

NLP Team Contribution

Juan Macias:

- Developed the python scripts for data preprocessing, analysis and result storage.
- Ran, annotated and evaluated Multistep Arithmetic operation tasks.
- Performed analysis for DeepSeek's thinking outputs against final outputs.
- Performed and contributed to literature review and selection of tasks from BigBenchHard.

Bickston Laenger:

- Ran Riddles, Translation Error, and Dyck Language Tasks
- Wrote Riddles, Translation Error, and Dyck Language parts in the paper
- Wrote introduction of the paper
- Created Explanation table in the paper
- Hosted and organized several group meetings

Alejandro Danies-Lopez

- Ran and annotated Web of Lies and Penguins, wrote corresponding analysis sections of final report
- Compiled results from all tests and models, calculated accuracies, generated table and plots
- Contributed to literature review and problem selection along with other members

Aditya Raghavan

- Ran Object Counting and Shuffle Seven Objects Tasks
- Annotated Objecting Counting and Formal Fallacy Tasks

- Manually prompted over one hundred SAT and arithmetic questions on each LLM
- Wrote the methodology of the paper
- Wrote Object Counting and Formal Fallacy parts in the paper

Araceli Rodríguez:

- Ran and annotated Shuffle objects and Adjective order tasks
- Wrote Shuffle objects and Adjective order parts in the report
- Wrote literature review in the paper
-