EMPLOYING MACHINE LEARNING FOR ENHANCED IMAGE RECOGNITION IN A FOOD-RECOMMENDATION CHATBOT

S. Amirtharaj[1] and S. Yuvasri[2]

[1] Associate Professor, [2]PG Student,
Department of Computer Applications,Mepco Schlenk Engineering College, Sivakasi, India

Abstract

In recent times, conversational AI has crucially transformed to user interaction in various domains, including food technology and services. The research paper covers the development of an intelligent food-ordering chatbot system that integrates image recognition with natural language processing (NLP) to improve the user experience. The proposed chatbot uses Dialogflow for intent detection and conversational flow management, Google Cloud Vision API for pre-trained image recognition, and a custom-trained Convolutional Neural Network (CNN) model for more accurate classification of local and various food items.The chatbot using natural language or by uploading food images which can make users can interact easily. When an image is uploaded, the system uses Google Vision API or a custom ML model to identify the food item based on visual features such as shape, color, and texture. Then, the system processes this information to generate corresponding responses, such as identifying the dish, suggesting similar items, or proceeding with order placement. A Rich Content response, including interactive cards and buttons, improves user experience and focuses the ordering process.

Keywords: *Chatbot, Image Recognition, Machine Learning, Dialogflow.*

1. Introduction

The improvements in artificial intelligence (AI) and machine learning (ML) have been developed how humans interact with technology, especially in the field of conversational systems. The chatbots intelligent agent is one of the most developed technologies in the field that capable of representing human-like conversations. They are increasingly being deployed across industries such as healthcare, education, e-commerce, and food services to automate responses, help customers and improves tasks.

In the terms of food industry, there is a growing demand for intelligent, interactive systems. In addition to understanding user queries it also interprets visual inputs such as food

images. This paper presents a food-oriented chatbot that combines natural language processing (NLP), image recognition, and response management to provide a continuous food discovery and ordering experience.

The important factors of this system are image recognition, which enables the chatbot to analyze a user-uploaded food image and finding the dish. For the above factors, the system using two approaches:

Google Cloud Vision API: A pre-trained image analysis tool it is capable of detecting multiple food-related labels and objects from an image with high accuracy.

Custom Machine Learning Model: A convolutional neural network (CNN)-based model trained on food datasets to recognize a wide range of local and diverse dishes. This allows for greater customization, improved recognition of local cuisine, and scalability.

The major component is Dialogflow, a cloud-based conversational platform developed by Google. Dialogflow enables the bot to understand and respond to user intents using NLP. It handles conversation flow, intent classification, entity extraction, and fulfillment through webhook integration. When a user engages with the chatbot - by giving queries or uploading images, Dialogflow interprets the user's intent and routes the request to a backend service that processes the image and returns a contextual response.

By integrating Dialogflow's conversational intelligence with image recognition models, it creates a multimodal chatbot that can be used toidentify food items based on images, understand natural language commands, suggest relevant dishes and place orders interactively.

2. Literature Survey

A Human-Chatbot Interaction platform is proposed in [1]. It uses data from a data set acquired from Kaggle, called Pima Indias, extracted from natives of Arizona in USA. Decision tree is used as the basis for the design, but from a different point of view, a web page will be designed that will have a chatbot, so that it will be the intermediary between the patient and a virtual doctor who will consult based on the data given by the user. The Chatbot have the ability to answer the questions of some fields and particular questions established in accordance with it, consequently served to collect the information necessary for the prediction.

A system that depicts technology is no longer just used to mimic human speech but also being used for more purposes, like, provides answers to queries in educational settings or for business purposes is presented in [2]. The advancements in artificial intelligence, machine

learning, and natural language processing (NLP) have led to the creation of chatbots by researchers and developers using a range of design approaches.

Another research focuses on creating an English conversation chatbot using speech recognition and Dialogflow as the artificial intelligence engine [3]. Expert reviews evaluated the chatbot's performance against specific indicators, and user feedback indicated high response accuracy. The availability of this chatbot is expected to enhance students' conversation skills.

Task-oriented chatbots serve specific purposes, such as automating tasks like restaurant reservations. The study aimed to understand users' perceptions and behaviours when interacting with chatbots for cafe takeout orders [4].

The review in [5] discusses how the chatbots are not limited by geographic boundaries and are available around-the-clock, in contrast to human personnel. They are an essential tool for everyday client engagement.

A learning-cum-assisted tool [6] focuses on a newly emerging tool for learning from chatbot. This paper reviews the technique, terminology, and different platforms used to design and develop the chatbot. It also presents some actual practical life typical applications and examples of chatbot.

A chatbot framework for object detection [7] is also found in literature. This proposed framework makes use of image processing and deep learning techniques.

The proposed system developed in [8] is an Informational cum Guidance Chatbot which tackles issues like what are the trending courses, finding new courses or courses of one's interest, providing guidance, finding colleges, etc. The Chatbot interacts with the user by asking basic questions, processes the query using Machine Learning Algorithms along with Natural Language Processing, and accordingly, serve the response or suggest the possible options.

A work on the mental model in Human Computer Interaction is presented in [9]. There are various approaches reviewed in this paper. It discusses the significance of the Chatbot approach, results and the trends in the human computer interaction and how it offers advantages over the earlier approaches.

The next-generation human-computer interaction (HCI) designs presented in [10] need to include the essence of emotional intelligence - the ability to recognize a user's affective states-in order to become more human-like, more effective, and more efficient.

The chatbot can respond dynamically and in a human-like way using contemporary algorithms such as Convolutional Neural Networks (CNNs) [11] by combining image recognition and natural language processing strategies.

Chatbots that can interpret the user questions and provide the right answers, in a fast and correct way is presented in [12]. Realization of a prototype of a Chatbot in educational domain is developed. This system provides support to university students for learning some courses.

Conversational image recognition chatbots proposed in [13] are dramatically changing how we use technology, making interactions feel both modern and intuitive. These systems integrate natural language processing (NLP) with visual recognition, enabling us to have real-time discussions about the content of images.This review explores the progress of these chatbots over time, focusing on how their core technologies—such as NLP and image recognition—combine to work effectively.

Building a conversational image recognition chatbot in python using CNN and LLM is proposed in [14]. It is a model where an AI agent can hold meaningful dialogue with humans in natural language about a visual content, specifically images.

The proposal in [15] exemplifies the effectiveness of image recognition technology when combined with AI interacting conversationally with the users or their images in practice ranging from the areas of customer support, education, and interactive media either with the out instance or with the high instance allow and include the others as part of the array.

The advanced conversational image recognition chatbot [16] is customized for the fashion industry. It is expected to revolutionize the way fashion related content is accessible for consumers online.

The chatbot framework proposed in [17] adopts a model which consists of natural language processing and image recognition technology. In this chatbot framework, neural encoder-decoder model is utilized with Late Fusion encoder 1 and 2 different decoders (generative and discriminate).

Training of the bot by exposure to both specific and generic questions would be crucial. Also, whenever a user isn’t satisfied with the answer given by the bot, he can raise an issue, and the answer for that question would be answered correctly in future updates. The bot also learns with each question the user asks, thereby making it better and better with each iteration [18]. The bot also has an image classifier feature for detecting objects.

3. Proposed Work

In recent times, there are several applications for conversational AI. The use cases are entail a user asking a specific question (intent) and the conversational experience or chatbot responding to the query by contacting a backend system, such as an API, database, or CRM. The core feature of the project is its image recognition capacity, which makes it simple for users to explore and find new cuisines, ingredients, and recipes from the chatbot. This system has a user interface in the form of web technology, through which users can interact with the chatbot and upload images for analysis. The chatbot engine will handle user interactions, process natural language queries, and provide responses. It will be responsible for initiating the image recognition process when prompted by the user. Image Recognition module helps to utilize computer vision and machine learning algorithms to analyse images uploaded by users. It may include pre-trained models for object detection, image classification, and text recognition. This proposed system provides a framework for building a robust and scalable system that can deliver accurate and valuable insights based on visual inputs from users.

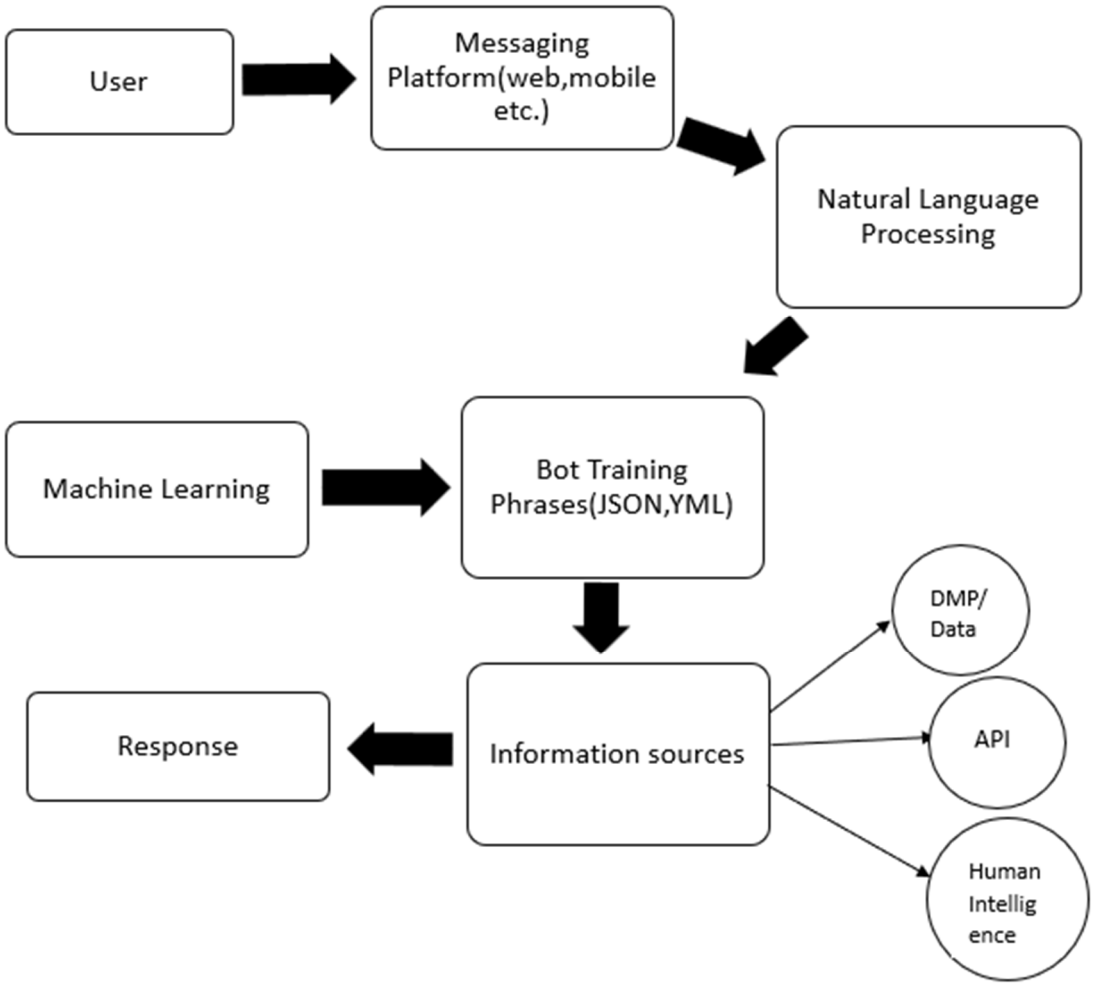

Fig.1 Chatbot WorkFlow

The workflow of the proposed Chatbot is illustrated in figure 1. As per the suggested framework, the user supplies the chatbot with input. The type of input might vary depending on the chatbot's capabilities, text and image-based. The chatbot parses user input using techniques for natural language comprehension and retrieves information. In order to understand the user's request or expression, tasks such as entity extraction, sentiment analysis, and intent recognition are necessary. The chatbot determines the purpose or intent of the user behind the input. For example, the user may be making a request, seeking information, or requiring assistance with a particular task. The chatbot remembers the context from previous conversations to help it understand the current topic more thoroughly. The chatbot can ensure that its responses are suitable and logical by controlling the context given in the current conversation.

In response to the user's input, the chatbot produces an output. Depending on the interactionthe response may be audio, text, or multimedia material. If the response is produced in text format, the chatbot might employ natural language generation methods to produce responses that are both appropriate for the context and human-like.The user receives the response from the chatbot via the messaging app or other application. After reviewing the response, the user can decide whether or not to carry on the conversation. In order to fulfil user requests, a chatbot need to communicate with external systems or APIs. This communication may require getting data from other sources, and processing information.

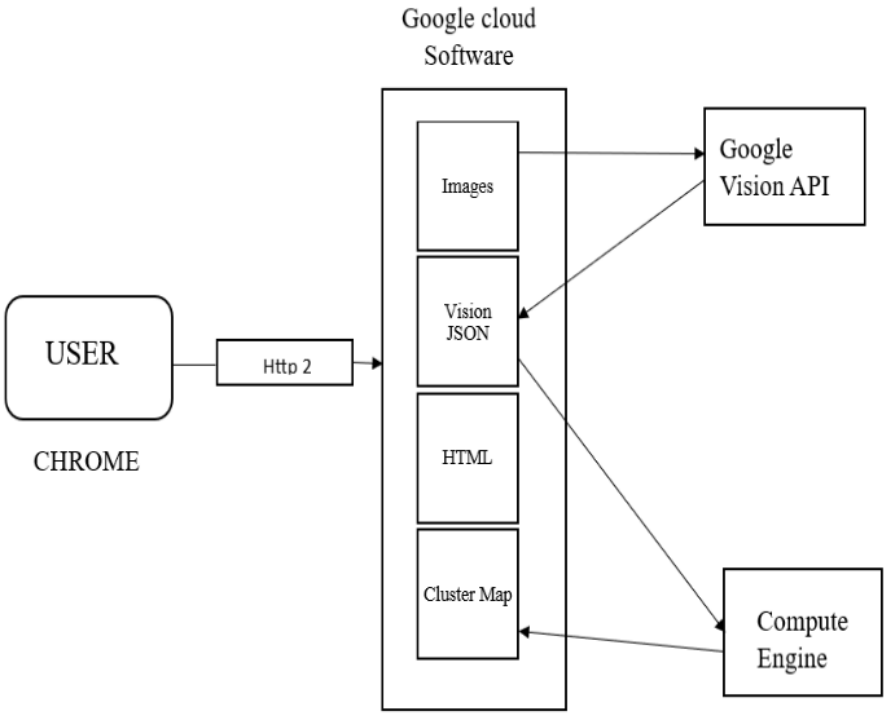

Fig. 2 Vision API Integration

Integrating Google Cloud Vision API with a chatbot is presented in figure 2. This integration allows the chatbot to analyze and understand images shared by users. This integration enables the chatbot to extract valuable information from images, such as objects, text, and faces, and incorporate this information into the conversation. First, create a Google Cloud Platform (GCP) project and enable the Cloud Vision API. Obtain the necessary API credentials (API key or service account key) to authenticate requests to the API. Depending on the chatbot platform you're using (e.g., Dialogflow, Microsoft Bot Framework, or custom-built), integrate the Google Cloud Vision API into your chatbot's backend code. Most chatbot platforms support HTTP requests, to integrate external APIs like Cloud Vision. Modify the chatbot's message processing logic to handle incoming images from users. When a user shares an image, extract the image data from the message and send it to the Cloud Vision API for analysis. Construct an API request to send the image data to the Cloud Vision API endpoint. Depending on the requirements, specify parameters such as the type of analysis (e.g., label detection, text detection, face detection) and any additional features you want to enable. Based on the analysis results from the Cloud Vision API, generate an response to the user. This could involve providing information related to the recognized objects, text, or faces in the image, asking follow-up questions, or executing specific actions based on the image content. Provide feedback to users if the chatbot encounters issues processing their images and offer alternative actions or assistance.

4. Results and Discussion

Rich content in chatbots is the integration of various multimedia features, such as images, videos, carousels, buttons, and interactive elements, to enhance user engagement and provide dynamic interactions. Add pictures to provide information, alternatives, or items with a visual representation, In this chatbot that shows pictures of food, ingredients, and cooking methods. To provide several options or products in a scrollable fashion, utilize carousels. Presenting product catalogs, restaurant menus, or recipes can all benefit from this. Provide links to other articles, websites, or resources so that readers can learn more or continue their research. It provides access to outside resources, such as many food-based websites, which can enhance the chatbot experience. A sample screenshot of the rich content presented to the user is presented in figure 3. A sample screenshot of the conversation responses offered by the system to the user is presented in figure 4.

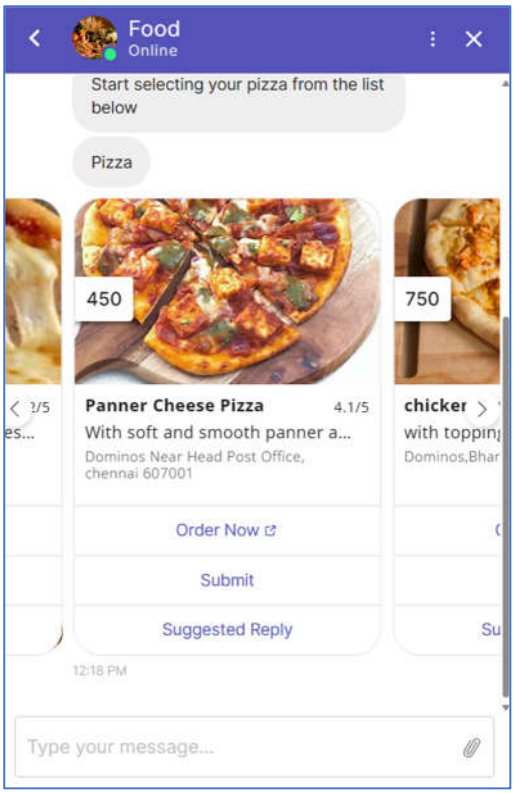

Fig. 3 Rich Content

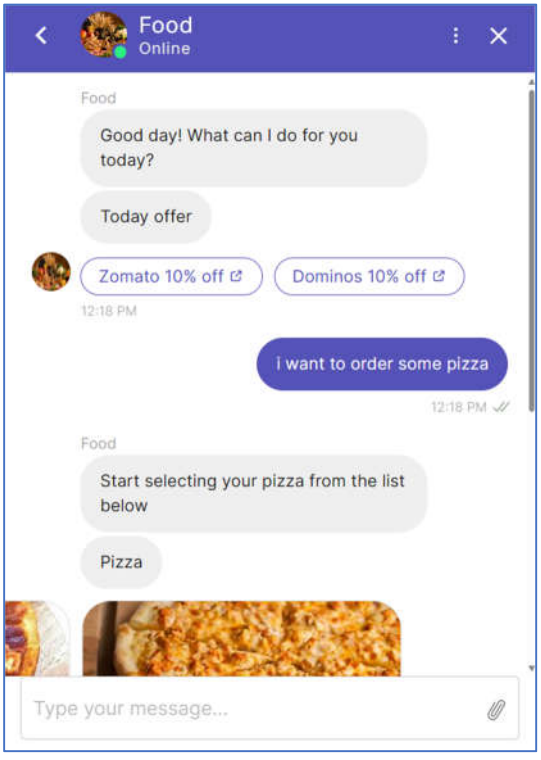

Fig. 4 Conversation Responses

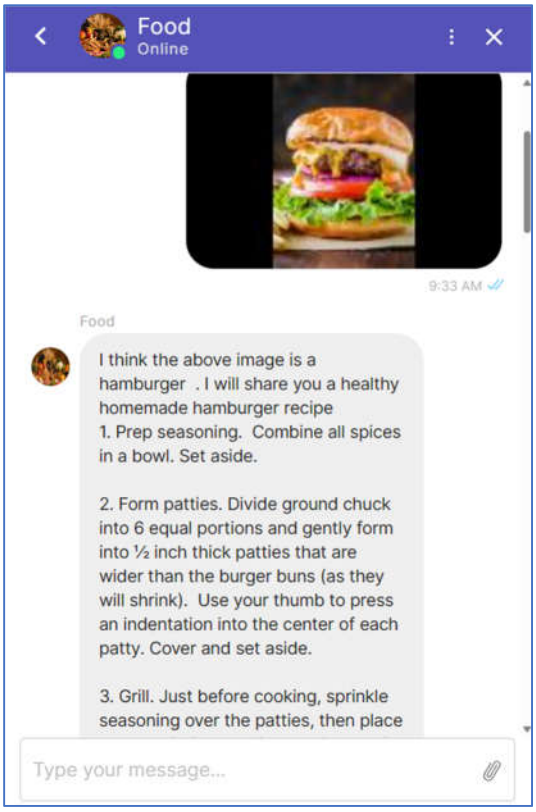

Fig. 5 Image Recognition and Responses

description: "burger", score: 0.9247923493385315, confidence: 0.0, topicality: 0.9247923493385315, bounds: 0, locations: 0, properties: {}
description: "pizza", score: 0.9149415493011475, confidence: 0.0, topicality: 0.9149415493011475, bounds: 0, locations: 0, properties: {}
description: "food", score: 0.899940550327301, confidence: 0.0, topicality: 0.899940550327301, bounds: 0, locations: 0, properties: {}
description: "veg salad", score: 0.8769422769546509, confidence: 0.0, topicality: 0.8769422769546509, bounds: 0, locations: 0, properties: {}

Fig. 6 Output of Vision API

Image recognition and response processes use advanced algorithms and techniques to enable chatbots to understand and respond to visual inputs from users in a meaningful way. Figure 5 presents a sample image recognition and response of the system according to our input image. The internal workflow of this chatbot is it receives an image input from the user. Before the image can be analyzed, preprocessing steps is applied, such as resizing, normalization, or noise reduction, to enhance the quality of the image and make it suitable for

analysis. Based on the classification and recognition results, the chatbot generates an appropriate response, as shown in figure 6. It's response is a description of what is illustrated in the image. A sample output of Intent Detection is presented in figure 7.


```
{
 "queryText": "Book a table at Monday 6PM",
 "intent": "BookOrderIntent",
 "parameters": {
  "Time": "6PM"
  "Day" : "Monday"
 },
 "fulfillmentText": "Booked Table in the restaurant"
}
```


Fig. 7 Intent Detection Output

Conclusion

This paper presents a food-oriented chatbot system that combines natural language processing, image recognition, and machine learning to deliver a more easy and interactive food ordering experience. With help of Dialogflow for conversational management and integrationof Google Cloud Vision API and a custom-trained ML model for image classification, the chatbot offers users to interact via text or food images. The system identifies food items, engages users in conversation, and simplifies order placement through rich content responses.

To provide further improvements the system's performance, scalability, and user experience, the following future enhancements may be included. (1) Multilingual Support: Expand the chatbot's capabilities to support multiple languages for global usability. (2) Nutritional Analysis: Integrate APIs or models that can estimate calorie content and nutritional value based on recognized food items for maintaining dietary food habits (3) User Feedback Loop: Collect feedback on food recognition accuracy to retrain and improve the custom ML model over time. (4) Payment Integration: Embed secure payment gateways to enable end-to-end order completion directly from the chatbot.

References

[1] Kohli, Bhaumik, et al. “A Platform for Human-Chatbot Interaction Using Python.” 2018 Second International Conference on Green Computing and Internetof Things (ICGCIoT). IEEE, 2018.

[2] Mohamad Suhaili, Naomie Salim, Mohamad Nazim Jambli “Service chatbots: A systematic review”,Expert Systems with Applications,Vol. 184, Dec. 2021.

[3] A. F. Muhammad, D. Susanto, A. Alimudin, F. Adila, M. H. Assidiqi and S. Nabhan, "Developing English Conversation Chatbot Using Dialogflow," *2020 International Electronics Symposium (IES)*, Surabaya, Indonesia, 2020, pp. 468-475, doi: 10.1109/IES50839.2020. 9231659.

[4] Rishma Garg; Riya Riya; Sahil Thakur; Nancy Tyagi; Kasunuru Nawaz Basha; Dinesh Vij; Gurjot Singh Sodhi, “NLP Based Chatbot for Multiple Restaurants” ,10th International Conference on System Modeling& Advancement in Research Trends (SMART),2021

[5] Amisha Gupta,HimanshuGupta,VaibhavRathore,SuyashAwasthi,HarshdeepSingh,“Impact of Chatbots on Customer Satisfaction in Food Delivery Apps”,International Journal for Scientific Research & Development, vol. 8, no. 7, 2020.

[6] Rohit Tamrakar,Niraj Wani “Design and Development of CHATBOT: A Review”,International Conference on Latest Trends in Civil, Mechanical and Electrical Engineering, 2021.

[7] A. Mahesh Babu, Malik Jawarneh, José Luis Arias Gonzáles, Meenakshi, Kishori Kasat, K.P. Yuvaraj, “Conversational Chatbot With Object Recognition Using Deep Learning and Machine Learning,” Conversational Artificial Intelligence, https://doi.org/10.1002/9781394200801.ch21

[8] Khan, Aysha, et al. “NEEV:An EducationInformational Chatbot,” International Research Journal of Engineering and Technology, vol. 6, no. 4, 2019, pp. 492–95.

[9] Bansal H. and KhanR., A review paper on human computer interaction. Int. J. Adv. Res. Comput. Sci. Software Eng., vol. 8, no. 53,2018.

[10] Maja Pantic, Leon J.M. Rothkrantz, “Towards an Affect – Sensitive Multimodal Human-Computer Interaction” in proceedings of the IEEE, September 2003, Vol. 91, No. 9 , pp. no.- 1370 - 1390.

[11] M. Buvana, P. Krishna Bharathi, H. Ammar, S. Balaji , "Conversational Image Recognition Chatbot" International Journal of Engineering Inventions,vol. 13, no. 11,pp. 133-138, Nov. 2024.

[12] Francesco Colace, Massimo De Santo, Marco Lombardi, Francesco Pascale, and Antonio Pietrosanto DIIn, "Chatbot for E-Learning: A Case of Study" International Journal of Mechanical Engineering and Robotics Research Vol. 7, No. 5, Sep. 2018.

[13] R.R. Kolte, Harsh Wanwe, Prajwal Sathawane, Sahil Kumbhare, Rohit Nagrikar, "Conversational Image Recogonition Chatbot," International Research Journal of Modernization in Engineering Technology and Science vol. 06, no. Oct. 2024.

[14] Devi Praba N , Kishor Kumar S , Vijay Surya R , Ananya R Melagiri , VijayaramanS, "Conversational Image Recognition Chatbot" International Journal of Advanced Trends in Engineering and Management, pp. 1-6, 2024.

[15] Sailesh R, Subiksha S, Yamini R, Dhinakaran,NaveenkumarK, "Convolutional and Image Recognition Chatbot" International Journal of Innovative Research in Technology,Vol. 11.No. 6, Nov. 2024.

[16] Harshavardhika K, Nejiya K S, Sujithra R, Reshma R, Mr. S. Janarthanan, "International Journal of Research Publication and Reviews", vol 6, Issue 3, pp 4511-4513 Mar. 2025.

[17] ShengyangSu, "Conversational and Image Recognition Chatbot" Stanford CS224N Natural Language Processing with Deep Learning, Stanford CS224N Custom Project.

[18] Pritham Sriram G, Prasana Venkatesh S, "Image classifying AI Chatbot," International Journal of Advanced Research, Ideas and Innovations in Technology, vol. 7, Issue 5 - V7I5-1378, 2021.