

AI CODE REVIEW ASSISTANT

Project Proposal

Team members:

Yogev Cohen

Dudi Ohayon

Romy Somkin

Problem Statement

Motivation

- Code reviews are critical for ensuring code quality but are time-consuming and subjective
- A Code Review Assistant automates feedback, detects issues, and suggests improvements.
 - Useful in professional teams, open-source contributions, and educational platforms.
- **Example:** The AI suggests improvements, style fixes, or potential bugs during development.
- **Application:** Helps developers improve code quality early, even without senior review.

NLP Task

Text generation (code → review comment)

Challenges:

- Understanding the logic and intent behind code
- Supporting multiple languages and coding styles
- Generating context-aware and useful feedback, not generic or repetitive

Input: Code snippet (e.g., Python, Java).

Output: Natural language comment with suggestions or feedback

Training And Test Data

Public Dataset

Source: <https://huggingface.co/datasets/fasterinnerlooper/codereviewer>

Entity Labels: patch,y,**oldf**,idx,id,**messege**,proj,**lang**

Labeled data: Batches of code snippets, associated review comments and including programming language.

Syntetic data

Use LLM to generate additional code review examples with similar structure .

Example

Input

```
def getSum(a,  
           b):  
    return a + b
```

Output

"Consider adding a docstring
and renaming to
'calculate_sum' for clarity."

Evaluation

Models

- Codet5-base
- CodeBERT
- Mistral-7B

Evaluation Metrics

- BLEU: Measures n-gram overlap with reference comments
- ROUGE: Measures recall of overlapping units like n-grams and sequences
- Mistral-7B (Evaluation only) Used as a large language model evaluator to assess the quality and helpfulness of the generated code review comments.

Evaluation Method

- Dataset split into train/test/validation

Baseline

- Static Template Comments-Always return a generic comment, regardless of input
- Rule-Based Commenting Use simple if-else logic based on string patterns.
- First-Line Classifier: Search for suspicious patterns and respond with canned message.
- Length-Based Heuristic
Make a judgment based on function length.