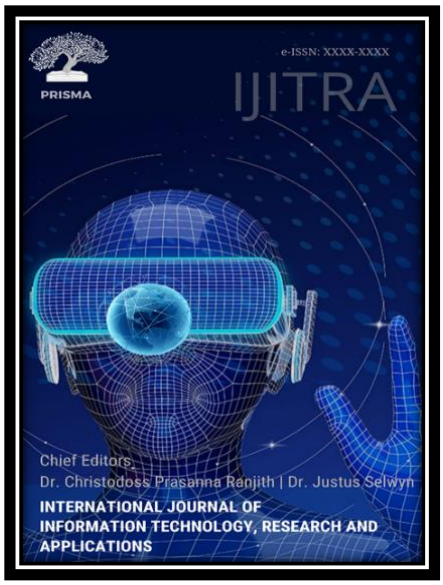

International Journal of Information Technology, Research and Applications (IJITRA)

Siddhant Sarthak, R. Vinayak, (2024). Whisperings From the Ordinary: A Conversational Artificial Intelligence (AI) Approach to Bridging the Human-Object Divide, 3(3), 01-06.

ISSN: 2583-5343

DOI: 10.59461/ijitra.v3i3.89

The online version of this article can be found at:
https://www.ijitra.com/index.php/ijitra/issue/archive

Published by:
PRISMA Publications

International Journal of Information Technology, Research and Applications (IJITRA) is a journal that publishes articles which contribute new theoretical results in all the areas of Computer Science, Communication Network and Information Technology. Research paper and articles on Big Data, Machine Learning, IOT, Blockchain, Network Security, Optical Integrated Circuits, and Artificial Intelligence are in prime position.

PRISMA
PUBLICATIONS

https://www.prismapublications.com/

Whisperings From the Ordinary: A Conversational Artificial Intelligence (AI) Approach to Bridging the Human-Object Divide

Siddhant Sarthak[1], R. Vinayak[2]
[1]Reverie (Reliance Jio Infocom), Bengaluru, Karnataka, India
[2]TEKsystems Global Services Pvt. Ltd, Kaikondrahalli, Bengaluru, India
[*]Corresponding Author: Siddhant Sarthak, Email: siddhant.sarthak06@gmail.com

Article Info

Article history:

Received March 19, 2024
Revised June 10, 2024
Accepted August 14, 2024

Keywords:

AIoT
Sustainability
Human-Object Interaction
Object Detection
Conversational AI
Smart Homes
Ethical Considerations

ABSTRACT

In a world increasingly dominated by connected devices, one realm remains silent: most inanimate objects that surround us. This stark contrast highlights a critical gap in human-environment interaction, one that the present work aims to bridge through the power of intelligent conversation and shared understanding. Traditional IoT methods, while efficient, raise crucial concerns regarding sustainability. Embedding electronics into every object risk exceeding costs and generating substantial e-waste. This necessitates exploration of alternative paradigms that unlock the communicative potential of everyday items without environmental burden. Enter the confluence of object detection and conversational AI. Object detection, encompassing techniques like image classification, empowers machines to perceive their surroundings and make informed decisions based on identified objects. Conversational AI refers to the ability of computers to engage in a natural, human-like dialogue. By fitting in object detection, conversational AI can transform lifeless objects into active and intelligent participants in our lives, fostering a richer relationship with the spaces we inhabit. This paper proposes an AI-powered mobile application that unlocks the communicative potential of everyday items, transforming them from passive elements into active participants in our lives. We are exploring to enhance domestic routines by streamlining daily tasks with object-initiated reminders, proactive maintenance suggestions, and personalized advice. This research delves into the technical aspects, potential applications, and ethical implications of granting voice to the voiceless.

Corresponding Author:

Siddhant Sarthak
Reverie (Reliance Jio Infocom)
Bengaluru
India
Email: siddhant.sarthak06@gmail.com

1. INTRODUCTION

In today's world, we are surrounded by an ever-growing number of connected devices, from smart speakers to wearable technology. The life force of these connected devices, also known as the Internet of Things (IoT), has revolutionized the way we interact with our environment, providing us with convenience, efficiency, and endless possibilities. These interconnected devices, equipped with sensors and advanced analytics capabilities, play a crucial role in transforming traditional buildings into intelligent, efficient, and sustainable spaces [1]. However, amidst this digital transformation, we are becoming more unsustainable as traditional object communication approaches often rely on embedding electronics into individual objects. The production of electronic equipment is causing potentially unbalanced waste of limited metals and resources in general, which could become a critical issue in the long run [2]. There exists a vast majority of inanimate objects that remain silent, unable to participate in the human-environment interaction. This stark contrast highlights a critical gap in our relationship with the spaces we inhabit. Traditional IoT can be highly

accurate and responsive, it raises concerns about sustainability, cost, and the potential environmental impact of manufacturing and powering countless chips. But what if we could bridge this gap and unlock the communicative potential of everyday items in a sustainable way? This is where object detection and conversational AI comes in.

Object detection is an umbrella term that encompasses various techniques, such as image classification, object localization, and object tracking. This allows machines to understand their surroundings and make informed decisions based on the objects present.

Conversational AI is the integration of natural language processing and context awareness, enabling computers to engage in a natural, human-like dialogue.

With the combination of both the aforementioned technologies, we can harness the ability to transform lifeless objects into active and intelligent participants in our lives. This paper proposes an AI-powered mobile application that aims to fill the gap in human-environment interaction by recognizing and giving a voice to the voiceless objects that surround us sustainably.

2. LITERATURE REVIEW

An increased implementation of IoT technologies would lead to a more intense utilization of fossil technologies to ensure the necessary energy supply for IoT production lines [2]. Unfortunately, the current electronics waste recycling rate of 20% [3] remains concerning, it also presents a significant question to produce more IoT products given the rise in the market demands. Recognizing the impracticality, unsustainability, and economic limitations of embedding chips in every everyday object, the study explores alternative solutions within AI to simulate the object and its capabilities offering a potentially more sustainable and cost-effective solution.

Today, it is estimated that more than 5 billion people have mobile devices, and over half of these connections are smartphones [4]. Hence, to bridge the human-object divide sustainably, we propose utilizing smartphones as a readily available platform. A dedicated application, developed using the cross-platform mobile frameworks like Flutter which will allow developers to build apps for multiple platforms from a single codebase, reducing costs [5] making it an ideal choice for maximizing accessibility and reach. The app will be communicating with YOLOv7 series of object detection systems, which receives the state-of-the-art results [6]. YOLOv7, will analyze the image and identify individual objects within the scene. The identified objects will be matched with a dictionary of known objects and their characters in a NoSQL database like Firebase's Firestore. Apart from offering database services, Firebase offers seamless integration with various Google services like Google Analytics, Google Ads, and Google Cloud. This allows developers to leverage a range of powerful tools for app marketing, analytics, and advanced cloud operations. Such integration also simplifies the task of managing services from a single platform. [7].

Upon successful object identification, the conversational transformer model, BART, takes the lead in the interaction. This powerful model is an agent capable of conversing about any domain [8]. Based on the knowledge from the database, it will assume the role of the identified object, serving as a bridge between the user and the physical world.

3. METHODOLOGIES

The user interface for this application will be implemented on a mobile platform using the Flutter framework, allowing deployment on both Android and iOS devices from a single codebase [5]. The interaction begins with the user scanning the room via their phone camera. This scan captures a panoramic image, providing a comprehensive view of the surrounding environment. (Figure 1)

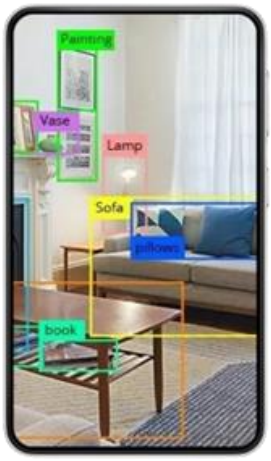

Figure 1. Overview of Scanning Objects

The panoramic image serves as the primary input for the object detection and conversational generation modules. Object detection models such as YOLOv7 which is considered superior to traditional Convolutional Neural Networks (CNNs) [9], will analyze the image and identify individual objects within the scene.

Once objects are identified, their specific attributes and potential functionalities are determined based on their type from the Firebase's Firestore. This information fuels the conversational generation module, potentially leveraging transformer-based models like BART [8], to formulate personalized and context-aware responses for each object. These responses can range from offering information about the object to suggesting actions based on the current state. The chosen language models must demonstrate proficiency in generating natural and engaging dialogue tailored to the specific object and its context. The user can then be directly addressed by the detected object (if allowed), This real-time dialogue will create a dynamic experience. (Figure 2)

Figure 2. Overview of Object's Chat with the User

By combining effective object detection and conversational generation, this user interaction design enables real-time communication with everyday objects within the user's immediate environment. Object recognition would allow the app to identify and classify everyday objects, such as books, utensils, and clothing. Natural language processing would enable the application to understand and respond to human speech, while context awareness would allow it to consider the surrounding environment and tailor its responses accordingly. (Figure 3)

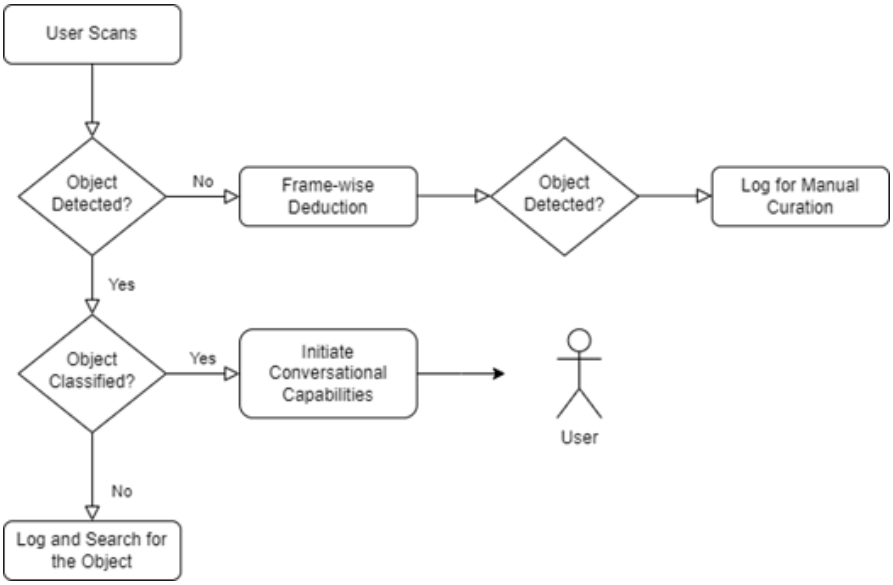

Figure 3. Overview of the Flow

4. EXPECTED FINDINGS AND DISCUSSION

We hypothesize that the proposed technology is quite feasible. Our chosen methodology, includes the YOLOv7's state of the art performance [6], we expect to achieve a good object detection accuracy on good lighting conditions (Table 1). The model has quite a significant improvement when compared to previous versions.

Table 1. Comparison of baseline object detectors

Model	#Param.	FLOPs	Size	AP^{val}	AP^{val}_{50}	AP^{val}_{75}	AP^{val}_{S}	AP^{val}_{M}	AP^{val}_{L}
YOLOv4 [2]	64.4M	142.8G	640	49.7%	68.2%	54.3%	32.9%	54.8%	63.7%
YOLOR-u5 (r6.1) [79]	46.5M	109.1G	640	50.2%	68.7%	54.6%	33.2%	55.5%	63.7%
YOLOv4-CSP [76]	52.9M	120.4G	640	50.3%	68.6%	54.9%	34.2%	55.6%	65.1%
YOLOR-CSP [79]	52.9M	120.4G	640	50.8%	69.5%	55.3%	33.7%	56.0%	65.4%
YOLOv7	36.9M	104.7G	640	**51.2%**	**69.7%**	**55.5%**	**35.2%**	**56.0%**	**66.7%**
improvement	-43%	-15%	-	+0.4	+0.2	+0.2	+1.5	=	+1.3
YOLOR-CSP-X [79]	96.9M	226.8G	640	52.7%	**71.3%**	57.4%	36.3%	57.5%	68.3%
YOLOv7-X	71.3M	189.9G	640	**52.9%**	71.1%	**57.5%**	**36.9%**	**57.7%**	**68.6%**
improvement	-36%	-19%	-	+0.2	-0.2	+0.1	+0.6	+0.2	+0.3
YOLOv4-tiny [76]	6.1	6.9	416	24.9%	42.1%	25.7%	8.7%	28.4%	39.2%
YOLOv7-tiny	6.2	5.8	416	**35.2%**	**52.8%**	**37.3%**	**15.7%**	**38.0%**	**53.4%**
improvement	+2%	-19%	-	+10.3	+10.7	+11.6	+7.0	+9.6	+14.2
YOLOv4-tiny-3l [76]	8.7	5.2	320	30.8%	47.3%	32.2%	**10.9%**	31.9%	51.5%
YOLOv7-tiny	6.2	3.5	320	**30.8%**	**47.3%**	**32.2%**	10.0%	**31.9%**	**52.2%**
improvement	-39%	-49%	-	=	=	=	-0.9	=	+0.7
YOLOR-E6 [79]	115.8M	683.2G	1280	55.7%	73.2%	60.7%	40.1%	**60.4%**	69.2%
YOLOv7-E6	97.2M	515.2G	1280	**55.9%**	**73.5%**	**61.1%**	**40.6%**	60.3%	**70.0%**
improvement	-19%	-33%	-	+0.2	+0.3	+0.4	+0.5	-0.1	+0.8
YOLOR-D6 [79]	151.7M	935.6G	1280	56.1%	73.9%	61.2%	**42.4%**	60.5%	69.9%
YOLOv7-D6	154.7M	806.8G	1280	56.3%	73.8%	61.4%	41.3%	60.6%	70.1%
YOLOv7-E6E	151.7M	843.2G	1280	**56.8%**	**74.4%**	**62.1%**	40.8%	**62.1%**	**70.6%**
improvement	=	-11%	-	+0.7	+0.5	+0.9	-1.6	+1.6	+0.7

Leveraging the capabilities of Flutter and Firebase, we anticipate seamless hosting of models and communication between users and identified objects through the application. We expect users to be able to initiate and maintain natural conversations with everyday objects, accessing their functionalities or receiving information readily.

BART retrieves the right memory passage 26% of the time on the seen test set as shown in Figure 4 [8], we hypothesize that users can be recommended best practices, reminded about different actions, or receive feedback specific to the object.

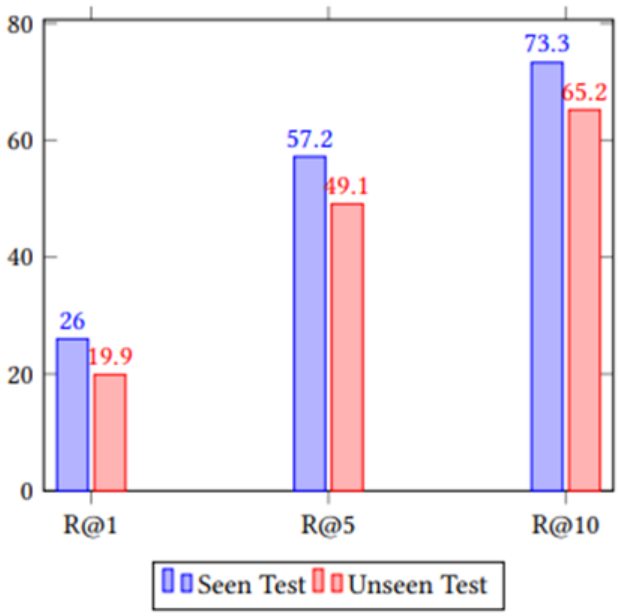

Figure 4. BART's Recall Metrics on Seen and Unseen Datasets

4.1. POTENTIAL APPLICATIONS

The proposed AI-powered mobile application could attract a wide range of stakeholders, including consumers, manufacturers, and retailers. It has the potential to enhance domestic routines by streamlining daily tasks with object-initiated reminders, proactive maintenance suggestions, and personalized advice. For example, a cluttered drawer could suggest organization strategies, and a plant could set a routine to remind the user to water it. This could help individuals stay organized and on top of their daily responsibilities. With the use of object usage data and insights, manufacturers can recommend personalized products to consumers, increasing customer satisfaction and loyalty. Retailers could also benefit from this technology by providing targeted advertisements and promotions based on the data.

Beyond efficiency and convenience, this technology eliminates the need for embedded chips in every object, which minimizes e-waste. It will also encourage responsible resource usage.

4.2. ETHICAL CONSIDERATIONS

While the concept of breathing a life into everyday object may seem futuristic and exciting, there would be challenges for widespread user acceptance. The idea of talking to everyday objects might seem strange and even uncomfortable to some people. Another major concern is the potential invasion of privacy. When AI collects personal data, it is essential to ensure that the collection, use, and processing of such data is done in compliance with the GDPR. AI algorithms should be designed to minimize the collection and processing of personal data and ensure that the data is kept secure and confidential [9].

However, with the increasing prevalence of voice assistants and smart devices, people are becoming more accustomed to conversing via technology. Additionally, the potential benefits of this application, such as increased efficiency and convenience, may encourage users to overcome their initial reservations.

5. CONCLUSION

The proposed AI-powered mobile application has the potential to transform the way we interact with our surroundings. By enabling conversations with everyday objects, we can foster a richer and more meaningful connection with our environment. Though, this technology also poses ethical challenges that must be addressed, the chip-free approach presents a more futuristic and sustainable path towards granting voice to the voiceless. With responsible development and usage, conversational AI can enhance our domestic routines, encourage responsible resource usage, and benefit various stakeholders. This paves the way for widespread adoption and fostering a richer connection with our physical surroundings in the longer run and contribute to a more sustainable future.

REFERENCES

[1] Gupta, Manish. "IoT Devices: Empowering Smart & Sustainable Buildings." Zenatix, May 26, 2023. https://www.zenatix.com/iot-devices-empowering-smart-sustainable-buildings/.

[2] Nižetić, Sandro, Petar Šolić, Diego López-de-Ipiña González-de-Artaza, and Luigi Patrono. "Internet of Things (IoT): Opportunities, Issues and Challenges towards a Smart and Sustainable Future." Journal of Cleaner Production 274 (November 20, 2020): 122877. https://doi.org/10.1016/j.jclepro.2020.122877.

[3] Web source: Rick Leblanc. Thebalancesmb.com/e-waste-recycling-facts-and-figures-2878189, (accessed, March 25, 2020).

[4] Silver, Laura. "Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally." Pew Research Center's Global Attitudes Project (blog), February 5, 2019. https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/.

[5] Web source: Tuhin Bhatt. "What Is Flutter? And What Is Flutter Development?" (accessed January 3, 2024). https://www.intelivita.com/resources/what-is-flutter/

[6] Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors." In 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7464–75. Vancouver, BC, Canada: IEEE, 2023. https://doi.org/10.1109/CVPR52729.2023.00721.

[7] Raj Sanghvi. "Why Choose Firebase for Your App Development in 2023," August 7, 2023. https://www.bitcot.com/why-choose-firebase-for-app-development/.

[8] De Bruyn, Maxime & Lotfi, Ehsan & Buhmann, Jeska & Daelemans, Walter. (2020). BART for Knowledge Grounded Conversations.

[9] Web source. "AI and Privacy: The privacy concerns surrounding AI, its potential impact on personal data," The Economic Times, Apr. 25, 2023. (accessed January 4, 2024) https://economictimes.indiatimes.com/news/how-to/ai-and-privacy-the-privacy-concerns-surrounding-ai-its-potential-impact-on-personal-data/articleshow/99738234.cms

BIOGRAPHIES OF AUTHORS

Siddhant Sarthak received the bachelor's degree in computer science and engineering from Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, Odisha in 2023. He demonstrates a history of ideation, tinkering, and leadership suggests a proactive approach to tackling challenges and contributing to the society. Siddhant possesses three years of diverse technical experience working in different startups, fostering a strong aptitude for problem-solving and innovation. With a proven track record of success in hackathons and competitions that solved real-life challenges, evidence indicates a talent for developing and executing ideas from conception. He is currently an R&D Software Engineer at Reverie (Reliance Jio Infocom). For any inquiries or collaboration opportunities he can be contacted at email: siddhant.sarthak06@gmail.com

R. Vinayak earned his bachelor's degree in computer science Engineer from Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, Odisha in 2023. He is currently working as a Data Analyst at TEKsystems Global Services. His research interests is focused on Machine Learning, and Big Data analytics. Vinayak is actively engaged in exploring and contributing to these fields. For any inquiries or collaboration opportunities, he can be reached via email at vinayakrao7784@gmail.com.