iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** II **Month of publication:** February 2024

DOI: https://doi.org/10.22214/ijraset.2024.58591

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

Artificial Intelligence Based Virtual Assistant for Vision Impaired Person

Navneet Kumar, Karishma Verma, Er. Asim Ahmad
[1, 2]*Computer Science and Engineering, SRMCEM, Lucknow, India*
[3]*(Assistant Professor) Computer Science and Engineering SRMCEM, Lucknow, India*

Abstract: This project is a brilliant example of how technology can be harnessed for a profoundly positive purpose, enhancing the quality of life for those who may face challenges in certain aspects of daily living. It's about creating a tool that complements and supports individuals rather than substituting human interaction or assistance.
That sounds like an incredible initiative! Combining technologies like object detection through YOLO algorithms, speech synthesis via text-to-speech, and incorporating them into smart glasses could significantly enhance the independence and mobility of visually impaired individuals. These glasses could provide real-time information about their surroundings, enabling them to navigate and interact more confidently with their environment.
Keywords: Visually impaired; Voice control; automate website; blind people

I. INTRODUCTION

About 285 million people worldwide have vision impairments. Even though technology has advanced a lot, it's still hard for people with disabilities, especially when using the internet. Nowadays, lots of things like shopping, ordering food, or booking train tickets happen online through websites. But using these websites can be really tough for visually impaired folks.

Visually impaired people need assistance in various day to day activities.Our project is like a smart helper for people who can't see well. It uses different cool things to understand what's around them. First, there's a voice assistant that talks to them and tells them about the stuff nearby. Then, it has a camera that spots and recognizes things, like naming objects. And to measure how far things are, we're trying to use ultrasonic sensors. The camera part is pretty neat—it uses this fancy algorithm called YOLOv3 to figure out what things are in real time. So, when someone uses this virtual assistant, it helps them know what's in front of them, like saying, "Hey, there's a chair in front of you," or "That's a bookshelf over there!" It's all about making it easier for them to understand what's going on around them. Web accessibility barriers can indeed pose significant challenges for visually impaired individuals. Many websites lack proper features that assistive technologies rely on, such as screen readers or magnifiers. This makes it difficult for them to navigate, access information, or perform tasks online independently.

Creating websites with accessibility features in mind is crucial. It involves implementing elements like alternative text for images, keyboard navigation, clear and descriptive headings, and ensuring compatibility with screen readers. Such measures can drastically improve the online experience for visually impaired users.

II. LITERATURE REVIEW

Virtual assistant that can do a bunch of tasks like reading emails, keeping a diary, and giving weather forecasts. It uses a Raspberry Pi with a voice hat, along with Google's text- to-speech and speech-to-text modules.

1) There have been various studies and opinions about how accessible the web is for people with disabilities. Some researchers explored technologies like Braille systems and screen magnifiers but faced technical challenges along the way.
2) Pilling et al. investigated whether the web helps disabled individuals perform tasks they couldn't do before or if it leads to more exclusion. access lack clear research.
3) Sinks and Kings pointed out that reasons behind disabled people's limited web access lack clear research. Meanwhile, Muller et al. highlighted economic and technical limitations as the primary barriers to accessibility.
4) Kirsty et al. highlighted issues with HTML code and PDFs, causing obstacles for the visually impaired online, despite the World Wide Web Consortium (W3C) providing guidelines for accessibility.
5) Power et al. found that only about half of the problems encountered by users were addressed by the Website Accessibility Guidelines (WCAG 2.0), and a small percentage of websites followed the recommended techniques without fully resolving the issues. These studies collectively emphasize the challenges and gaps in web accessibility for individuals with disabilities.

©IJRASET: All Rights are Reserved

6) Weal A. Ezat worked on training a smart model using YOLOv3 to detect objects. They got a pretty good accuracy of around 80%. However, it can only recognize specific objects that it was trained on.

III. PROBLEM STATEMENT

We want to create a system that helps blind people by recognizing objects around them. When the system spots an object, it will not only identify what it is but also tell the person about it using audio. Additionally, it will provide information on how far away the object is from the blind person. This way, they can understand their surroundings better and navigate safely.

A. Project Objective

The goal of our project is to support blind individuals in their daily tasks. We're working on a system that can recognize objects in front of them and describe these objects aloud. It can also read QR codes to provide more details about things like products or items. Additionally, it helps by giving information about nearby hospitals. And, to make things easier, we're designing it to allow the users to give input and control all these helpful features.

B. System Overview and Diagram

The system has a main menu that starts up when you use the software, along with modules on a website. When you interact with the system, you talk to it, and it talks back using speech recognition and text-to-speech technologies from Google and Python. The modules, written in Python, use tools like Selenium for automation and Beautiful Soup for gathering web content.

Each module has customized code that handles its features. For example, the Wikipedia module helps with questions, summaries, and reading articles using a special model trained on a dataset. These modules work together through APIs written in Flask. And importantly, this software can work on any operating system, making it easy for anyone to use hassle-free.

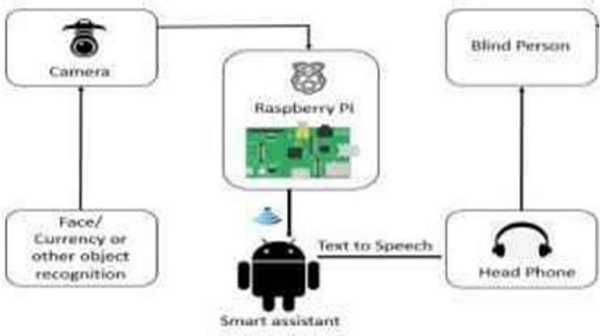
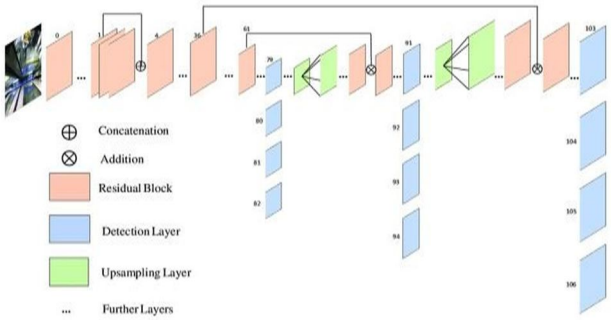


Figure 1: Diagram Of Visualization

C. Object Detection Module

We used a smart algorithm called YOLO that spots and identifies different things in a picture, all in real-time. Instead of breaking it down step by step, YOLO looks at the whole image once and figures out what's in it with just one go. It's like having two main parts: one part looks closely at the picture to understand its features, and the other part focuses on finding and marking the objects by putting boxes around them.

This way, it can quickly tell us what's in the image without needing to check multiple times.



IV. SYSTEM ARCHITECTURE

This system follows a typical setup where there's a main menu that starts when the software begins and separate website modules. The user interacts with these modules through a speech-to-text interface.

This means they talk to the software, and it understands their speech using a tool called Google's Speech Recognition in Python. The modules themselves aren't stored on the user's device but are accessed through a server, so when the user interacts with them, it's like they're talking to the server and getting responses back.

Figure 3: system Architecture

V. METHODOLOGY

When someone starts up their computer or laptop, they interact with the main menu of our software. They can do this by talking to an integrated voice assistant like Siri or by using a specific keyboard shortcut. The main menu shows different options like installed website modules, audio speed, and accent preferences.

Each website module has tools for speech-to-text and text-to-speech, along with a special script in Python that handles the website's features automatically. To help users speak more efficiently, the system gives a beep as a signal for when they can start talking. Whatever the user says gets recognized by the system and played back to the user for confirmation, reducing errors and letting them edit their input if needed.

Google, G mail, Wikipedia - and the main menu. They're designed to make it easy for users to interact with these websites using speech and customize their audio.

Each of the website modules contains a speech- to-text and text-to-speech bundle, a python script that automates the website and the features specific to the website

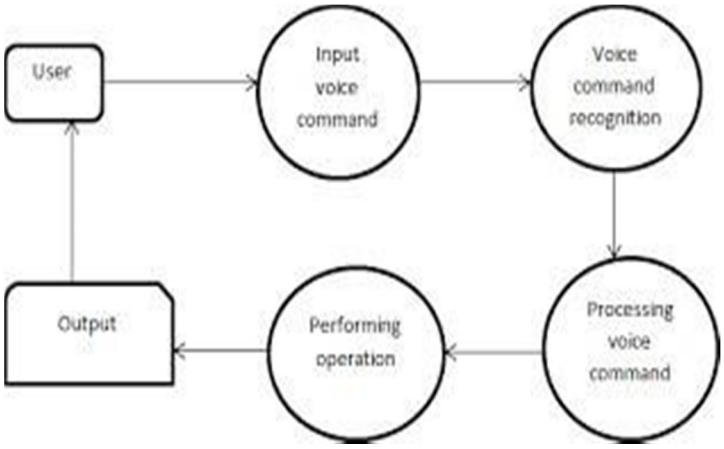


Figure 4: DFD (Data flow diagram)

A. *Minnesota State Chatbot System*

This section outlines the steps followed to support the Software Development Life Cycle (SDLC) [22], focusing on the Minnesota State Chatbot System. We begin by detailing the Requirement Engineering process, highlighting key system requirements. Subsequently, we provide an overview of the system design and architecture, concluding with an outline of the system security analysis.

VI. IMPLEMENTATION

Our virtual assistant is a cool system that has a Raspberry Pi camera and a text-to-speech tool. This camera captures real-time images, and we trained it using a dataset called COCO. When it sees something, it uses a smart system called YOLOv3 to figure out what it is and draws a box around it. We've added Google's text-to-speech tool (GTTS) to read any text it finds in the image.
This is really helpful for visually impaired folks. Plus, we've put in ultrasonic sensors to guess how far away the objects are from the user. So, when someone gives an image to the system, it identifies objects, measures distance, and then tells the user about it, all through audio. It's like having a smart assistant that describes what's in the image!

A. *DialogFlow*

DialogFlow is a powerful tool supporting Natural Language Processing (NLP) to detect keywords and intents within a user's sentence, facilitating the development of chatbots through Machine Learning algorithms.

1) **Agent:** DialogFlow enables the creation of Natural Language Understanding (NLU) modules known as agents, essentially the face of the bot. These agents connect to the backend and provide the necessary business logic.
2) **Intent:** Agents consist of intents, representing actions a user can perform. Intents map user statements to the appropriate actions, serving as entry points into a conversation. Users might express the same request in various ways, but all variations ultimately resolve to a single intent. For instance, "Where is Century College?" and "What does Dr. Baani teach?" could map to specific intents. Multiple intents can be created, and they can be correlated using contexts. Each intent determines which API to call, with what parameters, and how to respond to the user's request.
3) **Entity:** Entities play a crucial role in extracting values from user input. Any information vital to business logic becomes an entity, encompassing concepts like dates, distance, currency, etc. DialogFlow provides system entities for basic elements like numbers and dates, while developers can define custom entities tailored to specific needs.
4) **Context:** Context is integral to making the bot conversational. A context-aware bot can remember information and engage in a dialogue similar to humans. For example, in a conversation about course registration prerequisites, understanding the context is crucial. A context-aware bot can comprehend follow-up questions without explicit specification of the context in each user input.
5) **WebHook:** To implement complex actions, DialogFlow employs webhooks. Whenever DialogFlow matches an intent, it can send a request to a designated endpoint, typically coded to execute a customized process. This allows the chatbot to perform more intricate tasks beyond simple intent matching.

The Minnesota State Chatbot system utilizes DialogFlow in its software design and architecture, leveraging its capabilities for Natural Language Processing and intent recognition. This integration enhances the chatbot's ability to understand and respond to user queries effectively.

Technologies backing AI Virtual Assistants

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

VII. RESULT

The speech-to-text and text-to-speech tools we used in Python worked great, with a 96.25% accuracy in recognizing words. We tested this by using different voice samples and found it to be quite reliable.

For the Wikipedia module, we used a special BERT model on the SQuAD dataset to answer questions. It gave us accurate answers with an 80.88% Exact Match accuracy and an F1 score of 88.49%.

When we tried our software on popular sites like Google, G mail, and Wikipedia, it performed well. It could send emails based on user commands, give accurate answers from Wikipedia, and summarize text effectively. This means we've created a software that can make these websites easily accessible and efficient for people who are visually impaired.

Results showed that we were able to run our software on the three most popular sites: Google, Gmail and Wikipedia. The software was run on each of them separately. The software could send an email effectively using the commands from the user.

VIII. APPLICATIONS

The Virtual Assistant now has a cool feature! Instead of reading the whole text, users can ask a question, and the software, using machine learning, finds the answer from the text itself. It also offers a summary, so users don't need to read everything, making website access much simpler. With machine learning and speech-to-text, accessing websites, once tough, is now super easy, quick, and efficient. We think these virtual assistants for visually impaired users mark the start of what's coming in Web 3.0.

IX. CONCLUSION

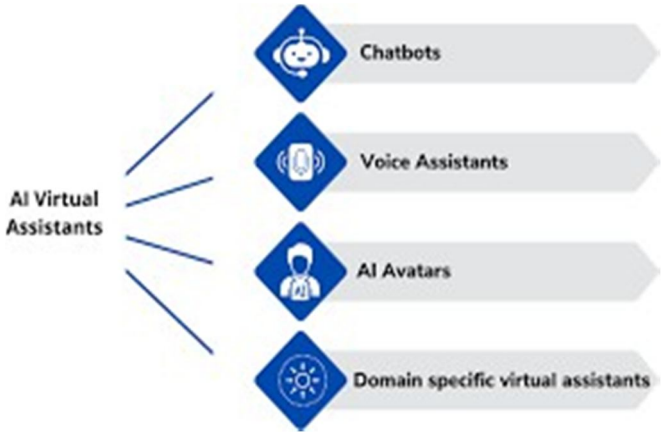
This paper introduces a helpful system designed to assist visually impaired individuals in their daily tasks. Our virtual assistant uses YOLOv3 to detect objects and is currently made up of four modules: object recognition, text recognition, distance estimation, and text-to-speech. In the future, our aim is to further enhance this virtual assistant to help people with visual impairments by identifying objects ahead of them. It will provide an audio output, using the text-to- speech module, that not only identifies the objects but also estimates the distance between the person and the object. This way, it can guide them more effectively in navigating their surroundings. The virtual assistant is a user-friendly way for visually impaired individuals to navigate websites. It removes the necessity of remembering complicated keyboard shortcuts or relying on screen readers. It's not just a convenient way to interact with websites; it's also efficient in doing so. This software acts as a bridge toward Web 3.0, where voice commands will be the main way of operating everything online

X. FUTURE SCOPE

Currently, our app understands commands only in English. But our goal is to broaden its reach by adding support for many other commonly used languages. This way, people worldwide can easily access the web without any language barriers.

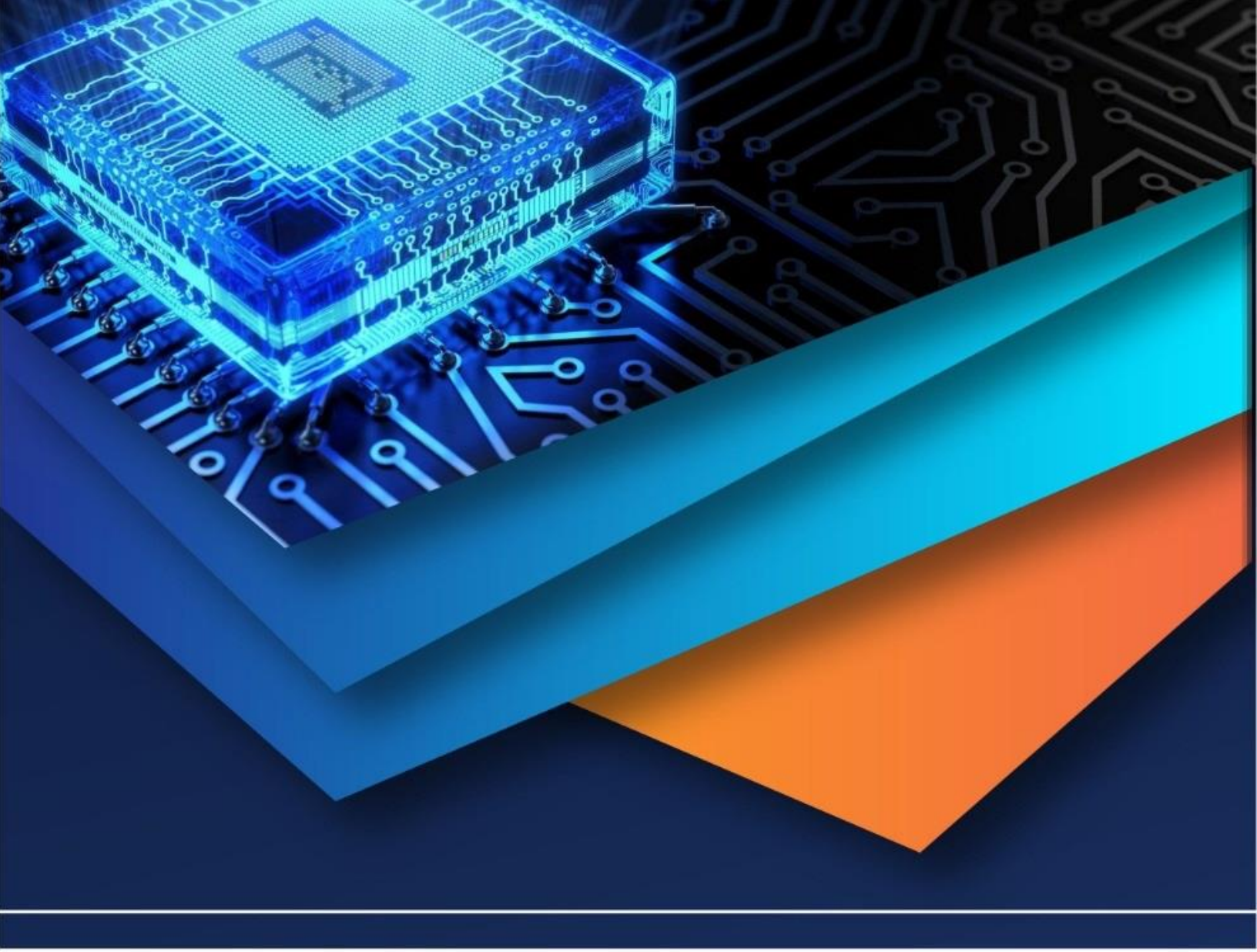
Additionally, we're working on a system that can fit into any website and transform into a browser extension. This will allow users, especially those with visual impairments, to switch between regular browsing and our specialized mode effortlessly. This will be especially helpful for educational websites, ensuring that visually impaired individuals can access online courses just as easily as everyone else.

©IJRASET: All Rights are Reserved



REFERENCES

[1] Pilling, D., Barrett, P. and Floyd, M. (2004). Disabled people and the Internet: experiences, barriers and opportunities. York, UK: Joseph Rowntree Foundation, unpublished.
[2] Porter, P. (1997) 'The reading washing machine', Vine, Vol. 106, pp. 34– 7
[3] JAWS -https://www.freedomscientific.com/products/sof tware/jaws/ accessed in April 2020
[4] Ferati, Mexhid & Vogel, Bahtijar & Kurti, Arianit & Raufi, Bujar & Astals, David. (2016). Web accessibility for visually impaired people: requirements and design issues. 9312. 79-96. 10.1007/978-3-319-45916- 5_6.
[5] Power, C., Freire, A.P., Petrie, H., Swallow, D.: Guidelines are only half of the story:accessibility problems encountered by blind users on the web. In: CHI 2012, Austin, Texas USA, 5–10 May 2012, pp. 1–10 (2012)
[6] Sinks, S., & King, J. (1998). Adults with disabilities: Perceived barriers that prevent Internet access. Paper presented at the CSUN 1998 Conference, Los Angeles, March. Retrieved January 24, 2000 from the World Wide Web
[7] Muller, M. J., Wharton, C., McIver, W. J. (Jr.), & Laux, L. (1997). Toward an HCI research and practice agenda based on human needs and social responsibility. Conference on Human Actors in Computing Systems. Atlanta, Georgia, 22–27 March.
[8] Kirsty Williamson, Steve Wright, Don Schauder, Amanda Bow, The internet for the blind and visually impaired, Journal of ComputerMediated Communication, Volume 7, Issue 1, 1 October 2001, JCMC712
[9] Deeppavlov documentation http://docs.deeppavlov.ai/en/master/features/mo dels/squad.html accessed in April 2020
[10] The website for American foundation for the blind https://www.afb.org/about- afb/what-we-do/afb-consulting/afbaccessibility- resources/challenges-web-accessibility accessed in April 2020
[11] Ryle Zhou, Question answering models for SQuAD 2.0, Stanford University, unpublished.
[12] Global data on visual impairments 2010 by World Health Organisation (WHO)https://www.who.int/blinness/GLOBAL DATAFINALforweb.p df?ua=1



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)