
VerifyStream

AI-Powered Video Content Verification¹

Kailash Kumar
Balasubramaniyam

With
Sage & CeSIA & CivAI & Apart Research

Abstract

VerifyStream is an innovative application that leverages advanced AI technologies to verify the truthfulness of content in YouTube videos. By converting audio to text, analyzing core concepts, and cross-referencing information with reliable online sources, VerifyStream provides users with a powerful tool to distinguish fact from fiction in online content. This project demonstrates the potential of AI in combating misinformation while also highlighting the ethical concerns surrounding AI's dual-use nature in both verifying and potentially creating convincing fake news. VerifyStream represents a significant step forward in empowering users to critically evaluate online information, addressing the pressing issue of misinformation in the digital age.

Keywords: AI, content verification, misinformation, fact-checking, YouTube, natural language processing

1. Introduction

In an era where misinformation spreads rapidly through digital channels, the need for efficient and accurate content verification tools has never been more critical. VerifyStream addresses this need by providing an accessible means for users to verify the truthfulness of YouTube video content. Our project contributes to the

¹ Research conducted at the AI risks and capabilities demonstrations jam, 2024

field of AI-powered fact-checking by demonstrating a practical application that combines audio processing, natural language understanding, and information retrieval technologies.

The importance of VerifyStream lies in its potential to:

1. Empower users to make informed decisions about the content they consume
2. Reduce the spread of misinformation by providing quick, AI-assisted fact-checking
3. Raise awareness about the dual nature of AI in both combating and potentially creating misinformation

By showcasing the capabilities and limitations of AI in content verification, VerifyStream serves as a visceral demonstration of both the promise and perils of advanced AI technologies in the realm of information integrity.

2. Overview

VerifyStream operates through a series of interconnected AI-driven processes:

1. Video Processing: The app takes a YouTube video link as input and downloads the audio content.
2. Speech-to-Text Conversion: Advanced AI converts the audio into text format.
3. Content Analysis: AI analyzes the core concepts and claims made in the video.
4. Prompt Generation: Based on the analysis, the system generates targeted prompts for fact-checking.
5. Web Scraping and Information Retrieval: The app searches the internet for relevant information related to the generated prompts.
6. Verification: An AI model compares the video's claims against the retrieved information from reliable sources.
7. Results Presentation: The app provides a clear verdict on the truthfulness of the video content.

[\[Link to a 2-minute video demonstration of VerifyStream\]](#)

[\[Try the app here\]](#) (Please be noted that YouTube asks for sign in while trying the app directly from Streamlit cloud. It is advisable to try it locally. More information on how to how to try it locally is shared in github)

We chose this approach to create a user-friendly tool that leverages the power of AI across multiple domains (speech recognition, natural language processing,

information retrieval, and analysis) to provide a comprehensive fact-checking solution. The step-by-step process ensures transparency and allows for potential improvements at each stage.

3. Code

Our project is built using a combination of open-source libraries and custom code. The main components include:

1. YouTube-DL for video downloading
2. OpenAI's Whisper for speech-to-text conversion
3. Custom natural language processing models (GPT-4omni) for content analysis and prompt generation
4. Google search api for searching the web
5. Web scraping tools (e.g., BeautifulSoup) for information retrieval
6. A fine-tuned language model (based on GPT-4omni or similar) for claim verification
7. A fine-tuned vision model (based on Dall E2 or similar) for image generation

The code is organized into modular components, each handling a specific part of the verification process. This architecture allows for easy updates and improvements to individual components without affecting the entire system.

GitHub repository: <https://github.com/kailash19961996/Verify-Stream>

[\[Try the app here\]](#) (Please be noted that YouTube asks for sign in while trying the app directly from Streamlit cloud. It is advisable to try it locally. More information on how to how to try it locally is shared in github)

We chose to use OpenAI's API for some components due to its robust performance in natural language tasks. However, we've implemented safeguards to ensure responsible use and to mitigate potential biases.

4. Discussion and Conclusion

VerifyStream demonstrates the potential of AI in combating misinformation, but it also raises important ethical considerations. While the app provides a powerful tool for fact-checking, it simultaneously highlights the ease with which AI could be used to create and spread convincing fake news.

Key observations and future directions:

1. Improving accuracy: Enhancing the AI's ability to understand context and nuance in video content.
2. Expanding language support: Developing capabilities to verify content in multiple languages.

3. Real-time verification: Exploring possibilities for live fact-checking during video playback.
4. Ethical AI use: Implementing stronger safeguards against potential misuse of the technology.
5. User education: Incorporating features to help users understand the verification process and develop critical thinking skills.

In conclusion, VerifyStream represents a significant step towards AI-assisted content verification, but it also serves as a reminder of the responsible development and use of AI technologies. As we continue to refine and expand this tool, we must remain vigilant about its potential impacts and strive to create a more informed and discerning digital citizenry.