


RAIF: Incentivizing Reasoning for Advanced Instruction-Following of Large Language Models

🤔 Incentivizing LLMs to Think Before They Follow


🔍 Problem: LLMs Struggle with Complex Instructions

- **Multi-constraint instructions** (And, Chain, Selection, Nested) are hard to follow
- **Vanilla CoT** often leads to **shallow reasoning** or **worse performance**
- **No scalable method** to train LLMs to reason for instruction-following

🎯 **Train LLMs to reason deeply for complex instruction-following, not just paraphrase.**



Complex Instruction with Atomic Constraints and their Compositions

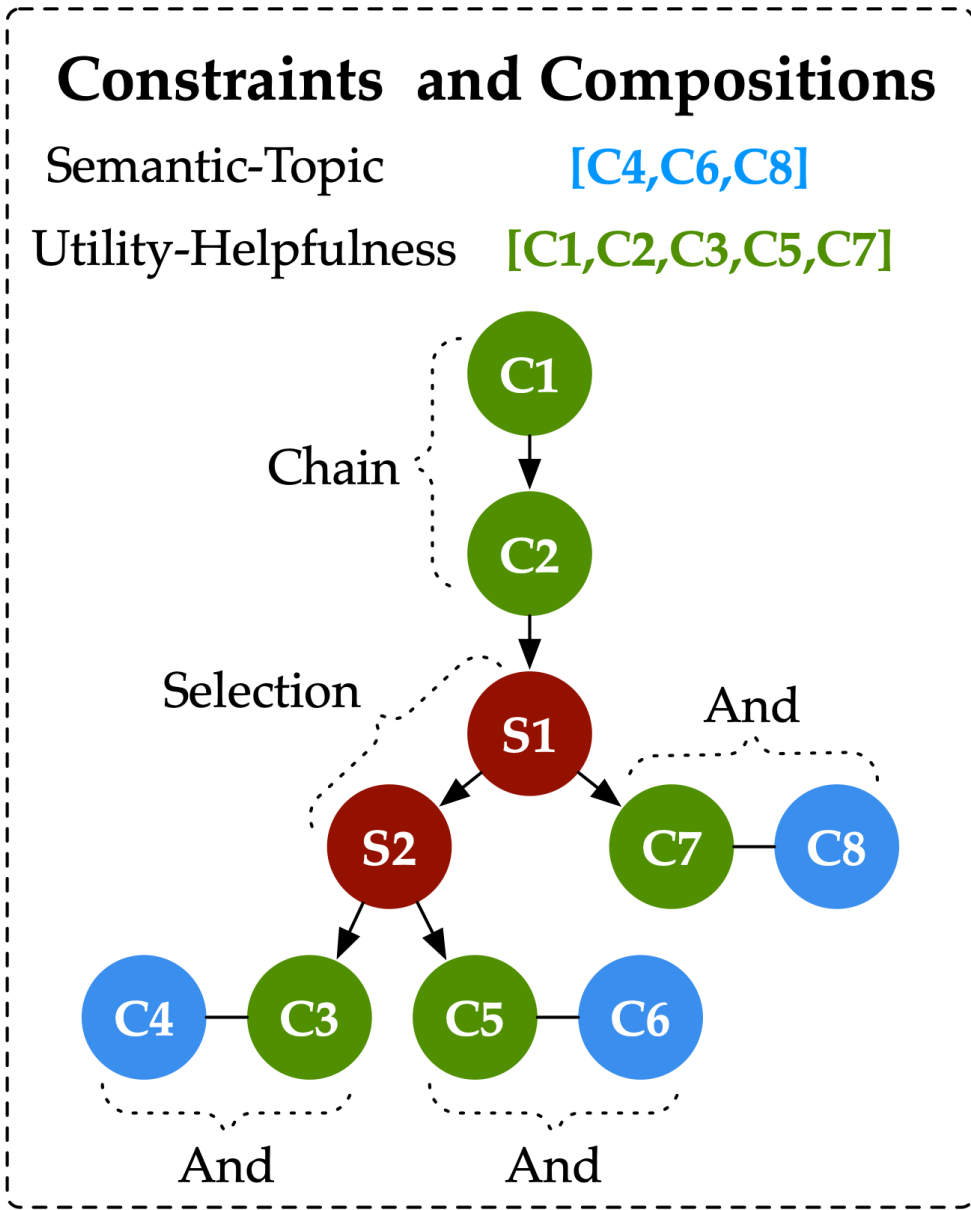



Data science is a field that combines various disciplines to extract insights and knowledge from data. It involves statistics, machine learning, data visualization, and domain expertise. In recent years, the importance of data science has grown exponentially, driven by the increasing availability of data and the need for data-driven decision-making. Data scientists play a crucial role in transforming raw data into actionable insights, which can significantly impact businesses and organizations.

- Determine if the above content is less than 100 words. [C1]
- Check if the above content mentions "machine learning". [C2]


- Based on the previous determinations:

- If the content is less than 100 words and mentions "machine learning", expand the content to a detailed article on the role of machine learning in data science. [C4]
- If the content is 100 words or more and mentions "machine learning", provide a brief summary of the key points in the content. [C6]
- If the content does not mention "machine learning," write a detailed introduction to data science, focusing on its applications and importance. [C8]





Instructed (Fast-Thinking) LLMs



Reasoning (Slow-Thinking) LLMs

CoT Prompting

Qwen2.5-1.5B-Instruct

17 Tokens

<think> The content is exactly 100 words and does not mention "machine learning." </think>.....

✗

CoT Prompting

Qwen2.5-7B-Instruct

156 Tokens

<think> 1. The content is 111 words, so it is more than 100 words. 2. The content does mention "machine learning." 3. Since the content is more than 100 words and mentions "machine learning," we need to provide a brief summary of the key points in the content. </think>.....

✗

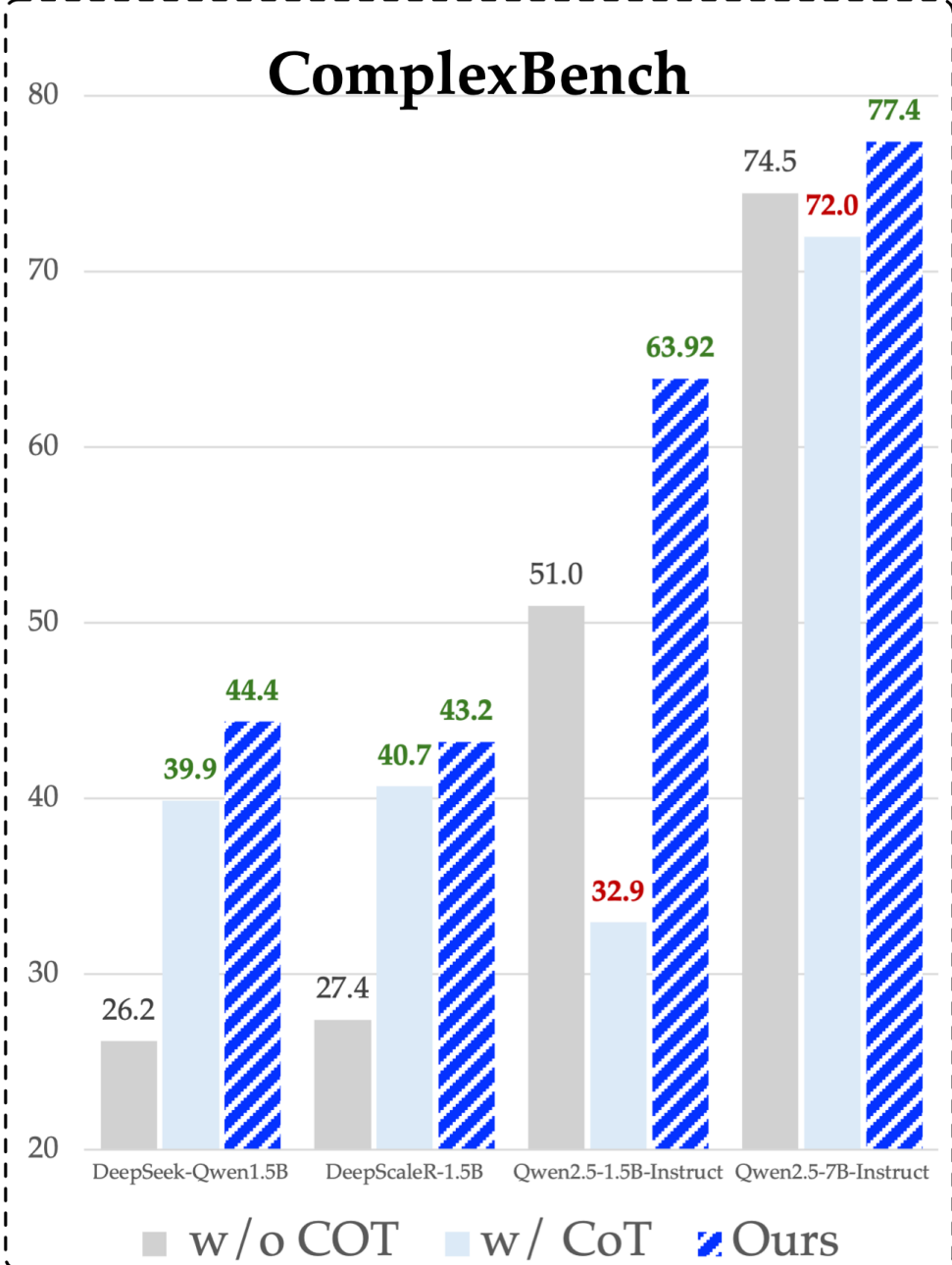
Direct I/O (CoT by default)

DeepSeek R1

451 Tokens

<think>Okay, let's tackle this query step by step.....First, I need to check if the content is under 100 words. Let me count. The original text has five sentences. Let me count each word. "Data science is a field that combines various disciplines to extract insights and knowledge from data." That's 18 words.....Total so far: 18+11+23+20=72. Wait, the original text actually has four sentences, not five. So total is 72 words. So yes, less than 100 words. Next, check if "machine learning" is mentioned. Scanning the text, the second sentence says "machine learning," so that's a yes. Now, based on the determinations: since it's less than 100 words and mentions ML, I need to expand the content into a detailed article focusing on ML's role in data science. So, the expansion should elaborate on how ML is integral to data science. Maybe start by defining data science and then dive into ML's part.....</think>.....

✓



2 Rule-Centric Reward Modeling with GRPO

- **Accuracy Reward:** Boolean checks on each atomic constraint

3 Experience Replay Buffer with Superior CoT Enforcement

- Filters out shallow reasoning with **sample-wise contrast**
- Guarantees **informative** group-based advantage computation

4 Policy Distribution Drift Control with Behavior Cloning

- Clones **expert reasoning patterns** to prevent **distribution drift**
- Avoids **reward hacking** and **semantic degradation**

📈 Results: RAIF Boosts Instruction-Following

Model	Base	RAIF	Gain
Qwen2.5-1.5B	50.61	62.35	+11.74%
DeepSeek-Qwen1.5B	39.03	49.58	+10.54%
Qwen2.5-7B	72.58	73.44	+0.85%

- **1.5B model** performs like 8B!
- **Outperforms SFT, CoT, SDC, and SOTA baselines** on seven benchmarks: IFEval, CELLO, CFBench, ComplexBench, FB-Bench, FollowBench, InfoBench.

🚩 Key Insight: Deep Reasoning > Shallow CoT

- **Vanilla CoT** often hurts performance on complex instructions.
- **RAIF trains real reasoning**, not just paraphrasing.
- **Small models benefit most** from test-time scaling via reasoning.

📊 Ablation: What Makes RAIF Work?

Component	Gain
+ Math data	+3.84%
+ Superior CoT filter	+2.59%
+ Behavior cloning	+3.76%
Full RAIF	+5.16%

🎓 Generalization: RAIF Works Beyond Training

- **OOD constraints:** up to +2.53% on IFBench (unseen rules) for 1.5B and +6.27% for 7B
- **Math tasks:** up to +16.2 on for 1.5B, +2.5% for 7B
- **Multi-task:** up to +0.74% for 1.5B, +1.81% for 7B

🕒 Future Directions

- **Larger models:** 32B, 70B, MoE
- **Better reward models:** scalar vs boolean
- **Beyond instructions:** reasoning for planning, coding, dialogue



WeChat





X





Paper





Code



References [1] Wen, B., Ke, P., Gu, X., Wu, L., Huang, H., Zhou, J., ... & Huang, M. (2024). Benchmarking complex instruction-following with multiple constraints composition. *Advances in Neural Information Processing Systems*, 37, 137610-137645.

Fig.1 Complex instructions with various atomic constraints and compositions pose great challenges to instruction-following capabilities of LLMs (The above example and its structure are from the ComplexBench [1]).