

TURINGQ DATASET

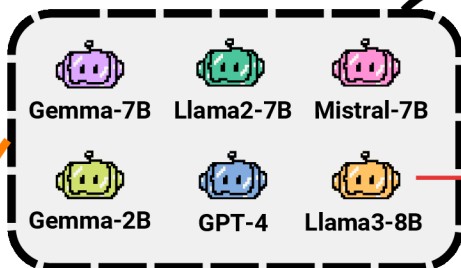
EVALUATION

Complexity
Theory
Turing Machine
Regular
Languages



Context-Free
Languages
Countability
Computability
Theory

MODELS



FINETUNING

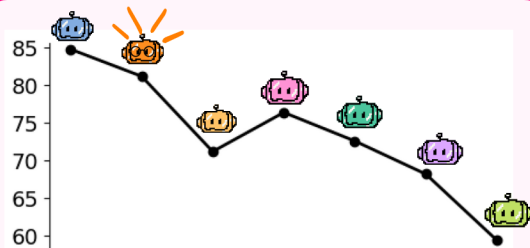
SFT

PEFT



Llama3-8B
TuringQ

Light Adapter
improved
Math skills!!



LLM AS A JUDGE

Similar Average Performance!



Llama3-8b

≈



Human Expert

You are an **automated grading system** for theory of computation and complexity answers. Assign a score of 1 to 4 based on **correctness and alignment** with the provided solution, following the specified **rubrics**...

Prefer Longer answers!
(May) Inject their own reasoning!

TURINGQ INSTANCES

Is the language $L = \{x \in \{a, b\}^* : x \text{ contains twice as many a's as b's}\}$ regular or not?

True/False: For all languages $L1$ and $L3$: $(L1^* \cup L3^*) = (L1^* \cup L3^*)^*$

Prove that the language $L = \{ \langle M1, M2 \rangle : M1, M2 \text{ are TMs and } L(M1) = L(M2) \}$, is undecidable.

The language L is **regular**.
 L can be described by the regular expression $aabb^*$.

True: The Kleene star operation distributes over union.

...
Since we have reached a **contradiction**, our initial assumption that L is decidable must be **false**. Therefore, L is undecidable.

Gemma-7b

GPT-4

Llama3-8B
TuringQ