Sample Design Engineering: An Empirical Study on Designing Better Fine-Tuning Samples for Information Extraction with LLMs

Biyang Guo[1†], He Wang[1†], Wenyilin Xiao[1†], Hong Chen[2†], Zhuxin Lee[3], Songqiao Han[4,1*], Hailiang Huang[1,5*]

[1]AI Lab, SIME, Shanghai University of Finance and Economics [2]Ant Group [3]Guangdong Yunxi Technology [4]Key Laboratory of Interdisciplinary Research of Computation and Economics, Ministry of Education, China [5]Shanghai University of Finance and Economics-Ant Group Joint Laboratory of Frontier Financial Intelligence

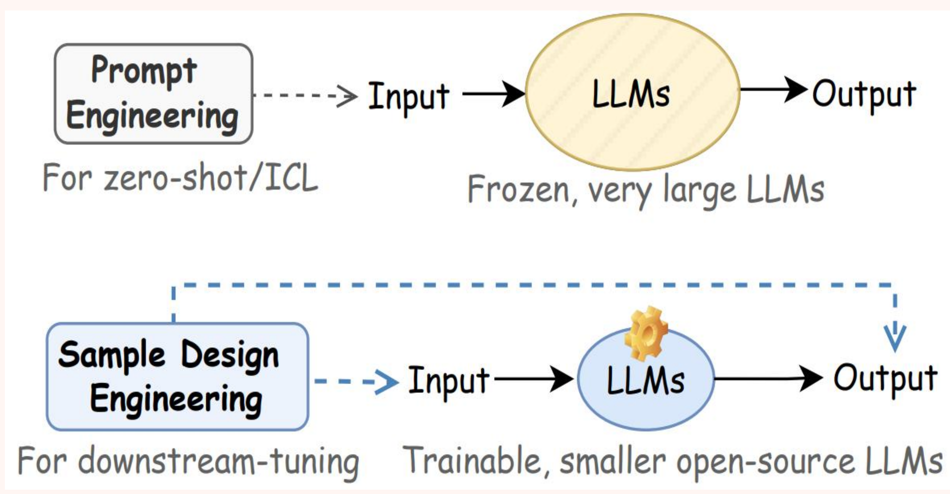
Motivation

Why do we need Sample Design Engineering(SDE) compared to Prompt Engineering(PE)

- The efficacy of PE relies on the size of LLMs.
- Companies prefer customizing smaller, open-source models for their needs due to high costs and privacy risks of large models.

Information extraction (IE) require our focus

- IE tasks is highly valuable in a wide range of industrial scenarios.
- There's a fundamental challenge arises from the discrepancy between the unstructured nature of the LLMs' generative paradigm and the requirement for structured output.

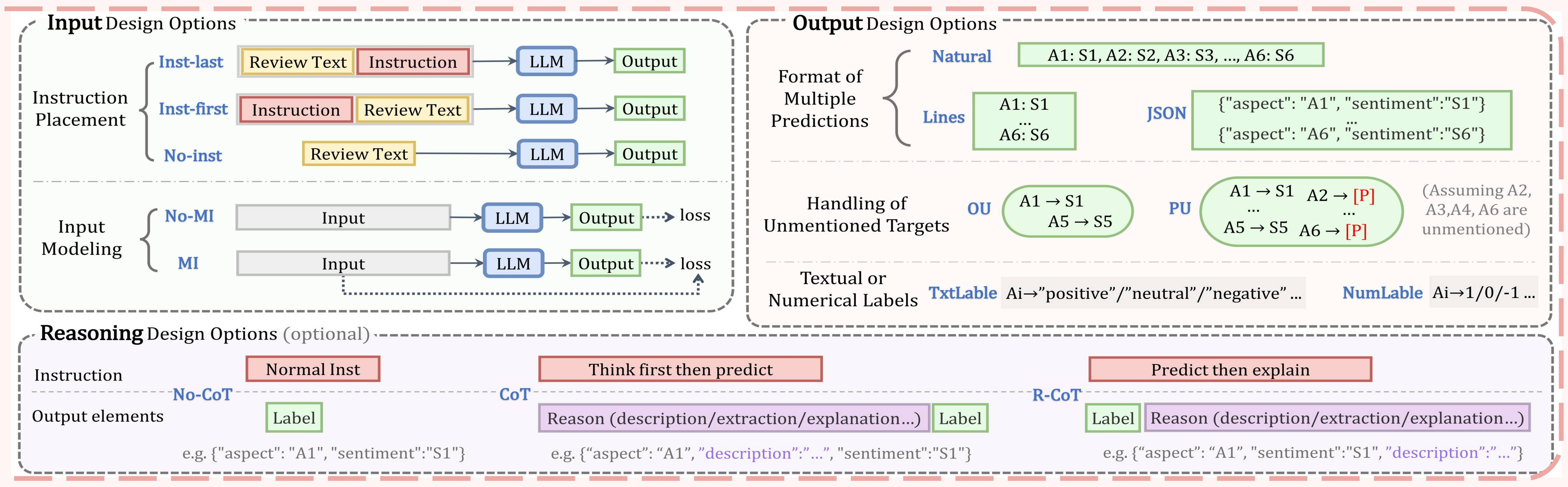


SDE Options

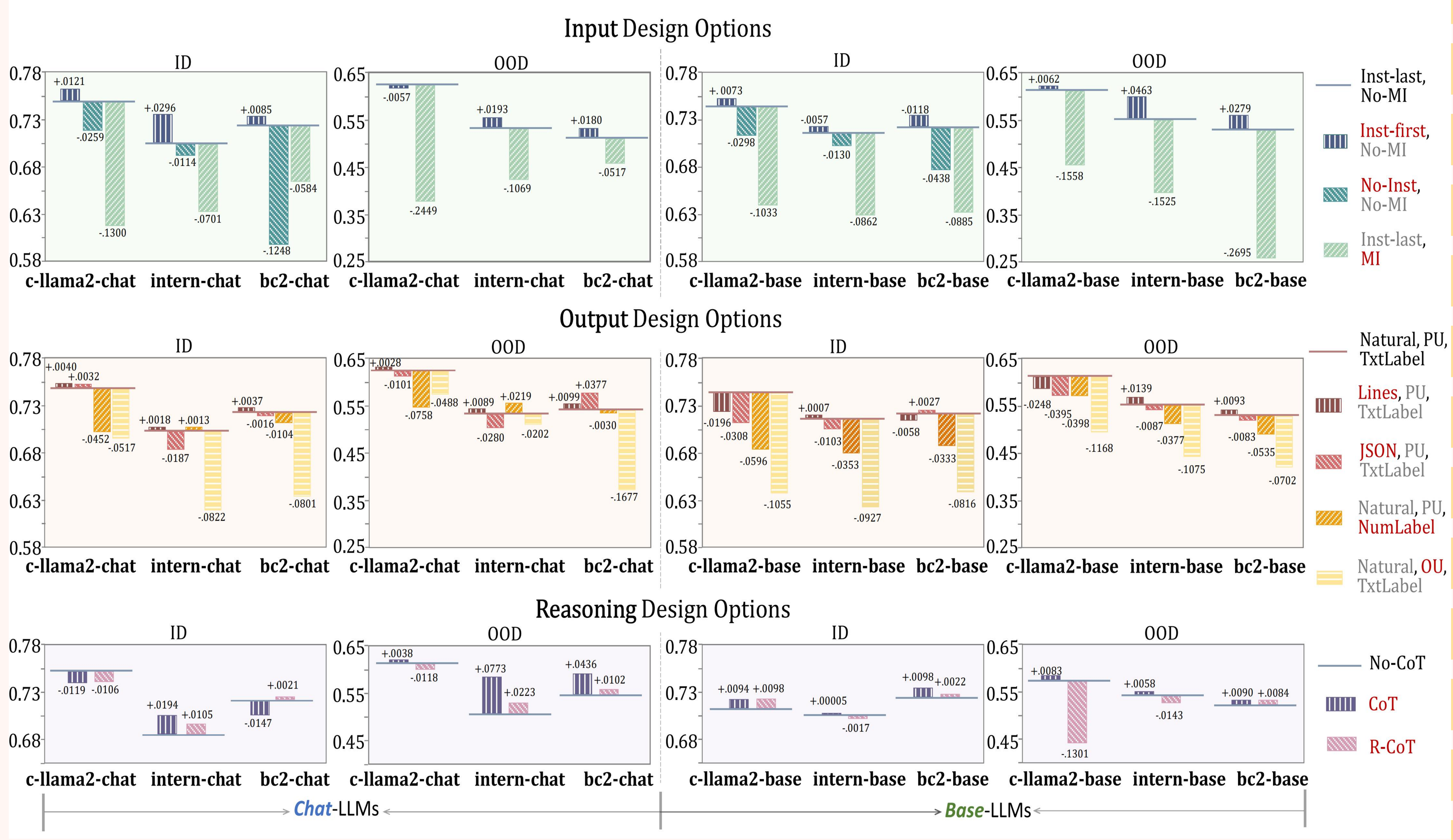
Input Design Options
(1)Instruction Placement
(2)Input Modeling

Output Design Options
(1)Multiple Predictions Formatting
(2)Handling of Unmentioned Targets
(3)Textual or numerical labels

Reasoning Design Options

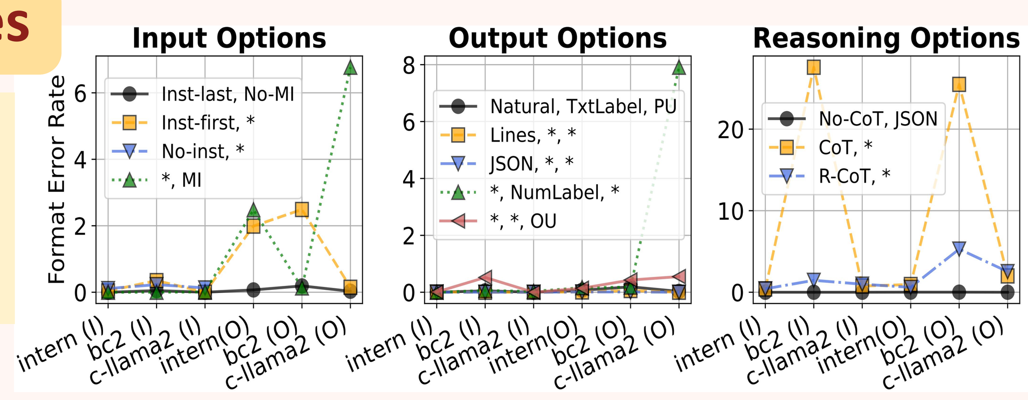


Experiments I: Evaluating The Impact of Each SDE Option



Evaluate from 2 perspectives

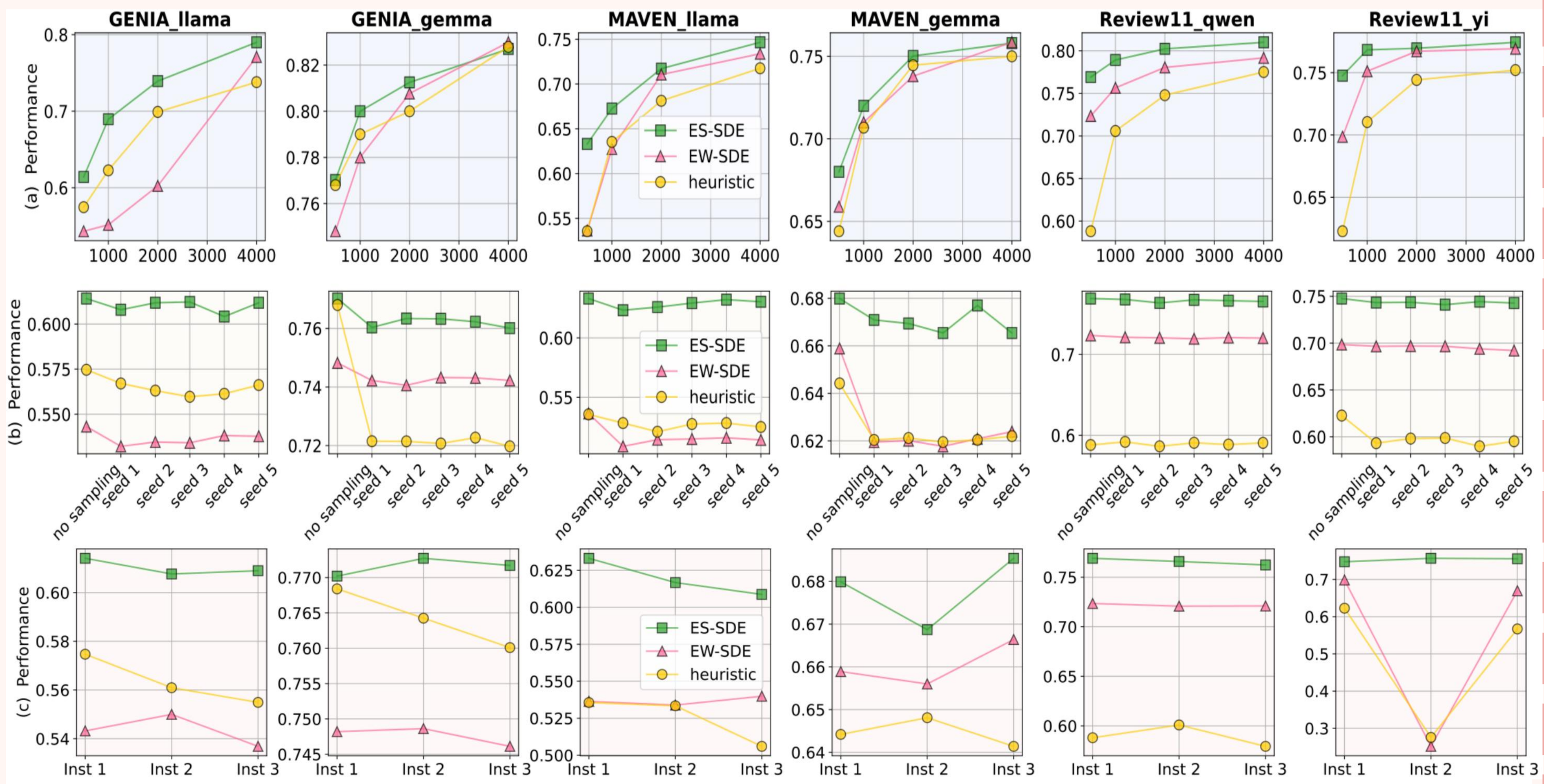
- Sentiment analysis performance
- Format adherence



KEY CONCLUSIONS:

- Better to place ***instruction first***
- ***Lines*** format is reliable
- ***Placeholders for Unmentioned*** targets is better than Omit them
- Subtle impact of ***CoT*** on ID, while significant on OOD

Experiments II: A Robust Integrated SDE Strategy



(a) Performance of different sample design strategies with increasing training sizes: 500, 1000, 2000 and 4000. (b) Robustness on decoding sampling randomness, training size = 500. (c) Robustness on instruction content variation, training size = 500.

Empirically Strong SDE Strategy (ES-SDE)

Use the well performing options in Experiments I:
Inst-first, ***No-MI*** input designs
Lines, ***PU(Placeholders for Unmentioned)***, ***TxtLabel*** output designs

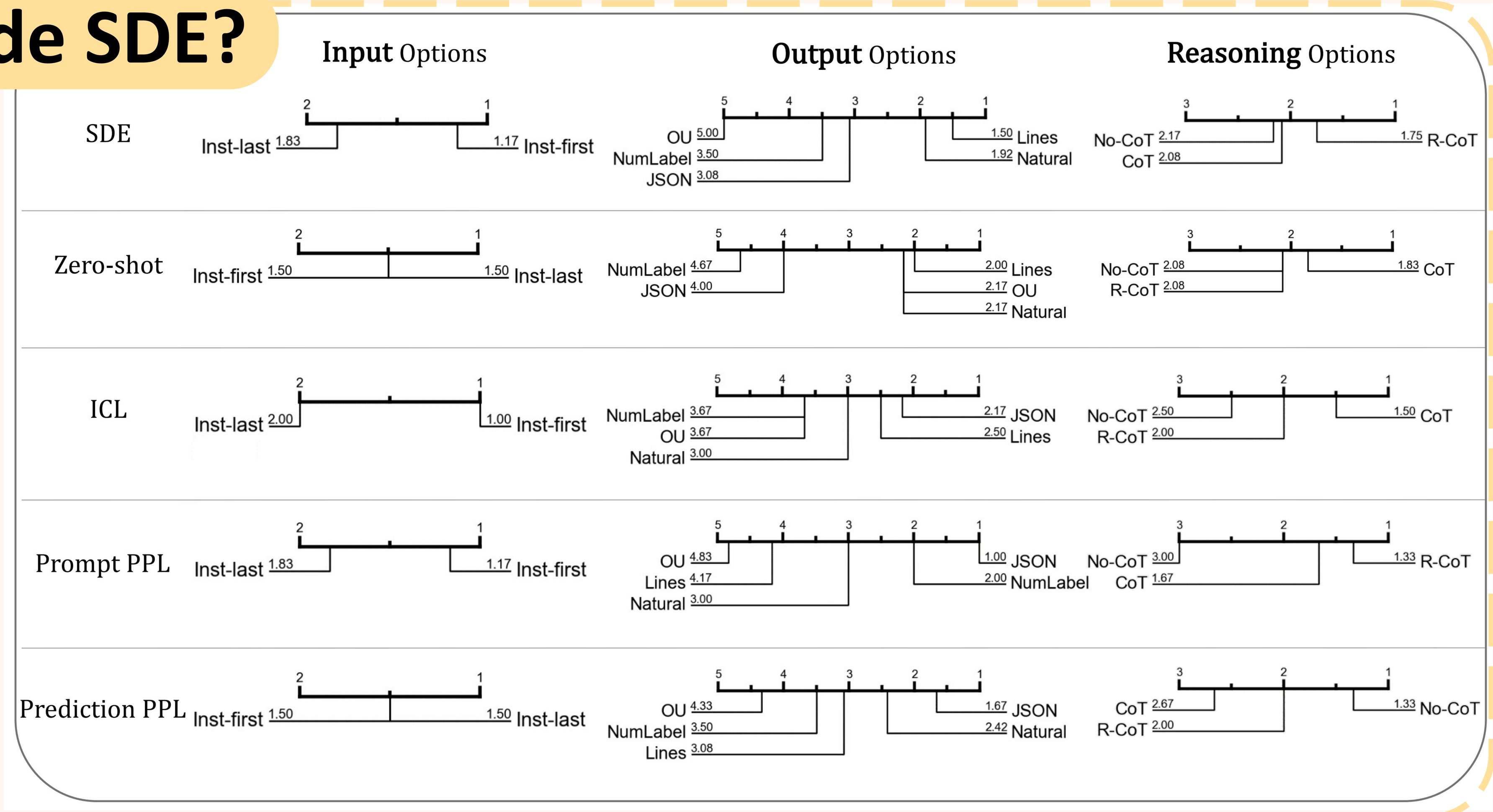
KEY CONCLUSIONS:

- ES-SDE maintains advantages across tasks and training sizes
- ES-SDE is stable on decoding randomness
- ES-SDE is robust to instruction variation

Can PE guide SDE?

PE patterns do not always align with SDE effectiveness.

For example, the performance of *OU (Omit Unmentioned)*, and the comparison between *Natural* and *Lines*.



Conclusion

- We propose a new data-centric perspective for enhancing the performance of LLMs in downstream tasks.
- We provide a comprehensive summary and systematic evaluation of various sample design strategies.
- Our experiments prove its necessity and effectiveness, and open up avenues for deeper SDE exploration in future studies.