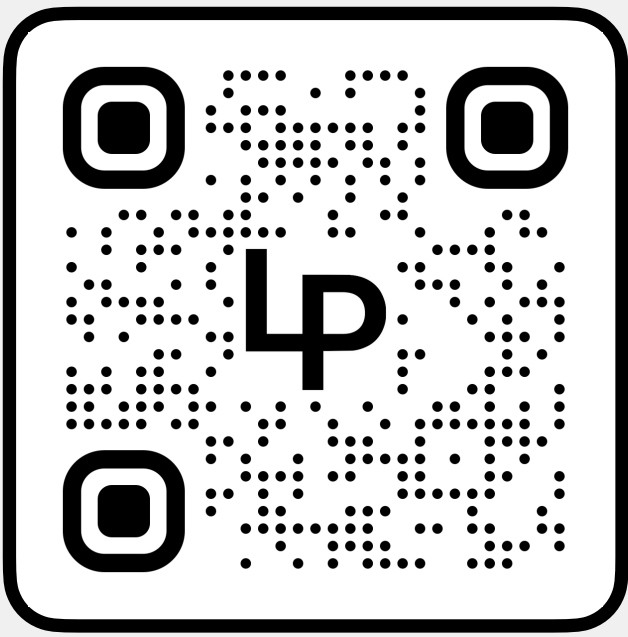


LeetPrompt

Leveraging Collective Human Intelligence to Study Large Language Models



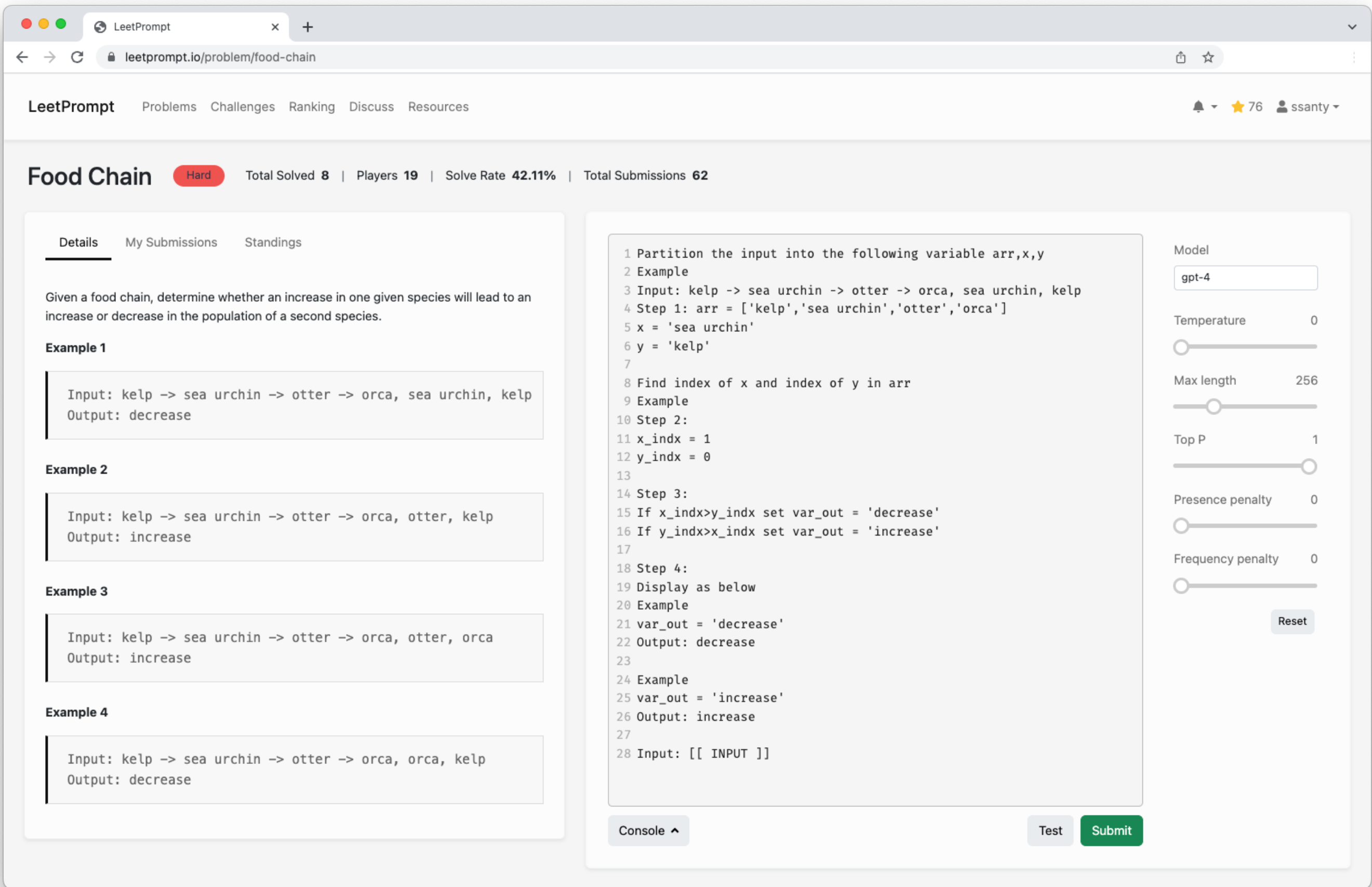
Sebastin Santy, Ayana Bharadwaj, Sahith Dambekodi, Alex Albert, Cathy Yuan, Ranjay Krishna

Writing good instructions aka “**prompt engineering**” is getting increasingly valuable:

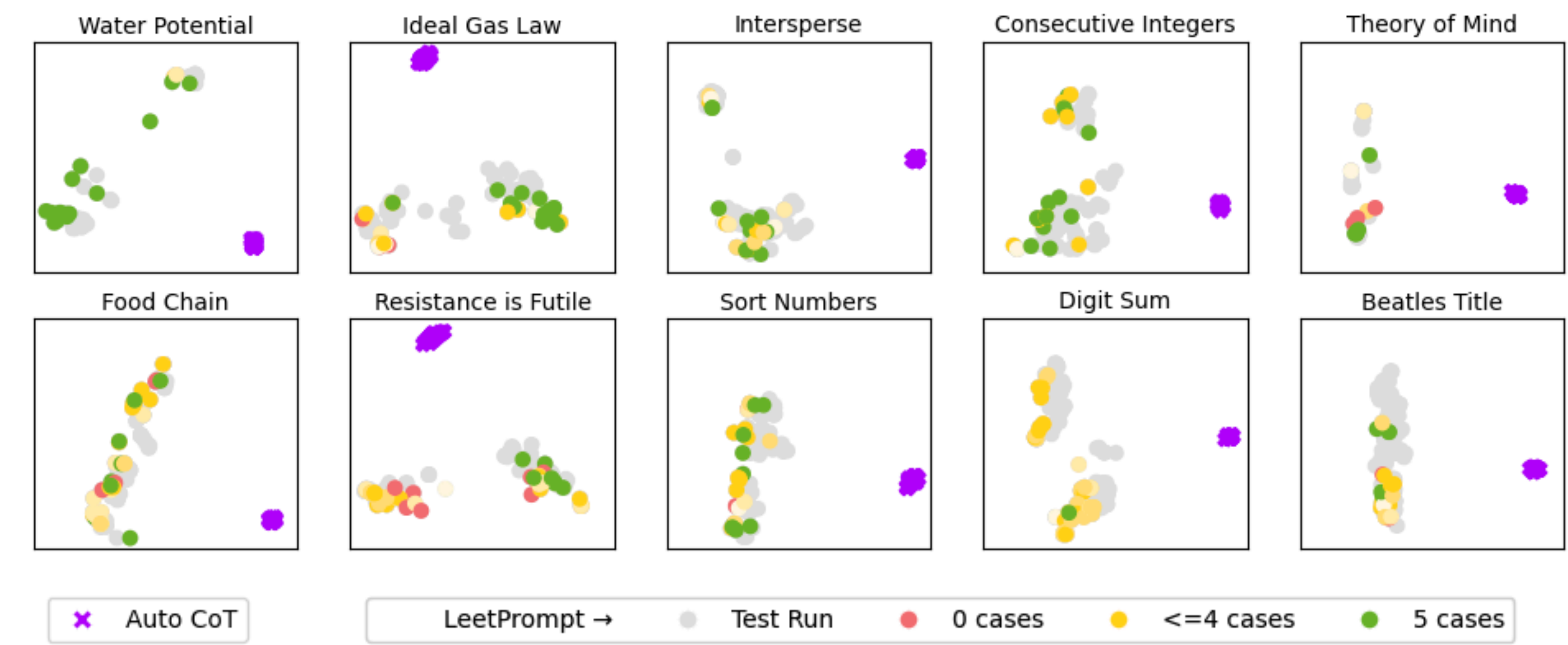
- 1. For users, to achieve their personal tasks.
- 2. For model developers, to train better models.

On the other hand, prompts are making it difficult to evaluate models:

*If the model cannot perform a task, is it because it **cannot understand the instruction?** or because it **does not have the ability to solve that task?***



User Study. 20 participants. 10 questions.



- 1. Unsolvable questions got solved by people.
- 2. Multiple common strategies to solve a problem. Some not identified by GPT-4.

Dual Incentives.

Lay Person		Solve Puzzles & having fun Learn Prompt Engineering Hangout with LM community
Researcher		Source Instruction Data Benchmark Language Models Analyze complex model behaviors

For more

leetcode.com/play/
 leetcode.com/paper/