Multi-Modal Assistant

# **Project Overview**

This document tracks the learning journey, challenges faced, and solutions implemented while developing the Vision AI Assistant application.

# **Running the Application**

To run the application:

***streamlit run app.py***

# **Key Features**

* Image upload and display
* AI-powered image analysis
* Simple web interface
* Google's Gemini AI Integration
* Image Processing

# **Issues Faced and Solutions**

1. Layout Issues

- Problem: Image and text not aligned

- Fix: Used `st.columns()` for side-by-side layout

2. Image Upload

- Problem: Only certain image formats worked

- Fix: Added support for JPG, JPEG, PNG

3. API Key Security

- Problem: Hardcoded API key

- Note: For production, use environment variables

# **Development Notes**

* Keep the virtual environment activated while working on the project
* Always pull the latest changes before starting work
* Write tests for new features
* Document API endpoints and their expected request/response formats

# **Future Improvements**

* Add support for batch image processing
* Implement user authentication
* Add more customization options for the AI model
* Include error handling for API rate limits
* Add a feature to save analysis results