Using reports created by the IILAP

Table of Contents

Introduction

When a teacher clicks the Evaluate button, the system automatically generates two reports that summarize student performance: a summary report and a per-participant report. In addition to creating these files, the interface also displays the file paths showing where each report has been saved. The following sections explain the contents and purpose of each report.

Figure 1 Example output showing report paths

Evaluation Status

☑ Evaluation completed.

Evaluation Output

Reports created
☑ Summary report saved: reports/summary_participant_overview_no_names.xlsx
☑ Per-participant report saved: reports/per_participant_report_no_names.xlsx

File location example:
reports/summary_participant_overview_no_names.xlsx
- The report is generated within the IILAP project folder, under the reports directory.

Summary report

The Summary Report provides an overview of participant performance across the entire group. It includes aggregated data without giving information on individual participants. This report is useful for analyzing overall trends, identifying common strengths and weaknesses, and comparing performance across different tasks or criteria.

Colum information

The summary contains 38 sheets with different information.

name

- Contains the list of names of all students that took part in the survey

familiarity

- Contains the counts, percentages, summary statistics and distribution graph of participants familiarity with digital tools

information_sources

- Contains counts and graph of the information sources the participants reported using

Ai_experience

- Contains the counts, percentages, summary statistics and distribution graph of participants prior ai experience

Ai_frequency

- Contains the counts, percentages, summary statistics and distribution graph of participants frequency of ai use

Ai_satisfaction

- Contains the counts, percentages, summary statistics and distribution graph of participants prior ai satisfaction

ai_reason

- Contains the list of all the reasons participants gave for their prior ai experience

source_misattribution

Evaluation metrics for the classification task related to source misattribution.
Columns:

Total TP: Total true positives (correctly identified misattributions).

Total FP: Total false positives (incorrectly flagged as misattributed).

Total FN: Total false negatives (missed misattributions).

Precision: TP / (TP + FP).

Recall: TP / (TP + FN).

F1: Harmonic mean of precision and recall.

correct_info_red

Evaluation metrics for the classification task related to correct_info_red.
Columns:

Total TP: Total true positives (correctly identified instances).

Total FP: Total false positives (correct information incorrectly flagged).

Total FN: Total false negatives (missed cases).

Precision: TP / (TP + FP).
Recall: TP / (TP + FN).
F1: Harmonic mean of precision and recall.

wrong_info_green

Evaluation metrics for the classification task related to wrong_info_green.
Columns:
> **Total TP**: Total true positives (correctly identified errors).
> **Total FP**: Total false positives (incorrectly flagged as errors).
> **Total FN**: Total false negatives (missed errors).
> **Precision**: TP / (TP + FP).
> **Recall**: TP / (TP + FN).
> **F1**: Harmonic mean of precision and recall.

wrong_credibility_score

Evaluation metrics for the classification task related to wrong_credibility_score.
Columns:
> **Total TP**: Total true positives (correctly identified scoring errors).
> **Total FP**: Total false positives (incorrectly flagged scores).
> **Total FN**: Total false negatives (missed scoring errors).
> **Precision**: TP / (TP + FP).
> **Recall**: TP / (TP + FN).
> **F1**: Harmonic mean of precision and recall.

misinfo_in_unhighlighted

Evaluation metrics for the classification task related to misinfo_in_unhighlighted.
Columns:
> Total TP: Total true positives (correctly identified missinformation).
> Total FP: Total false positives (incorrectly flagged errors).
> Total FN: Total false negatives (missed errors).
> Precision: TP / (TP + FP).
> Recall: TP / (TP + FN).
> F1: Harmonic mean of precision and recall.

AI was easy to use, AI gave clear answers, Highlights improved understanding, AI helped critical thinking, I noticed source or content errors, I checked the sources, I asked follow-up questions, positive experience, Useful for academic use

- Contains the counts, percentages, summary statistics, and a distribution graph of participant responses.

Improvements, liked_disliked, thought_process,

- List of full participants responses

participant_scores

- Contains summary statistics, a boxplot, and a distribution of participants' scores for error detection (a composite of all error detection types).

Chat Submission

- Contains the counts, percentages, summary statistics, and a distribution graph of number of asked questions

Source Click

- Contains the counts, percentages, summary statistics, and a distribution graph of number of clicked sources

Button Click

- Contains the counts, percentages, summary statistics, and a distribution graph of number of clicked buttons

total_interaction_time

- Contains the counts, percentages, summary statistics, and a distribution graph of total interaction with the system

quiz_percentage_correct

- Contains the counts, percentages, summary statistics, and a distribution graph of participants quiz scores

score_1

- Contains summary statistics, a boxplot, and a distribution of participants' scores1
- Score 1 is equal to the coefficient the teacher put under "Number of clicks 1+ " if the user clicked on at least one source, else it's 0

score_2

- Contains summary statistics, a boxplot, and a distribution of participants' scores2
- Score 2 is equal to the coefficient the teacher put under "Total number of questions 8+ " if the user asked at least 8 questions, else it's 0

score_3

- Contains summary statistics, a boxplot, and a distribution of participants' scores3
- Score 3 is equal to the coefficient the teacher put under " More than 10 s spent on source " if the user on average spent at least 10s per source checked, else it's 0

score_4

- Contains summary statistics, a boxplot, and a distribution of participants' scores4
- Score 4 is equal to the coefficient the teacher put under " More than 20 s spent on source " if the user on average spent at least 20s per asked question, else it's 0

score_5

- Contains summary statistics, a boxplot, and a distribution of participants' scores5
- Score 5 is participants quiz answer multiplied with the coefficient under "Quiz score"

score_6

- Contains summary statistics, a boxplot, and a distribution of participants' scores6
- Score 6 is participants error identification score multiplied with the coefficient under "Correctly identified errors"

score

- Contains summary statistics, a boxplot, and a distribution of participants' total scores
- Total score is sum of all of the previous scores

Per participant report

The Per Participant Report provides information on individual participants. This report is useful for analyzing specific student's performance, identifying strengths and weaknesses, and comparing performance across different tasks.

Per participants report contains one sheet for each participant. It contains all of the data mentioned above as well as full logs of chatbot interactions and individual per error type performance