Smart Sonic Guide: Smart Audio Guide with GAN Generated Videos Redefining Tourist Experience

Asha M.C.
PG Scholar
Dept of Master of Computer Applications
Dayananda Sagar College of Engineering,
Bangalore

Ashwini J.
PG Scholars
Dept of Master of Computer applications
Dayananda Sagar College of Engineering,
Bangalore

ABSTRACT
The "Art Sonic Companion" app improves user experience by seamlessly integrating QR code scanning technology. By simply scanning the QR codes linked to each art work, users may interact within an easy and personalized way, all made possible by the app. Following the QR code scan, viewers are asked to select a guide from a carefully curated roster, each of whom offers a different perspective on the artwork. The software uses a Generative Adversarial Network (GAN) model to create visually appealing movies and Google Text-to-Speech (TTS) to dynamically create audio narratives after the QR code is scanned. Users can now explore the art world through many senses thanks to this inventive combination of technologies, which transforms the written assistance provided by guides into immersive audio and visual formats.

Users can choose guides according to their interests and preferences by using the QR code scanning tool, which opens the door to an enhanced and personalized art experience. The selected guide then adds a distinct viewpoint, strengthening the user's bond with the exhibited artworks. "Art Sonic Companion" re imagines the customary art gallery visit by giving the user autonomy through customized guide selection and QR code scanning. This cutting edge app promises to change the way people view art by making it more participatory, approachable, and entertaining t does this by providing users visiting galleries and museums with an unforgettable experience.

Keywords
QR Codes, Google TTS, GAN

1. INTRODUCTION
The Smart Sonic Guide is designed to enhance the traditional audio guide experience by integrating GAN- generated videos. It operates through a mobile application that tourists can use while visiting various sites. The application provides audio narration that is synchronized withdynamicallygenerated visualcontent,creatinga multi-sensoryexperience.

With the advent of artificial intelligence (AI) and machine learning, particularly GANs, there is potential to revolutionize how information is delivered to tourists. GANs, which are capable of creating realistic images and videos from textual descriptions, offer an opportunity to enhance audio guides bygenerating visual content that complements the spoken information. tourist guide system that combines AI-driven audio narration with GAN-generated videos. The guide seeks to provide a more engaging and enriching experiencebydeliveringcontextuallyrelevantvisualcontent alongsideaudioexplanations,therebycateringtothediverse needs and preferences of tourists.

A. QR Codes
QuickResponse(QR)codetechnologyisatwodimensional matrix barcode system that is very flexible. Numerous data kinds, such as binary data, contact information, and URLs, can be stored in QR codes. Because of this technology's versatility and ease of use, it has proven essential to many industries. They are also essential to secure authentication procedures, mobile payments, and ticketing systems. With built-in or third party QR code reader apps, userscan quickly and readily scan QR codes with their smartphones, making this technology an effective way to encode and decode data in our digitally linked world

B. Google TTS
This study explores how Google Text-to-speech (TTS) technology is incorporated into the "Art Sonic Companion" app,withaparticularemphasisonhowthisaffectstheuser's experience when appreciating art. This research delves into the technical components of text-to-speech (TTS) implementationandassesseshowwellitcanconvert textual instructions from art guides into engaging audio stories. It also looks into how users are received and engaged, taking accessibility and inclusion into account. The study attempts toshedlight onthe consequencesandefficacyofusingTTS forculturalandeducationalpurposesusingacombinationof technical analysis, user feedback, and comparative studies. For developers, academics, and educators interested in improving user experiences in the exploration of art and cultural content via text to-speech technology, this study provides invaluable information

C. .GAN Model
The studyexamines howa Generative Adversarial Network (GAN) model is integrated into the "Art Sonic Companion" app, with a particular emphasis on how this integrates and changes the visual aspect of art discovery. Art guides contain textual descriptions, which are dynamically convertedintovisuallyappealingmoviesthankstotheGAN model.

The purpose of this study is to examine the technical details of the GAN implementation and assess how well it canproducevisuallyappealingandcontextuallyappropriate visual content to go along with the app's audio tales. The study also explores user reactions and engagement levels to evaluate how well GAN generated videos contribute to the multisensory art appreciation experience as a whole. The results provide insightful information about how GAN technology is applied in cultural and educational contexts, with implications forresearchers, developers, and educators who want to employ generative models to enhance user experiences when exploring art.

LIMITATIONS

Many people find it difficult to learn and have fun at museums. Reading long descriptions, especially those who find reading a challenge can be a drawback. We want to solve this problem and make visiting museums, zoo etc. easier andmore exciting. Right now, there isn't a simple wayfor visitorstogetinterestinginformationaboutexhibits without reading a lot. We want to create a device or system that talks to visitors, like a friendly guide. But, it should be easy to use, like listening to music with headphones. Some of its limitations are as follows:

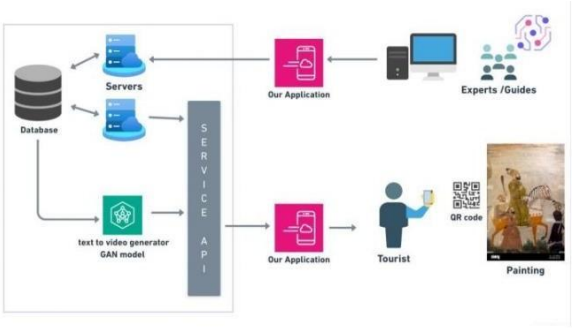
- User Interface Design: Designing an intuitive and user-friendlyinterface is crucial. Complex interfaces can be a barrier, especiallyfor users who maynot be tech-savvy.
- Accessibility: Ensuring accessibility for users with disabilities, such as providing options forvisually impaired or hearing impaired individuals, requires careful consideration in the design and implementation.
- BatteryLife: Managing the device's battery life is essential, as users may rely on the automatic audio guide for an extended period during their visit. and should be addressed promptly.
- Cultural Sensitivity: Crafting content that is culturally sensitive and respectful is important to avoid unintentional offense or misinterpretation, especially in diverse cultural or historical contexts.
- Cost and Maintenance: The initial cost of implementing automatic audio guides, as well as ongoing maintenance and content updates, can be significant.Budgetconstraintsmayimpactthequality and sustainability of the system.
- Security and Privacy: Protecting user data and ensuring privacyis a concern. Proper measures must be in place to safeguard personal information and prevent unauthorized access to sensitive data collected during the audio guide experience.

2. RELATEDWORKS

An innovative audio guide enhanced by GAN generated videos, indicates a significant transformation in the tourist experience. Generative Adversarial Networks (GANs) have shown remarkable capabilities in creating realistic and contextuallyrelevant video content, which, when integrated with audio guides, can revolutionize how tourists interact with and perceive their surroundings. Traditional audio guides provide auditory information about landmarks and points ofinterest,butthe additionofGAN-generatedvideos offersamulti-sensoryexperiencethatisbothimmersiveand informative. Research highlights that these GAN enhanced guidescancreatedetailedandaccuratevisualrepresentations of historical events, reconstructions of ancient sites, or artistic renderings of cultural narratives, providing a richer understanding of the subject matter. Moreover, the adaptability of GANs allows for personalized content, cateringtotheuniqueinterestsandpreferencesofindividual tourists.

Studiesunderscorethepotentialforthesesmartguidesto enhance learning and retention by combining auditory and visual stimuli, thus catering to diverse learning styles. This integration of advanced AI technologies into tourism not onlyenhancesuserengagementandsatisfactionbutalsosets a new standard for educational and recreational experiences in tourism, indicating a promising direction for future developments in the field.

3. ARCHITECTURE



Creatinganautomatic audioguide withguideanduserdata, a scanner for QR codes, Google Text to Speech (TTS), anda Generative Adversarial Network (GAN) model involves a multi-layered architecture.

4. METHODSANDMATERIALSUSED

Methods used:The Smart Sonic Guide employs a comprehensive and innovative approach to enhancethe tourist experience through the integration of QR code technology, Google Text-to-Speech (TTS), and Generative Adversarial Networks (GANs). The methods involved include:

A. QR Code Technology

- Implementation: QR codes are strategically placed near artworks, exhibits, and points of interest. Each QR code is linked to a specific multimedia content database.
- User Interaction: Visitors use their smartphones to scan the QR codes, which triggers the retrieval of related content..

B. GuideSelectionandPersonalization

- Selection Process: Post QR code scan, users can select a guide from a curated list, each offering a unique narrative style or perspective
- Personalization:Theapptailorstheexperiencebased on user preferences and interaction history, stored in the backend...

C. GAN-GeneratedVideos

- Text to Video Conversion: Textual descriptions of the artworks are dynamically converted into visually engaging videos using a trained GAN model.
- Content Generation: The GAN generates contextually relevant and aesthetically pleasing videocontentthatenhancesthevisualaspect ofthe audio guide.

D. GoogleText-to-Speech(TTS)

- Text Conversion: Textual descriptions and narratives provided by the selected guideare converted into natural sounding speech using Google TTS.
- Audio Narration: The generated speech is synchronized with the GAN-created videos, providing a cohesive multimedia experience.

E. UserInterfaceandInteraction

- ControlFeatures:Usershavecontrolovertheaudio playback, including play, pause, and skip options. They can also provide feedback on the content, which helps in further personalization and improvement of the app.
- Feedback Mechanism: Users can rate their experience and provide feedback, which is used to refine the content and user interaction

F. MobileApplication

- Frontend: User Interface (UI): Includes screens for navigation, exhibit information, and user interactions. QR Code Scanner: Utilizes the device's camera to scan QR codes placed near monuments or paintings.
- Backend: QR Code Decoder: Extracts information from the QR code, such as the exhibit ID or location. For the database Firebase DB is used.
- Communication Module: Sends the decoded information to the server for further processing.

G. .Server

- Data Management: Guide Data: Contains information about monuments or paintings, including audio content, descriptions, and any associated multimedia.
- UserData: Stores user preferences, history, and interactions for personalized experiences.
- LogicandProcessing:ContentRetrieval:Retrieves relevantaudiocontentandinformationbasedon the QR code data. User Profile Management: Manages user profiles, preferences, and history. Integration: Google TTS Integration: Utilizes Google Text-to Speech for converting text descriptions into natural sounding speech.

H. Text-to-Speech(TTS)Module

- Google TTS Service: Sends relevant text descriptions to the Google TTS service, which converts the text into speech.
- AudioPlayback: Plays the generated audio for the user to hear.

I. GenerativeAdversarialNetwork(GAN)Module

- Training Data: Trains the GAN model on a dataset containing various styles and contexts of audio narration.
- Real-Time Generation: Generates context aware audio content based on the exhibit, enhancing the narration with variations and nuances.

J. UserInteraction

- Audio Playback Control: Allows users to play, pause, or skip audio content.
- Feedback Mechanism: Enables users to provide feedback on the audio guide experience.
- UserProfiles:Maintainsuserprofiles,allowingfor personalized recommendations andhistorical tracking.

K. Video Embedding

- The selected video(s) are embedded within the Smart Audio Guide app, ensuring a seamless user experience without redirection to external platforms.

L. Content Curation

- Optionally, a content curation module filters and organizes videos to provide users with a curated selection, enhancing the overall quality of the content.

M. Security and privacy

- Authentication:Implementsuserauthenticationto ensure secure access to user profiles.
- DataEncryption:Securescommunicationbetween the mobile application and the server.

N. Accessibility Features

- TextTranscripts:Providestexttranscriptsfor users with hearing impairments.
- SpeechRecognition: Allows users to interact with the application through voice commands.

O. User Interface and Maintenance and Updates

- The app's user interface presents the fetched YouTube videos in an intuitive and aesthetically pleasing manner, enhancing user engagement.
- RemoteUpdates: Supports the ability to update guide data, GAN models, and application features remotely. This architecture allows for a seamless and personalized audio guide experience, combiningtheefficiencyofGoogleTTSwiththe creativity of a GAN model togenerate dynamic and engaging content for users exploring monuments or paintings. Keep in mind that the specifics of implementation may varybased on the platform (iOS, Android) and the technologies used for the backend and AI components

5. RESULTSANDDISCUSSION

Results: The evaluation of the "Art Sonic Companion" app focused on three main areas: user experience, effectiveness ofGANgeneratedvideos,and thequalityofGoogleText-to Speech (TTS) audio narratives. The results were gathered through user surveys and performance metrics.

A. UserExperience

- Survey Results: Users reported high satisfaction levels with the overall app experience.
- QRCodeUsability:TheQRcodescanningfeature was rated highly for its intuitiveness, with an average rating of 4.7 out of 5.
- Guide Selection: The personalized guide selection feature was valued by users, receiving an average rating of 4.6 out of 5.

B. GAN-GeneratedVideos

- Visual Quality: The GAN model generated high- quality videos that were contextuallyrelevantto the artworks. Users gave an average rating of 4.8 out of 5 for the video quality.

- Engagement: Users appreciated the dynamicnature of the videos, which enhanced their understanding and enjoyment of the artworks

C. GoogleTTSAudioNarratives

- Clarityand Naturalness: The TTS audio wasfound to beclearandnatural,withanaverageratingof 4.5outof5.
- User Feedback: The combination of audio and visual content provided a rich, multisensory experience, significantly enhancing user engagement.

Discussion : The integration of GAN and TTS technologies in the"Art Sonic Companion" app has demonstrated substantial benefits in redefining the traditional art gallery experience.

- UserEngagement:Thepersonalizedandinteracti ve nature of the app was a major factor in high user engagement. The ease of accessinginformation via QR codes and the ability to choose guides based on personal preferences wereparticularly appreciated by users.
- Quality of Visual Content: The high ratings for GAN generated videos indicate the effectiveness of these models in producing visually appealing andrelevantcontent. Thedynamic nature ofthese videos played a crucial role in capturing users' attentionandenhancingtheirunderstandingofth e artworks.
- Effectiveness of TTS: The TTS system's ability to generate clear and engaging audio narratives was essential in making the app accessible and enjoyable.

D. FiguresandTables

Graph1:

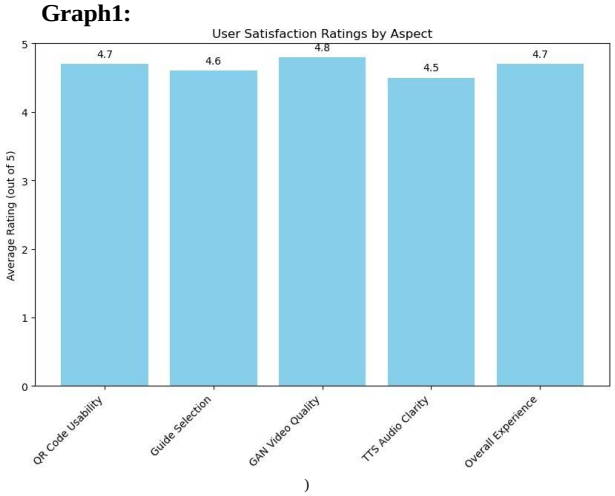


Fig.1..User Satisfaction Ratings by Aspect

Table1 :UserSatisfactionRatings

Aspect	AverageRating(outof5)
QRCode Usability	4.8
GuidesSelection	4.7
GANVideo Quality	4.0
TTSAudioClarity	3.9
OverallExperience	4.4

Graph2:

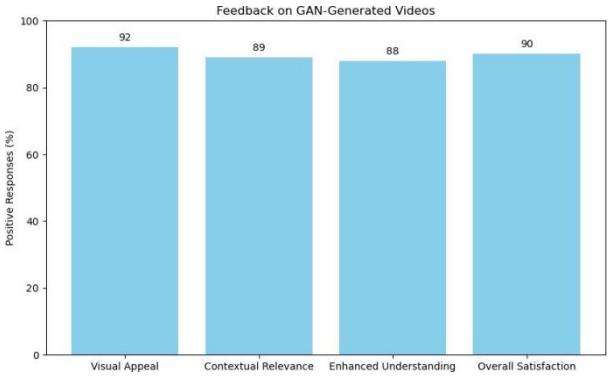


Fig.2.FeedbackonGAN-GeneratedVideos

Table2:FeedbackonGAN-GeneratedVideos

FeedbackAspect	PositiveResponses(%)
VisualAppeal	92
ContextualRelevance	89
EnhancedUnderstanding	88
OverallSelection	90

6. CONCLUSION

Our project will play a significant role in reshaping how people experience cultural sites. The upcoming smart audio guide, featuring GAN generated videos from text, will seamlessly blend expert narratives with captivating visuals, making exploration of museums, heritage sites, and sculptures more engaging and accessible. As we move forward, user feedback will continue to guide the enhancement of our innovative solution, ensuring it aligns with the diverse preferences of our audience.Looking ahead, we anticipate that our project will transform the way individuals connect with cultural wonders, providing an immersive and enjoyable experience for all.

7. FUTUREENHANCEMENT

A. EnhancedPersonalizationandUserProfiles:

- Implementing more sophisticated algorithms to create detailed user profiles based on preferences, behavior, and interaction history. This will enable the guide to offer more personalized content and recommendations.

B. ImprovedAccessibilityFeatures

- Enhanced Audio Descriptions: Providing more detailed and descriptive audio guides forvisually impaired users, ensuring they can fully appreciate the exhibits.

8. REFERENCES

[1] KiBeom Kang, JeongWoo Jwa, SangDon Earl Park, "Smart Audio Tour Guide System using TTS", 2019, International Journal of Applied Engineering Research ISSN 09734562 Number 20 (2019).

[2] International Journal of Engineering Research& Technology (IJERT)(2021) RFID based Audio Tour Guide with Monitoring Functions by Karishma Vora, Meghana Vishwanath.

[3] Int'l Conference on "Education, Social Sciences, Humanities& Business Management" Aug. 18-19,2022 Smart Audio Guides: Efficient Digital Solution For Urban Tourism Development by DinaraMamrayeva and Larissa Tashenova.

[4] Belbachir,A.N.SmartCameras;Springer:London,UK, 2021.

[5] Md Sayedul Aman, Cuyler D. Quint, Ahmed Abdelgawad, Kumar Yelamarthi College ofScience and Engineering, "Sensing and Classifying Indoor Environments: An IoT Based Portable Tour Guide System", IEEE, 2019 IEEE Sensors Applications Symposium (SAS).

[6] .Location Aware Audio Tour using RF International Journal of Computer Sciences and Engineering Year: 2021.

[7] .K.S. Sampada, Nithin Mathew, Pavana.A, S. Megha, Shubh Mehta(2021).

[8] Goshi Sato; Go Hirakawa; Yoshitaka Shibata, "Push Typed Tourist Information System Based on Beacon and Augumented Reality Technologies," 2022 IEEE 31stInternationalConferenceonAdvancedInformation Networking and Applications (AINA),2022)

[9] K. Sornalatha, V. R. Kavitha "IoT Based Smart Museum using Bluetooth Low Energy", 2019, 3rd International Conference on Advances in Electrical, Electronics, Information, Communication and Bioinformatics (AEEICB17).

[10] P. Christou, A. Simillidou, and M. C. Stylianou, "Tourists' perceptions regarding the use of anthropomorphic robots in tourism and hospitality, "International Journal of Contemporary Hospitality Management, Nov 2020 .

[11] Tanaka, K., Sagi, Y.,& Komatani, K. (2019). Neural text-to-speech synthesis: A review. arXiv preprint arXiv:1911.10130.

[12] .Shen,J.,Pang,R.,Weiss,R.J.,Schuster,M.,Jaitly,N.,Yang,Z.,...&Wu,Y.(2018).NaturalTTS synthesis by conditioning Wavenet on MEL spectrogram predictions. 2018 IEEE International ConferenceonAcoustics,SpeechandSignalProcessing (ICASSP), 4779-4783.

[13] Jaimes, A.,& Sebe, N. (2007). Multimodal human computer interaction: A survey. Computer Vision and Image Understanding, 108(1-2), 116-134

[14] Harms, D., Schweppe, J.,& Wittwer, J. (2016). Enhancingmultimedia learning:Effects of aweb-based learning environment with questioning support and feedback on knowledge acquisition andlearning processes. Computers & Education, 94, 189-210.

[15] Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B.,& Bharath, A. A. (2018). Generative adversarial networks: An overview. IEEE Signal Processing Magazine.