## Problem Framing and Approach

I framed this as a standard classification problem. Considering the available labeled data and its statements, I tried SVM, Random Forest, and Logistic Regression. While approaches like neural networks and LLM-based classification were considered, the limited dataset, independent nature of data, and lightness of the model made me choose logistic regression.

## Model Development

The implementation followed these key steps:

* Used logistic regression as the base model: accuracy of ~80%
* Used LLM to generate additional test data using prompt engineering
* Retrained the logistic regression model with augmented dataset, and the accuracy increased to 85%

The complete code and implementation details are available in my GitHub repository (<https://github.com/JayaManasa/nlp-intent-detection/tree/main>).

## Results Analysis & Improvements

The model shows promising results with the available data, giving reasonable predictions with up to 85% accuracy. Logistic Regression appeared strong as the data came out to be linear, where the length of the text and intent are linearly related, such as COD and Cash on Delivery fell into the same label consistently. This showed linear relationship. This is where SVM and Random Forest didn't perform well.

Potential improvements:

* Each label's data has to be consistent, varied, and in higher numbers
* Using other text tokenization and vectorization methods could possibly give higher accuracy
* Experimenting with different classification models like Neural Networks, LLM classification model
* Using larger LLMs (like Sonnet 3.5) for generating more consistent and varied training data

These suggested improvements could boost model performance while maintaining simplicity and interpretability.