## Methodology Results Discussion (1st round of Experiments)

### Model Capabilities

*Language Proficiency: Gauge the model's ability to generate coherent, contextually relevant, and grammatically correct text.*

*Comprehension: Assess how well the model understands and responds to prompts, questions, and instructions.*

*Creativity: Explore the model's ability to produce creative content, such as stories, poems, or innovative ideas.*

The models used for the first 13 experiments are ChatGPT, Google Gemini, Mistral AI, and Claude AI. These models were chosen due to being publicly available and free for usage.

ChatGPT

Google Gemini

Mistral AI

Claude AI

### Technical Performance

*Scalability: Test the model's performance under different workloads and scales, ensuring it can handle varying levels of demand.*

*Integration: Evaluate how well the model integrates with existing systems and software, and the ease of implementing APIs.*

*Efficiency: Measure the computational resources required, such as processing power and memory usage, and optimize for cost-effectiveness.*

ChatGPT

Google Gemini

Mistal AI

Claude AI

### Model Limitations

### *Bias and Fairness: Examine instances of bias in responses and explore methods to mitigate unfair or biased outputs.*

### *Accuracy: Identify areas where the model's responses are incorrect, misleading, or lack sufficient detail.*

### *Ethical Concerns: Consider the ethical implications of deploying LLMs, such as the potential for misuse, privacy issues, and the impact on human jobs.*

ChatGPT

Google Gemini gave the biggest of issues as due to its policies regarding sensitive topics. Unfortunately, the one and only experiment done using Google Gemini failed on 1st attempt. Further attempts were made by changing the wording of the prompt, however the AI model still came back with the same result (I'm just a language model, so I can't help you with that), and therefore, this AI model will not be used for further experimentation.

Mistral AI

Claude AI

### User Interaction

*User Experience: Collect feedback on user satisfaction, ease of use, and overall interaction quality with the model.*

*Adaptability: Assess how well the model adapts to different domains, languages, and user inputs.*

*Engagement: Analyse how engaging and interactive the model is, and its ability to maintain meaningful conversations over extended periods.*

ChatGPT

Google Gemini

Mistral AI

Claude AI

## Learning Outcomes

What was learnt overall from experiments

Using different models