

《Machine Learning》Project·2024

Mining Misconceptions in Mathematics

1

Introduction

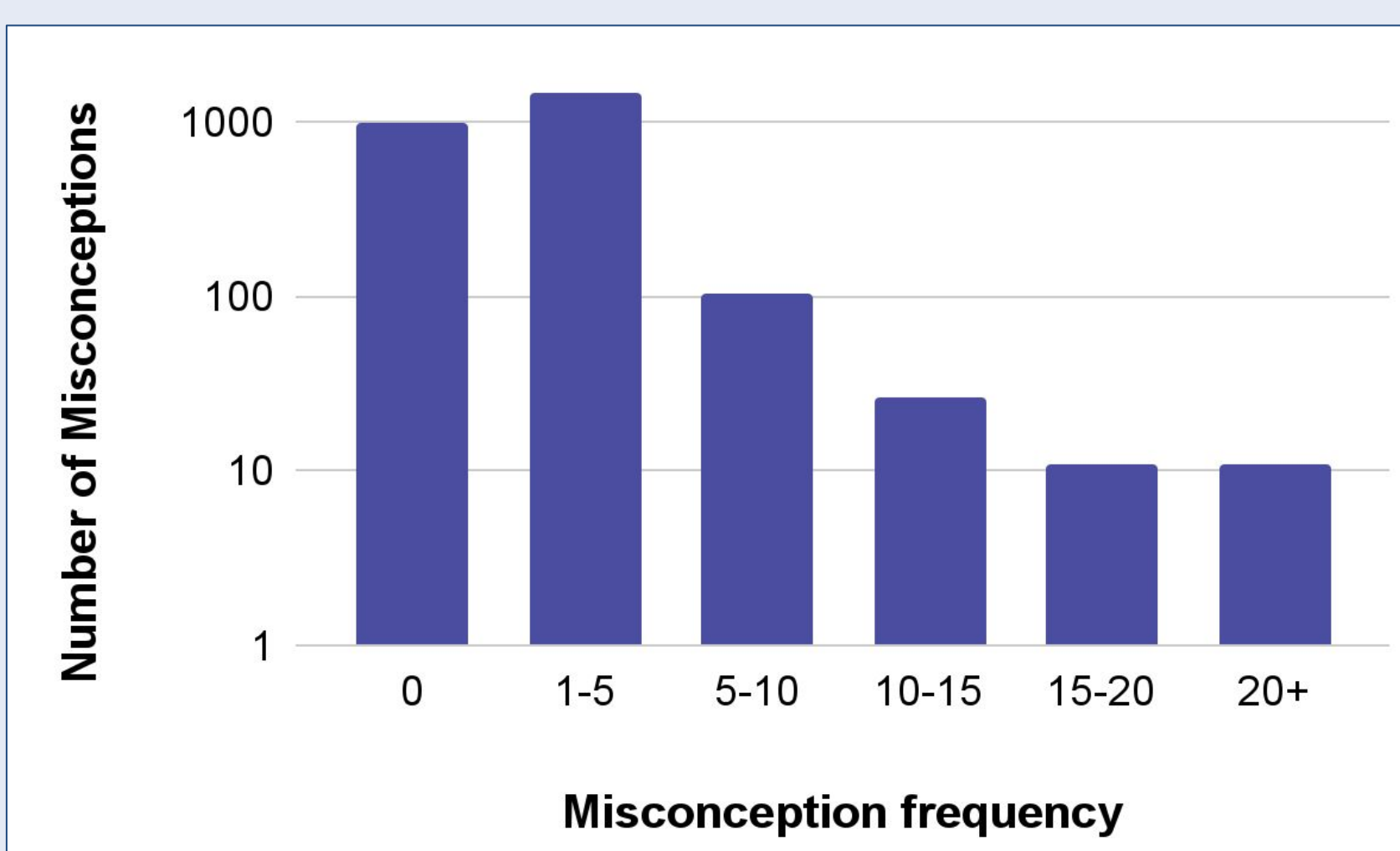
This project identifies the **mathematical misconceptions** that are associated to the incorrect answers in Multiple Choice Questions (MCQs) to improve the learning experience of the students. It is inspired by the *Kaggle EEDI* challenge.

Metrics

$$\text{MAP@25} = \frac{1}{U} \sum_{u=1}^U \sum_{k=1}^{\min(n,25)} \text{Precision}(k) \times \text{rel}(k)$$

U - Number of observations, n - Number of predictions
 $\text{rel}(k)$ - Relevance indicator (1 or 0)

Misconception Distribution



Kaggle EEDI Dataset

Question	Subject	Correct Answer	Answers	Misconception A, B, C, D
$x_{n+1} = 5 - \frac{1}{x_n}$ Using $x_1 = 2$ as the starting value, what is the solution to 2 decimal places?	Trial and Improvement and Iterative Methods	D	A. 4.78 B. 4.79128 C. 4.8 D. 4.79	A. 1775, Stops iteration too early when solving equations B. 1248, Rounds to the wrong degree of accuracy C. 1529, Rounds to the wrong degree of accuracy

Challenges

- Limited **dataset size**
- LLMs struggle with **misconceptions**
- Misconception similarity

2

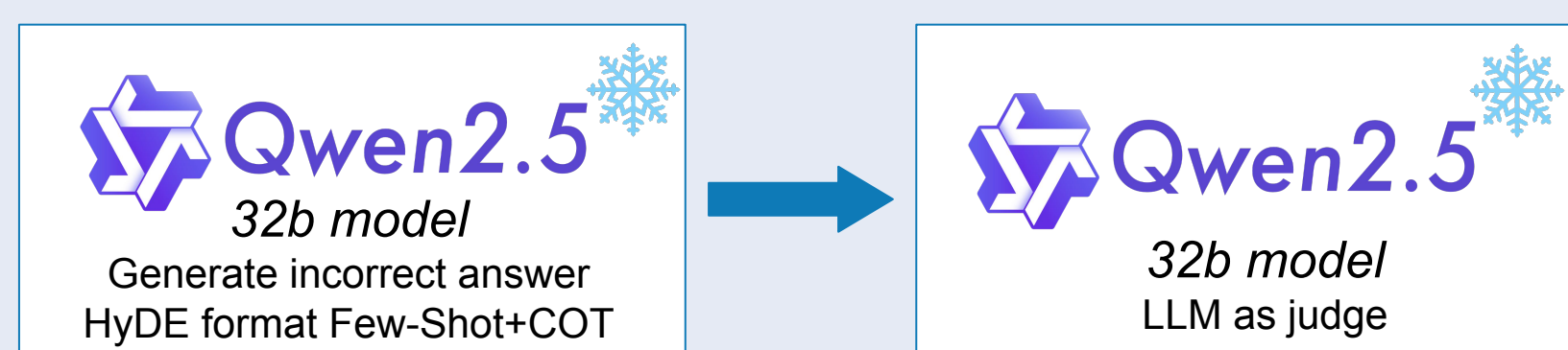
Methods

Synthetic Dataset

Subject	Question	Solution
...

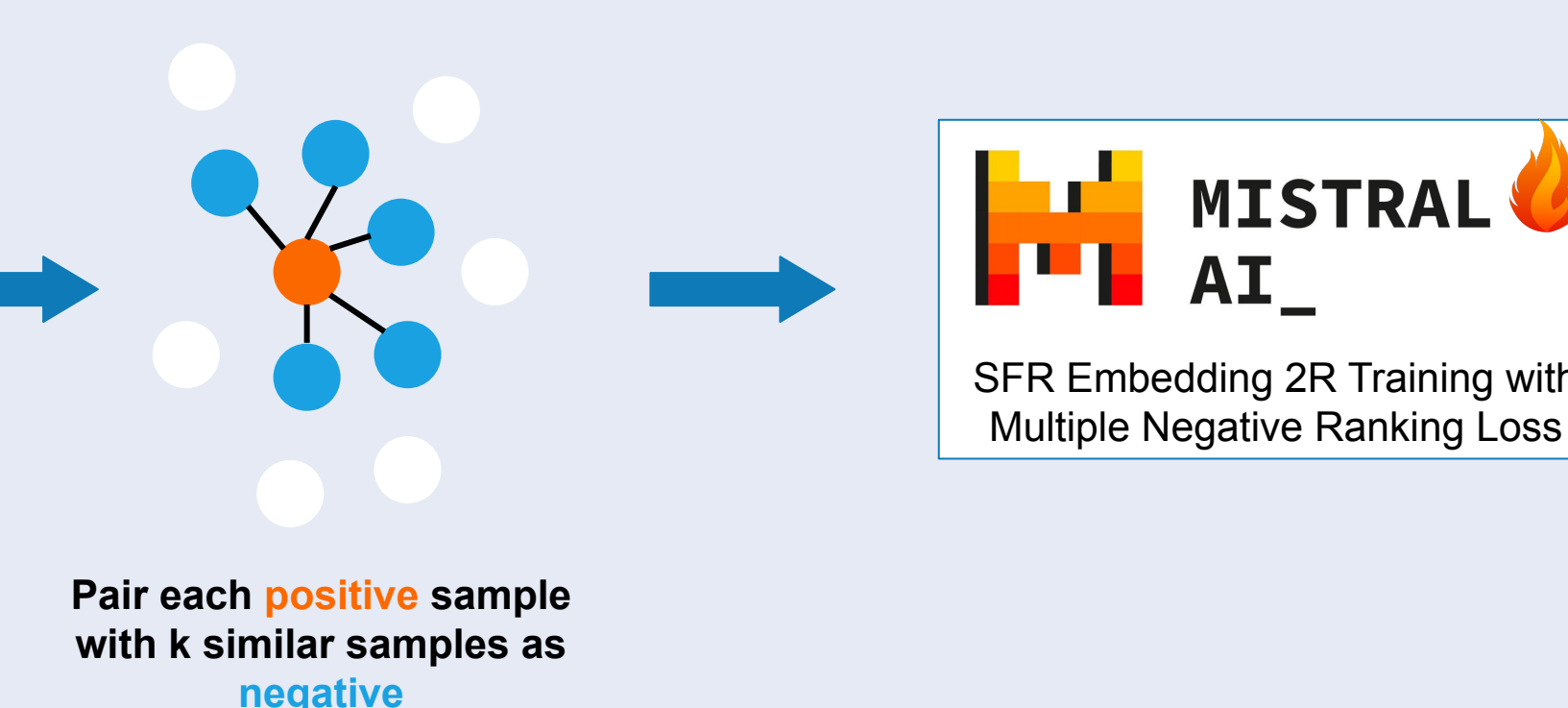
Subject Mapping	
Subject	EEDI Subject
...	...

Cluster of Misconceptions	
EEDI Subject	Misconceptions
...	...



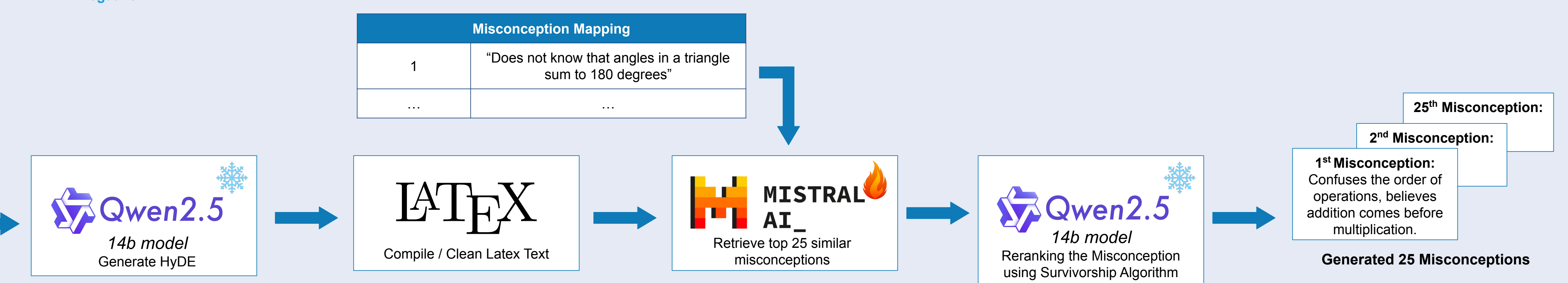
LoRA Fine-tuning the Embedding Model

Question	Incorrect Answer	HyDE Format	Misconception Label
3 x 2 + 4 - 5 Where do the brackets need to go to make the answer equal to 13?	Does not need brackets	Mistakenly believes that addition comes before multiplication.	Confuses the order of operations, believes addition comes before multiplication.
...

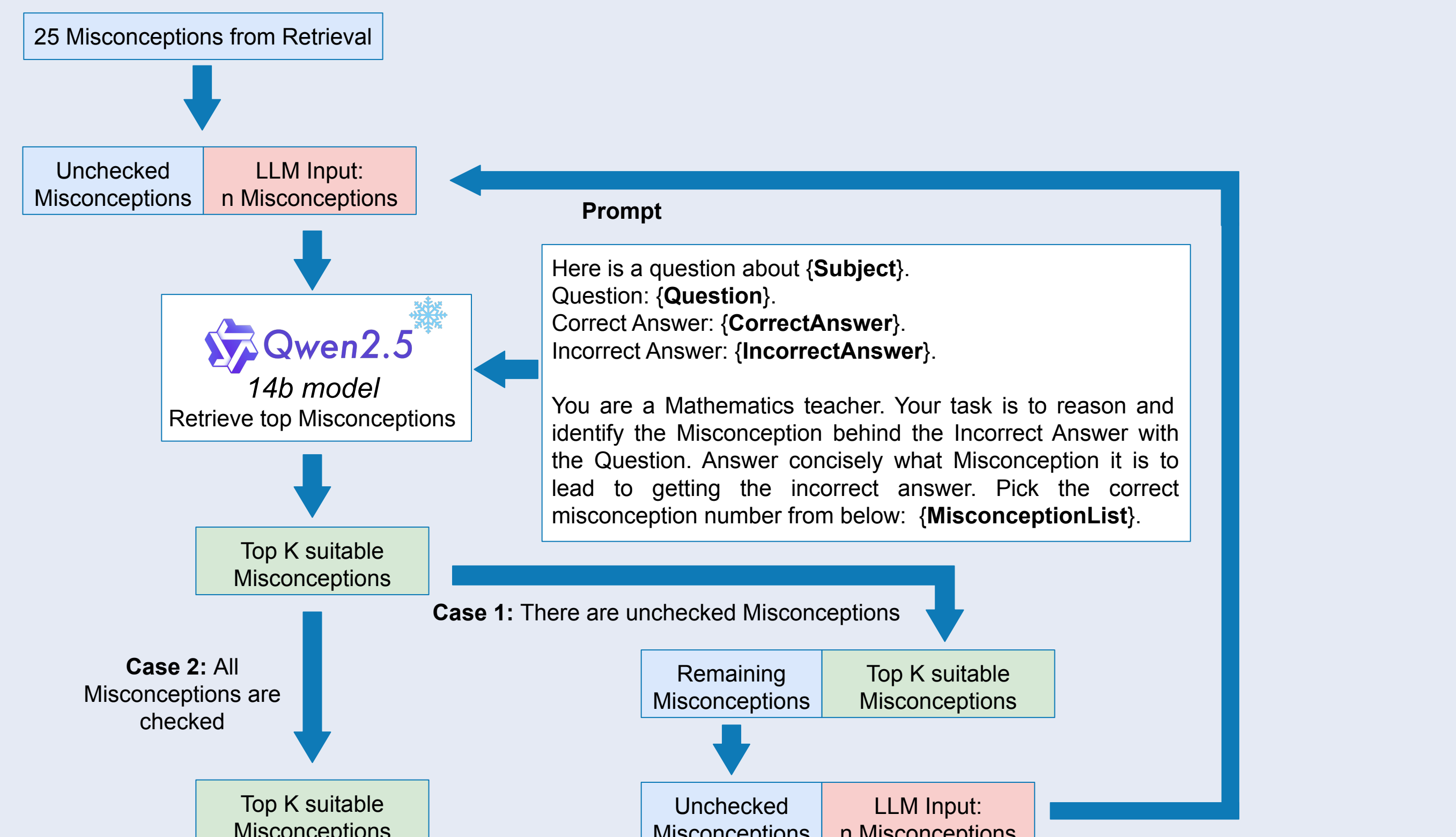


Inference Pipeline

Question	Incorrect Answer	Correct Answer	Subject Name
3 x 2 + 4 - 5 Where do the brackets need to go to make the answer equal to 13?	Does not need brackets	3 x (2 + 4) - 5	BIDMAS
...



Logits processor based Reranking



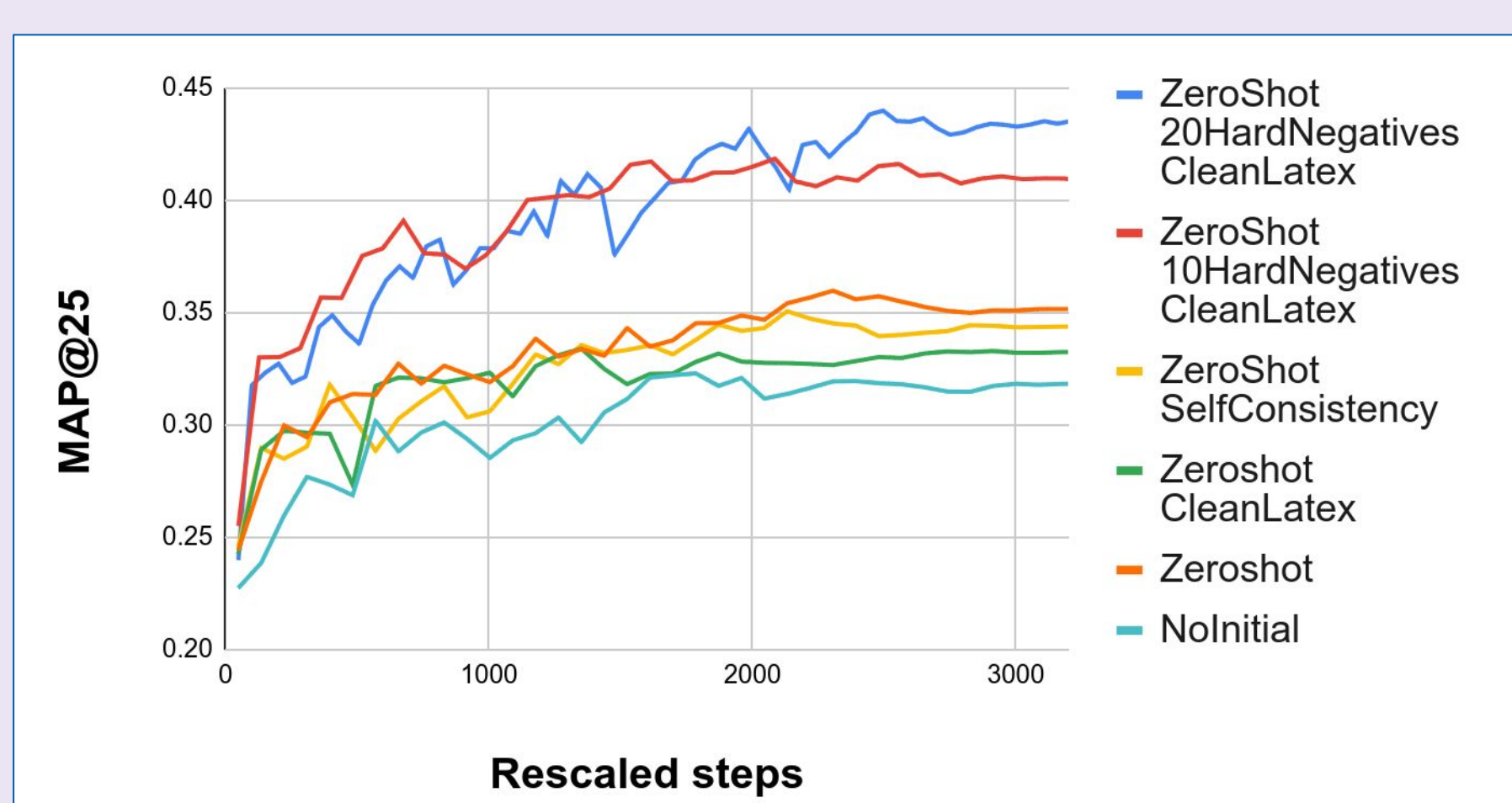
3

Results

- Training and inference were performed on a single *NVIDIA H100 GPU*
- Self-consistency** marginally increased fine-tuned performance
- Hard Negative Mining** significantly increased performance.

Model Configuration	MAP@25
Baseline (No Fine-tuning, no reranking)	0.18
Best Fine-tuning	0.43
Best Fine-tuning, rerank top-1 Misconception	0.45
Best Fine-tuning, rerank top-3 Misconception	0.46

Fine-tuning MAP@25 evaluation



4

Conclusion

- Mining misconception for MCQs is feasible
- Multiple Negative Ranking Loss** is critical for fine-tuning embedding search model
- First** (Faster Improved Listwise Reranking with Single Token Decoding) improves accuracy
- Self-Consistency** yields negligible gains

Future work

- Additional experiments with synthetic data
- Evaluate performance with different Reranking techniques
- Evaluate **larger models** for HyDE generation and Reranking

Members :

Bryan Constanstine Sadihin, Matteo Jiahao Chen, Hector Rodriguez Rodriguez
Master in Advanced Computing, Department of Computer Science, Tsinghua University