

Ai-Powered Virtual News Anchors: An Integrated Framework for Automated News Acquisition, Speech Synthesis and Lip-Synchronized Presentation

Ankesh Agrawal

B.Tech in Computer Science and Engineering [Artificial Intelligence and Machine Learning]

Lakshmi Narain College of Technology & Science

Abstract - The rapid advancement of artificial intelligence has transformed the way news content can be collected, processed, and presented. Traditional news broadcasting requires human anchors, production resources, and considerable preparation time, while existing digital news platforms primarily provide textual or audio-based information. This research presents an AI-powered Virtual News Anchor framework that integrates automated news acquisition, article extraction, text-to-speech synthesis, avatar-based video generation, and audio-driven lip synchronization into a single desktop application. The proposed system automatically retrieves recent news according to a user-specified topic and number of articles through Google News, extracts article content using newspaper3k, converts the resulting news text into speech using Google Text-to-Speech (gTTS), and generates a talking-avatar video using the Wav2Lip deep learning model. A Tkinter-based graphical user interface provides users with two modes of operation: automatic news retrieval and manually supplied news content.

The system was experimentally tested approximately 55–60 times using geopolitical news, Indian political news, trending news, and technology-related topics. Experiments were conducted using both automatically retrieved and manually entered news. The reported mean operating ranges were approximately 45–50 seconds for news fetching, 1–1.5 minutes for TTS, 1–

2.5 minutes for Wav2Lip, and 2–2.5 minutes for end-to-end generation. Approximately 94–96% of article extraction attempts were successful, while the reported video-generation success rate was 99%. The developer observed no visible lip-sync artifacts and qualitatively rated synchronization at 9/10. The study demonstrates the feasibility of integrating multiple AI and multimedia technologies into a low-cost, modular virtual news presentation system while identifying limitations related to internet dependency, processing time, source reliability, and the absence of formal human-subject evaluation.

Keywords—Artificial Intelligence, Virtual News Anchor, Automated Journalism, Text-to-Speech, Wav2Lip, MediaPipe, Facial Landmark Detection, Lip Synchronization, News Automation, AI-Generated Media, Digital News Presentation.

I. INTRODUCTION

News broadcasting has traditionally depended on human journalists, presenters, editors, camera operators, and production teams. Although this model provides human judgment and editorial oversight, it also requires time and operational resources. The emergence of artificial intelligence has introduced new possibilities for automating portions of the news production workflow.

Recent literature shows rapid growth in AI-supported journalism, including automated content production, personalization, and new forms of news presentation. Systematic reviews identify efficiency gains alongside unresolved concerns involving transparency, accountability, misinformation, trust, and the preservation of context. These issues become particularly important when AI is used not only to generate or transform news text but also to present information through a human-like virtual anchor.

Virtual presenters add an audiovisual layer to automated journalism. A virtual presenter can transform text-based information into a more familiar broadcast format, while audio-driven facial animation makes it possible to synchronize a synthetic face with generated speech. Wav2Lip, for example, was developed specifically for accurate speech-to-lip synchronization in

unconstrained videos.

This research develops and evaluates an end-to-end AI Virtual News Anchor prototype that combines news retrieval, article extraction, speech synthesis, avatar preparation, and neural lip synchronization in a single Python desktop application. The contribution is not a new TTS or lip-synchronization algorithm; instead, the contribution is the practical integration and evaluation of these technologies in an accessible news-presentation workflow.

II. LITERATURE REVIEW

A. Artificial Intelligence in Journalism

Research on automated journalism has evolved from structured-data-to-text generation toward broader generative-AI workflows. Recent systematic studies report substantial growth in AI journalism research and identify recurring themes including automation, journalistic role changes, credibility, transparency, accountability, misinformation, and human-machine collaboration.

B. Virtual and AI News Anchors

Research on virtual human presenters and AI anchors has examined audience engagement, credibility, trust, and the differences between synthetic and traditional news presentation. These studies suggest that virtual presenters can provide consistency and scalability while also creating challenges involving perceived authenticity and trust.

C. News-to-Video Systems

Recent systems have explored transforming news articles into scripts, storyboards, or audiovisual content. Such work demonstrates a growing research interest in connecting textual journalism with multimedia generation and in designing human-AI workflows for editorial production.

D. Audio-Driven Lip Synchronization

Prajwal et al. introduced Wav2Lip, a neural framework that uses a learned lip-sync discriminator to generate accurate mouth movements corresponding to target speech. The model is used in this project as a pretrained component rather than retrained from scratch.

E. Trust, Transparency and Synthetic Media

AI-generated news presentation introduces risks involving misinformation, source bias, deepfakes, and uncertainty about whether viewers are interacting with a human or synthetic presenter. Responsible deployment therefore requires clear disclosure, editorial oversight, and mechanisms for source verification.

III. Research Gap and Motivation

The literature contains substantial work on individual components such as automated journalism, text generation, virtual presenters, speech synthesis, and lip synchronization. However, the practical integration of these components into a lightweight, GUI-driven desktop workflow remains a useful engineering research problem, especially for a system that supports both automatically retrieved and manually supplied news.

The research gap addressed in this study is therefore system-level integration and empirical prototyping rather than algorithmic novelty. The proposed framework connects news retrieval and extraction with speech generation and a neural lip-synchronization stage, providing an end-to-end route from news text to a playable virtual-anchor video.

IV. RESEARCH QUESTIONS AND OBJECTIVES

- RQ1: Can existing AI and multimedia technologies be integrated into a functional end-to-end virtual news presentation system?
- RQ2: How reliably can the proposed system retrieve and extract news articles from online sources?
- RQ3: What processing time is required to transform news text into a lip-synchronized virtual-anchor video?
- RQ4: Can Wav2Lip provide visually synchronized lip movement for generated news speech within the proposed system?

The objectives are to automate news acquisition, extract article content, convert news text into speech, synchronize speech with a virtual presenter, provide a user-friendly GUI, and experimentally assess reliability and processing performance.

V. PROPOSED SYSTEM ARCHITECTURE

End-to-end architecture of the proposed AI Virtual News Anchor

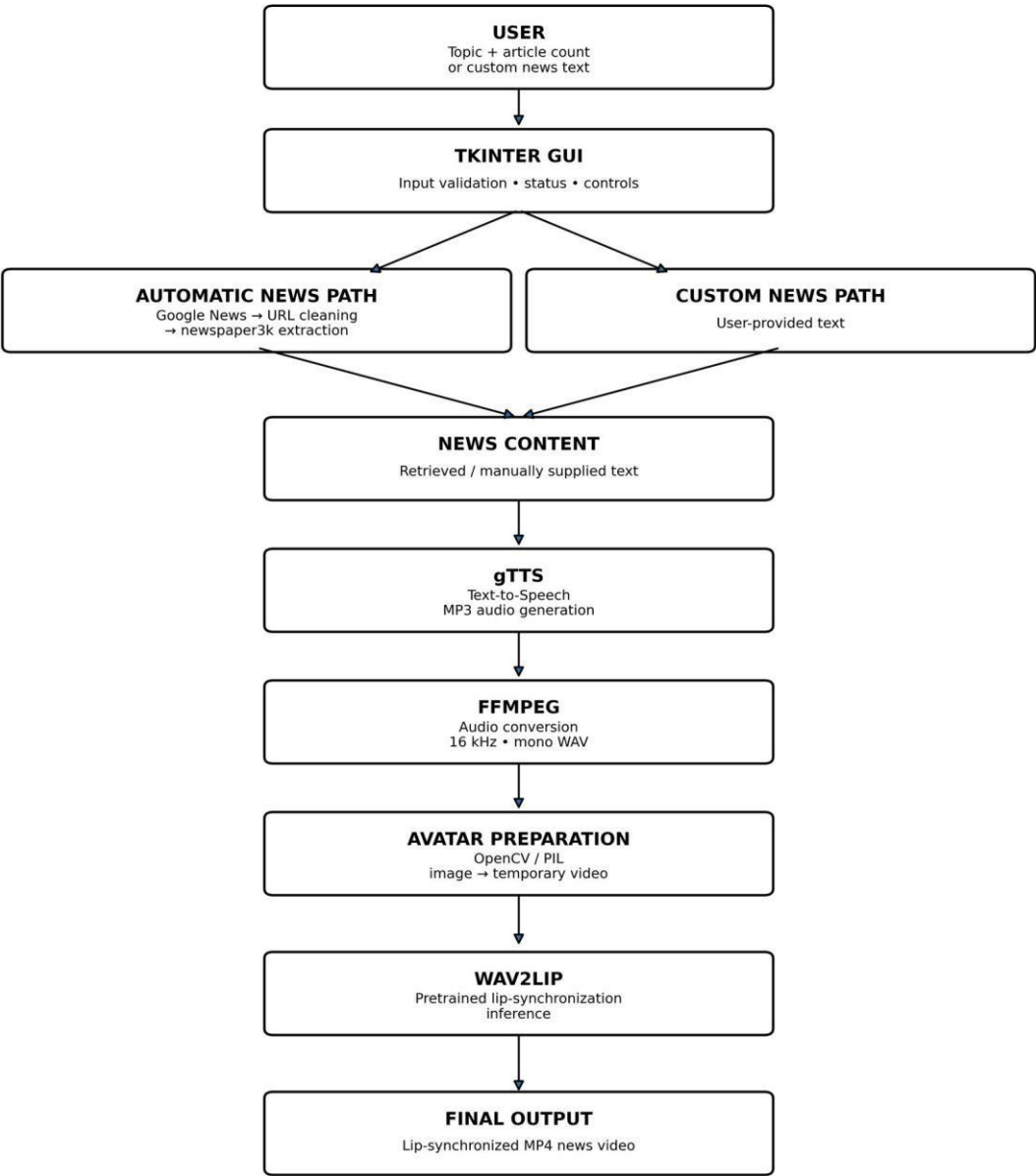


Fig. 1. End-to-end architecture of the proposed AI Virtual News Anchor.

The system begins with a user interaction layer implemented in Tkinter. Users can either specify a topic and article count for automatic news retrieval or enter custom news text. The two paths converge at the text-processing stage, after which gTTS produces speech, FFmpeg converts the audio into a Wav2Lip-compatible WAV format, an avatar image is prepared as a video stream, and Wav2Lip generates the final lip-synchronized MP4 output.

VI. TECHNOLOGY SELECTION

Component	Technology	Purpose
Programming	Python 3.11.6	Core development
GUI	Tkinter	User interaction and workflow control

News retrieval	Google News / GoogleNews	Topic-based news discovery
Article extraction	newspaper3k	Download and parse article content
TTS	gTTS	Convert text into speech
Image/video processing	OpenCV and PIL	Avatar and media preparation
Facial landmark detection	MediaPipe (Face Mesh)	Detect lip, chin, and cheek landmarks to define the avatar's lip-sync region
Audio processing	FFmpeg	Convert MP3 to 16-kHz mono WAV
Lip synchronization	Wav2Lip	Generate speech-synchronized mouth movement
Deep learning runtime	PyTorch	Run Wav2Lip inference

VII. DATA ACQUISITION AND PREPARATION

The project does not train a new supervised model on a fixed labeled dataset. Instead, it uses live news retrieval. Experiments covered geopolitical news, Indian politics, trending news, and technology. The number of requested articles was user-controlled through the GUI, with approximately 3–4 articles typically processed per automatic run.

The automatic retrieval stage searches for the requested topic, retrieves article URLs, cleans the URLs, downloads articles, and extracts their content using newspaper3k. The developer manually verified the extracted news and reported an observed article-extraction success rate of approximately 94–96%. The reported 4–6% failures occurred during article retrieval/extraction and should be interpreted as observed prototype behavior rather than a universal web-extraction guarantee.

The system also supports custom text, enabling the audiovisual generation pipeline to be tested independently of the online news retrieval stage.

VIII. SYSTEM IMPLEMENTATION

A. Graphical User Interface

A Tkinter-based GUI provides topic input, article-count selection, custom-news input, controls for automatic retrieval and custom text, status feedback, and video playback. The GUI forms the main interaction layer between the user and the processing modules.

B. Automatic News Retrieval

For automatic mode, the system searches Google News according to a user-defined topic and article count. Retrieved URLs are cleaned and passed to newspaper3k for article downloading and parsing. Successfully extracted articles are combined into the text used by the downstream audiovisual pipeline.

C. Custom News Input

Users can bypass online retrieval by supplying their own news text. This mode is useful for controlled testing and demonstrates that the audiovisual pipeline is independent of the news-source acquisition stage.

D. Text-to-Speech

The generated news text is converted into MP3 speech using gTTS. The observed TTS time was approximately 1–1.5 minutes under the tested conditions.

E. Avatar and Audio Preparation

The system validates and loads the avatar image using PIL/OpenCV-based handling. FFmpeg converts the generated MP3 into 16-kHz mono WAV audio, which is then supplied to the Wav2Lip inference stage.

Prior to lip-synchronization, a facial landmark detection step based on MediaPipe Face Mesh [10] is used to locate key landmarks on the avatar image, including points along the lips, chin, and cheeks. These landmarks are used to compute a bounding region that localizes the mouth area on the avatar's face. This region guides where the generated lip movements are applied during Wav2Lip inference, helping ensure that the synthesized mouth motion is correctly aligned with the avatar's facial geometry rather than being applied to a fixed or manually specified area.

F. Lip Synchronization and Video Generation

A pretrained Wav2Lip checkpoint is used to synchronize the avatar's mouth movement with the generated speech. The final output is stored as an MP4 video and can be played from the application.

IX. EXPERIMENTAL METHODOLOGY

Approximately 55–60 experimental runs were performed. Tests covered geopolitical news, Indian political news, trending news, and technology topics. Both automatically retrieved news and manually supplied text were tested. The primary evaluation dimensions were article extraction reliability, processing time, video-generation reliability, and qualitative lip synchronization.



Fig. 2. Generated virtual news-anchor output used in the project.

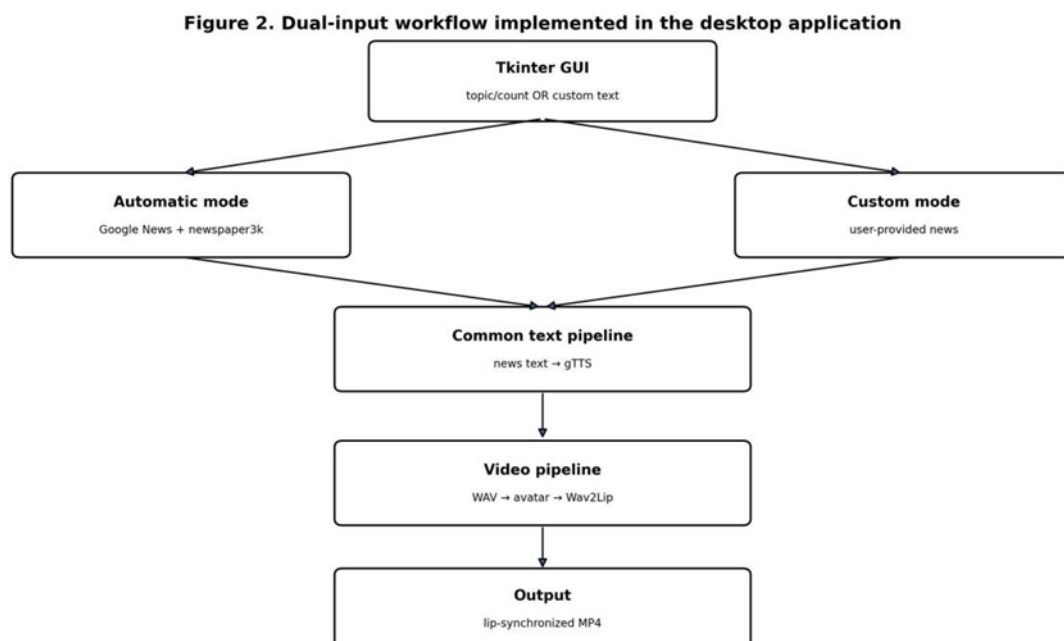


Fig. 3. Dual-input workflow of the implemented desktop application.

X. EVALUATION METRICS

Article extraction success rate is defined as the proportion of article extraction attempts that completed successfully. Video-generation success rate is the proportion of video-generation attempts that produced the required output. Processing time is recorded for the news-fetching, TTS, Wav2Lip, and overall generation stages.

Because the project did not train a new classifier or summarizer, classification metrics such as accuracy, precision, recall, and F1-score are not directly applicable to the complete system. The present evaluation instead focuses on functional reliability and processing performance.

The developer's 9/10 lip-synchronization score is reported as a qualitative assessment and is not treated as a statistically validated human-evaluation score.

XI. RESULTS

The reported mean operating ranges were approximately 45–50 seconds for news fetching, 1–1.5 minutes for TTS, 1–2.5 minutes for Wav2Lip, and 2–2.5 minutes for total end-to-end generation. The overall observed minimum and maximum generation times were approximately 1 minute and 3 minutes, respectively.

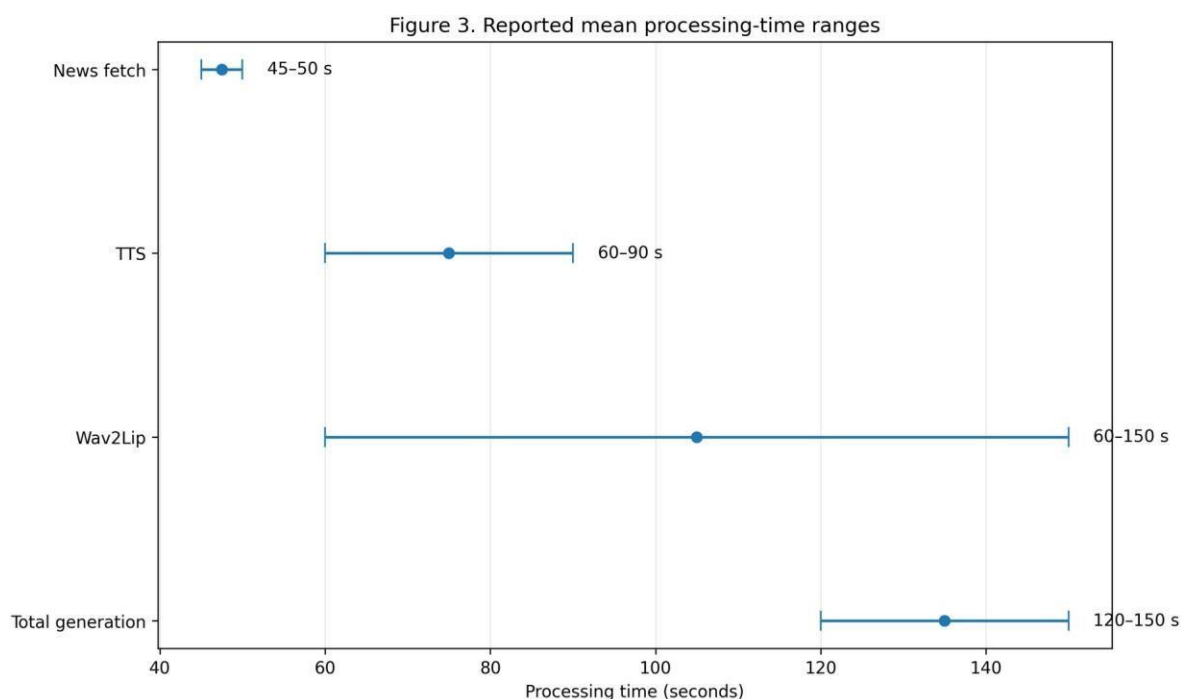


Fig. 4. Reported mean processing-time ranges. Ranges are shown as supplied from the experimental observations.

The article extraction success rate was reported at approximately 94–96%, while the video-generation success rate was reported at 99%. Wav2Lip was reported to work successfully during the conducted tests, with no noticeable visual artifacts.

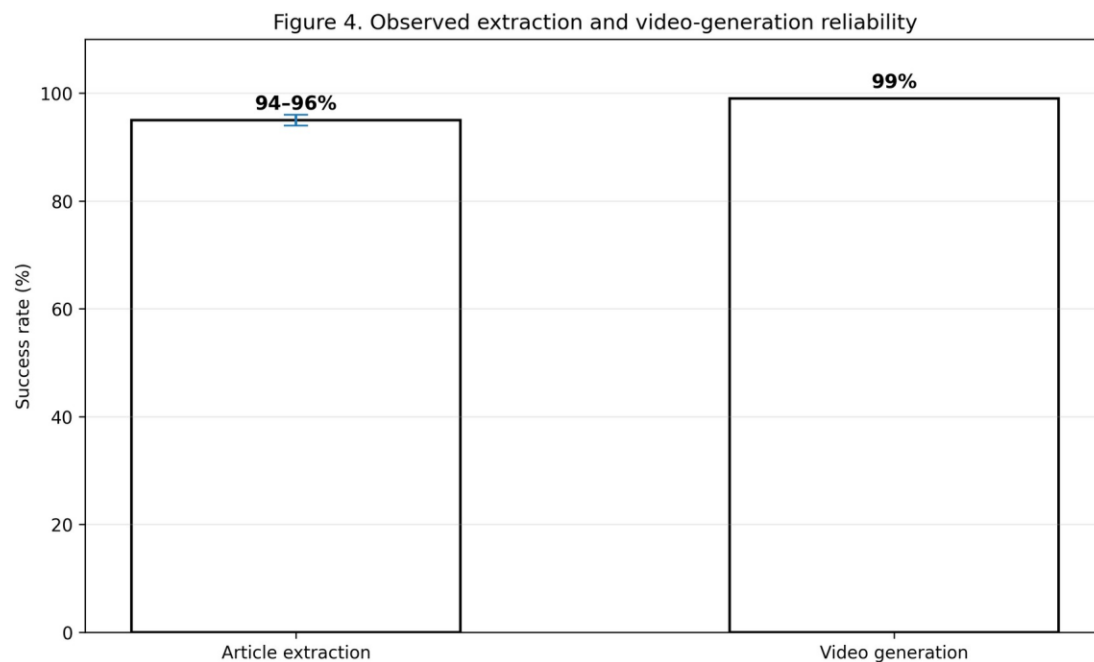


Fig. 5. Observed extraction and video-generation reliability.

Evaluation parameter	Observed result
Experimental runs	Approximately 55–60
News categories	Geopolitical, Indian politics, trending, technology
Articles per automatic run	Typically 3–4, user-controlled
Mean news-fetching time	45–50 s
Mean TTS time	1–1.5 min
Mean Wav2Lip time	1–2.5 min
Mean total generation time	2–2.5 min
Overall observed time	Approximately 1–3 min
Article extraction success	94–96%
Video-generation success	99%
Lip synchronization	Developer qualitative assessment: 9/10
Visible artifacts	None observed

XII. DISCUSSION

The results demonstrate that the proposed system can integrate multiple independent technologies into a functional virtual news presentation pipeline. The observed 94–96% extraction success rate indicates that the online retrieval and parsing process generally operated successfully during the tests, although failures remain possible because the system depends on external web pages and network conditions.

The reported 99% video-generation success rate indicates high functional reliability of the downstream video pipeline in the conducted experiments. Wav2Lip also produced visually synchronized results without noticeable artifacts according to the developer's observations. However, the 9/10 synchronization assessment should be interpreted as a qualitative developer judgment rather than a controlled perceptual evaluation.

Processing time remains a limitation. The Wav2Lip stage is computationally intensive, and network-dependent stages introduce additional variability. The reported component-time ranges should not be arithmetically summed to reconstruct the end-to-end mean because the measurements were obtained as operational estimates under varying input sizes and network conditions.

The current implementation should therefore be characterized as an automated or near-automated video-generation prototype rather than a real-time broadcasting system. It also does not currently implement an independent fact-checking or AI summarization stage; retrieved article content is extracted and passed to the audiovisual pipeline.

XIII. ETHICAL CONSIDERATIONS

- Transparency: The final system should clearly disclose that the presenter is AI-generated.
- Misinformation: Online source content may contain errors, bias, or incomplete context, and the current prototype does not independently fact-check claims.
- Synthetic-media misuse: Talking-avatar technology can be misused for impersonation or fabricated statements.
- Human oversight: Sensitive political, emergency, medical, and other high-impact news should receive human editorial review before publication.
- Copyright and attribution: Production deployment should respect source licensing, attribution, and applicable copyright requirements.

XIV. LIMITATIONS

- Dependence on internet connectivity for news retrieval and gTTS.
- Processing time is not suitable for strict real-time broadcasting.
- No dedicated automatic fact-checking mechanism.
- No dedicated AI summarization model in the current implementation.
- The presenter is based on a static avatar image, limiting natural body and facial movement.
- No formal human-subject evaluation was conducted.
- Current experiments focus primarily on English-language news.

XV. FUTURE SCOPE

- Integrate transformer/LLM-based summarization to create concise broadcast-ready scripts.
- Add multi-source fact verification and source credibility scoring.
- Support Hindi and other regional Indian languages.
- Introduce emotion-aware voice and facial-expression control.
- Replace the static avatar with a fully animated 2D/3D digital human.
- Optimize inference and GPU utilization for faster or near-real-time generation.
- Introduce a human-in-the-loop editorial approval stage.
- Conduct a controlled user study measuring trust, credibility, naturalness, comprehension, and lip synchronization.

XVI. CONCLUSION

This research presented an AI-powered Virtual News Anchor system integrating automated news acquisition, article extraction, speech synthesis, avatar preparation, and neural lip synchronization in a single Python desktop application. The system supports both automatically retrieved and manually supplied news and provides a GUI-driven workflow from news text to playable lip-synchronized video.

Approximately 55–60 experimental runs were conducted across geopolitical, Indian political, trending, and technology topics. The observed article extraction success rate was approximately 94–96%, and the reported video-generation success rate was 99%. Mean operating ranges were approximately 45–50 seconds for news fetching, 1–1.5 minutes for TTS, 1–2.5 minutes for Wav2Lip, and 2–2.5 minutes for total generation. The developer observed no noticeable lip-sync artifacts and qualitatively rated synchronization at 9/10.

The study demonstrates the practical feasibility of integrating established AI and multimedia components into an end-to-end virtual news presentation prototype. The principal contribution is the implementation and integration of these components, including a dedicated GUI and Wav2Lip-based synchronization workflow. Future work should strengthen factual verification, summarization, multilingual support, human evaluation, and real-time performance.

REFERENCES

- [1] A. F. Sonni, H. Hafied, I. Irwanto, and R. Latuheru, "Digital Newsroom Transformation: A Systematic Review of the Impact of Artificial Intelligence on Journalistic Practices, News Narratives, and Ethical Challenges," *Journal. Media*, vol. 5, no. 4, pp. 1554–1570, 2024.
- [2] "Artificial intelligence and journalism: A systematic bibliometric and thematic analysis of global research," *Computers in Human Behavior Reports*, vol. 20, 100830, 2025.
- [3] K. R. Prajwal, R. Mukhopadhyay, V. P. Namboodiri, and C. V. Jawahar, "A Lip Sync Expert Is All You Need for Speech to Lip Generation in the Wild," *Proc. 28th ACM International Conference on Multimedia*, pp. 484–492, 2020.
- [4] "Generative AI and the New Landscape of Automated Journalism: A Systematized Review of 185 Studies (2012–2024)," *Journal. Media*, 2026.
- [5] S. Wang et al., "ReelFramer: Human-AI Co-Creation for News-to-Video Translation," *arXiv preprint*, 2023.
- [6] T. Jain, Y. Gao, S. Vanga, and K. Singla, "News Reporter: A Multi-lingual LLM Framework for Broadcast T.V News," *arXiv preprint*, 2024.
- [7] L. Hawthorne, "Talking Heads: The Use of Virtual Human Presenters for the Delivery of News Content," *Doctoral thesis*, City, University of London, 2024.
- [8] I. Chae and L. Kim, "The Impact of AI Anchor News: The Comparison of AI vs. Human Anchor Programs in News Industry," *EMAC Conference Proceedings*, 2024.
- [9] Reuters Institute for the Study of Journalism, "Digital News Report 2024: Public Attitudes Towards the Use of AI in Journalism," *University of Oxford*, 2024.
- [10] C. Lugaresi et al., "MediaPipe: A Framework for Building Perception Pipelines," *arXiv preprint arXiv:1906.08172*, 2019.