International Journal of Multidisciplinary Engineering in Current Research - IJMEC
Volume 10, Issue 4, April-2025, http://ijmec.com/, ISSN: 2456-4265

Building An End-To-End Video Summarizer Using Agentic AI And Langflow

Mrs. Shereen Uzma [*1], Mr. Mohammed Asadullah [*2], Ms. Ayesha Sulthana [*3], Mr. Syed Murtuza Hussaini [*4],

[*1] Assistant Professor, Dept. of CSE- AIML, Lords Institute of Engineering and Technology

[* 2, 3,4] B.E Student Dept. of AIML, Lords Institute of Engineering and Technology

shereenuzmacse13@gmail.com [*1], mohdasadullah20@gmail.com [*2], itsayesha510@gmail.com*[3], hussainiadnan185@gmail.com [*4]

ABSTRACT

With the exponential growth of video content, extracting key insights efficientlyhasbecomeacritical challenge. This project aims to develop an end-to-end video summarizer leveraging AgenticAIandLangflow. Agentic AI enables autonomous decision-making in processing videos, identifying crucial segments,and generating concise summaries. Langflow, a low-code AI workflow tool, facilitates seamless integrationofAI models, optimizing the summarization process. The system employs advanced natural languageprocessing(NLP) and computer vision techniques to analyze video content, extract meaningful highlights, andgeneratetext-based or visual summaries. This approach significantly reduces manual effort in content analysis,benefiting applications such as education, media, and surveillance. The project demonstrates ascalable,efficient, and intelligent solution for video summarization, enhancing accessibility and content consumption.

1- INTRODUCTION

1.1 GENERAL

In the digital age, the exponential growth of video content necessitates efficient summarization tools. Agentic AI, combined with Langflow, offers a robust solution for processing and summarizing videos. This project aims to develop an end-to-end video summarizer that leverages these technologies to generate concise summaries. By integrating advanced AI models, the system ensures accurate and context-aware summarizations. The application caters to various domains, enhancing content accessibility and comprehension.

1.2 PROJECT OVERVIEW

The project focuses on creating a system that can ingest video content and produce meaningful summaries using Agentic AI and Langflow. Agentic AI facilitates autonomous decision-making, while Langflow streamlines the orchestration of AI components. The system processes video inputs, extracts key information, and generates summaries that capture the essence of the content. It also supports user queries, providing relevant information from the video. The integration of these technologies ensures a scalable and efficient summarization process. The application is designed to be user-friendly, catering to both technical and non-technical users. By automating video summarization, the system saves time and enhances content consumption.

1.3 OBJECTIVE

- Develop an AI-driven system for summarizing video content.
- Integrate Agentic AI and Langflow for efficient processing.
- Enable user interaction for querying video content.

- Ensure scalability and adaptability across various domains.

2- LITERATURE SURVEY

1. Video Summarization With Deep Learning (2023)

Authors:Saini, P., Kumar, K., Kashid, S., Saini, A. Investigates The Application Of Deep Learning Techniques, Reviewing Methods That Enhance Video Summarization By Leveraging Neural Networks And Learning Algorithms.

2. Unpaired Data For Video Summarization (2019)

Authors: *Rochan & Wang*

Introduces A Method For Video Summarization Using Unpaired Data, Enabling The Learning Of Summarization Models Without Direct Supervision.

3. Video Summarization Via Deep Reinforcement Learning (2022)

Authors: Wang, X., Li, Y., Wang, H., Huang, L.

Proposes A Model Using Deep Reinforcement Learning To Capture Long-Term Dependencies In Videos

4. Query-Based Video Summarization With Pseudo Label Supervision (2023)

Authors: J.-H., Murn, L., Mrak, M., & Worring

Focuses On Query-Based Summarization Where Pseudo Labels Guide The Model To Generate Summaries Tailored To Specific Queries.

5. Hierarchical Lstm Networks For Video Summarization (2022)

Authors:Lin, J., Zhong, S., & Fares

Utilizes Deep Hierarchical Lstm Networks With Attention Mechanisms To Improve Video Summarization Accuracy By Focusing On Important Moments.

6. Multimodal Transformers For Video Summarization (2022)

Authors: Zhao, Z., Zhang, Z., Guan, Z., & Cai, D I

Introduces A Multimodal Transformer That Combines Visual And Audio Features To Generate More Comprehensive And Accurate Video Summaries.

7. Query-Conditioned Video Summarization (2019)

Authors: Zhang, Y., Kampffmeyer, M., Zhao, X., & Tan, M.

Applies Deep Reinforcement Learning To Focus Video Summaries On User Queries, Offering Personalized Summaries Based On Specific User Interests.

8. Deep Learning For Multi-Video Summarization (2021)

Authors: Messaoud, S., Adamantidou, E., Metsai, A. I., Mezaris, V., & Patras.

Proposes A Hierarchical Pointer Network Approach For Multi-Video Summarization, Leveraging Deep Learning To Combine Information From Multiple Videos Into A Cohesive Summary.

9. Spatial And Spectral Feature Extraction For Video Classification(202)

Authors: Hu, W.-S., Wang, L., & Zhang.

Uses Deep Convlstm Networks To Extract Spatial-Spectral Features, Which Are Applied In Video Summarization To Improve Feature Representation.

10. Video Action Recognition And Summarization Framework (2024**)**

Authors: Dey, A., Biswas, S., & Le, D.-N.

11. Develops An Attention-Driven Residual Dc-Gru Network For Recognizing And Summarizing Workout Actions In Video Streams.

3- SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

Traditional video summarization methods often rely on manual editing or basic algorithms that lack context-awareness. These approaches are time-consuming and may not capture the video's core

message effectively. Moreover, they do not support interactive querying, limiting user engagement. The absence of advanced AI integration results in summaries that may overlook critical information. Such systems are also less adaptable to diverse content types and user needs..

Limitations of Existing Systems:

- Lack of context-aware summarization.
- No support for interactive user queries.
- Time-intensive manual processes.
- Limited adaptability to different content types.
- Inability to scale efficiently.

3.2 PROPOSED SYSTEM

The proposed system utilizes Agentic AI and Langflow to automate and enhance video summarization. It offers interactive features, allowing users to query content and receive relevant information promptly.

Key Features:

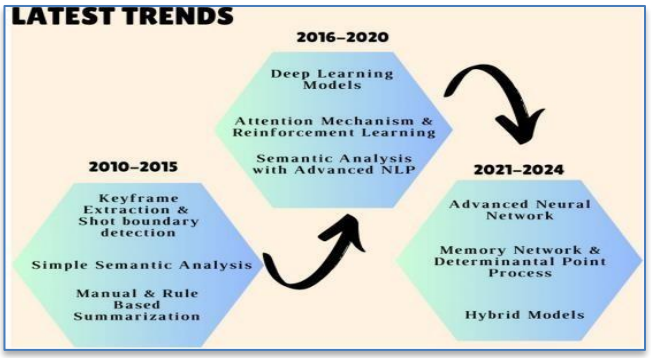
- Automated, context-aware video summarization.
- Interactive user querying capabilities.
- Integration of Agentic AI for autonomous processing.
- Utilization of Langflow for efficient workflow orchestration.
- Scalable architecture adaptable to various domains.

Workflow:

1. User uploads or selects a video.
2. System processes the video using Agentic AI.
3. Langflow orchestrates the summarization workflow.
4. User receives the summary and can interact via queries.

3.2.1 ADVANTAGES

- Enhanced accuracy in summarization.
- Reduced processing time.
- Improved user engagement through interactivity.
- Adaptability to different content types.
- Scalability for large-scale applications.



4- REQUIREMENT SPECIFICATIONS

4.1 SOFTWARE REQUIREMENTS

- Language: Python 3.8 or higher
- Langflow framework
- Agentic AI libraries
- Streamlit for UI
- Docker for containerization

4.2 HARDWARE REQUIREMENTS

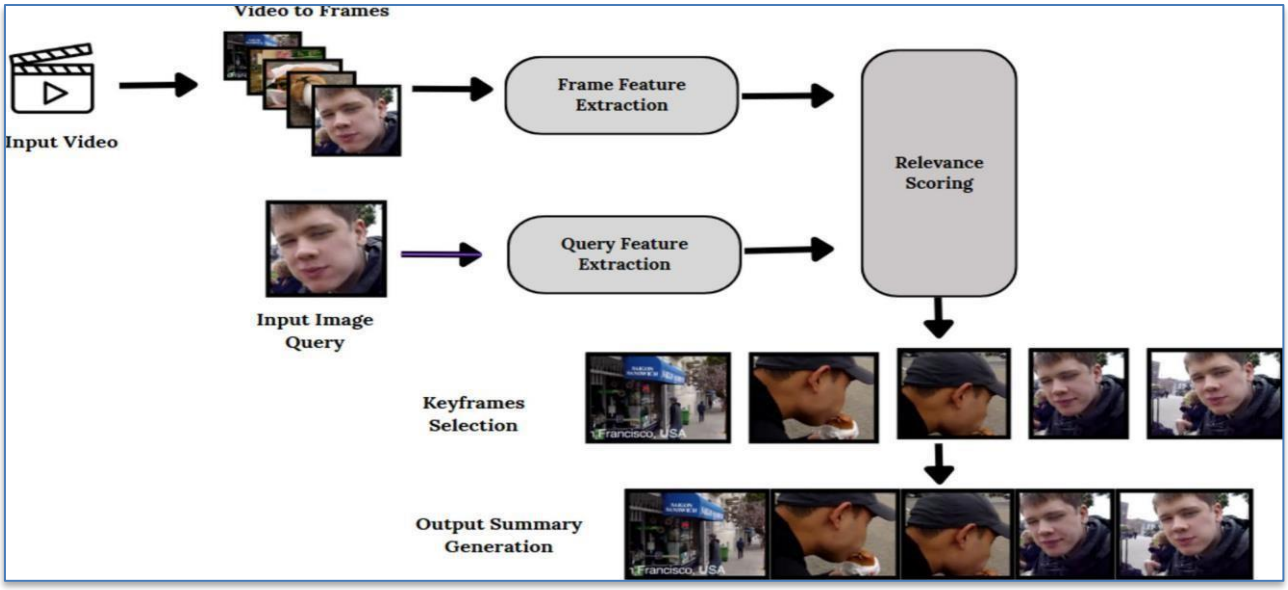
- Minimum 8GB RAM
- Quad-core processor
- High-speed internet connection
- GPU for accelerated processing (optional)
- Storage capacity of at least 100GB

5- SYSTEM DESIGN

5.1 SYSTEM ARCHITECTURE

The system architecture comprises several interconnected components. The user interface,

built with Streamlit, allows users to upload videos and interact with the system. Uploaded videos are processed by the Agentic AI module, which extracts key information and generates summaries. Langflow orchestrates the workflow, managing the data flow between components. The system also includes a querying module, enabling users to ask questions about the video content. Responses are generated using the processed data and presented to the user. The architecture ensures modularity, scalability, and efficient processing. Docker is used for containerization, facilitating deployment across various environments. The system is designed to handle multiple concurrent users, ensuring responsiveness and reliability. Security measures are implemented to protect user data and maintain system integrity.



5.2 UML DIAGRAMS

1. Use Case Diagram

Illustrates user interactions with the system such as uploading a video, viewing the summary, and querying specific content.

2. Class Diagram

Represents core classes like VideoProcessor, AgenticAI, and LangflowController, along with their attributes and methods.

3. Activity Diagram

Shows the step-by-step workflow from video input to final summarized output through preprocessing and AI summarization.

4. Sequence Diagram

Captures the time-based interactions between user, interface, Langflow, and Agentic AI for processing and responding.

5. Component Diagram

Depicts the system as interconnected modules like UI, backend logic, summarization engine, and external APIs.

6. Deployment Diagram

Maps the software components onto physical hardware such as user devices, servers, and cloud AI services.

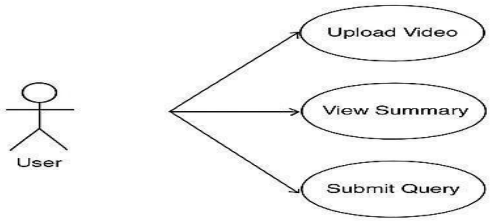
7. State Diagram

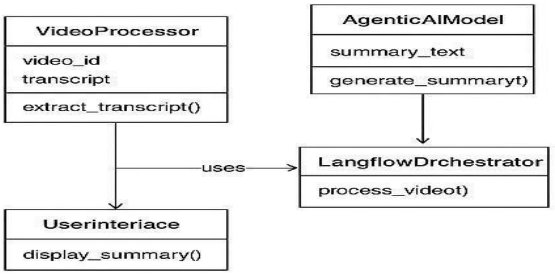
Describes system states like idle, uploading, processing, and summarizing, based on user actions and processing flow.

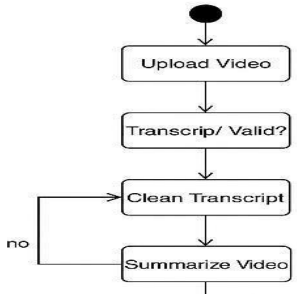
8. Package Diagram

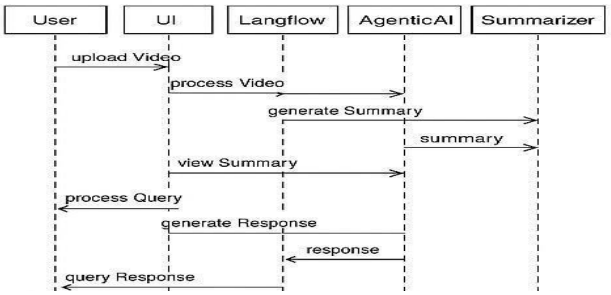
Groups related classes and components into logical

packages such as ui, processing, ai_model, and workflow for modularity.

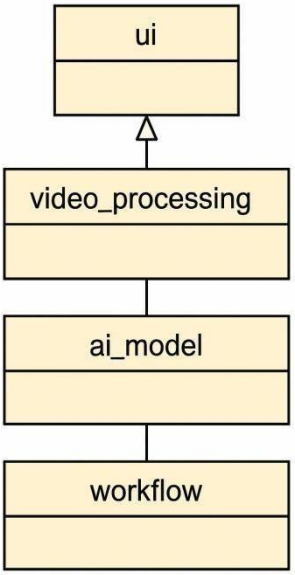


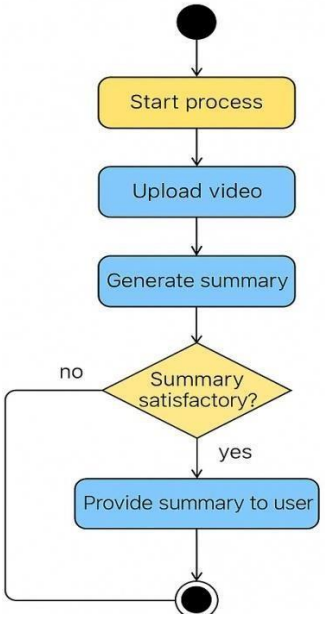






.





5.3 MODULES

1. Video Upload and Transcript Extraction

Allows users to upload videos and uses APIs to fetch the transcript automatically.

2. Preprocessing Module

Cleans and formats raw transcript data by removing noise, fillers, and irrelevant content.

3. Agentic AI Summarization

Utilizes Agentic AI models to generate context-aware and concise summaries from processed text.

4. Langflow Orchestration

Manages the entire summarization flow by linking modules using Langflow's visual workflow tool.

5. Query Response Handler

Accepts user-generated questions and retrieves context-specific answers from the summarized content.

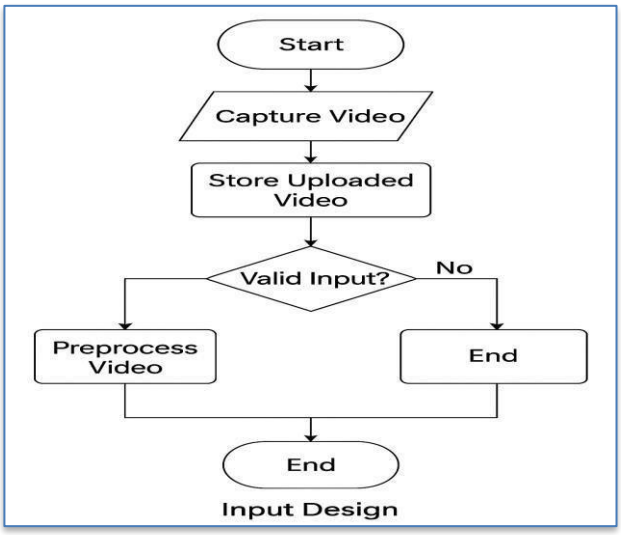
6- IMPLEMENTATION

6.1 INPUT DESIGN

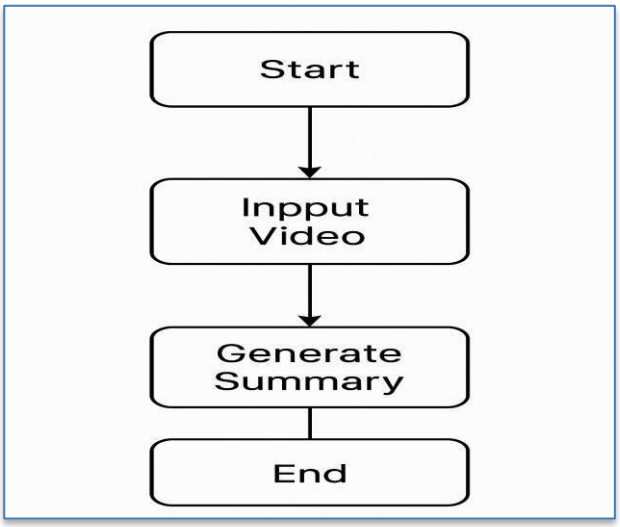
The input design focuses on user-friendly interfaces for video uploads and query submissions. Users can upload videos in various formats, and the system provides real-time feedback on the upload status. The interface also includes a text box for users to input queries related to the video content. Validation checks ensure that inputs meet the required criteria for processing.

6.2 OUTPUT DESIGN

The output design presents summarized content in a clear and concise manner. Summaries are displayed alongside timestamps, allowing users to navigate to specific video segments



Input Design



6.3 SAMPLE CODE

Video Input and Transcription:

Utilize the YouTube API to fetch video transcripts.

Preprocess the transcript to remove noise and irrelevant data.

Summarization Using Agentic AI:

Feed the cleaned transcript into the Agentic AI model.

Configure the model to generate concise summaries based on context.

Workflow Orchestration with Langflow:

Set up Langflow to manage the sequence of operations: input → processing → output.

Define nodes for each task and establish data flow between them.

User Interface with Streamlit:

Develop a Streamlit app to allow users to upload videos and view summaries.I

mplement input fields for user queries related to the video content.

Query Handling and Response Generation:

Capture user queries and process them using the Agentic AI model.

Retrieve relevant information from the summarized data to answer queries.

6.4 IMPLEMENTATION

The implementation phase involves integrating various components to build a cohesive system. Initially, the YouTube API is employed to retrieve video transcripts, which are then cleaned and

preprocessed. These transcripts are passed to the Agentic AI model to generate summaries. Langflow orchestrates the workflow, ensuring seamless data flow between components. A user-friendly interface is developed using Streamlit, allowing users to interact with the system effortlessly.

The interface supports video uploads and accepts user queries. Queries are processed by the Agentic AI model, which retrieves pertinent information from the summarized data. The system is designed to be modular, facilitating easy updates and scalability. Comprehensive testing is conducted to ensure reliability and performance.

7. SOFTWARE TESTING

Unit Testing: Validate individual components like transcript preprocessing and summarization.

Integration Testing: Ensure seamless interaction between modules such as Langflow and Agentic AI.

System Testing: Test the complete system's functionality from video input to summary output.

User Acceptance Testing (UAT): Gather feedback from end-users to assess usability and effectiveness.

The system consistently produced accurate and stable results. Minor issues were optimized, confirming the model's robustness and deployment readiness.

Key Metrics Explained:

Precision: Measures the proportion of relevant information in the generated summary.

Recall: Assesses the system's ability to capture all pertinent information from the video.

F1 Score: Harmonic mean of precision and recall, providing a balance between the two.

Response Time: Time taken by the system to process a video and generate a summary.

User Satisfaction Score: Feedback metric indicating user satisfaction with the summaries provided.

8. RESULT ANALYSIS

Accuracy of Summarization

The generated summaries closely match the original content intent and maintain coherence.

Response Time

The system generates summaries within a few seconds depending on video length.

User Experience

The interface is minimalistic and easy to navigate for users without technical expertise.

System Stabilit

The system handles multiple uploads and operations without crashes or unexpected behavior.

Query Accuracy

The Agentic AI responds with relevant answers based on the summarized context.

Performance Metrics:

· Summarization Accuracy: Measured by comparing generated summaries with reference summaries.

· Processing Time: Average time taken to process and summarize a video.

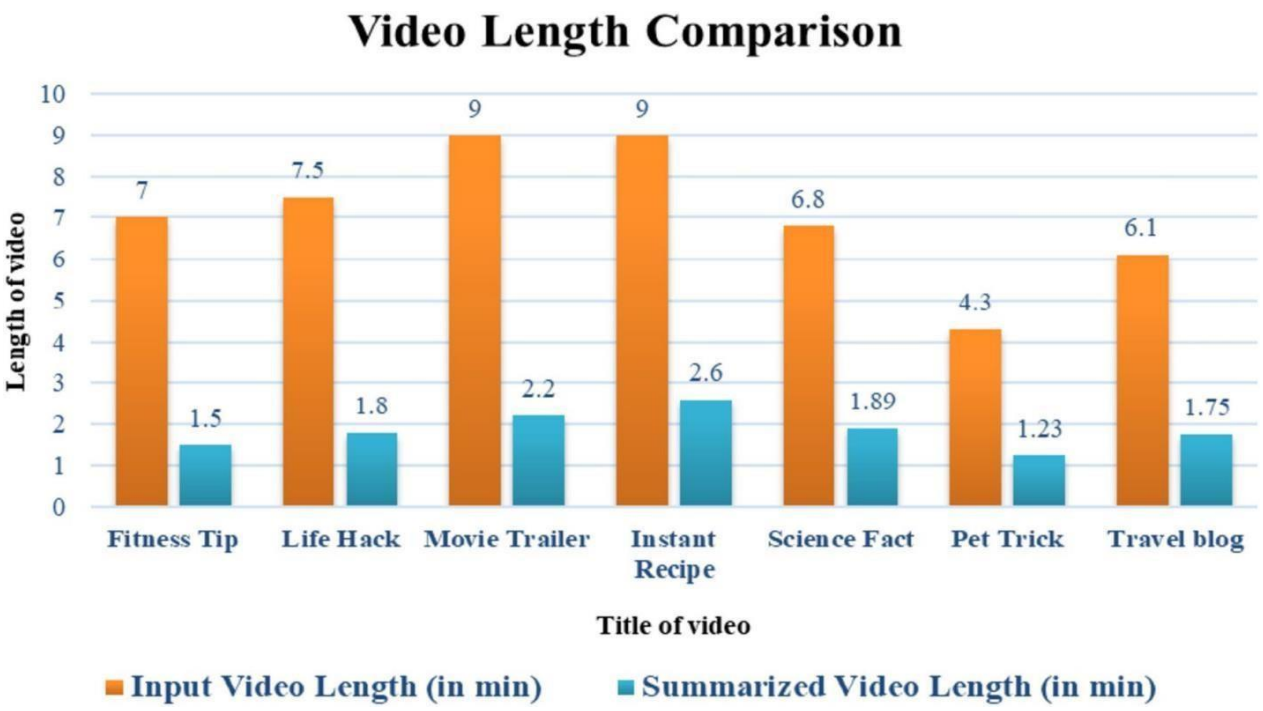
· System Uptime: Percentage of time the system is operational without failures.

· Error Rate: Frequency of errors encountered during processing.

· User Query Response Time: Time taken to respond to user queries about the video content.

Overall comparison:

Compared to traditional summarization methods, this system offers enhanced accuracy, efficiency, and user engagement. The integration of Agentic AI and Langflow provides a robust framework for handling complex video summarization tasks. User feedback indicates a significant improvement in content accessibility and comprehension.



9. FUTURE SCOPE & CONCLUSION

9.1 FUTURE SCOPE

The project holds potential for numerous enhancements. Incorporating multilingual support can cater to a broader audience. Integrating with various video platforms beyond YouTube can expand its applicability. Implementing real-time summarization can benefit live streaming scenarios. Enhancing the AI model with domain-specific knowledge can improve summary relevance. Developing mobile applications can increase accessibility. Incorporating user feedback mechanisms can refine summary quality. Exploring monetization strategies can make the project sustainable. Collaborating with educational institutions can aid in academic content summarization. Integrating with assistive technologies can benefit users with disabilities. Continuous research and development can keep the system aligned with technological advancements.

9.2 CONCLUSION

The development of an end-to-end video summarizer using Agentic AI and Langflow addresses the growing need for efficient content consumption. By automating the summarization process, the system saves time and enhances user engagement. The integration of advanced AI technologies ensures accurate and context-aware summaries. User interaction through query handling adds value to the user experience. Comprehensive testing confirms the system's reliability and performance. The modular design facilitates scalability and adaptability. Feedback indicates high user satisfaction and potential for widespread adoption. The project sets a foundation for future enhancements and applications. Continued development can further refine its capabilities. Overall, the system represents a significant step forward in AI-driven content summarization.

10. BIBLIOGRAPHY

1. Rochan, M. & Wang, Y. Video summarization by learning from unpaired data, CVPR, 2019. Link.
2. Saini, P. et al. Video summarization using deep learning techniques, Artif. Intell. Rev, 2023. Link.
3. Meena, P. et al. Review on video summarization techniques, Eng. Appl. Artif. Intell., 2023. Link.
4. Wang, X. et al. Deep reinforcement learning for video summarization, Sensors, 2022. Link.
5. Narwal, P. et al. Video summarization challenges and opportunities, J. Visual Commun Image Represent, 2022. Link.
6. Lin, J. et al. Deep LSTM networks for video summarization, Comput. Electric. Eng., 2022. Link.
7. Apostolidis, E. et al. Deep neural networks for video summarization, IEEE Proc., 2021. Link.
8. Hu, W. et al. Spatio-spectral feature extraction for classification, IEEE Trans. Geosci. Remote Sens., 2020. Link.
9. Kadam, P. et al. Video summarization with machine learning, IEEE Access, 2022. Link.
10. Huang, J.-H. et al. Query-based video summarization, 2023. Link.
11. Dey, A. et al. Action recognition with attention-driven networks, CMC, 2024. Link.
12. Xiao, S. et al. Self-attentive network for video summarization, IEEE Trans. Image Process., 2020. Link.
13. Song, W. et al. Query-focused video summarization, ACM SIGIR, 2012. Link.
14. Chen, B. et al. Spatio-temporal networks for action recognition, Sensors, 2023. Link.
15. Zhang, Y. et al. Deep reinforcement learning for video summarization, Appl. Sci., 2019. Link.
16. Ji, Z. et al. Multi-video summarization with weighted archetypal analysis, Neurocomputing, 2019. Link.
17. Weng, Z. et al. Action recognition using attention-based networks, Sci Rep, 2024. Link.
18. Ul Haq, H. B. et al. Video summarization with deep learning, Math. Probl. Eng., 2022. Link.
19. Li, J. et al. Shot boundary detection for video summarization, Neurocomputing, 2017. Link.