Supplemental Materials Details

Amit Kumar Das[*] Klaus Mueller[†]

We have uploaded all supplemental materials in a file named Supplemental Materials Details.zip. Upon unzipping this file, you will find 4 different folders:

1. 1_MisVisFix Final Dashboard
2. 2_Expert User Study
3. 3_Original VLAT
4. 4_All Misleading Graphs

We will describe the contents of each folder, including file types, required software for viewing/execution, and how the material relates to the paper. Below is a detailed description of each folder and its contents:

1 1_MisVisFix Final Dashboard

This folder contains all the source code for the MisVisFix interactive dashboard system described in our paper.

1.1 Contents

- Python source code files for the complete dashboard implementation
- HTML/CSS/JavaScript files for the frontend interface
- Configuration files for the system setup
- LLM prompting templates incorporated within the code
- Requirements.txt file listing all dependencies

1.2 File Types

- Python (.py) files
- HTML (.html) files
- CSS (.css) files
- JavaScript (.js) files
- Requirements (.txt) file
- Configuration files

[*] Department of Computer Science, Stony Brook University, Stony Brook, NY 11794, USA. Email: amitkumar.das@stonybrook.edu

[†] Department of Computer Science, Stony Brook University, Stony Brook, NY 11794, USA. Email: mueller@cs.stonybrook.edu

1.3 Software Requirements

To run the MisVisFix dashboard, you will need:

1. **Python 3.8+** with the following packages (all specified in requirements.txt):
   - Django
   - Pandas
   - NumPy
   - Matplotlib
   - OpenAI (for GPT API access)
   - Anthropic (for Claude API access)
   - Various other dependencies listed in requirements.txt
2. Modern web browser (Chrome, Firefox, Safari, or Edge)
3. Internet connection for API access to LLMs

1.4 Installation and Running Instructions

1. Go to the folder ("1_MisVisFix Final Dashboard")
2. Create a virtual environment:

   ```
   python -m venv myenv
   ```

3. Now activate environment:
   - For Windows:

     ```
     myenv\Scripts\activate
     ```

   - For Mac:

     ```
     source myenv/bin/activate
     ```

4. Install additional required packages:

   ```
   pip install -r requirements.txt
   ```

5. Now install dependency:

   ```
   pip install anthropic
   pip install djangorestframework
   ```

6. Set proper file permissions:

   ```
   chmod 775
   chmod 664 db.sqlite3
   ```

7. Now go to the project folder:

```
cd my_chart_app
```

8. Run project:

```
python manage.py runserver
```

Note: Upload misleading graph and test, Install other packages if necessary. The system automatically manages thread IDs, but if you encounter problems, check the browser console for any JavaScript errors and ensure proper API key configuration.

1.5 Relation to the Paper

This code implements the complete MisVisFix system described in Section 3 of the paper. It includes all components of the dashboard shown in Figure 1, including:

- The visualization upload and display (Panel A)
- Corrected visualizations generation by Claude and GPT (Panels B and C)
- Dataset upload functionality (Panel D)
- Issue detection panels for both models (Panels E and F)
- Interactive chat interface (Panel G)

The code also implements the issue localization technique described in Section 3.4 and the learning mechanism shown in Figure 5.

2 2_Expert User Study

This folder contains materials related to the expert evaluation described in Section 4.4 of the paper, including both pre-study and post-study data and survey instruments.

2.1 1_PreStudy

2.1.1 Contents

- **MisVisFix_Pre-Study.csv**: Contains the raw data collected from expert users during the pre-study phase.
- **PreStudy_Survey.pdf**: A sample view of the Qualtrics survey used in the pre-study. (Note: University names and identifying information have been removed to maintain anonymity for review.)

2.1.2 File Types

- CSV (.csv) file
- PDF (.pdf) file

2.1.3 Software Needed

- Any spreadsheet software (e.g., Microsoft Excel, Google Sheets) for viewing the CSV file
- Any PDF viewer for viewing the survey file

2.1.4 Relation to the Paper

The pre-study data corresponds to the expert evaluation described in Section 4.1.3, capturing information about the five visualization experts (3 male, 2 female) with a mean of 13.4 years of experience in data visualization who assessed the system. These experts had diverse backgrounds: four held PhDs in visualization or related fields, and one had a Master's degree. This file contains the demographic and background information collected before the experts used the MisVisFix dashboard.

2.2 2_PostStudy

2.2.1 Contents

- **MisVisFix_Post-Study.csv**: Contains the raw data collected from expert users during the post-study phase.
- **PostStudy_Survey.pdf**: A sample view of the Qualtrics survey used in the post-study. (Note: University names and identifying information have been removed to maintain anonymity for review.)

2.2.2 File Types

- CSV (.csv) file
- PDF (.pdf) file

2.2.3 Software Needed

- Any spreadsheet software (e.g., Microsoft Excel, Google Sheets) for viewing the CSV file
- Any PDF viewer for viewing the survey file

2.2.4 Relation to the Paper

The post-study data was collected after the experts used our MisVisFix dashboard and contains their detailed feedback and comments. This data supports the quantitative results reported in Section 4.4 of the paper, including the ratings for Detection Accuracy (8.5/10), Usefulness of Suggested Improvements (8.0/10), and Likelihood to Use in Professional Work (8.0/10). The file includes experts' qualitative comments that provided insights into the system's strengths and limitations as discussed in the paper.

3 3_Original VLAT

This folder contains materials related to our experiments with the original VLAT charts:

3.1 Charts-of-Thought Prompt

3.1.1 Images

- **Contents:** PNG files of all original VLAT visualizations
- **File type:** PNG
- **Software needed:** Any standard image viewer

3.1.2 Results

This folder contains four subfolders, one for each LLM tested:

3.1.2.1 1_Claude 3.7_Final Results with Explanation

- **Contents:**
  - Python notebooks: `VLAT_run.ipynb`, `Claude_Results.ipynb`
  - CSV files (e.g., `VLAT_1740626876.csv`) containing the model responses, time taken, and correctness
  - Corresponding DOCX files (e.g., `VLAT_1740626876.docx`) containing the detailed explanations from the LLM
- **File types:** IPYNB, CSV, DOCX
- **Software needed:** Jupyter Notebook, Microsoft Word or any word processor that can open DOCX files

3.1.2.2 2_GPT 4.5_Final Results with Explanation

- **Contents:**
  - Python notebooks: `VLAT_run.ipynb`, `GPT_Results.ipynb`
  - CSV files (e.g., `GPT_VLAT_1740709963.csv`) containing the model responses, time taken, and correctness
  - Corresponding DOCX files (e.g., `GPT_VLAT_1740709963.docx`) containing the detailed explanations from the LLM
- **File types:** IPYNB, CSV, DOCX
- **Software needed:** Jupyter Notebook, Microsoft Word or any word processor that can open DOCX files

3.1.2.3 3_GEMNI 2.0 Pro_Final Results with Explanation

- **Contents:**
  - Python notebooks: `VLAT_run.ipynb`, `Gemni_Results.ipynb`
  - CSV files (e.g., `Gemini_VLAT_1741123345.csv`) containing the model responses, time taken, and correctness
  - Corresponding DOCX files (e.g., `Gemini_VLAT_1741123345.docx`) containing the detailed explanations from the LLM
- **File types:** IPYNB, CSV, DOCX
- **Software needed:** Jupyter Notebook, Microsoft Word or any word processor that can open DOCX files

3.1.2.4 4_GPT 4.o_Final Results with Explanation

- **Contents:**
  - Python notebooks: `VLAT_run.ipynb`, `GPT_Results.ipynb`
  - CSV files (e.g., `GPT_VLAT_1743188849.csv`) containing the model responses, time taken, and correctness
  - Corresponding DOCX files (e.g., `GPT_VLAT_1743188849.docx`) containing the detailed explanations from the LLM
- **File types:** IPYNB, CSV, DOCX
- **Software needed:** Jupyter Notebook, Microsoft Word or any word processor that can open DOCX files

3.1.3 Additional Files

- `VLAT Questions Metadata.csv`: CSV file containing metadata about the VLAT questions
- `VLAT Questions.csv`: CSV file containing the VLAT questions and answer options

4 4_All Misleading Graphs

This folder contains the comprehensive benchmark dataset used to evaluate MisVisFix's detection capabilities across all 74 misleading visualization categories identified by Lo et al. [1].

4.1 Organization Structure

The benchmark dataset is organized into 11 category-specific folders based on issue types:

- **1_Truncated Axis**: Charts with manipulated axis ranges
- **2_3D Effects**: Visualizations using misleading three-dimensional representations
- **3_Dual Axis**: Charts with problematic dual y-axis implementations
- **4_Missing Title**: Visualizations lacking essential titles
- **5_Missing Axis Labels**: Charts with missing or inadequate axis labeling
- **6_Inappropriate Color Use**: Visualizations with color-related problems
- **7_Inconsistent Scale**: Charts with inappropriate scaling choices
- **8_Selective Data Presentation**: Visualizations with cherry-picked or filtered data
- **9_Data Manipulation**: Charts with altered or manipulated data values
- **10_Misrepresentation of Findings**: Visualizations that misrepresent conclusions or relationships
- **11_Others**: Additional miscellaneous misleading techniques

4.2 Contents per Folder

Each category folder contains the following:

4.2.1 Visualization Images

- **Contents**: PNG/JPG files of misleading visualizations numbered sequentially (1.png, 2.png, 3.png, etc.)
- **File types**: PNG, JPG
- **Software needed**: Any standard image viewer

4.2.2 Evaluation Results

- **Contents**: Excel file named after the category (e.g., Truncated_Axis.xlsx)
- **File type**: XLSX
- **Software needed**: Microsoft Excel, Google Sheets, or compatible spreadsheet software

4.3 Excel File Structure

Each Excel file contains the following columns:

- **Row**: Sequential number corresponding to image filename
- **major_claude**: Major issues detected by Claude 3.7
- **minor_claude**: Minor issues detected by Claude 3.7
- **potential_claude**: Potential issues detected by Claude 3.7
- **major_gpt**: Major issues detected by GPT-4.5
- **minor_gpt**: Minor issues detected by GPT-4.5
- **potential_gpt**: Potential issues detected by GPT-4.5
- **Benchmark_dataset**: Ground truth labels from original benchmark

4.4 Evaluation Methodology

Our evaluation approach treats each issue instance as a separate test case. When a single visualization contains multiple issues, we evaluate each issue independently through multiple system runs. For cases where our system detects related but differently named issues (e.g., detecting "color overuse" for a benchmark labeled "color misuse"), we mark these as correctly solved when the detected issue's explanation addresses the original problem.

4.5 Alternative Access Methods

- **Online Gallery**: All benchmark results are accessible through our online gallery at http://167.71.222.168/
- **Dashboard Integration**: Users can examine original misleading images, generated corrections, and detailed explanations for each test case

4.6 Relation to the Paper

This benchmark dataset supports the quantitative evaluation reported in Section 4.2 of the paper, including:

- Detection performance metrics (Table 2)
- Issue categorization accuracy (Table 4)
- Performance comparison between GPT-4.5 and Claude 3.7 (Table 5)
- Detailed analysis by issue category (Table 3)

The comprehensive documentation enables researchers to verify our detection accuracy across all 74 issue categories and reproduce our evaluation methodology.

5 Overall Software Requirements

To fully utilize these supplemental materials, the following software is required:

1. **Python 3+** with packages specified in requirements.txt
2. Web browser (Chrome, Firefox, Safari, or Edge)
3. Spreadsheet software (e.g., Microsoft Excel, Google Sheets) for viewing CSV files

4. PDF viewer for survey documents

Note: The file names are self-explanatory, indicating their specific purposes within the research project. These materials allow for complete replication of the experiments and analysis described in the paper.

6 Public Access and Gallery

MisVisFix is publicly accessible at http://167.71.222.168/. The hosted system includes all features described in the paper: issue detection, explanation, correction, and interactive chat functionality.

The online gallery contains curated examples of misleading visualizations from real-world sources. Each example shows:

- Original misleading visualization
- Issues detected by both GPT and Claude
- Corrected versions generated by each model
- Detailed explanations of why specific elements are problematic

Users can upload their visualizations for analysis or explore the gallery to understand standard misleading techniques. The gallery serves as an educational resource for visualization literacy and demonstrates the system's capabilities across diverse chart types.

7 System Behavior with Non-Misleading Visualizations

During development, we observed interesting behavior when users upload good (non-misleading) visualizations. While the system occasionally flags good charts as having minor issues, the correction module consistently generates alternative good visualizations rather than misleading ones. The LLMs tend to recreate charts using different visual approaches while maintaining data integrity. This suggests the models have learned to default toward best practices rather than generate deceptive visualizations, which aligns with responsible AI principles.

8 Performance and Latency Considerations

System response time varies from 2-3 minutes, depending on visualization complexity and network conditions. The latency stems from several factors: sequential processing through both Claude and GPT models, comprehensive analysis across all 74 issue categories, code generation for corrected visualizations, and API response times. Complex visualizations with multiple detected issues require additional processing time. Future optimizations could include parallel LLM processing, caching for common issues, and progressive result display.

9 Detailed Expert Evaluation Protocol

Expert evaluation followed a structured three-phase protocol: (1) Introduction phase where experts viewed a demonstration video and received system overview, (2) Hands-on testing phase where experts uploaded both provided misleading visualizations and their own examples, exploring detection accuracy and correction quality, and (3) Interview phase with structured questions about system utility, accuracy assessment, and professional applicability. Each session lasted 30-60 minutes. Experts tested various chart types, used interactive chat features, and evaluated both GPT and Claude outputs for comparison.

References

[1] L. Y.-H. Lo, A. Gupta, K. Shigyo, A. Wu, E. Bertini, and H. Qu, “Misinformed by visualization: What do we learn from misinformative visualizations?” *Computer Graphics Forum*, vol. 41, 2022.