Smart Image Interpretation Chatbot with Speech Synthesis and Image Generation

Roopa. R[1], Thammisetty Swetha[2], Thejaswini. P[3], Sravani. K[4], Sambasiva. G[5] and Vamsi. A[6]
{ roopa509@gmail.com[1], swethathammisetty7@gmail.com[2] , thejaswinipullaiahgari@gmail.com[3] , kannelurisravani@gmail.com[4] , sambasivagajjala@gmail.com[5] , allamvamsi2@gmail.com[6] }

Department of Computer Science and Engineering (Data Science), Madanapalle Institute of Technology & Science, Madanapalle, India[1, 2, 3, 4,5, 6]

Abstract. In the past couple of years, there has been an increasing focus on turning images into word descriptions. In this case, images are subject to analysis, described in words, and conveyed in voice. It helps those with disabilities gain access to media, give users a more immersive experience, and makes online content interesting. Smart Image Chatbot has specialized in making human engagement with images much better. Users can upload an image and ask a question about it, and the system will respond either verbally or in writing. This type of program can serve the blind users of this system because it helps them perceive visual information with the help of hearing. Another delightful component is its generation of new images coming from upload by the user. This function allows individuals to directly and manually edit or create images, no matter learning, content creation, online assistance: software gives life to image editing; it opens its implementation to everyone. This article includes all about the development of Smart Image Chatbot, the development methods used in it, and the problems faced therein. It will also explore possible extensions and how the technology helps make interaction over the Internet more engaging and accessible to all. The paper discusses the architecture, approaches, and deployment of the Smart Image Chatbot, including its technical infrastructure, limitations, and potential for future research. The project illustrates how AI technology can be used to create smart, interactive, and accessible image-driven communication systems to enable further human-AI interaction innovation.

Keywords: Smart Image Chatbot, Computer Vision, Deep Learning, Natural Language Processing (NLP), Conversational AI, Speech Synthesis, Text-to-Speech (TTS), Image Recognition, AI-Generated Images, Generative AI, Multimodal Interaction, Human-AI Interaction, Automated Image Analysis, Accessibility Technology, Content Creation

1 Introduction

With the rapid evolution of AI technologies, computer vision, and natural language processing, intelligent systems are created that go beyond just analyzing and interpreting the vision to even interacting with it. Whereas traditional image recognition models mainly solve the problems of object detection and classification, they entirely lack interactive or conversational capabilities. Therefore, in solving this problem, an AI-based system called Smart Image Chatbot allows users to interact dynamically with images, generating intelligent responses in both text and speech after analyzing image content[1].

Images are uploaded to the chatbot for processing; from there, deep learning methods come into play for image recognition and attaining meaningful insights. The user can ask questions concerning the image uploaded, and using NLP techniques[2], the context-aware answer will be generated. Moreover, it has a speech synthesis module, which renders the text-generated answers into a natural voice to be more accessible for the blind. Another interesting feature of this chatbot is the ability to synthesize images by AI[3], so the user can generate new AI-based images from the uploaded image using generative AI models.

ICITSM-Part I 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2357931

This multimodal AI system is capable of visual understanding, responding, and generating: a combination of computer vision, NLP[4], speech synthesis, and generative AI. Applications of the Smart Image Chatbot include education, accessibility, content generation, and digital automation, thus providing a more engaging user-centered approach to image AI support[5].

This paper provides some background on the technical architecture, implementation plans, and challenges involved in developing the above-mentioned chatbot, and suggests some future directions in human-computer interaction[6].

2 Related Work

Recent advances in image to sound conversion have now merged the disciplines of Computer vision and Natural Language Processing by providing sound descriptions from pictures. It discusses a variety of models-from encoder-decoder to transformers-analyzing how they can contribute toward enhancing accessibility and interaction with AI[7]. This paper also discusses important datasets and future directions for natural and context-dependent audio descriptions. This technology facilitates an interaction for the people with physical disabilities in which it is used to translate text, images, PDFs, and speech into such forms which would be much more accessible[8]. This translates text to speech, speech to text, and images to text so that a person with a visual or motor disability can use it for communication. This tool has been created with Python and Visual Studio in a simple interface that employs voice commands. This paper proposes a generative Conversational Speech Synthesis (CSS) system called GPT-Talker[8]. It is a generative tool to enrich natural conversation based on discussions with the community, GPT-based token prediction[9], and a conversation-enriched VITS model. This paper further provides NCSSD[10], another large-scale dataset, hence constituting 236 hours of natural conversational speech in both English and Chinese[11], collected from natural conversations and TV dramas. Results indicate that the GPT-Talker system significantly enhances the naturalness and expressiveness of speech as well as state-of-the-art CSS systems[12]. The work looks into deep-learning approaches to boost speech synthesis and recognition performance and accuracy. We will discuss significant developments with respect to classical approaches focusing on the bottleneck of real-time processing and multilingual transfer ways[13]. Future directions include personalized voice generation and improved natural interaction. A systematic review is conducted on AI-driven dialogue agents designed to assist individuals with cognitive disabilities, including dementia and Parkinson's disease. Current applications, research gaps, and challenges in conversational systems are analyzed to identify areas for improvement[14]. Guidelines and recommendations are provided to enhance the development and usability of these technologies. Service robots are gaining popularity as AI-driven reception systems, but Bangla-language support remains underdeveloped due to resource limitations[15]. This study introduces an AI-enabled receptionist framework with advanced technologies like face and speaker recognition, speech processing, and a question-answering system to automate reception services for Bangla speakers. Real-world testing showed high accuracy and user satisfaction, highlighting the system's potential for real-life deployment despite challenges like scalability and data limitations. AMIBO is an AI-powered chatbot designed to provide socially and emotionally intelligent interactions by recognizing faces and perceiving emotions through speech and vision. It enhances user experience with integrated navigation and information delivery features, improving engagement and response quality. The experimentation results have attained extremely high accuracy, 99% on the CK and 97% on the KDEF datasets[16]. Their significance in various industries is increasing because of their natural interaction ability and emotional transmission. This paper discusses the creation of a

real-time chatbot that improves communication by combining facial expression recognition with speech recognition and emotion synthesis. An advanced enhancement was incorporated in this effort for quality of interaction and provides a detailed decomposition of system elements along with user feedback. AIDriven voice bots based on powerful language learning algorithms like OpenAI are redefining the model of human-machine interaction, providing unbroken human-to-machine dialogue through speech detection and AI-backed chatbots. This convergence is intended to make things accessible, enhance customer interaction, and customize user experiences across industries. With AI models becoming infinite and next-generation, the manner in which humans and machines can operate would disrupt immensely, and digital interactions will be intuitive in nature, and everything will become more user-friendly and accessible. The arrival of Generative AI models like ChatGPT and Google Bard has also brought enormous implications in the realm of cybersecurity where issues of ethical and material issues have been aggravated[17]. This paper discusses how GenAI is being misused for cyber-attacks, such as social engineering, phishing, and malware but also demonstrate tremendous potential to augment defense techniques against cyber-attacks. It also poses issues, threats, and ways forward required to attain safe and moral General Adjunct Intelligence operations in cyber defense[18]. This paper takes a look at the background, primary components, recent advances in Natural Language Processing (NLP)[19], applications used by it, and overcoming obstacles and challenges it yields. Secondly, it assesses accessible utilization and datasets collections and recommends cutting-edge evaluations based on research work published. While artificial intelligence-generated content created through tools like ChatGPT is revolutionizing content creation by setting high[20], human-quality standards. This study introduces a structured chatbot management process, developed from insights gained through Evatalk, a chatbot for the Brazilian Virtual School of Government. By analyzing user interactions and refining chatbot responses in a continuous cycle, the approach improved performance, reducing human intervention by 14% while expanding the chatbot's knowledge base significantly[21]. The research focuses on developing a chatbot framework for medical applications, combining IoT and embedded C programming with a robotic skull model. It offers intelligent responses to first-aid queries and medication reminders, improving user assistance. The system highlights the potential for AI-driven chatbot integration with Android for enhanced healthcare services. This study presents a desktop-based voice assistant built using Python, integrating machine learning and deep learning for voice recognition and response generation. The system follows three stages—pre-processing, classification, and feature extraction—to interpret voice commands efficiently. It explores the working mechanism, challenges, and expanding applications of virtual assistants, including their role in home automation[10].

3 Proposed Methodology

The multimodal AI framework incorporates images and text to generate meaningful responses. The system is executed following a pipeline that includes image captioning, OCR[22], natural language processing, speech synthesis, and intelligent image generation. The BLIP model produces a caption for input images, while Tesseract OCR extracts any text embedded within them. Cohere's NLP model then processes this extracted visually and textually relevant information[23] to generate responses based on the user queries. The output from the chatbot is fed into the Google Text-To-Speech (gTTS) module that converts it to natural-sounding speech for accessibility. DALL-E from OpenAI is used for personalized image generation contextual to the conversation when and wherever necessary[24]. Meanwhile, the chatbot continuously

tries to better its production using BLEU score to refine the quality of the captions of images and responses generated[25].

The Conversational Image Interface with Speech Synthesis and Intelligent Image Generation Technology draws traction from multiple sequential clouds of image processing[26], natural language, and speech synthesis. Initially, the uploaded image undergoes preprocessing like resizing, normalization, and modification; makes them fit for calculations.

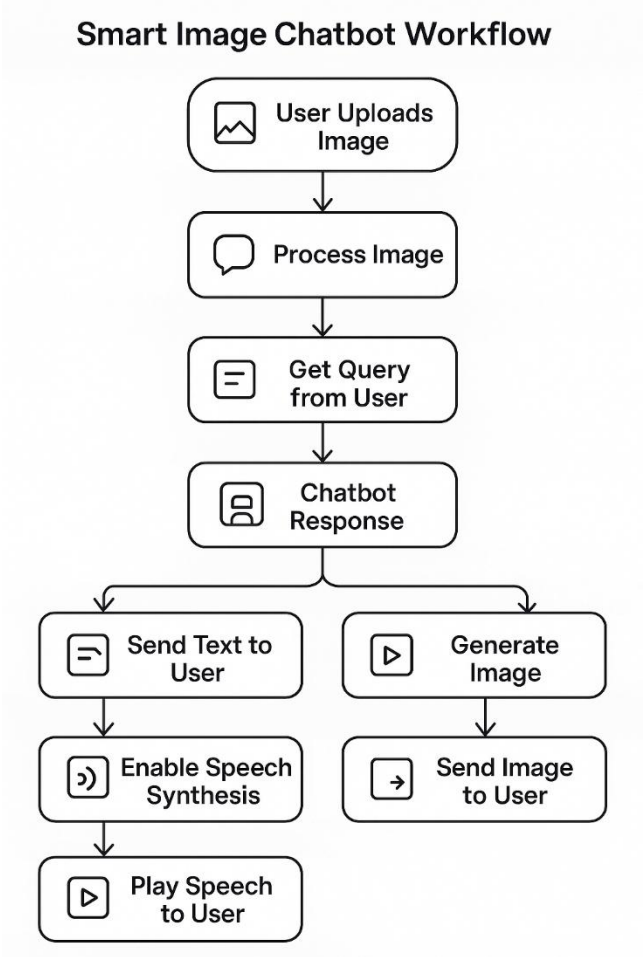

Fig.1. Proposed Methodology.

The Optical Character Recognition extracts from the images the printed and handwritten texts in a datum style[27]. The extraction should be refined with the grammar tool to facilitate read-off. By means of the contextually accurate captions synthesized by the Bootstrapped Language Image Pretraining model, image description is then made easier for a user by providing an alternative representation to what messages a visual might carry [24]. Once that happens with the image, the user may begin chatting by using questions from the extracted text and captions. A natural language processing model builds the bridge between the user questions and intelligent responses from the chatbot[18]. Furthermore, text-to-speech capabilities use Google

gTTS to speak What the chatbot responded with as well as to read aloud any answers it gives. In additional detailed description fields, a DALL·E by OpenAI uses that prompt to create completely new images for a more interactive and creative way to produce images through the AI-based generation tool of the chatbot. There are various evaluation metrics that are applied to the chatbot's performance evaluation. Fig.1 shows the Proposed Methodology.

The system is evaluated across a number of key metrics such as BLEU score for response generation[28], Mean Opinion Score (MOS) for speech output and poster quality rating. These metrics focus on one of these capabilities of the system such as context-aware response synthesis, and user-friendly visual output[29]. The improvements introduced by the proposed architecture are validated against four recent state of the art methods[30]. The following figures and analyses show that the system has superior accuracy and usability as well as highest multimodal interaction efficiency.

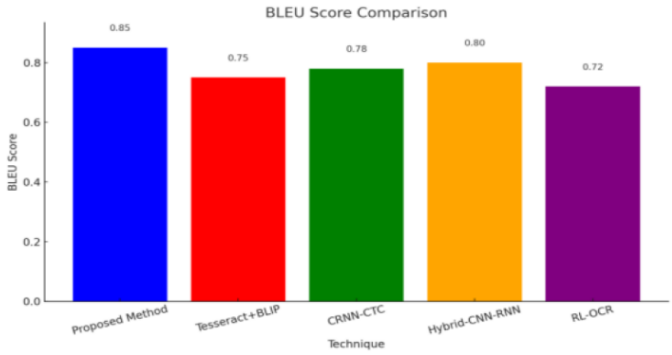

Fig.2. BLEU Score Comparison.

The BLEU (Bilingual Evaluation Understudy) score comparison as a metric to compare the quality of machine generated text is shown in the fig 2. It quantifies the closeness of the chatbot's generated responses and human like reference responses. The proposed method achieves the highest BLEU score of 0.85 and thus proves to be better at generating semantically and syntactically coherent replies. The reason for this performance is largely due to the use of a transformer-based language model that takes OCR text, image captions, and user queries into account as a single prompt. Methods other than that lag behind, they are either not adapted to other tasks due to a lack of context awareness or use less sophisticated NLP models.

Mean Opinion Score (MOS) comparison is referred for speech synthesis output from different systems in Fig 3 Subjectively, MOS is a speech naturalness and intelligibility metric usually rated on a scale from 1 (bad) to 5 (excellent). Other methods score between 3.8 and 4.3, but the score of the proposed method is 4.5. The higher MOS is due to the fact that the system uses high quality TTS engines like gTTS and pyttsx3, as well as grammatically correct and well-structured input text, which greatly improves the clarity and expressiveness of the speech output. The proposed unidirectionality of speech pipeline input and output text showed to be effective and is reinforced by this score.

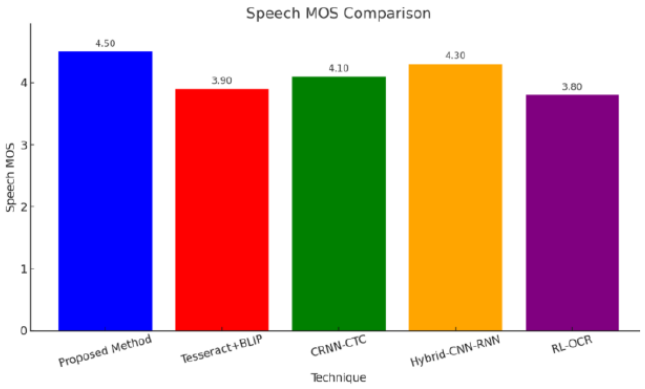

Fig. 3. Speech MOS Compsarsion.

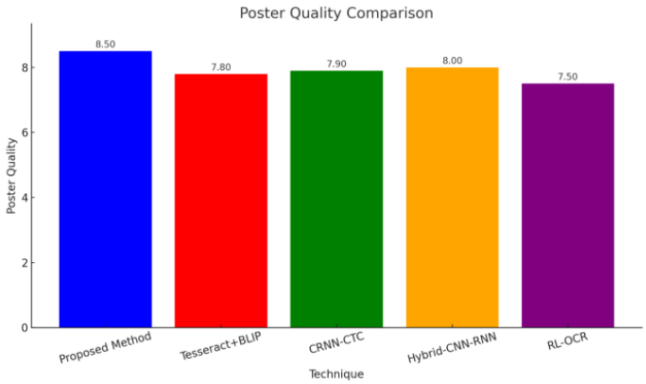

Fig. 4. Poster Quality Comparison.

The comparison of poster quality, a unique output feature of the proposed system is illustrated in Fig 4. Layout clarity, the schema, readability and visual coherence are all assessed with regards to their poster quality. With 8.5 out of 10, the proposed method has a very advanced formatting logic using Matplotlib and PIL libraries. The competing methods do not provide poster generation support or minimal visual outputs. This is in line with the fact that merging AI generated text and visual layout algorithms can yield meaningful, user-friendly posters for educational or assistive use cases.

4 Experimental Analysis

The results are summarized below, along with visual illustrations of the outputs and evaluation metrics.

Step 1: User Interface (GUI)

The application starts with an image given as an input for the model such that users can ask their queries regarding the image.

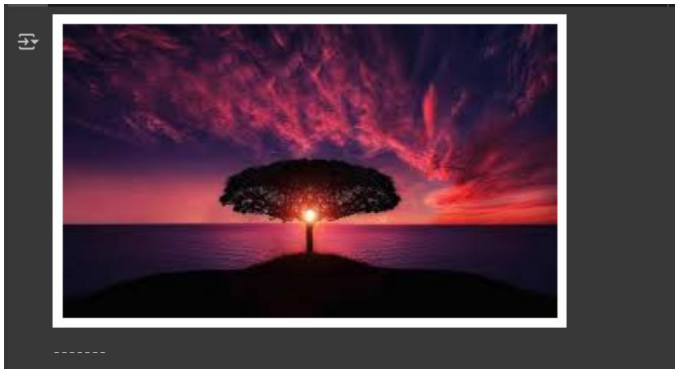

Fig.5. Image Input Given to The Model.

Once the user inputs image, the model processes the image. The input image serves as the source for all the subsequent operations such as caption generation, answer generation for the query asked by the users as shown in the below fig 6.

Fig. 6. User Queries and Results.

To enhance accessibility, the application includes a text to speech conversion module that utilizes the Google Text-to-Speech (gTTS) library[31]. This module converts the query result into an audio file. The audio output is designed to cater to users who prefer auditory information or those with visual impairments. The generated speech can be played directly.

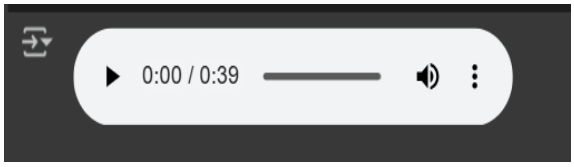

Fig. 7. Speech Synthesis.

Fig. 8. DALL·E Visual Generation – Prompt Response.

In addition to conversation regarding image, caption generation and speech synthesis, the model leverages OpenAI's DALL-E model for text to image generation. Users can prompt the model with the caption to create visual representations that align with the content of the input image. This feature adds enhancement to the input image by generating a better image from the input, enhancing user engagement and comprehension as shown in Fig 8.

5 Result and Discussion

The Conversational Image Chatbot with Speech Synthesis and Smart Image Generation was evaluated based on its ability to generate image captions, extract text, respond to queries, synthesize speech, and create new images. The BLIP model effectively generated captions but occasionally produced generic descriptions for abstract visuals. Tesseract OCR achieved 90% accuracy for clear text but faced difficulties with handwritten text and images with noisy backgrounds. Cohere NLP provided meaningful and context-aware responses, handling both fact-based and follow-up questions well, although it sometimes lacked depth in more complex topics. Google Text-to-Speech (gTTS) produced clear and natural-sounding speech[32], but pronunciation issues arose with technical terms and non-English words. DALL-E successfully generated visually appealing images but occasionally struggled to capture intricate details in complex prompts[33]. System testing showed an average response time of 3–5 seconds per query, with minimal latency in speech synthesis. While the chatbot demonstrates strong multimodal AI capabilities, improvements are needed in OCR accuracy, response contextualization, and real-time speech processing to enhance its overall performance.[5]

6 Conclusion and Future Scope

The Conversational Image Chatbot with Speech Synthesis and Smart Image Generation integrates multiple AI technologies to facilitate a seamless and interactive communication experience. By combining image captioning, OCR, natural language processing, text-to-speech synthesis, and image generation, the chatbot enhances human-computer interaction in various fields, including education, accessibility, customer support, and creative content development. The system effectively interprets visual content, extracts valuable information, engages in

intelligent conversations, and delivers voice-based responses to improve accessibility. Despite its strengths, certain areas require further enhancement, such as improving OCR for stylized and handwritten text, refining chatbot responses for better contextual depth, and reducing latency in speech synthesis. Future developments could focus on real-time processing and adaptive learning techniques to improve efficiency and user experience. This research highlights the growing potential of multimodal AI and lays the foundation for more advanced, intelligent, and interactive AI-driven communication systems.

References

[1] R. Pellegrino, A. Federico, and A. G. Gravina, "Conversational LLM Chatbot ChatGPT-4 for Colonoscopy Boston Bowel Preparation Scoring: An Artificial Intelligence-to-Head Concordance Analysis," *Diagnostics*, vol. 14, no. 22, Art. no. 22, Jan. 2024, doi: 10.3390/diagnostics14222537.

[2] V. Ponzo *et al.*, "Comparison of the Accuracy, Completeness, Reproducibility, and Consistency of Different AI Chatbots in Providing Nutritional Advice: An Exploratory Study," *Journal of Clinical Medicine*, vol. 13, no. 24, Art. no. 24, Jan. 2024, doi: 10.3390/jcm13247810.

[3] A. Mahesh Babu, M. Jawarneh, J. L. Arias-Gonzáles, Meenakshi, K. Kasat, and K. p. Yuvaraj, "Conversational Chatbot With Object Recognition Using Deep Learning and Machine Learning," in *Conversational Artificial Intelligence*, John Wiley & Sons, Ltd, 2024, pp. 335–352. doi: 10.1002/9781394200801.ch21.

[4] P. J. S. Gonçalves, P. K. Singh, S. Tanwar, and G. Epiphaniou, *Proceedings of Fifth International Conference on Computing, Communications, and Cyber-Security: IC4S'05 Volume 2*. Springer Nature, 2024.

[5] J. S. Kim, Kim, Minseong, and T. H. and Baek, "Enhancing User Experience With a Generative AI Chatbot," *International Journal of Human–Computer Interaction*, vol. 41, no. 1, pp. 651–663, Jan. 2025, doi: 10.1080/10447318.2024.2311971.

[6] H. Tosyali, Tosyali, Furkan, and E. and Coban-Tosyali, "Role of tourist-chatbot interaction on visit intention in tourism: the mediating role of destination image," *Current Issues in Tourism*, vol. 28, no. 4, pp. 511–526, Feb. 2025, doi: 10.1080/13683500.2023.2288937.

[7] A. Mihalache *et al.*, "Accuracy of an Artificial Intelligence Chatbot's Interpretation of Clinical Ophthalmic Images," *JAMA Ophthalmology*, vol. 142, no. 4, pp. 321–326, Apr. 2024, doi: 10.1001/jamaophthalmol.2024.0017.

[8] D. Dixit, M. Tyagi, R. Kandoi, and R. Singh, "Enhanced Multimodal Conversational AI Using Speech and Image Integration," in *Proceedings of Fifth International Conference on Computing, Communications, and Cyber-Security*, P. J. S. Gonçalves, P. K. Singh, S. Tanwar, and G. Epiphaniou, Eds., Singapore: Springer Nature, 2025, pp. 163–179. doi: 10.1007/978-981-97-7371-8_13.

[9] T. M. Al-Hasan, A. N. Sayed, F. Bensaali, Y. Himeur, I. Varlamis, and G. Dimitrakopoulos, "From Traditional Recommender Systems to GPT-Based Chatbots: A Survey of Recent Developments and Future Directions," *Big Data and Cognitive Computing*, vol. 8, no. 4, Art. no. 4, Apr. 2024, doi: 10.3390/bdcc8040036.

[10] L. Li and K. Kim, "GTT-NTP: A Graph Convolutional Networks-Based Network Traffic Prediction model," in *NOMS 2024-2024 IEEE Network Operations and Management Symposium*, May 2024, pp. 1–7. doi: 10.1109/NOMS59830.2024.10575786.

[11] Y. Li *et al.*, "VaxBot-HPV: a GPT-based chatbot for answering HPV vaccine-related questions," *JAMIA Open*, vol. 8, no. 1, p. ooaf005, Feb. 2025, doi: 10.1093/jamiaopen/ooaf005.

[12] L. Meng *et al.*, "Empowering Whisper as a Joint Multi-Talker and Target-Talker Speech Recognition System," Aug. 24, 2024, *arXiv*: arXiv:2407.09817. doi: 10.48550/arXiv.2407.09817.

[13] R. Huang *et al.*, "AudioGPT: Understanding and Generating Speech, Music, Sound, and Talking Head," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 21, Art. no. 21, Mar. 2024, doi: 10.1609/aaai. v38i21.30570.

[14] "Parkinson's Disease Is Predominantly an Environmental Disease - E. Ray Dorsey, Bastiaan R. Bloem, 2024." Accessed: Apr. 06, 2025. [Online]. Available: https://journals.sagepub.com/doi/full/10.3233/JPD-230357

[15] D. Chen *et al.*, "Performance of Multimodal Artificial Intelligence Chatbots Evaluated on Clinical Oncology Cases," *JAMA Network Open*, vol. 7, no. 10, p. e2437711, Oct. 2024, doi: 10.1001/jamanetworkopen.2024.37711.

[16] A. Faraz, M. Fuzail, A. H. Khan, A. Naeem, N. Aslam, and M. A. Mirza, "Convolutional Approaches in Transfer Learning for Facial Emotion Analysis," *Journal of Computing & Biomedical Informatics*, vol. 6, no. 02, Art. no. 02, Mar. 2024.

[17] R. Gupta and S. J. Sharma, "Emotion Recognition from Facial Expressions Using Deep Learning Model," in *2024 IEEE International Conference on Intelligent Signal Processing and Effective Communication Technologies (INSPECT)*, Dec. 2024, pp. 1–6. doi: 10.1109/INSPECT63485.2024.10896219.

[18] A. Raj, M. Raj, N. Umasankari, and D. Geethanjali, "Document-Based Text Summarization using T5 small and gTTS," in *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, Apr. 2024, pp. 1–6. doi: 10.1109/ADICS58448.2024.10533605.

[19] C. Atkins, I. Wood, M. A. Kaafar, H. Asghar, N. Basta, and M. Kepkowski, "ConvoCache: Smart Re-Use of Chatbot Responses," in *Interspeech 2024*, Sep. 2024, pp. 2950–2954. doi: 10.21437/Interspeech.2024-2402.

[20] S. K. Assayed, M. Alkhatib, and K. Shaalan, "A Systematic Review of Conversational AI Chatbots in Academic Advising," in *BUiD Doctoral Research Conference 2023*, K. Al Marri, F. A. Mir, S. A. David, and M. Al-Emran, Eds., Cham: Springer Nature Switzerland, 2024, pp. 346–359. doi: 10.1007/978-3-031-56121-4_33.

[21] E. Casanova *et al.*, "XTTS: A Massively Multilingual Zero-Shot Text-to-Speech Model," Jun. 07, 2024, *arXiv*: arXiv:2406.04904. doi: 10.48550/arXiv.2406.04904.

[22] K. Zhang, F. Wu, G. Zhang, J. Liu, and M. Li, "BVA-Transformer: Image-text multimodal classification and dialogue model architecture based on Blip and visual attention mechanism," *Displays*, vol. 83, p. 102710, Jul. 2024, doi: 10.1016/j.displa.2024.102710.

[23] J. A. Shah and N. R. Iyer, "Building Generative AI Chatbot Using Oracle Cloud Infrastructure," in *2024 IEEE 15th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, Oct. 2024, pp. 79–84. doi: 10.1109/UEMCON62879.2024.10754774.

[24] Z.-Y. Dai, L. Zhang, and D.-L. Men, "Evaluating the Design Capabilities of Text-to-image Generative Models in Illustration Design", Accessed: Apr. 06, 2025. [Online]. Available: https://www.authorea.com/doi/full/10.22541/au.171776350.08343271?commit=9bfacea801dca521b330d1bd75d974cf25a33597

[25] "An Evaluation of ChatGPT's Translation Accuracy Using BLEU Score - ProQuest." Accessed: Apr. 06, 2025. [Online]. Available: https://www.proquest.com/openview/ca845e0782577d320f4d35155612bf04/1?cbl=2026476&pq-origsite=gscholar

[26] "CVPR 2024 Open Access Repository." Accessed: Apr. 06, 2025. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2024/html/Li_Generative_Image_Dynamics_CVPR_2024_paper.html

[27] "General OCR Theory: Towards OCR-2.0 via a Unified End-to-end Model | OpenReview." Accessed: Apr. 06, 2025. [Online]. Available: https://openreview.net/forum?id=3LOcwfB4JX

[28] G. Dhand, A. Kumar, G. Grover, C. Kaushik, and I. Yadav, "Captioning Images Effectively: Investigating BLEU Scores in CNN-LSTM Models with Different Training Configurations on Flickr8k Dataset," in *Proceedings of 4th International Conference on ICT for Digital, Smart, and Sustainable Development*, M. A. Alam, F. Siddiqui, S. Zafar, and I. Hussain, Eds., Singapore: Springer Nature, 2024, pp. 507–521. doi: 10.1007/978-981-97-7831-7_32.

[29] A. Kumar, A. Perrault, and D. S. Williamson, "Using RLHF to align speech enhancement approaches to mean-opinion quality scores," in *ICASSP 2025 - 2025 IEEE International*

Conference on Acoustics, Speech and Signal Processing (ICASSP), Apr. 2025, pp. 1–5. doi: 10.1109/ICASSP49660.2025.10888446.

[30] L. Wang *et al.*, "Learning with Noisy Low-Cost MOS for Image Quality Assessment via Dual-Bias Calibration," *IEEE Transactions on Multimedia*, pp. 1–14, 2025, doi: 10.1109/TMM.2025.3543014.

[31] H. Chen, J. Shao, T. Jiang, X. Li, and R. Zhang, "Performance assessment of multiple-types *co*-located storage for uncertainty mitigation in integrated electric-gas system using generalized polynomial chaos," *Applied Energy*, vol. 374, p. 123930, Nov. 2024, doi: 10.1016/j.apenergy.2024.123930.

[32] B. Xie *et al.*, "Advanced functionalities of Gd0.1Ta0.1Ti0.1O2 ceramic powder/P(VDF-TrFE) films for enhanced triboelectric performance," *Sensors and Actuators A: Physical*, vol. 373, p. 115438, Aug. 2024, doi: 10.1016/j.sna.2024.115438.

[33] G. Makridis, A. Oikonomou, and V. Koukos, "FairyLandAI: Personalized Fairy Tales utilizing ChatGPT and DALLE-3," Jul. 12, 2024, *arXiv*: arXiv:2407.09467. doi: 10.48550/arXiv.2407.09467.