

Guide Through Complexity

What Makes Good Supervision for Hard Reasoning Tasks?

¹ Xuan He*, ² Da Yin*, ² Nanyun (Violet) Peng,

¹Tsinghua University, ²University of California, Los Angeles

*Equal Contribution

Corresponding to: hexuan21@mails.tsinghua.edu.cn, da.yin@cs.ucla.edu, violetpeng@cs.ucla.edu

- Large Language Models (LLMs) are reaching the point of solving hard math reasoning problems that challenge human experts.

- Large Language Models (LLMs) are reaching the point of solving hard math reasoning problems that challenge human experts.

Model (up to 2024 Sept)	AIME/%	OlympicArena/%
Claude-3.5-Sonnet	10.0	23.2
GPT-4o (2024-08-06)	14.0	28.3
Grok-2	15.2	-
GPT-4o-mini	11.5	27.4
Gemini-1.5-Flash	17.5	-
Gemini-1.5-Pro	13.7	20.0
Llama-3-70B-Instruct	-	11.9

- How can we further effectively supervise these AI systems on super-hard tasks?
- Or, what supervision data should “weak teacher models”, like humans and current LLMs, annotate or synthesize?

- From perspective of data, two directions:

Strategy 1:
annotate on hard
task for supervision



Strategy 2:
annotate on easy
task for supervision

- From perspective of data, two directions:

Strategy 1:
annotate on hard
task for supervision



Strategy 2:
annotate on easy
task for supervision

Pros: The reasoning difficulty matches with our studied downstream tasks

Cons: Quality may be low since they're challenging for humans/current models to annotate

- From perspective of data, two directions:

Strategy 1:
annotate on hard
task for supervision



Strategy 2:
annotate on easy
task for supervision

Pros: The reasoning difficulty matches with our studied downstream tasks

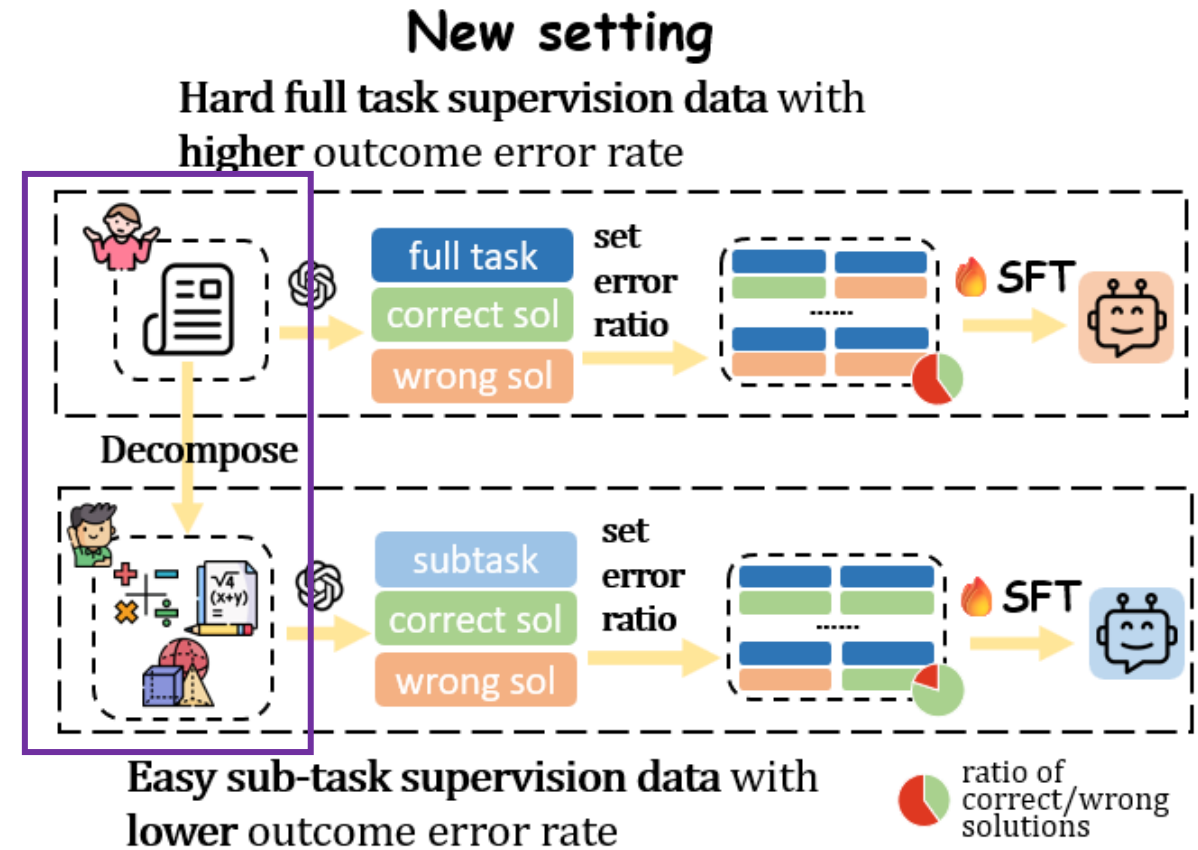
Cons: Quality may be low since they're challenging for humans/current models to annotate

Pros: Quality will be higher since they're simpler to handle

Cons: Difficulty level of training data does not match with test

2. Simulated Setting

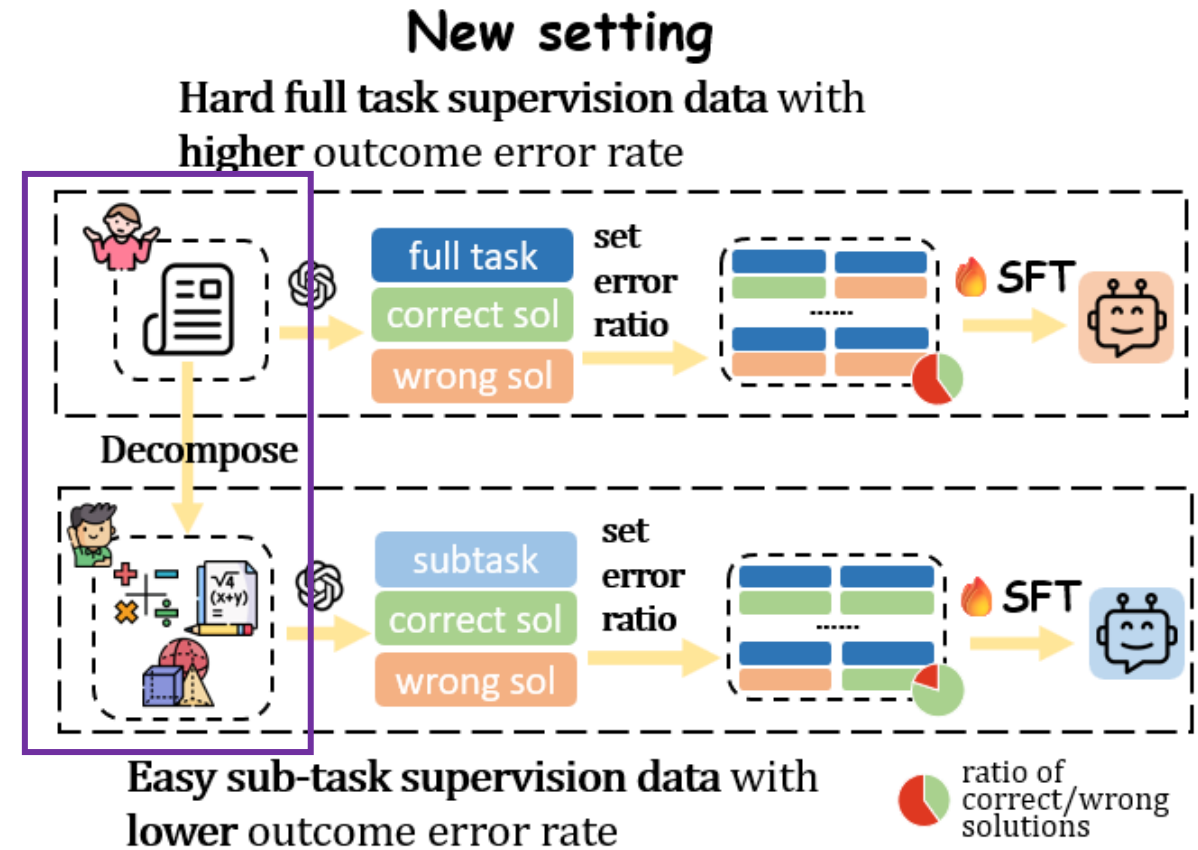
- Compare the 2 strategies in a simulated setting
 - 1. **Decompose subtasks** from hard math problems, original task \rightarrow hard task (strategy 1), subtask \rightarrow easy task (strategy 2)



2. Simulated Setting

- Compare the 2 strategies in a simulated setting
 - 1. **Decompose subtasks** from hard math problems, original task -> hard task (strategy 1), subtask -> easy task (strategy 2)

Key point 1: Hard task is **strictly more complex** than easy task due to decomposition

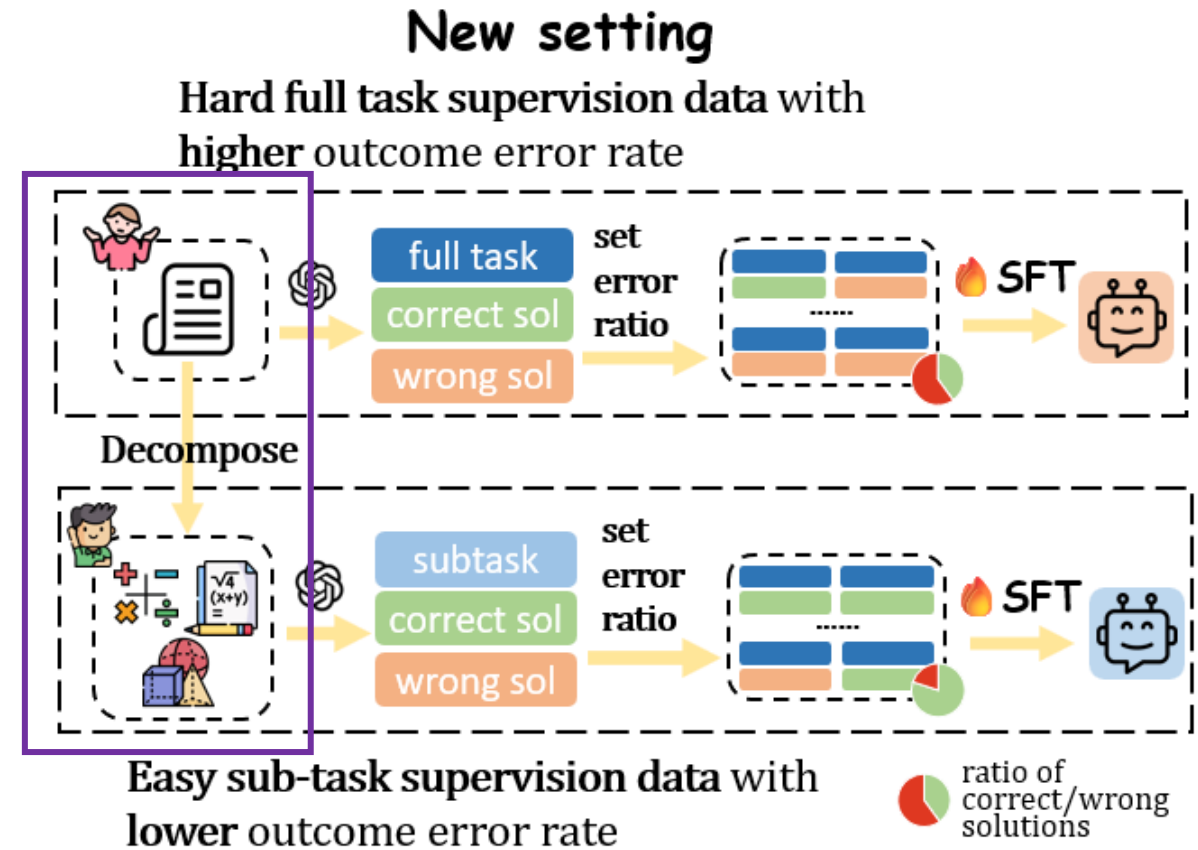


2. Simulated Setting

- Compare the 2 strategies in a simulated setting
 - 1. **Decompose subtasks** from hard math problems, original task \rightarrow hard task (strategy 1), subtask \rightarrow easy task (strategy 2)

Key point 1: Hard task is **strictly more complex** than easy task due to decomposition

Key Point 2: “Knowledge” or “Information” of easy task **is fully covered** by the hard one, **without introducing new info**



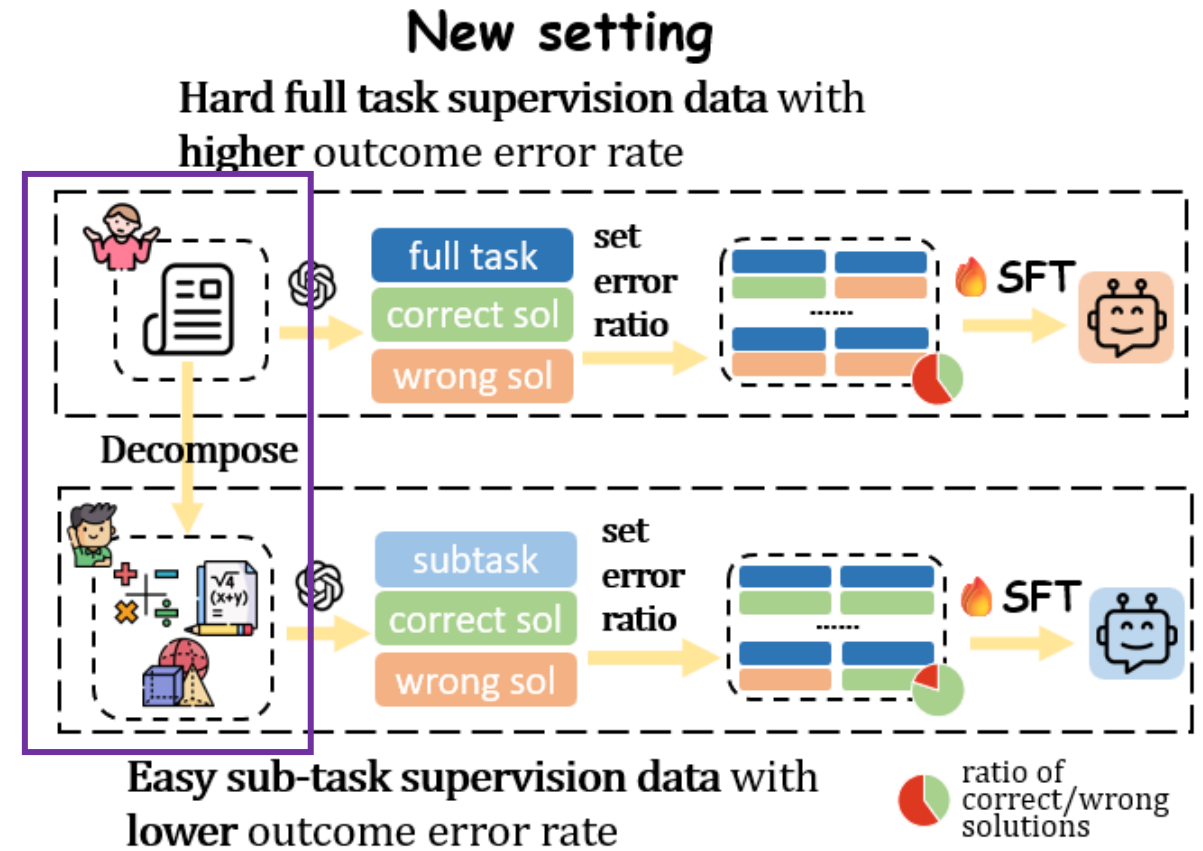
2. Simulated Setting

- Compare the 2 strategies in a simulated setting
 - 1. **Decompose subtasks** from hard math problems, original task \rightarrow hard task (strategy 1), subtask \rightarrow easy task (strategy 2)

Key point 1: Hard task is **strictly more complex** than easy task due to decomposition

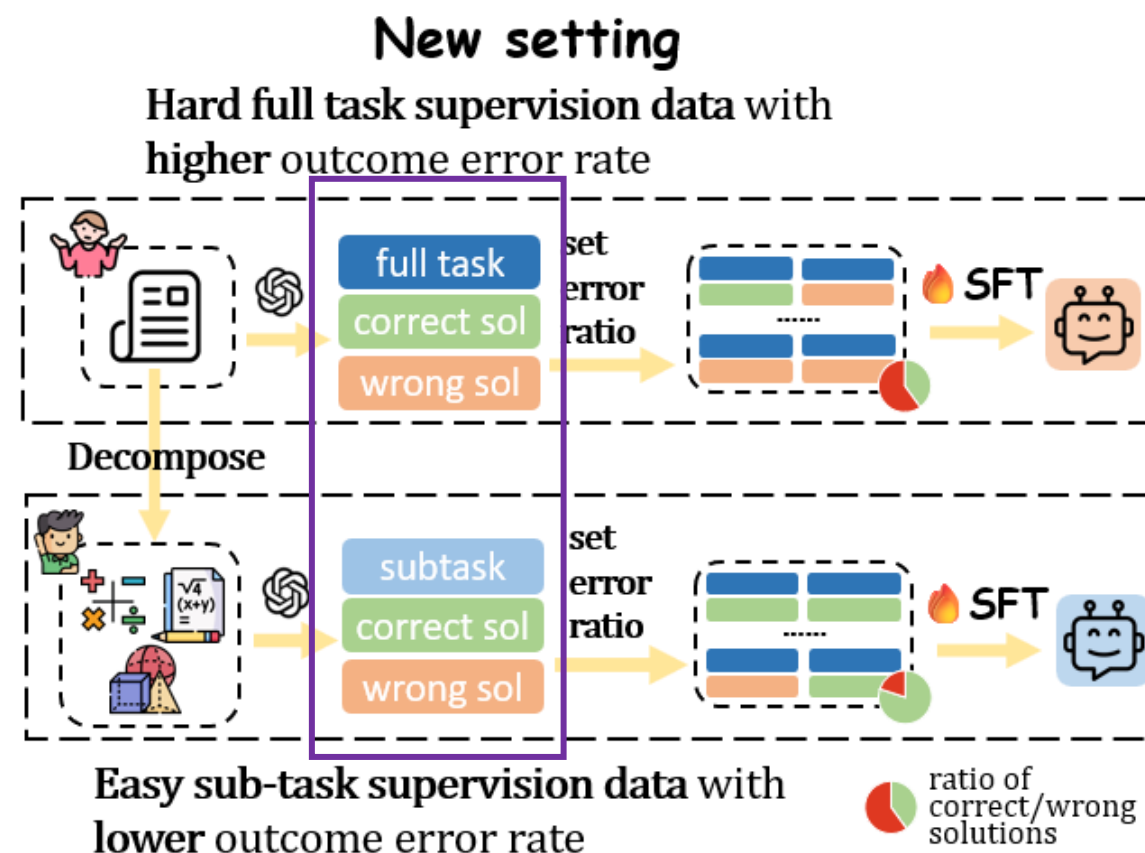
Key Point 2: “Knowledge” or “Information” of easy task **is fully covered** by the hard one, **without introducing new info**

- K1&2 make the comparison fairer, which are ignored by most previous works



2. Simulated Setting

- Compare the 2 strategies in a simulated setting
 - 2. **Generate** correct & incorrect solutions (check final result of the solution to judge the correctness)

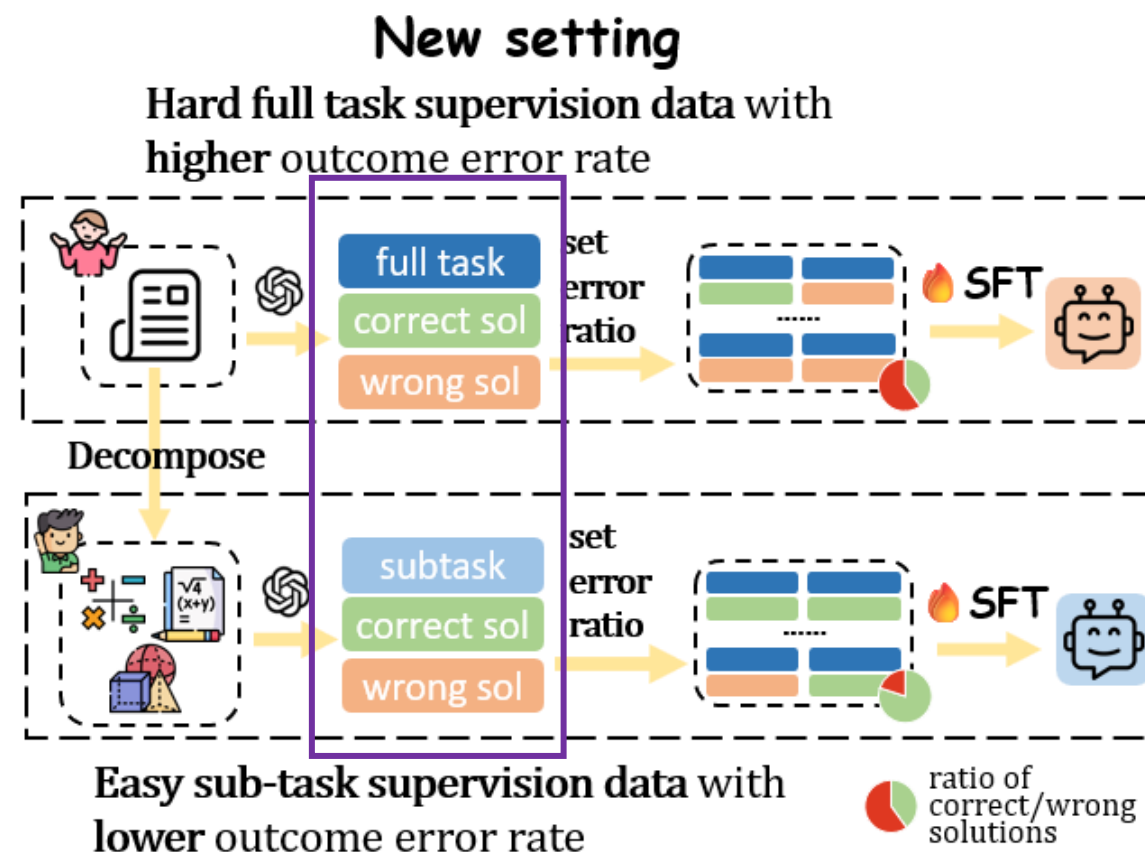


2. Simulated Setting

- Compare the 2 strategies in a simulated setting

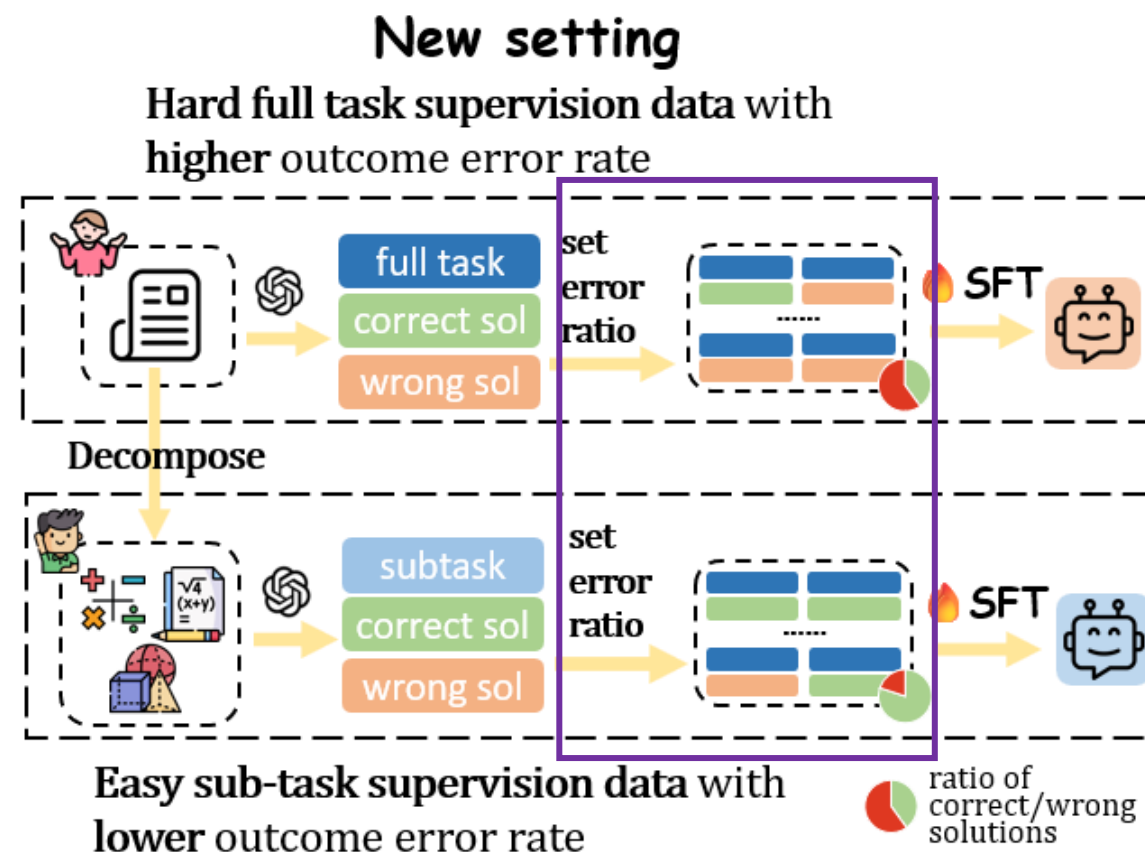
- 2. **Generate** correct & incorrect solutions (check final result of the solution to judge the correctness)

Make sure all full-task and subtask have one correct solution and one wrong solution.



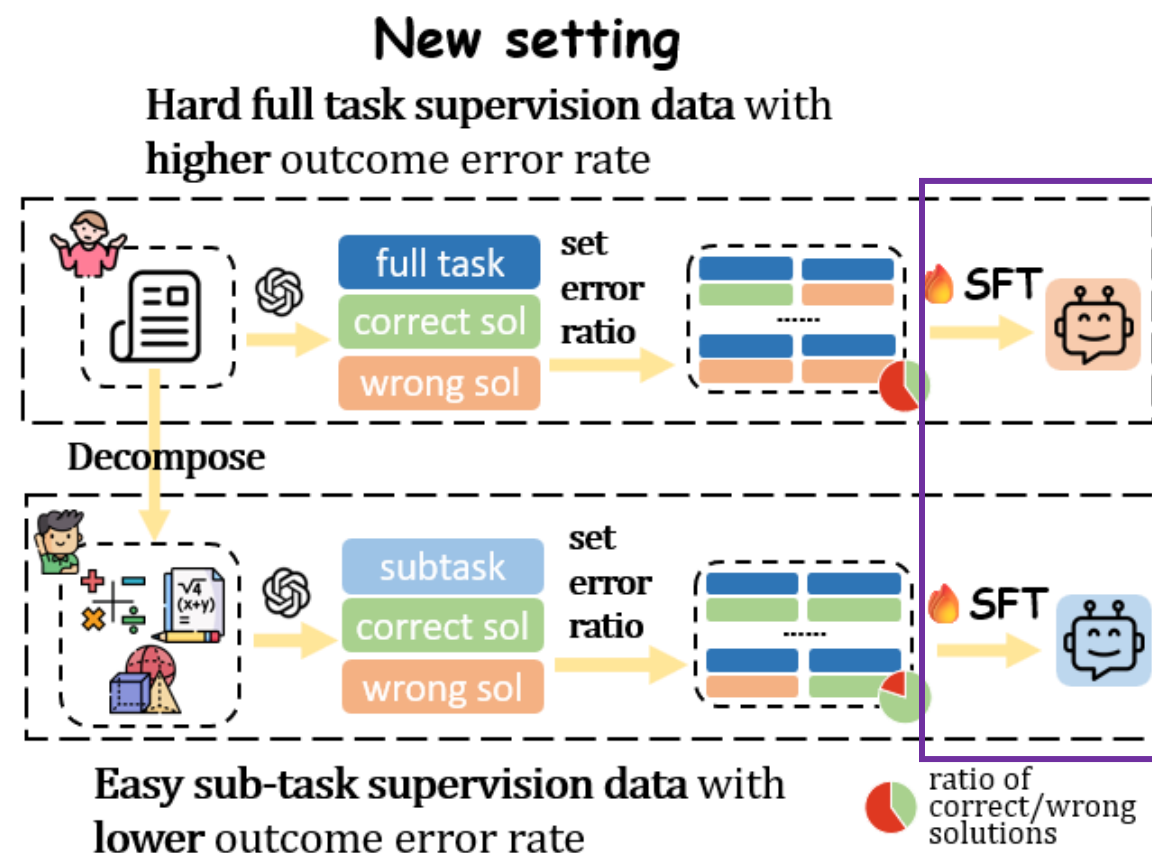
2. Simulated Setting

- Compare the 2 strategies in a simulated setting
 - 2. **Generate** correct & incorrect solutions (check final result of the solution to judge the correctness)
 - 3. **Control supervision quality** by replacing part of correct sols with incorrect ones



2. Simulated Setting

- Compare the 2 strategies in a simulated setting
 - 2. **Generate** correct & incorrect solutions (check final result of the solution to judge the correctness)
 - 3. **Control supervision quality** by replacing part of correct sols with incorrect ones
 - 4. Conduct **Supervised Fine-Tuning** (SFT) on the two types of data.



3. Compare hard and easy supervision data

- **Test finetuned models**

- 5 hard math test sets: MATH-500, Olympic-Arena, SAT, JEE-Bench, Gaokao-math-cloze

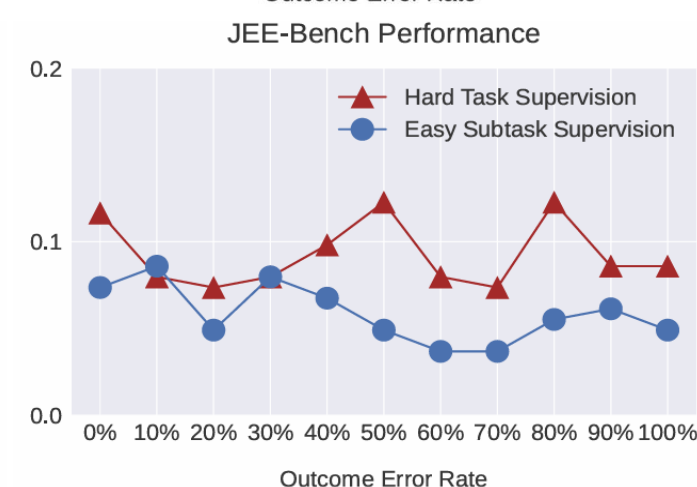
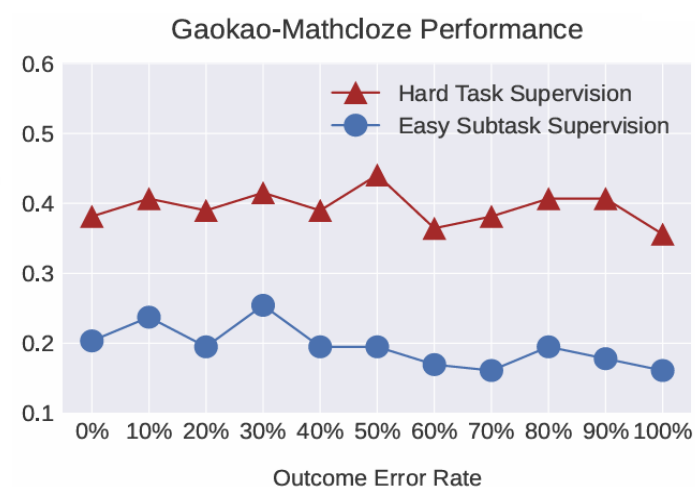
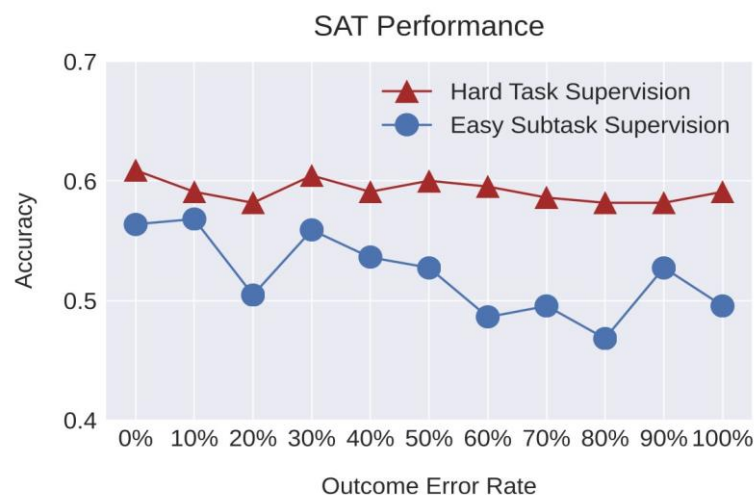
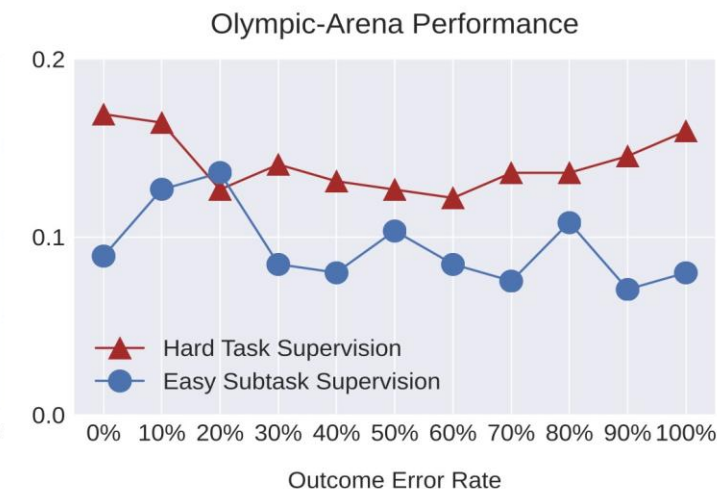
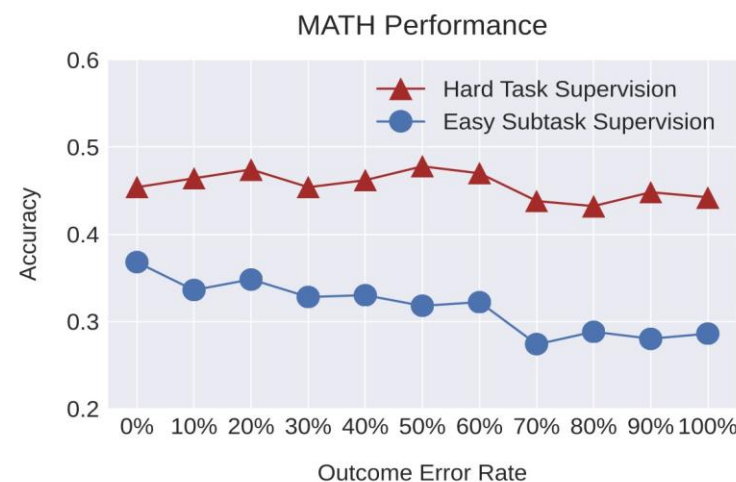
Test Tasks	Sizes	Difficulty
MATH	500	high school competition
JEE-Bench	163	high school & college
Olympic-Arena	213	Olympiad competition
SAT-Math	220	college entrance exam
Gaokao-Mathcloze	118	college entrance exam

3. Compare hard and easy supervision data

- Test finetuned models

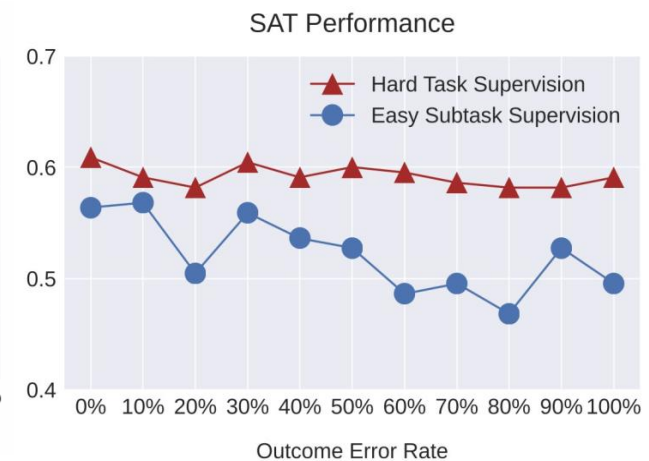
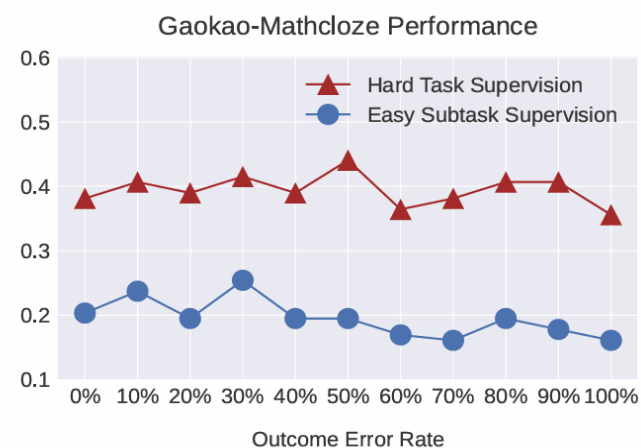
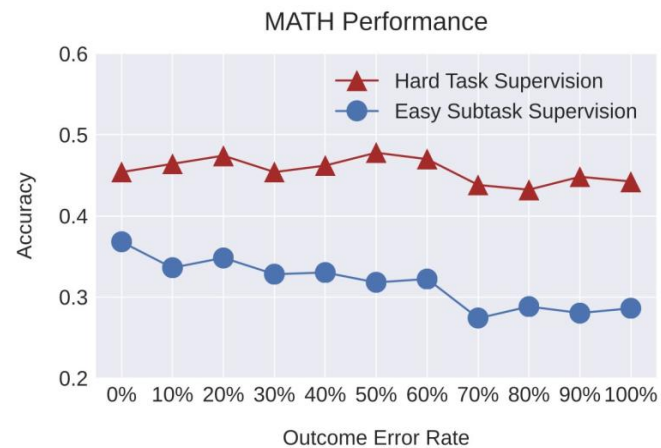
- 5 hard math test sets: MATH-500, Olympic-Arena, SAT, JEE-Bench, Gaokao-math-cloze

Test Tasks	Sizes	Difficulty
MATH	500	high school competition
JEE-Bench	163	high school & college
Olympic-Arena	213	Olympiad competition
SAT-Math	220	college entrance exam
Gaokao-Mathcloze	118	college entrance exam



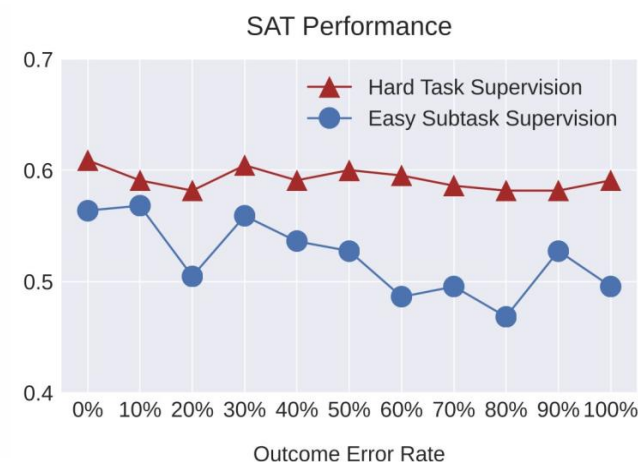
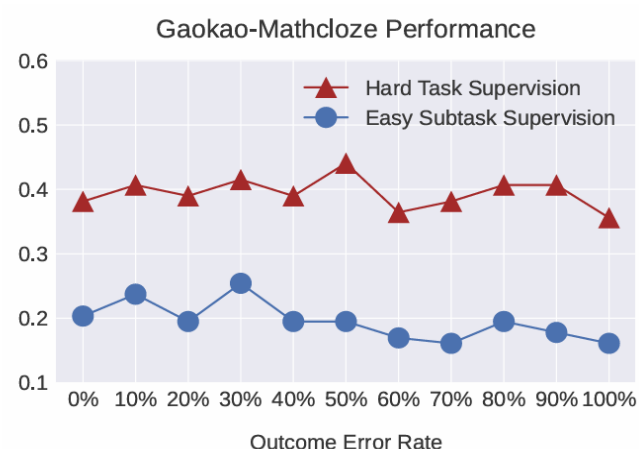
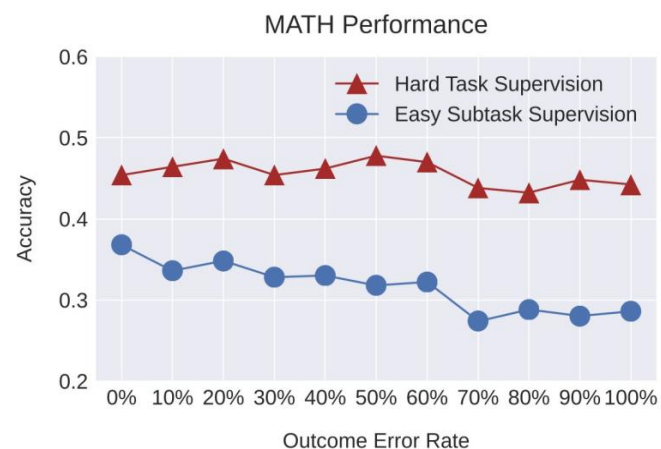
3. Compare hard and easy supervision data

- Observation-1: Hard task supervision **consistently outperforms** subtask supervision, even with **higher outcome error rates**



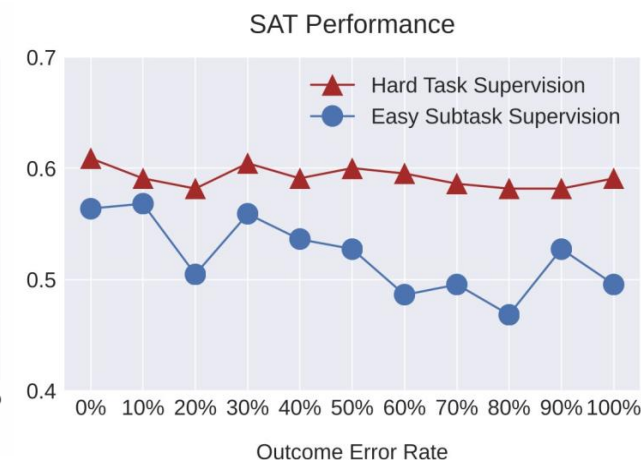
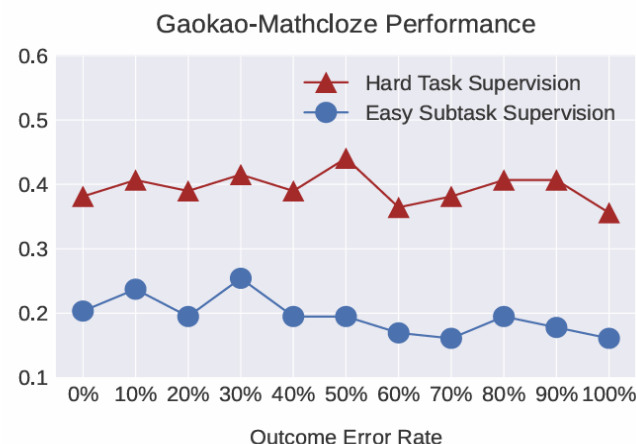
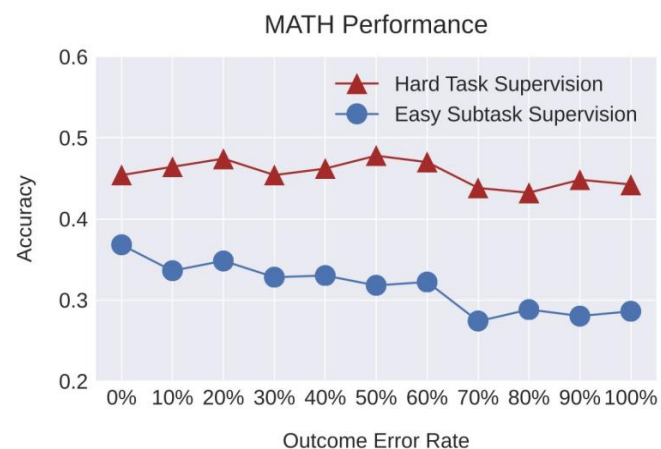
3. Compare hard and easy supervision data

- Observation-1: Hard task supervision **consistently outperforms** subtask supervision, even with **higher outcome error rates**
- Observation-2: Performance **does not consistently degrade** with increasing outcome error rates



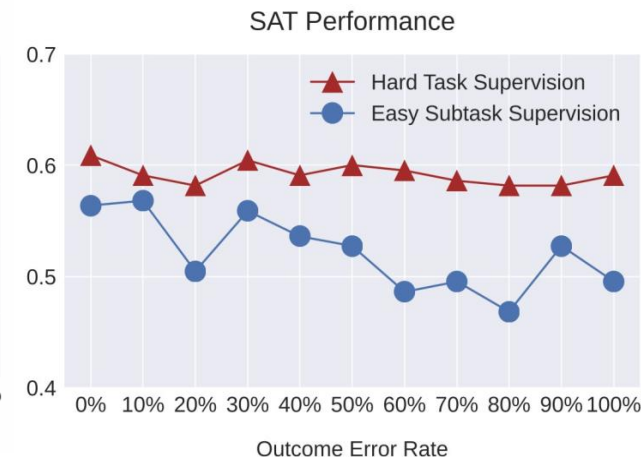
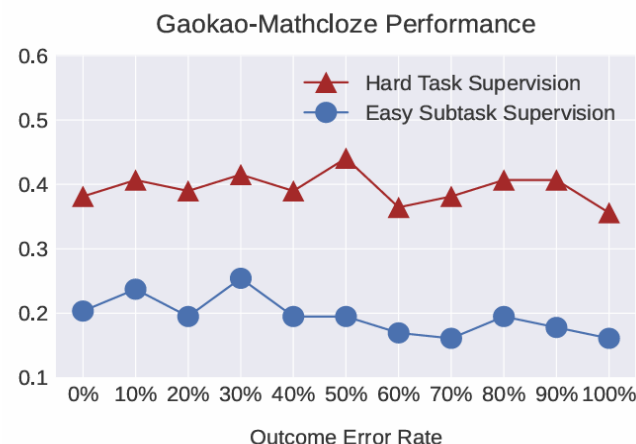
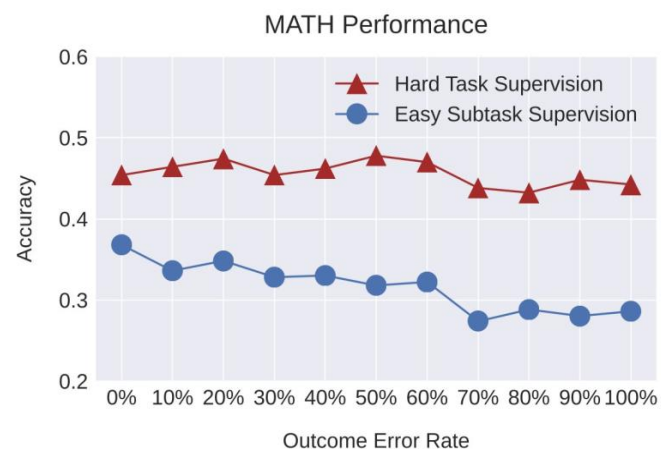
3. Compare hard and easy supervision data

- Observation-1: Hard task supervision **consistently outperforms** subtask supervision, even with **higher outcome error rates**
- Observation-2: Performance **does not consistently degrade** with increasing outcome error rates
- Observation-3: Changes in outcome error rates have a **greater impact on subtask supervision** than on hard task supervision



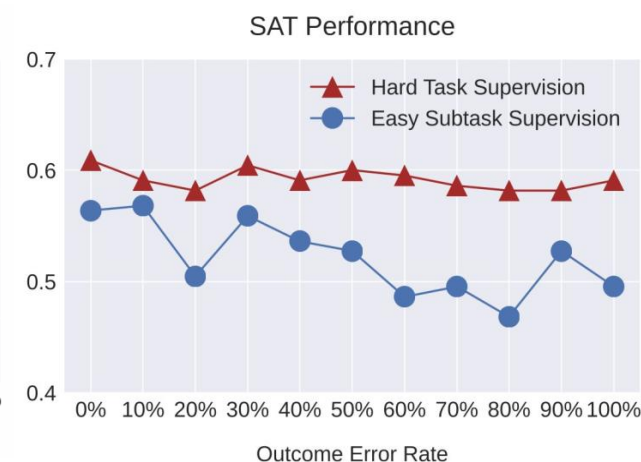
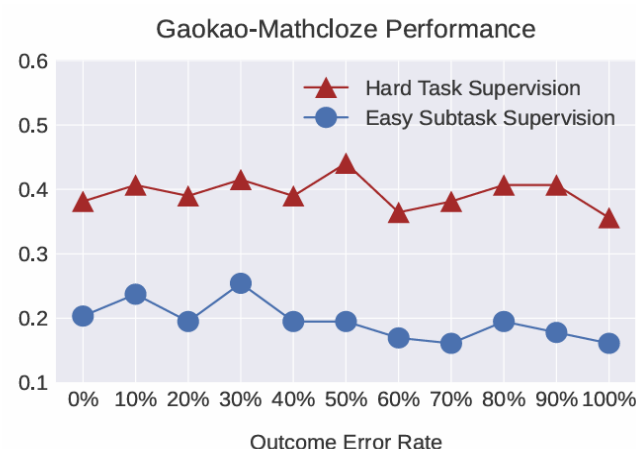
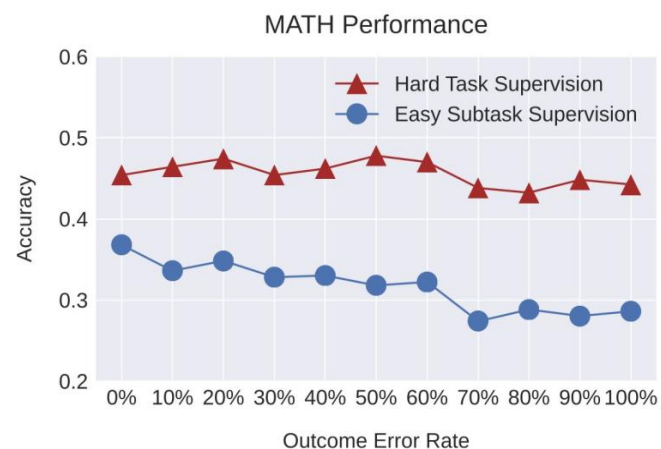
3. Compare hard and easy supervision data

- Observation-1: Hard task supervision **consistently outperforms** subtask supervision, even with **higher outcome error rates**
- Observation-2: Performance **does not consistently degrade** with increasing outcome error rates
- Observation-3: Changes in outcome error rates have a **greater impact on subtask supervision** than on hard task supervision
- **Why does these happen?**



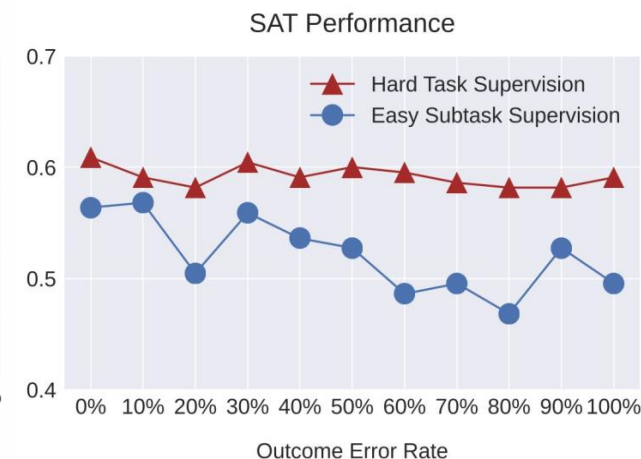
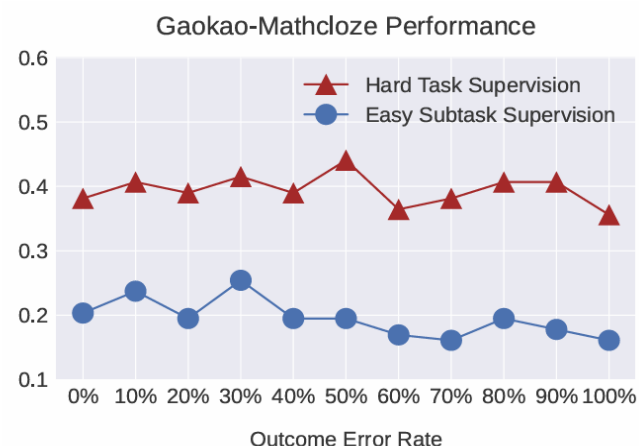
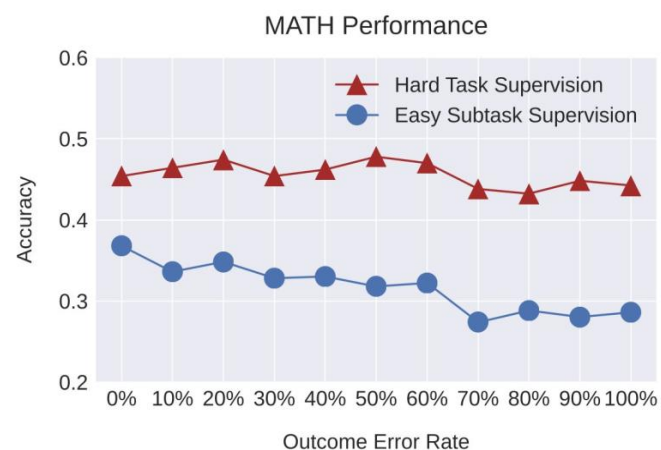
3. Compare hard and easy supervision data

- Observation Summary
 - Irrespective of outcome ER, hard task supervision outperforms subtask one and is robust across varying quality levels.



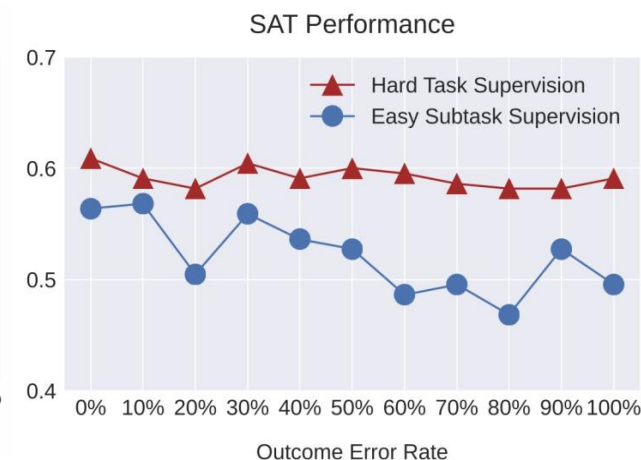
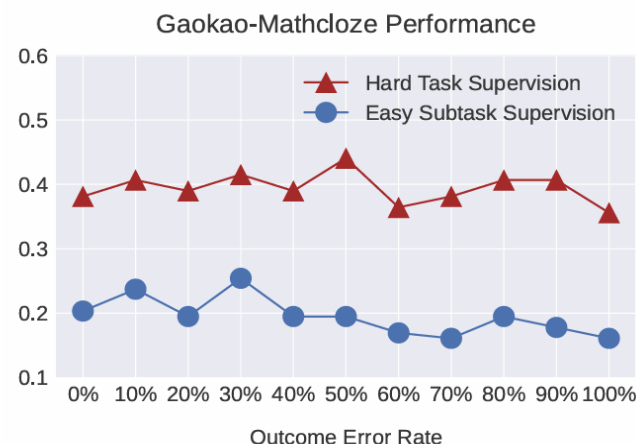
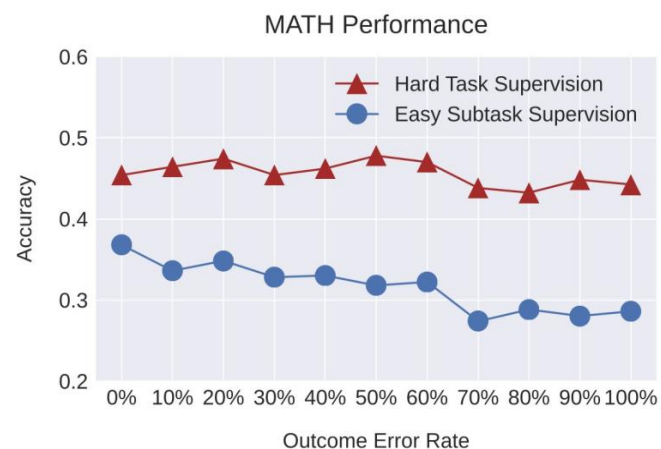
3. Compare hard and easy supervision data

- Observation Summary
 - Irrespective of outcome ER, hard task supervision outperforms subtask one and is robust across varying quality levels.
- Hypothesis & Insights
 - In domains requiring **long-form reasoning**, outcome error rate **MAY NOT** be a reliable indicator for supervision quality.



3. Compare hard and easy supervision data

- Observation Summary
 - Irrespective of outcome ER, hard task supervision outperforms subtask one and is robust across varying quality levels.
- Hypothesis & Insights
 - In domains requiring **long-form reasoning**, outcome error rate **MAY NOT** be a reliable indicator for supervision quality.
 - Look deeper into “**severity of wrong solution**”. Try using ratio of incorrect steps or “**step-wise error rate**”!



4. Severity of wrong solution matters

- **Control the previous metric** – outcome error rate, but **vary the new metric** – step-wise error rate.
 - Use “**weaker teacher models**” like GPT-3.5-turbo, Command-R to generate solutions **with more incorrect intermediate steps**.

$$\text{step wise ER} = \frac{\#(\text{erroneous steps})}{\#(\text{all steps})}$$

4. Severity of wrong solution matters

- **Control the previous metric** – outcome error rate, but **vary the new metric** – step-wise error rate.
 - Use “**weaker teacher models**” like GPT-3.5-turbo, Command-R to generate solutions **with more incorrect intermediate steps**.

$$\text{step wise ER} = \frac{\#(\text{erroneous steps})}{\#(\text{all steps})}$$

Human evaluation result on 50 problems:

Teacher models	Step-wise Error Ratio (%)
GPT-4o-mini	11.7
Llama-3-70B-Instruct	30.2
Command-R-2024-08	40.0
Command-R-2024-03	62.6
GPT-3.5-turbo	34.7

4. Severity of wrong solution matters

- **Control the previous metric** – outcome error rate, but **vary the new metric** – step-wise error rate.
- Use “**weaker teacher models**” like GPT-3.5-turbo, Command-R to generate solutions **with more incorrect intermediate steps**.

$$\text{step wise ER} = \frac{\#(\text{erroneous steps})}{\#(\text{all steps})}$$

Human evaluation result on 50 problems:

Teacher models	Step-wise Error Ratio (%)
GPT-4o-mini	11.7
Llama-3-70B-Instruct	30.2
Command-R-2024-08	40.0
Command-R-2024-03	62.6
GPT-3.5-turbo	34.7

Solution:

Let's denote the position vectors of points A, B, and C as \vec{A} , \vec{B} , and \vec{C} respectively. The position vector of the centroid G of triangle ABC is given by:

$$\vec{G} = \frac{1}{3}(\vec{A} + \vec{B} + \vec{C}) \quad (\text{Step 1: correct})$$

We can express \vec{GA} , \vec{GB} , and \vec{GC} as follows:

$$\vec{GA} = \vec{A} - \vec{G} = \vec{A} - \frac{1}{3}(\vec{A} + \vec{B} + \vec{C}) = \frac{2}{3}\vec{A} - \frac{1}{3}\vec{B} - \frac{1}{3}\vec{C} \quad (\text{Step 2: correct})$$

$$\vec{GB} = \vec{B} - \vec{G} = \frac{2}{3}\vec{B} - \frac{1}{3}\vec{A} - \frac{1}{3}\vec{C} \quad (\text{Step 3: correct})$$

$$\vec{GC} = \vec{C} - \vec{G} = \frac{2}{3}\vec{C} - \frac{1}{3}\vec{A} - \frac{1}{3}\vec{B} \quad (\text{Step 4: correct})$$

Substituting these into the given equation:

$$\frac{2}{3}\vec{GA} + \frac{2}{3}\vec{GB} + \frac{2}{3}\vec{GC} = \vec{0} \quad (\text{Step 5: correct})$$

This becomes:

$$\frac{2}{3}\left(\frac{2}{3}\vec{A} - \frac{1}{3}\vec{B} - \frac{1}{3}\vec{C}\right) + \frac{2}{3}\left(\frac{2}{3}\vec{B} - \frac{1}{3}\vec{A} - \frac{1}{3}\vec{C}\right) + \frac{2}{3}\left(\frac{2}{3}\vec{C} - \frac{1}{3}\vec{A} - \frac{1}{3}\vec{B}\right) = \vec{0} \quad (\text{Step 6: correct})$$

Combining the terms gives:

$$\frac{20}{27}\vec{A} - \frac{8}{27}\vec{B} - \frac{8}{27}\vec{C} + \frac{20}{27}\vec{B} - \frac{8}{27}\vec{A} - \frac{8}{27}\vec{C} + \frac{20}{27}\vec{C} - \frac{8}{27}\vec{A} - \frac{8}{27}\vec{B} = \vec{0} \quad (\text{Step 7: correct})$$

Collecting the coefficients of \vec{A} , \vec{B} , and \vec{C} :

$$\left(\frac{20}{27} - \frac{8}{27} - \frac{8}{27}\right)\vec{A} + \left(-\frac{8}{27} + \frac{20}{27} - \frac{8}{27}\right)\vec{B} + \left(-\frac{8}{27} - \frac{8}{27} + \frac{20}{27}\right)\vec{C} = \vec{0} \quad (\text{Step 8: correct})$$

For this equation to hold, each coefficient must equal zero:

$$\frac{20}{27} - \frac{8}{27} - \frac{8}{27} = 0, -\frac{8}{27} + \frac{20}{27} - \frac{8}{27} = 0, -\frac{8}{27} - \frac{8}{27} + \frac{20}{27} = 0$$

(Step 9: wrong, since \vec{A} , \vec{B} and \vec{C} are all 2-D vectors, the expression $\left(\frac{20}{27} - \frac{8}{27} - \frac{8}{27}\right)\vec{A} + \left(-\frac{8}{27} + \frac{20}{27} - \frac{8}{27}\right)\vec{B} + \left(-\frac{8}{27} - \frac{8}{27} + \frac{20}{27}\right)\vec{C} = \vec{0}$ does not imply all the coefficients are zeros, there exist non-zero solutions for those coefficients.)

After solving, we can use the Law of Cosines to find the angle B.

Finding Angle B:

To find B using the cosine rule:

$$\cos B = \frac{a^2 + c^2 - b^2}{2ac} \quad (\text{Step 10: correct})$$

From the relationships derived from the above equations, substituting a, b, and c will yield the angle B.

After computation, we find:

$$B = 90^\circ \quad (\text{Step 11: correct})$$

Thus, the measure of angle B is:

$$\boxed{90^\circ}$$

from
GPT-4o-mini

4. Severity of wrong solution matters

- Control the previous metric – outcome error rate, but vary the new metric – step-wise error rate.
- Use “weaker teacher models” like GPT-3.5-turbo, Command-R to generate solutions with more incorrect intermediate steps.

$$\text{step wise ER} = \frac{\#(\text{erroneous steps})}{\#(\text{all steps})}$$

Human evaluation result on 50 problems:

Teacher models	Step-wise Error Ratio (%)
GPT-4o-mini	11.7
Llama-3-70B-Instruct	30.2
Command-R-2024-08	40.0
Command-R-2024-03	62.6
GPT-3.5-turbo	34.7

Solution:

Let's denote the position vectors of points A, B, and C as \vec{A} , \vec{B} , and \vec{C} respectively. The position vector of the centroid G of triangle ABC is given by:

$$\vec{G} = \frac{1}{3}(\vec{A} + \vec{B} + \vec{C}) \text{ (Step 1: correct)}$$

We can express \vec{GA} , \vec{GB} , and \vec{GC} as follows:

$$\vec{GA} = \vec{A} - \vec{G} = \vec{A} - \frac{1}{3}(\vec{A} + \vec{B} + \vec{C}) = \frac{2}{3}\vec{A} - \frac{1}{3}\vec{B} - \frac{1}{3}\vec{C} \text{ (Step 2: correct)}$$

$$\vec{GB} = \vec{B} - \vec{G} = \frac{2}{3}\vec{B} - \frac{1}{3}\vec{A} - \frac{1}{3}\vec{C} \text{ (Step 3: correct)}$$

$$\vec{GC} = \vec{C} - \vec{G} = \frac{2}{3}\vec{C} - \frac{1}{3}\vec{A} - \frac{1}{3}\vec{B} \text{ (Step 4: correct)}$$

Substituting these into the given equation:

$$\frac{2}{3}\vec{GA} + \frac{1}{3}\vec{GB} + \frac{2}{3}\vec{GC} = \vec{0} \text{ (Step 5: correct)}$$

This becomes:

$$\frac{2}{3}(\frac{2}{3}\vec{A} - \frac{1}{3}\vec{B} - \frac{1}{3}\vec{C}) + \frac{1}{3}(\frac{2}{3}\vec{B} - \frac{1}{3}\vec{A} - \frac{1}{3}\vec{C}) + \frac{2}{3}(\frac{2}{3}\vec{C} - \frac{1}{3}\vec{A} - \frac{1}{3}\vec{B}) = \vec{0} \text{ (Step 6: correct)}$$

Combining the terms gives:

$$\frac{20}{27}\vec{A} - \frac{6}{27}\vec{B} - \frac{6}{27}\vec{C} + \frac{2}{27}\vec{B} - \frac{1}{27}\vec{A} - \frac{2}{27}\vec{C} + \frac{20}{27}\vec{C} - \frac{6}{27}\vec{A} - \frac{6}{27}\vec{B} = \vec{0} \text{ (Step 7: correct)}$$

Collecting the coefficients of \vec{A} , \vec{B} , and \vec{C} :

$$(\frac{20}{27} - \frac{6}{27} - \frac{6}{27})\vec{A} + (-\frac{6}{27} + \frac{2}{27} - \frac{6}{27})\vec{B} + (-\frac{6}{27} - \frac{6}{27} + \frac{20}{27})\vec{C} = \vec{0} \text{ (Step 8: correct)}$$

For this equation to hold, each coefficient must equal zero:

$$\frac{20}{27} - \frac{6}{27} - \frac{6}{27} = 0, -\frac{6}{27} + \frac{2}{27} - \frac{6}{27} = 0, -\frac{6}{27} - \frac{6}{27} + \frac{20}{27} = 0$$

(Step 9: wrong, since \vec{A} , \vec{B} and \vec{C} are all 2-D vectors, the expression $(\frac{20}{27} - \frac{6}{27} - \frac{6}{27})\vec{A} + (-\frac{6}{27} + \frac{2}{27} - \frac{6}{27})\vec{B} + (-\frac{6}{27} - \frac{6}{27} + \frac{20}{27})\vec{C} = \vec{0}$ does not imply all the coefficients are zeros, there exist non-zero solutions for those coefficients.)

After solving, we can use the Law of Cosines to find the angle B.

Finding Angle B:

To find B using the cosine rule:

$$\cos B = \frac{a^2 + c^2 - b^2}{2ac} \text{ (Step 10: correct)}$$

From the relationships derived from the above equations, substituting a, b, and c will yield the angle B.

After computation, we find:

$$B = 90^\circ \text{ (Step 11: correct)}$$

Thus, the measure of angle B is:

$$90^\circ$$

from
GPT-4o-mini

Solution:

From the given information, we have:

$$\frac{a}{5}\vec{GA} + \frac{b}{7}\vec{GB} + \frac{c}{8}\vec{GC} = \vec{0}$$

$$\implies \frac{a}{5} \cdot \frac{2\vec{HA}}{\sin A} + \frac{b}{7} \cdot \frac{2\vec{HB}}{\sin B} + \frac{c}{8} \cdot \frac{2\vec{HC}}{\sin C} = \vec{0} \text{ (Step 1: wrong, } \vec{HA}, \vec{HB}, \vec{HC} \text{ are not defined.)}$$

$$\implies \frac{2a}{5\sin A} \cdot \vec{HA} + \frac{2b}{7\sin B} \cdot \vec{HB} + \frac{2c}{8\sin C} \cdot \vec{HC} = \vec{0} \text{ (Step 2: correct)}$$

Since the three vectors \vec{HA} , \vec{HB} and \vec{HC} are coplanar and sum to $\vec{0}$, the only way for their scaled versions above to also sum to $\vec{0}$, is for each of them to be zero individually. (Step 3: correct)

$$\text{This means that: } \frac{2a}{5\sin A} = 0 \implies \sin A = \frac{5a}{2} \text{ (Step 4: wrong, } \sin A = \frac{5a}{2} \text{ can not be inferred from } \frac{2a}{5\sin A} = 0.)$$

$$\text{This also implies } \sin B = \frac{5b}{2}, \sin C = \frac{5c}{2}. \text{ (Step 5: wrong, the reason is same as above step.)}$$

Using the fact that $\sin A + \sin B + \sin C = 1$, we get: (Step 6: wrong, $\sin A + \sin B + \sin C = 1$ does not always hold for general cases.)

$$\frac{5a}{2} + \frac{5b}{2} + \frac{5c}{2} = 1 \implies \frac{5}{2}(a + b + c) = 1 \text{ (Step 7: correct)}$$

Since $a + b + c$ is just twice the area of the triangle, we get that the area of $\triangle ABC$ is $\frac{1}{10}$. (Step 8: wrong, it's not mentioned in the problem or we cannot infer that $a + b + c$ is twice the area of the triangle)

In particular, this means that the angle B, which measures $\frac{2\pi}{10}$, or 12° , is the smallest angle of the three. (Step 9: wrong, we cannot infer from the above steps that angle B is $\frac{2\pi}{10}$. Also, $\frac{2\pi}{10} \neq 12^\circ$)

from
Command-R

4. Severity of wrong solution matters

- Calculate the outcome error rate of “weaker teacher’s supervision” (e.g. 87.7%), and select the supervision data from previous teacher (GPT-4o-mini) with closest outcome error rate (e.g. 90%).

from previous teacher
(GPT-4o-mini)



from new “weaker”
teacher (like Command-R)

4. Severity of wrong solution matters

- Calculate the outcome error rate of “weaker teacher’s supervision” (e.g. 87.7%), and select the supervision data from previous teacher (GPT-4o-mini) with closest outcome error rate (e.g. 90%).

from previous teacher
(GPT-4o-mini)



from new “weaker”
teacher (like Command-R)

**Almost the same outcome ER,
but different step-wise ER**

4. Severity of wrong solution matters

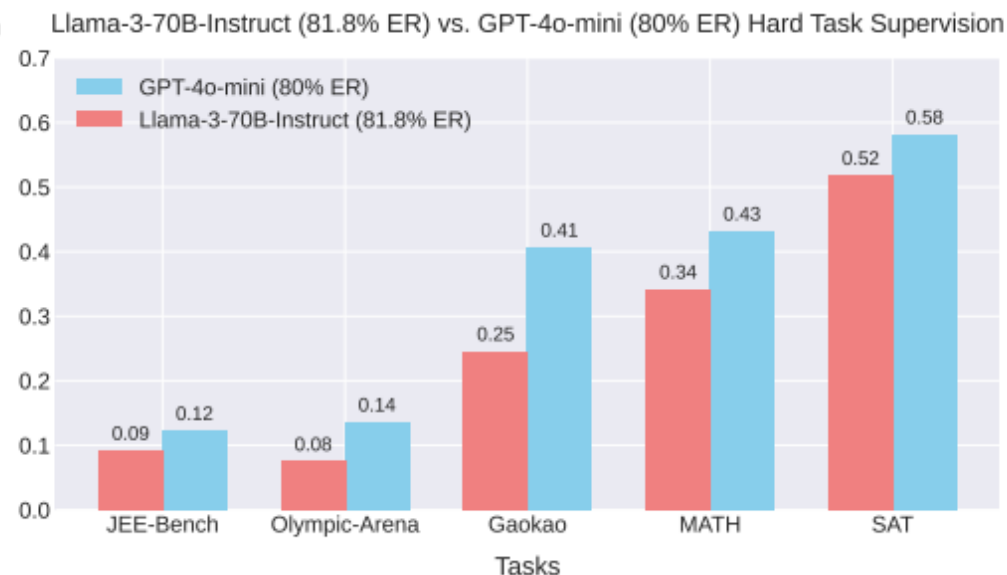
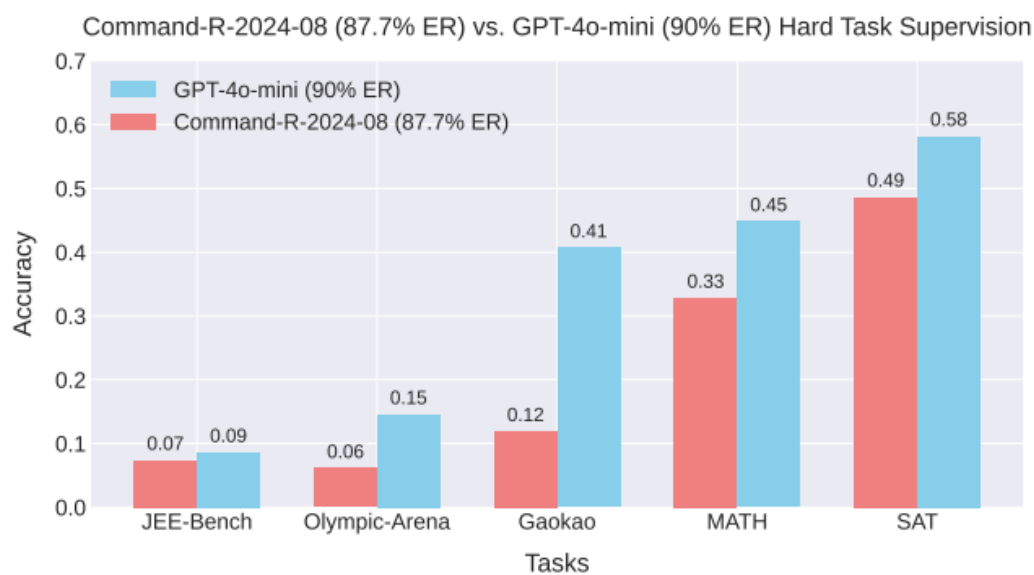
- Calculate the outcome error rate of “weaker teacher’s supervision” (e.g. 87.7%), and select the supervision data from previous teacher (GPT-4o-mini) with closest outcome error rate (e.g. 90%).

from previous teacher
(GPT-4o-mini)



from new “weaker”
teacher (like Command-R)

Almost the same outcome ER,
but different step-wise ER



5. Improve on sole hard supervision: Combination

- Since hard full task supervision is a better choice, can we make further improvement **without introducing new tasks or new information?**

5. Improve on sole hard supervision: Combination

- Since hard full task supervision is a better choice, can we make further improvement **without introducing new tasks or new information?**

Combination Strategies!

5. Improve on sole hard supervision: Combination

- Since hard full task supervision is a better choice, can we make further improvement **without introducing new tasks or new information**?

Combination Strategies!

- (1) Combine with decomposed subtask supervision, with multiple ER combinations,

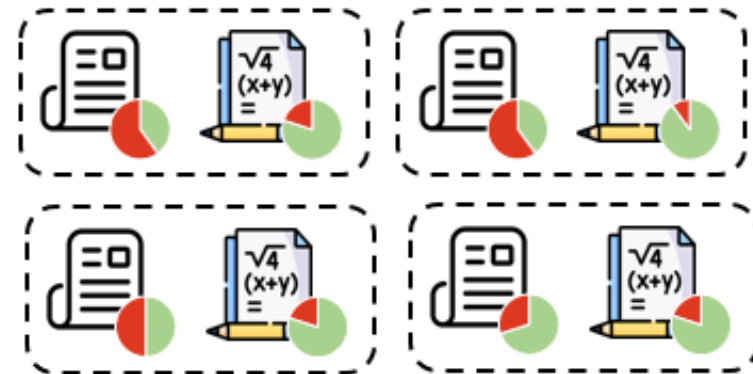


5. Improve on sole hard supervision: Combination

- Since hard full task supervision is a better choice, can we make further improvement **without introducing new tasks or new information**?

Combination Strategies!

(1) Combine with decomposed subtask supervision, with multiple ER combinations,



Tasks	Combination 1			Combination 2			Combination 3			
	(20%, 10%)	(20%,0%)	20%	(50%, 10%)	(50%,0%)	50%	(80%, 40%)	(80%, 10%)	(80%, 0%)	80%
MATH	43.2	48.0	47.4	48.6	45.8	47.8	43.8	46.8	47.2	43.2
Olympic-Arena	13.6	16.4	12.7	12.7	18.8	12.7	14.1	19.2	17.8	13.6
JEE-Bench	12.3	10.4	7.4	11.7	9.2	12.3	12.3	9.8	7.4	12.3
Gaokao-Mathcloze	40.7	40.7	39	45.8	43.2	44.1	42.4	37.3	43.2	40.7
SAT-Math	58.2	60.5	58.2	61.0	63.2	60.0	60.0	66.8	62.3	58.2
Average	33.6	35.2	32.9	36.0	36.0	35.4	34.5	36.0	35.6	33.6

5. Improve on sole hard supervision: Combination

(2) Combine with decomposed subtask supervision from different weaker teacher models

Outcome ER Combinations	Combination		
	GPT-4o-mini (80%, 10%)	Llama-3-70B-Instruct (81.7%, 10%)	Command-R-2024-03 (87.7%, 10%)
Step-Wise ER of Hard Task Supervision	11.7%	30.2%	62.6%
MATH	46.8	31.2	24.6
Olympic-Arena	19.2	12.2	5.2
JEE-Bench	8.7	10.4	5.5
Gaokao-Mathcloze	37.3	23.7	5.1
SAT-Math	61.0	54.5	47.7

5. Improve on sole hard supervision: Combination

(2) Combine with decomposed subtask supervision from different weaker teacher models

Outcome ER Combinations	Combination		
	GPT-4o-mini (80%, 10%)	Llama-3-70B-Instruct (81.7%, 10%)	Command-R-2024-03 (87.7%, 10%)
Step-Wise ER of Hard Task Supervision	11.7%	30.2%	62.6%
MATH	46.8	31.2	24.6
Olympic-Arena	19.2	12.2	5.2
JEE-Bench	8.7	10.4	5.5
Gaokao-Mathcloze	37.3	23.7	5.1
SAT-Math	61.0	54.5	47.7

(3) combine with rephrased hard supervision to increase task diversity;

Combine with
rephrased hard tasks



(4) combine with itself – double the training epoch

Combine with
itself - Double
training epochs



5. Improve on sole hard supervision: Combination


(2) Combine with decomposed subtask supervision from different weaker teacher models

Outcome ER Combinations	Combination		
	GPT-4o-mini (80%, 10%)	Llama-3-70B-Instruct (81.7%, 10%)	Command-R-2024-03 (87.7%, 10%)
Step-Wise ER of Hard Task Supervision	11.7%	30.2%	62.6%
MATH	46.8	31.2	24.6
Olympic-Arena	19.2	12.2	5.2
JEE-Bench	8.7	10.4	5.5
Gaokao-Mathcloze	37.3	23.7	5.1
SAT-Math	61.0	54.5	47.7

combination method
(2)(3)(4) are not better
than (1)


(3) combine with rephrased hard supervision to increase task diversity;

Combine with
rephrased hard tasks



(4) combine with itself – double the training epoch

Combine with
itself - Double
training epochs



Tasks	Combination		
	(50%, 10%)	50% (Doubled Epochs)	50% (Merge Rephrased.)
MATH	48.6	45.0	48.4
Olympic-Arena	12.7	12.7	12.7
JEE-Bench	11.7	9.2	10.4
Gaokao-Mathcloze	45.8	42.4	44.1
SAT-Math	61.0	55.5	57.3

- Summarize data-driven guidance for improving LLMs in future reasoning-heavy applications:

- Summarize data-driven guidance for improving LLMs in future reasoning-heavy applications:

(1) Prioritize hard task supervision for better performance on challenging reasoning tasks

- Summarize data-driven guidance for improving LLMs in future reasoning-heavy applications:

(1) Prioritize hard task supervision for better performance on challenging reasoning tasks

(2) Step-wise error rates of teacher models is more critical than outcome error rates for supervision quality

- Summarize data-driven guidance for improving LLMs in future reasoning-heavy applications:

(1) Prioritize hard task supervision for better performance on challenging reasoning tasks

(2) Step-wise error rates of teacher models is more critical than outcome error rates for supervision quality

(3) Consider sampling or annotating hard tasks and their associated subtasks then use the combined supervision for further enhancement

Thanks For
Watching