

PODGEN: AI SaaS Podcast Web Application

Prof. S.M. Shelke¹, Pranav Mahamuni², Jami-ul-Meraj Lukde³, Dinesh Kolhe⁴, Parthavi Jagtap⁵

1 Professor, Department of Computer Engineering, Sinhgad Academy of Engineering, Pune, India

2,3,4,5 Students, Computer Engineering, Sinhgad Academy of Engineering, Pune, India

Abstract: Podgen is the subsequent-era (SAAS) AI-based totally software program platform that simplifies the procedure of making podcasts. conventional podcast production includes considerable time, technical abilities and get admission to to professional gear. Podgen uses extended artificial intelligence to get rid of those limitations, permitting customers to generate podcasts from textual content requests, remodel them into human voices, or even create engaging podcast dumplings with AI. It helps multilingual content and affords actual-time dashboard get admission to. This system integrates personnel to exclude them for authentication and subscription management. The backend is provided with the aid of convex surfaces and vercel to ensure seamless scalability and real-time overall performance. Podgen no longer only goals to make podcasting more on hand, however additionally objectives to promote traits by way of helping users with language disorders. This paper introduces Podgen's methodology, gadget design, structure, and implementation, even as highlighting its capacity and destiny enlargement options..

Key Words: PodGen, SaaS, Podcasting, AI, Text-to-Speech, Thumbnail Generation, Stripe, Clerk, Convex, Multilingual

1.INTRODUCTION

Podcasting has become a global vehicle for communication, entertainment and education. Despite its popularity, many intensive creators are unable to participate due to technical obstacles associated with the production of podcasts. Podgen deals with AI-controlled SaaS solutions that democratize podcast creation. The platform provides tools for converting scripts using Text 2 Check (TTS) in audio, generates podcast miniature views using CNN-based AI, and manages episodes with real-time data sync. Podgen allows creators to focus on storytelling by repeating themselves and automating technical tasks. With multilingual support, secure authentication and subscription-based access, Podgen ensures global reach and friendly user interaction.

2. Methodology

2.1 System Architecture
PodGen is structured around a modular architecture that allows seamless podcast creation:

1. Text input: users offer enter as a set off or script.
2. TTS Module: AI techniques the textual content the usage of Recurrent Neural Networks (RNNs) to generate sensible audio.
3. Image technology Module: A Convolutional Neural network (CNN) translates the text and creates a unique thumbnail.
4. Content material assembly: Audio and image components are included into a complete podcast.
5. Garage & control: content material is stored in Convex and controlled through at ease dashboard.

2.2 Technology Stack

1. Frontend: Next.js, React.js, Tailwind CSS, ShadCN UI
2. Backend: Node.js, Convex (for real-time updates)
3. Authentication: Clerk (supports OAuth and email/password)
4. Payments: Stripe (for subscriptions and billing)
5. AI Integration: OpenAI APIs (for text-to-speech and text-to-image)
6. Deployment: Vercel (serverless CI/CD hosting)

2.3 Algorithms

2.3.1 TEXT TO IMAGE GENERATION ALGORITHM:

Create dynamic podcast thumbnails with CNN :

CNN is particularly well-suited for this task as it enables the extraction of functions from visual data. The algorithm begins by converting the text description into a numerical representation. Following that, we utilize this encoded text to assist the CNN in generating relevant images. CNN acquires the ability to combine specific

words and phrases with corresponding visual elements, including objects, scenes, and emotions. As soon as CNN processes text input, it generates an image that corresponds to the content described. This image can be modified in various ways, including different styles, colors, and layouts, to ensure that each thumbnail view is distinct and visually pleasing. By automating the process of positioning the miniature, the PodGen platform saves creators time and effort, and produces high-quality images that attract listeners.

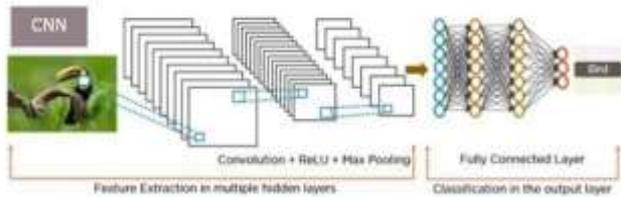


Figure 1: Convolutional Neural Network

2.3.2 TEXT TO SPEECH GENERATION ALGORITHM:

Enhancing Podcast Audio with RNN-Powered Text to Speech:

The PodGen platform contains recurrent neural networks (RNNs) to generate highquality audio from text. this article-to-speech functionality allows creators to fast and effortlessly produce podcasts without the need for expert voice actors or recording studios. The RNN utilized in PodGen is trained on a huge dataset of human speech, permitting it to generate herbal-sounding audio. The set of rules procedures the textual input and predicts the corresponding audio waveform. The RNN's potential to seize lengthy-time period dependencies inside the information ensures that the generated speech is coherent and flows smoothly. PodGen offers a ramification of six distinct voices, each with its own specific characteristics and fashion. This allows creators to choose the voice that first-class fits their podcast's tone and target audience. by way of leveraging OpenAI's superior textual content-to-speech era, PodGen guarantees that the generated audio is of the best first-rate and indistinguishable from human speech.

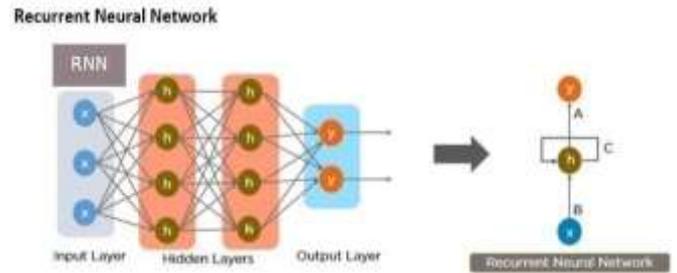


Figure 2: Recurrent Neural Network

2.3.4 UNIQUE FEATURES

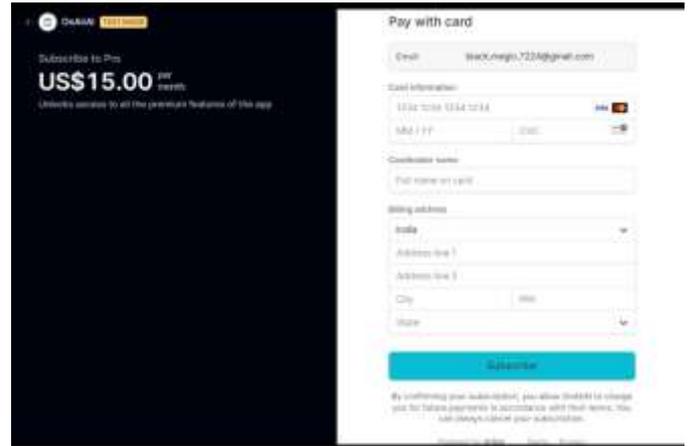
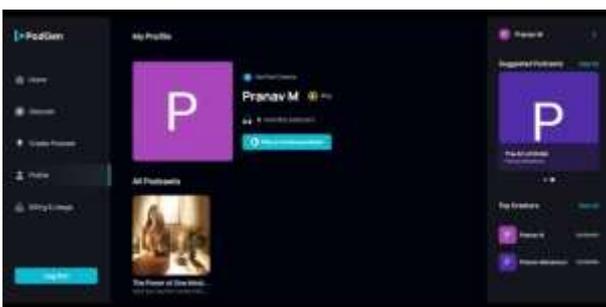
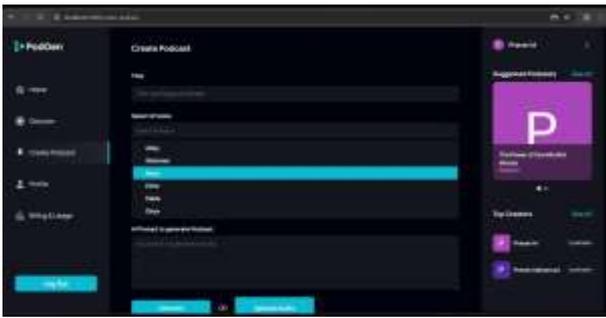
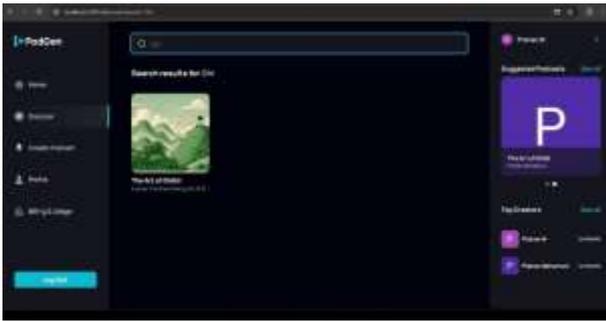
- 1.The advent of multilingual content has greatly benefited the podcast industry, allowing it to reach a global audience in various languages.
- 2.Voice variety: provides six unique AI voices, each with its own distinct accents and tones.
- 3.Secure dashboard: real-time access to podcasts, which include editing, previewing, and publishing features.
- 4.Podgems gamification (future work): customers accumulate "podgems" for expanding episodes, which can be exchanged for premium voices, unique designs, or valuable analytics.

2.3.4 COMPARISON ANALYSIS

PodGen	Traditional Podcast Platforms
PodGen is a fully AI-powered SaaS platform designed to simplify podcast creation through automation and intelligent tools.	Most traditional platforms lack built-in AI capabilities and rely heavily on manual podcast creation and editing processes.
Users can generate podcast episodes using OpenAI's six advanced AI voices, eliminating the need for voice recording.	Users must manually record their voice, which requires time, confidence, and proper audio recording equipment.
PodGen includes AI-powered text-to-image generation to create custom thumbnails and artwork for podcast episodes.	Creators have to manually design or upload images using external tools like Canva or Photoshop.
Even users who are mute or experience speaking anxiety can produce professional-sounding podcasts without ever needing to speak.	These platforms require users to record their own voice, making it difficult for people with speech impairments or anxiety.
A fully integrated Stripe-based subscription system lets users unlock premium features and scale access—no manual integrations or plugins required.	Subscription systems are often basic, limited, or require users to set up and manage third-party services themselves.
PodGen offers a responsive, modern interface that works seamlessly on mobile, tablet, and desktop devices.	User interfaces vary widely and may not be fully optimized for all device types or screen sizes.
As a web-based platform, PodGen can be accessed from any location with an internet connection, with no installation needed.	Some traditional platforms require mobile apps or desktop applications and may not be accessible from every device.

Table 1: Comparison and Analysis

2.3.5 SCREENSHOTS



3. CONCLUSION

Podgen offers a flexible, user-friendly, AI-powered platform for podcast productions. Eliminate traditional barriers by using modern technologies such as speech from text, image generation, and cloud-based hosting. By ensuring your podcast, Podgen promotes a more integrated environment for creators, regardless of your financial situation or language skills. Podgen is ready to change the podcasting industry with future improvements such as humanized voice, podiums and joint characteristics.

ACKNOWLEDGEMENT

We sincerely thank our project guide, Prof. S.M Shelke, for her valuable support and guidance throughout this project. We also appreciate our institution and peers who contributed directly and indirectly to PodGen's development.

REFERENCES

1. Amsaleg, L., Chelly, O., and Romero, A. "Deep learning-based speech-to-text services for next-gen applications." Proceedings of the 2021 ACM Conference on Audio-Visual Speech Processing, 2021, pp. 12–19.
2. Brown, T. B., et al. "Language models are few-shot learners." Advances in Neural Information Processing Systems, 33, 1877–1901, 2020.
3. Cunningham, S., and McMahon, T. "Artificial intelligence in content creation: A study of AI-based podcast tools." Journal of Creative Technology, 14(2), pp. 45–60, 2022.
4. Devlin, J., Chang, M. W., Lee, K., and Toutanova, K. "BERT: Pre-training of deep bidirectional transformers

for language understanding." Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1, 4171–4186, 2019.

5. Dhir, R., and Singh, J. "Podcasting platforms: Analysing subscription models and user engagement." *International Journal of Digital Media & Marketing*, 7(3), 105–117, 2021.

6. Goodfellow, I., et al. "Generative adversarial nets." *Advances in Neural Information Processing Systems*, 27, 2672–2680, 2014.

7. Gruber, M., and Hartley, R. "Vercel and Next.js: Optimizing cloud infrastructure for scalable SaaS applications." *Cloud Computing and SaaS Systems Journal*, 8(2), 150–162, 2023.

8. Kang, H., and Lee, S. "Next.js framework: A case study on server-side rendering and static site generation for web applications." *Journal of Web Engineering*, 12(1), 23–36, 2023.

9. Karras, T., Laine, S., Aila, T. "A style-based generator architecture for generative adversarial networks." *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

10. O'Reilly, P., and Wilson, J. "Subscription-based AI tools for content creation." *AI and Society*, 36(1), 91–110, 2022.

11. Ping, W., Peng, K., and Chen, J. "Deep voice 3: Scaling text-to-speech with convolutional sequence learning." *arXiv preprint arXiv:1710.0765*, 2017.

12. Ramachandran, K., and Mathew, D. "Convex as a serverless backend for real-time applications." *Journal of Software Architecture*, 5(4), 78–90, 2023.

13. Reed, S. E., Akata, Z., Yan, X., et al. "Generative adversarial text-to-image synthesis." *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, 2016.

14. Schmidt, F., and Martin, D. "Leveraging Clerk for seamless user authentication in SaaS platforms." *International Journal of Cybersecurity & Authentication*, 11(1), 25–38, 2023.

15. Thompson, A., and Zhou, Y. "AI-generated media: Ethical implications of text-to-speech technologies." *AI Ethics and Governance Journal*, 10(2), 134–145, 2023.

16. Wang, Z., and Zhao, L. "Building scalable SaaS platforms using Stripe for subscription management." *Journal of Cloud Computing Solutions*, 9(3), 208–221, 2022.

17. Oord, A. v. d., Dieleman, S., Zen, H., et al. "WaveNet: A generative model for raw audio." *arXiv preprint arXiv:1609.03499*, 2016.

18. Kumar, S., Pathak, S., and Kumar, B. "Automated detection of eye-related diseases using digital image processing." *Handbook of Multimedia Information Security: Techniques and Applications*, pages 513–544, 2019.

19. Kim, H. E., et al. "Transfer learning for medical image classification: A literature review." *BMC Medical Imaging*, 22(1):69, 2022.

20. Gulshan, V., et al. "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs." *JAMA*, 316(22):2402–2410, 2016.