Humanoid Robotic Head with Two Active Senses using Python and Arduino

Muhammed Rizwan P. S., Sahil Rao, Sonal Balayan, Neeraj Bhandari, Gyanender Kumar

Department of Computer Science and Engineering
HMRITM, Plot No. 326, Hamidpur, Delhi-110036.

Abstract—**This research paper studies and presents the research and development of a humanoid robotic head with two active senses for input. The outputs produced by the robot is primarily auditory and behavior based. The project develops and proposes several features and functions that were developed using Python and Arduino, integrated with the hardware part to perform some basic behavioral tasks like moving lips, blinking, looking, etc. Functions include image captioning to recognize the scene, chatbot feature to talk and interact using audio, functions to control the servos to control eyes, mouth and eyelids.**

Index Terms- Humanoid Robot, Computer Vision, Natural language Processing, OCR, Chatbot, OpenCV, OpenAI, Arduino, Interactive Machine, Talking Android.

I. Introduction

Humanoid robots are advanced machines designed to mimic the physical characteristics and movements of humans. These robots are typically equipped with sensors, actuators, and other advanced technologies that allow them to perform a wide range of tasks, from walking and running to manipulating objects and communicating with humans. Humanoid robots are often used in research and development, as they provide a valuable platform for studying human-like behaviours and movements. Researchers can use these robots to explore topics such as balance, coordination, and cognitive processing, as well as to test new technologies and algorithms [1, 2].

Humanoid robots have a wide range of potential applications in various fields such as healthcare, education, entertainment, and manufacturing. They can be used to assist doctors and nurses in hospitals and clinics, providing support for patient monitoring and medicine delivery. In education, humanoid robots can be used as interactive tools for teaching children, especially those with learning difficulties. In entertainment, they can be used as performers or as interactive exhibits in museums and theme parks. In manufacturing, humanoid robots can be used for tasks such as assembly, packaging, and quality control [1, 2, 3].

One of the key benefits of humanoid robots is their ability to interact with humans in a natural and intuitive way [3, 4]. This makes them well-suited for a range of applications, from healthcare and education to entertainment and hospitality. For example, humanoid robots can be used to assist patients with physical therapy or to teach children with special needs. However, designing and building humanoid robots is a complex and challenging task. It requires expertise in a range of fields, including mechanical engineering, computer science, and artificial intelligence. Nevertheless, the development of humanoid robots continues to advance rapidly, with new breakthroughs in technology and design emerging all the time.

Humanoid robots represent a fascinating area of research and development with enormous potential to transform many aspects of our lives [4]. While there are still many challenges to overcome, the advances made in this field are providing new opportunities for innovation and progress.

In this paper, we are proposing an easy to develop humanoid robotic head that can facilitate up to two active senses which are sense of vision and sense of audition with the help of the latest computer vision and natural language processing technologies. To make the robot itself, we will be bringing Arduino into the picture and integrating everything with Python and C++.

The final outcome, according to our expectations, is to showcase the sense of vision, sense of audition, ability to talk and chat, ability to interact using behaviours of eyes, eyelids and lips. Also conducting some experiments for research purposes would be a high priority task.

II. Taxonomy on the Basis of Functions

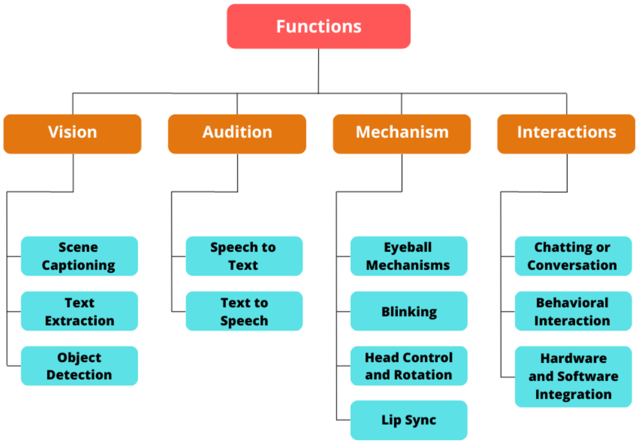


Fig. 1: Functional Taxonomy

1. Vision
    - Scene Captioning and Image Captioning
    - Text Extraction or Optical Character Recognition
    - Object Detection
2. Audition
    - Speech to Text
    - Text to Speech
3. Mechanisms

- Eyeball Mechanisms
- Blinking
- Head Control and Rotation
- Lip Sync

4. Interactions
    - Chatting or Conversation
    - Behavioral Interaction
    - Hardware and Software Integration
5. Miscellaneous
    - Chatting or Conversation
    - Behavioral Interaction
    - Hardware and Software Integration

III. SENSES

The sense of vision is a remarkable and intricate sensory system that allows human beings to perceive and interpret the visual world around them [5, 6]. It is through our eyes that we receive visual information, enabling us to navigate our environment, recognize objects and faces, appreciate the beauty of nature, and engage in various activities. This article aims to provide a comprehensive exploration of the sense of vision, covering its anatomy, the process of visual perception, the mechanisms of color vision, depth perception, and visual illusions [8, 9].

The sense of audition, or hearing, is a remarkable sensory system that allows humans to perceive and interpret the intricate world of sound. It plays a fundamental role in communication, music, and our overall understanding of the environment. This article aims to provide a comprehensive exploration of the sense of audition, covering its anatomy, the process of auditory perception, the mechanisms of sound localization and pitch perception, the impact of hearing impairments, and the applications of audition in various fields [10, 21, 22].

IV. ROBOT MECHANISMS

Robotics has emerged as a groundbreaking field that combines engineering, computer science, and artificial intelligence to create intelligent machines capable of performing various tasks. At the core of these machines lies the field of robot mechanics, which focuses on the design, construction, and control of the mechanical systems that enable robots to move, manipulate objects, and interact with their environment. This article aims to provide a comprehensive exploration of robot mechanics, covering the fundamental components of robot systems, types of robot joints and actuators, robot kinematics and dynamics, and the advancements in robot mechanics. Through a detailed examination of these aspects, we can gain a deeper understanding of the mechanical foundations that drive robotic technologies [11, 12].

V. HUMAN-ROBOT INTERACTION

Human-Robot Interaction (HRI) is a multidisciplinary field that focuses on studying and designing interfaces and interactions between humans and robots. It aims to create seamless and intuitive communication channels between humans and intelligent machines, enabling effective collaboration, cooperation, and understanding. This article delves into the realm of HRI, exploring its significance, challenges, interaction modalities, and applications in various domains. Through a comprehensive analysis of these aspects, we can gain a deeper understanding of the dynamics and potential of human-robot interaction [13, 23].

VI. DESIGN PLAN

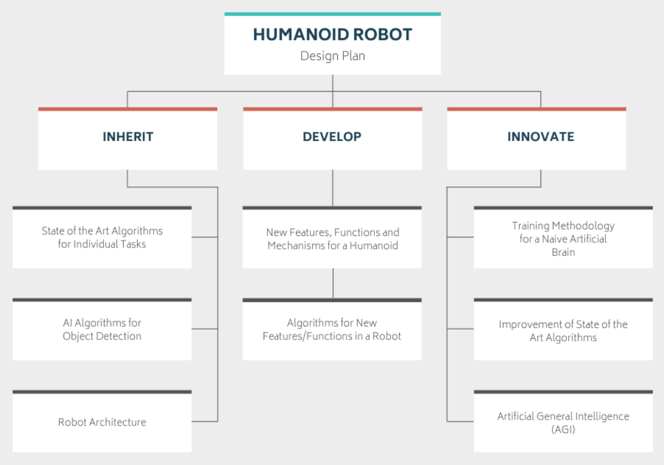

HUMANOID ROBOT
Design Plan
INHERIT
DEVELOP
INNOVATE
State of the Art Algorithms for Individual Tasks
AI Algorithms for Object Detection
Robot Architecture
New Features, Functions and Mechanisms for a Humanoid
Algorithms for New Features/Functions in a Robot
Training Methodology for a Naive Artificial Brain
Improvement of State of the Art Algorithms
Artificial General Intelligence (AGI)

Fig. 2: Design Planning

This project due to it’s vastness, has been developed partially using the above proposed methodology. The adoption of such a procedure helps in saving time, resources and effort. We followed a 3 phase development process,

1. **Inheritance:** Use present day state of the art algorithm that are already at it’s best and results in better accuracy. Working on algorithms that are already available is a waste of time. Hence using and altering the existing is preferable.
2. **Development:** For introduction of new features, we develop our own algorithms for these features and test them for acceptable accuracy.
3. **Innovation:** Introduction of new algorithms, AI’s and upgradation of existing state of the art algorithms in this field. It is completely research based and hence cannot be attributed to have groundbreaking results.

VII. PROJECT WORK

Languages used: Python, C++

Platforms used: Jupyter notebook, Anaconda, Arduino IDE [26], Hugging face [25].

Libraries: OpenAI [27], OpenCV, PyFirmata [28], MediaPipe, Transformers, Speech Recognition

Hardware [20]: Arduino Uno and Programming Cable * 1, MG90S Micro Servo * 4, SG90 Micro Servo * 3, Male to Male Jumper Wires with Square Head * 24, Breadboard * 1, Screwdriver Set, Adhesives, Metal Wires * 37, Safety Pin, Cardboard Boxes, Medicine Bottle, Empty Pens

VIII. FUNCTIONS INVOLVED

1. rotate_servo()
    - Purpose: Operating individual servos.

- Libraries Involved: pyfirmata.
- Arguments: Pin and angle.
- Outputs: None.
- Inherited Functions: None.
- Process: Using the direct function from Pyfirmata, we use an object to rotate a servo connected to a specific pin in Arduino Uno to required angles.

2. talk()
   - Purpose: Looping code for talking mechanism.
   - Libraries Involved: time and random.
   - Arguments: None.
   - Outputs: None.
   - Inherited Functions: rotate_servo
   - Process: This function controls a single servo connected with the mouth pieces and makes the lips move at predefined angles to facilitate the lip sync mechanism.

3. blinking()
   - Purpose: Looping code for blinking mechanism.
   - Libraries Involved: time.
   - Arguments: None.
   - Outputs: None.
   - Inherited Functions: rotate_servo
   - Process: Makes the blinking mechanism possible by using predefined angles for the eyelids positions and hence mimicking a human blink.

4. idle_looking()
   - Purpose: Looping code for the eyes that look randomly and blink randomly to mimic a human behavior.
   - Libraries Involved: time and random.
   - Arguments: None.
   - Outputs: None.
   - Inherited Functions: rotate_servo and blinking.
   - Process: Makes use of random angles between a predefined range to randomly look here and there.

5. wish_master()
   - Purpose: Greets the user.
   - Libraries Involved: datetime and random.
   - Arguments: None.
   - Outputs: None.
   - Inherited Functions: engine_speak.
   - Process: Using datetime module, we get the time of the day and use it to decide whether the user should be greeted with a good morning, afternoon or evening.

6. engine_speak()
   - Purpose: Speaks the given texts.
   - Libraries Involved: pyttsx3.
   - Arguments: audio in the form of text.
   - Outputs: None.
   - Inherited Functions: None.
   - Process: This function speaks the texts passed as an argument, it converts text to speech.

7. engine_listen()
   - Purpose: Listens to the users sound inputs.
   - Libraries Involved: speech_recognition.
   - Arguments: None.
   - Outputs: Query in the form of text.
   - Inherited Functions: engine_speak and engine_listen.
   - Process: Detects, recognizes and records the speech from the user and converts the speech to text.

8. camera()
   - Purpose: Open camera for various purposes.
   - Libraries Involved: cv2.
   - Arguments: None.
   - Outputs: None.
   - Inherited Functions: None.
   - Process: Using OpenCV module, we access the camera and take a photo of the scene to detect objects and written texts in the image for OCR.

9. textextractor()
   - Purpose: Detects texts from the images.
   - Libraries Involved: easyocr.
   - Arguments: None.
   - Outputs: Prompt in the form of text.
   - Inherited Functions: None.
   - Process: Using OCR, we extract the texts in an image to use it to read out loud or to pass it to the chatbot for processing.

10. chatbot()
    - Purpose: An interactive and talking chatbot.
    - Libraries Involved: openai.
    - Arguments: Query in the form of text.
    - Outputs: Response in the form of text.
    - Inherited Functions: None.
    - Process: Chatbot which can answer questions, engage in casual talks using the openai's pre trained models. We are using an API key to access the pre trained model.

11. ImageCaptioning()
    - Purpose: Image or scene captioning.
    - Libraries Involved: cv2 and transformers [14, 15, 16].
    - Arguments: None.
    - Outputs: None.
    - Inherited Functions: engine_speak.
    - Process: The function developed using transformer to caption a scene or an image.

12. Decypher()
    - Purpose: Main function acting as the voice assistant.
    - Libraries Involved: threading.
    - Arguments: None.
    - Outputs: None.
    - Inherited Functions: All of the above.
    - Process: This is the main function that makes the robot alive by combining and integrating the software and hardware together. Decypher is the robot's code name.

IX. Result

1. Head Mechanisms [17, 18, 19, 29]

   Expectations:
   - Human like eye mechanism
   - Human like mouth mechanism.
   - Rotating neck.

- Mimicking natural human behaviors in combination.

Results/Outputs:

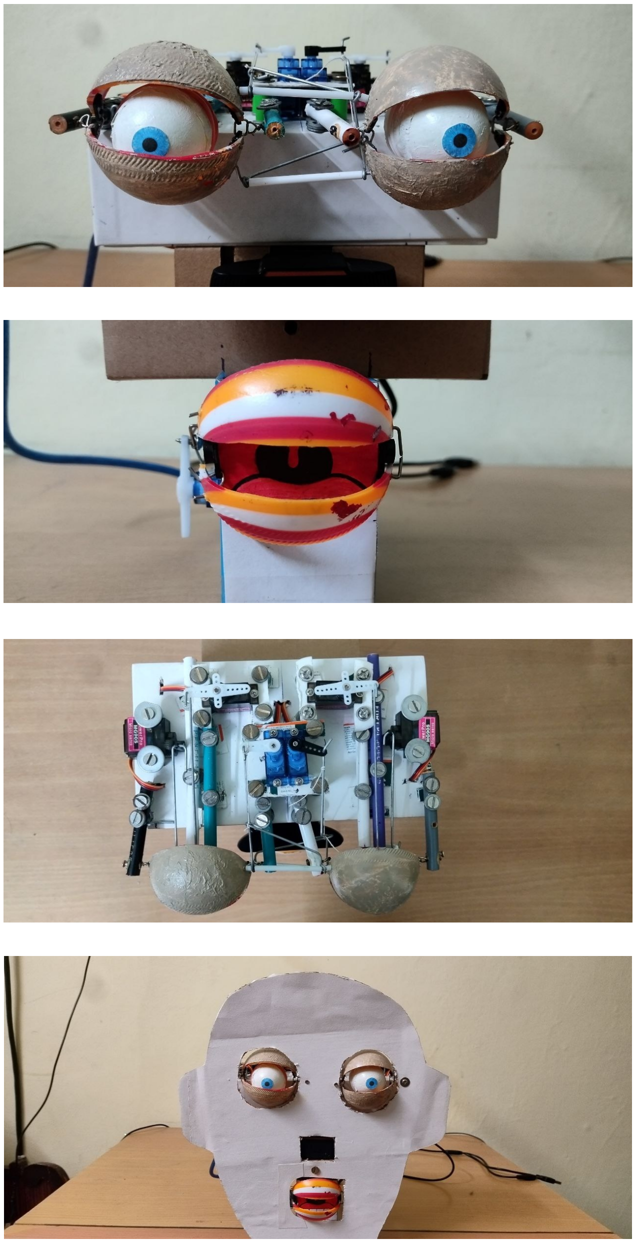


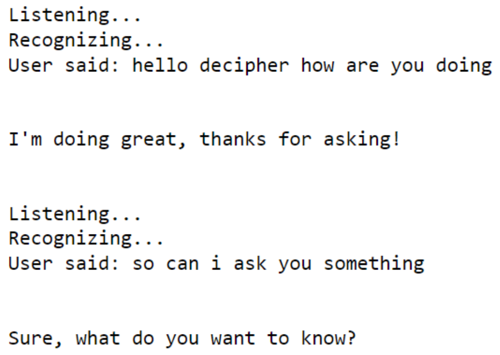
Fig. 3: Mechanisms for the Head

2. Chatbot Features

Expectations:

- Casual conversation.
- Searching things that user asks.
- Solving questions.
- Searching from different functions for example the text returned from the OCR function.

Results/Outputs:



```
Listening...
Recognizing...
User said: hello decipher how are you doing


I'm doing great, thanks for asking!


Listening...
Recognizing...
User said: so can i ask you something


Sure, what do you want to know?
```

Fig. 4: A Casual Conversation

3. Reading and Answering

Expectations:

- Ability to extract texts from the environment.
- Ability to order the extracted texts to make it prompt ready.

Results/Outputs:

- The image was taken in a way to replicate a natural environmental scenario.

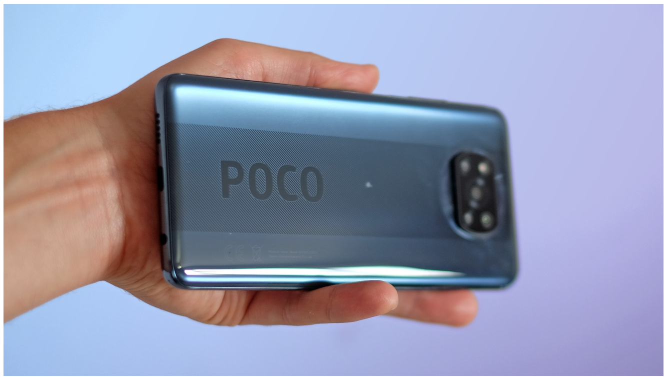
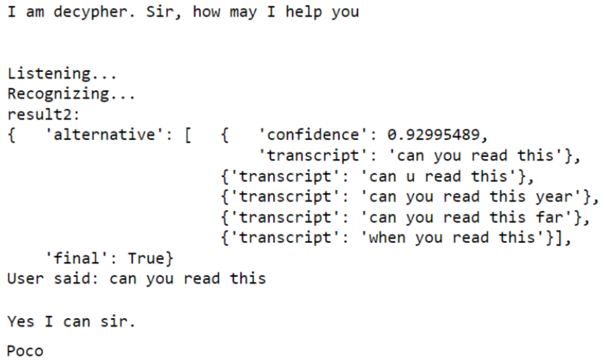


Fig. 5: Experimental Image for OCR



```
I am decypher. Sir, how may I help you


Listening...
Recognizing...
result2:
{   'alternative': [   {   'confidence': 0.92995489,
                           'transcript': 'can you read this'},
                       {'transcript': 'can u read this'},
                       {'transcript': 'can you read this year'},
                       {'transcript': 'can you read this far'},
                       {'transcript': 'when you read this'}],
    'final': True}
User said: can you read this


Yes I can sir.
Poco
```

Fig. 6: Detecting the Text on the Phone

4. Voice Assistant

Expectations:

- Greeting and giving goodbyes.
- Invoking appropriate functions.
- Providing an interface for the functions to be integrated.
- The starting and ending of the robot.

Results/Outputs: Shown in chatbot section.

5. Object Detection

   Expectations:
   - Detecting objects present in the environment.
   - Naming the objects according to the users request.

   Results/Outputs: Under development.

6. Scene Captioning

   Expectations:
   - Generating captions of the environment.
   - Telling the users about how the environment looks to the robot.

   Results/Outputs: The photo took for the scene is as given below:

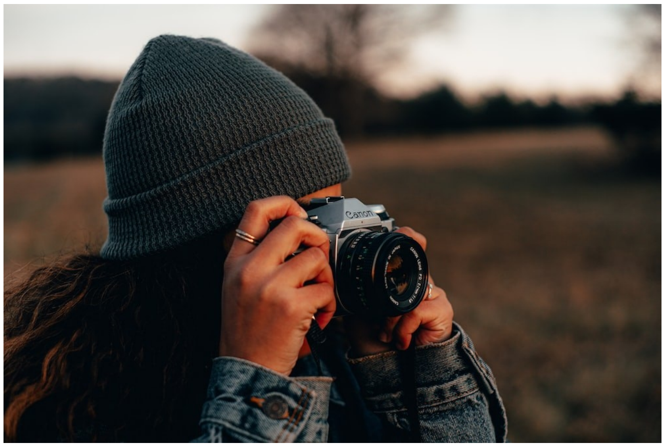
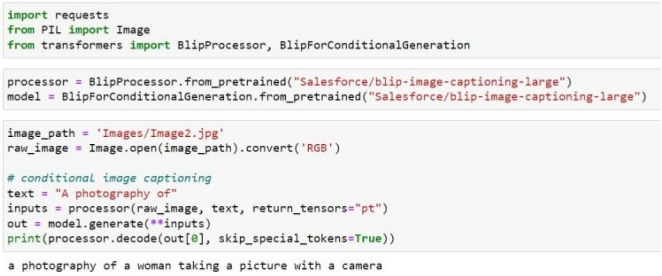


Fig. 7: Experimental Image for Scene Captioning



```
import requests
from PIL import Image
from transformers import BlipProcessor, BlipForConditionalGeneration
```

```
processor = BlipProcessor.from_pretrained("Salesforce/blip-image-captioning-large")
model = BlipForConditionalGeneration.from_pretrained("Salesforce/blip-image-captioning-large")
```

```
image_path = 'Images/Image2.jpg'
raw_image = Image.open(image_path).convert('RGB')

# conditional image captioning
text = "A photography of"
inputs = processor(raw_image, text, return_tensors="pt")
out = model.generate(**inputs)
print(processor.decode(out[0], skip_special_tokens=True))
```

```
a photography of a woman taking a picture with a camera
```

Fig. 8: Generating Caption

- Generated caption: A photography of a women taking a picture with a camera.
- Comment: Highly accurate and explanatory.

The project repository can be found in the reference section of this research paper [29].

X. CONCLUSION

This research was done with the sole purpose of

ACKNOWLEDGMENT

This is to acknowledge the valuable contributions made by several people within the faculty of our institution and outside. Special thanks to our guide, Mr. Gyanender Kumar, for his time and energy to help us develop and provide with the resources we needed. Thanks to all the people who voluntarily or involuntarily got involved and helped us in this project.

REFERENCES

[1] I. Pavan Raju, S. Sikka, A. Garg, and M. Pandey, “A Brief Review of Recent Advancement in Humanoid Robotics Research.”

[2] Dr. S. V. Viraktamath, “Humanoid Robot: A Review,” *Int J Res Appl Sci Eng Technol*, vol. 9, no. 8, pp. 2884–2894, Aug. 2021, doi: 10.22214/ijraset.2021.37890.

[3] N. Rani, “Humanoid Robotics.” [Online]. Available: www.ijert.org

[4] R. Mahum, F. S. Butt, K. Ayyub, S. Islam, M. Nawaz, and D. Abdullah, “A review on humanoid robots,” *International Journal of ADVANCED AND APPLIED SCIENCES*, vol. 4, no. 2, pp. 83–90, Feb. 2017, doi: 10.21833/ijaas.2017.02.015.

[5] Fulkerson M. Rethinking the senses and their interactions: the case for sensory pluralism. Front Psychol. 2014 Dec 10;5:1426. doi: 10.3389/fpsyg.2014.01426. PMID: 25540630; PMCID: PMC4261717.

[6] Nandagopal R, R. R. (2015). A Study on the Influence of Senses and the Effectiveness of Sensory Branding. Journal of Psychiatry, 18(2). https://doi.org/10.4172/Psychiatry.1000236

[7] U. Schmidt-Erfurth, A. Sadeghipour, B. S. Gerendas, S. M. Waldstein, and H. Bogunović, “Artificial intelligence in retina,” *Prog Retin Eye Res*, vol. 67, pp. 1–29, Nov. 2018, doi: 10.1016/j.preteyeres.2018.07.004.

[8] Perkins, E. S. and Davson, Hugh. "human eye." Encyclopedia Britannica, April 19, 2023. https://www.britannica.com/science/human-eye.

[9] L. Gu *et al.*, “A biomimetic eye with a hemispherical perovskite nanowire array retina,” *Nature*, vol. 581, no. 7808, pp.

[10] Hawkins, J. E.. "human ear." Encyclopedia Britannica, March 29, 2023. https://www.britannica.com/science/ear.

[11] W. Wei, B. Zhou, B. Fan, M. Du, G. Bao, and S. Cai, “An Adaptive Hand Exoskeleton for Teleoperation System,” *Chinese Journal of Mechanical Engineering (English Edition)*, vol. 36, no. 1, Dec. 2023, doi: 10.1186/s10033-023-00882-w.

[12] H. H. Poole, “Introduction to Robot Mechanics,” in *Fundamentals of Robotics Engineering*, Dordrecht: Springer Netherlands, 1989, pp. 55–76. doi: 10.1007/978-94-011-7050-5_3. 278–282, May 2020, doi: 10.1038/s41586-020-2285-x.

[13] J. M. Beer, A. D. Fisk, and W. A. Rogers, “Toward a Framework for Levels of Robot Autonomy in Human-Robot Interaction,” *J Hum Robot Interact*, vol. 3, no. 2, p. 74, Jun. 2014, doi: 10.5898/jhri.3.2.beer. 49

[14] K. Xu *et al.*, “Show, Attend and Tell: Neural Image Caption Generation with Visual Attention,” Feb. 2015, [Online]. Available: http://arxiv.org/abs/1502.03044

[15] A. Vaswani *et al.*, “Attention Is All You Need,” Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.03762

[16] Y. Bisk *et al.*, “Experience Grounds Language,” Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.10151

[17] M. A. Haleem *et al.*, “Amigo (A Social Robot):Development of a robot hearing system Amigo (A Social Robot):Development of a Humanoid Robot Head.” [Online]. Available: https://www.researchgate.net/publication/354272025

[18] Ikkalebob, “DIY Compact 3D Printed Animatronic Eye Mechanism.” [Online]. Available: https://amzn.to/374seRU

[19] D. Sati, S. Avkirkar, R. Pandey, and A. Somnathe, “Human Following Robot Using Arduino,” *International Journal of Advanced Research in Science, Communication and Technology*, pp. 347–350, Apr. 2021, doi: 10.48175/ijarsct- 1025.

[20] J. A. Rojas-Quintero and M. C. Rodriguez-Linan, “A literature review of sensor heads for humanoid robots,” *Rob Auton Syst*, vol. 143, Sep. 2021, doi:10.1016/j.robot.2021.103834.

[21] https://www.techbriefs.com/component/content/article/tb/pub/briefs/machinery-and-automation/47126

[22] https://towardsdatascience.com/human-like-machine-hearing-with-ai-1-3-a5713af6e2f8

[23] https://www.scientificamerican.com/article/deep-learning-networks-rivalhuman-vision1/

[24] https://en.wikipedia.org/wiki/Human%E2%80%93robot_interaction

[25] https://huggingface.co/

[26] https://www.arduino.cc/

[27] https://openai.com/blog/openai-api

[28] https://pyfirmata.readthedocs.io/en/latest/

[29] https://github.com/MUHAMMED-RIZWAN-P-S/Humanoid_Robotic_Head

Authors

First Author – Muhammed Rizwan P. S., B.Tech. in Computer Science and Engineering, HMRITM, work.rizwan912@gmail.com.
Second Author – Sahil Rao, B.Tech. in Computer Science and Engineering, HMRITM, raosahil2719024@gmail.com.
Third Author – Sonal Balayan, B.Tech. in Computer Science and Engineering, HMRITM, Sonalbalayan2000@gmail.com.
Fourth Author – Neeraj Bhandari, B.Tech. in Computer Science and Engineering, HMRITM, bhandarin007@gmail.com.
Fifth Author – Gyanender Kumar, Assistant Professor, Department of Computer Science and Engineering, HMRITM, email