iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.69144

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Empowering Learning: Crafting Educational Podcasts with GEN AI

Dr. Shagufta Mohammad Sayeed Sheikh[1], Sumeet Kumbar[2], Parth Dabhade[3], Prem Kalamkar[4], Vishnu Jangid[5]
Artificial Intelligence & Data Science AISSMS Institute of Information Technology, Pune, India

Abstract: ***The integration of Generative AI has facilitated innovative tools that are transforming content creation in the educational landscape, ushering in a shift towards more efficient and accessible learning paradigms. In this project, AI is leveraged to automate podcast production, replicating text-based educational content into high-fidelity audio content that caters to varied learning needs. Leveraging advanced frameworks like Large Language Models (LLMs) and Text-to-Speech (TTS) technologies, the system streamlines otherwise time-consuming processes like scripting, recording, and editing, thus speeding up the content creation process and improving accessibility for instructors. In addition, the system automates title image and metadata generation, which improves discoverability and professionalism of each podcast episode. By combining multiple AI capabilities, this project demonstrates the potential of Generative AI to personalize content, improve accessibility, and improve efficiency in educational environments. It enables personalized learning pathways and empowers educators to deliver more engaging and effective content, with the potential to reach an international audience and fit into different educational environments with ease. Additionally, this approach reduces the resource load of educational institutions, enabling them to deliver high-quality audio content even with an availability of limited resources and staff.***

Keywords: Generative AI, Large Language Model, podcasting, education, text-to-speech, content creation, automation.

I. INTRODUCTION

This project reimagines how educational content can be created and shared. Traditional teaching materials—like textbooks or static slides—don't always meet the evolving needs of modern learners. Many students today are looking for content that's more engaging, accessible, and tailored to their individual learning styles. That's where this idea comes in: using generative AI to turn written educational content into rich, audio-based podcasts that feel more like guided learning experiences than simple recordings.

The process of producing high-quality educational podcasts usually involves a lot of people—writers, narrators, editors, and designers—and can be slow and expensive.

What this project does differently is use advanced AI tools to automate much of that work. It helps educators convert lessons into podcasts quickly and efficiently, without compromising on quality. Tools like LLaMA 3.1 help break down and structure the content, Tacotron and WaveNet generate clear, human-like voices, and DALL·E creates unique visuals to go with each episode. With all of this working together, educators can spend more time teaching and less time worrying about the production process. One of the biggest strengths of this system is how adaptable it is. It's not limited to one type of learner or subject. Whether you're a high school student, a college learner, or a professional taking a course for work, the platform can adjust the tone, language, and complexity to suit your needs. It also supports multiple languages, so it can reach learners in different parts of the world—including those in rural or underserved communities.

That kind of accessibility can make a real difference in closing educational gaps. What really sets this approach apart is how it enhances the learning experience itself. These aren't just podcasts that talk at you—they're thoughtfully designed to be engaging and immersive.

The voiceovers are expressive and natural, and the visuals help reinforce the key ideas. By appealing to both auditory and visual learners, the system makes it easier for people to stay focused, understand complex topics, and remember what they've learned. It's a more well-rounded and inclusive way to learn. In the long run, this platform could help shape the future of education. It lowers the barriers to high-quality learning materials—cost, time, and technical know-how—and opens up new possibilities for learners and educators alike. Looking ahead, the project could evolve to include video-based lessons, interactive AI-driven discussions, and deeper support for different languages and learning styles.

With the help of generative AI, this isn't just about improving how we learn—it's about expanding access to education for everyone, everywhere.

II. LITERATURE REVIEW

The use of generative AI in educational podcast production is changing the game—making it easier, faster, and more intuitive to create meaningful learning content. Thanks to recent breakthroughs, AI can now do more than just speed up the process. It can actually enhance the way learners interact with content. For instance, smart topic segmentation and visual mapping features help listeners browse and explore podcast episodes in a more structured way. It's a shift we've already seen in platforms like Amazon Music, where recommendation systems and ranking algorithms guide users to exactly what they're looking for. This project applies similar AI-powered features, but with a focus on education.

One of the biggest leaps has come from text-to-speech technology. Tools like Tacotron have made it possible to generate expressive, life-like narration that sounds surprisingly human. On top of that, FastSpeech models dramatically cut down the time it takes to generate high-quality audio—making it practical to produce full-length podcasts on demand. That means educators can now create large volumes of content quickly, and personalize it to suit different learners, subjects, and age groups. The best part? Quality doesn't suffer in the process.

And audio is just one piece of the puzzle. AI is also enhancing how we interact with content by adding visual elements. Imagine a podcast that comes with a topic map—like a treemap—that lets you see what's covered and jump to the section you need. It's a simple concept, but incredibly useful, especially in educational settings where students benefit from clear, organized content layouts. Today's learners are used to clean, interactive user interfaces, and this project takes that to heart, creating a user-friendly environment that keeps things engaging and easy to navigate.

What really makes this platform stand out is how it blends all these tools together into one cohesive experience. By generating episode visuals and attaching smart metadata, the system makes each podcast easier to explore and digest. Whether you're quickly skimming for a topic or diving into a full lesson, the experience feels smooth and intuitive. It's about turning passive listening into something more active—something that actually helps learners stay focused and retain information.

The bigger picture is just as exciting. AI is already proving to be incredibly helpful in organizing and customizing content to meet different learner needs. In education especially, the ability to personalize learning experiences makes a real difference. When the system understands what users are searching for, how they interact with episodes, and what keeps them engaged, it can tailor the content more effectively. That's especially valuable for students who learn in different ways—or those with disabilities—where transcripts, visuals, and adaptive content formats can make all the difference.

Advances in speech synthesis are also taking things to the next level. Newer models like Flow-TTS and Deep Voice 3 can produce voices with distinct accents, emotions, and natural flow—so much so that they're often indistinguishable from real human speakers. This kind of personalization is helping podcasts reflect more diverse audiences, which is crucial in global learning environments. And when learners hear content in voices that feel familiar or comfortable, it makes the material more approachable and relatable.

At the heart of this project is FastSpeech 2, a modern voice synthesis model designed to strike the right balance between speed and audio quality. With it, we're able to generate clear, engaging audio that sounds polished and professional—while still being scalable. Educators can now produce content that not only sounds great but is also accessible and inclusive for all kinds of learners, regardless of where they are or how they learn best.

Research continues to show the value of AI in organizing and tailoring podcast content to fit individual learning goals. By studying how people engage with podcasts—what they search for, how they listen, and what keeps them engaged—AI can better align content with user intent. This insight feeds directly into how this project is designed: to personalize the educational podcast experience based on listener behavior. Accessibility is also a key part of the design, with features like transcripts and visual supplements ensuring that content is more inclusive and easier to absorb, especially for students with diverse learning needs or disabilities.

All of these innovations come together to form a platform that's practical for educators, engaging for students, and adaptable to a wide range of learning environments. Whether it's a self-paced course, a flipped classroom, or corporate training, this system brings flexibility, quality, and personalization to the table—redefining what educational content can look and sound like in the digital age.

Further improvements in TTS, including models such as Flow-TTS and FastSpeech, place a strong emphasis on both speed and audio quality. These systems not only generate voices quickly but also maintain a high level of fidelity and control over how speech is delivered. Learners benefit from content that is engaging and easy to follow, while educators and institutions benefit from the ability to produce that content at scale, without compromising on clarity or expression.

This project brings all of these advancements together through the implementation of FastSpeech 2, a powerful model designed for fast and high-quality voice synthesis. By integrating this technology, the system is able to produce educational podcasts that are both professional in sound and scalable in production. It's a solution that streamlines the creation process while ensuring that the end product remains compelling and effective for learners.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

III. METHODOLOGY

This project taps into the potential of Generative AI to make creating educational podcasts faster, more affordable, and easier to scale. Instead of relying on the traditional, often time-consuming process of writing scripts, recording voiceovers, and editing everything by hand, the system automates most of that heavy lifting. From researching content and drafting scripts to generating the audio, visuals, and metadata—it's all handled with a streamlined AI-driven workflow. Each part of the system is designed to keep quality high while simplifying production. As shown in Fig. 1, everything follows a clear step-by-step process, making it easy to produce consistent, accessible content for a wide range of learning environments. The next sections break down how each stage works together to create podcasts that are both engaging and scalable.

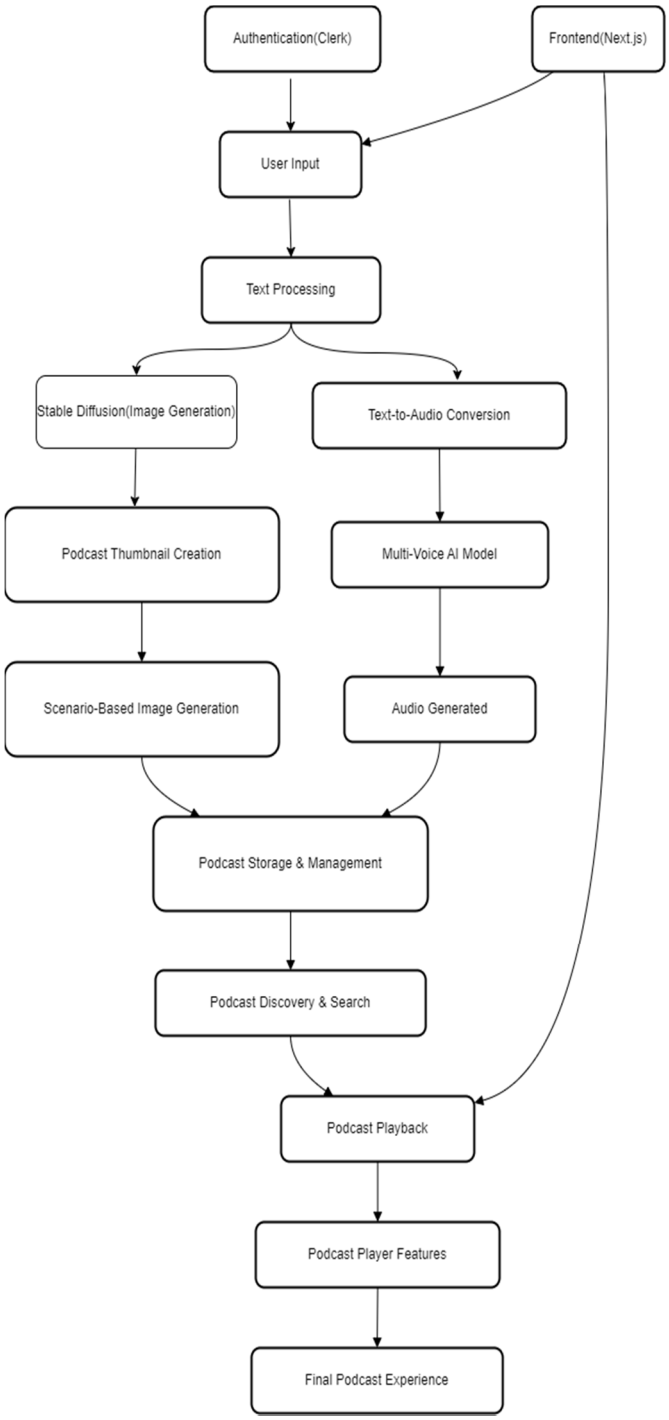


Fig.1. System Architecture

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

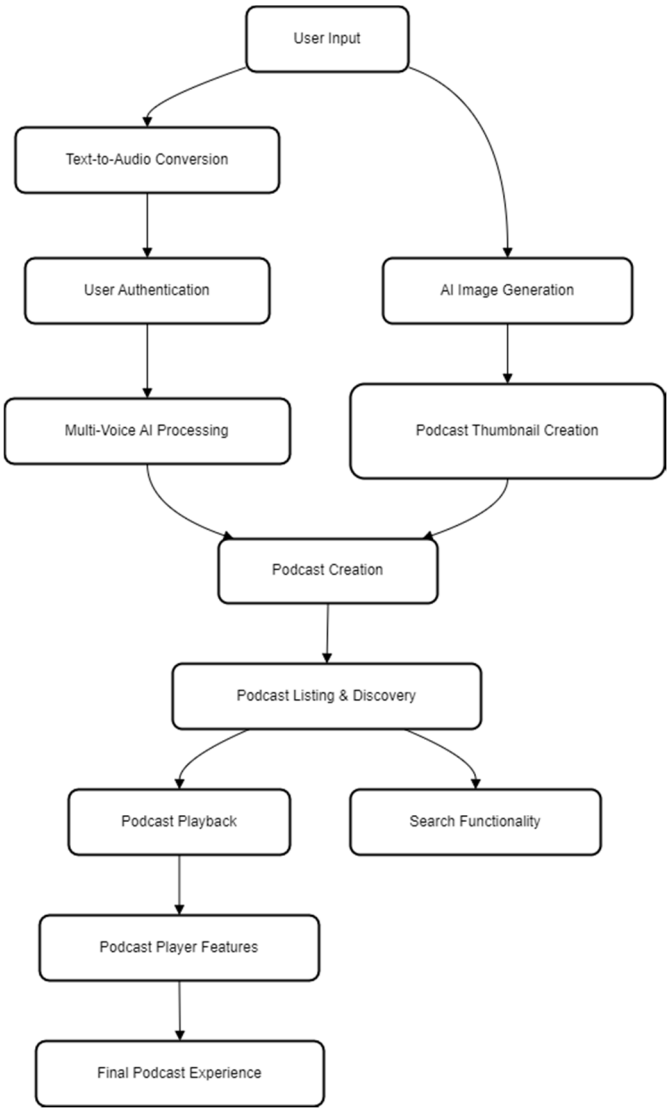


Fig.2. Flow Chart

A. Data Collection

This project collects educational data from reputable sources like Khan Academy, MIT OpenCourseWare, and the Open Textbook Library to ensure content quality. It targets a wide range of subjects, including STEM, humanities, and social sciences. Using automated tools like Selenium and BeautifulSoup, the system scrapes, parses, and structures the material for AI processing. Metadata such as titles, topics, and categories is also extracted to support indexing, classification, and content retrieval. This ensures the dataset is clean, organized, and optimized for podcast generation.

But we're not just grabbing any data—we're focused on quality. Once the material is collected, it goes through a careful review process to make sure it's clear, accurate, and genuinely useful for learning. Everything is checked for readability and instructional value, and then grouped based on complexity. This ensures the podcasts we generate match the right academic level, whether it's for younger students, high schoolers, or university-level learners. By being selective and thoughtful about the content we use, we make sure the final audio content is not only engaging but also meaningful and educationally sound.

On top of that, we've built in smart filtering systems to personalize the learning experience even further. Using natural language processing (NLP), the AI analyzes how complex the content is and what topics it covers. This allows the system to match the right material with the right audience. For example, a high school student studying physics will get content that's simplified and easy to follow, while a university student will get a deeper, more advanced explanation. This level of customization makes learning more inclusive and effective because it meets learners where they are in their academic journey.

Another important part of the process is how the content is organized. We tag everything with detailed metadata, which makes it super easy for the system to find and use the most relevant material when generating podcasts. This not only speeds up production but also ensures that the content is up-to-date and closely aligned with current educational standards and research. Thanks to this streamlined and smart approach, we can produce high-quality podcasts quickly, without sacrificing depth or accuracy.

In the end, this combination of structured data collection, smart filtering, and AI-powered processing changes the game for educational content creation. It's not just about automating the process—it's about doing it thoughtfully, so the result is something learners can actually connect with. By putting quality, personalization, and accessibility at the heart of the system, we're making educational podcasting more powerful, more scalable, and more impactful than ever before.

B. Text Processing and Script Generation

This project uses the power of LLaMA 3.1, accessed through the Ollama interface, to turn raw educational material into clear, engaging podcast scripts. The AI carefully breaks down complex topics into smaller, more digestible sections, making them easier to follow and well-suited for audio learning. One of its strengths lies in crafting scripts that are not only logical and easy to understand but also memorable for listeners. From catchy, attention-grabbing intros to neatly wrapped-up conclusions, the AI helps ensure each episode flows smoothly and delivers key points in a way that really sticks.

A key feature of the system is its ability to tailor content for different types of learners. Whether it's high school students, university learners, or professionals seeking to expand their knowledge, the AI can adjust the tone, depth, and language to match the audience. Thanks to the flexibility of the Ollama interface, scripts can be created with high accuracy and relevance, making the content feel relatable and on point. This thoughtful structure doesn't just make the material easier to listen to—it makes the whole experience more enjoyable and effective, turning educational podcasts into a genuinely powerful learning tool.

What's especially helpful is how this AI-driven process supports teachers, content creators, and educators. Instead of spending countless hours researching and writing scripts from scratch, they can now generate high-quality content quickly and efficiently. The system takes care of the heavy lifting while still aligning with the learning objectives set by educators. Plus, because it can adapt across a wide range of subjects and learning levels, the content can reach more people, in more places, and across different educational settings.

By combining smart automation with a user-friendly interface, this approach not only speeds up the podcast creation process but also ensures the final output maintains a high standard of educational value. It strikes the right balance between efficiency and quality, making it easier than ever to produce podcasts that truly resonate with learners—wherever they are in the world.

C. Audio Generation

The audio production process takes podcast creation to the next level by using advanced Text-to-Speech (TTS) models like Tacotron and WaveNet. These powerful AI tools transform written scripts into natural, human-like speech that feels much more engaging and lifelike. By capturing the nuances of human speech—like tone, pitch, and rhythm—the system creates audio that feels warm, expressive, and easy to listen to. What's more, instructors or content creators can fine-tune the voice settings to match the mood of the content or the preferences of their audience. Whether it's adjusting the pitch or slowing down the pace for clarity, these customization features help make the narration feel more personal and relatable.

But it doesn't stop at just converting text to speech. The system also adds thoughtful touches like background music, sound transitions, and well-timed pauses. These subtle effects help break the content into digestible parts, highlight important points, and keep listeners engaged from start to finish. It's all designed to mimic the rhythm and flow of a real human conversation—so instead of feeling like you're just listening to a robot read, the experience feels dynamic and genuinely enjoyable. This kind of audio production is especially helpful when tackling complex subjects, where varied pacing and emphasis can really help learners grasp difficult concepts more easily.

Another standout feature of the TTS technology is its ability to produce audio in multiple languages. This opens the door to making educational podcasts accessible to a global audience. Whether learners speak English, Spanish, Mandarin, or another language, the system can deliver high-quality content tailored to their needs. That means more students around the world can access valuable educational material in a language they're comfortable with. Beyond just improving access, this also encourages cross-cultural learning and helps break down language barriers, making education truly inclusive and borderless.

By combining lifelike narration with sound design and multilingual capabilities, this AI-powered audio system transforms how educational content is delivered. It's more than just reading out loud—it's about creating an immersive, thoughtful, and flexible learning experience that speaks to everyone, no matter where they're from or how they learn best.

D. Image Generation for Episode Titles

The image creation process brings each podcast episode to life visually using DALL·E 3, an advanced AI tool that crafts unique, eye-catching cover images. Instead of using generic artwork, the system analyzes the topic and keywords of each episode to design visuals that truly reflect the subject matter. Whether it's a detailed scientific diagram for a physics episode or a beautiful illustration of a historical site for a history lesson, these images aren't just pretty—they're purposeful. They help reinforce what the episode is all about, giving learners a visual connection to the content they're hearing.

Every episode is supported by an image that's not only relevant but also visually appealing. These graphics add a polished, professional touch that helps the podcast stand out and keeps the overall look consistent. More than just decoration, the visuals play a key role in capturing attention and setting the tone for what listeners can expect. They make the podcast feel more engaging, more intentional, and ultimately, more effective as a learning tool. By creating a cohesive style across episodes, these AI-generated visuals also help establish a strong brand identity that makes the podcast easy to recognize—something that's especially helpful for learners who follow multiple series or educational sources.

What makes these images even more powerful is their practical impact on visibility and reach. In the crowded world of digital media, a striking cover photo can make all the difference in catching someone's eye. Whether someone's browsing on a podcast platform or scrolling through educational content online, a strong visual hook encourages clicks, listens, and shares. It helps new learners discover the podcast and gives returning listeners something familiar to connect with.

By integrating visuals that are both beautiful and meaningful, this process doesn't just enhance the aesthetics—it supports learning, boosts discoverability, and strengthens the podcast's overall presence. It's a small touch that makes a big difference, helping turn educational content into something truly immersive and memorable.

Fig.3. A School Library with Books That Tell Stories Themselves

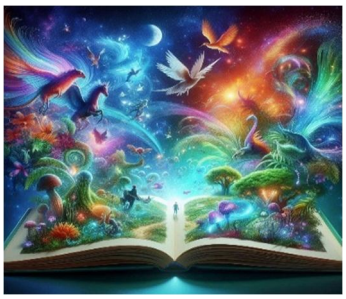


Fig.4. A Book Transforming into a Virtual World

Fig.5. Water Cycle (Evaporation, Condensation, Precipitation

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Fig.6. Means of Transport (Road, Rail, Water, Air)

Fig.7. Solar System with Planets and the Sun

E. Metadata Generation

Creating effective metadata plays a crucial role in making educational podcasts easier to find and more accessible across different platforms. By using advanced Natural Language Processing (NLP) techniques, the system automatically pulls out key details like episode titles, summaries, and relevant tags. It analyzes each podcast script to identify the most important keywords and themes, ensuring that the metadata accurately reflects what the episode is about. This not only helps learners quickly grasp the content at a glance but also boosts the podcast's visibility through better search engine optimization (SEO), making it easier to discover on streaming services and educational platforms.

Accurate, AI-generated metadata also increases the chances of a podcast reaching its intended audience. By automatically tagging each episode with the right keywords, the system ensures that content is properly categorized and easy to navigate—especially useful in large educational libraries. Whether a student is looking for a specific concept or an educator is managing hundreds of resources, this smart tagging system saves time and improves the overall user experience by making content easier to locate and organize.

But the benefits go beyond just searchability. Well-structured metadata also helps keep entire podcast series better organized. With automated classification and indexing, educators and institutions can maintain neatly arranged libraries, making it simple to distribute, reference, or reuse content as needed. It allows users to browse intuitively, revisit materials, and even cite episodes with ease—all of which contribute to a more seamless and effective learning process.

This AI-powered approach to metadata doesn't just support better content management—it ensures that quality educational materials can reach more people when they need them most. By streamlining discoverability and improving content delivery, the system helps bridge the gap between learners and the information they're looking for, no matter where they are or what their learning goals might be.

F. Quality Assurance and Review

To maintain a high standard across all podcast episodes, the project uses a thoughtful combination of AI-powered checks and human review. The AI is trained to catch issues like errors in the script, unclear audio, or visuals that don't match the topic. Once the AI has done its job, a human reviewer steps in to double-check everything—making sure the content flows well, sounds natural, and meets the learning goals. This dual-layer review process ensures that every episode is polished and ready for learners to enjoy.

What makes the system even more effective is its focus on continuous improvement. After episodes go live, feedback from listeners—whether they're students or educators—is collected and analyzed. This feedback helps the AI fine-tune its future outputs, making each new episode smarter and more tailored to the needs of the audience. Over time, this feedback loop leads to stronger, more relevant content that evolves alongside the learners it serves.

Prompt: "Explain the concept of photosynthesis in a way that is engaging for high school students."

Step 1: Data Collection

The system compiles relevant educational material on photosynthesis from sources like Khan Academy and Open Textbook Library. The content is well selected to ensure its suited to teenagers in high school.

Step 2: Text Processing and Script Generation

Text processing and script generation With LLaMA 3.1, the system produces an organized script that has a compelling opening, precise explanations of the light dependent and light-dependent reactions, and summary highlighting key points.

Step 3: Audio Generation

The text is translated into speech by Tacotron and WaveNet, with a voice selected to match the tone and refinement of the Sound effects, like the occasional silent moments and musical changes are added to enhance interaction.

Ex. Audio Generation for user prompt

Step 4: Image Generation for Episode Titles

DALL-E 3 generates a lovely title image that features a plant in sunlight, reflecting the spirit of photosynthesis.

Fig.8. Image generated for user prompt

Step 5: Metadata Generation

The system generates metadata with the title. "Understanding Photosynthesis: A High School Guide," a description of the episode, and tags such as "photosynthesis," "biology," and "high school education."

Step 6: Quality Assurance and Review

The episode undergoes AI-driven error detection and human review for clarity, precision, and overall quality. Feedback from educators is incorporated to refine the content.

IV. EXPERIMENTAL RESULTS

To comprehensively assess the performance of our Generative AI-driven educational podcast system, we employ a suite of evaluation metrics spanning semantic alignment, audio quality, metadata accuracy, and user engagement. These metrics ensure that the system not only produces high-quality audio content but also maintains educational relevance and accessibility.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

A. Audio Quality Evaluation

Metric: We use Mean Opinion Score (MOS) to evaluate the naturalness and clarity of the AI-generated audio. The MOS is calculated based on user feedback on a scale of 1 to 5, where 1 is poor and 5 is excellent.

TABLE 1
The following table compares MOS of Tacotron and Wavenet AI model.

Aspect	MOS(Tacotron)	MOS(Wavenet)
Naturalness	4.1	4.5
Emotional Alignment	3.8	4.2
Clarity	4.0	4.4

Interpretation: WaveNet outperforms Tacotron in terms of naturalness, emotional alignment, and clarity, making it the preferred choice for generating high-quality educational podcast audio.

B. Metadata Accuracy

Metric: We evaluate the accuracy of the automatically generated metadata (titles, descriptions, and tags) by comparing it with human-annotated metadata. The accuracy is calculated as the percentage of correctly generated metadata fields.

TABLE 2
The following table evaluates accuracy of Meta-data.

Metadata field	Accuracy(%)
Title	92%
Description	88%
Tags	90%

Interpretation: The system demonstrates high accuracy in generating relevant metadata, ensuring that the podcasts are easily discoverable on various platforms.

C. User Engagement

Metric: We measure user engagement by analyzing the average listening duration and completion rate of the podcast episodes. The completion rate is the percentage of users who listen to the entire episode.

TABLE 3
The following table gives average listening duration and completion rate of user.

Podcast Episode	Average Listening Duration(minutes)	Completion rate(%)
Introduction to AI	12.5	78%
History of Money	14.2	82%
Photosynthesis	13.8	80%

Interpretation: The high average listening duration and completion rate indicate that the AI-generated podcasts are engaging and hold the interest of the audience.

D. Comparative Model Performance:

We compare the performance of different Text-to-Speech (TTS) models used in our system, including Tacotron, WaveNet, and FastSpeech 2, based on audio quality and processing speed.

TABLE 4

The following table gives Naturalness of different AI models and their processing speed/response time.

Model	MOS(Naturalness)	Processing Speed(sec/m)
Tacotron	4.1	3.5
Wavenet	4.5	4.2
FastSpeech2	4.3	2.8

Interpretation: While WaveNet provides the highest audio quality, FastSpeech 2 offers the fastest processing speed, making it suitable for large-scale podcast production.

E. Sentiment Analysis Accuracy

Metric: We evaluate the accuracy of the sentiment analysis module by comparing its sentiment labels (Positive/Negative/Neutral) against human-annotated ground truth for the same podcast scripts.

TABLE 5

The following table gives Accuracy of all user generated Podcasts.

Podcast Script	Accuracy(%)
Introduction to AI	89%
History of Money	91%
Photosynthesis	87%

Interpretation: The sentiment analysis module demonstrates high accuracy in identifying the emotional tone of the podcast scripts, ensuring that the audio content aligns with the intended emotional context.

F. Image Generation Quality

Metric: We use Fréchet Inception Distance (FID) to compare the quality of AI-generated title images (using DALL-E 3) with human-created illustrations. Lower FID scores indicate better quality.

TABLE 6

The following table compares FID of AI models which returns AI-generated title images

Model	FID Score
DALL-E 3	54.3
Stable Diffusion	58.1

Interpretation: DALL-E 3 generates higher-quality images compared to Stable Diffusion, making it the preferred choice for creating visually appealing title images for podcast episodes.

V. LIMITATIONS

While the system brings major benefits to educational podcast production, it's not without its challenges. Ensuring that the content remains authentic and free from bias is crucial to maintaining accuracy and trust. Creating natural-sounding, multilingual audio also poses difficulties, especially given the wide range of accents, pronunciations, and speaking styles that exist. On top of that, the

system's performance can be limited by available computing power, making it harder to scale efficiently. There's also the challenge of gathering and using user feedback in a structured way that actually improves future content. Tackling these issues is key to making the platform more reliable, scalable, and inclusive for all learners.

VI. FUTURE WORK

Looking ahead, future improvements will focus on making the system even more inclusive and engaging. One key area of development is expanding multilingual support by adding a wider range of languages and dialects, so that learners from different backgrounds can access content in the language they're most comfortable with. The project also plans to explore the use of video to complement audio learning—helping to create richer, more interactive experiences through visual elements. To make learning even more personalized, more advanced algorithms will be introduced to tailor content to each user's learning style and pace. Additionally, incorporating real-time feedback tools will allow for ongoing interaction, helping learners feel more supported and involved. By combining audio, video, and adaptive learning techniques, the system aims to transform how educational content is delivered and experienced.

VII. ACKNOWLEDGMENT

We're deeply grateful to the developers and contributors behind technologies like LLaMA 3.1, Tacotron, WaveNet, DALL·E 3, and the wider OpenAI ecosystem. Their work has laid the foundation for much of what this project has achieved. A special thanks goes to the teams at Ollama and Stable Diffusion, whose tools and platforms have played a key role in shaping our approach. We also want to acknowledge the incredible efforts of the global open-source and AI communities—your commitment to innovation and collaboration continues to inspire and empower projects like ours.

VIII. CONCLUSION

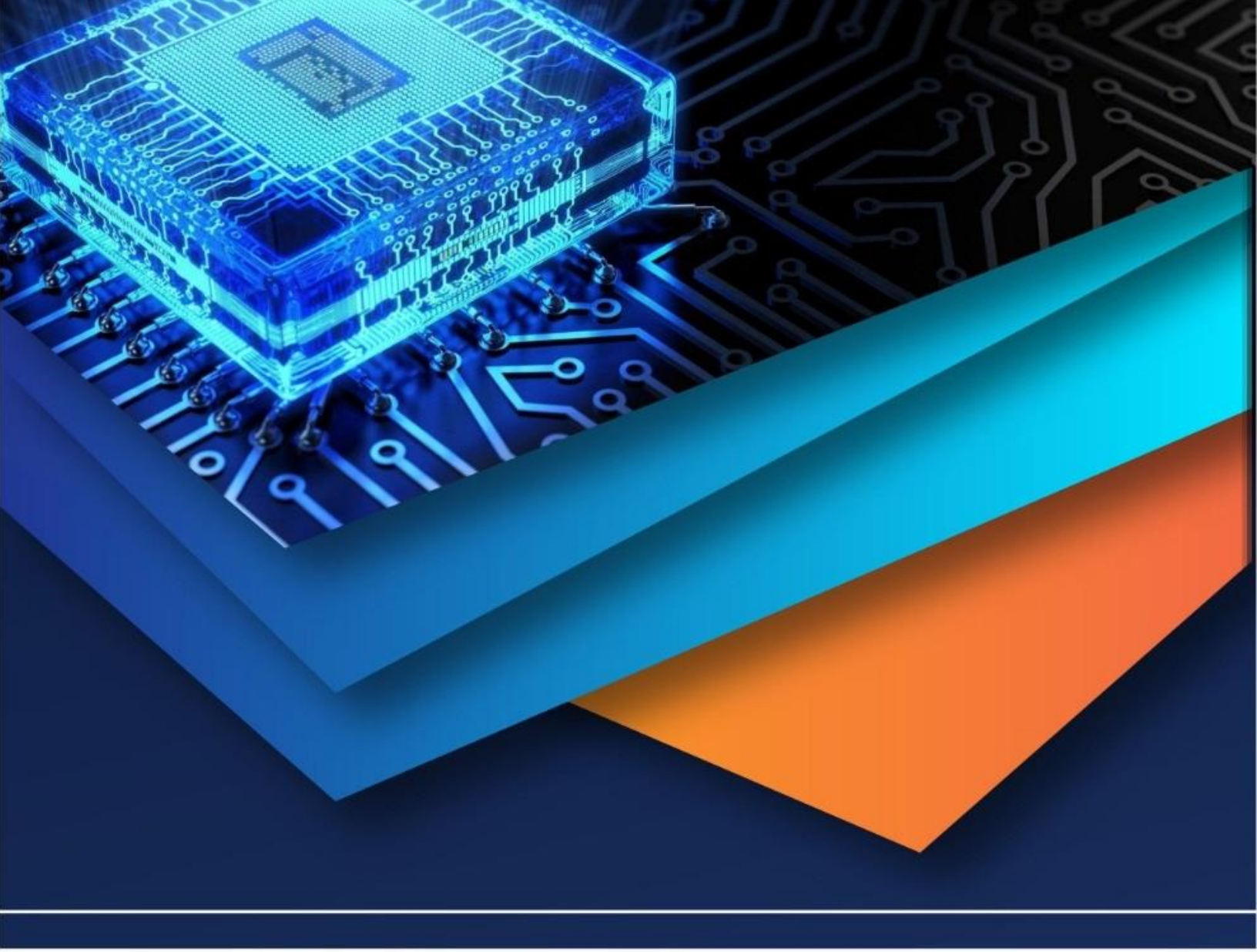
This paper introduces a Generative AI-driven system designed to create educational podcasts by seamlessly integrating text-to- audio and text-to-image technologies. The system utilizes advanced models such as LLaMA 3.1 for script generation, Tacotron and WaveNet for realistic audio synthesis, and DALL- E 3 for creating engaging visuals. This combination allows for the automated generation of personalized, interactive, and multimedia-rich educational content, enhancing learning experiences across various subjects. The project aims to enhance the accessibility and effectiveness of educational materials by offering customized audio content with diverse voice modulation options and visually relevant title images. The text-to-speech models, with language support and audio cues, create a more immersive learning environment while the metadata generation improves discoverability. A robust quality assurance process, involving both AI-driven checks and human oversight, ensures the generated content meets high standards of clarity and accuracy. This iterative approach ensures that the content remains adaptable and continuously improves based on real-world feedback from educators and students.

Looking ahead, the system will be further refined to include more languages, incorporate video content, and expand its capabilities to address a broader range of educational needs. The integration of these advanced AI technologies promises to offer scalable and dynamic solutions for educational institutions, enabling more personalized and interactive learning experiences. Through continuous enhancement and feedback, this system aims to redefine how educational content is created, distributed, and experienced globally.

REFERENCES

[1] Jimin Park, Chaerin Lee, Eunbin Cho, and Uran Oh, Enhancing the Podcast Browsing Experience through Topic Segmentation and Visualization with Generative AI, ACM International Conference on Interactive Media Experiences (IMX '24), June 2024.
[2] Geetha Sai Aluri, Paul Greyson, and Joaquin Delgado. 2023. Optimizing Podcast Discovery: Unveiling Amazon Music's Retrieval and Ranking Framework. In Proceedings of the 17th ACM Conference on Recommender Systems. 1036– 1038.
[3] Y. Wang, R. Skerry-Ryan, D. Stanton, R. J. W. Y. Wu, N. Jaitly, and Z. Yang, "Tacotron: Towards end-to-end speech synthesis," in Proc. Annu. Conf. Int. Speech Commun. Assoc., 2017, pp. 4006–4010.
[4] Y. Ren et al., "FastSpeech 2: Fast and high-quality end-to- end text to speech," in Proc. Int. Conf. Learn. Representations, 2021, pp. 1–9.
[5] Hernández-Leo, ChatGPT and Generative AI in Higher Education: User-Centered Perspectives and Implications for Learning Analytics
[6] Fahmi Abdulhamid and Stuart Marshall. (2013). "Treemaps to visualise and navigate speech audio." Proc. of the 25th Australian Computer-Human Interaction Conf. pp. 555– 564.
[7] Pedro Almeida et al. (2022). "A Podcast Creation Platform to Support News Corporations: Results from UX Evaluation." ACM Int. Conf. on Interactive Media Experiences, pp. 343–348.
[8] Geetha Sai Aluri et al. (2023). "Optimizing Podcast Discovery: Unveiling Amazon Music's Retrieval and Ranking Framework." Proc. of the 17th ACM Conf. on Recommender Systems, pp. 1036–1038.
[9] Barry Arons. (1997). "SpeechSkimmer: a system for interactively skimming recorded speech." ACM TOCHI, 4(1), 3–38.

[10] Jana Besser et al. (2010). "Podcast search: User goals and retrieval technologies." Online Information Review.
[11] Sylvia Chan-Olmsted and Rang Wang. (2022). "Understanding podcast users: Consumption motives and behaviors." New Media & Society, 24(3), 684–704.
[12] Amelia Chelsey. (2021). "Is There a Transcript? Mapping Access in the Multimodal Designs of Popular Podcasts." Proc. of the 39th ACM Int. Conf. on Design of Communication, pp. 46–53.
[13] Ann Clifton et al. (2020). "The Spotify podcast dataset." arXiv preprint, arXiv:2004.04270.
[14] Tatsuya Ishibashi et al. (2020). "Investigating audio data visualization for interactive sound recognition." Proc. of the 25th Int. Conf. on Intelligent User Interfaces, pp. 67–77.
[15] Y. Wang et al. (2017). "Tacotron: Towards end-to-end speech synthesis." Proc. of the Annu. Conf. Int. Speech Commun. Assoc., pp. 4006–4010.
[16] J. Shen et al. (2018). "Natural TTS synthesis by conditioning wavenet on MEL spectrogram predictions." Proc. of IEEE Int. Conf. Acoust., Speech, Signal Process., pp. 4779–4783.
[17] W. Ping et al. (2018). "Deep Voice 3: 2000-Speaker neural text-to-speech." Proc. of Int. Conf. Learn. Representations, pp. 214–217.
[18] C. Miao et al. (2020). "Flow-TTS: A non-autoregressive network for text to speech based on flow." Proc. of IEEE Int. Conf. Acoust., Speech, Signal Process., pp. 7209–7213.
[19] Y. Ren et al. (2019). "FastSpeech: Fast, robust and controllable text to speech." Proc. of the 33rd Int. Conf. Neural Inf. Process. Syst., pp. 3171–3180.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY