**InfoBot(A robot that recognizes objects/faces and extracts information about them)**

Chigiri Varsha  
*Computer Science and Engineering-Artificial Intelligence and Machine Learning*   
*MLR Institute of Technology*Hyderabad, India   
[varshapoornachandar@gmail.com](mailto:varshapoornachandar@gmail.com)

Vinjravath Arun Kumar  
*Computer Science and Engineering-Artificial Intelligence and Machine Learning*  
*MLR Institute of Technology*Hyderabad, India  
[arunpwr2712@gmail.com](mailto:arunpwr2712@gmail.com)

Matam Jahnavi  
*Computer Science and Engineering-Artificial Intelligence and Machine Learning*  
*MLR Institute of Technology*Hyderabad, India  
[matams1108@gmail.com](mailto:matams1108@gmail.com)

Surukunti Sai Krishna Reddy  
*Computer Science and Engineering-Artificial Intelligence and Machine Learning*  
*MLR Institute of Technology*Hyderabad, India  
[surukuntisaikrishnareddy@gmail.com](mailto:surukuntisaikrishnareddy@gmail.com)

**Abstract- This paper aims to develop “Infobot”, a friendly chatbot robot with advanced object and face recognition capabilities. The robots’ primary functions include recognizing specific objects and faces of certain users and extracting relevant information to enable personalized interactions. Additionally, Infobot is an “infobot” for known faces, providing tailored information and assistance. Moreover, it can identify unknown faces of visitors, gathering information about their visit and the person they intend to meet. Our project uses cutting-edge AI and computer vision technologies to create an intelligent, user-friendly robot for enhanced human-robot interactions. Enhancing human-robot interactions in diverse scenarios by developing an intelligent and adaptive chatbot with advanced object and face recognition capabilities, implementing cutting-edge AI and computer vision technologies to enable friendly chat interactions, personalized assistance, and efficient face and object recognition.**

**Keywords- AI, Chatbot, Computer Vision, Face Recognition, Human-Robot Interaction, Infobot, Object Recognition.**

I.Introduction

In the evolution of robotics and artificial intelligence (AI), the improvement of sensible and adaptive chatbots has emerged as a point for researchers and engineers. The combination of computer imagination and prescient and the slicing facet has produced an efficient manner for face and item reputation capabilities. One such project at the forefront of this technological frontier is “Infobot”. Infobot is a robot equipped with object and face recognition functionalities. The aim is to develop a personalized chatbot with human-robotic interactions. The users can perceive gadgets, and faces and extract relevant records about them via this robotic. This robot is beyond the traditional chatbot, as it goes one step ahead by identifying unknown faces of visitors and extracting their purpose of visit. This capability enables a robot to gather all the information regarding known and unknown persons and extract information about their visit individually. Infobot is a seamless and intelligent human-robot interaction experience using advanced facial recognition algorithms and real-time processing.

The base of this project lies in utilizing edge-cutting technologies, By using superior facial recognition algorithms and real-time processing, Infobot is a continuing and shrewd human-robotic interplay. using AI and computer vision, to develop a user-friendly and intelligent robot. To personalize assistance, enhance face and object recognition, and create friendly conversation will be modified to lead to altered scenarios by Infobot. We are going to explore the techniques in detail by inspecting their first-rate packages, technological components, and implications. In this, we're capable of contributing to the growing literature on artificial intelligence and human-robotic relationships with a watch on clever robots.

II. Literature Survey

[1] [Abhishek Sarda, Dr. Shubhra Dixit, Dr. Anupama Bhan] The YOLO (You Only Look Once) method is used by researchers who have studied item reputation for self-sustaining driving. This technique employs multistage algorithms which have been advanced for quicker results and additionally utilizes single-shot detector strategies to beautify accuracy. Through the use of bounding boxes, the YOLO set of rules is skilled in unique instructions to become aware of objects in photos accurately. Increasing annotations, formatting them in a YOLO manner, creating a specially crafted dataset, and training the algorithms are all part of the model. To lessen the number of bounding bins, the look shows the use of the open pix dataset and enforcing non-maximal suppression. After about 12 hours of education, the YOLO set of rules has been determined to have an impressive 74.6% accuracy in detecting gadgets, including automobiles.

[2] [Muhammad Zeeshan Khan, Shahid Mumtaz, Muhammad Imran] Introducing a current solution, a comprehensive and revolutionary method to stand reputation using CNN and area computing. The problem of managing large volumes of data produced by IoT devices is resolved by this upgraded and unified version. By harnessing the capability of area computing, it effectively reduces statistics latency and complements actual imagination and president, this model can understand people even in crowded frames. It is likewise resilient to special lighting conditions, as its performance is minimally impacted with the aid of mild sources. The proposed method acknowledges 30 of the 35 diagnosed faces with an accuracy of 97.7%. Their paintings’ barriers embody the want for a strong schooling dataset and the problems related to facial popularity in actual international instances. They propose paths of capability to improve the encouraged collection of policies, as well as new completeness and splendid architectures to discover.

[3] [Kanakamedal Deepika, VeerankiTilekya, Jatroth Mamatha] The paper proposes a “jollity chatbot” – a conversational AI system to provide companionship and emotional support to users, especially those feeling depressed or going through tough times. The chatbot is designed to converse with users and suggest uplifting content like positive quotes, motivational videos, etc. It uses Rasa, an open–source conversational AI framework, with custom intents, sample conversations, and responses defined to understand user needs and maintain context. The paper presents a conversational agent focused on providing companionship and emotional support, especially to depressed individuals. The experimental results demonstrate its ability to understand user intents and maintain context in conversations at a reasonably good accuracy.

[4] [Md. Mahmud Hasan, Lim Hooi Jiun, Ng Wei Chuen] This paper explains the design of a Smart Telephone system that includes Caller identification, Tele-security, and Answering Machine features. The system is based on a low-cost microcontroller and utilizes a photo coupler device to detect the ringing signal without overloading telephone company lines. The caller Identification system is supported by Signaling System No.7(SS7) and it uses Frequency Shift Key (FSK) modem tones to transmit information on the subscriber loop. The Answering Machine is designed to record incoming messages if the person is not available. The system is simple and cost-effective, making it a potential household device in the future.

[5] [Neelkumar P. Patel, Devangi R. Parikh, Darshan A. Patel, Ronak R. Patel] The paper presents “Unibot” – a web–based conversational agent or chatbot to provide college/university-related information to users. It has a messaging application-like interface for a friendly user experience. Users can ask questions in natural language and get relevant answers. The backend is built using PHP and MySQL database containing question–answer pairs on topics like fees, admission process, syllabus, etc. It uses an efficient algorithm to preprocess user input, extract keywords, and query the database to retrieve the most appropriate response. Results demonstrate that Unibot provides customized, variant responses based on user questions with minimal response time. The paper presents the design and implementation of Unibot – an intuitive, efficient, and interactive chatbot acting as an information assistant to provide university-related information to students/visitors in a conversation-style interface.

[6] [Tao Feng, Yao Yu, Y Bai] This paper presents a study on a human-tracking robot using ultra-wideband (UWB) technology. To monitor the target, the research suggests using UWB technology in conjunction with a virtual spring model that is integrated into the robot and a modified hyperbolic positioning algorithm. The paper also covers the difficulties that human-tracking robots face in identifying and determining the target person’s position and orientation, as well as how the improved hyperbolic positioning algorithm addresses the issue of measurement inaccuracies in the robot’s tracking system. The experimental results and more discussion of the technology’s possible uses in robotics and human-robot interaction are included in the paper’s conclusion.

[7] [Menaka R, Archana N] The PDF file presents an Enhanced Missing Object Detection System that utilizes the YOLO algorithm to identify missing objects in real time. The YOLO method is used by the PDF file’s enhanced missing object detection system to find missing objects instantly. An object detector as quick and precise as the YOLO algorithm can be interfaced directly on entire images and is simple to construct. Applications for the suggested method include facial recognition, character recognition, and surveillance systems. Object co-detection and GSM alerts for missing objects can be added to the system to make it better.

[8] [Muhammad Iqbal, Bhakti YudhoSuprapto, SuciDwijayanti] The study proposes a modified convolutional neural network architecture that achieves high accuracy in recognizing faces and emotions. The authors proposed a method that measures the distance between the position of the robot and the object's face, which enables the robot to identify the position of the human. The study uses data that is obtained from male and female students where some students wore spectacles and some female students wore hijab. The authors compared the performances of CNN architecture i.e., VGG16 and Alex Net, with this of modified architecture. The study does not mention any negative impact of the proposed system, but it is important to consider the potential ethical concerns related to the collection and storage of facial data, as well as the unintended consequences of using robots for human-robot interactions.

[9] [Prasnurzaki Anki, Alhadi Bustamam, Herley Shaori Al-Ash, DevviSarwinda] The paper provides a practical demonstration of how chatbots can be implemented in Python using deep learning models like BiSTM. This can help developers and researchers build real-world chatbot applications. By testing different model architectures and hyperparameters, the research compares the performance of BiLSTM versus other options for chatbots. The research achieves high accuracy scores, reaching over 99% with the optimal BiLSTM model. This shows the potential for accurate chatbots using this approach. The BiLSTM model architecture provides more interpretability compared to “black box” models like neural networks.

[10] [M. Swathi, K. Gopal Reddy] The paper proposed a developing chatbot using OpenAI’s GPT-3 language model that can generate human-like responses. GPT-3 is trained on a vast amount of data to understand patterns in language and nuances.It builds on earlier GPT models with more parameters and data. It can be used for various applications like customer service, education, and research beyond just chatbots. The future scope involves enhancing chatbots to facilitate work, handle conversations with multiple users, and interact via natural language based on machine learning. The paper proposes using the advanced AI model GPT-3 to create more capable and human-like chatbots while discussing its advantages, limitations, and applications. It provides an initial framework to build GPT-3 chatbots and suggestions for future development.

[11] [Reagan L. Galvez] The paper discusses the risk of manually finding unstable items in X-rays and gives an automated method for doing so, an object detector based totally on the YOLO model. The authors showed that training YOLO from scratch works better than transferring getting to know the short identity of risk objects via comparing the overall performance of two methods with the use of an IED X-ray dataset. Then try to discover the most reliable version that balances inference speed and achieves high accuracy with the usage of the YOLOv3 architecture. The paintings additionally cover schooling and assessment using the IED X-ray dataset and Darknet-53 architecture. The artwork additionally consists of evaluation and guidance concerning the use of Darknet-53 architecture and the IED X-ray dataset. The findings show that training from scratch outperformed transfer learning, with the help of achieving implied common accuracy (mAP) of 45.89% in a 416x416 photo, 51.48 in a 608x608 photo, and 52.40 in a multiscale picture. A thorough exam of the training method, outcomes, and common performance assessment is given within the paper, which emphasizes the promise of YOLO-based standard object detection for computerized risk detection in X-ray photos.

[12] [Shashi Kant Singh, Shubham Kumar, Pawan Singh Mehra] The paper provides an intense evaluation of ChatGPT and Google Bard AI, superior chatbots based totally on Artificial Intelligence. It cited the ancient historical past, training, functions, and barriers of these chatbots. The paper moreover includes an evaluation between ChatGPT and Google Bard AI. The assessment highlights the primary applications of ChatGPT, together with thinking-answering, revolutionary writing, and language translation. It additionally said the capability to utilize those chatbots for precise customers, medical docs, engineers, and business corporation professionals. It concludes with insights into the destiny scope of these chatbots, emphasizing the need to cope with limitations, enhance accuracy, and overcome biases for additional notable capability.

[13] [ Y. Ye, Y. Zhang, Y. Li, and Y. Liu] The paper presents the issue of trust in human-robot cooperation and suggests using the ChatGPT to enhance trust and communication between people and machines. The study’s methodology, which includes the Robo GPT system’s design, a human-subject experiment, and data analysis techniques, is described by the authors. The study’s findings show that ChatGPT improves human-robot collaboration through higher performance ratings, less cognitive load, and greater confidence in robot assistance.Overall, the proposal provides valuable insights into the potential of ChatGPT in improving human-robot collaboration and lays the foundation for future research in this field.

[14] [IlmanShazhaev, Arbi Tularov, Dmitry Mikhaylov, Islam Shazhaev, Abdulla Shafeeg] The paper noted the combination of a voice assistant with the GPT chatbot era, specializing in its packages within the gaming organization. The authors emphasize the growing characteristics of Artificial Intelligence (AI) in normal life, specially in the form of voice assistants and chatbots. They spotlight the development and capabilities of the ChatGPT version, that is designed to have interaction in meaningful conversations and perform several obligations together with code writing, trojan horse identification, and generating content material. The remark indicates a method to leverage the ability of ChatGPT technology with the aid of combining it with voice assistant algorithms. The authors suggest the integration of voice assistant capabilities with ChatGPT to beautify the gaming experience, offering help in gaming mechanics, and personalized services for gamers. The authors advocate leveraging advanced AI ideas to provide customized and complex conversational AI, specially in the gaming industry.

[15] [Mazin Hnewa, Hayder Radha] This paper proposes a novel Multi-Scale Domain Adaptive YOLO framework for area adaption in object detection, especially focusing on the YOLOv4 architecture. With several area classifiers and a Domain Adaptive Network (DAN) with multiscale feature inputs, the proposed approach objectives to beautify YOLOv4’s overall performance in cross-area object detection. The counseled technique outperformed the authentic YOLOv4 in an evaluation of large-scale trials the authors achieved the usage of well-known datasets. The dialogue of the counseled framework’s capability to automatically adapt YOLO to goal domains and outperform modern-day YOLOv4 in loads of checking out situations for self-reliant driving packages involves a close inside the have a look at.

[16] [Shaji Thorn Blue, M. Brindha] The study offers an approach to enhance the accuracy of boundary boxes by identifying objects and growing the diagnosed object’s precision within the YOLOv3 gadget. The precision of the boundary boxes improves through the use of pixel values and area detection. The summary of the work in object identification, including several algorithms is also included in the study. The proposed algorithm offers a promising approach to enhancing object detection precision within the YOLOv3 system, with potential applications in diverse fields such as unmanned aerial vehicles, pedestrian detection, and video surveillance. The study concludes by highlighting the algorithm’s ability to produce more accurate boundary boxes than YOLOv3 and suggests future research to address the identified limitations and achieve even greater accuracy.

[17] [Shraddha Mane, Supriya Mangale] The paper proposed an approach for moving object detection and tracking using CNN. It includes object detection which uses tensor flow object detection API which is robust in detecting objects in complex scenes and background conditions. Uses a CNN-based tracking algorithm instead of traditional computer vision techniques. Loads weights of a pre-trained model capable of incorporating temporal information. Able to track objects at 150 FPS and handle occlusion scenarios. The main innovation is in using deep learning models to improve generalization across challenging uncontrolled video settings. Both detection and tracking leverage CNN models for state-of-the-art performance.

[18] [Anis Koubaa] The paper presents ROSGPT, a concept that integrates large language models (LLMs) such as ChatGPT with the robot operating system (ROS) to advance human-robot interaction. The system uses prompt engineering in the conversion of unstructured human language commands into structured robotic instructions, enabling seamless communication between robots and humans. The implementation includes proof-of-concept for robot navigation and demonstrates the use of an ontology to guide ChatGPT in generating context-specific robotic commands. Overall, the paper shows a step towards artificial general intelligence and offers collaboration between natural language processing communities and robotics in developing human-robot interactions.

[19] [Xiao-Feng Zhao, Zi-Heng Chen, He-Feng Yin, Xiao-Jun Wu] The paper proposes a design of an intelligent visitor system based on edge collaborative computing and cloud computing. The system addresses the management requirements for visitors in enterprises and government departments, aiming to improve efficiency and accuracy while reducing labor costs. With the use of AI cloud and area computing, the system incorporates face recognition and speech reputation generation to allow abilities together with traveler verification, and voice self-registration. The gadget’s cloud-thing collaboration enhances computing ability, improves actual-time common normal overall performance, and guarantees excessive accuracy. The trying-out effects showed excessive recognition accuracy and short popularity delays. The limitations of traditional visitor manipulation strategies are overcome by intelligent tourists and also expose the cooperation used to enhance traveler manipulation strategies.

[20] [W.S. Mada Sanjaya, DyahAnggraeni, Kiki Zakaria, Atip Juwardi, MadinatulMunawwaroh] The social robot called SyPEHUL was developed, which uses image-processing techniques to detect human faces. The Arduino microcontroller and 12-degree-of-freedom motor servos are used to control the hardware configuration of the robotic head and its face. The OpenCV library and Python 2.7 are the foundation for the Cascade Classification and LBPH (Local Binary Pattern Histogram) Face Recognizer techniques used in the face identification and tracking process. With the help of this system, human faces may now be accurately tracked and recognized, facilitating human-robot interaction. The references include studies in human-robot interaction and facial recognition, as well as the thorough use of the Viola-Jonas detection approach.

1. COMPARISION TABLE

| **Author** | **Year** | **Approach** | **Description** |
| --- | --- | --- | --- |
| Abhishek Sarda,  Dr. Shubhra Dixit,  Dr. Anupama Bhan | 2021 | Training, Data Collection, Non-Max Suppression, Calculation, Augmentation, Techniques | Training the YOLO algorithm on custom classes using the open images dataset, implementing non-max suppression, and applying specific techniques for object detection in the context of autonomous driving |
| M. Z. Khan, Shahid Mumtaz, Muhammad Khurram Khan | 2019 | UsingMachinelearning, edgedetectiontechniques,and CNN, thefaces arerecognized | The suggested solution makes use of CNN, a cutting-deep learning technique in computer vision. Even with several faces in the shot, thesystem can still identify persons. Due to the system's lack of sensitivity to light, it can also recognize individuals in a variety of situations and lighting scenarios. |
| Md. Mahmud Hasan, Lim Hooi Jiun, Ng Wei Chuen | 1998 | tele-security system, hardware design, Signaling System, Caller ID system | The development and deployment of a smart telephone device equipped with Caller ID, an answering machine, and a tele-security system. It details the hardware design, including the utilization of a photo coupler for ring signal detection and a multifunctional IC for FSK signal reception and decoding.  It elucidates the transmission formats, operational aspects of the smart telephone device and indoor sensors, and the Caller ID system within digital telephone networks, supported by Signaling System No. 7 (SS 7). |
| T. Feng,  Y. Yao,  L. Wu,  Y. Bai,  Z. Xiao,  Z. Lu | 2018 | Ultrawideband (UWB) Technology, hyperbolic algorithm, virtual spring model in the robot to track target person. | The process for tracking the target person is based on Ultrawideband (UWB) technology, a modified hyperbolic positioning algorithm, and a virtual spring model implemented in the robot to track the person. The use of a moving-average filter to reduce the noise in the UWB data. The potential use of saliency methods for object identity agnostic models. |
| Menaka R  Archana N  Dhana gopal  Ramesh R | 2020 | YOLO algorithm, Intersection over Union (IoU), Non-max suppression, anchor boxes | It uses the convolutional layers to reduce the feature space and uses frame analysis and anchor boxes for object detection. The data augmentation and dropout are used to avoid overfitting. The YOLO algorithm is pre-trained using the related layer and then fine-tuned for object detection. |
| Muhammad Iqbal, Bhakti YudhoSuprapto, SuciDwijayanti | 2022 | Convolutional neural network (CNN), architectures like Alex Net VGG16, accuracy measurements | Real-time face emotion recognition in a humanoid robot is achieved by the use of convolutional neural network designs such as VGG16 and Alex Net, in this technique. The performance is assessed based on accuracy in real-time interactions, and the system is built to detect faces and emotions simultaneously. |
| Reagan L. Galvez, Elmer P. Dadios, Argel A. Bandala, Ryan Rhay P. Vicerra | 2019 | Transfer learning involves using pre-trained models and fine-tuning them on a specific dataset, training from scratch with pre-existing weights, mean average precision(mAP) used for specifications, IED X-ray dataset for X-rays images | The authors show that training YOLO from scratch works better than transfer learning in the rapid identification of threat objects by comparing the performance of the two methods using an IED X-ray dataset. The work also covers training and evaluation using the IED X-ray dataset and the Darknet-53 architecture. |
| Y. Ye, Y. Zhang, Y. Li, Y. Liu | 2023 | ChatGPT to enhance communication, designing RoboGPT, NLP, accuracy or performance scores and cognitive load measures | The method used in this involves designing the RoboGPT system, which utilized ChatGPT to enable natural language communication between humans and robots. A human-subject experiment was conducted to validate the effectiveness of solution. The performing a simple assembly task with the robot arm under conditions and was assessed using performance scores, self-evaluated performance, and cognitive load measures. |
| Mazin Hnewa, Hayder Radha | 2021 | development of a Multi Scale Domain Adaptive YOLO, Domain Adaptive Network (DAN) with multiscale feature inputs, performance of YOLOv4 for cross-domain object detection by producing robust domain-invariant features that reduce the impact of domain shift. | The method outlined in this study entails creating a Multi-scale Domain adaptive YOLO framework, which combines the YOLOv4 object detector within a domain adaptive network (DAN). When learning domain invariant features from the multiscale feature inputs, the DAN is connected to YOLOv4 architecture. To predict the domain class, the DAN combines many domain classifiers with neural layers. This method seeks to achieve domain invariance during domain adaption-based training and resealing features against domain shifts at various sizes. |
| V. Amala Rani and S. Lalitha KumariW.S. Mada Sanjaya, DyahAnggraeni, Kiki Zakaria, Atip Juwardi, MadinatulMunawwaroh | 2017 | Development of a Social Robot named SyPEHUL, Cascade Classification and LBPH (Local Binary Pattern Histogram), Real-time face detection and recognition, with the system being capable of mentioning the person's name upon identification | The creation of a social robot called SyPEHUL, combines cutting-edge image processing techniques to follow and identify human faces. For precise face tracking and recognition, the system makes use of LBPH and cascade classification techniques which are based on the OpenCV library and python2.7. The hardware configuration comprises 12 Degree of Freedom(DoF) motor servos to move the robot's head and face. It is based on an Arduino microcontroller. A database of photos of faces is used to train the system, and its ability to accurately track and recognize faces is evaluated. It is possible to do real-time face detection and recognition, and the system can mention an individual's name after successful identification. |
| Prasnurzaki Anki, Alhadi Bustamam, Herley Shaori Al-Ash, DevviSarwinda | 2020 | Chatbot, program, BiLSTM, accuracy, input, output | The research provides a practical demonstration of how chatbots can be implemented in Python using deep learning models like BiLSTM. This can help developers and researchers build real-world chatbot applications. The BiLSTM model architecture provides more interpretability compared to “black box” models like neural networks. The use of the Cornell Movie Dialog corpus provides a diverse, real-world dataset for training the models. This enhances applicability. |
| Kanakamedala Deepika, VeerankiTilekya, Jatroth Mamatha, Subetha T | 2020 | Chatbot, Contextual AI Assistant, Rasa, Rasa NLU, Rasa Core | The chatbot is designed to converse with users, suggest uplifting content like positive quotes, motivational videos, etc. based on the user’s mood, and act as a virtual friend. It uses Rasa, an open-source conversational AI framework, with custom intents, sample conversations, and responses defined to understand user needs and maintain context. |
| Neelkumar P. Patel, Devangi R. Parikh, Darshan A. Patel, Ronak R. Patel | 2019 | Artificial Intelligence, Chatbot, Human-like interactive, Machine Learning, University Chatbot | The overall chatbot system architecture and interface for interacting with users. The custom algorithm was developed for processing user queries and retrieving responses from the database. This algorithm involves steps like preprocessing, keyword matching, spell checking, option display, learning new Q&A pairs, etc. Expanding the chatbot’s knowledge base, deploying it on servers to handle multiple users, and maintaining performance is key for broader implementation. |
| M. Swathi, K.Gopal Reddy | 2023 | Chatbot, OPENAI, Artificial Intelligence, Agents, GPT-3 | The chatbot, known as ChatGPT, is based on a language model that excels in natural language processing tasks is capable of generating human-like text responses. The model underwent extensive training on a large corpus of text data, enabling it to understand language nuances, generate accurate responses, and handle multiple inputs and languages. The destiny scope of chatbots includes interacting with multiple users and expanding inside the numerous applications inclusive of studies, schooling, and creative writing. |
| Shashi Kant Singh, Shubham Kumar, Pawan Singh Mehra | 2023 | Artificial Intelligence, Chatbots, Chat GPT, Google Bard AI | It especially specializes in the evaluation of Chat GPT, it is a first-rate innovation within the location of Artificial Intelligence. It recognizes the restrictions of language fashions like GPT-2 and GPT-3, collectively with biases and restrained accuracy. It highlights the want for interactive language fashions like Chat GPT to cope with those limitations and decorate easy normal performance. Enables the chatbot to recognize and reply to complicated queries. It permits making the chatbot available online and integrating it into programs. |
| IlmanShazhaev, Arbi Tularov, Dmitry Mikhaylov, Islam Shazhaev, Abdulla Shafeeg | 2023 | Artificial Intelligence, ChatGPT, digital community, voice assistant | Neural Networks allow the chatbot and voice assistant to perform tasks that include answering questions, writing code, and composing essays. OpenAI’s GPT language the model has been trained to generate meaningful responses to questions and can perform various tasks, including finding bugs in code and composing poetry. The system has filters to keep away from debatable subjects and save you from growing texts about illegal or immoral sports. Farcana has integrated the functionality of the GPT chatbot and voice assistant to enhance player experience, familiarize players with game mechanics, and manage gaming accounts. |
| Shaji Thorn Blue, M. Brindha | 2019 | Boundary Box prediction, Edge Detection, Object Detection, YOLO, YOLO9000, YOLOv3 | The proposed algorithm demonstrates improved precision in drawing boundary boxes compared to YOLOv3. Results show the proposed method outperforms YOLOv3 in terms of accuracy, as indicated by the Intersection over Union (IOU) comparison. The proposed work exhibits potential for object detection with precise boundary boxes, although limitations exist, particularly in scenarios involving sharp objects and high image noise. |
| Shraddha Mane, Supriya Mangale | 2018 | CNN, Object detection, Tensorflow, Tracking | The model uses Tensorflow object detection API which is robust in detecting objects in complex scenes and background conditions. Detects object location (x,y,w,h) which is then passed to the tracking algorithm. Uses a CNN-based tracking algorithm instead of traditional computer vision techniques. Loads weights of a pre-trained model capable of incorporating temporal information. Able to track objects at 150 FPS and handle occlusion scenarios. The main innovation is in using deep learning models to improve generalization across challenging uncontrolled video settings. |
| Anis Koubaa | 2023 | Human-Robot Interactions, ROS, ROS2, ChatGPT, Large Language Model, Robot Operating System | The system utilizes prompt engineering to convert unstructured human language commands into structured robotic instructions, enabling seamless communication between humans and robots. The implementation includes a proof-of-concept for robot navigation and demonstrates the use of an ontology to guide ChatGPT in generating context-specific robotic commands. It also emphasizes the importance of prompt engineering and ontology development, as well as the limitations and potential biases of using LLMs for human-robot interaction. |
| Xiao-Feng Zhao, Zi-Heng Chen, He-Feng Yin, Xiao-Jun Wu | 2023 | Artificial Intelligence cloud, edge computing, face recognition, speech recognition | The system addresses the management requirements for visitors in enterprises and government departments, aiming to improve efficiency and accuracy while reducing labor costs. By utilizing AI cloud and edge computing, the system incorporates face recognition and speech recognition technologies to enable functions such as visitor verification, and voice self-registration. The system is composed of a cloud service layer and an edge computing layer, with hardware components including RK3399 motherboards, microphone array, camera, and IR temperature sensor. The system’s cloud-edge collaboration enhances computing capacity, improves real-time performance, and ensures high accuracy. The testing results showed high recognition accuracy and short recognition delays. |

# III. Problem Formulation

Human-robot interplay is confined because of bad conversational skills, emotional expertise, and face recognition. Existing robots are not gifted sufficiently to interact in nuanced conversations or apprehend human feelings, hindering customized assistance. Robotic systems additionally struggle to effectively grow to be privy to items and faces. These barriers make it hard for robots to come to be beneficial and friendly partners.

Addressing these problems calls for modern answers that beautify the conversational emotional, and perceptual abilities of robots, in the long run fostering advanced utility and friendliness in human-robotic interactions. This survey paper's goalis to discover and observe cutting-edge upgrades and proposed methodologies inside the realm of enhancing robot verbal exchange, comprehensive, and popularity to provide valuable insights and instructions for destiny studies inside the place.

Develop a smart robot system incorporating, superior conversational and cognitive skills. Implement revolutionary computational techniques to decorate the gadget’s abilities. Enable the robotic to apprehend and bear in mind the human beings it encounters.

The memory retention skills to keep and preserve in thoughts pertinent statistics, ensuring a comprehensive expertise of patron history. Achieve customized interactions with the aid manner of leveraging stored information about users and Enhance context in conversations to create a more human-robot interaction.

Integrate object detection functionality to permit the robotic to perceive and interact with gadgets in its environment. Implement face detection to effectively find out and interact with users, even at the end of preliminary encounters.

Integrate a chatbot to facilitate seamless conversation in the robotic tool. Assess how the chatbot has superior general communication abilities of the system.

# IV. Solution To The Problem

Our mission is a floor-breaking aggregate of human-targeted layout and modern technology, resulting in a robot with exquisite interactive and conversational abilities. Driven via GPT, a sophisticated language model, our device is prepared to transform human-robot interactions by fusing computational electricity with intelligence, flexibility, and knowledge. Our goal has constantly been to construct a robotic that can converse obviously and meaningfully, however also has the mental capacity to pick out up new skills through the years and alter to unique situations, making each interplay particular.

The robot's conversational prowess is powered by way of GPT, a complex language version designed to have interaction in natural and human-like conversations. Leveraging GPT's computational strategies complements conversational capabilities, making interactions extra fluid and significant. From its inception, our purpose has been to create a robot that now not simplest engages in natural and meaningful conversations however also possesses the cognitive skills to study, adapt, and personalize interactions over time. The chatbot flow is shown in fig 1.



Fig 1. GPT chatbot workflow

                One special feature of the device is its memory retention and consider, facilitated by GPT. It lets in the robotic to understand and recollect people, even at some point of preliminary encounters. Information approximately past interactions is saved, contributing to a more personalized and context-aware engagement. GPT allows hold pertinent information approximately patron encounters, ensuring complete data of the individual's records. This information aids in tailoring responses and interactions based on the user's preferences and past engagements, all facilitatedseamlessly through the GPT interface at the front end.

Essential functionalities such as object detection and face recognition are integrated into the system. Object detection skills permit the robotic to grow to be aware of and have interaction with physical gadgets in its environment. Here first the video frames are taken from a camera and then each frame is sent to the object detection model which uses YOLOv8 in the background. YOLOv8 model run on the concept of CNN (convolutional neural network). After processing the image frames using YOLOv8 model we get a list of class names where all the detected objects are included. The architecture of YOLOv8 can be seen in fig 2.

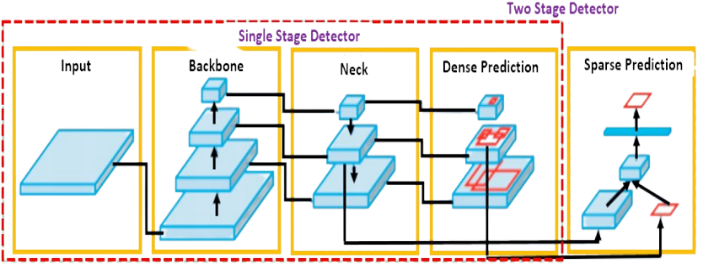


Fig 2. Object detection architecture

Face popularity complements the robot's capacity to understand and consider individuals, fostering a greater personalised experience. Here also the video frames are taken as input and each frame is processed using the face recognition model. The uses face recognition library of python and identifies the faces of individuals by comparing the features of current user face with the faces in its database. After comparing it gives an output with a boolean value. An integrated chatbot component is incorporated to facilitate seamless communication. You can see how the face is detected in fig 3.

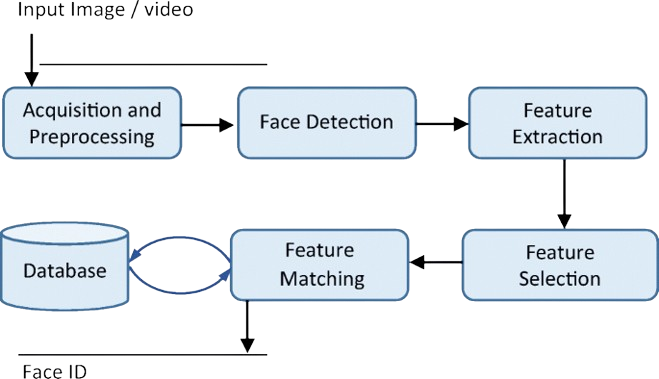


Fig 3. Face Recognition architecture

Our machine capabilities a visitor monitoring machine the use of the face recognition version to pick out recognised and unknown individuals. Video frames are processed to compare facial features with stored data. If recognized, the individual is identified otherwise, they're categorized as unknown. This complements personalization and safety, permitting the robotic to adapt interactions for that reason. The architecture visitor tracking system can be seen in fig 4.

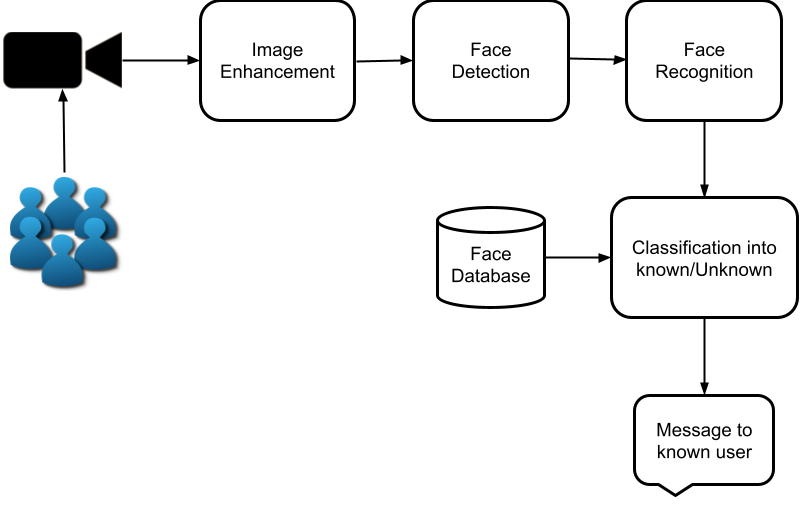


Fig 4. Visitor tracking system architecture

This chatbot enhances the robot's conversational abilities, enabling it to handle a wide range of queries and engage in meaningful dialogues with users. Object detection and face recognition lay the foundation for more sophisticated interactions. The robot can respond to visual stimuli, identify people, and interact with objects, expanding its range of capabilities. With its ability to recognize individuals, interact with physical objects, and track visitors, our system heralds a new era of personalized and context-aware robotics, poised to redefine the boundaries of human-robot coexistence in diverse domains. The infobots architecture can be seen in fig 5.

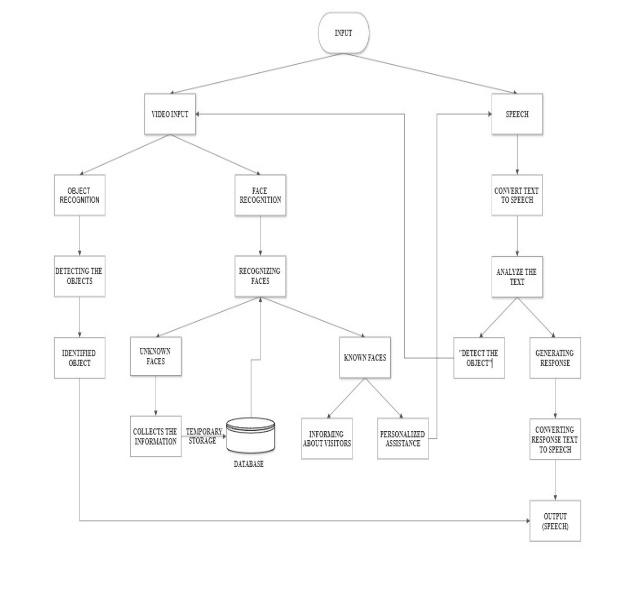


Fig 5. Infobot architecture

In summary, our project embodies the convergence of innovation and practicality, showcasing the transformative potential of artificial intelligence in enhancing human-robot interactions. Through the integration of GPT for conversation, YOLOv8 for object detection, and face recognition technology, our robot stands as a testament to the power of interdisciplinary collaboration and technological advancement. With its ability to recognize individuals, interact with physical objects, and track visitors, our system heralds a new era of personalized and context-aware robotics, poised to redefine the boundaries of human-robot coexistence in diverse domains.

V. PREDICTED OUTPUT

* Object Detection output:

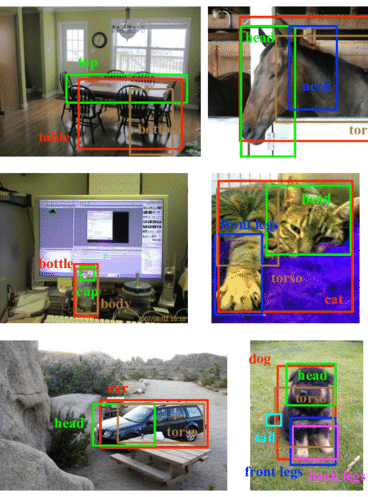


Fig 6. Objects detected

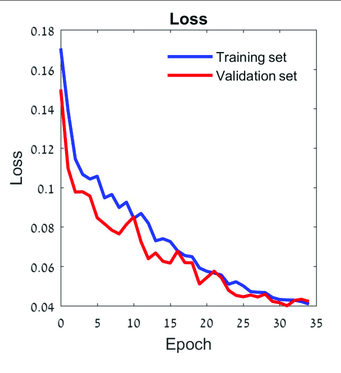


Fig 7. Loss function graph of object recognition

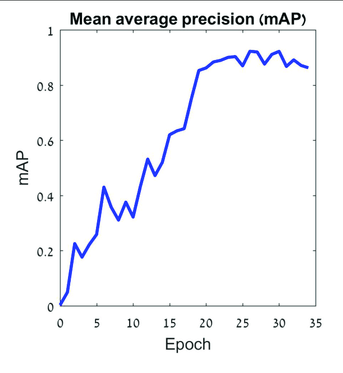


Fig 8. mAP graph of object recognition

* Face Recognition:



Fig 9. Identified Faces and unknown faces

* Visitor tracking system:



Fig 10. Visitor report generated by visitor tracking system

# VI. Conclusion

Our ultimate goal is to develop an AI robot companion that is intelligent and helpful. To gain this we advocate a solution that focuses on regions such, as detecting gadgets spotting faces, maintaining memories, and integrating chatbot talents. By taking a method we aim to enhance the abilities of the robot and make it an integral and personable assistant in numerous situations. Through those efforts, we goal to increase the field of human-robot interplay and provide users with an enjoyable revel. Our task objective is to increase a chatbot robotic that excels in communication and popularity skills. By addressing demanding situations in human-robotic interaction we attempt to push the limits of cooperation amongst people and robots. This venture can convert generations and open the door to a time when robots would easily enhance human connections and sell dating among human beings and current technical systems.

##### References

1. N. Zsoter et al., "PET-CT based automated lung nodule detection," 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA, 2012, pp. 4974-4977, Doi: 10.1109/EMBC.2012.6347109.
2. C. Hima Bindu and K. Veera Swamy, "Medical image fusion using content based automatic segmentation," International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014), Jaipur, India, 2014, pp. 1-5, Doi: 10.1109/ICRAIE.2014.6909206.
3. Himanshi, V. Bhateja, A. Krishn and A. Sahu, "An improved medical image fusion approach using PCA and complex wavelets," 2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), Greater Noida, India, 2014, pp. 442-447, Doi: 10.1109/MedCom.2014.7006049.
4. Z. Guo, X. Li, H. Huang, N. Guo and Q. Li, "Medical image segmentation based on multi-modal convolutional neural network: Study on image fusion schemes," 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 2018, pp. 903-907, Doi: 10.1109/ISBI.2018.8363717.
5. M. B. Abdulkareem, "Design and Development of Multimodal Medical Image Fusion using Discrete Wavelet Transform," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 2018, pp. 1629-1633, Doi: 10.1109/ICICCT.2018.8472997.
6. K. Vanitha, D. Satyanarayana and M. N. G. Prasad, "Multimodal Medical Image Fusion Based on Hybrid L1- L0 Layer Decomposition Method," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 2019, pp. 1-5, Doi: 10.1109/ICCCNT45670.2019.8944896.
7. Jiaxin Li, Houjin Chen, Yanfeng Li, and Yahui Peng. 2019. A Novel Network Based on Densely Connected Fully Convolutional Networks for Segmentation of Lung Tumors on Multi-Modal MR Images. In Proceedings of the 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM 2019). Association for Computing Machinery, New York, NY, USA, Article 69, 1–5. https://doi.org/10.1145/3358331.3358400
8. K. S. Asish Reddy, K. Kalyan Kumar, K. N. Kumar, V. Bhavana and H. K. Krishnappa, "Multimodal Medical Image Fusion Enhancement Technique for Clinical Diagnosis," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 586-589, Doi: 10.1109/ICCMC.2019.8819840.
9. H. Yan and Z. Li, "A Multi-modal Medical Image Fusion Method in Spatial Domain," 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China, 2019, pp. 597-601, Doi: 10.1109/ITNEC.2019.8729143.
10. V. A. Rani and S. Lalitha Kumari, "A Hybrid Fusion Model for Brain Tumor Images of MRI and CT," 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2020, pp. 1312- 1316, Doi: 10.1109/ICCSP48568.2020.9182371.
11. Kaur M, Singh D, Multi-modality medical image fusion technique using multi-objective differential evolution based deep neural networks. J Ambient Intell Human Comput 12, 2483–2493 (2021). Doi: https://doi.org/10.1007/s12652-020-02386-0
12. L. Wang, J. Zhang, Y. Liu, J. Mi, and J. Zhang, "Multimodal Medical Image Fusion Based on Gabor Representation Combination of Multi-CNN and Fuzzy Neural Network," in IEEE Access, vol. 9, pp. 67634- 67647, 2021, Doi: 10.1109/ACCESS.2021.3075953.
13. C. Gao, C. Song, Y. Zhang, D. Qi, and Y. Yu, "Improving the Performance of Infrared and Visible Image Fusion Based on Latent Low-Rank Representation Nested with Rolling Guided Image Filtering," in IEEE Access, vol. 9, pp. 91462-91475, 2021, Doi: 10.1109/ACCESS.2021.3090436.
14. X. Fu, L. Bi, A. Kumar, M. Fulham, and J. Kim, "Multimodal Spatial Attention Module for Targeting Multimodal PETCT Lung Tumor Segmentation," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 9, pp. 3507- 3516, Sept. 2021, Doi: 10.1109/JBHI.2021.3059453.
15. K. Kusram, S. Transue and M. -H. Choi, "Two-Phase Multimodal Image Fusion Using Convolutional Neural Networks," 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 2021, pp. 1874-1878, Doi: 10.1109/ICIP42928.2021.9506703.
16. Barrett J., &amp; Viana, T. (2022). EMM-LC Fusion: Enhanced Multimodal Fusion for Lung Cancer Classification. AI, 3(3), 659–682. Doi: https://doi.org/10.3390/ai3030038
17. Y. Zhang, H. Zhang, L. Xiao, Y. Bai, V. D. Calhoun and Y. -P. Wang, "Multi-Modal Imaging Genetics Data Fusion via a Hypergraph-Based Manifold Regularization: Application to Schizophrenia Study," in IEEE Transactions on Medical Imaging, vol. 41, no. 9, pp. 2263-2272, Sept. 2022, Doi: 10.1109/TMI.2022.3161828.
18. Das K. P., &amp; Chandra, J. (2022). Multimodal Classification on PET/CT Image Fusion for Lung Cancer: A Comprehensive Survey. ECS Transitions, 107(3649). Doi: 10.1149/10701.3649ecst
19. Maha M. Althobaiti, Amal Adnan Ashour, Nada A. Alhindi, Asim Althobaiti, Romany F. Mansour, Deepak Gupta, Ashish Khanna, &quot;Deep Transfer Learning-Based Breast Cancer Detection and Classification Model Using Photoacoustic Multimodal Images&quot;, BioMed Research International, vol. 2022, Article ID 3714422, 13 pages,2022. Doi:<https://doi.org/10.1155/2022/3714422>
20. Haribabu M., &amp; Guruviah, V. (2023). An Improved Multimodal Medical Image Fusion Approach Using Intuitionistic Fuzzy Set and Intuitionistic Fuzzy Cross-Correlation. Diagnostics, 13(14), 2330. Doi: https://doi.org/10.3390/diagnostics13142330.
21. Wang, Kunpeng; Zheng, Mingyao; Wei, Hongyan; Qi, Guanqiu; Li, Yuanyuan (2020). Multi-Modality Medical Image Fusion Using Convolutional Neural Network and Contrast Pyramid. Sensors, 20(8), 2169–. Doi:10.3390/s20082169.
22. M. Qiguang and W. Baoshu, "A Novel Image Fusion Method Using Contourlet Transform," 2006 International Conference on Communications, Circuits and Systems, Guilin, China, 2006, pp. 548-552, Doi: 10.1109/ICCCAS.2006.284696.
23. X. Zhang, "Deep Learning-Based Multi-Focus Image Fusion: A Survey and a Comparative Study," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 9, pp. 4819-4838, 1 Sept. 2022, Doi: 10.1109/TPAMI.2021.3078906.
24. Li, Shutao; Kang, Xudong; Fang, Leyuan; Hu, Jianwen; Yin, haitao (2016). Pixel-level image fusion: A survey of the state of the art, Information Fusion, (), S1566253516300458-. Doi:10.1016/j.inffus.2016.05.004
25. Malik, S.S., Shivprasad, B.J., & Maruthi, G.B. (2013). Feature Level Image Fusion.
26. Hu, Qiu; Hu, Shaohai; Zhang, Fengzhen (2020). Multi-modality medical image fusion based on separable dictionary learning and Gabor filtering. Signal Processing: Image Communication, (), 115758–Doi:10.1016/j.image.2019.115758
27. H. Li, X. -J. Wu and J. Kittler, "MDLatLRR: A Novel Decomposition Method for Infrared and Visible Image Fusion," in IEEE Transactions on Image Processing, vol. 29, pp. 4733-4746, 2020, Doi: 10.1109/TIP.2020.2975984.
28. Balasubramaniam, P.; Ananthi, V.P. (2014). Image fusion using intuitionistic fuzzy sets. Information Fusion, 20(), 21–30. Doi: 10.1016/j.inffus.2013.10.011
29. B. Yang, S. Li and F. Sun, "Image Fusion Using Nonsubsampled Contourlet Transform," Fourth International Conference on Image and Graphics (ICIG 2007), Chengdu, China, 2007, pp. 719-724, doi: 10.1109/ICIG.2007.124.
30. Xia, Kai-jian; Yin, Hong-sheng; Wang, Jiang-qiang (2018). A novel improved deep convolutional neural network model for medical image fusion. Cluster Computing, (), –. doi:10.1007/s10586-018-2026-1.