iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.58787

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 12 Issue III Mar 2024- Available at www.ijraset.com

Conversational AI

Raunak Kandoi[1], Deepali Dixit[2], Mihul Tyagi[3], Raghuraj Singh Yadav[4]
Deptt. Of Comp. Sc. and Engg., School of Engg. And Tech, Sharda University, Gr. Noida, Uttar Pradesh, 201306

Abstract: ***Conversational AI systems are becoming increasingly popular across many industries and are transforming the way people interact with technology. For a more authentic, human-like connection and a smooth user experience, these systems should combine text-based interactions with multimodal capabilities. The authors of this work suggest a new approach to improving conversational AI systems' usability by combining speech and visual analysis. By combining visual and auditory processing capabilities, AI systems can better understand human inquiries and instructions. Both visual data and speech can be better understood with the use of computer vision algorithms and natural language processing techniques, respectively. Conversational AI systems can provide more accurate and tailored replies by integrating many modalities to better grasp human intent and context. The development of multimodal conversational AI presents a significant difficulty in ensuring the smooth integration of voice and visual processing units. A strong architectural design and advanced algorithms are necessary for the simultaneous synchronization and comprehension of data from several modalities in real-time. The system needs to keep track of the conversation's context even when it switches between different forms of communication so it can keep providing fair and relevant responses all through the engagement. Customization is key to making multimodal conversational AI better for users. Based on user data and preferences, the system may tailor interactions to offer more relevant ideas and support. Users are more invested in the AI system over time, and they have a better experience overall because to customization. Ensuring the privacy and security of important audiovisual data is of the utmost importance while building multimodal conversational AI. Strong encryption, anonymization technologies, and compliance with data protection regulations are vital for user privacy and system confidence. Continuous improvement is key to the success of multimodal conversational AI systems. The feedback from users can help the developers improve the system and add new features. Thanks to this iterative technique, the AI system stays flexible and can adjust to changing consumer preferences. By combining voice and picture processing, conversational AI systems have a great deal of promise for improving the user experience. Through the integration of visual and auditory signals, these systems have the ability to comprehend user intent more accurately, provide customized experiences, and completely transform the way humans engage with technology.***

Keywords: Conversational AI, Speech Processing, Image Processing, Multimodal Integration, Natural Language Processing (NLP), Computer Vision, User Experience, Personalization, Privacy, Continuous Improvement.

I. INTRODUCTION

As a subfield of AI, conversational agents are powerful resources for developing human-computer interactions. Chatbots and virtual assistants are agents that enable users to engage in natural language conversations for the purpose of acquiring information, completing tasks, and receiving assistance. Conventional conversational AI systems often ignore vital modalities like speech and images in favor of text interactions, despite significant advancements in natural language understanding and production. A sea change in human-computer interaction is on the horizon with the advent of conversational AI systems that incorporate image and speech processing. These systems are able to achieve a higher level of immersion, intuitiveness, and personalization by incorporating many modalities. The purpose of this introduction is to provide background on conversational AI, why it's important to combine voice and image processing, what problems and benefits multimodal interaction can bring, and how the research will be structured and executed. The integration of visual and auditory processing in conversational AI has immense promise for transforming the way humans engage with computers. More genuine, immersive, and personalized user experiences are possible with these technologies since they use numerous modalities. In order to help this ever-changing field advance, this study aims to examine the technical foundations, challenges, opportunities, and real-world applications of multimodal conversational AI.

Argument in Favor of Multimodal Communication

By combining many modalities, conversational AI systems aim to imitate human-like interactions. When people are having a conversation, they convey meaning and purpose through written words, vocal intonations, body language, and visual clues. Better understanding and reactions to user inputs are possible with conversational AI systems because they combine visual and audio processing capabilities. This leads to more effective communication and higher user satisfaction.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

A. Difficulties and Possibilities

The integration of visual and auditory processing in conversational AI presents both technical challenges and exciting opportunities. Ensuring the privacy and security of sensitive data, coordinating and comprehending inputs from several modalities, and retaining context across modalities are all technical concerns. Once these obstacles are overcome, there will be more opportunities for personalized interactions, improved comprehension of user intent, and heightened accessibility for users with different needs.

B. What is Multimodal Conversational AI

A huge step forward in AI, multimodal conversational AI enables sophisticated and natural-sounding interactions between intelligent systems and humans through a variety of communication channels. The goal of multimodal conversational AI is to create AI systems that can mimic human conversation by understanding and responding to user inputs in a variety of modalities, including as text, voice, visuals, and gestures. To effectively understand and respond to user inquiries and directives, these systems leverage advancements in speech recognition, natural language understanding, picture processing, and multimodal fusion methods.
What follows is a breakdown of the main components and characteristics of multimodal conversational AI:-

1) Integrating Modalities In order to fully understand what the user is saying, multimodal conversational AI integrates data from multiple sources. Consolidating user intent and context from audio, text, and visual inputs into a single, cohesive whole.
2) Understanding Natural Language (NLU) Conversational AI systems that support several modalities make use of advanced natural language understanding techniques to understand the intent and meaning of user inputs. In order to fully grasp the user's inquiries, requests, or instructions, it is necessary to analyze text, recognize speech, interpret images, and comprehend movements.
3) Understanding the Environment Better conversational cohesion and personalization are possible with multimodal conversational AI systems because they keep context across many interactions and modes. They enhance the user experience by remembering previous talks, understanding references, and adapting their responses to the current discourse.
4) Creating an answer Conversational AI systems that support many modes of user input can better understand their context and provide more relevant responses. A combination of textual, spoken, and visual elements can be used to generate responses that are informative, interesting, and genuine.
5) Modification and Personalization Conversational AI systems that support many modalities can personalize interactions based on user tastes, past actions, and current trends. They learn from users' input, adapt to new environments, and personalize responses, leading to better interactions overall.
6) Use case Virtual assistants, educational platforms, healthcare applications, entertainment, and customer service chatbots are just a few of the many industries that make use of multimodal conversational AI. Facilitating tasks, providing information, and enhancing user engagement, these technologies enable natural, intuitive, and organic interactions between people and machines.

C. Importance of Speech and Image Integration

Conversational AI systems need to incorporate visual and auditory modalities to enhance usability, context awareness, application scope, accessibility, and response comprehensiveness. By incorporating input from several modalities, these systems can enhance their understanding of user intent and deliver more interactive experiences. There are a number of crucial reasons why conversational AI systems that combine visual and audio modalities are absolutely necessary:-

1) Enhancement of the User Experience Conversational AI systems integrate visual and auditory senses to create a more intuitive and natural user experience. By combining spoken language with visual cues, users can have more natural and fruitful conversations with the system, just like they would with a real person.
2) Enhancing Our Understanding of User Intent By combining visual and auditory cues, a conversational AI system can gain a deeper understanding of the user's intent. As an example, a verbal input may provide background and explanation, while an image input provides visual context and additional details. When both methods are utilized simultaneously, the system is able to comprehend and respond to user demands more effectively.
3) Improving Resilience in Critical Situations When speech and visuals are combined, the system is able to remember more details from one conversation to the next. By providing additional context, images can enhance the system's ability to understand references, answer questions more precisely, and tailor its responses to each individual user. Consequently, discussions take on a deeper, more personalized quality.

4) Increasingly Diverse Applications The combination of voice and visual modalities opens up new possibilities for use in several fields. For example, in customer service, clients may choose to bolster their spoken comments with visual evidence such as images or screenshots in order to facilitate more effective problem resolution. By analyzing medical images and patient reports simultaneously, multimodal AI systems might enhance the precision of healthcare diagnosis.
5) Improving the Ease of Use The integration of voice and visuals can improve accessibility for individuals with different needs. When it comes to communicating, certain persons with disabilities, including vision impairments, may do better with spoken language, while others may find that visual cues and gestures work better. Combining the two methods will make the AI system more versatile and suitable for a wider range of users.
6) Additional Details By integrating information from both auditory and visual sources, conversational AI systems may generate responses that are both comprehensive and informative. For example, with a shopping assistant app, users can provide a more comprehensive description of the things they're pursuing while simultaneously showing an image of a similar item. This allows the system to offer more informed recommendations.
7) Versatility for a Range of Use Cases Conversational AI systems may adapt to different use cases and contexts by integrating speech and pictures. No matter what kind of user interface you're looking for—visual, auditory, or a combination of the two—this system is built to meet your needs and provide a seamless experience.

D. Examples of Speech and Image Integration in Conversational AI

These examples demonstrate how several domains might benefit from conversational AI systems that combine visual and aural modalities to enhance user experiences and enable more effective interactions.
These systems can enhance user engagement and satisfaction by responding to their particular needs and situations using a combination of audio and visual cues. The integration of visuals and speech in conversational AI opens up a world of possibilities for creating more intuitive and effective user experiences.

Several methods exist for conversational AI systems that combine visual and auditory cues:-

1) A Retail Assistant Enabled by AI

- ✓ Users can interact with virtual assistants while shopping online by speaking to them and submitting images of products.
- ✓ The algorithm takes into account both the user's spoken requests and the supplied photographs to understand their interests and make suitable product recommendations.
- ✓ Someone can say something along the lines of, "I need a blue dress for a summer wedding," and then attach a picture of the outfit they're lusting over.
- ✓ By integrating the user's spoken request with visual signals from the provided image, the system provides individualized recommendations that take into account their color and style preferences.

2) A Chatbot for Arranging a Vacation

- ✓ Itineraries and hotel reservations can be handled by a chatbot that assists with trip planning.
- ✓ In addition to sharing photographs of places they admire, users may also voice-describe their ideal vacation locale.
- ✓ Posts featuring tropical scenery might be accompanied by the following statement: "I want to go somewhere with beautiful beaches and lush greenery."
- ✓ By combining the user's verbal description with visual cues from the photos, the chatbot can better tailor the holiday experience by suggesting possible places, accommodations, and activities.

3) Help Desk for Medical Diagnosis

- ✓ When it comes to medical imaging interpretation and diagnosis, healthcare practitioners could use a diagnostic assistance system.
- ✓ On top of that, doctors can even express patients' issues verbally, and medical images like X-rays or MRIs can be uploaded.
- ✓ It is possible for a doctor to send a picture of an abnormal X-ray while talking to a patient over the phone about their symptoms.
- ✓ This technology helps doctors make quick and accurate diagnoses and treatment plans by combining the patient's spoken description with the visual data from medical images.

4) Learning Support System

- ✓ Students can study and practice concepts through an educational tutoring platform's interactive sessions.
- ✓ Students can not only explain their questions or issues verbally, but they can also include relevant textbook pages or problem statements as screenshots.
- ✓ If a student writes something like, "I'm having trouble understanding this math problem," they might also include a textbook screenshot of the task.
- ✓ The platform's visual and verbal explanations of its answers, recommendations, and step-by-step guidance are designed to aid users in learning and comprehension.

5) Interactive Chatbot for Enhanced Visual Customer Support

- ✓ A customer service chatbot is available to help users with technical issues.
- ✓ When reporting an issue, users have the option to include visual aids such as screenshots of error messages or device setups in addition to spoken descriptions.
- ✓ Consider the following scenario: a user submits a bug report along with a screenshot displaying the error message.
- ✓ The chatbot understands the problem description and provides suitable troubleshooting steps and solutions based on the submitted image by integrating speech recognition with computer vision skills.

II. THIS PAPER'S OBJECTIVES

The goal of this article is to look into conversational AI systems that use image and audio processing to enhance the user experience. Our focus will be on:

Learn about the newest approaches and tools for conversational AI image and audio processing.

Look at the pros and cons of multimodal interaction, including personalization, privacy, and continuous improvement.

Display real-world examples and case studies that prove multimodal conversational AI works.

Make recommendations for where this rapidly developing field could go from here in terms of both research and opportunities for innovation.

III. BACKGROUND

Conversational AI systems have revolutionized how people engage with technology. They enable natural language communication for a range of tasks, including task automation and information retrieval. By and large, these systems rely on text-based communication, analyzing user inputs in textual form and generating responses accordingly. This approach has worked in some cases, but it can't compare to the nuance and complexity of multimodal communication. There is hope for a solution to this problem in conversational AI systems that incorporate image and audio processing. These systems can better understand user intent, provide more contextually relevant replies, and make the user experience more immersive by integrating visual and auditory modalities. Several technical hurdles necessitate an in-depth understanding of both speech and image processing modalities and their processing approaches in order to successfully combine the two. The integration of speech and image processing represents a significant leap forward in conversational AI, offering the potential to elevate user experiences through the creation of more engaging, natural, and efficient interfaces. With an emphasis on the substantial impact of linking various modes in human-computer interaction, this study will explore the technical foundation, challenges, opportunities, and practical applications of multimodal conversational AI.

A. Processing Speech from Conversational AI

Speech processing involves using Automatic Speech Recognition (ASR) to transcribe spoken words into text and then using Natural Language Understanding (NLU) to decipher that text for meaning. Systems for Automatic Speech Recognition (ASR) make use of state-of-the-art Deep Learning techniques including Transformer models and Recurrent Neural Networks (RNNs), in addition to Hidden Markov Models (HMMs). In order to deduce the user's intent and context from their spoken inputs, Natural Language Understanding (NLU) techniques use parsing, semantic analysis, and entity recognition.

B. Visual AI for Conversations

In order to identify things, scenes, and patterns, as well as extract useful information from visual data, image processing is used. In order to extract useful information from images, computer vision methods including object detection, image categorization, and semantic segmentation are used. One form of Deep Learning, Convolutional Neural Networks (CNNs), have demonstrated remarkable performance in picture processing, enabling accurate and efficient interpretation of visual data.

C. Problems with Integrating Multiple Modes

Harmonizing modalities, maintaining context across modalities, and accommodating various user inputs are some of the challenges that arise when combining voice and picture processing abilities. When dealing with sensitive visual and auditory data, privacy and security concerns arise, necessitating robust encryption and anonymization techniques to safeguard user data.

D. Possibilities and Implementations

The integration of speech and image processing has a wealth of opportunities for enhancing conversational AI systems, despite the challenges it faces. Improved accessibility for users with different needs, improved comprehension of user intent, and personalized replies are all possible outcomes of multimodal interaction. applications are used for a variety of purposes, such as virtual assistants, chatbots for customer support, educational platforms, healthcare apps, and more.

IV. LITERATURE OF REVIEW

There is a wealth of information available in the existing literature on multimodal conversational AI, including methods for integrating speech and images, developing context-aware responses, exploring domain-specific applications, and dealing with ethical problems. Research in the area of multimodal conversational AI is important for the future of human-computer interaction in many domains. Existing work on multimodal conversational AI covers a wide range of topics, including methods for fusing several modalities, context-aware response generation, applications in numerous disciplines, and the integration of speech and visuals. Some notable examples of research are:-

Methods for integrating visual and audible inputs into conversational AI systems are a topic of continuing research. Scientists are trying to figure out how to combine information from picture processing and voice recognition to decipher what people mean. Anderson et al. (2018) proposed a multimodal navigation system in their article "Listen, Attend and Walk: Neural Mapping of Navigational Instructions to Action Sequences" that uses both visual and aural signals to generate navigational actions.

Multimodal fusion approaches, which combine data from many modalities, are the focus of most research. These approaches make use of techniques including attention processes, multimodal transformers, and graph-based fusion. According to Li et al.'s "MuSe-CAR Multimodal Sentiment Analysis with Context-Aware Regression" (2020), multimodal fusion is a technique for sentiment analysis. To combine text, audio, and visual data, this technique use context-aware regression.

Current research focuses on ways to generate responses that are both sensitive to context and capable of integrating data from several sources. In light of what is being said, these approaches aim to generate rational and relevant responses. To improve the accuracy of speech recognition, "Contextual Speech Recognition Using Multimodal Fusion of Audio and Video" (2019) by Zhou et al. proposes a system that takes into account the surrounding environment while making decisions about speech recognition.

Multimodal conversational AI is being researched for its possible applications in various fields, including healthcare, education, customer service, and entertainment. Using multimodal inputs, these apps may provide personalized and interesting experiences. Based on the work of Gaur et al. (2020), "SmartChat: A Conversational Agent for Patient Care and Health Education" serves as an example of a healthcare conversational agent. This agent provides patients with personalized health education and treatment by utilizing voice and visual inputs. Responsible and ethical AI activities are the subject of research as it pertains to developing multimodal conversational AI. Research looks at accountability, transparency, privacy, and fairness to make sure AI systems are developed and used responsibly. For instance, in order to improve the equality and fairness of AI systems, "Towards Fairness in Multimodal Classification: A Study on Bias Detection and Removal" (Chowdhury et al., 2021) explores ways for finding and removing prejudice in multimodal classification tasks.

V. ANALYSIS

To better understand the present and future of this dynamic technology, your team should do an in-depth study of multimodal conversational AI's technical difficulties, possibilities, uses, and future paths. Thinking about the technological hurdles, prospects, applications, and future directions of multimodal conversational AI is crucial for a comprehensive understanding. For your consideration, here is a well-organized analysis:

A. Difficulties Related to Technology

1) The Difficulty of Integration Technical difficulties arise when attempting to integrate voice and picture processing modules into a single system because of the disparities in data encoding and processing approaches.
2) Timely matching Maintaining context and coherence during talks requires synchronization of speech and picture inputs, which is no easy feat.

3) Understanding in a Context It is still difficult to create algorithms that can comprehend context across modalities and adjust responses appropriately.
4) Protecting Personal Information Ensuring privacy and security while handling sensitive data from many modalities is vital yet complex.

B. Opportunities

1) Better Experience for Users The potential for multimodal conversational AI to generate interactions that are more intuitive and natural has the potential to increase user happiness.
2) Personalization Utilizing several modalities allows AI systems to deliver solutions that are more customized to each user's preferences and the specific circumstances in which they are encountered.
3) Expanded Scope of Use Opportunities for healthcare, education, customer service, and entertainment-related multimodal conversational AI applications are expanding rapidly.
4) Diversity and inclusion Multimodal conversational AI can make AI more accessible and inclusive by supporting different modalities.

C. Use case

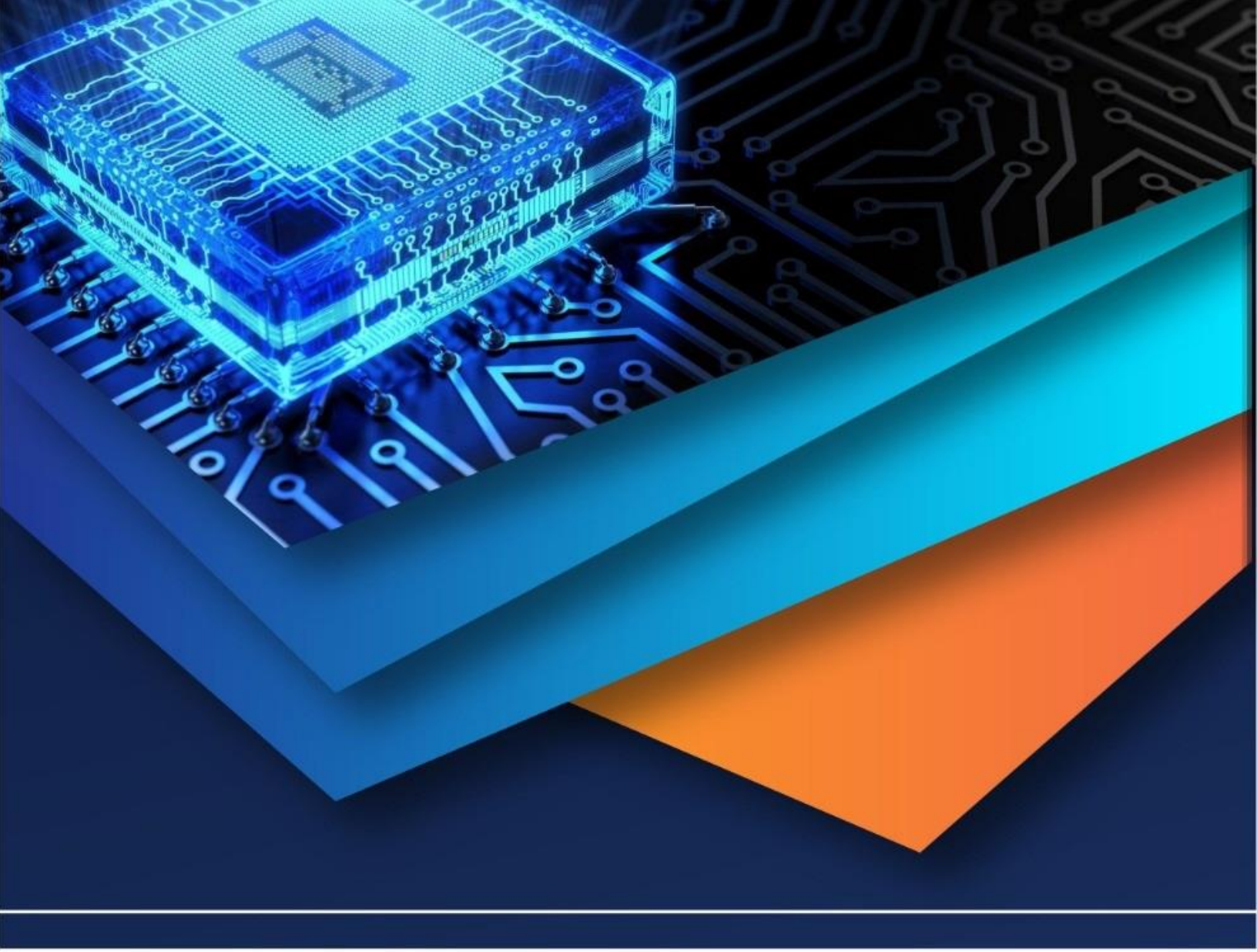
1) Help Desk Software The application of multimodal conversational AI has the potential to enable virtual assistants that can comprehend and react to user inquiries through visuals, text, and voice.
2) Support for Customers Chatbots that can understand and respond to consumer issues in a more holistic way are able to deliver better customer support.
3) Online Resources for Learning By combining voice and visual inputs to deliver individualized instruction and feedback, multimodal conversational AI can enhance interactive learning environments.
4) Use Cases in Healthcare By evaluating medical pictures, interpreting patient data, and offering clinical decision assistance, multimodal conversational AI systems can help healthcare workers.
5) Cutting-Edge Fusion Methods Improving fusion methods so they more efficiently combine data from different modalities can be a focus of future studies.
6) Generating Responses Based on Context To make context-aware response generation algorithms better at using multimodal inputs to provide more meaningful and consistent responses, additional research is required.
7) Fairness, transparency, privacy, and responsibility should be at the forefront of future advancements in multimodal conversational AI.
8) Cooperation with New Technology Enhancing the capabilities and uses of multimodal conversational AI can be achieved by integration with new technologies like AR and VR.

VI. CONCLUSION

A more organic, intuitive, and immersive user experience is offered by multimodal conversational AI, which represents a revolutionary approach to human-computer interaction. These systems can comprehend and react to user inputs in a variety of ways thanks to their integration of voice and picture processing capabilities; this improves context awareness, allows for more personalized interactions, and opens up new domains of use. We have covered the technological aspects, potential uses, obstacles, and future developments of multimodal conversational AI in this analysis. We have highlighted possibilities for a more personalized, inclusive, and extensive user experience as well as opportunities to address critical issues including data protection, contextual understanding, integration difficulty, and synchronization. Multimodal conversational AI has several potential uses in many fields, such as virtual assistants, chatbots for customer service, educational platforms, healthcare, and even the entertainment industry. By facilitating better communication, tailored support, and effortless incorporation into everyday life, these technologies may completely transform the way we engage with technology. Improving context-aware response generation, integrating with upcoming technologies, prioritizing ethical considerations, and expanding fusion techniques will be the focus of future research and development in multimodal conversational AI. Multimodal conversational AI may rise to the occasion by taking note of these potential and threats; doing so would open up fresh avenues for innovation and improve human interaction with AI. A new frontier in AI, multimodal conversational AI has the ability to revolutionize user empowerment in many sectors and reshape human-computer interaction. This technology will be essential in determining how humans and machines work together in the future as it develops further.

REFERENCES

[1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 2015, pp. 1– 14. doi: 10.48550/arXiv.1409.1556.
[2] Z. Yu, J. Yu, J. Fan, and D. Tao, "Multi-modal Factorized Bilinear Pooling with Co-Attention Learning for Visual Question Answering Key Laboratory of Complex Systems Modeling and Simulation ," in International Conference on Computer Vision, 2017, pp. 1821–1830. doi: 10.1109/ICCV.2017.202.
[3] H. Pham, T. Manzini, P. P. Liang, and B. Poczós, "Seq2Seq2Sentiment: Multimodal Sequence to Sequence Models for Sentiment Analysis," in Proceedings of the First Grand Challenge and Workshop on Human Multimodal Language (Challenge-HML), 2019, pp. 53–63. doi: 10.18653/v1/w18-3308.
[4] M. Firdaus, N. Thakur, and A. Ekbal, "MultiDM-GCN: Aspect-guided Response Generation in Multi-domain Multi-modal Dialogue System using Graph Convolutional Network," in Findings of the Association for Computational Linguistics: EMNLP 2020, 2020, pp. 2318–2328. doi: 10.18653/v1/2020.findings-emnlp.210.
[5] R. Bhushan et al., "Odo: Design of multimodal chatbot for an experiential media system," Multimodal Technol. Interact., vol. 4, no. 4, pp. 1–16, 2020, doi: 10.3390/mti4040068.
[6] S. Moon et al., "Situated and Interactive Multimodal Conversations," pp. 1103–1121, 2021, doi: 10.18653/v1/2020.coling-main.96.
[7] Q. Sun et al., "Multimodal Dialogue Response Generation," in Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 2022, vol. 1, pp. 2854–2866. doi: 10.18653/v1/2022.acl-long.204.
[8] D. Ramachandram and G. W. Taylor, "Deep multimodal learning: A survey on recent advances and trends," IEEE Signal Process. Mag., vol. 34, no. 6, pp. 96–108, 2017, doi: 10.1109/MSP.2017.2738401.
[9] K. Bayoudh, R. Knani, F. Hamdaoui, and A. Mtibaa, "A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets," Vis. Comput., vol. 38, no. 8, pp. 2939–2970, 2022, doi: 10.1007/s00371-021-02166-7.
[10] S. Ali, S. Tanweer, S. Khalid, and N. Rao, "Mel Frequency Cepstral Coefficient: A Review," in Proceedings of the 2nd International Conference on ICT for Digital, Smart, and Sustainable Development, ICIDSSD 2020, 2021, pp. 1–10. doi: 10.4108/eai.27-2-2020.2303173.
[11] A. Ahmed et al., "A review of mobile chatbot apps for anxiety and depression and their self- care features," Comput. Methods Programs Biomed. Updat., vol. 1, no. March, p. 100012, 2021, doi: 10.1016/j.cmpbup.2021.100012.
[12] I. Papaioannou and O. Lemon, "Combining chat and task-based multimodal dialogue for more engaging HRI: A scalable method using reinforcement learning," ACM/IEEE Int. Conf. Human-Robot Interact., pp. 365–366, 2017, doi: 10.1145/3029798.3034820.
[13] L. Liao, Y. Ma, X. He, R. Hong, and T. S. Chua, "Knowledge-aware multimodal dialogue systems," MM 2018 - Proc. 2018 ACM Multimed. Conf., pp. 801–809, 2018, doi: 10.1145/3240508.3240605.
[14] A. Saha, M. M. Khapra, and K. Sankaranarayanan, "Towards building large scale multimodal domain-aware conversation systems," 32nd AAAI Conf. Artif. Intell. AAAI 2018, pp. 696–704, 2018.
[15] C. Cui, M. Huang, W. Wang, X. S. Xu, X. Song, and L. Nie, "User attention-guided multimodal dialog systems," SIGIR 2019 - Proc. 42nd Int. ACM SIGIR Conf. Res. Dev. Inf. Retr., pp. 445–454, 2019, doi: 10.1145/3331184.3331226.
[16] L. Tavabi, K. Stefanov, S. N. Gilani, D. Traum, and M. Soleymani, "Multimodal learning for identifying opportunities for empathetic responses," ICMI 2019 - Proc. 2019 Int. Conf. Multimodal Interact., pp. 95–104, 2019, doi:ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195 10.1145/3340555.3353750.
[17] L. Nie, W. Wang, R. Hong, M. Wang, and Q. Tian, "Multimodal dialog system: Generating responses via adaptive decoders," MM 2019 - Proc. 27th ACM Int. Conf. Multimed., pp. 1098–1106, 2019, doi: 10.1145/3343031.3350923.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY