Cairo University

Faculty of Computers
and Artificial Intelligence

Thotron – AI Narrated Video Compilation

Thotron – AI NVC

Under the Supervision of:

DR: Ihab El-Khodary
DR: Wael Wahby
Eng. Hamza Emad

Academic Year 2024/2025
Artificial Intelligence Department
Jul *5th*2025 ,

Cairo University

Faculty of Computers
and Artificial Intelligence

Thotron - AI Narrated Video Compilation

Thotron – AI NVC

Under Supervision of:

DR: Ihab El-Khodary
DR: Wael Wahby
Eng. Hamza Emad

Prepared by:

Abdelrahman Elsayed	**20210222**
Ahmed Alaa	**20210027**
Farah Walid	**20210290**
Howida Adel Abd El-Halim	**20210446**
Yahia Rashed	**20210454**

Academic Year 2024/2025
Artificial Intelligence Department
Jul 5th 2025

Table of Contents

List of Figures

Acknowledgment

First and foremost, we would like to thank **ALLAH,** who has always sustained us throughout our lives.

This graduation project would not have been accomplished without the help of numerous people. Some contributed directly to the work and some indirectly, and all their contributions were valuable.

We gratefully acknowledge ***Dr. Ihab El-Khodary*** and ***Dr. Wael Wahby*** for their advice, supervision, and crucial contributions, which made them the backbone of this graduation project. Their involvement with their originality has triggered and nourished our intellectual maturity, which we will benefit from for a long time to come.

thanks to ***Eng***. ***Hamza Emad*** for supporting our idea and helping us throughout our project.

Many thanks to our parents and friends for their support and motivation throughout our lives. There are no words that can fully express our gratitude to them. We are grateful in every possible way and hope to continue our collaboration in the future with our supervisors.

Finally, we would like to thank everyone who contributed to the success of this work.

Abstract

The project "Thotron - AI Narrated Video Compilation " aims to brainstorming, knowledge sharing, and educational content creation face multiple challenges. Discussions often become inefficient, leading to fragmented ideas “Polarization” and the loss of valuable insights. Educators and content creators struggle with managing large volumes of video content, while time constraints and information overload hinder effective communication and learning.

Thotron was born from the need to solve these challenges. It is an AI-enhanced platform designed to streamline idea organization, enhance collaboration, and revolutionize video content creation. By integrating AI-powered automation with YouTube APIs, Thotron allows users to effortlessly retrieve, edit, and narrate video content.

Thotron offers a set of innovative features to enhance idea merging and video-based learning. The AI-powered Map of Ideas automatically groups and maps similar ideas, providing a structured format for discussions. The YouTube Video Organizer enables AI-driven retrieval, editing, and categorization of video content for enhanced accessibility. TipTalk AI Narration allows AI-generated avatars and real-character voiceovers to create engaging video presentations. Finally, AI-Generated Storytelling converts raw user prompts into structured, AI-generated video narratives, streamlining content creation.

Developing Thotron into a fully functional and scalable platform requires the implementation of advanced AI and machine learning algorithms for natural language processing, text analysis, and video processing. Additionally, seamless API integration with YouTube and other content repositories is essential for efficient content retrieval and management. A robust front-end and back-end system will ensure smooth user experience, while scalability and optimization techniques will be implemented to facilitate real-time AI recommendations and fast processing speeds.

Chapter 1: Introduction

In today's fast-paced digital environment, collaboration, idea-sharing, and educational content creation often suffer from inefficiency, lack of structure, and high time demands. Unorganized discussions frequently lead to the loss of valuable insights and make it challenging to synthesize diverse ideas into cohesive outcomes. Traditional tools fall short in providing effective solutions for knowledge organization and presentation, resulting in information overload and reduced productivity for educators, content creators, and professionals alike.

Thotron — an AI-driven platform built to redefine how ideas are shared and knowledge is structured. Acting as a smart creative partner, Thotron intelligently merges concepts, organizes conversations, and enhances the video-production experience.

At its core, Thotron is committed to:

- Simplifying the brainstorming journey
- Empowering educators and video creators to produce high volumes of quality content efficiently
- Elevating the way knowledge is communicated — making it more engaging, accessible, and impactful

Thotron AI-NVC is a cutting-edge AI solution designed to redefine the video production experience through intelligent automation and assistance. More than just a functional tool, it fosters structured dialogue and purposeful idea exchange — offering a transformative platform that elevates how information is developed, delivered, and experienced.

We aim to simplifying video production, enabling users to effortlessly retrieve, edit, and narrate content with the support of advanced AI. Beyond streamlining workflows, it underscores the importance of organized discussions and collaborative idea-sharing — providing an innovative environment that reshapes the way knowledge is presented and consumed.

1.1. Problem

With the rising demand for video content in digital marketing, education, and entertainment, there is a need for a tool that simplifies video content creation while maintaining professional quality and trusted content. to merge multiple ideas into a cohesive concept Search manually for each video and go through the editing process is often inefficient, information overload, unstructured, and time-consuming.

Furthermore, traditional tools do not provide an efficient means to organize and present editing video tools effectively.

As a result, educators, content creators, and professionals struggle with information overload and time constraints, making knowledge management and video-based learning less effective.

Thotron aims to solve these challenges by offering an AI-driven platform that systematically organizes discussions, merges ideas, and enhances video-based knowledge sharing. By structuring conversations and utilizing AI-powered tools, the platform ensures that every valuable insight is captured and efficiently presented.

1.2. Scope

Thotron is an AI-powered system that restructures knowledge-sharing by mapping ideas, optimizing video retrieval and narration, and enabling AI-driven storytelling. It enhances collaboration, streamlines educational content creation, and simplifies video management, allowing users to curate, edit, and present information efficiently.

1.3. Target Users

- Content creators
- Students and researchers
- Video editors
- General users looking for condensed video insights

1.4. Objectives

Thotron aims to redefine collaborative thinking by integrating AI-driven automation into content creation and knowledge management. It structures ideas intelligently, enhances video-based learning with AI-powered search and retrieval, and enables AI-generated storytelling from text prompts. Designed for scalability and real-world adaptability, the platform ensures seamless performance

1.5. Solution

Thotron provides a comprehensive AI-enhanced ecosystem to streamline idea organization and video content creation through the following key features:

- AI-Powered Idea Mapping: Automatically structures and categorizes discussions, and provides a clear visual representation of connected ideas.
- YouTube API Integration: Retrieves and organizes video content efficiently, and enhances accessibility and interactivity of educational resources.
- User-Friendly Interface: Features an intuitive drag-and-drop functionality, allowing seamless arrangement and customization of content.

1.6. Methodology definition

For this system, we are going to follow a structured and sequential approach to ensure clarity, efficiency, and scalability. We have chosen the Waterfall Model, a linear development methodology where each phase must be fully completed before moving to the next. This approach provides a clear roadmap, making it easier to manage and implement the project.

The development process consists of the following key phases:

- **Requirement Gathering:** Understanding the problem domain and defining the core requirements, engaging with users to collect insights and refine system needs.
- **Analysis:** Detailing system requirements and user expectations, identifying technical and functional specifications to guide development.
- **Design:** Creating the system architecture and technical framework
- Selecting suitable technologies, programming languages, and database structures.
- **Implementation:** Developing the frontend with an intuitive UI/UX for smooth user interaction, Building the backend, including APIs and databases, for efficient content management and retrieval, Integrating AI models for text analysis, recommendation systems, and video processing.
- **Testing & Optimization**: Ensuring system performance, reliability, security, implementing real-time AI recommendations to enhance the user experience.
- **Deployment & Maintenance:** Launching the platform and ensuring seamless functionality across different environments, scaling features while maintaining high performance and user satisfaction.

  By following this structured methodology, Thotron ensures a well-organized, efficient, and scalable system that meets user needs while maintaining a high standard of performance.

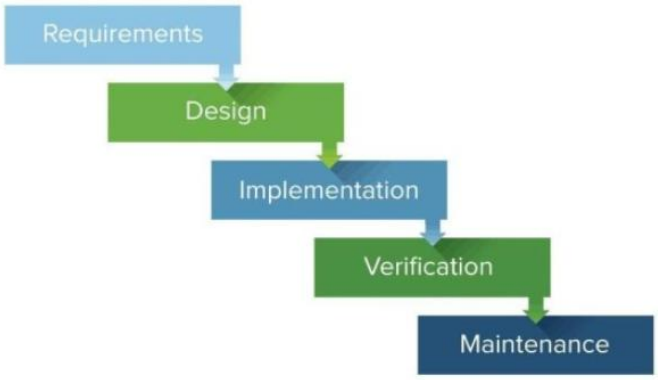

Figure 1-1:waterfall methdology

Chapter 2: Related work

2.1. Current Tools

2.1.1. Wondershare Filmora

Wondershare Filmora is a popular video editing software designed for beginners and professionals alike. It provides a simple yet powerful interface with drag-and-drop functionality.

Features:

- AI-powered tools for background removal and motion tracking.
- Built-in stock media, effects, and transitions.
- Advanced audio editing features, including noise reduction.
- Supports 4K video editing and exporting.
- AI-driven subtitle generation and text-to-speech options.

2.1.2. Clipchamp

Clipchamp is an online video editor by Microsoft, offering cloud-based editing with a focus on ease of use.

Features:

- Templates for various video formats (social media, presentations, ads).
- AI-powered text-to-speech and auto-captions.
- Free stock footage, audio, and customizable transitions.
- Cloud-based editing with real-time collaboration.
- Direct export to platforms like YouTube, TikTok, and OneDrive

2.1.3. ImageFX

ImageFX is an AI-powered tool for generating and editing visual content, often used for creative effects.

Features:

- AI-driven image enhancement and transformation.
- Text-to-image and AI-generated artwork.
- Advanced filters and special effects.
- Real-time preview and customization options.
- Integration with other design tools for workflow automation.

2.1.4. Lovo AI

Lovo AI is an advanced text-to-speech and AI voiceover platform designed for content creators.

Features:

- AI-generated voices with natural speech patterns.
- Multiple languages and voice customization options.
- Custom branding for commercial use.
- Voice cloning for personalized AI narrations. Seamless integration with video editing software.

2.1.5. Vidnoz AI

Vidnoz AI is an AI-powered video creation tool designed for businesses, marketing, and educational purposes.

Features:

- AI-generated avatars for video narration.
- Text-to-video automation for quick content creation.
- Library of templates for different industries.
- Voiceover customization and lip-syncing technology.
- Cloud-based platform for easy access and sharing.

2.1.6. Captions.ai

Captions.ai is an AI-driven tool designed to automate the process of adding captions to videos.

Features:

- Automatic speech recognition and subtitle generation.
- Customizable captions with different fonts and colors.
- AI-powered translation for multilingual subtitles.
- Supports multiple video formats and social media platforms.
- High accuracy with punctuation and speaker detection.

2.2. Thotron AI NVC vs Current Tools

Thotron bridges the gaps left by current tools—filling in what they weren't built to do as it introduces a new layer of intelligence on top of traditional tools.it's designed to orchestrate content creation in a smarter, faster, more cohesive way. Think of it as your creative AI co-director.

2.2.1. Automated Video Compilation

- Thotron AI-NVC: Automatically assembles videos from media assets based on a defined goal, mood, or audience. It chooses transitions, highlights, pacing—no editing skill needed.

- Others: Manual stitching of clips. Users need to understand storytelling, editing timelines, and formats.

- Why it matters: Thotron saves hours by skipping manual editing while producing emotionally compelling narratives.

2.2.2. Creator-Centric AI Tools

- Thotron AI-NVC: Focused on helping vedio creators, with recurring video needs (e.g., tutorials, product showcases, explainers).

- Others: Canva focuses on general design; CapCut gears toward casual, trendy videos.

- Why it matters: Thotron delivers depth and consistency for creators—not just flashy edits.

2.2.3. YouTube API Integration

- Thotron AI-NVC: Deep integration allows uploading, metadata tagging, thumbnail selection, and title optimization—right from Thotron.

- Others: Users create content separately and manually upload to YouTube.

- Why it matters: You can strategize, optimize, and publish in a single flow—like having your own content manager.

2.2.4. Workflow Automation

- Thotron AI-NVC: Automates repetitive tasks like clipping videos and repurposing long videos into short clips.

- Others: Often focus on one feature (editing, captioning) in isolation.

- Why it matters: You get a studio-level pipeline without the overhead. Ideal for video editing beginners.

2.2.5. User Experience & Customization

- Thotron AI-NVC: Minimal learning curve, intuitive interface, and modular templates tailored for different content goals.
- Others: Steeper learning curve or rigid templates.
- Why it matters: Thotron speaks the user's intent, not just their inputs.

- Overview of how Thotron-AI NVC uniquely enhances the Video creation landscape.

Feature	Thotron	Wondershare Filmora	Clipchamp	Image FX	Captions.ai	Canva	CapCut
AI-Powered Editing	✅	✅	✅	❌	❌	✅	✅
Auto Video Compilation	✅	❌	❌	❌	❌	❌	❌
YouTube API Integration	✅	❌	❌	❌	❌	❌	❌
User-Friendly Interface	✅	✅	✅	✅	✅	✅	✅
Project Downloading	❌	✅	✅	✅	✅	✅	✅
Share project	✅	✅	✅	✅	✅	✅	✅

Figure 2-1:Comparison between Thotron AI- NVC and others

Chapter 3: AI Strategy and Implementation

3.1. AI Plan for Thotron AI-NVC

Thotron's AI-driven automation is divided into multiple modules, each designed to optimize video production. The two main modules currently under development are **Transcript Analysis** and **Project Summarization**, with additional AI modules planned for future enhancements.

3.2. API Transcript Extraction

3.2.1. Objective

To extract human-readable captions from YouTube videos using official and fallback methods for robustness. This forms the data source for all downstream AI processes, including summarization, video clipping, and transcript visualization.

3.2.2. Primary Method: YouTube Transcript API

- Uses `YouTube_transcript_api` to retrieve available subtitle tracks for a video.
- Language fallback: attempts to retrieve English or Arabic subtitles.
- Handles common issues like:
  - Transcripts disabled.
  - Video unavailable.
  - Rate limiting.

3.2.3. Fallback Method: Supadata API

- If the YouTube Transcript API fails, `whisper model` is used as a backup.
- `whisper model` retrieves full transcripts and provides additional metadata if available.
- Helps bypass edge cases and enhances reliability for long videos or regional content.

3.2.4. Transcript Segmentation

The /GetTranscript endpoint enables selective extraction:

- Takes optional start_time and end_time parameters.
- Returns only transcript segments within that range.
- Helps frontend or auto-clipping logic focus on relevant parts.

3.3. **Module 1**: Auto-Clipping Module

Ensuring smooth transitions between cropped video segments is essential for coherence. To address this, we are implementing video embeddings to compare segments beyond just textual content.

This method ensures: Contextual continuity by assessing the visual and audio similarity of video segments. Smooth transitions by minimizing abrupt scene changes. Logical storytelling flow that aligns with the intended message.

3.3.1. Overview

The Auto-Clipping Module is designed to intelligently extract the most relevant segments from a YouTube video transcript based on a user-provided prompt. It leverages T5 transformer-based embeddings and cosine similarity to locate the best-matching sequence of transcript segments. This module enables dynamic, context-aware video clip generation—laying the foundation for auto-editing, highlight extraction, and personalized content summaries.

3.3.2. Functional Objective

Given a set of YouTube video IDs and a semantic prompt (with optional headings), the system:

- Retrieves transcripts from the videos
- Embeds both the prompt and transcript segments using the T5 model
- Calculates semantic similarity between the prompt and each caption
- Selects the best sequence of transcript segments (based on score and duration)
- Returns the matching video segment's start/end timestamps and key quotes.

3.3.3. Architecture Flow

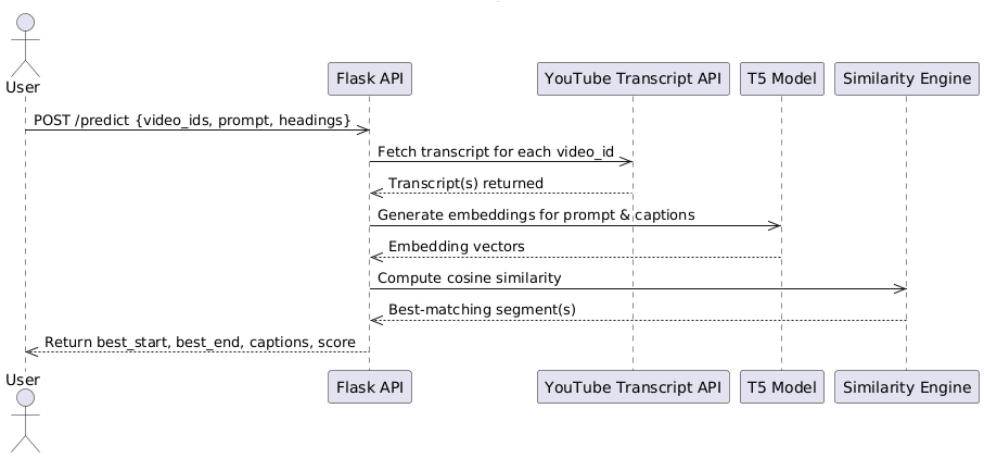

Figure 3-1: Auto-clipping module Sequence

3.3.4. Core Components

A. Semantic Embedding (T5)

- Uses the T5-base model to transform both the user prompt and each transcript segment into a high-dimensional vector.
- Embeddings represent the semantic meaning of each sentence, enabling meaningful comparison between the prompt and the caption.

B. Cosine Similarity

- Computes pairwise cosine similarity between prompt embedding and all caption embeddings.
- Scores are used to rank and select the most semantically aligned caption or caption sequence.

C. Segment Optimization

- Once a top match is identified, the algorithm continues to expand the segment to satisfy a **minimum duration threshold**.
- Thresholds are dynamically calculated based on video length (e.g., 20□30% of total duration).
- Ensures returned clip is informative, coherent, and long enough for practical use.

D. Multi-Video Ranking

- After processing all videos, the system ranks them based on the **average semantic alignment** between the prompt and their best segments.
- The most relevant videos surface first for downstream tasks (e.g., summarization, editing).

3.3.5. Use Cases

- Extracting highlights for video summarization
- Contextual video recommendation
- Auto-generating content for short-form platforms (e.g., Reels, Shorts)

3.4. Module 2: Project Summarization

A significant challenge in video content creation is copyright restrictions, which often prevent users from downloading edited videos. As an alternative, Thotron provides a text-based summarization feature powered by Large Language Models (LLMs).

3.4.1. Overview

The summarization module is a core component of our AI-driven video transcript processing system. It transforms long-form YouTube-style transcripts into concise, structured summaries using a fine-tuned LLaMA 3.2-1B language model. This module not only reduces cognitive load for end users but also enables effective content extraction, knowledge organization, and downstream content generation (e.g., HTML, PDF).

3.4.2. Key Features

- Converts raw transcripts into clean, organized Markdown summaries.
- Generates multi-level structured outputs: main topic, list of subtopics, summaries, key terms, and representative quotes.
- Ensures output maintains the original language (Arabic or English) and tone of the transcript.
- Uses a schema-guided summarization approach with Pydantic models for consistency and validation.

3.4.3. Architecture and Flow

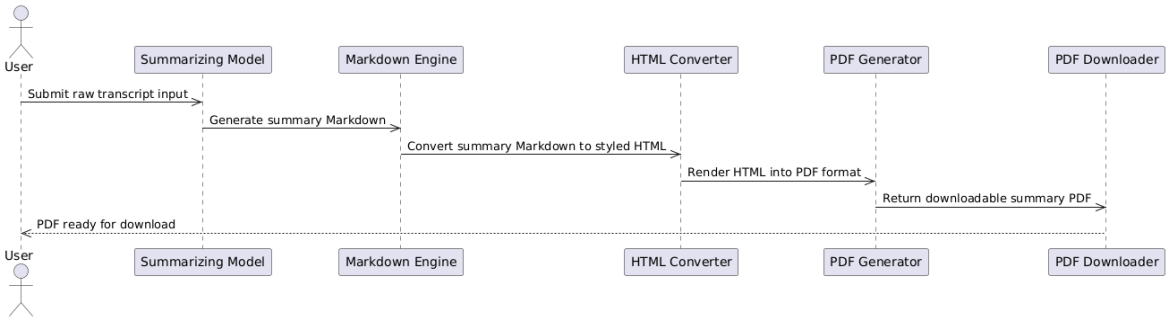

Figure 3-2: Summarizing module Sequence

3.4.4. Model Configuration

- **Base Model:** meta-llama/Llama-3.2-1B-Instruct
- **Fine-Tuning:** LoRA (Low-Rank Adaptation) applied using domain-specific summarization data
- **Framework:** VLLM for optimized inference
- **Tokenizer:** HuggingFace's AutoTokenizer
- **Inference Settings:**
  - max_tokens: 4096

3.4.5. Prompt Strategy

We utilize a system-user chat-style prompt format:

- **System message** defines summarization behavior and constraints.
- **User message** includes the raw transcript and the JSON schema definition.
- The model responds strictly with the structured Markdown output conforming to the schema.

3.4.6. Pydantic Schemas

Each SubtopicSummary contains:

- Title
- Concise summary
- List of key terms
- List of direct quotes from the transcript.

3.4.7. Language Detection

Language is auto-inferred from content (Arabic vs English) to ensure natural phrasing and correct directionality in output. Summaries are always returned in the same language as the original transcript, preserving accessibility for the target audience.

3.4.8. Integration with Other Modules

The summarization module feeds directly into:

- **HTML Generation Module:** Parses Markdown to multilingual HTML.
- **PDF Module:** Converts HTML to print-ready PDF via WeasyPrint.
- **Frontend Viewers:** Allows users to view and download summaries.

3.4.9. Conclusion

The Summarization Module provides an intelligent, multilingual summarization capability that sits at the heart of the system. By leveraging fine-tuned transformer models and structured schema prompts, it delivers coherent, informative summaries that greatly enhance transcript readability, user engagement, and downstream utility (e.g., report generation, learning materials, etc.).

3.5. Module 3: Narration (Avatar)

A frequent challenge in modern e-learning and storytelling tools is maintaining engagement without burdening users with complex video editing or manual voiceovers. Thotron addresses this by providing a streamlined AI-powered narration experience through expressive digital avatars.

3.5.1. Overview

The Avatar Narration Module is responsible for transforming text summaries into dynamic, lip-synced video clips. It uses external AI services to first synthesize high-quality speech from textual input and then synchronize that speech with a user-selected avatar image. The result is a visually engaging narrated video that conveys content more effectively and accessibly.

3.5.2. Key Features

- Converts project summaries into realistic speech using voice cloning.
- Supports user selection of voice type (male/female) and avatar image.
- Generates lip-synced avatar videos with expressive facial movement.
- Provides downloadable video files rendered in real-time.
- Automatically copies the result to a system-specific destination folder.

3.5.3. Architecture and Flow

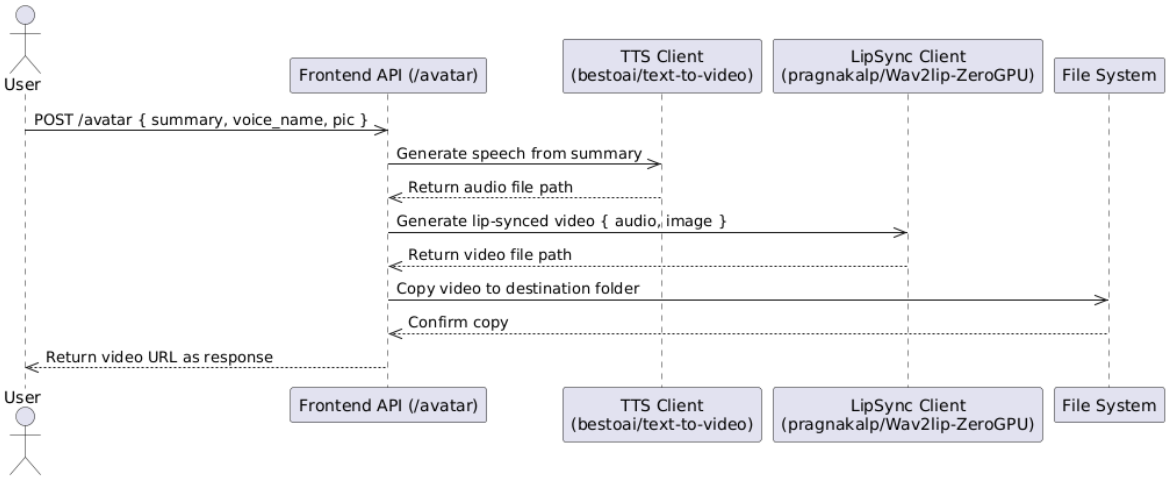

Figure 3-3: Narration module Sequence.

3.5.4. Model Configuration

- TTS Client: Gradio API (bestoai/text-to-video)
- Lip-Sync Engine: Gradio API (pragnakalp/Wav2lip-ZeroGPU)
- Destination Handling: Video is stored in a fixed path on the server (C:\Users\...\Downloads)
- Flask Framework: Hosts API endpoints for triggering the avatar workflow.

3.5.5. Error Handling Strategy

- Checks for missing or invalid input fields.
- Returns JSON-formatted error messages if TTS or video rendering fails.
- Wraps the video copy operation in exception handling for robustness.

3.5.6. Integration with Other Modules

- The output from this module is triggered after a project has been summarized.
- Accepts inputs from the Summarization Module (summary text).
- Generated avatar videos can be shared using the Project Sharing Module.
- The result may be included in final video compilation pipelines for future export options.

3.5.7. Conclusion

The Avatar Narration Module adds a unique dimension to Thotron's video capabilities by transforming static summaries into personable, engaging content. Its integration with AI voice synthesis and lip-syncing technologies makes it an intuitive and powerful storytelling tool for educators, content creators, and learners.

3.6. Future AI Modules in Development

We are actively developing additional AI-powered modules to further enhance automation, improve video editing quality, and refine storytelling structures in Thotron. More details on these upcoming modules will be shared as development progresses.

Chapter 4: Thotron AI-NVC Analysis

4.1. Project stakeholders

Stakeholders are individuals or groups who have a vested interest in the development, deployment, and impact of the Thotron platform. In this system, there are three primary stakeholder categories:

First, the target users, which include content creators, educators, and learners who seek more effective, intelligent ways to produce, organize, and learn from video content. They benefit from Thotron's automated summarization, narration, and storytelling capabilities.

Second, the project team members and technical contributors, encompassing AI engineers, software developers, designers, and project managers responsible for conceptualizing, building, and maintaining the system. This also includes academic or strategic advisors who guided the project's vision and ensured alignment with pedagogical or technological standards.

Third, any third-party service providers or integration partners, such as API providers (e.g. YouTube), voice synthesis platforms, or cloud infrastructure services, which Thotron relies on to operate its core functionalities efficiently

4.2. Functional requirements

- **Users can search for YouTube videos** by entering a topic manually, then select desired video segments by defining start and end timestamps for each, and rearrange these clips in a timeline to create a custom video (Manual Clipping Scenario).
- **Users can choose the Auto-Clipping feature** by providing a topic (prompt), a related heading, and several videos to retrieve.

  The system will:

  - Fetch transcripts for the selected videos.
  - Embed both the prompt and transcript segments using the T5 model.
  - Compute semantic similarity between the prompt and the transcript.
  - Select and merge the most relevant video segments based on score and duration.
  - Return a compiled video with time-aligned highlights and key quotes.

- **Users can save a project** after either scenario for future review or continued editing.
- **Users can generate a transcript summary** of the selected video parts and receive a downloadable PDF file (Summarization Module).
- **Users can create AI-narrated videos** by:
    - Selecting a character avatar (image).
    - Choosing a narration voice (male or female).
    - Submitting the project summary for rendering.
    - Receiving a narrated video with synced lip movement based on the selected voice and avatar (Narration Module).
- **Users can share projects via link access**, which opens the project in a new browser window using its unique project ID. Shared projects are view-only and cannot be edited by recipients.

4.3. Non-functional requirements

- **Usability**: Thotron is designed to be intuitive, accessible, and user-friendly. Users should be able to easily navigate its features and understand its capabilities without needing extensive guidance.
    - The interface will follow familiar patterns found in modern web platforms.
    - A tutorial will guide users through advanced features such as auto-clipping, narration, and summarization.
    - All available features will be clearly presented to users during onboarding.
    - The website will offer streamlined navigation to ensure every task is easy to locate and
- **Response Time**: The system will aim for a maximum response time of **10 minutes** for AI-intensive operations such as transcript analysis, narration rendering, and summarization. Performance will be supported by efficient backend infrastructure to minimize delays.
- **Platform Support**: Thotron will be delivered as a **web-based platform**, accessible via modern desktop and mobile browsers (Chrome, Firefox, Edge, Safari), eliminating the need for native app installations while ensuring responsive design across devices.
- **Security Requirements**:
    - All user credentials will be encrypted using **SHA-256** hashing algorithms.
    - The system architecture will resist high-load access attempts, withstanding up to **1000 login requests per second** without degrading performance.
    - Project-sharing links will include strict access controls, ensuring shared content remains **view-only** to protect creator integrity.

4.4. User Profiles

Characteristic	Decision Makers
Age	16 and above
Language	Primarily English; potential for future multilingual expansion
Education	Secondary school and above; comfortable with media-rich digital platforms
Computer Experience	Medium – able to navigate websites, upload files, and explore basic content tools
Domain Expertise	Medium to High – educators, content creators, or learners with a topic-specific focus.
Task Knowledge	Medium – familiar with video creation goals but not necessarily with editing tools.
Expectations	Fast processing, AI-assisted support, minimal manual editing, clear and simple UI

4.5. Personas

1) Nour: 19 years old, first-year media student exploring digital storytelling for the first time.
What she needs:
She wants a simple way to turn her creative ideas into videos without needing to learn editing software. She needs a tool that guides her through idea mapping, video selection, and automatic narration without requiring technical skills.

---

2) Kareem: 24 years old, YouTube content creator producing weekly explainers in science and technology.
What he needs:
He's looking to speed up his content production. He wants to input a topic, receive curated clips, and auto-generate narration with minimal editing. Seamless integration with YouTube and fast turnaround are essential for his workflow.

---

3) Leila: 33 years old, instructional designer at an online education platform.
What she needs:
She needs to generate structured educational videos from long lectures and content libraries. She values transcript summarization, idea grouping, and PDF exports to build companion materials for her lessons.

---

4) Rami: 21 years old, computer science student working on a student-led project about knowledge sharing.
What he needs:
He wants a collaborative video workspace where team members can search, clip, and narrate ideas together. He's drawn to Thorton's AI-enhanced flow for combining content into one cohesive story without heavy manual work.

4.6. User Stories

User Story ID	US #1
User Story Name	User Register
Actors	User
Description	As a User I like to be able to register for the website So that I can register and access its services
Per condition	The user is not already registered, The user has a valid email for OTP verification.
Post condition	The user receives an OTP and verifies it successfully. The user should go to the login page.
Acceptance Criteria	Given I'm a User and I'm on the register page, When I fill in the register information fields with my credentials, and I click register Then the system registers me

User Story ID	US #2
User Story Name	User Login
Actors	User
Description	As a User, I like to be able to log in to the website So that I can use its functionalities
Per condition	The user is already registered and has a valid account: email and password.
Post condition	
Acceptance Criteria	Given I'm a user with an account and I'm on the login page, When I fill in the login information fields with my credentials, and I click login Then the system logs me in.

User Story ID	US #3
User Story Name	Creating a New Project using Manual Scenario
Actors	User
Description	As a User, I like to be able to create a new project so that I can use YouTube videos of my selection and montage them.
Per condition	I should be logged in to the website.
Post condition	The new project is saved with the selected video clips and sequence. The project is visible in the user's dashboard.
Acceptance Criteria	Given I'm a user with an account and I'm on the Video Editor page, I can use YouTube videos of my selection, I can trim them into clips of my interest and rearrange the order of the video clips. The project is saved and listed on the dashboard. The project can be reopened and edited later.

User Story ID	US #4
User Story Name	Creating a New Project using AI Scenario
Actors	User
Description	As a User, I like to be able to create a new project So that I can use YouTube videos selected and trimmed and rearranged by AI.
Per condition	I should be logged in to the website. Insert the project name, the relevant topics of the project.
Post condition	A new project is created and opened. The selected YouTube videos are added to the timeline. The project appears in the user's dashboard after being saved.
Acceptance Criteria	Given I am logged in and on the Video Editor page, When I enter a project name and topics and click "Generate", Then, a new project is created with the AI-processed videos in the timeline, and visible in my dashboard.

User Story ID	US #5
User Story Name	Create a Summary for the project using AI
Actors	User
Description	As a User, I want to create a summary for my project using AI, So that I can convert YouTube video content into concise bullet points.
Per condition	I must be logged in to the website. The videos must have available transcripts.
Post condition	The summary is generated from the videos in the timeline and displayed in a modal.
Acceptance Criteria	Given I'm a user with an account and I'm on the Video Editor page, I can click a button to generate a summary using AI. The output is a concise summary and can be downloaded in PDF format.

User Story ID	US #6
User Story Name	Delete Project
Actors	User
Description	As a User, I like to be able to delete a project So that I can remove projects I no longer need from my dashboard.
Per condition	I must be logged in to the website. The project must exist in my dashboard.
Post condition	The project is permanently deleted from the dashboard and is no longer accessible.
Acceptance Criteria	Given I'm a user with an account and I'm on the dashboard page, When I click the delete icon on a specific project card, Then I saw a confirmation prompt. After confirmation, the project is removed from my dashboard.

User Story ID	US #7
User Story Name	Search Project
Actors	User
Description	As a user, I want to search for a project by title, So that I can quickly find a specific project in my dashboard.
Per condition	I must be logged in to the website. The project must exist. I must enter the project title into the search field.
Post condition	The matching project is displayed and other projects are temporarily hidden until I press the "Escape" key.
Acceptance Criteria	Given I am logged in and on the dashboard page, When I search using a project title, Then the matching project is shown and others are hidden. When I press "Escape", all projects reappear.

User Story ID	US #8
User Story Name	Share Project
Actors	User
Description	As a User, I like to be able to share my project So that anyone with the link can view the project, even without logging in.
Per condition	I must be logged in to the website. The shareable project needs to exist or be created and then saved before sharing.
Post condition	A shareable link is generated. The project is accessible to anyone with the link.
Acceptance Criteria	Given I am logged in and on the Video Editor page, I can click a button to share the project. Anyone with the link can view the project montage.

User Story ID	US #9
User Story Name	Insert Avatar into Project
Actors	User
Description	As a User, I like to be able to put an avatar in my project So that the avatar is narrating the summary produced from the summarization model.
Per condition	I must be logged in to the website. The videos must have available transcripts. A summary must have already been generated.
Post condition	A video of the avatar narrating the summary is created and inserted into the Video Editor page.
Acceptance Criteria	Given I am logged in and have a project with a generated summary, When I select an avatar and click "TipTalk", Then a video is generated with the avatar narrating the summary, And it appears in the timeline on the Video Editor page.

4.7. Use Case Diagram

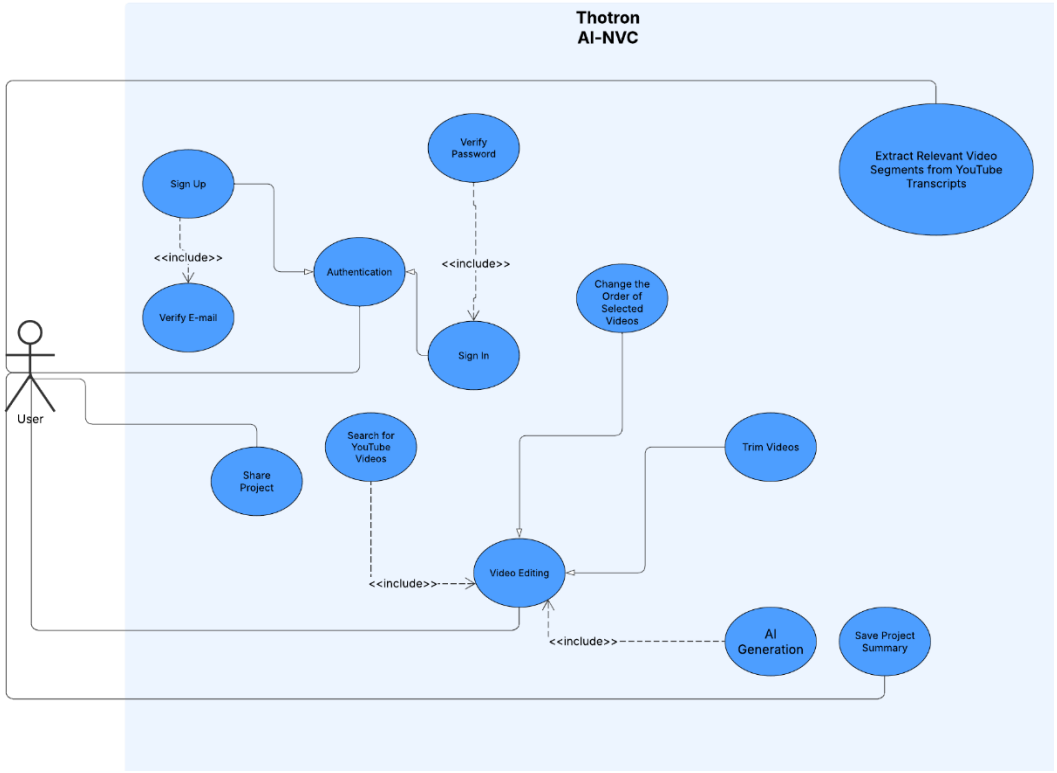

Figure 4-1: Thotron AI-NVC Use Case Diagram

4.8. Class Diagram

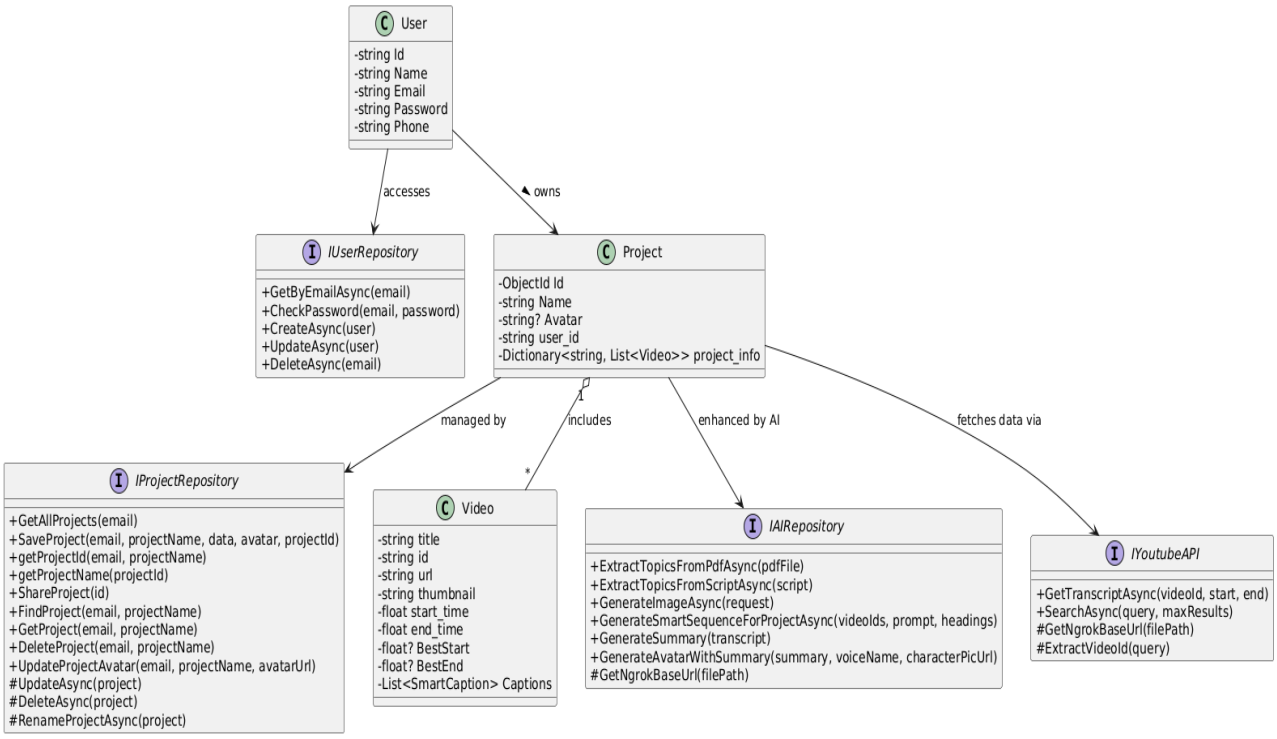

Figure 4-2: Thotron AI-NVC Class Diagram

4.9. Entity Relationship Diagram

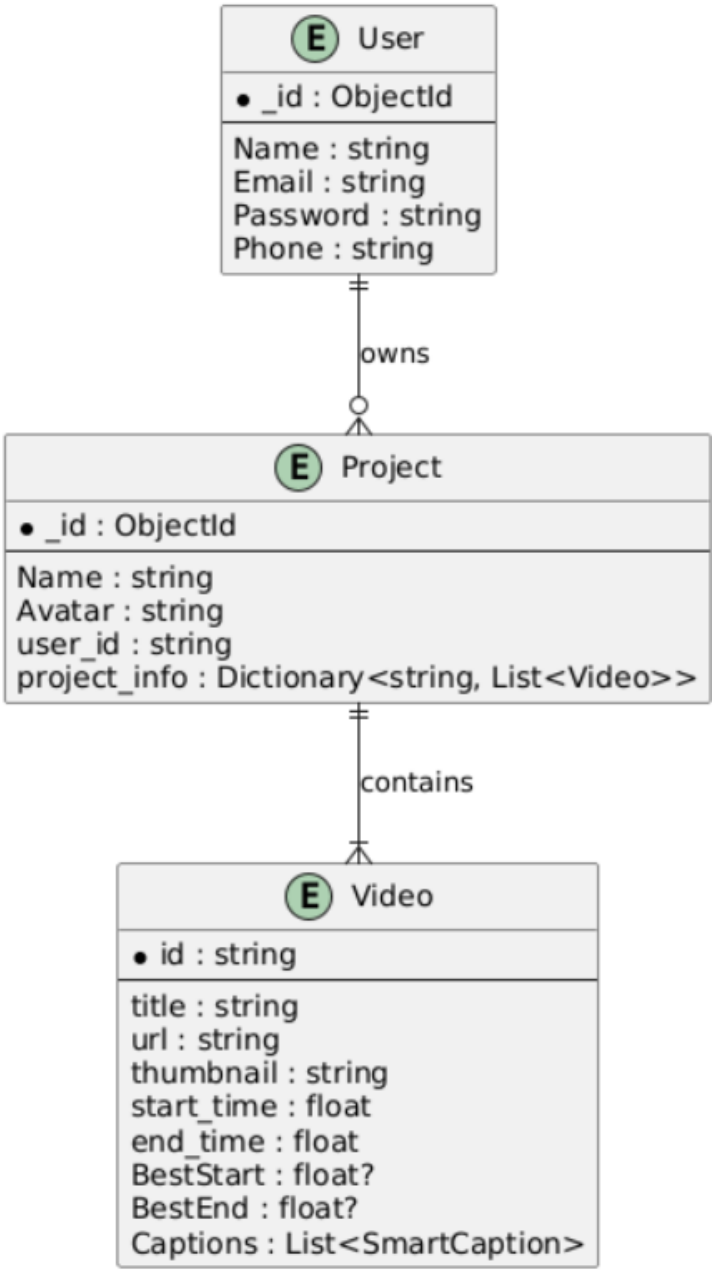

Figure 4-3: Thotron AI-NVC Entity Relationship Diagram

4.10.Sequence Diagrams

4.10.1. Manual Scenario

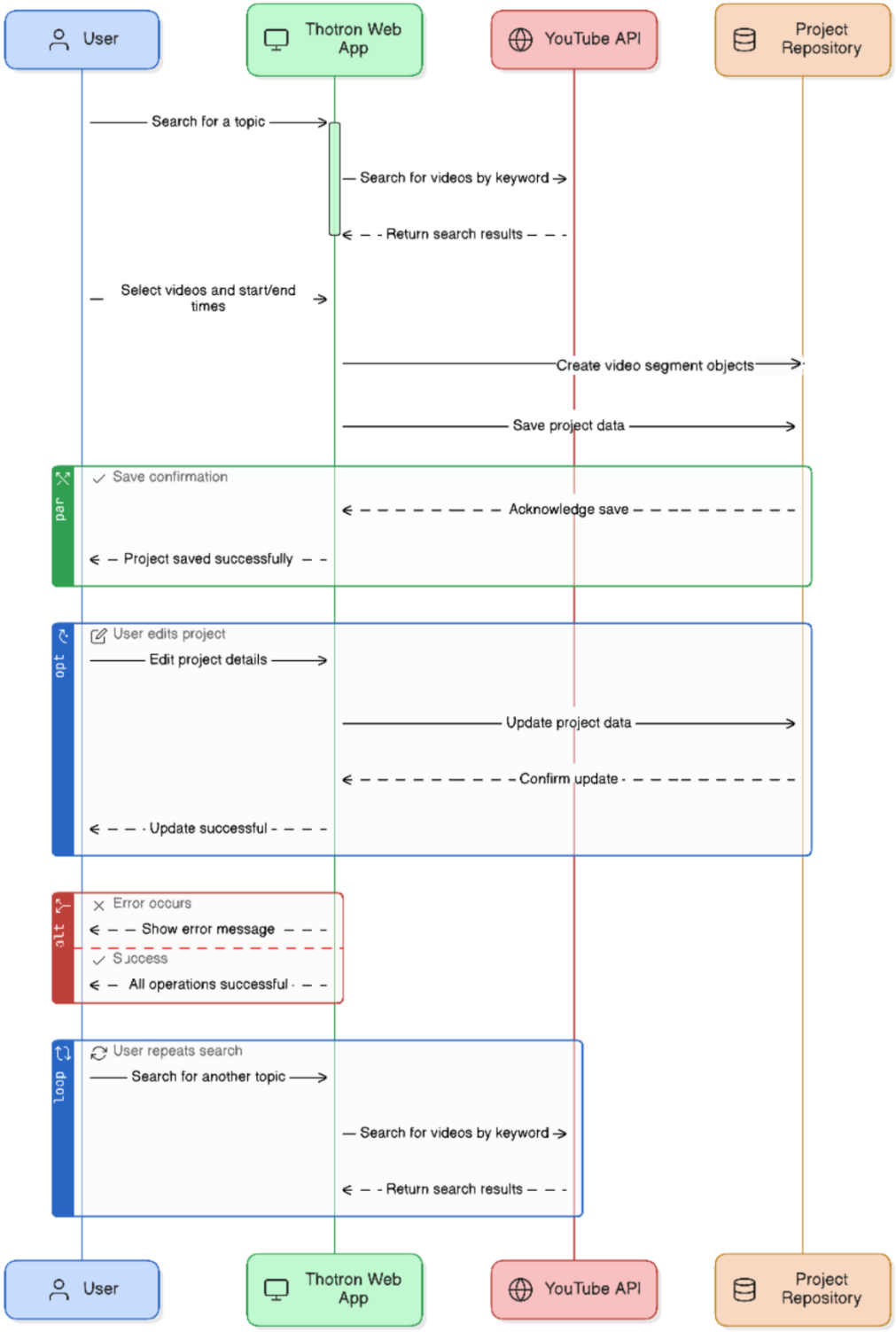

Figure 4-4: Thotron AI-NVC Manual Scenario Sequence Diagram

4.10.2. Auto Clipping Scenario

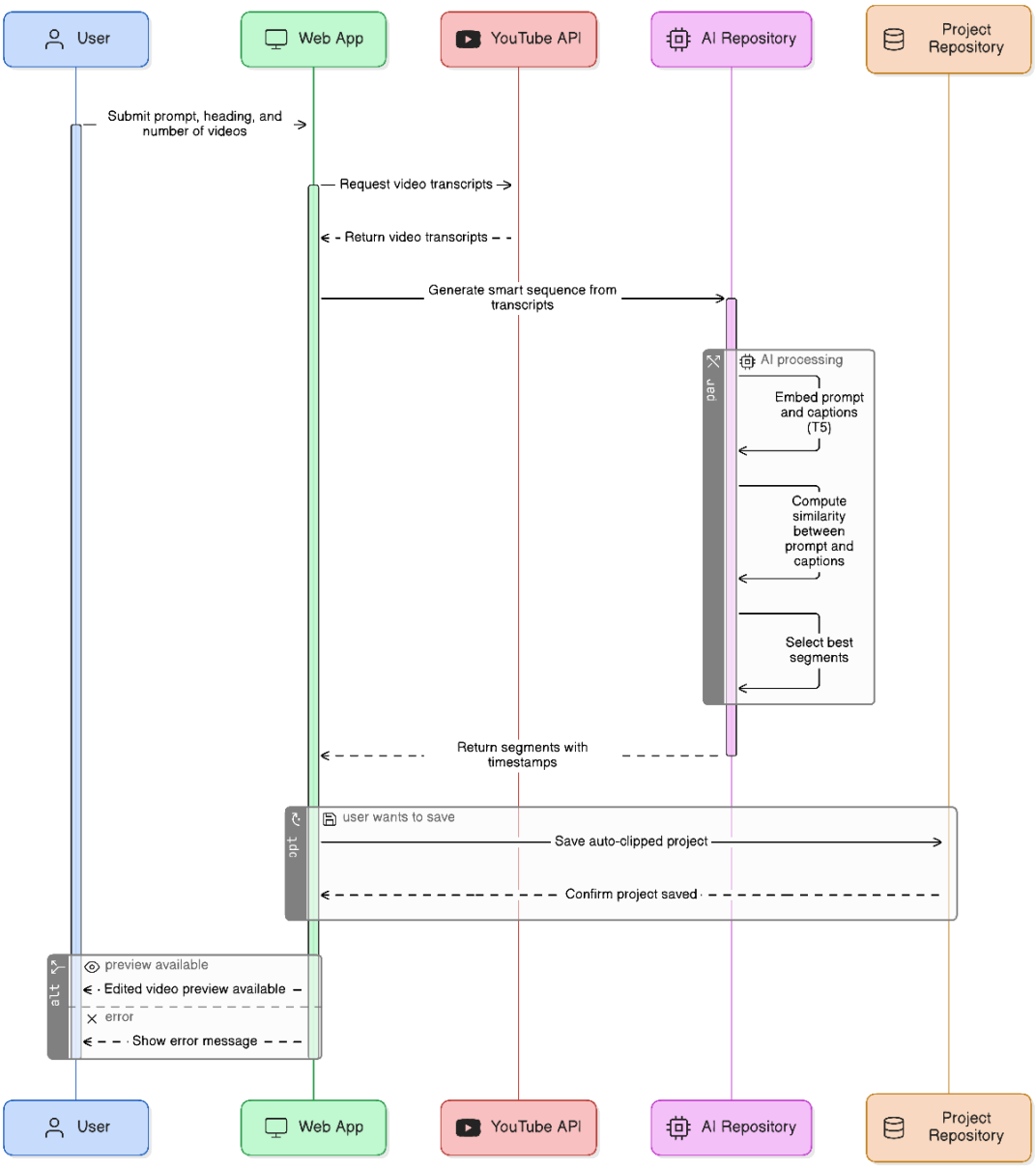

Figure 4-5: Thotron AI-NVC Auto Clipping Scenario Sequence Diagram

4.10.3. Summarization Scenario

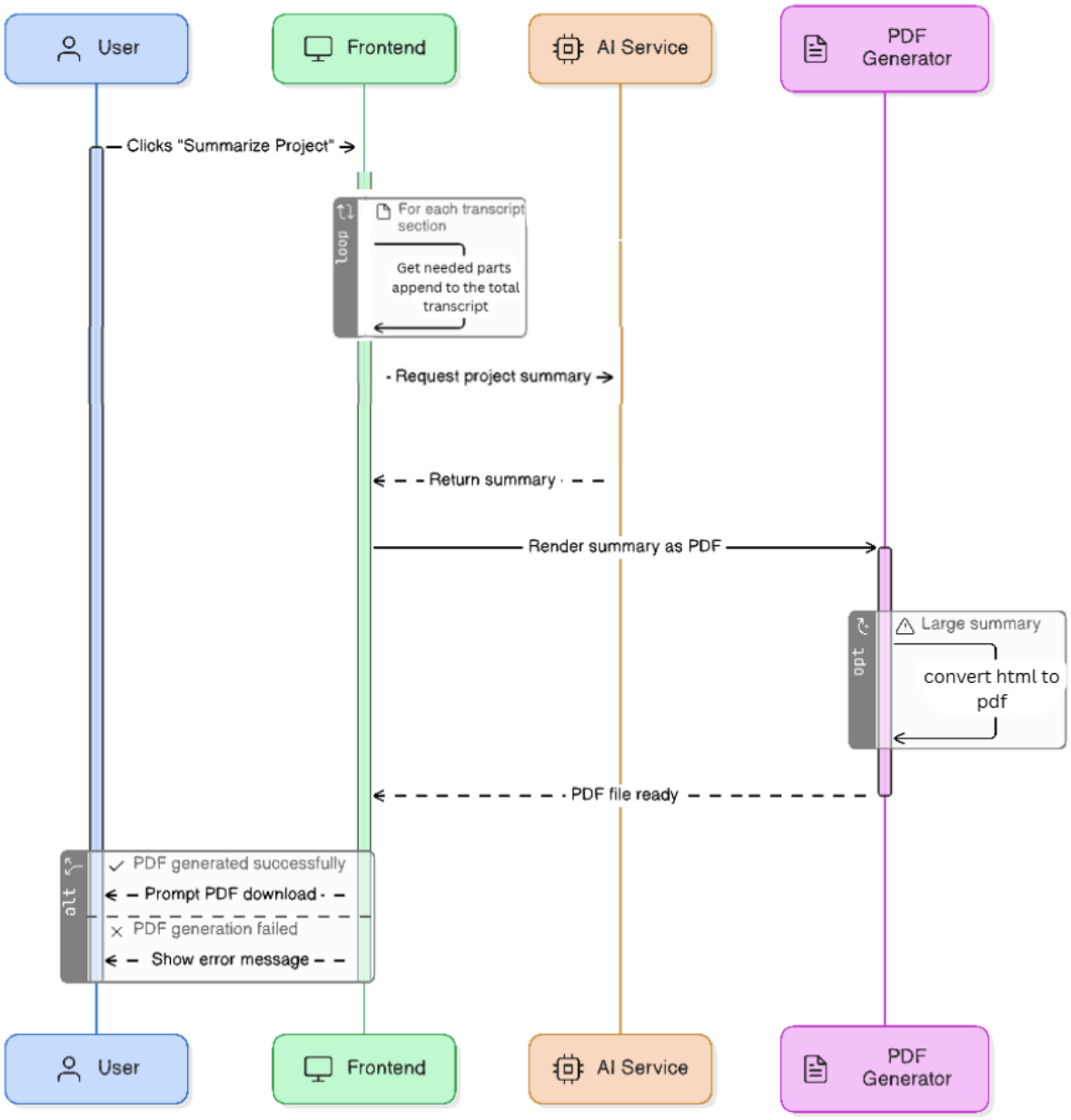

Figure 4-6: Thotron AI-NVC Summarization Scenario Sequence Diagram

4.10.4. Narration Scenario

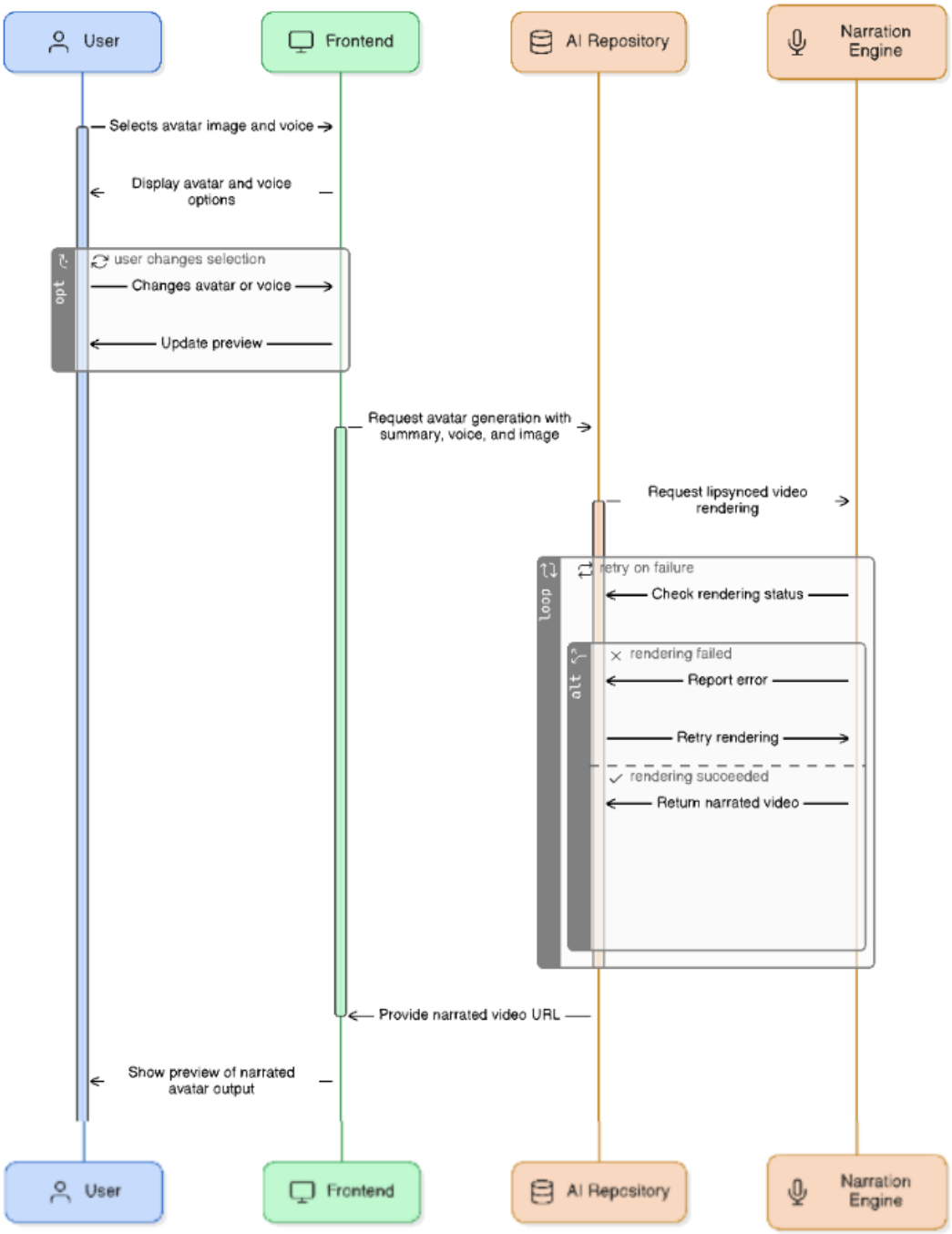

Figure 4-7: Thotron AI-NVC Narration Scenario Sequence Diagram

4.11. Gantt Chart

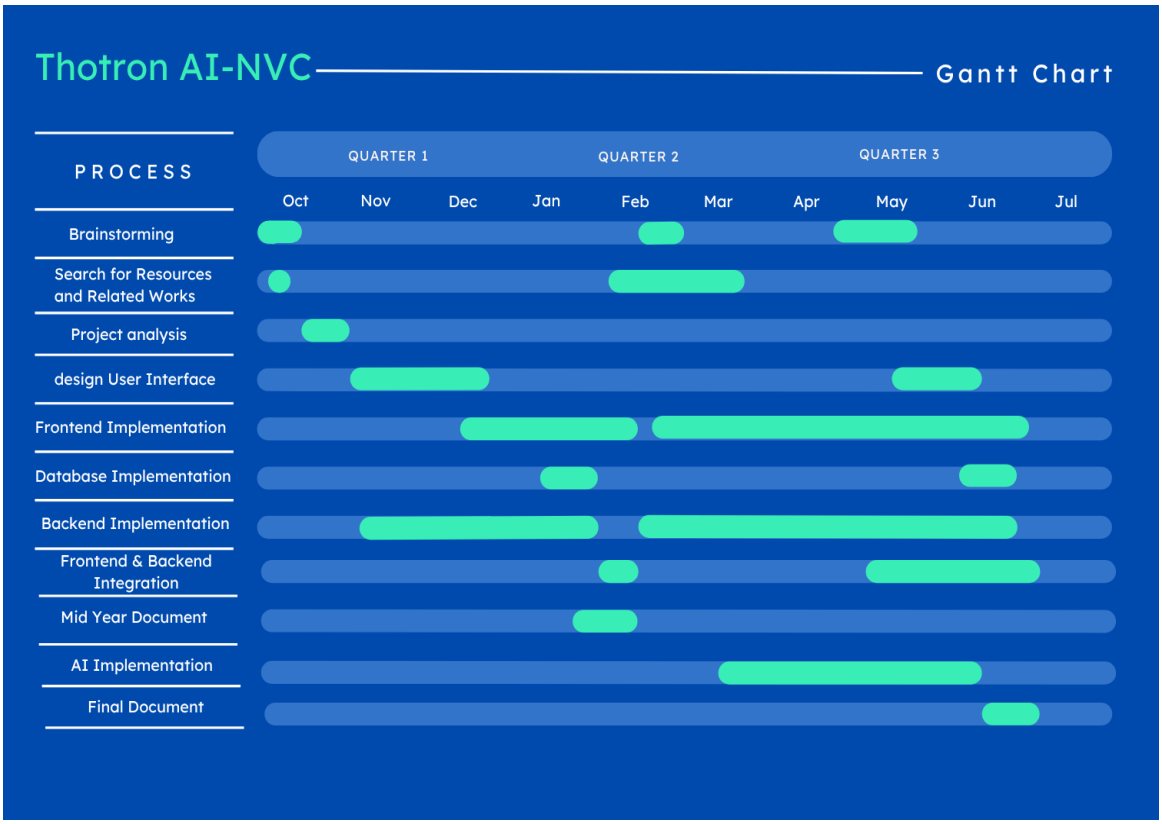

Figure 4-8: Thotron AI-NVC Gantt Chart.

Chapter 5: Thotron AI-NVC Design

5.1 Web UI Design

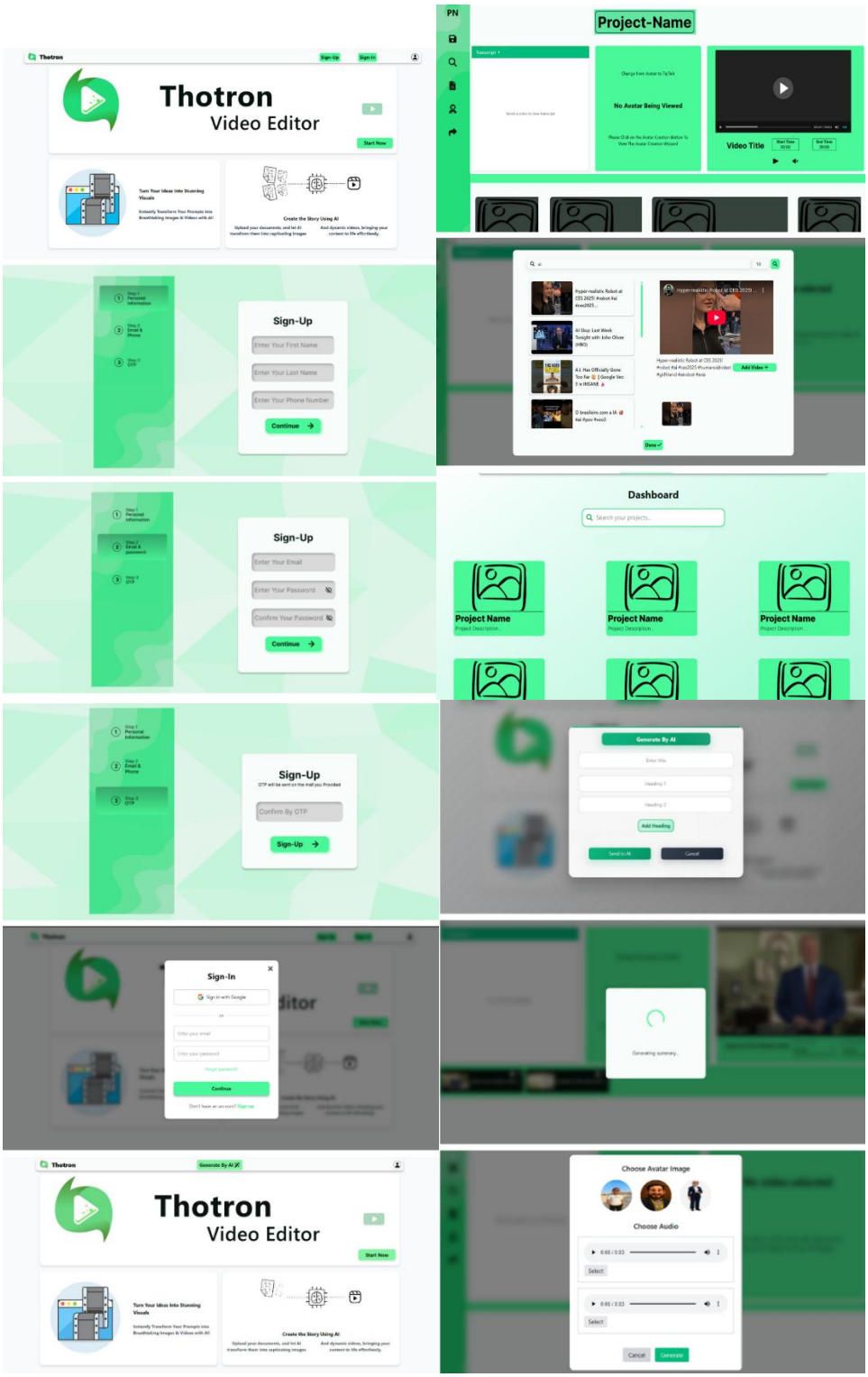

Figure 4-9: Thotron AI-NVC UI Design

References

- **FigmaUI:https://www.figma.com/design/9NAdttZgBQWqWugJ4HfATd/Thotron?node-id=0-1&t=YC7EmjiSYmmJCtti-0**
- **LLAMA 3.2-1B:https://huggingface.co/meta-llama/Llama-3.2-1B**
- **Text To Speech: https://huggingface.co/spaces/bestoai/text-to-video**
- **Gradio: https://huggingface.co/spaces/pragnakalp/Wav2lip-ZeroGPU**
- **Whisper: https://huggingface.co/openai/whisper-base**
- **T5: https://huggingface.co/google-t5/t5-base**