# EXPERIMENTAL PLATFORM FOR RED TEAMING MULTIMODAL LARGE LANGUAGE MODELS IN ROAD SAFETY

**Client:** Matt Albrecht, matthew.albretch@uwa.edu.au (Lead Contact), Muhammad Hussain muhammad.hussain@uwa.edu.au **TEAM:** Gnaneshwar Reddy Bana (23832048), Kanishk Kanishk (23959947), Pedro Wang (23870387), Sarath Pathari (22941307), Yuanfu Cao (23633858), Yuxin Gu (23743373)

## AIM & BACKGROUND

The Multimodal LLM Image Analysis Platform aims to develop an AI-driven system for analyzing images from driving simulations. By integrating image processing with Large Language Models (LLMs), the platform will assist researchers and engineers in analyzing driving scenarios, including simulated conditions.

## PROJECT OBJECTIVES:

- **Advanced AI-Powered Image Analysis:** Analyze driving simulation images, with and without distortions, using LLMs to interpret various driving scenarios.
- **Simulated Driving Conditions:** Implement image processing techniques to recreate diverse driving conditions for enhanced analysis.
- **User-Friendly Interface:** Develop a Streamlit interface for seamless image upload, analysis, and result display.

## PROJECT DELIVERABLES:

The Multimodal LLM Image Analysis Platform will include the following key features:

- **Web Application:** A Streamlit-based platform for road safety image analysis, supporting batch processing, structured csv output, various image distortion effects (e.g., None, Blur, Brightness, Contrast, Sharpness, Color, Rain, Overlay, Warp), and integrated AI analysis using the Google Gemini model.
- **User Interaction & Results Display:** A user-friendly interface with an "Analyze" button, supporting secure API key input, clear result presentation, and CSV export of structured analysis data.
- **Documentation & Deployment:** A comprehensive user guide and deployment on Streamlit Community Cloud.

## PROJECT PLAN:

The following work is planned into three Sprints with 6 user stories for each:

1. **CLI Tool Development (3 weeks, 28th AUG – 15th SEP)**
2. **GUI Development (3 weeks, 16th Sep – 6th Oct) and**
3. **Finalize GUI and Complete Image Distortion Feature (2 weeks, 7th OCT – 18th OCT).**

The corresponding activities have been created in Trello (Please check the appendix for Trello link.) Also, responsibilities and deadlines for each activity are defined clearly. When it comes to the beginning of each sprint, we will breakdown the user stories into subtasks.

**Time Cost:**

Estimated time cost:

- Requirement Gathering: 100 hours
- Design & Prototyping: 100 hours
- Development: 300 hours
- Testing: 160 hours
- Client Meetings: 40 hours
- Documentation: 60 hours
- Project Management: 40 hours

## PROTOTYPE:

https://www.figma.com/design/XaY1Gj4GGDYnQT3a7rHhKs/Multimodal-LLM-Road-Safety-Platform?node-id=0-1&t=m9u1DMpXtYAX23DD-1

## TECHNICAL REQUIREMENTS:

The following table outlines the key technical specifications and components necessary for the development and implementation of the Multimodal LLM Image Analysis Platform:

| Requirement | Description |
|---|---|
| **Platform** | Python-based environment |
| **AI Model** | Google Gemini 1.5 Flash |
| **Data Source** | DriveSim dataset: https://github.com/sreeramsa/DriveSim/tree/main/Eval_LLM_Drive |
| **Image Manipulation Capabilities** | Blur, brightness, contrast, sharpness, color, rain effect |
| **User Interface** | Streamlit |
| **Image Processing** | Pillow (PIL), SciPy |
| **API Integration** | Google Generative AI API |

**Reasons for choice of the above requirements:**

A Python-based environment offers versatility and extensive libraries for machine learning, image processing, and UI development, making it ideal for our project. The Gemini 1.5 Flash model is cutting-edge, seamlessly combining vision and language, crucial for road safety, and is free. Streamlit and Pillow are effective and well-documented, streamlining development without sacrificing performance.

## STRIDE ANALYSIS:

Key risks and mitigations:

1. **Spoofing**: Unauthorized API key use.

*Mitigation*: Secure keys in environment variables.

2. **Tampering**: Data alteration.
   *Mitigation*: Use HTTPS.
3. **Repudiation**: Users denying submissions.
   *Mitigation*: Log metadata.
4. **Information Disclosure**: Unintended exposure of sensitive data.

*Mitigation*: Use HTTPS and restrict access to sensitive data.

5. **DoS**: Platform overwhelmed.
   *Mitigation*: Streamlit's DoS protection.
6. **Elevation of Privilege**: Unauthorized feature access.
   *Mitigation*: Input validation.

## APPENDIX

### Project Gantt Chart:

Timeline of development, task completions, client meetings, and deadlines are detailed in the Gantt Chart.

https://github.com/AlteredOracle/CITS5206/blob/main/Project%20Documents/Deliverables/Deliverable%201/Ganttchart.png

### Project Management:

Tasks and progress are tracked using a Trello Board.
https://trello.com/b/LpLXq0zV/capstone-project

### Client Interaction:

All client meetings, discussions, and feedback are documented in the Client Interaction folder.

https://github.com/AlteredOracle/CITS5206/tree/main/Project%20Documents/Client%20Interactions

### Meeting Minutes:

The meeting minutes are documented in our repo:
https://github.com/AlteredOracle/CITS5206/tree/main/Project%20Documents/Minutes%20Of%20Meeting

### Project Repository:

All files and code are maintained in the Project Repository.
https://github.com/AlteredOracle/CITS5206