2 \).

Looking at this equation, it's a difference of squares.

Wait, \( m^2 - n^4 = (m - n^2)(m + n^2) = 17^2 = 289 \).

So, we have \( (m - n^2)(m + n^2) = 289 \).

Since \( m \) and \( n \) are positive integers, both \( m - n^2 \) and \( m + n^2 \) are positive integers and factors of 289. Also, \( m + n^2 > m - n^2 \).

So, since 289 is 17^2, its positive factor pairs are (1, 289) and (17,17).

Hence, the factor pairs are:

1. \( m - n^2 = 1 \) and \( m + n^2 = 289 \).
2. \( m - n^2 = 17 \) and \( m + n^2 = 17 \).

But the second case gives \( m - n^2 = 17 \) and \( m + n^2 = 17 \).

Subtracting the equations: \( (m + n^2) - (m - n^2) = 17 - 17 \) => \( 2n^2 = 0 \) => \( n^2 = 0 \). But \( n \) is a positive integer, so this case is invalid.

So, we only consider the first factor pair: \( m - n^2 = 1 \) and \( m + n^2 = 289 \).

Adding the two equations: \( (m - n^2) + (m + n^2) = 1 + 289 \) => \( 2m = 290 \) => \( m = 145 \).

Subtracting the two equations: \( (m + n^2) - (m - n^2) = 289 - 1 \) => \( 2n^2 = 288 \) => \( n^2 = 144 \) => \( n = 12 \) since \( n \) is positive.

Therefore, \( n = 12 \).

Before I conclude, let me verify that \( S_n = \sqrt{12^4 + 17^2} = \sqrt{20736 + 289} = \sqrt{21025} = 145 \), which is indeed an integer.

Is this the only \( n \)? The question says there's a unique positive integer \( n \), so this must be the one.

**Final Answer**
The unique positive integer \( n \) is \boxed{12}.
</think>

To solve the problem, we need to find a positive integer \( n \) such that \( S_n \), the minimum value of the sum \( \sum_{k=1}^n \sqrt{(2k-1)^2 + a_k^2} \), is an integer. Here, \( a_1, a_2, \ldots, a_n \) are positive real numbers whose sum is 17.

First, we recognize that each term \( \sqrt{(2k - 1)^2 + a_k^2} \) represents the hypotenuse of a right-angled triangle with legs \( (2k - 1) \) and \( a_k \). The total vertical distance is the sum of the vertical components, which is \( \sum_{k=1}^n (2k - 1) = n^2 \). The total horizontal distance is the sum of \( a_k \), which is 17.

To minimize the sum, we use the method of Lagrange multipliers. We find that each \( a_k \) is proportional to \( (2k - 1) \). Thus, \( a_k = \frac{17(2k - 1)}{n^2} \).
```