Supplemental Materials

This document describes the structure and content of the supplemental materials.

/pipeline

This folder contains the code for the VILA pipeline.

- /config_files
  - This folder contains the three config files for the three stages of the pipeline. Each config file contains the full prompt to the LLM in that stage. The prompt summary of these prompts are used in Figure 3 of the paper.
- /generated_code_session_1
  - This folder is where the R code from Stage I is stored. There are 108 (9 contexts x 12 chart types) files.
- /generated_dataset_session_1
  - This folder is where the datasets from Stage I are stored. There are 108 (9 contexts x 12 chart types) files.
- /generated_vis_session_1
  - This folder is where the visualizations from Stage I are stored. There are 108 (9 contexts x 12 chart types) files.
- /generated_text_session_2
  - This folder is where the textual components from Stage II are stored. There are 1,404 (108 x 13 tasks) files.
- /generated_text_session_3
  - This folder is where the textual components from Stage III are stored. There are 1,404 (108 x 13 tasks) files.
- stage1_generator.py, stage2_composer.py, stage3_checker.py
  - Each of the three files contains a Python function that uses the prompts from the config files and makes a call to OpenAI's API.
- main_stage_1.py, main_stage_2.py, main_stage_3.py
  - The three files use the Python functions from stage1_generator.py, stage2_composer.py, and stage3_checker.py. Running these three files sequentially will create the R code files, the datasets, and the visualizations in Stage I, the textual components in Stage II, and the checked textual components in Stage III.
- requirements.txt
  - This file contains the libraries needed to run the pipeline.

- utils.py
  - This file contains two string formatting functions to extract the desired outputs from the LLM's response.

/candidate_bank

This folder contains the candidate bank of 1,404 items

- /108_visualization_component
  - This folder contains 108 visualizations (9 contexts x 12 chart types) of the candidate bank.
- /1404_textual_component
  - This folder contains 1,404 textual components (108 x 13 tasks) of the candidate bank.
  - In each file, "part 1" is the question stem, "part 2" is the options, and "part 3" is the LLM-selected correct answer.

/evaluation_I_and_II

This folder contains the data from the two-phase evaluation of the candidate bank.

- selection_of_representative_sample.py, 156_names.txt
  - The Python file selects the sample of the 156 representative items from the candidate bank, and the txt file contains the names of the 156 items.
- expert_11.csv, two_authors.csv
  - These two csv files contain the 11 experts' and two author-evaluators' ratings for the 156 representative items.
- phase1.Rmd
  - This file contains the data processing code for the 11 experts' and two author-evaluators' ratings for the 156 representative items. The code identifies items with disagreements for authors to discuss and computes the suggested ratings for "Relevance" and "Clarity" for each (chart type, task) combinations.
  - Results of this file can be directly seen in phase1.nb.html.
- phase_1_author_discussion_on_answerable.csv
  - This csv file contains the representative items where at least one of the 11 experts and two author-evaluators rated "No" for "Answerable", everyone's comments (if any), and the two author-evaluators' tentative ratings after reading all the comments and ratings and their reasons for the tentative ratings.

- phase_1_author_discussion_on_correctness.csv
  - This csv file contains the representative items where at least one of the 11 experts and two author-evaluators rated "No" or "More than one option could be considered correct" for "Correctness", everyone's comments (if any), and the two author-evaluators' tentative ratings after reading all the comments and ratings and their reasons for the tentative ratings.
- rulebook_suggested_ratings.csv, rulebook_combinations_prone_to_error.pdf
  - The csv file contains the suggested ratings for "Relevance" and "Clarity" for each (chart type, task) combinations, which are part of the rulebook.
  - The pdf contains combinations the author-evaluators identified to be prone to certain types of errors to help them make decisions for Phase II, which are part of the rulebook.
- phase_2_data.csv
  - The csv file contains the two author-evaluators' ratings of the entire candidate bank of 1,404 items, as well as ratings of the versions of these items before they were checked by Stage III: Checker.
- phase2.Rmd
  - This file processes the data in phase_2_data.csv. It selects the final bank of items based on the inclusion criteria. It computes the number of items in each (chart type, task) combination in the final bank. It generates the numbers in the confusion matrix for the performance of Stage III: Checker, which are used in Figure 6 of the paper.
  - Results of this file can be directly seen in phase2.nb.html
- error_classification.csv
  - The csv file contains our classification of errors that all the items that did not get selected into the final bank have.

/final_bank

- final_bank.csv
  - This csv file contains the names of 1,103 items in the final bank. The visualization and textual components of these items can be found in /candidate_bank
- final_bank_count_by_charttypetaskcombo.csv
  - This csv file contains the number of items in each (chart type, task) combination in the final bank. This is used for Figure 5 of the paper.

/vila_vlat

- reclassification_of_vlat_tasks.js
  - This file contains the 53 VLAT items and our reclassification of their tasks using our revised task taxonomy.
- /6_items_treemap_hierarchical
  - This folder contains the 6 visualization and textual components for the special item (tree map, identify the hierarchical structure)
- /22_items_for_scatterplot_bubblechart_anomalies_clusters
  - This folder contains the 22 visualization and textual components for the special items (scatterplot, identify anomalies), (scatterplot, identify clusters), (bubble chart, identify anomalies), (bubble chart, identify clusters)
- data_cleaning_and_anonymization.Rmd
  - This file contains code that cleans and anonymizes the data of the convergent validity study. The un-anonymized data that contains participants' identifiable information is not provided in the supplemental materials.
  - Results of this file can be directly seen in data_cleaning_and_anonymization.nb.html
- /data
  - This folder contains two cleaned and anonymized files of data of participants' responses to VILA-VLAT and VLAT.
- convergent_validity.Rmd
  - This file contains the pre-registered analysis code that computes the correlation coefficient for convergent validity.
  - Results of this file can be directly seen in convergent_validity.nb.html

Revisions of Burns Taxonomy.pdf

This document contains our revisions of the visualization task taxonomy in Burns et al.

Design Guidelines for Creating Multiple-Choice Visualizations Items.pdf

This document contains the operational design guidelines that we collected and organized from the prompts of the pipeline and the error classification from the evaluation.