

Project - High Level Design on Multimodal Automotive Creator

Course Name: Gen-AI

Institution Name: Medicaps University – Datagami Skill Based Course

Student Name(s) & Enrollment Number(s):

Sr no	Student Name	Enrolment Number
1.	Ayush Yadav	EN22CS301255
2.	Ayush Bangar	EN22CS301245
3.	Bhumika Gupta	EN22CS301277
4.	Bhupendra Yadav	EN22CS301282
5.	Charu Arya	EN22CS301289
6.	Rituraj Kumar Verma	EN23CS3L1015

Group Name : D3G12

Project Number : GAI-36

Industry Mentor Name : Mr. Suraj Nayak

University Mentor Name : Prof. Vineeta Rathore

Academic Year : 2022-26

Table of Contents

1. Introduction.

The **Multimodal Automotive Creator** is an AI-powered platform designed to simplify and accelerate vehicle design and development. It integrates multiple input modes such as text, images, voice, and technical data to generate complete automotive concepts. Using advanced technologies like generative AI, computer vision, and simulation models, the system can create exterior designs, interior layouts, and performance estimations. This project aims to reduce development time, enhance innovation, and enable intelligent, data-driven automotive creation.

1.1. **Scope of the document.**

This document defines the overall scope and boundaries of the Multimodal Automotive Creator system. It explains the objectives, system architecture, and core functionalities of the platform. The document describes how multiple input modalities such as text, images, voice, and technical data are processed within the system. It outlines the major system components, including the user interface, backend modules, AI models, and simulation engines.

Functional requirements such as vehicle design generation, 3D modeling, and performance simulation are clearly specified. Non-functional requirements, including scalability, performance, security, and usability, are also addressed. The complete workflow, from user input to the final automotive concept output, is clearly defined.

Target users such as designers, engineers, researchers, and students are identified, along with potential applications in rapid prototyping and AI-assisted automotive development. This document focuses on software-based design and simulation and excludes physical manufacturing and regulatory processes.

1.2. **Intended Audience**

This document is intended for stakeholders involved in the design, development, and evaluation of the **Multimodal Automotive Creator** system. The primary audience includes software developers, AI engineers, and system architects responsible for implementing and maintaining the platform. It is also relevant to automotive designers and engineers who will use the system for concept creation and performance analysis.

Additionally, the document serves researchers and academic students working in the fields of artificial intelligence, automotive engineering, and smart mobility systems. Project managers and technical evaluators can use this document to understand system requirements, scope, and feasibility. Overall, it provides clear guidance for both technical and non-technical stakeholders involved in the project.

1.3. **System Overview**

The Multimodal Automotive Creator is an intelligent application designed to generate automotive content using multiple input modes such as text, images, and voice. It allows users to express design ideas in a flexible and user-friendly manner. The system processes each input type individually and then combines them to understand user intent accurately. A multimodal fusion mechanism integrates all inputs into a unified request. Based on this, the system generates customized automotive designs or descriptions. The application follows a modular architecture for easy maintenance and scalability. It ensures secure data handling and efficient performance. Overall, the system enhances automotive content creation through multimodal interaction.

2. System Design.

The System Design of the **Multimodal Automotive Creator** defines the structure of the application, flow of processes, data movement, internal components, and APIs used. The design supports **multimodal inputs** such as text, image, and voice to generate automotive-related content efficiently.

2.1. Application Design

The application follows a layered design approach to ensure clarity, modularity, and scalability.

Layers of the Application

1. User Interface Layer
 - Accepts text prompts, image uploads, and voice inputs.
 - Displays generated automotive designs and descriptions.
2. Processing Layer
 - Processes text, image, and audio inputs individually.
 - Combines multimodal inputs into a unified request.
3. Backend & Storage Layer
 - Stores user data, inputs, and generated outputs.
 - Manages API requests and responses.

2.2. Process Flow.

The process flow explains how the system works step by step.

1. User logs into the system.
2. User provides input in the form of text, image, or voice.
3. Input is validated by the system.
4. Text, image, and audio data are processed in their respective modules.
5. Processed data is sent to the multimodal fusion engine.
6. The automotive content generator creates the output.
7. Output is stored in the database.
8. Generated result is displayed to the user.

2.3. Information Flow.

Information flow describes how data moves within the system.

1. User input data flows from the UI layer to the processing modules.
2. Processed data is transferred to the multimodal fusion engine.
3. Combined data is forwarded to the content generation module.
4. Generated information is saved in the database.
5. Final information is returned to the user interface.

2.4. Components Design

The system is divided into the following components:

1. User Interface Component
 - Collects multimodal inputs.
 - Displays system output.
2. Text Processing Component
 - Analyzes user text prompts.
 - Extracts automotive-related keywords.

3. Image Processing Component
 - Analyzes uploaded vehicle images.
 - Identifies design elements and patterns.
4. Audio Processing Component
 - Converts voice input into text.
 - Improves accessibility.
5. Multimodal Fusion Component
 - Integrates text, image, and audio inputs.
 - Resolves conflicts between inputs.
6. Automotive Content Generator
 - Generates automotive designs or descriptions.
 - Produces final output.
7. Database Component
 - Stores user details and history.
 - Stores generated content.

2.5. Key Design Considerations

The following factors were considered during system design:

1. Scalability: System can handle more users and data.
2. Modularity: Easy to add or modify components.
3. Performance: Fast processing of multimodal data.
4. Security: Safe handling of user data.
5. Usability: Simple and intuitive user interface.
6. Reliability: Accurate and consistent output generation.

2.6. API Catalogue.

The API Catalogue lists the APIs used for communication between frontend and backend.

API Name	Method	Description
/login	POST	Authenticates user
/uploadText	POST	Accepts text input
/uploadImage	POST	Accepts image input
/uploadAudio	POST	Accepts voice input
/processInput	POST	Processes multimodal data
/generateContent	POST	Generates automotive output
/getResult	GET	Fetches generated result
/saveData	POST	Stores data in database

3. Data Design.

Data Design defines how data is structured, stored, accessed, and maintained in the **Multimodal Automotive Creator** system. Since the application works with multimodal inputs such as text, images, and audio, proper data organization is essential for smooth system operation, accuracy, and future scalability. The data design ensures consistency, security, and efficient handling of large volumes of diverse data.

3.1. **Data Model**

The data model of the Multimodal Automotive Creator represents the logical structure of data and the relationships between different entities in the system. The primary entities include User, Input Data, Generated Output, and System Logs. Each user has a unique identity and can submit multiple inputs in the form of text, images, or audio. These inputs are linked to a specific user and are processed to generate automotive-related content.

The Input Data entity stores details such as input type, input content, timestamp, and processing status. The Generated Output entity maintains the results produced by the system, including automotive descriptions or design information, along with references to the original inputs. Relationships between these entities ensure traceability, allowing the system to associate generated outputs with their respective user inputs. This structured data model helps in maintaining data integrity and supports easy retrieval of historical records.

3.2. **Data Access Mechanism**

The data access mechanism defines how data is retrieved, stored, and updated within the system. The Multimodal Automotive Creator uses a controlled access approach where the application layer interacts with the database through secure backend services or APIs. Direct access to the database is restricted to maintain data security and consistency.

Whenever a user submits text, image, or audio input, the data is first validated and then stored in the database using backend services. During processing, the system retrieves the required input data, performs multimodal fusion, and saves the generated output back into the database. User data and history are accessed only after proper authentication, ensuring that sensitive information remains protected. This mechanism improves system reliability and prevents unauthorized data manipulation.

3.3. **Data Retention Policies**

Data retention policies define how long different types of data are stored in the system and when they are removed. In the Multimodal Automotive Creator, user-related data such as profile information is retained for as long as the user account remains active. Multimodal input data and generated outputs are stored to provide users with access to their previous creations and to improve system performance through historical analysis.

Temporary data, such as intermediate processing files or cache data, is removed automatically after processing is completed to save storage space. Logs and system activity records are retained for a limited period for monitoring and debugging purposes. These retention policies ensure optimal use of storage resources while maintaining compliance with data privacy and security requirements.

3.4. **Data Migration**

Data migration in the Multimodal Automotive Creator refers to the process of transferring data from one system or storage structure to another without data loss or corruption. This becomes necessary when the system is upgraded, when database schemas are modified, or when the application is moved to a new server or cloud environment. The migration process is planned carefully to ensure continuity of service and data accuracy.

During data migration, existing user data, multimodal inputs, and generated outputs are extracted from the current database and transformed to match the new data structure if required. Validation

checks are performed to ensure that relationships between users, inputs, and outputs remain intact after migration. Backup copies of the original data are maintained to prevent data loss in case of failure. By following a structured migration approach, the system ensures minimal downtime and preserves the reliability and integrity of stored data.

4. Interfaces

Interfaces define how different parts of the Multimodal Automotive Creator system interact with users and with each other. A well-designed interface ensures smooth communication between the user, the application, and backend services. The system mainly includes User Interfaces, External Interfaces, and Internal Interfaces, all of which work together to provide a seamless multimodal experience.

4.1 User Interface

The User Interface is the front-facing part of the system through which users interact with the Multimodal Automotive Creator. It is designed to be simple, intuitive, and responsive so that users with minimal technical knowledge can easily use the application. The interface allows users to enter text prompts describing automotive features, upload reference images, and provide voice input for commands or ideas.

The interface displays generated automotive outputs such as vehicle descriptions or design concepts in a clear and visually appealing format. Input fields, buttons, and navigation elements are arranged logically to reduce user effort. Error messages and validation prompts are shown whenever incorrect or incomplete input is provided, ensuring better usability and user experience.

4.2 External Interfaces

External interfaces define how the system interacts with third-party services and tools. The Multimodal Automotive Creator communicates with external services for tasks such as voice-to-text conversion, image processing, and content generation. These interactions are handled using APIs that allow the system to send requests and receive responses in a standardized format.

The system may also interact with external storage services or cloud platforms for saving multimedia data such as images and audio files. These interfaces help extend the system's functionality without increasing internal complexity and allow easy integration of advanced features in the future.

4.3 Internal Interfaces

Internal interfaces describe communication between different modules within the system. In the Multimodal Automotive Creator, internal interfaces connect components such as the text processing module, image processing module, audio processing module, multimodal fusion engine, and the content generation module.

Each component communicates through well-defined interfaces to exchange processed data. This modular interaction ensures that changes in one module do not affect the entire system. Internal interfaces improve maintainability, make debugging easier, and support future enhancements by allowing new modules to be added with minimal changes to existing components.

4.4 Interface Design Considerations

While designing interfaces, several factors were considered to ensure system effectiveness. The interfaces are designed to be consistent in layout and behavior to avoid user confusion. Performance is optimized so that user interactions and system responses occur with minimal delay. Security is also considered, ensuring that data exchanged between interfaces is protected from unauthorized access.

Scalability is another important factor, allowing the system to support additional features or increased user load in the future. By following these design considerations, the interfaces of the Multimodal Automotive Creator provide a reliable and user-friendly interaction environment.

5. State and Session Management

State and Session Management plays a crucial role in maintaining continuity and personalization in the Multimodal Automotive Creator system. Since users interact with the application through multiple inputs such as text, images, and audio, the system must remember user actions and data during a session. Proper state and session management ensures smooth navigation, secure user interaction, and consistent system behavior throughout the user's interaction with the application.

5.1 State Management

State management refers to how the application maintains and tracks the current condition of the system during user interaction. In the Multimodal Automotive Creator, the state includes information such as user login status, selected input mode, uploaded files, processing status, and generated results. This state information allows the system to provide a seamless experience without requiring users to repeatedly enter the same data.

The application maintains state at both the frontend and backend levels. On the frontend, temporary state is used to track user inputs and interface changes, such as active input fields or progress indicators. On the backend, the system maintains processing states to manage multimodal data handling and content generation. Efficient state management ensures that partial inputs are not lost and that the system can recover smoothly from minor interruptions.

5.2 Session Management

Session management handles user identity and interaction duration within the system. When a user logs into the Multimodal Automotive Creator, a session is created to uniquely identify that user throughout their interaction with the application. This session allows the system to associate user inputs and generated outputs with the correct user account.

Sessions are maintained using secure session identifiers, which are validated with every request sent to the backend. The session remains active until the user logs out or the session times out due to inactivity. Session expiration helps protect user data and prevents unauthorized access. By managing sessions effectively, the system ensures secure access control and personalized user experiences.

5.3 Handling Multimodal Sessions

In a multimodal environment, session management becomes more complex due to multiple input types. The Multimodal Automotive Creator maintains all text, image, and audio inputs under a single active session. This allows the system to correctly combine different inputs provided by the user during the same session.

For example, if a user uploads an image and later adds a text description, both inputs are stored and processed together within the same session context. This unified session handling ensures accurate

multimodal fusion and consistent output generation, enhancing the overall effectiveness of the system.

5.4 Security and Reliability in Session Management

Security is a key consideration in state and session management. The system uses authentication mechanisms to ensure that only authorized users can create and maintain sessions. Sensitive session data is protected against unauthorized access, and session identifiers are refreshed periodically to reduce security risks.

Reliability is achieved by maintaining session consistency even during temporary network issues. If a session is interrupted, the system can restore essential state information once the user reconnects. This approach prevents data loss and improves user trust in the application.

6. Caching

Caching is an important performance optimization technique used in the Multimodal Automotive Creator system. Since the application processes large amounts of multimodal data such as text, images, and audio, repeated processing of the same or similar requests can increase system load and response time. Caching helps reduce this overhead by temporarily storing frequently accessed or recently generated data so that it can be reused without reprocessing.

In the Multimodal Automotive Creator, caching is primarily used to store intermediate results, frequently generated automotive outputs, and commonly accessed user data during active sessions. When a user submits a request that is similar to a previously processed request, the system first checks the cache before performing full processing. If the required data is found in the cache, it is returned immediately, significantly improving response time and overall user experience.

The system implements caching at multiple levels. At the application level, temporary data such as user session information and processing states are cached to ensure smooth interaction and quick transitions between different stages of input and output. At the backend level, generated automotive content and frequently accessed reference data are cached to minimize repeated database queries. This multi-level caching approach improves system efficiency and scalability.

Cache consistency and validity are carefully maintained to ensure that outdated or incorrect data is not served to users. Cached data is assigned an expiration time after which it is automatically removed or refreshed. When a user updates their input or modifies preferences, the related cache entries are invalidated to ensure that new outputs are generated based on the latest information. This balance between performance and data accuracy ensures reliable system behavior.

Caching also plays a role in reducing server load and optimizing resource usage. By limiting unnecessary processing and database access, the system can support more users simultaneously without performance degradation. Overall, the caching strategy enhances speed, reliability, and scalability, making the Multimodal Automotive Creator more efficient and user-friendly.

7. Non-Functional Requirements

Non-functional requirements describe the quality attributes of the Multimodal Automotive Creator system rather than its specific functionalities. These requirements focus on how well the system performs, how secure it is, and how reliably it operates under different conditions. Meeting non-

functional requirements ensures that the system is efficient, secure, and user-friendly while handling multimodal data such as text, images, and audio.

7.1. **Security Aspects**

Security is a critical non-functional requirement for the Multimodal Automotive Creator because the system handles user information and multimedia data. The system ensures secure access by implementing user authentication mechanisms so that only authorized users can log in and use the application. Each user session is uniquely identified, preventing unauthorized access to user data and generated outputs.

Data security is maintained by protecting stored information from unauthorized modification or misuse. Sensitive user data, including login credentials and session details, is handled securely to prevent data breaches. The system also validates all user inputs to protect against malicious data and unauthorized requests. Secure communication between the frontend and backend ensures that data exchanged within the system remains protected.

Regular monitoring and controlled access to system resources further strengthen security. By following these security measures, the Multimodal Automotive Creator ensures data confidentiality, integrity, and safe user interaction.

7.2. **Performance Aspects**

Performance is another key non-functional requirement of the Multimodal Automotive Creator. Since the system processes multiple types of data, it must deliver quick and accurate responses to maintain a smooth user experience. The application is designed to handle multimodal inputs efficiently, minimizing delays during processing and output generation.

Techniques such as caching and optimized data access are used to reduce response time and system load. The system is capable of handling multiple users simultaneously without significant performance degradation. Efficient resource management ensures optimal use of processing power and storage, even during peak usage.

Scalability is also considered as part of performance design, allowing the system to grow and support additional features or increased user traffic in the future. By focusing on performance optimization, the Multimodal Automotive Creator provides fast, reliable, and consistent results to users.

8. **References**

- **Software Engineering**, Ian Sommerville, Pearson Education.
- **GeeksforGeeks**, Articles on system design and multimodal systems.
- **IEEE**, Research papers on multimodal applications.
- **OpenAI**, Documentation on AI and multimodal models.
- **W3Schools**, Web technologies reference.