StructChem:

Structured Chemistry Reasoning with Large Language Models

Homepage - https://ozyyshr.github.io/StructChem/

Presenter: Siru Ouyang

Siru Ouyang

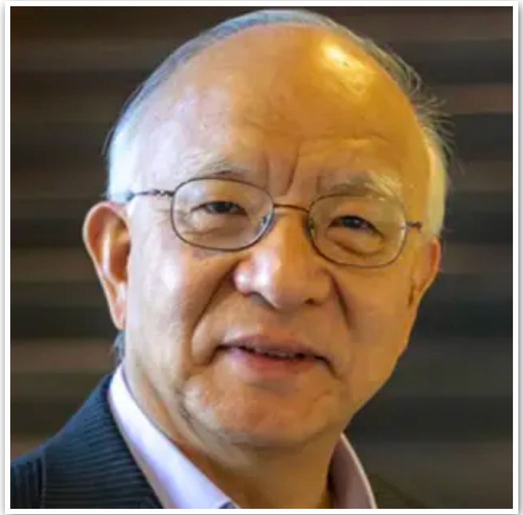
Zhuosheng Zhang

Bing Yan

Xuan Liu

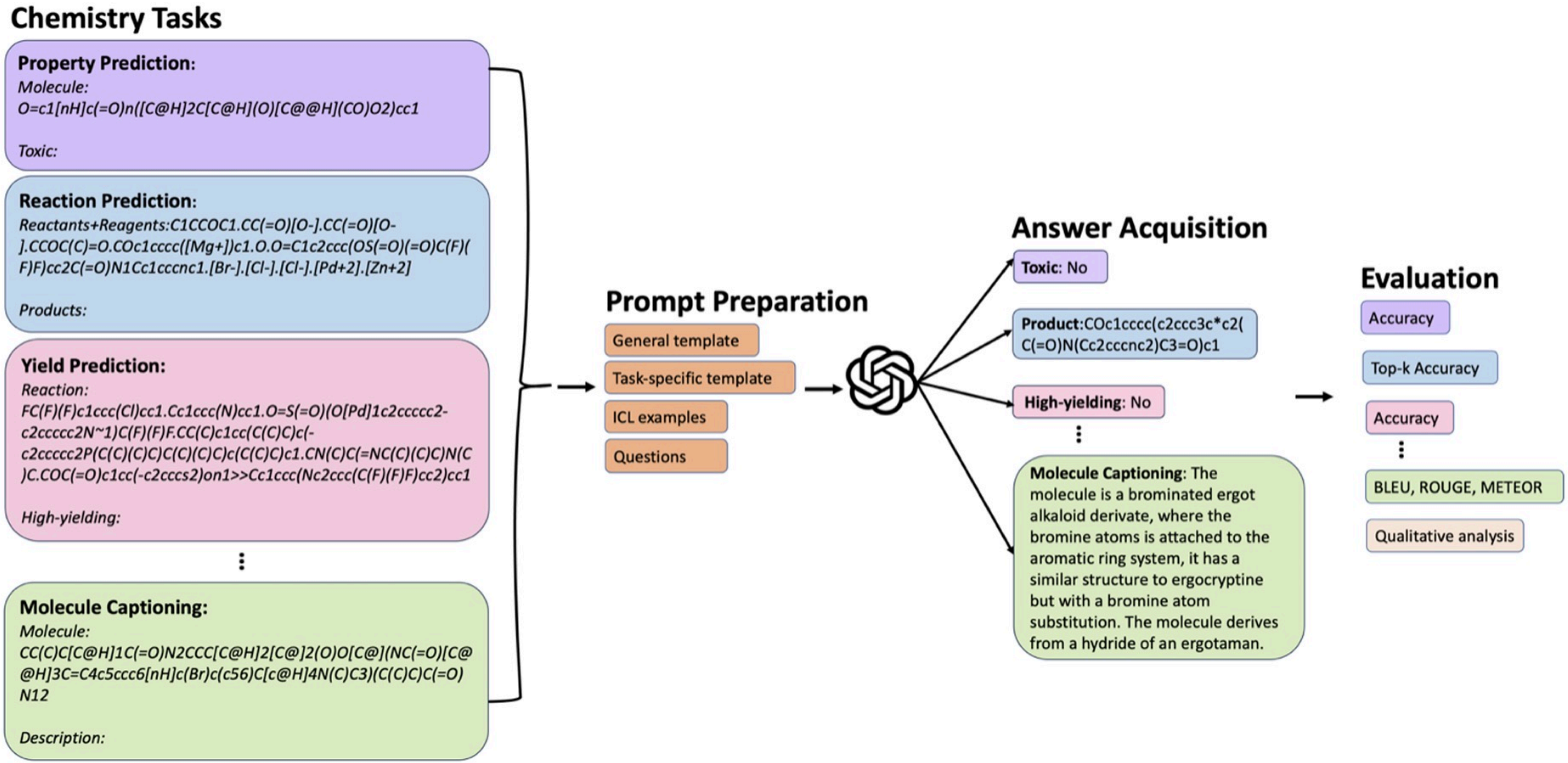
Yejin Choi

Jiawei Han



Lianhui Qin

Previous chemistry tasks



Previous chemistry tasks

Direct question-answering;
Lack complex reasoning process

Bleu	Description-guided molecule design	Retrosynthesis	Forward reaction prediction
Alpaca	0.4	6.3	6.5
Baize	0.6	9.5	4.4
ChatGLM	0.4	11.7	18.3
Llama	0.3	3.6	2.0
Vicuna	0.6	5.7	5.7
Galactica	19.2	45.2	46.8
Text+Chem T5	50.8	76.5	78.2

Unifying textual and molecule representation via multi-tasking.

Mol-Instructions: A Large-Scale Biomolecular Instruction Dataset for Large Language Models. Fang et al., arXiv 2306.

Chemistry Problem

At a particular temperature, a 2.00L flask at equilibrium contains 2.80×10^{-4} mol N_2, 2.50×10^{-5}mol O_2, and 2.00×10^{-2} mol N_2O. How would you calculate K_c at this temperature for the following reaction: $N_2(g) + O_2(g) \rightarrow N_2O(g)$?

K_c means the equilibrium constant.

How will chemicals behave at equilibrium?

Understanding of this certain chemistry reaction.

SciBench: Evaluating College-Level Scientific Problem-Solving Abilities of Large Language Models Wang et al., arXiv 2023.

GPT-4 solution with CoT prompting

For the given reaction:

$$N_2(g) + O_2(g) \rightarrow N_2O(g)$$

The equilibrium constant K_c is defined as follows:

$$K_c = \frac{[N_2O]}{([N_2] \times [O_2])}$$

where $[N_2]$, $[O_2]$, and $[N_2O]$ are the molar concentrations of N_2, O_2, and N_2O respectively at equilibrium.

$$[N_2] = \frac{2.80 \times 10^{-4} mol}{2.00L} = 1.40 \times 10^{-4} M$$

Then plug these values into the formula for K_c:

$$K_c = \frac{(1.00 \times 10^{-2})}{1.40 \times 10^{-4} \times 1.25 \times 10^{-5}} = 5.71 \times 10^6 K$$

Chemistry Problem

At a particular temperature, a 2.00L flask at equilibrium contains 2.80×10^{-4} mol N_2, 2.50×10^{-5}mol O_2, and 2.00×10^{-2} mol N_2O. How would you calculate K_c at this temperature for the following reaction: $N_2(g) + O_2(g) \rightarrow N_2O(g)$?

GPT-4 solution with CoT prompting

For the given reaction:

$$N_2(g) + O_2(g) \rightarrow N_2O(g)$$

The equilibrium constant K_c is defined as follows:

$$K_c = \frac{[N_2O]}{([N_2] \times [O_2])}$$

where $[N_2]$, $[O_2]$, and $[N_2O]$ are the molar concentrations of N_2, O_2, and N_2O respectively at equilibrium.

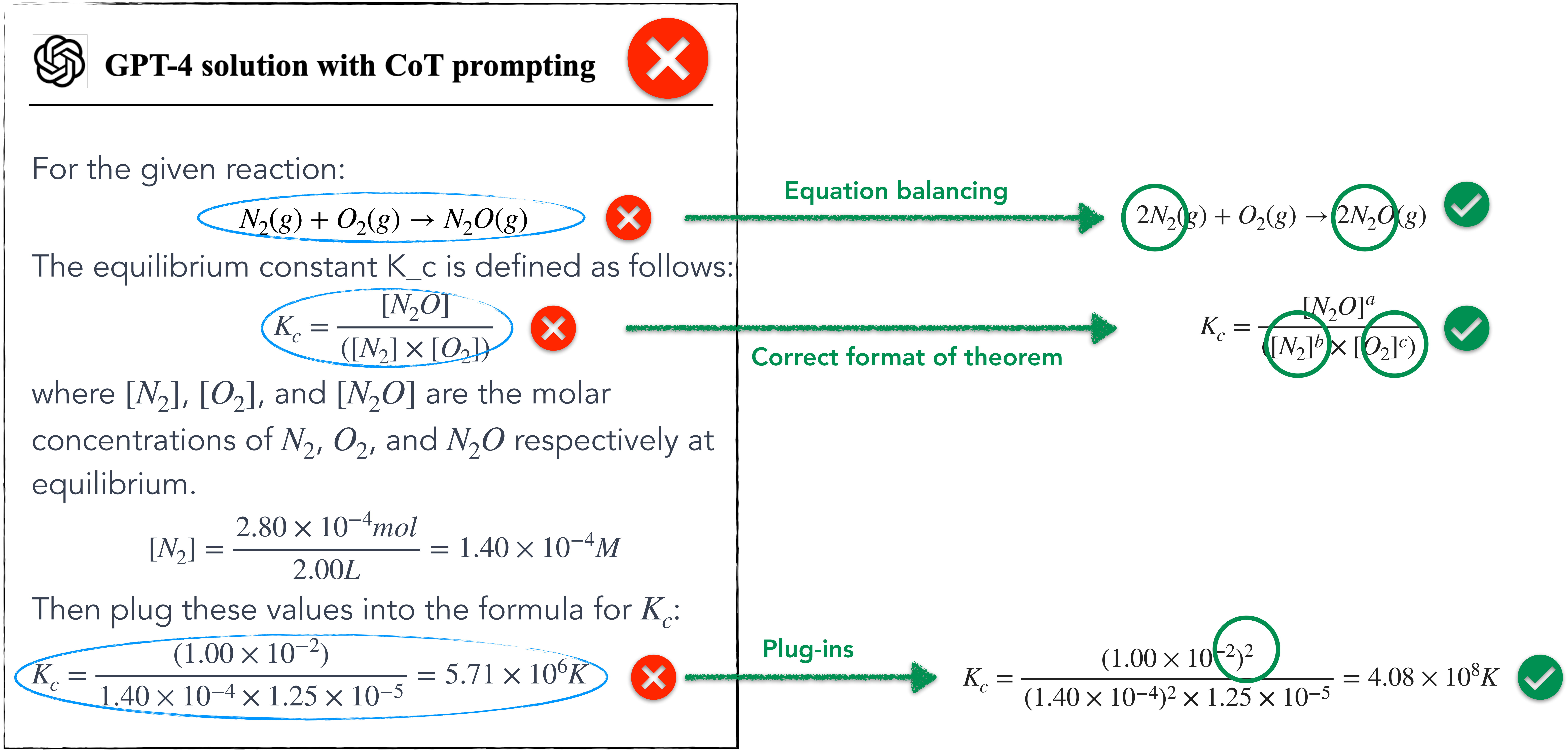
$$[N_2] = \frac{2.80 \times 10^{-4} mol}{2.00L} = 1.40 \times 10^{-4} M$$

Then plug these values into the formula for K_c:

$$K_c = \frac{(1.00 \times 10^{-2})}{1.40 \times 10^{-4} \times 1.25 \times 10^{-5}} = 5.71 \times 10^{6} K$$

Chemistry Problem

At a particular temperature, a 2.00L flask at equilibrium contains 2.80×10^{-4} mol N_2, 2.50×10^{-5}mol O_2, and 2.00×10^{-2} mol N_2O. How would you calculate K_c at this temperature for the following reaction: $N_2(g) + O_2(g) \rightarrow N_2O(g)$?


GPT-4 solution with CoT prompting
For the given reaction:
$N_2(g) + O_2(g) \rightarrow N_2O(g)$
The equilibrium constant K_c is defined as follows:
$K_c = \frac{[N_2O]}{([N_2] \times [O_2])}$
where $[N_2]$, $[O_2]$, and $[N_2O]$ are the molar concentrations of N_2, O_2, and N_2O respectively at equilibrium.
$[N_2] = \frac{2.80 \times 10^{-4} mol}{2.00 L} = 1.40 \times 10^{-4} M$
Then plug these values into the formula for K_c:
$K_c = \frac{(1.00 \times 10^{-2})}{1.40 \times 10^{-4} \times 1.25 \times 10^{-5}} = 5.71 \times 10^6 K$
Equation balancing
$2N_2(g) + O_2(g) \rightarrow 2N_2O(g)$
Correct format of theorem
$K_c = \frac{[N_2O]^a}{([N_2]^b \times [O_2]^c)}$
Plug-ins
$K_c = \frac{(1.00 \times 10^{-2})^2}{(1.40 \times 10^{-4})^2 \times 1.25 \times 10^{-5}} = 4.08 \times 10^8 K$

GPT-4 solution with CoT prompting ✗

For the given reaction:

$$N_2(g) + O_2(g) \rightarrow N_2O(g)$$ ✗

The equilibrium constant K_c is defined as follows:

$$K_c = \frac{[N_2O]}{([N_2] \times [O_2])}$$ ✗

where $[N_2]$, $[O_2]$, and $[N_2O]$ are the molar concentrations of N_2, O_2, and N_2O respectively at equilibrium.

$$[N_2] = \frac{2.80 \times 10^{-4} mol}{2.00L} = 1.40 \times 10^{-4} M$$

Then plug these values into the formula for K_c:

$$K_c = \frac{(1.00 \times 10^{-2})}{1.40 \times 10^{-4} \times 1.25 \times 10^{-5}} = 5.71 \times 10^{6} K$$ ✗

Correct solution ✓

Firstly, we need to balance the given reaction:

$$2N_2(g) + O_2(g) \rightarrow 2N_2O(g)$$ ✓

The equilibrium constant K_c is defined as follows:

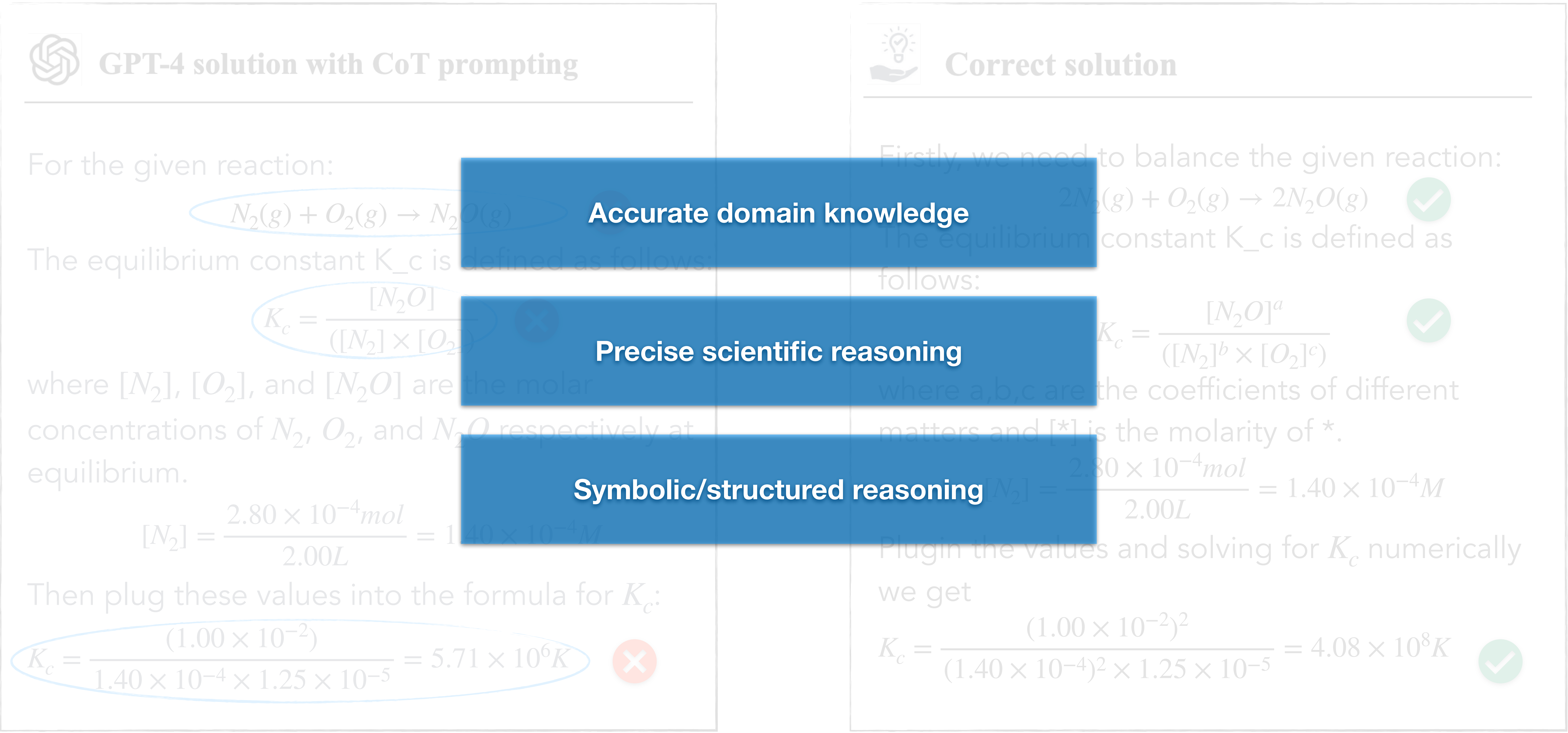
$$K_c = \frac{[N_2O]^a}{([N_2]^b \times [O_2]^c)}$$ ✓

where a,b,c are the coefficients of different matters and [*] is the molarity of *.

$$[N_2] = \frac{2.80 \times 10^{-4} mol}{2.00L} = 1.40 \times 10^{-4} M$$

Plugin the values and solving for K_c numerically we get

$$K_c = \frac{(1.00 \times 10^{-2})^2}{(1.40 \times 10^{-4})^2 \times 1.25 \times 10^{-5}} = 4.08 \times 10^{8} K$$ ✓


GPT-4 solution with CoT prompting
Correct solution
Accurate domain knowledge
Precise scientific reasoning
Symbolic/structured reasoning

It’s really challenging task…

- **Accurate domain knowledge**
- **Precise scientific reasoning**
- **Symbolic/structured reasoning**

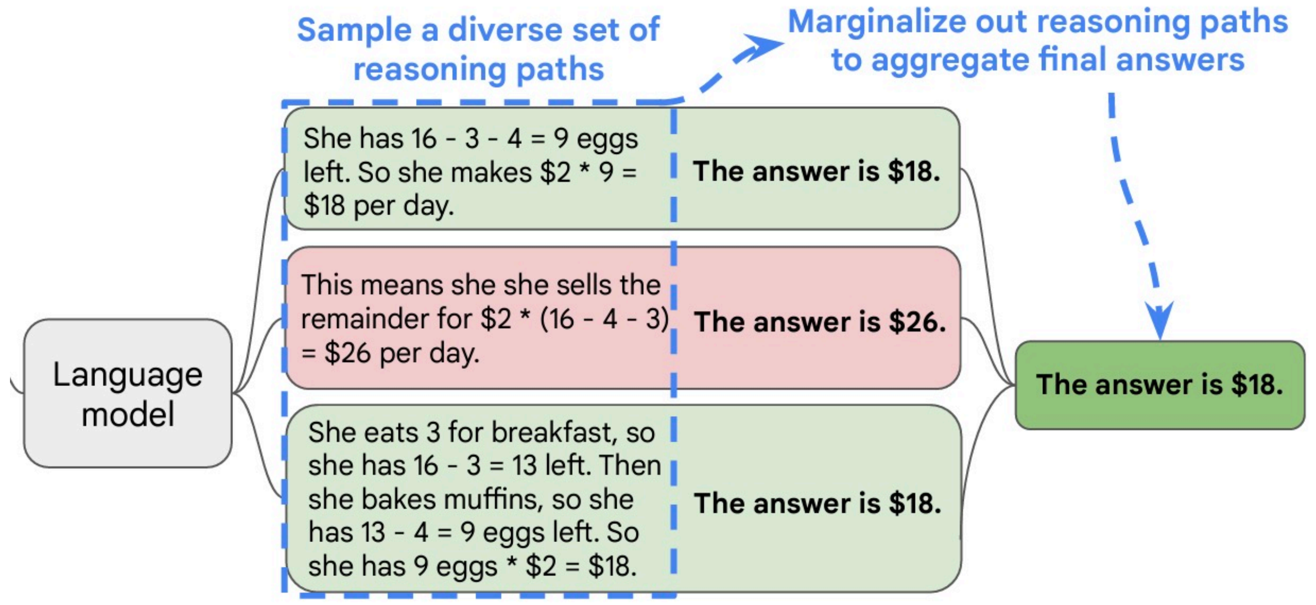
Undergrad students from top university

Fail on the task

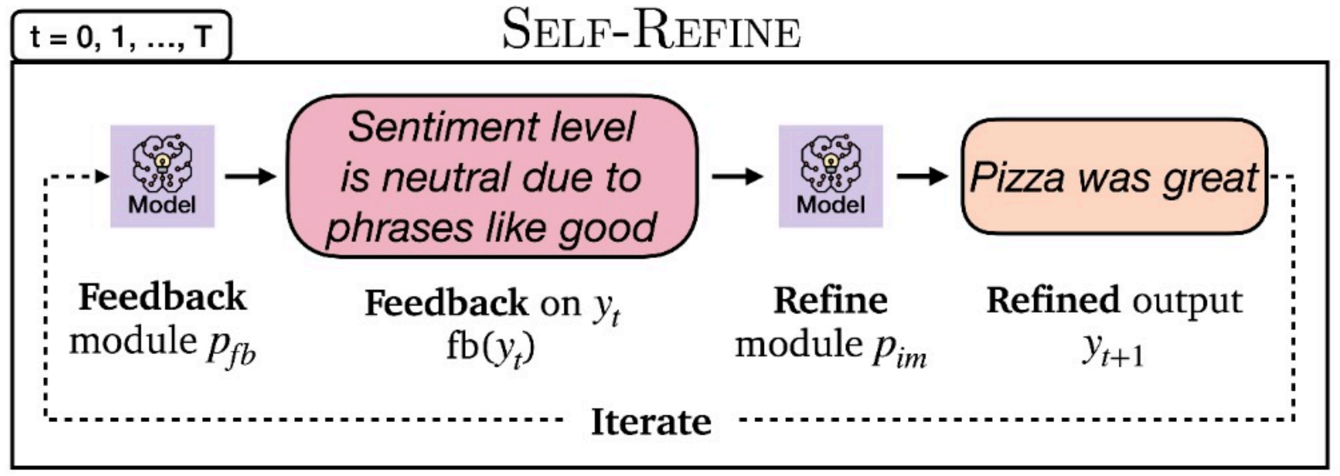
Score: 27/100

Recap on related works

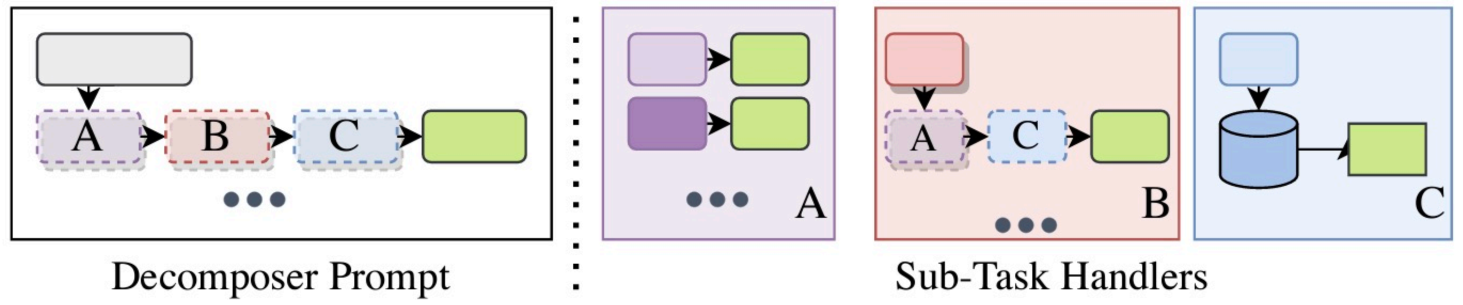
CoT (Wei et al., 2022)



Self-Consistency (Wang et al., 2022): majority vote



Self-Refine (Madaan et al., 2023): feedback machinist

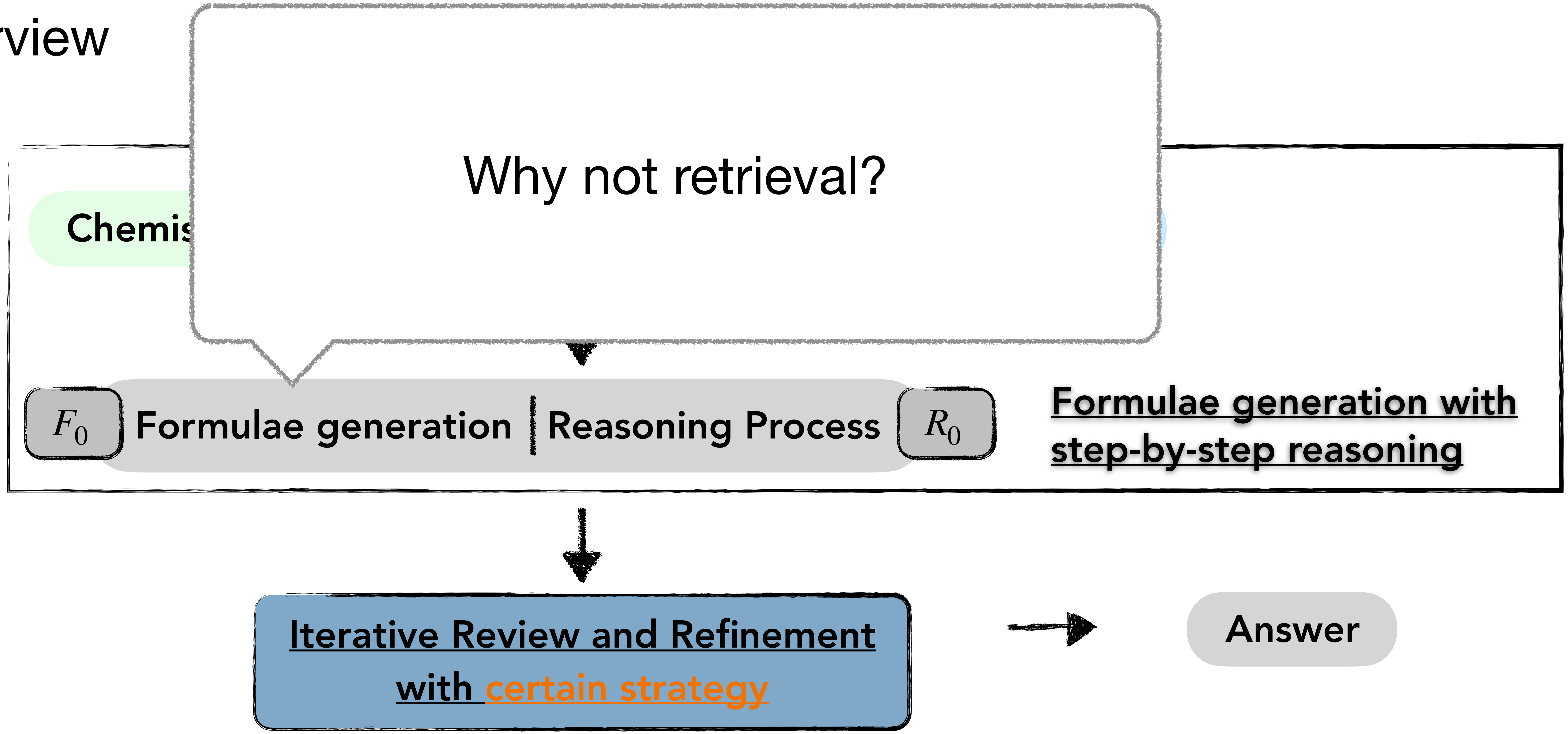


Decomposed Prompting (Khot et al., 2022): modular

StructChem:

A structural approach that elicits chemistry reasoning in LLMs

- Overview



Structured instruction

Instruction

Please provide a clear and step-by-step solution for a scientific problem in the categories of Chemistry. The problem will specify the unit of measurement, which should not be included in the answer. Express the final answer as a decimal number with three digits after the decimal point. Conclude the answer by stating "The answer is therefore \\boxed{[ANSWER]}."

For each instance, you need to three things. Firstly, for "formulae retrieval", you need to identify the formulae explicitly and implicitly entailed in the problem context. Then there is a "reasoning/calculation process" where you are required to reason step by step based on the identified formulae and problem context. Finally, conclude the answer. For each problem, the output format should incorporate the following components in the corresponding format:

Output Format

**Formulae retrieval: **
[Formula 1] (the formula required to solve the problem)
[Formula 2] (the second formula required to solve the problem, if any)
...
[Formula n] (the n-th formula required to solve the problem, if any)

Reasoning/calculation process:
[step 1] (the first step for solving this problem)
.....
[step n] (the n-th step for solving the problem, if any)

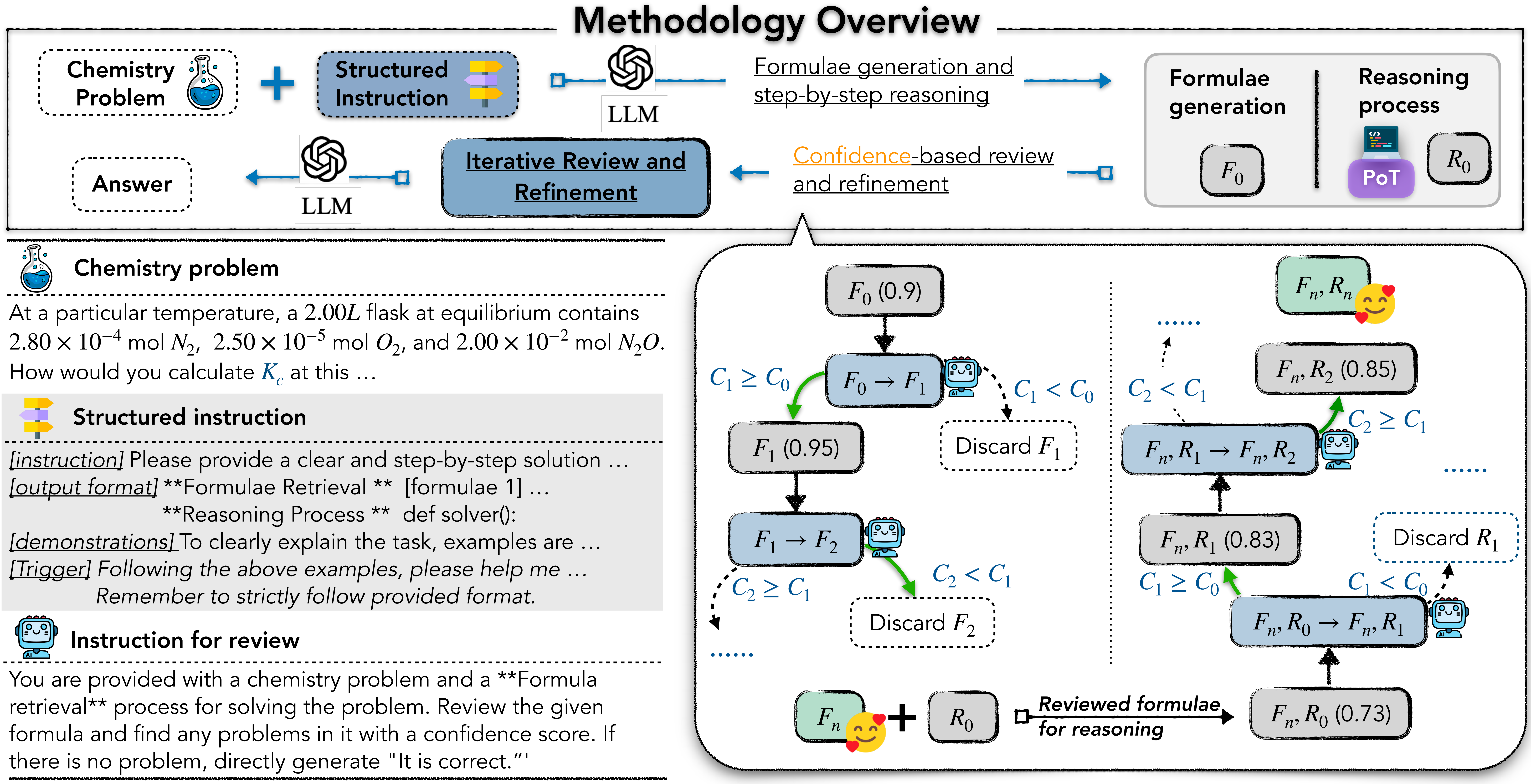
Answer conclusion:
[answer] The answer is therefore \\boxed{[ANSWER]}.

Demonstration (5 samples)

To clearly explain the task, we provide the following example:

Problem:
Assume that all gases are perfect and that data refer to 298.15 K unless otherwise stated. Calculate the change in chemical potential of a perfect gas when its pressure is increased isothermally from $1.8 \\mathrm{~atm} $ to $29.5 \\mathrm{~atm}$ at $40^{\\circ} \\mathrm{C}$. The unit of the answer is $\\mathrm{kJ} \\mathrm{mol}^{-1}$.

Response:
In order to solve this problem, we will use the formula for the change in chemical potential \(\Delta \mu \) of a perfect gas due to a change in pressure. Given that the temperature is constant (isothermal), the chemical potential of a perfect gas is given by:
……


Methodology Overview
Chemistry Problem
Structured Instruction
LLM
Formulae generation and step-by-step reasoning
Formulae generation
Reasoning process
F_0
PoT
R_0
Answer
LLM
Iterative Review and Refinement
Confidence-based review and refinement
Chemistry problem
At a particular temperature, a $2.00L$ flask at equilibrium contains 2.80×10^{-4} mol N_2, 2.50×10^{-5} mol O_2, and 2.00×10^{-2} mol N_2O. How would you calculate K_c at this …
Structured instruction
[instruction] Please provide a clear and step-by-step solution …
[output format] **Formulae Retrieval ** [formulae 1] …
**Reasoning Process ** def solver():
[demonstrations] To clearly explain the task, examples are …
[Trigger] Following the above examples, please help me …
Remember to strictly follow provided format.
Instruction for review
You are provided with a chemistry problem and a **Formula retrieval** process for solving the problem. Review the given formula and find any problems in it with a confidence score. If there is no problem, directly generate "It is correct."'
F_0 (0.9)
$F_0 \rightarrow F_1$
$C_1 \geq C_0$
$C_1 < C_0$
F_1 (0.95)
Discard F_1
$F_1 \rightarrow F_2$
$C_2 \geq C_1$
$C_2 < C_1$
Discard F_2
F_n, R_n
F_n, R_2 (0.85)
$C_2 < C_1$
$C_2 \geq C_1$
$F_n, R_1 \rightarrow F_n, R_2$
F_n, R_1 (0.83)
Discard R_1
$C_1 \geq C_0$
$C_1 < C_0$
$F_n, R_0 \rightarrow F_n, R_1$
F_n
R_0
Reviewed formulae for reasoning
F_n, R_0 (0.73)

Iterative Review and Refinement

F_0

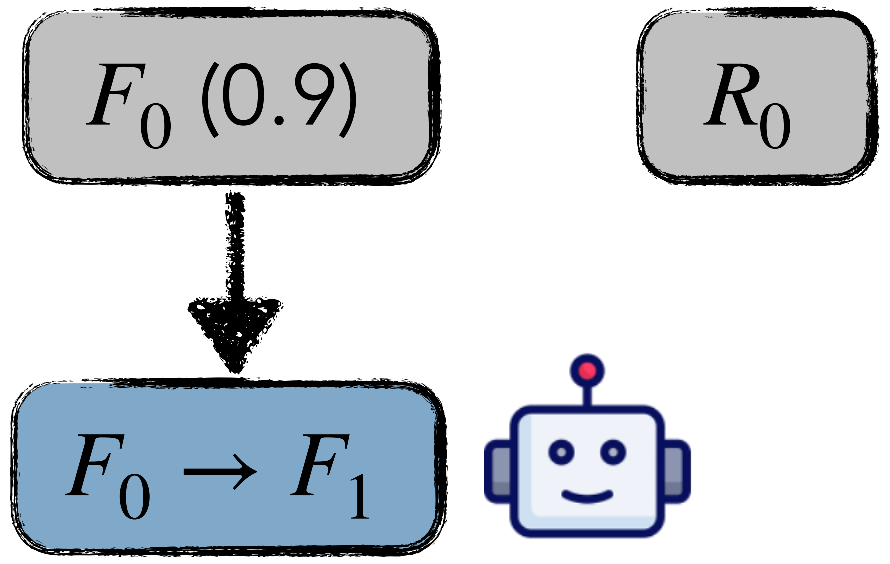
[Formulae 1] $2N_2(g) + O_2(g) \rightarrow 2N_2O(g)$

[Formulae 2] $K_c = \frac{[N_2O]}{([N_2] \times [O_2])}$

Confidence score C_0: 0.9

Reviewer 1

Iterative Review and Refinement



F_0

[Formulae 1] $2N_2(g) + O_2(g) \rightarrow 2N_2O(g)$

[Formulae 2] $K_c = \frac{[N_2O]}{([N_2] \times [O_2])}$

Confidence score C_0: 0.9

Higher

Reviewer 1

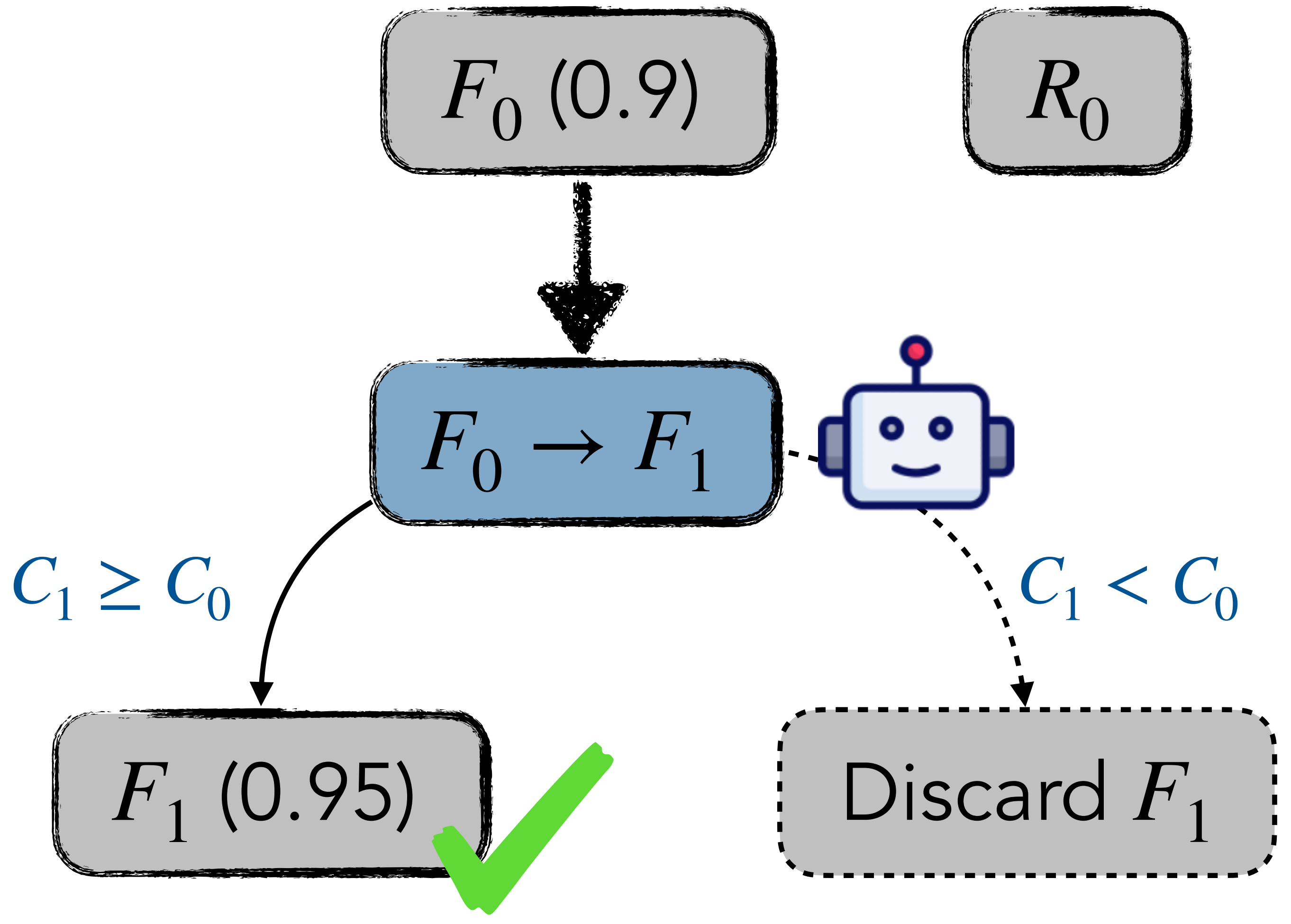
F_1

[Formulae 1] $2N_2(g) + O_2(g) \rightarrow 2N_2O(g)$

[Formulae 2] $K_c = \frac{[N_2O]^a}{([N_2]^b \times [O_2]^c)}$

Confidence score C_1: 0.95

Iterative Review and Refinement



F_0
[Formulae 1] $2N_2(g) + O_2(g) \rightarrow 2N_2O(g)$
[Formulae 2] $K_c = \frac{[N_2O]}{([N_2] \times [O_2])}$
Confidence score C_0: 0.9
Higher
Reviewer 1
F_1
[Formulae 1] $2N_2(g) + O_2(g) \rightarrow 2N_2O(g)$
[Formulae 2] $K_c = \frac{[N_2O]^a}{([N_2]^b \times [O_2]^c)}$
Confidence score C_1: 0.95
Lower
Reviewer 2
F_2
[Formulae 1] $N_2(g) + O_2(g) \rightarrow N_2O(g)$
[Formulae 2] $K_c = \frac{[N_2O]^a}{([N_2]^b \times [O_2]^c)}$
[Formulae 3] $M(g) = \frac{mol(g)}{C(g)}$
Confidence score C_2: 0.70
Iterative Review and Refinement
F_0 (0.9)
R_0
$F_0 \rightarrow F_1$
$C_1 \geq C_0$
$C_1 < C_0$
F_1 (0.95)
Discard F_1
$F_1 \rightarrow F_2$
$C_2 \geq C_1$
$C_2 < C_1$
Discard F_2
......

F_0
[Formulae 1] $2N_2(g) + O_2(g) \rightarrow 2N_2O(g)$
[Formulae 2] $K_c = \frac{[N_2O]}{([N_2] \times [O_2])}$
Confidence score C_0: 0.9
Higher
Reviewer 1
F_1
[Formulae 1] $2N_2(g) + O_2(g) \rightarrow 2N_2O(g)$
[Formulae 2] $K_c = \frac{[N_2O]^a}{([N_2]^b \times [O_2]^c)}$
Confidence score C_1: 0.95
Lower
Reviewer 2
F_2
[Formulae 1] $N_2(g) + O_2(g) \rightarrow N_2O(g)$
[Formulae 2] $K_c = \frac{[N_2O]^a}{([N_2]^b \times [O_2]^c)}$
[Formulae 3] $M(g) = \frac{mol(g)}{C(g)}$
Confidence score C_2: 0.70
......
Reviewer n
F_n
Iterative Review and Refinement
F_0 (0.9)
R_0
$F_0 \rightarrow F_1$
$C_1 \geq C_0$
$C_1 < C_0$
F_1 (0.95)
Discard F_1
$F_1 \rightarrow F_2$
$C_2 \geq C_1$
$C_2 < C_1$
Discard F_2
F_n

F_n, R_0

[Formulae 1] $2N_2(g) + O_2(g) \rightarrow 2N_2O(g)$

[Formulae 2] $K_c = \frac{[N_2O]}{([N_2] \times [O_2])}$

[Formulae 3] …

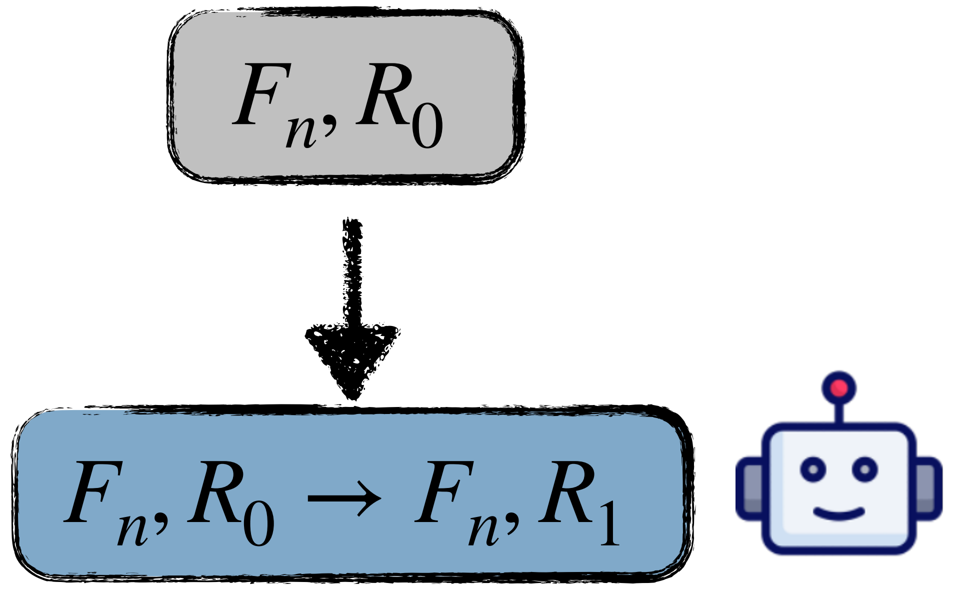
[Step 1] First calculate the molarity of each gas based on Formulae 3, …

[Step 2] …

Confidence score C_0: 0.76

Reviewer 1

Iterative Review and Refinement



F_n, R_0
[Formulae 1] $2N_2(g) + O_2(g) \rightarrow 2N_2O(g)$
[Formulae 2] $K_c = \frac{[N_2O]}{([N_2] \times [O_2])}$
[Formulae 3] …
[Step 1] First calculate the molarity of each gas based on Formulae 3, …
[Step 2] …
Confidence score C_0: 0.76
Reviewer 1
Higher
F_n, R_1
[Formulae 1] $2N_2(g) + O_2(g) \rightarrow 2N_2O(g)$
[Formulae 2] $K_c = \frac{[N_2O]}{([N_2] \times [O_2])}$
[Formulae 3] …
[Step 1] Based on the balanced equation in Formulae 1, …
[Step 2] …
Confidence score C_1: 0.83
Reviewer 2
Iterative Review and Refinement
F_n, R_0 (0.76)
$F_n, R_0 \rightarrow F_n, R_1$
$C_1 \geq C_0$
$C_1 < C_0$
F_n, R_1 (0.83)
Discard R_1

F_n, R_0
[Formulae 1] $2N_2(g) + O_2(g) \rightarrow 2N_2O(g)$
[Formulae 2] $K_c = \frac{[N_2O]}{([N_2] \times [O_2])}$
[Formulae 3] …
[Step 1] First calculate the molarity of each gas based on Formulae 3, …
[Step 2] …
Confidence score C_0: 0.76
Reviewer 1
Higher
F_n, R_1
[Formulae 1] $2N_2(g) + O_2(g) \rightarrow 2N_2O(g)$
[Formulae 2] $K_c = \frac{[N_2O]}{([N_2] \times [O_2])}$
[Formulae 3] …
[Step 1] Based on the balanced equation in Formulae 1, …
[Step 2] …
Confidence score C_1: 0.83
Reviewer 2
……
F_n, R_n
Iterative Review and Refinement
F_n, R_0 (0.76)
$F_n, R_0 \rightarrow F_n, R_1$
$C_1 \geq C_0$
$C_1 < C_0$
F_n, R_1 (0.83)
Discard R_1
$F_n, R_1 \rightarrow F_n, R_2$
$C_2 \geq C_1$
……
……
F_n, R_n

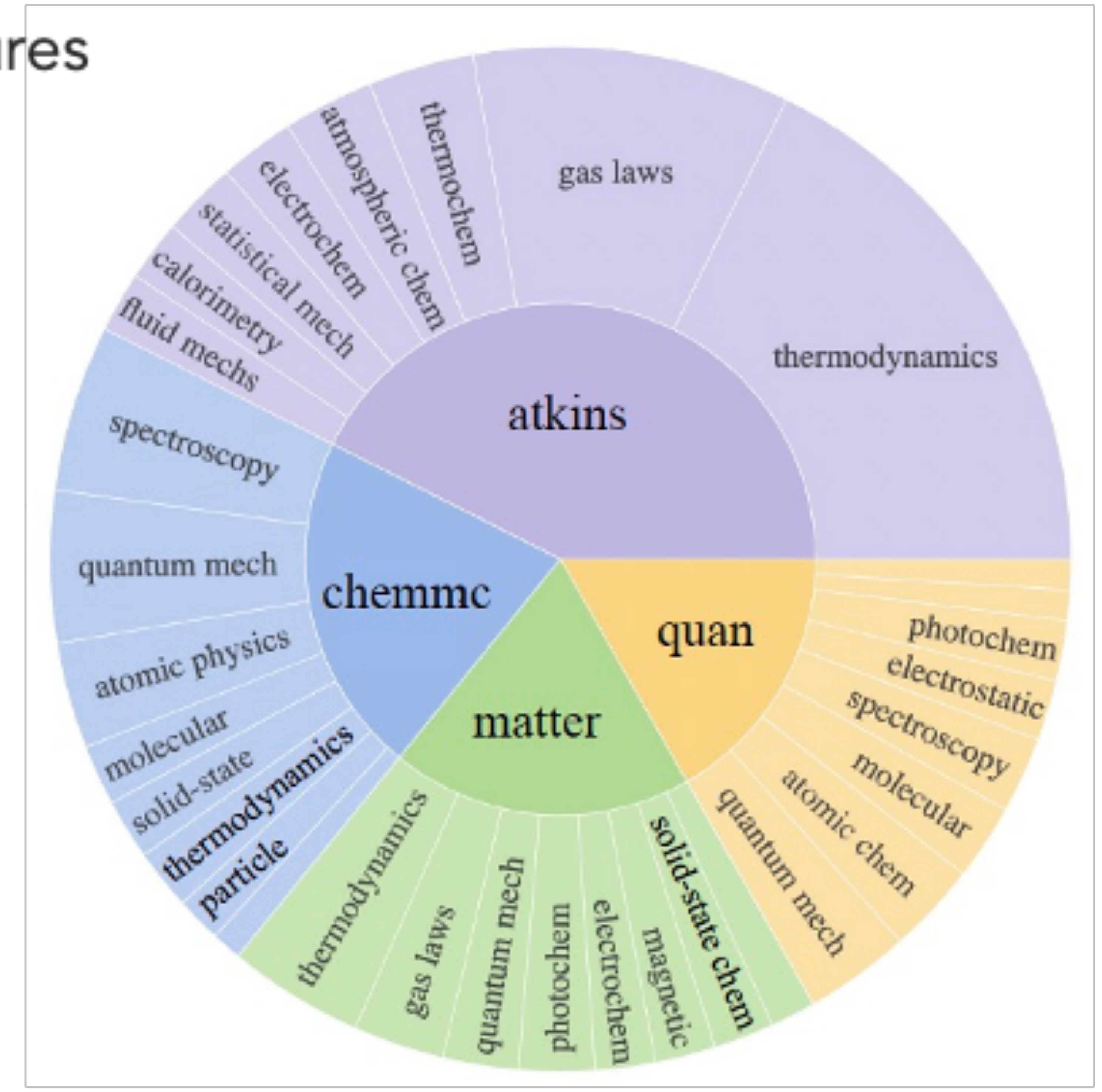
Experiments

SciBench: College-level complex chemistry problems spanning various advanced topics

- Macroscopic Equilibrium Nature: With thermodynamic theory, different states of matter (solid, liquid, gas, dissolved, etc.) reach equilibrium. atkins matter
- Microscopic Structure and Properties: With quantum theory, how microstructures affect the exhibited physical properties. quan matter
- Dynamic Properties: Mechanism of chemical reactions. chemmc matter

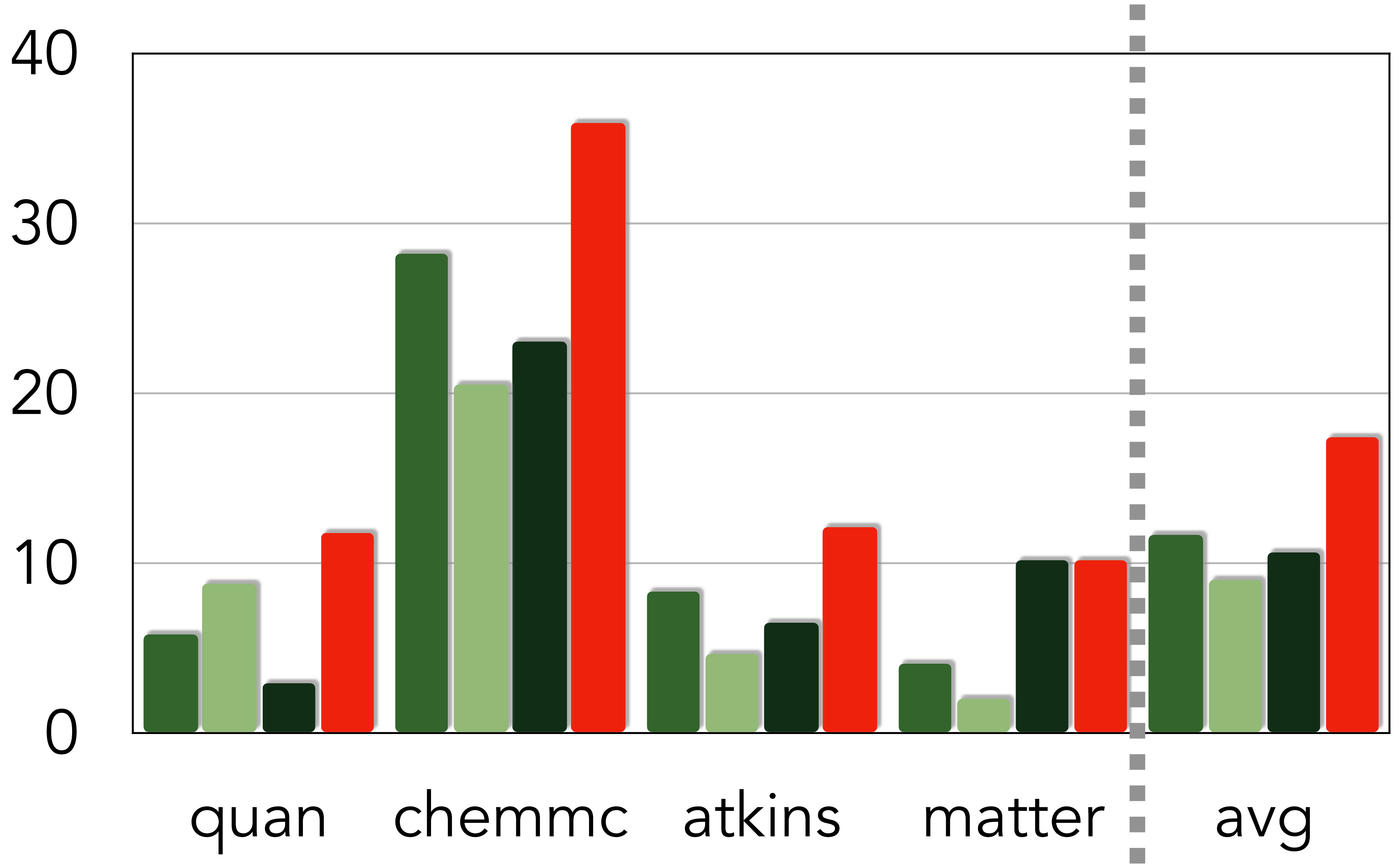
Datasets	Subfields/Topics	# $\mathcal{P}_w(\mathcal{P}_s)$	# F	# RS
quan	Quantum chemistry	34 (8)	1.93	3.94
chemmc	Quantum mechanics	39 (9)	1.88	3.95
atkins	Physical chemistry	107 (16)	1.65	4.33
matter	Chemistry kinetics	49 (10)	1.89	4.43

#F denotes the average number of formulae required, #RS is the average number of reasoning steps.

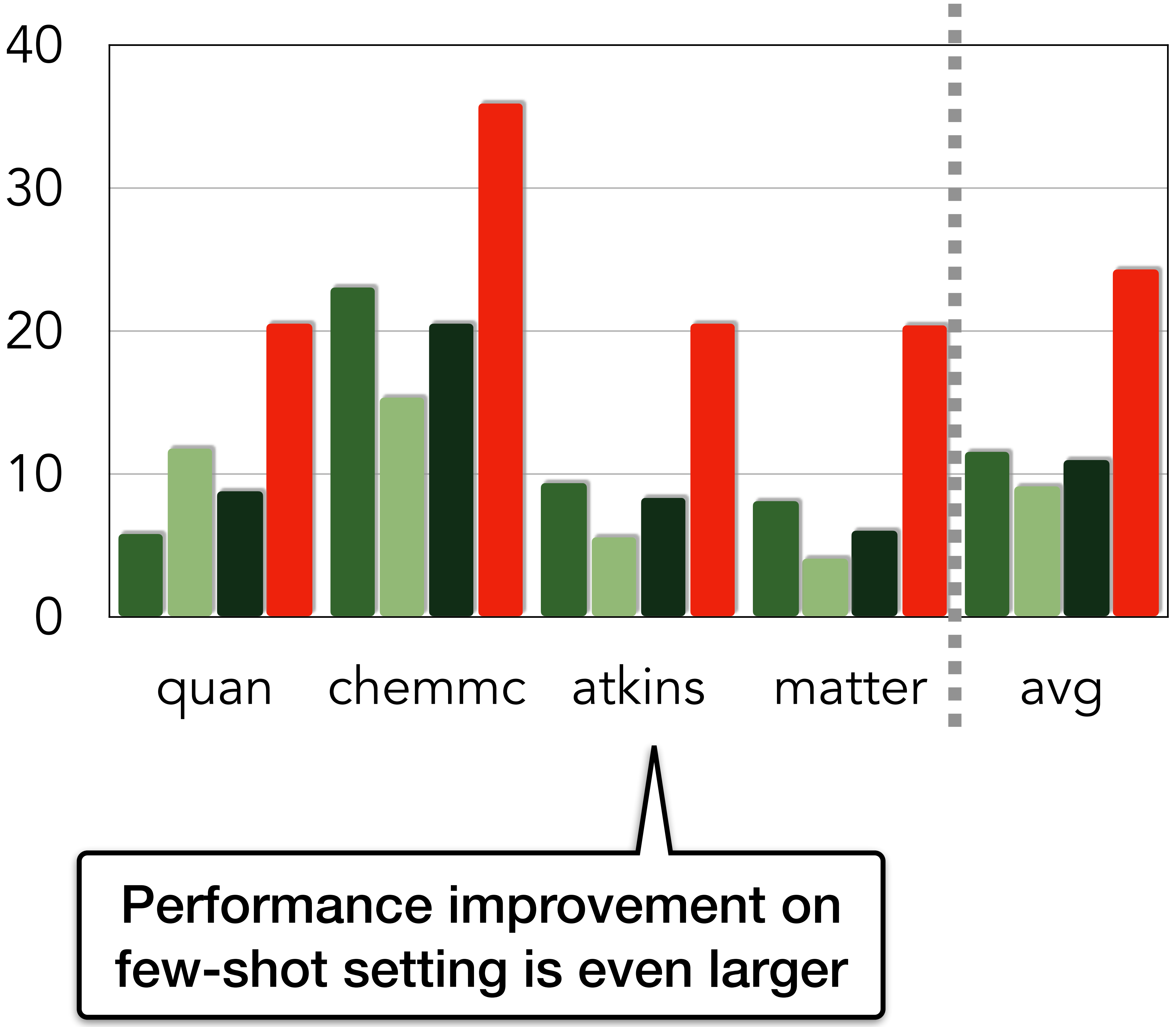


Benchmark performance (GPT-3.5)

Zero-shot setting

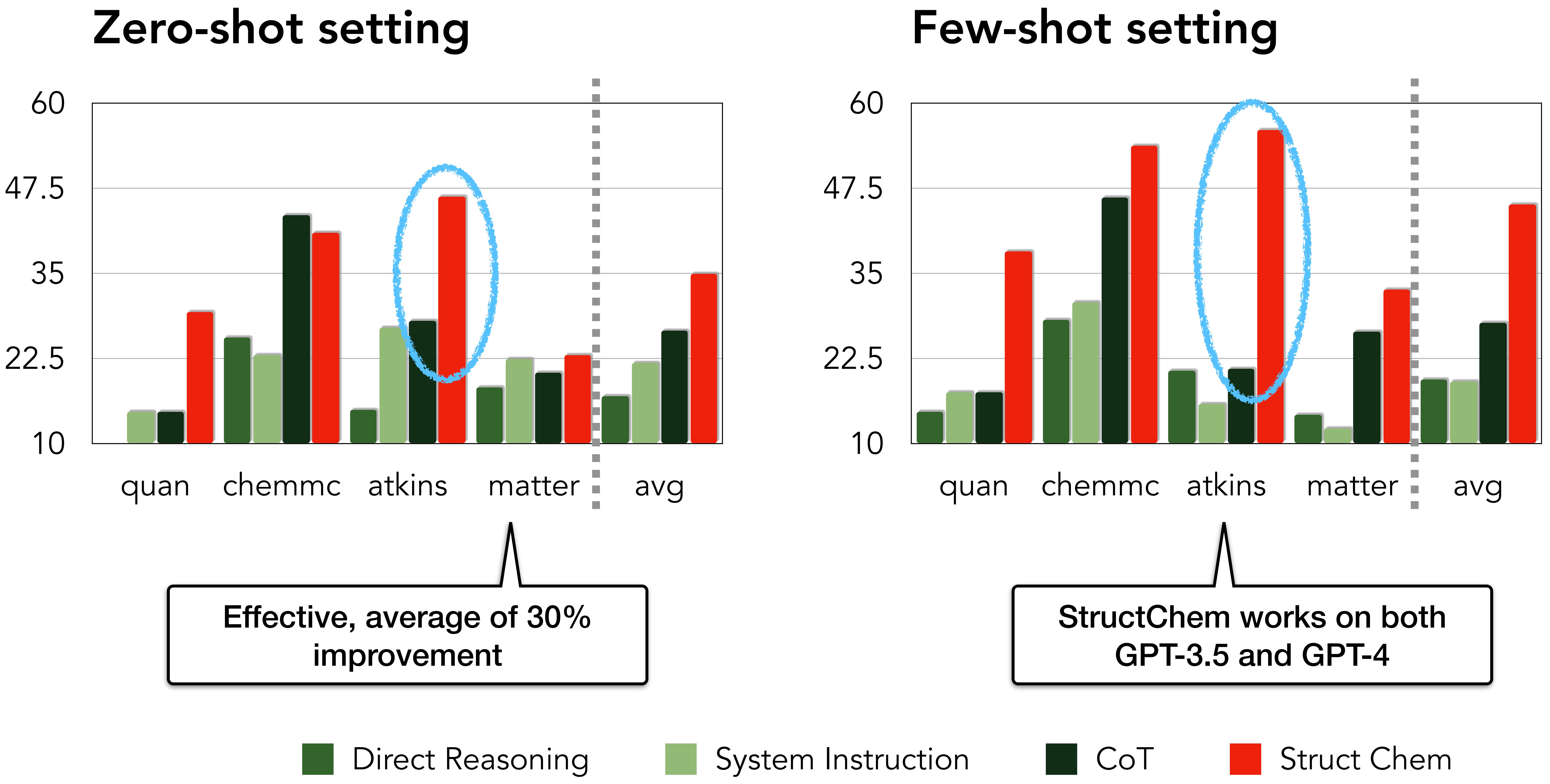


Few-shot setting

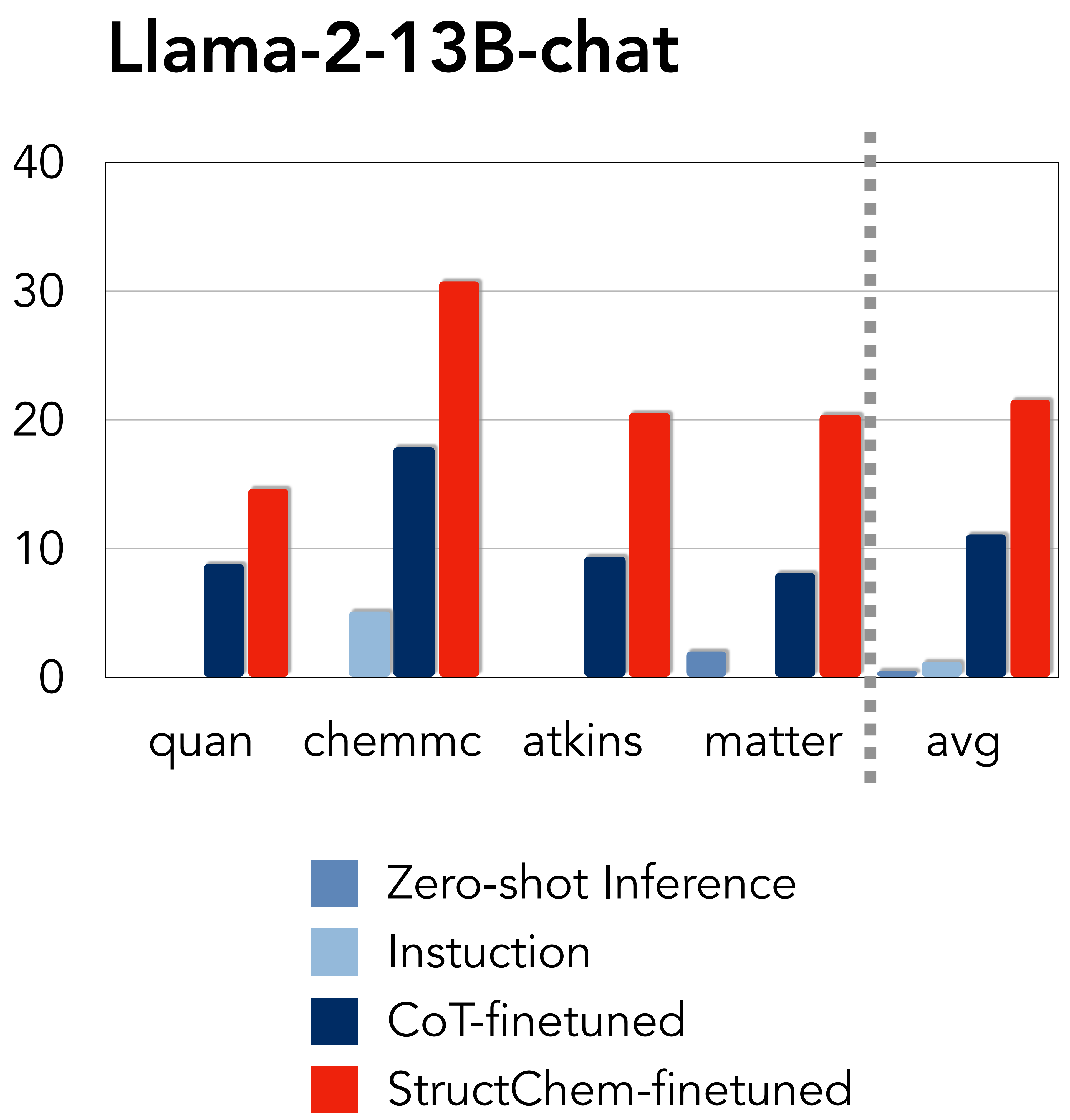


Direct Reasoning   System Instruction   CoT   Struct Chem

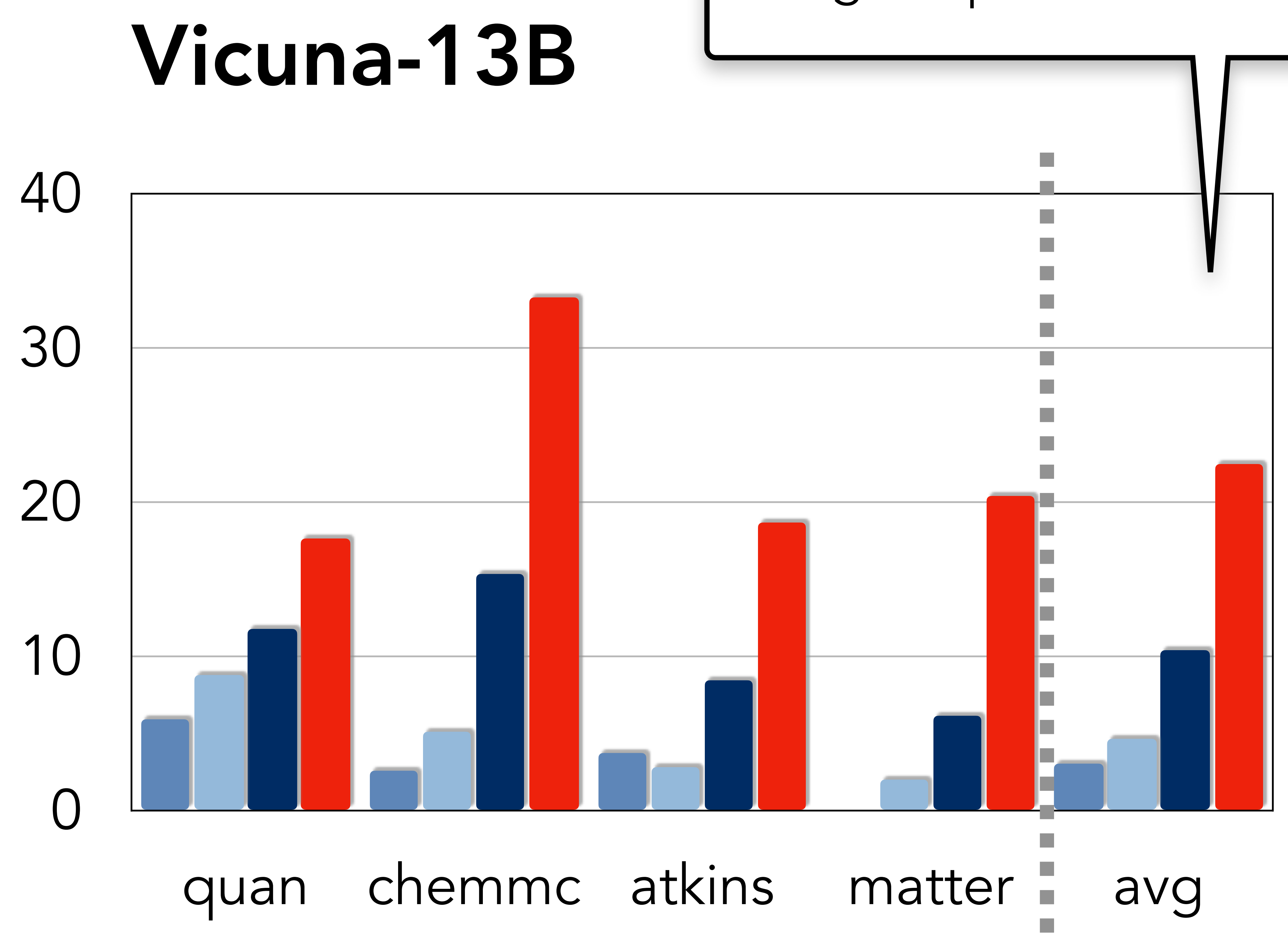
Benchmark performance (GPT-4)



Validating reasoning quality



Huge improvement over baselines.

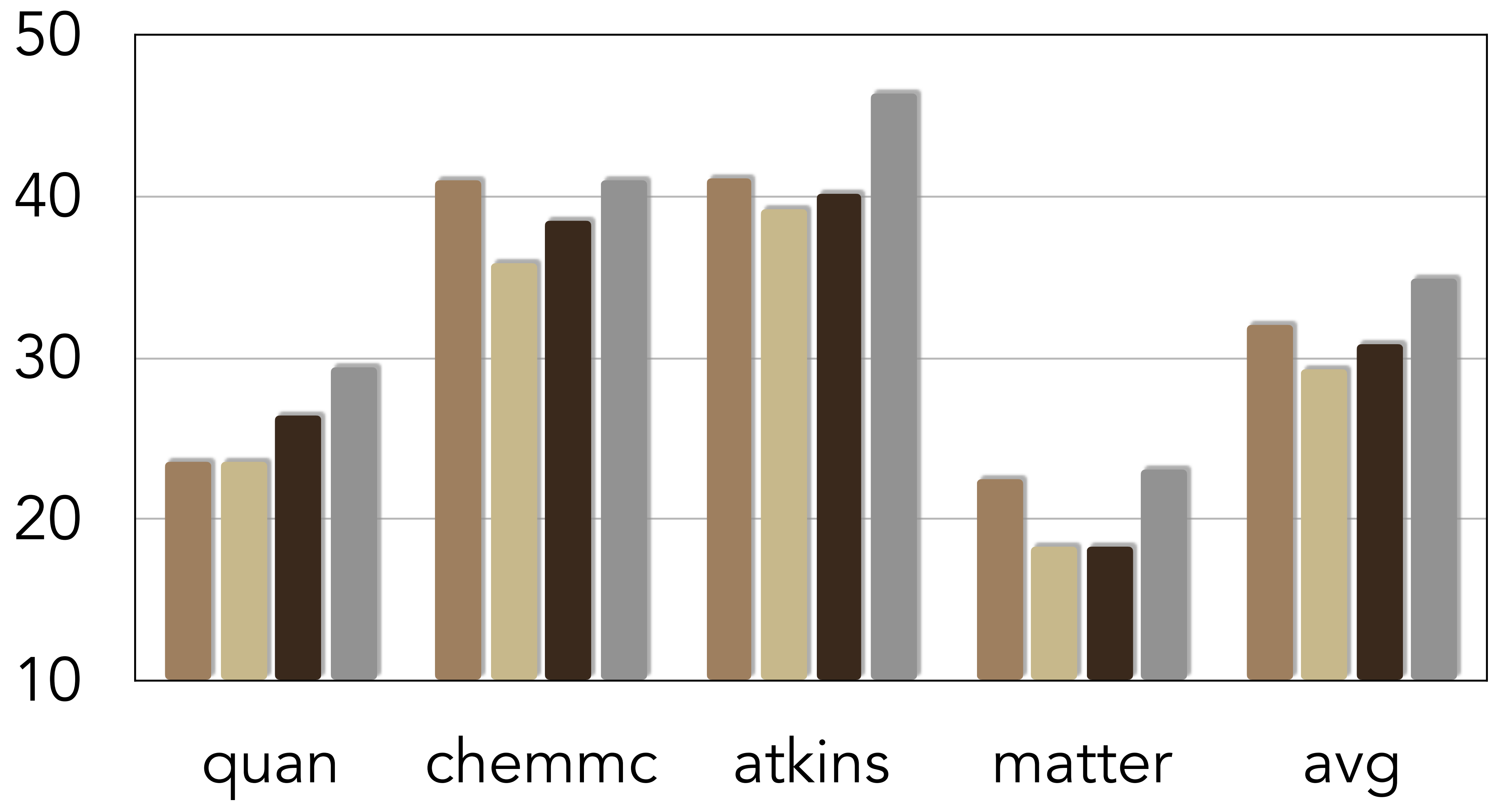


Teach smaller open-sourced models how to reason:

- Chemistry problems generated by GPT-4 as input
- Reasoning processes generated by StructChem as output

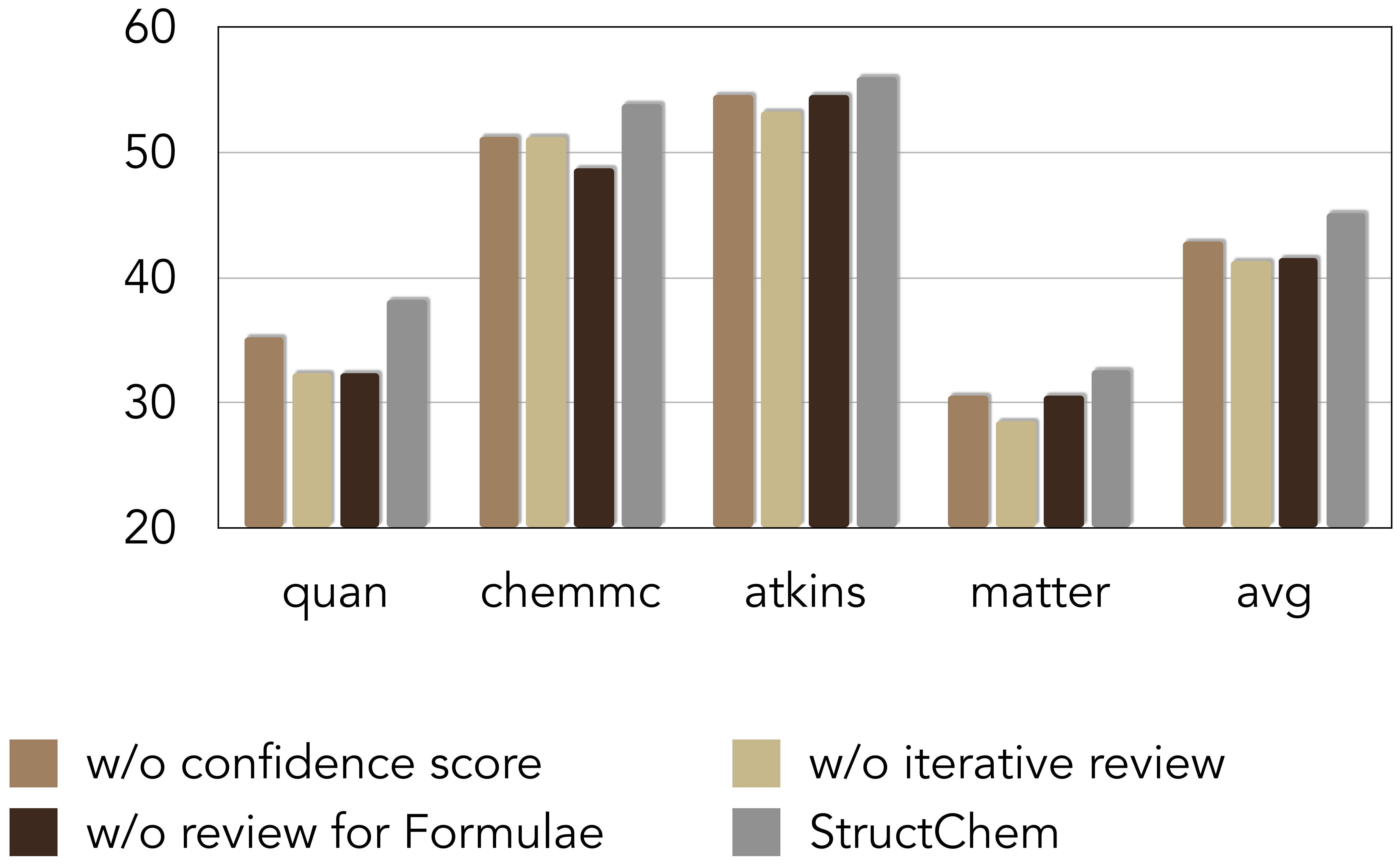
Ablations

Zero-shot setting



Effectiveness of all components.

Few-shot setting



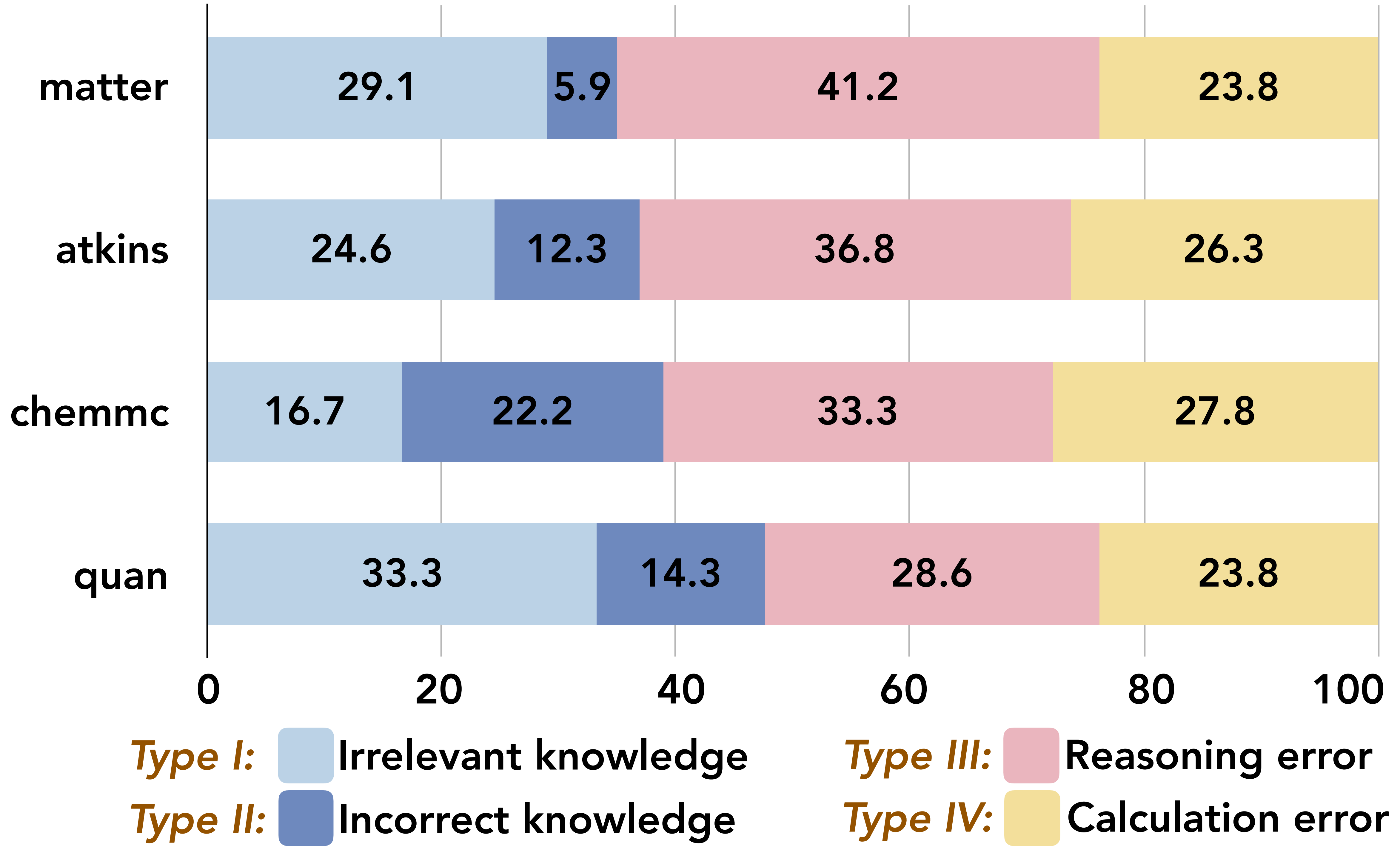
Error analysis

Formulae collection

- **Irrelevance:** irrelevant formulae collected to solving the problem.
- **Incorrectness:** incorrectness inherent in the formula itself.

Reasoning

- **Reasoning error:** errors made during the intermediate reasoning steps.
- **Calculation error:** mathematical computation mistakes made during reasoning process.

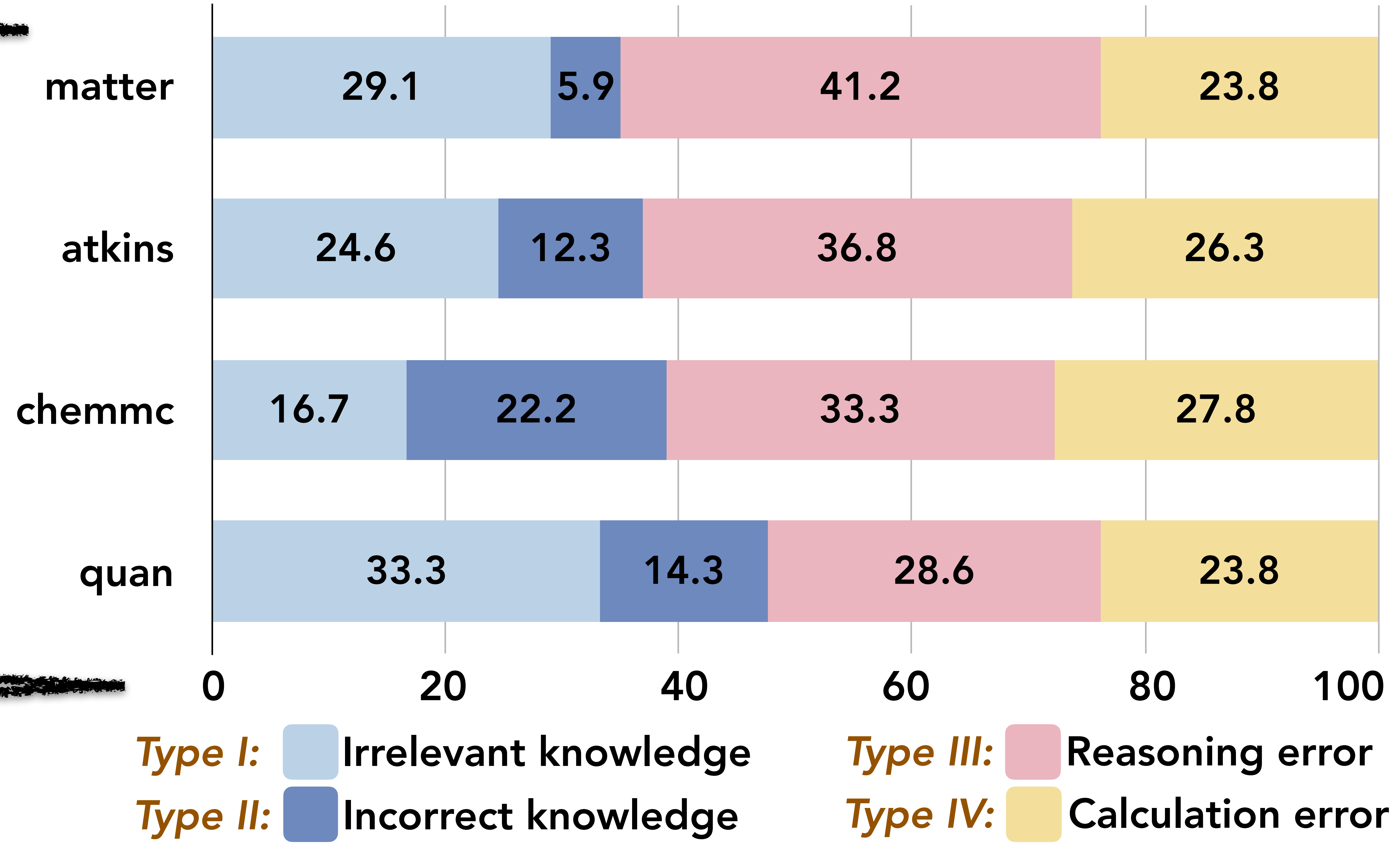


Error analysis

LLMs are more likely to be irrelevant than inaccurate

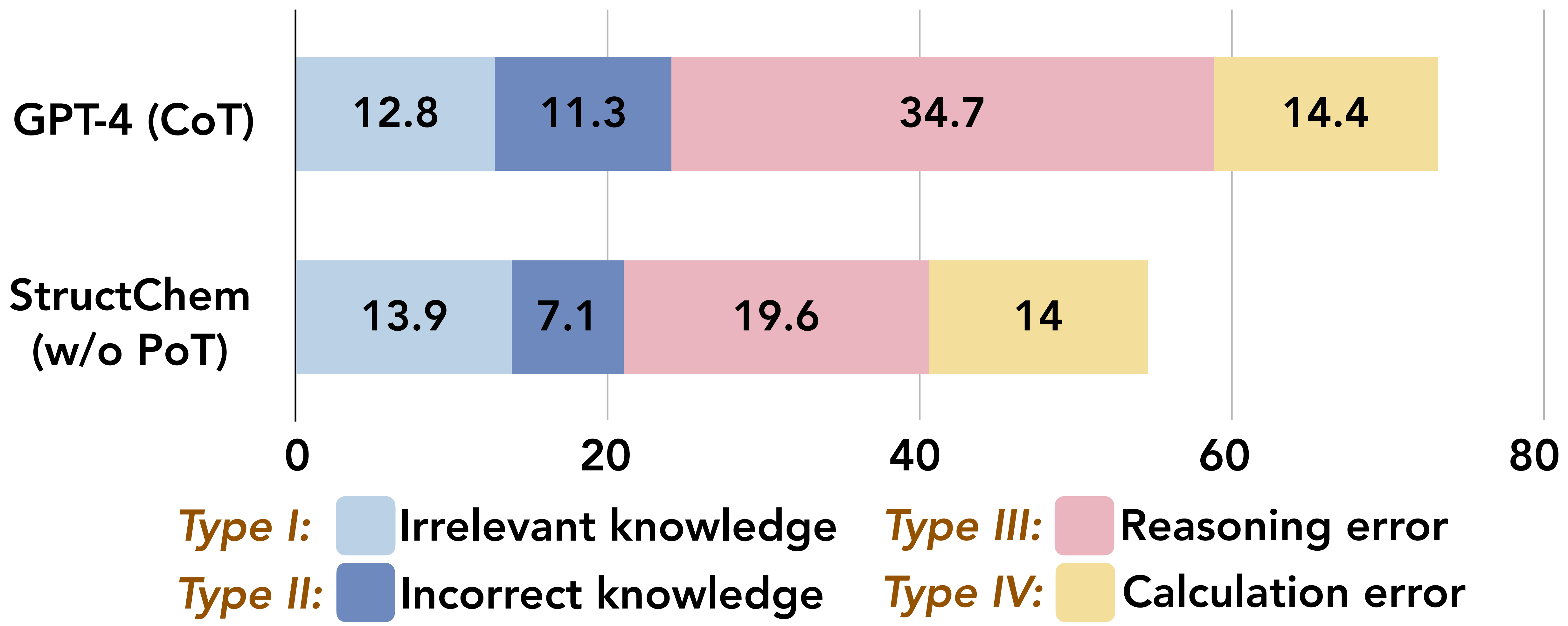
Complex reasoning is still the bottleneck.

Calculation errors are a significant issue as well.



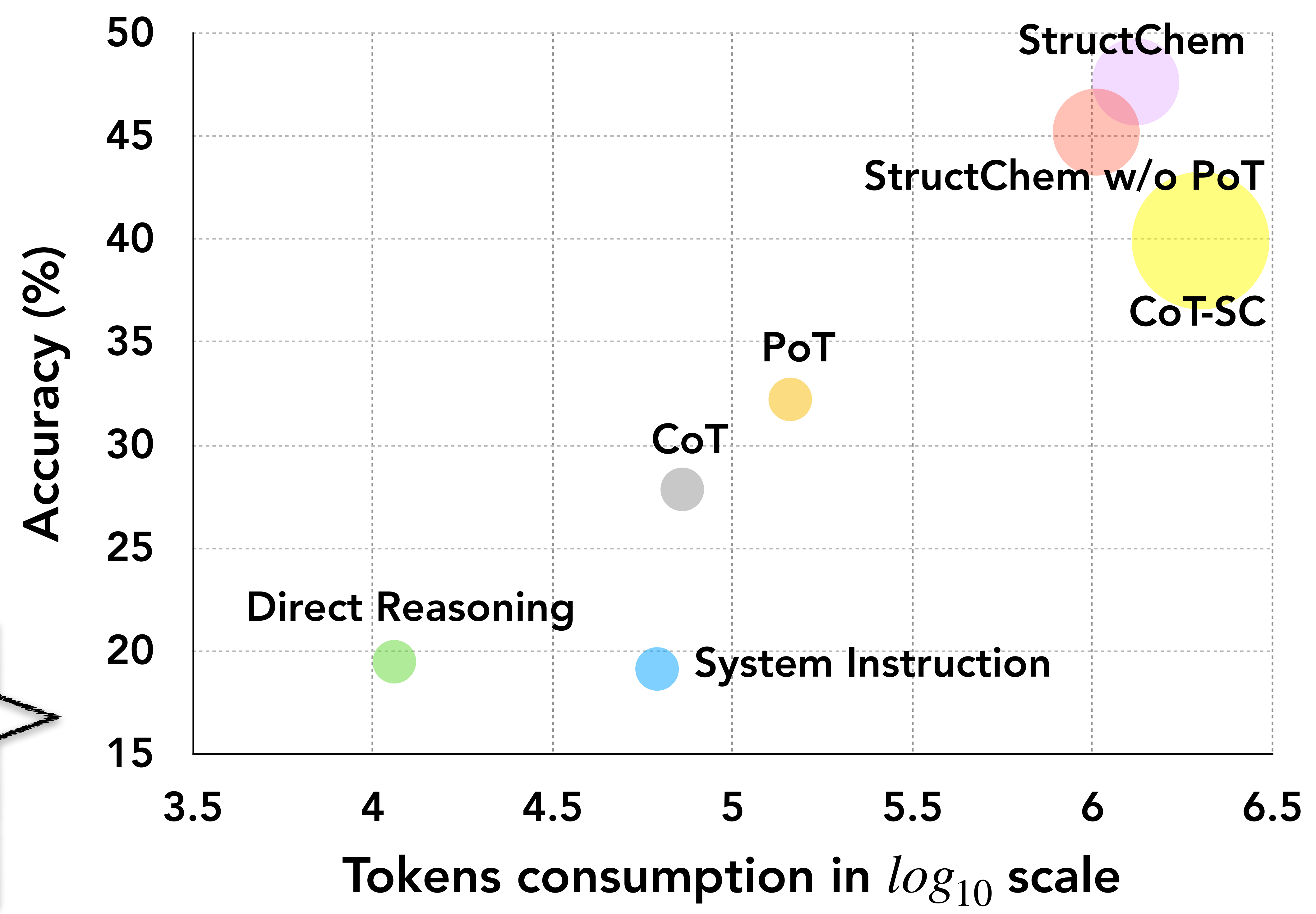
Error analysis

Reduced error rates overall, specifically for reasoning errors.



Slightly higher irrelevant knowledge: StructChem provides a sider range of knowledge

Cost-effectiveness analysis



Size of dot is proportional to number of API calls.

Improvement brought by StructChem does not depend on token consumptions.

Discussion

Whether/How Can LLMs Self-Correct Reasoning?

LARGE LANGUAGE MODELS CANNOT SELF-CORRECT REASONING YET

Jie Huang[1,2*] Xinyun Chen[1*] Swaroop Mishra[1] Huaixiu Steven Zheng[1] Adams Wei Yu[1] Xinying Song[1] Denny Zhou[1]
[1]Google DeepMind [2]University of Illinois at Urbana-Champaign
jeffhj@illinois.edu, {xinyunchen, dennyzhou}@google.com

SELFCHECK: USING LLMS TO ZERO-SHOT CHECK THEIR OWN STEP-BY-STEP REASONING

Ning Miao[1*] Yee Whye Teh[1] Tom Rainforth[1]

Step-by-step decomposition

Use external feedback

Use memory mechanisms

Summary

(1) study the various ways the precise scientific reasoning can fail with frontier LLMs

(2) the importance of symbolic/structured reasoning

(3) acknowledging that this type of precise reasoning remains to be a major challenge, suggesting a more fundamental research into this direction

Thank you!

StructChem:

Structured Chemistry Reasoning with Large Language Models

https://arxiv.org/abs/2311.09656

https://ozyyshr.github.io/StructChem

Siru Ouyang

Zhuosheng Zhang

Bing Yan

Xuan Liu

Yejin Choi

Jiawei Han

Lianhui Qin