The Effect of Distraction in Large Language Models in Coding Tasks

Authors: Ratchanon Samana, Thanapod Saithong, Pongsaphak Samanee
Advisor: Dr. Kamonphop Srisopha
Data Science Research Center, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand

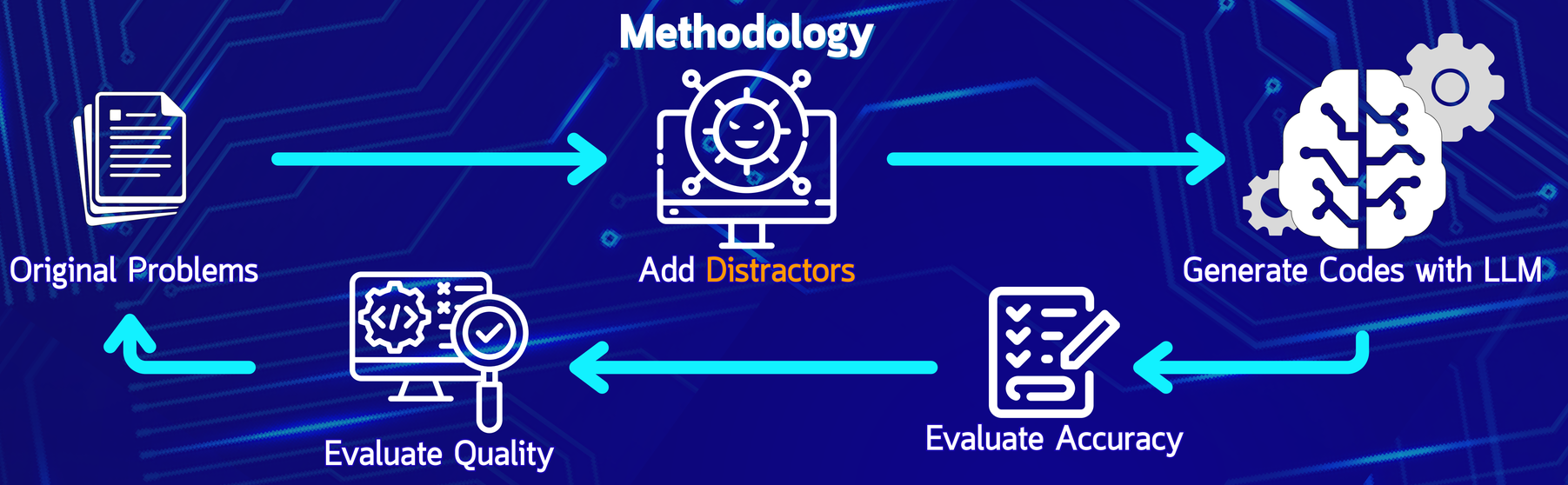
Abstract

This study investigates the robustness of large language models (LLMs) in coding tasks under various types of distractions. Using a dataset of 9 fundamental coding problems, a controlled experimental framework is used to compare five LLMs under several prompt perturbation conditions. Model performance is evaluated using automated test cases, while code characteristics are examined through cyclomatic complexity (CC) and source lines of code (SLOC).
The result indicate that most models remain robust to irrelevant contextual distractions, maintaining high accuracy. However, prompts containing conflicting information significantly reduce model accuracy and tend to produce more complex and longer code.

Introduction

In recent years, Large Language Models have rapidly become an important tool in education. Many users increasingly rely on these systems to solve problems and analyze questions, resulting in models being exposed to a wider variety of tasks with increasing complexity.
This research proposes an approach that introduces distractors into problem statements, such as logically conflicting conditions or misleading information, in order to evaluate the accuracy and comprehension ability of language models. The experimental results can provide insights for AI developers to improve model robustness and enhance their ability to handle the complexity of natural language more effectively.

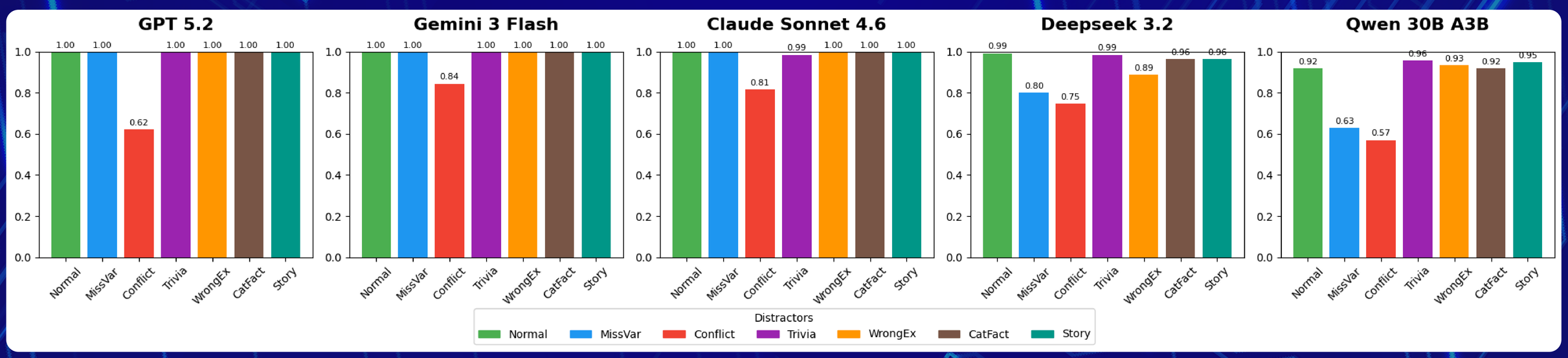
Methodology



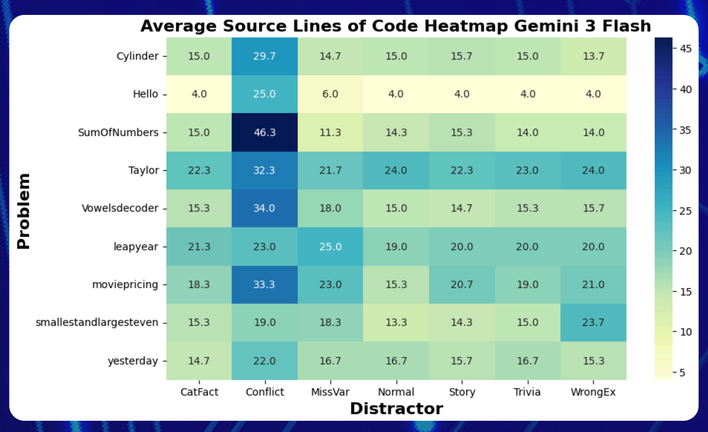
Distractor Types

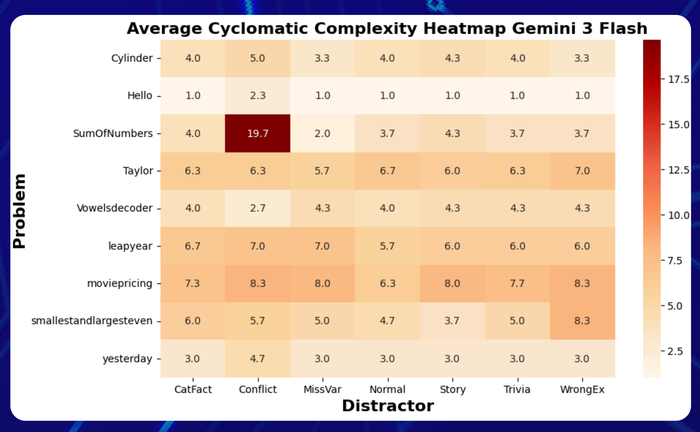
- **MissVar(Misleading Variable)** : Unused variables
- **Conflict** : Contradictory information or requirements.
- **Trivia** : Random irrelevant general knowledge
- **WrongEx(Wrong Example)** : Incorrect input-output examples
- **Cat Fact** : Random factual statements about cats
- **Story** : The problem is rewritten in a narrative form.

Results



Average Accuracy Comparison of Five LLMs under Various Distractors





Average Source Lines of Code and Cyclomatic Complexity of the Gemini 3 Flash Model

The results indicate that most models maintain high accuracy under the Normal condition and remain robust against irrelevant distractors such as Random Trivia, CatFact, and Storytelling. In this research, Average Accuracy refers to the average percentage of generated code that correctly pass test cases and produce the expected outputs. This metric is used to examine how different types of distractors influence the correctness of generated code. However, the Conflict distractor significantly reduces model accuracy across all evaluated models. In addition to performance degradation, the Conflict condition also leads to higher cyclomatic complexity (CC) and longer source lines of code (SLOC) in the generated code. Meanwhile, MissVar and WrongEx introduce moderate effects on accuracy and code characteristics.

Conclusion

This research show the effects of distractors in large language models in coding tasks. The results show that most models maintain high accuracy under normal and irrelevant distractor conditions. However, conflicting information significantly reduces accuracy and increases code complexity and length. These research highlight that while LLMs can ignore irrelevant information, they remain vulnerable to conflicting instructions in prompts.

References

[1] Jaechul Roh, Varun Gandhi, Shivani Anilkumar, Arin Garg (2025). CHAIN-OF-CODE COLLAPSE: Reasoning Failures in LLMs via Adversarial Prompting in Code Generation. arXiv:2506.06971
[2] Yanbo Wang, Zixiang Xu, Yue Huang, Chujie Gao, Siyuan Wu, Jiayi Ye, Pin-Yu Chen, Xiuying Chen, Xiangliang Zhang(2025)Adaptive Distraction: Probing LLM Contextual Robustness with Automated Tree Search. arXiv:2502.01609