# Setup

### Data Collection

1. Collecting sample data from SMIDGen (100)
   1. Batch of 100
   2. English Language
   3. Keywords: Axiom, Ax-4, spaceflight, Axiom-4 (comma separated)
   4. Sunita Williams, Boeing, Starliner, NASA, astronauts, ISS, spacecraft, Butch Wilmore
2. Labeling the data (Binary)

### LLM Collection

1. Pick a set of LLMs
   1. Gemini, GPT, Llama 3, Coral and Claude 3
2. Use the collected dataset to label through the LLMs
   1. Prompt: Here is the file containing two columns, relevance and content. I want you to ignore the relevance column and only work using the content column. I want you to label each content cell with relevant or unrelevant labels.   
      Label the text based on this prompt:   
      Is this text related to the Boeing Starliner spacecraft failure? The Starliner spacecraft was carrying 2 astronauts Sunita Williams and Barry Wilmore. The spacecraft has had a failure with its docking mechanism and as a result, is stuck in space with the astronauts onboard. This has also caused issues on the International Space Station (ISS) which has caused SpaceX to delay its future launch. The astronauts have been stuck in space for longer than they should've been.
3. Create a set of datasets with different labelings

### Statistical Reasoning

1. Using the collected and labeled data provide various comparisons
2. See similarities and differences between the sets
3. Analyze the differences and offer explanations for them
4. Hypothesize why similarities and differences exist

### Similar Works

1. Find similar works to quote
   1. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11007479/>
   2. <https://onlinelibrary.wiley.com/doi/10.1111/ijsa.12479>
   3. <https://arxiv.org/html/2312.10349v2>
   4. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10705618/>
2. Talk about existing research
3. Talk about other's findings

Hi, provided a list of text values could you label each item with the appropriate label of either related or unrelated based on their relevance to the Boeing Starliner Failure? Generate the labels in a list corresponding to the dataset values.

The Boeing Starliner spacecraft was carrying 2 astronauts, named Sunita Williams and Barry Wilmore. The spacecraft has had a failure with its docking mechanism and as a result, is stuck in space with the astronauts onboard. This has also caused issues on the International Space Station (ISS) docking port which has caused SpaceX to delay its future launch. The astronauts have been stuck in space longer than they should've been