Eval

### Compilation success rate

| **Models** | **Success rate** | |
| --- | --- | --- |
|  | **GPT 4o** | **Gemini** |
| **Direct prompting** | 100 | 100 |
| **Few shot** | 100 | 100 |
| **CoT** | 100 | 71.4 |
| **Neurosymbolic API synthesis** | 85.7 | 100 |

### Code quality

<https://docs.google.com/spreadsheets/d/1tDEe0I8JBt8dDW6eNF89yBv7_KXgnnDE1mIL55Z0Yzs/edit?usp=sharing>

### Test cases

NA

### Generalization

1. **Create a python script that uses imbalanced-learn to do a binary classification model on Breast Cancer Wisconsin (Diagnostic) Dataset (WDBC), where your major class to minor class ratio is 99:1. You should use techniques like SMOTE to balance the dataset. Use a general model like XGboost. Do an analysis of metrics (like accuracy, precision, recall, f1 score, auc score and confusion matrix) before and after using imbalanced-learn techniques.**
2. **Write a script in python that uses tensorflow and keras library. Your task is to work on Semantic Segmentation using Oxford Pets Dataset and a U-Net Model. The semantic segmentation task would be to differentiate between background and the pet. Also calculate pixel-wise classification accuracy, model loss and IoU.**

Runtime evaluation analysis:

Task 1: