ACL 2025 VIENNA

RATIONALYST: Mining Implicit Rationales for Process Supervision of Reasoning

Dongwei Jiang, Guoxuan Wang, Yining Lu, Andrew Wang,
Jingyu Zhang, Chuyu Liu,
Benjamin Van Durme, Daniel Khashabi

Center for Language and Speech Processing

JOHNS HOPKINS UNIVERSITY

Introduction

What is reasoning and why is it important for large language model?

Reasoning is the ability to logically connect ideas, draw conclusions, and solve problems.

Importance of reasoning in Large Language Models (LLMs):

- **Enhances accuracy:** Helps LLMs generate correct and reliable responses.
- **Improves coherence:** Allows models to maintain logical flow in generated text.
- **Supports generalization:** Enables LLMs to tackle new problems not seen during training.
- **Increases interpretability:** Makes the model's outputs understandable and trustworthy.
- **Facilitates complex tasks:** Allows step-by-step problem solving in domains like math and coding.

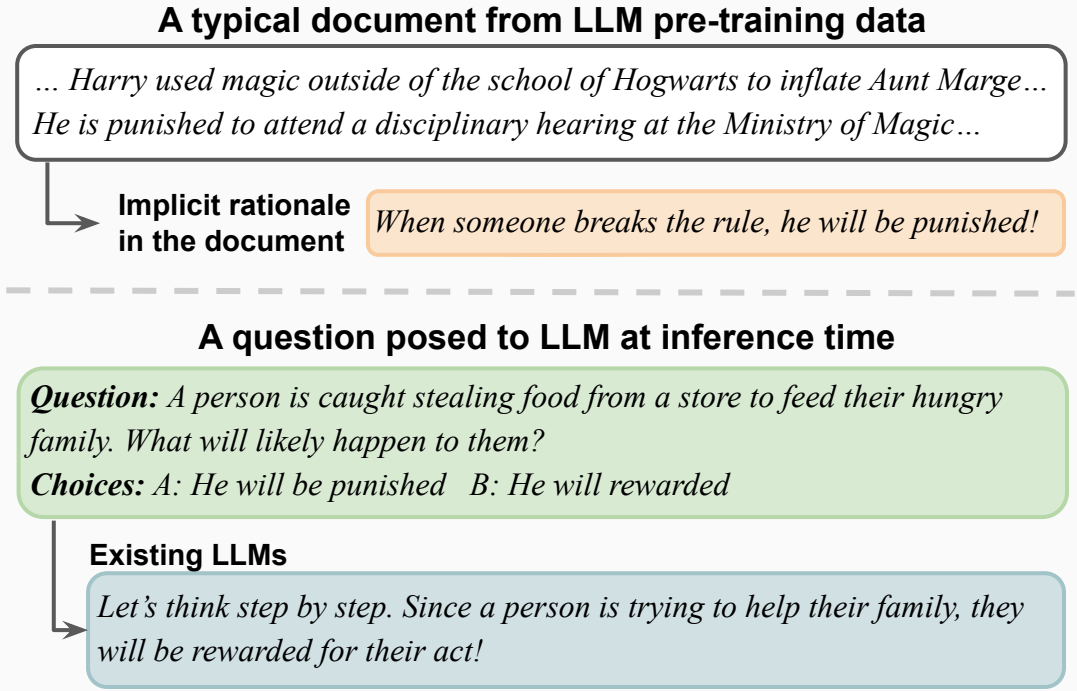
https://chat.openai.com/chat

Motivation

Problem: LLM reasoning steps are often incomplete

Root cause: They mimic logical leaps from their training data

Result: Existing LLMs will have difficulty surfacing these implicit statements during the reasoning process, which can lead to flawed conclusions.



Motivation

Our solution: RATIONALYST trained on implicit rationales from pre-training data

How it works: RATIONALYST works by making these implicit rationales explicit and using them to guide the reasoning process at inference time.

A typical document from LLM pre-training data

... Harry used magic outside of the school of Hogwarts to inflate Aunt Marge... He is punished to attend a disciplinary hearing at the Ministry of Magic...

Implicit rationale in the document: *When someone breaks the rule, he will be punished!*

A question posed to LLM at inference time

Question: *A person is caught stealing food from a store to feed their hungry family. What will likely happen to them?*
Choices: *A: He will be punished B: He will rewarded*

Existing LLMs

Let's think step by step. Since a person is trying to help their family, they will be rewarded for their act!

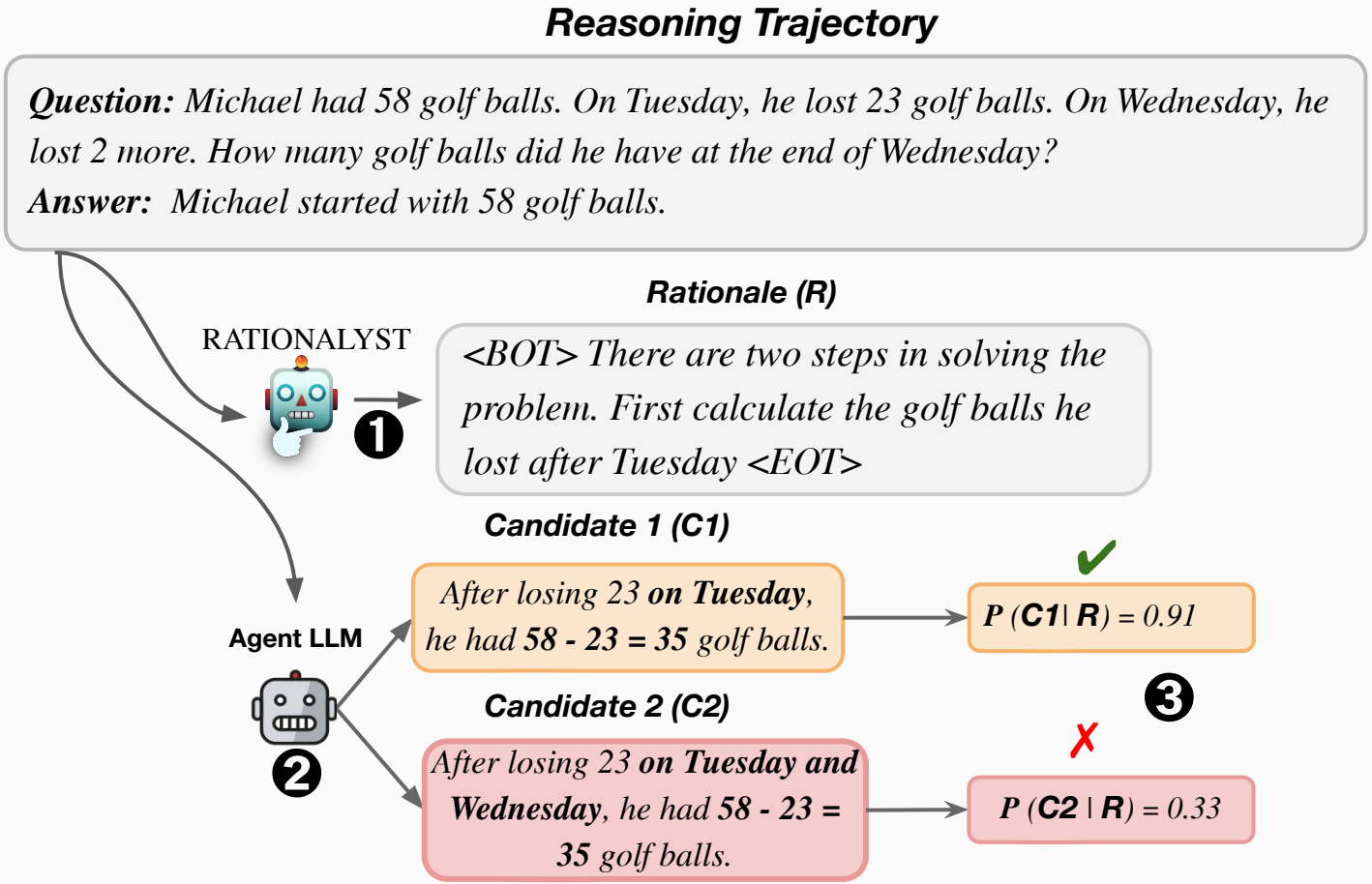
Existing LLMs + rationale supervision via RATIONALYST 🤖

Let's think step by step. Although this stealing has good intentions, ***stealing from a store breaks the rule of society, so it should be punished!***

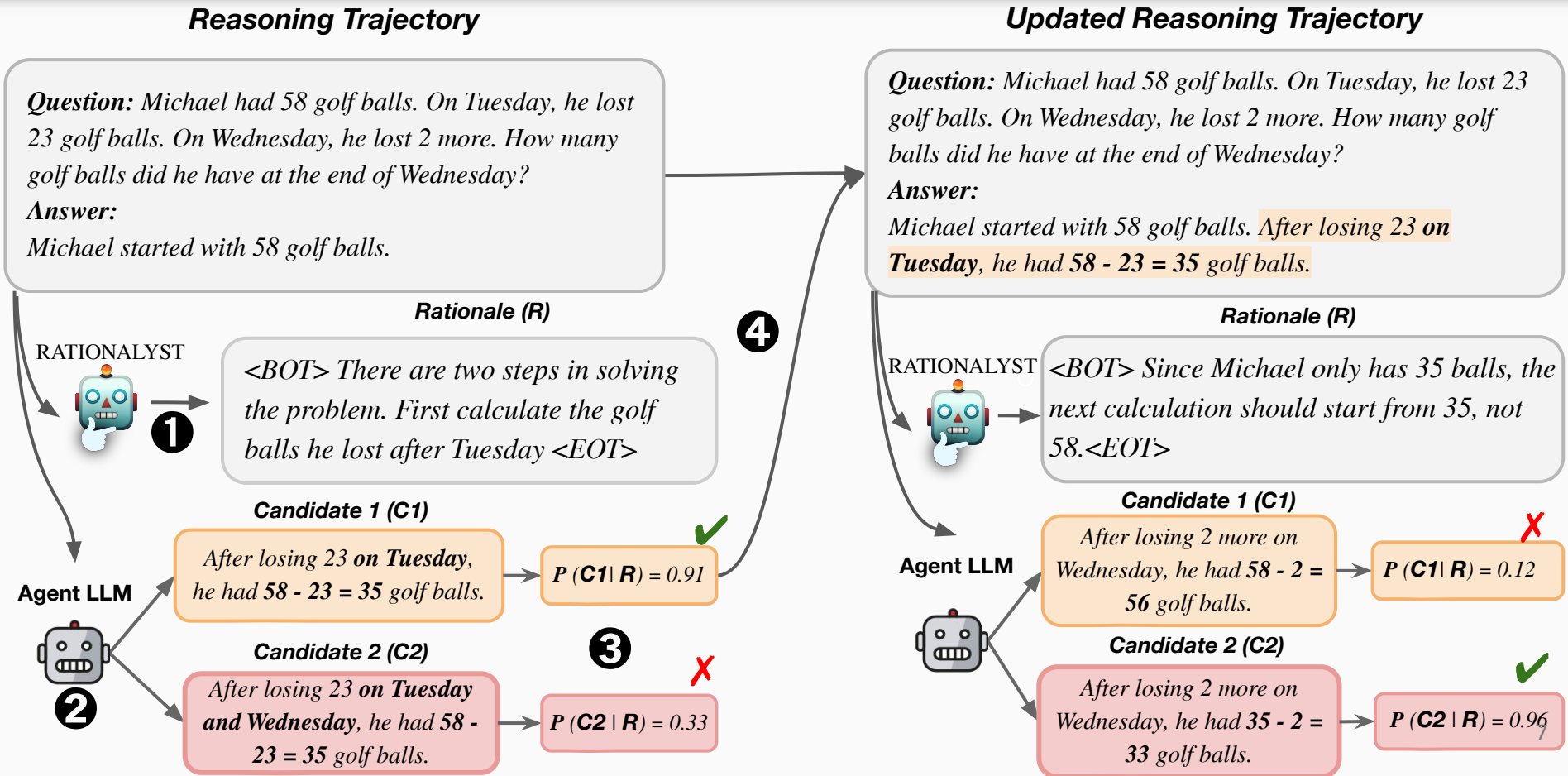
RoadMap

- How is RATIONALYST actually used during inference
- How to mine a dataset of implicit rationales and train RATIONALYST?
- Evaluations

Inference-time supervision



Inference-time supervision cont.



RoadMap

- How is RATIONALYST actually used during inference
- How to mine a dataset of implicit rationales and train RATIONALYST?
- Evaluations

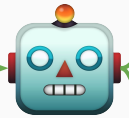
Rationale extraction

Your task is to identify implicit reasoning steps in text - the unstated logical connections that bridge ideas and help predict what comes next. Look for logical leaps where important reasoning steps are assumed but not written out. Add these implicit rationales by writing "<BOT>rationale< EOT >".

Unlabelled Data from different sources

… Harry used magic outside of the school of Hogwarts to inflate Aunt Marge… He is punished to attend a disciplinary hearing at the Ministry of Magic…

LLM (LLaMa 3 8B Instruct)



Implicit Rationales Extracted by LLMs from Unlabeled Data

… Harry used magic outside of the school of Hogwarts to inflate Aunt Marge… ***<BOT>When someone breaks the rule, he will be punished<EOT>*** *…*

… Harry used magic outside of the school of Hogwarts to inflate Aunt Marge… ***<BOT>Hogwarts magic is incredibly versatile, capable of a wide range of effects<EOT>*** *…*

Filtering via Future Text Perplexity

Adding implicit rationale significantly reduces future text perplexity!

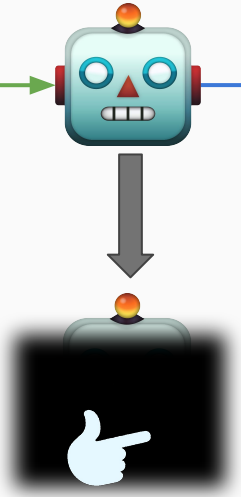
Adding implicit rationale has little effect on future text perplexity!

Model training

Input

… Harry used magic outside of the school of Hogwarts to inflate Aunt Marge

LLaMa 3 8B Instruct



RATIONALYST

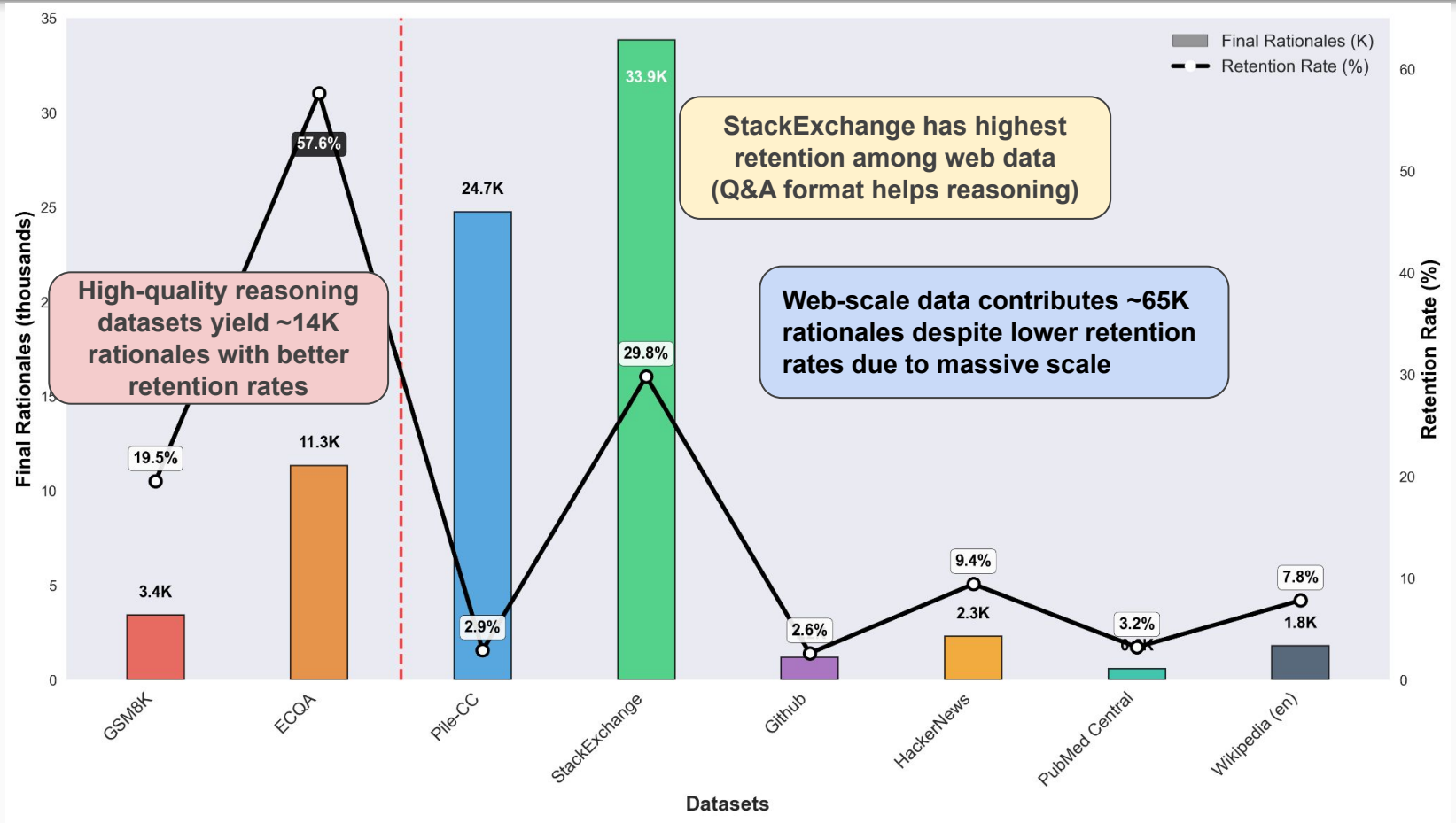
Output

When someone breaks the rule, he will be punished

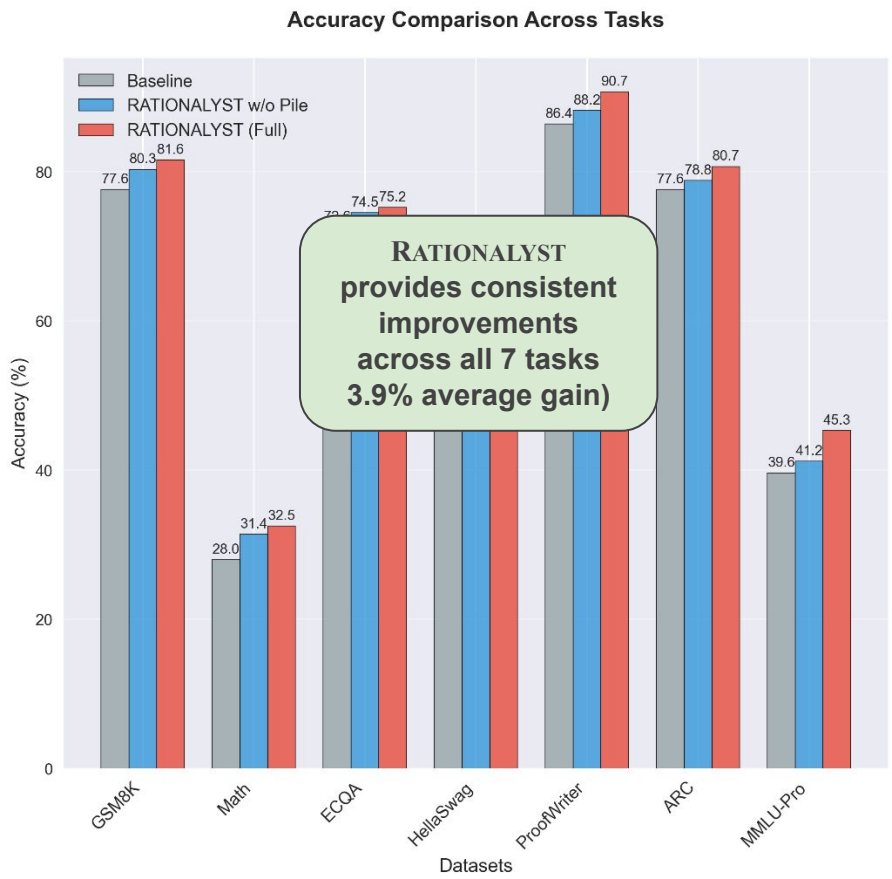
RoadMap

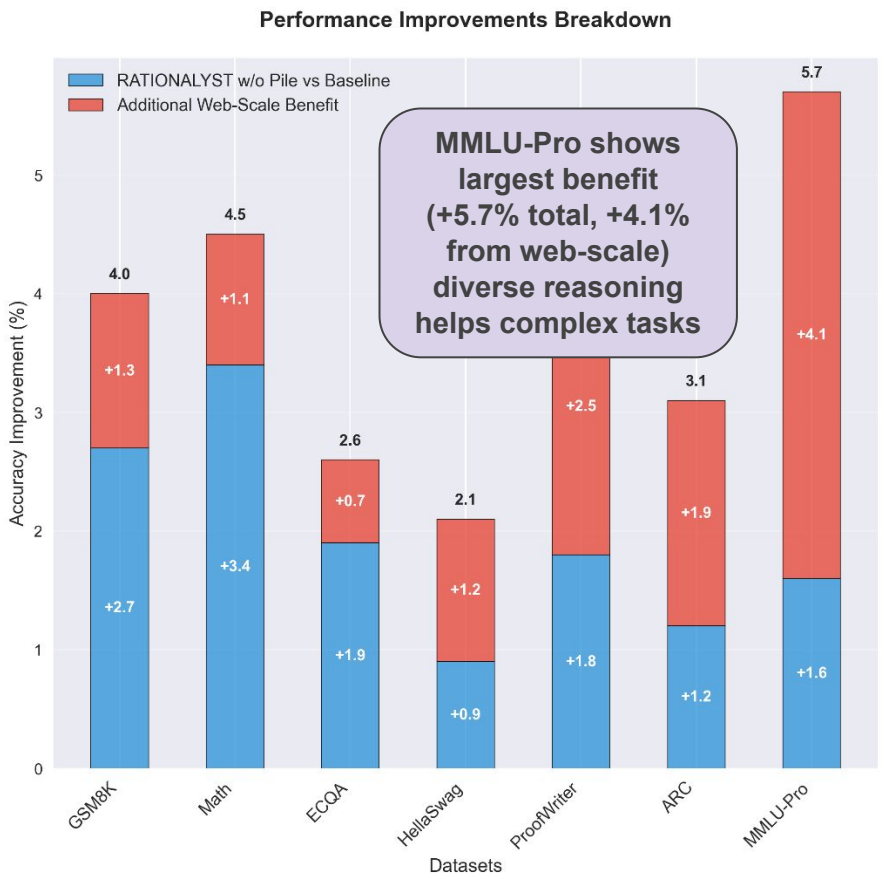
- How is RATIONALYST actually used during inference
- How to mine a dataset of implicit rationales and train RATIONALYST?
- Evaluations

Results of our large-scale rationale extraction

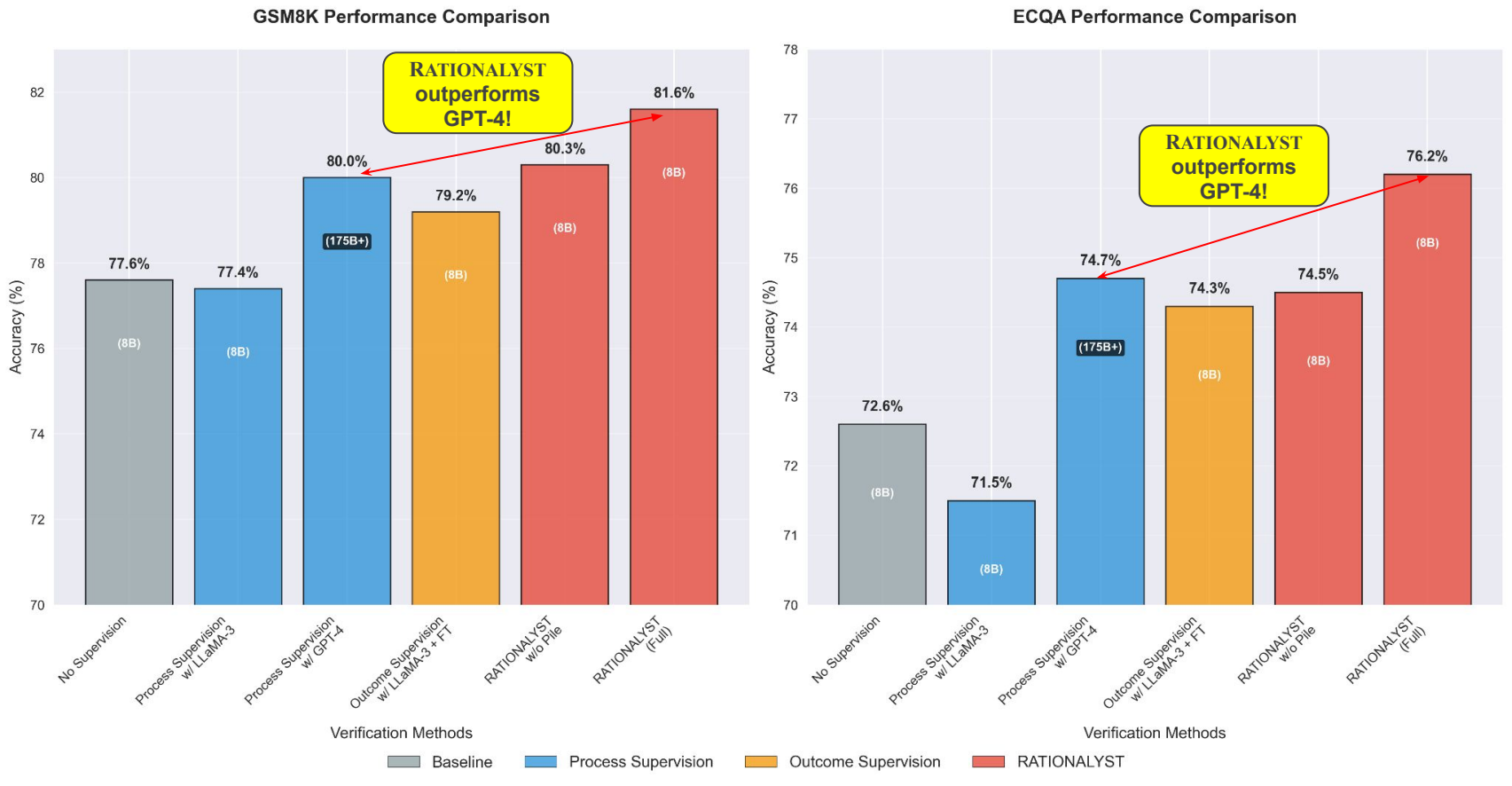


The benefits of adding RATIONALYST





How does Rationalyst compare against other verifiers



Summary

- **Motivation:** The reasoning steps generated by LLMs might be incomplete because they mimic logical leaps common in everyday communication.
- **What we did:** We extract 79K implicit reasoning steps from unlabeled text, train a specialized rationale generation model called RATIONALYST, and uses it to supervise reasoning at inference time.
- **What we found:** RATIONALYST show +3.9% average improvement across 7 reasoning tasks, outperforming even GPT-4 verification.
- **Future Work:**
  - **Scale up**: Use stronger models (GPT-4, LLaMA-70B) and larger datasets
  - **Better integration**: Combining RATIONALYST with test-time compute and preference fine-tuning
- **Parallel Works:**

arxiv 2025

Reasoning to Learn from Latent Thoughts

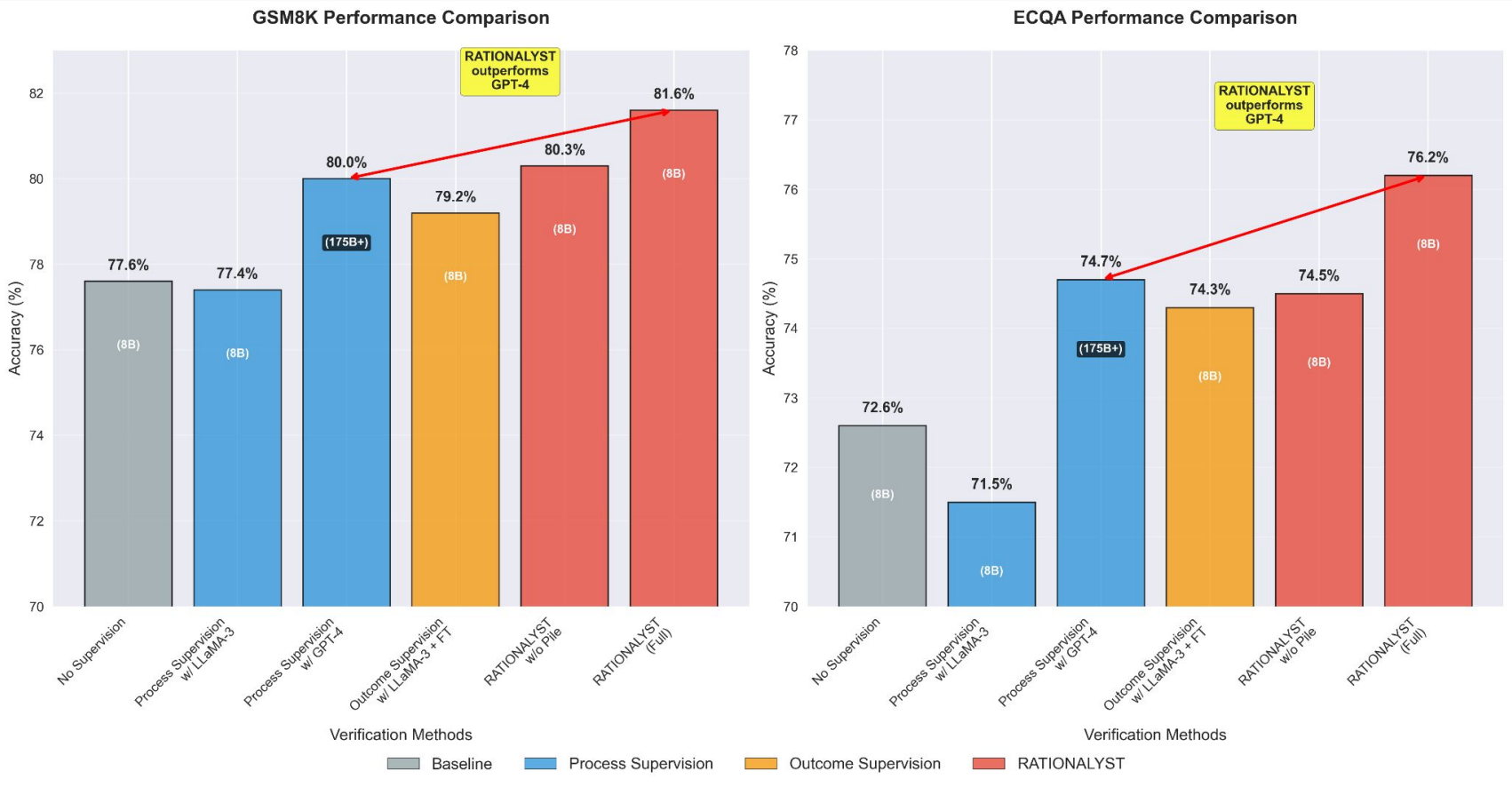
Yangjun Ruan[1,2,3], Neil Band[1], Chris J. Maddison[2,3†], Tatsunori Hashimoto[1†]
[1]Stanford University [2]University of Toronto [3]Vector Institute
{yjruan,cmaddis}@cs.toronto.edu, {nband,thashim}@stanford.edu

arxiv 2025

Reinforcement Pre-Training

Qingxiu Dong[*†‡] Li Dong[*†]
Yao Tang[†] Tianzhu Ye[†§] Yutao Sun[†§] Zhifang Sui[‡] Furu Wei[†◇]
[†] Microsoft Research
[‡] Peking University
[§] Tsinghua University
https://aka.ms/GeneralAI

How does Rationalyst compare against other verifiers



Future Work

Scaling Up

- Use stronger models (LLaMA-3-70B, GPT-4) for better rationale extraction
- Train on larger datasets like OpenWebMath
- Scale to more reasoning tasks and benchmarks

Technical Improvements

- Integrate with test-time compute techniques (beam search, look-ahead)
- Add preference tuning (DPO) to distinguish valid/invalid rationales
- Combine with existing methods (self-consistency, STEP-BACK prompting)

Research Directions

- Study optimal rationale mixing strategies across datasets
- Understand what makes rationales effective for reasoning
- Investigate rationale transfer across different domains

Roadmap

- What to do at inference-time
- How to mine a dataset of implicit rationales that are turned explicit?
- Training on this data
- Evaluations

Using Rationalyst at inference time

Motivation

Our solution: RATIONALYST trained on implicit rationales from pre-training data
How it works: Provides supervision for reasoning processes
Key benefit: Makes implicit rationales explicit
Outcome: More robust reasoning and conclusions

This is where our approach comes in. Our solution is RATIONALYST, which is trained on a vast collection of implicit rationales extracted from pre-training data to provide supervision for reasoning. RATIONALYST works by making these implicit rationales explicit and using them to guide the reasoning process at inference time.

RATIONALYST provides an additional supervision mechanism to guide LLMs' reasoning processes, resulting in more robust conclusions.

A typical document from LLM pre-training data

... Harry used magic outside of the school of Hogwarts to inflate Aunt Marge... He is punished to attend a disciplinary hearing at the Ministry of Magic...

Implicit rationale in the document: *When someone breaks the rule, he will be punished!*

A question posed to LLM at inference time

Question: *A person is caught stealing food from a store to feed their hungry family. What will likely happen to them?*
Choices: *A: He will be punished B: He will rewarded*

Existing LLMs

Let's think step by step. Since a person is trying to help their family, they will be rewarded for their act!

Existing LLMs + rationale supervision via RATIONALYST 🤖

Let's think step by step. Although this stealing has good intentions, ***stealing from a store breaks the rule of society, so it should be punished!***

Motivation

The reasoning steps generated by LLMs might be incomplete!

They mimic logical leaps common in everyday communication that's found in their pre-training data

Underlying rationales are frequently left implicit (unstated).

As a result, existing LLMs trained

A typical document from LLM pre-training data

... Harry used magic outside of the school of Hogwarts to inflate Aunt Marge... He is punished to attend a disciplinary hearing at the Ministry of Magic...

Implicit rationale in the document

When someone breaks the rule, he will be punished!

A question posed to LLM at inference time

Question: *A person is caught stealing food from a store to feed their hungry family. What will likely happen to them?*
Choices: *A: He will be punished B: He will rewarded*

Existing LLMs

Let's think step by step. Since a person is trying to help their family, they will be rewarded for their act!

Existing LLMs + rationale supervision via RATIONALYST 🤖

Let's think step by step. Although this stealing has good intentions, ***stealing from a store breaks the rule of society, so it should be punished!***