EgoErrorVQA

QA-pairs Generation

Error annotations + Action annotations + Task + Step sequence

LLM

Reviewed by 3 human reviewers → New QA-pairs is added by 1 human annotator

QA-pairs for Open-end QA task

Open-end VQA Evaluation

Green agent

Procedure, Task_id, Video_path, Questions

White agent

Watch videos

Answers

Green agent

Score Answer-GT similarity with LLMs

Judge 1 + Judge 2 = 10 Avg

Human annotators

Reliability analysis

Evaluation result

Multiple-choice VQA Evaluation

Green agent

Procedure, Task_id, Video_path, Questions, Error types

White agent

Watch videos

Error type

Green agent

Evaluation metrics

Accuracy | Precision

Recall | F1 Score

Quantitative evaluation result