International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2024: 7.101

AI-Powered Podcast

Akhilesh K

Department of Computer Applications

Abstract: ***The AI-Powered Podcast Generation Platform introduces a revolutionary approach to audio content creation by integrating generative AI models like text-to-speech synthesis, and image generation. It automates the podcast production lifecycle-from idea generation to audio synthesis and visual branding-enabling content creators, educators, and professionals to publish rich, engaging audio material rapidly and with minimal manual effort. This paper details the system architecture, tools used, security considerations, user workflow, and future scope for scalability and innovation.***

Keywords: Podcast Automation, Text-to-Speech, Generative AI, Convex, Clerk, Vercel, Next.js, Audio Content Creation

1. Introduction

The democratization of audio content is a hallmark of the modern internet age. Podcasting, in particular, offers users the flexibility to consume content passively while multitasking. However, creating a podcast traditionally involves scriptwriting, professional narration, editing tools, and publishing infrastructure, which can be time-consuming and expensive.

With advances in artificial intelligence, it is now possible to delegate several components of the production pipeline to machines. Tools such as GPT-4 can now generate humanlike text, and TTS engines are capable of producing natural sounding voices. Our platform aims to leverage this power to assist in the automatic generation of podcast episodes from a simple text prompt.

This project represents a shift in how media is created and consumed by reducing the friction associated with audio production and allowing creators to focus on content strategy rather than technical implementation.

2. Objectives

The main objectives of this AI-powered podcast system are:

- To eliminate manual scriptwriting by using GPT-4 for dynamic and creative script generation.
- To provide realistic text-to-speech audio from generated scripts using TTS APIs.
- To generate a visual thumbnail using DALL·E 3 for better reach and recognition of podcast episodes.
- To manage user authentication and content history using Clerk and Convex.
- To create a fully responsive, secure, and scalable frontend using Next.js and deploy it seamlessly using Vercel.

3. Proposed System

The proposed system comprises a web platform that allows users to generate full podcast episodes by submitting a topic. Upon submission, the system executes a multi-step workflow:

1) The system forwards the user's topic to GPT-4 to generate a structured podcast script.
2) The generated script is converted into human-like speech using a TTS engine.
3) A visual thumbnail is generated using the DALL·E 3 model for branding and promotion.
4) All content is saved in a centralized database (Convex), viewable through the user's dashboard.

The key innovation lies in automating all aspects of media generation with minimum user interaction, while maintaining quality and customizability.

4. System Architecture

The architecture is divided into five layers:

1) Frontend Layer: Built using React and Next.js, it presents a clean interface where users can submit prompts, view episode history, and manage accounts.
2) Authentication Layer: Powered by Clerk, it supports sign-up/sign-in, session management, and secure route protection.
3) Backend/API Layer: Hosted on Convex, this layer handles routing, calling OpenAI's APIs, and managing user data.
4) AI Services Layer: Includes GPT-4 for script generation, TTS for voice conversion, and DALL·E 3 for thumbnail creation.
5) Deployment Layer: Vercel hosts the application and provides edge caching for performance.

5. Module Descriptions

a) User Authentication Module

Handles secure user sign-up and login using Clerk. Sessions are managed using JSON Web Tokens and are required to access protected content generation features.

b) Prompt Handling and Script Generation

Accepts user prompts and sends them to the GPT-4 API. The returned script is formatted and presented for review. Prompts are validated to avoid abuse or irrelevant content.

c) Audio Synthesis Module

Utilizes a TTS API to convert scripts into audio files. The generated audio is streamed or saved to the user's library for future playback or download.

d) Thumbnail Generation Module

Sends a request to DALL·E with keywords derived from the topic or script. Returns a high-resolution image as a podcast cover.

e) Content Storage and Dashboard

Stores user history, generated scripts, audio files, and thumbnails using Convex. The dashboard interface presents these assets in an organized, downloadable, and playable format.

6. Workflow

Step 1: User registers/logs in via Clerk.

Step 2: User enters a podcast topic (e.g., "How AI is transforming education").

Step 3: Backend sends the topic to GPT-4, receives a script.

Step 4: Script is forwarded to TTS engine to generate speech.

Step 5: Keywords are extracted and sent to DALL·E to generate a thumbnail.

Step 6: Convex stores all files; frontend dashboard allows users to listen, view, and download.

7. Security Considerations

Security is critical since user-generated content and private data must be protected:

- Role-based access control via Clerk ensures only authenticated users can generate or view content.
- API keys and session tokens are secured using environment variables and Vercel vaults.
- Rate-limiting and abuse detection are enforced at the backend.
- All communication is secured via HTTPS.

8. Performance and Scalability

- Stateless architecture allows for horizontal scaling with zero downtime.
- Convex handles real-time data transactions without requiring traditional backend hosting.
- OpenAI APIs scale automatically, ensuring consistent latency.
- Vercel ensures global edge delivery, reducing content load times.

9. Use Cases

- Education: Teachers generate audio lessons easily.
- Business: Companies produce automated updates for customers.
- Influencers: Rapid content creation for social media engagement.
- News Outlets: Real-time news delivered in podcast form.

10. Future Enhancements

- Support for regional languages and multilingual podcasts.
- Voice cloning to maintain consistent hosts.
- Live collaboration and editing interface.
- Integration with Spotify, Apple Podcasts, and RSS feeds.
- Analytics dashboard to measure audience engagement.

11. Conclusion

The AI-Powered Podcast Generation Platform stands as a transformative tool in the digital media space. By leveraging GPT-4, TTS, and image synthesis technologies, it automates end-to-end podcast production. The modular, scalable design powered by modern web technologies ensures future growth and real-world applicability. It opens up possibilities for creators, educators, and enterprises to enter the podcasting space with minimal friction and maximum creativity.

Acknowledgment

I would like to utilize this opportunity to express gratitude to each person who made it possible for me to complete my project successfully. Thus, I would like to remark few people, whom I want to thank and express sincere gratitude. I convey our truthful gratitude to BMSITM Management for providing good infrastructure and educational support in lighting our career. I would like to show my sincere gratitude to our Principal Dr. Sanjay H A for his kind support in completing this project. I take this opportunity to thank our Head of Department, Dr. M. Sridevi who supported me with her valuable inputs on this project. I express my deep sense of gratitude to my internal guide Assistant Professor, Ashwitha K, and project coordinator Assistant Professor Head, Dept. of MCA Dr. M. Sridevi for her support, encouragement, and valuable input throughout the completion of this project. I also thank all my professors and non-teaching staff members, who contributed their help and support directly or indirectly in completing this project. Last but not least, I thank my parents and friends who stood with me as a moral support and encouraged me in accomplishing this project.

References

[1] OpenAI, "GPT-4 Technical Report," OpenAI, 2023. https://openai.com/ research/gpt-4
[2] Vercel Documentation. https://vercel.com/docs
[3] Clerk Authentication Guide. https://clerk.com/docs
[4] Convex Developer Docs. https://docs.convex.dev https://platform.openai.
[5] com/docs/guides/images