Enabling Autonomy Through Voice Control: AI-Assisted Mobile Platform

Ivelina Balcheva(✉) and Yassen Gorbounov

New Bulgarian University, Sofia, Bulgaria
ivab1414@gmail.com, ygorbounov@nbu.bg

Abstract. The rapid development of the technological sector and the need to automate a wide range of human-aiding activities defines the development and research of autonomous mobile platforms as an increasingly urgent problem. At the same time, artificial intelligence plays the role of an innovative catalyst that increases the functionality, efficiency and convenience of mobile platforms, making them more intelligent and adaptive. The present paper focuses on the study of an autonomous mobile platform that uses voice control and artificial intelligence techniques for effective object recognition in the environment. This platform aims to improve the functionality of robotic systems, providing a more intuitive way to communicate and control. The research combines theoretical analysis, simulation modeling, software algorithm development, and experimental testing. The development process includes studying existing voice recognition and artificial intelligence methods, as well as their applications in the field of autonomous systems. Improving human-machine communication and decision-making processes facilitates the implementation of voice-assisted control. The proposed approach can find application in the improvement of robotic home assistants, in the control of mobile robots in dangerous or hard-to-reach environments, in the construction of remote laboratories, as well as in a number of other fields of technology.

Keywords: Autonomous Platform · Artificial Intelligence · Human-Machine Interface · Voice Control · Machine Learning

1 Introduction

The abundance of information that surrounds us, combined with the rapid development of modern technology, predetermines the growing need to automate human activities. This affects both industrial production and everyday work. The application of advanced proprioceptive and exteroceptive sensor technologies [11, 14] is an increasingly relevant direction in the field of robotics and autonomous navigation, which improves the safety of decision-making in dynamic environments. There is a growing interest in intelligent systems capable of interpreting commands in natural language such as Samsung SmartThings, Apple HomeKit, Amazon Alexa, etc. [4, 8], applicable primarily for home use. A promising approach on a larger scale is the combination of autonomous mobile platforms with artificial intelligence (AI) and voice control systems [1, 15]. This, along

T. Zlateva and G. Tuparov (Eds.): CSECS 2024, LNICST 609, pp. 50–61, 2025.
https://doi.org/10.1007/978-3-031-84312-9_3

with the addition of surveillance tools such as cameras and LiDAR (Light Detection and Ranging) laser scanners, provides rich opportunities for environmental perception. For example, in [13] the LOAM (Lidar Odometry and Mapping) method [26] is discussed. A mobile robot with a 16-row LiDAR installed on it is considered for moving in densely populated areas. In [21], the authors go further with improving the accuracy in robot position estimation by applying Rao Blackwell Particle Filter (RBPF-SLAM) and Extended Kalman Filter (EKF). In [24], the same algorithm, but without the EKF, is used to improve localization and mapping accuracy. The authors of [10] have built a robot guide for the blind under the control of an RTOS (Real-Time Operating System) using a LiDAR sensor and voice control. A LiDAR-based system assisted by voice control is proposed as well in [3]. The article claims that in Level 5 autonomous vehicles, human intervention with the car will be reduced to the use of voice commands. In [9], a focus is placed on the analysis of the voice control of a collaborative robot by conducting experiments with different voice commands, while testing their repeatability and reliability. In [20], the software capabilities of the voice control system available in ROS (Robotic Operating System) and integrated in the SLAM (Simultaneous Localization and Mapping) algorithm are discussed. A real-time experiment was conducted with the ROSWITHA robot. In [5] and [18], attention is paid to the integration of artificial intelligence systems, including chatbots, to achieve remote and intelligent control of household electrical appliances over the Internet. The authors have applied natural language processing (NLP) techniques, which makes such systems highly adaptable. In [12], a system using advanced algorithms for automatic speech and voice recognition, based on deep neural networks (DNN), was designed and verified in real life scenarios.

The above said paves the way for the ever-expanding future research in the field of enabling autonomy through voice control and motivates the object of consideration in this article. Perfecting home assistants, automating robots control, and increasing safety in dangerous or hard-to-reach environments aim to improve the quality of human life and will remain a topical trend for the years to come. The paper's main goal is to continue the search for opportunities to combine autonomous mobile platforms with voice control systems and add recognition and automatic decision-making capabilities. It discusses a framework for a voice-controlled autonomous mobile platform with integrated camera sensor. It outlines the major communication protocols, sensor fusion algorithms, and machine learning techniques. The result of the proposed development could lead to improved terrain navigation and reliable decision making in complex environments, as well as improved machine response to user commands. A voice-control software is designed and discussed, in which the decision-making system for autonomous mobile platforms equipped with a camera is connected to the Vertex AI API (Application Programming Interface). Voice control uses natural language processing to convert spoken commands into executable code, which could be particularly useful for people with visual impairments or limited mobility. Connecting to the Vertex AI API involves using libraries and authentication methods to access its AI models for pattern recognition, text classification, and object detection. For example, using the Google Cloud SDK allows developers to interact with Google Cloud services and leverage Vertex AI capabilities.

The remainder of this work is organized as follows. Chapter II provides an overview of some major advantages of voice control in the field of robotics and the integration of

AI for decision-making in autonomous systems. A technology comparison is made in Chapter III. Chapter IV describes the core algorithm and methodology of the proposed AI-assisted mobile platform. The conducted experiments are given in the same section. Finally, Section V summarizes the results of the research and concludes the paper.

2 Advantages of Voice Control and AI Integration in Robotics

Voice control in robotics [7, 22] can offer many potential advantages that can enhance the efficiency of robotic systems. One key advantage is simpler communication between humans and machines. Voice control allows robots to respond more quickly in dynamic environments, as they can quickly adjust their actions based on real-time verbal cues from their operators. In this paper, an attempt is made to address and possibly remove the need for complex visual programming or manual control interfaces. Vocal commands could also allow for greater versatility in how robots are deployed and utilized. Machines may be more easily repurposed for different tasks without the need for extensive reconfiguration or retraining if giving the robots the ability to receive verbal instructions. Only limited by the operators' imagination, this flexibility cannot only increase the overall efficiency of robotic systems but it can also reduce downtime and minimize costs associated with reprogramming or replacing the conventional manual controls. Furthermore, by reducing the necessity for direct physical interactions with the machines, voice control can improve worker safety while using robots. That is why delivering commands vocally, allows operators to stay away from potentially hazardous surroundings or situations while still efficiently regulating the machine's actions [19]. In addition to safeguarding human operators, the lower accident risk would increase the longevity of robotic systems by reducing physical manipulation-related wear and tear.

Another significant advantage of implementing voice control in robotics is the potential for increased accessibility [17]. Without the need for specialized training or equipment, people with physical disabilities or restrictions can readily interface with and control robotic systems through voice commands. Therefore, one major aim of this research is trying to widen access and create new opportunities for people who are excluded from enjoying the full range of cutting-edge technology. There are many more benefits to voice control in robots than those mentioned above. Voice-controlled robots can provide a range of advantages that grant greater autonomy and automation in a variety of ways, from enhancing accessibility and safety to enhancing human-machine communication.

In order to enable the level of autonomy that comes with voice-controlled robotics, the integration of artificial intelligence for autonomous decision-making is paramount. The AI algorithms will play a major role in ensuring the robot is able to analyze vast amounts of data quickly and make some of the less important decisions to allow the operator more freedom to focus on the bigger choices. These AI algorithms help the robot adapt to changing environments, identify patterns and predict possible future outcomes accurately. With this ability, the machine will be able to adjust to complex tasks and complete scenarios independently, enhancing their overall efficiency. On the topic of the adaptation to changing environments, AI assisted robots could be able to continually accrue data as they work for their algorithms. This means that every task will become exponentially more efficient. Over time, the additional data and constant adaptation in response to it will allow machines to evolve in their task-solving prowess, requiring even less hands-on intervention.

The reduced need for human intervention and increase in self-sufficiency due to the assistance of AI will also offer significant benefits in terms of safety and risk management. Integrating these AI algorithms allows the machines to assess potential hazards in real-time and communicate this with the operators. As a result, the possibility for accidents that may harm both the human and the robot will decrease. AI algorithms implemented in machines for autonomous decision-making is crucial and extremely useful for the integration of voice control in robotics. From adapting to rough terrain to harm reduction and going as far as to open opportunities for humans to focus their efforts on the pursuit of advancement, this technology has a lot of potential in the future of technological progress.

3 AI Instruments Comparison

A short comparison between GPT, BlackBox AI and Vertex models is given in Table 1. The three models compare in terms of their architecture, training data, use cases, open source availability, customization, scalability, model size, performance and deployment. Each model advantages and limits are briefly outlined and a conclusion of model's suitability is made.

ChatGPT [25] is an extension of the GPT (Generative Pre-trained Transformer) model designed for text generation, language understanding, and interaction with users in natural language. The biggest advantages of this model are its powerful and flexible algorithm from text generation and meaning understanding, ability to provide personalized answers, and the use of voice inputs, which makes it suitable for voice control. However, the GPT model does not support object detection and tracking in photos, videos and live streams. BlackBox AI [16] is a machine learning and artificial intelligence system specialized in image processing and object recognition. This model possesses high accuracy in real-time object recognition, and can be easily integrated with different types of mobile platforms. Its flexibility and scalability make it a good candidate for the purpose of this paper, unfortunately, the lack of optimized voice control and natural language user interaction may be a big problem. Vertex [6] is a comprehensive AI development platform that includes tools for training, optimization, and model deployment. The possibility of creation of custom models for object recognition and voice control, its flexibility to integrate with different types of data and sensors, and opportunity to be optimized for mobile platforms and real time are only part of the advantages. The only possible problem is that it requires more development and configuration time than other models due to the ability to be fine-tuned. The most suitable model for the proposed mobile platform is the Vertex AI model. It provides a combination of flexibility and power, allowing the creation of custom models that combine both voice and high accuracy object recognition.

From the hardware-point of view, the core capabilities of two possible boards, namely the ESP32-CAM and NVIDIA Jetson Nano, compare in Table 2.

Both NVIDIA Jetson Nano [23] and the ESP32-CAM [2] could be good choices for a robot that has to connect to an external AI model, recognize objects, and do voice

Table 1. AI model comparisons.

Feature/model	GPT	BlackBox AI	Vertex
Architecture	Transformer	Proprietary	Custom
Training data	Large corpus of text data from the internet	Internal and external data sources	Custom datasets, potentially from Google Cloud services
Use cases	Natural language understanding and generation, text completion, question answering, language translation	General-purpose AI solutions, potentially with a focus on specific industries or tasks	Customized AI solutions, potentially integrated with Google Could services
Open Source	OpenAI provides pre-trained models but underlying architecture is not fully open source	No	No
Customization	Fine-tuning available for specific tasks or domains	Possible customization but the client needs to provide the datasets and the training may be insufficient	Customization is possible with custom and Google integrated datasets
Scalability	Limited by hardware and resources	Scalability may depend on infrastructure and resources	Scalable using Google Cloud infrastructure
Model size	Large, with billions of parameters	Size may vary based on implementation and data quality	Size may vary based on implementation and data quality
Performance	State-of-the-art in various NLP tasks	Performance may vary based on implementation and data quality	Performance may vary based on implementation and data quality
Deployment	Can be deployed on various platforms	Can be deployed as a service or integrated into specific applications	Can be deployed through Google Cloud services

control. The ESP32-CAM is appropriate for simple jobs and has basic processing capabilities. Its inferior processing capacity limits its ability to do voice control and object recognition. The ESP32-CAM would mostly be responsible for capturing images and sounds, processing them to some amount, and then sending the data to an external AI model for more analysis, since the external AI model will be handling the intense processing. Due to its lower cost, it might be a desirable choice for this project with limited funding. Compared to the ESP32-CAM, the Jetson Nano delivers a lot more processing

Table 2. Comparing core capabilities of ESP32-CAM and NVIDIA Jetson Nano.

Feature/board	ESP32-CAM	NVIDIA Jetson Nano
Processor	ESP32 (dual-core, 32-bit, 240 MHz)	Quad-core ARM Cortex-A57 @ 1.43 GHz
GPU	None	128-core NVIDIA Maxwell GPU
RAM	520 KB SRAM, 448 KB PSRAM	4 GB 64-bit LPDDR4
Camera Interface	OV2640 (2 MP), OV7670 (0.3 MP)	MIPI CSI-2
AI Capabilities	Limited due to lower processing power	High, can run deep learning models efficiently
Object Recognition	Can perform basic object recognition tasks	Excellent support for complex object recognition
Voice Control	Limited due to processing power	Possible but may require external processing
Connectivity	Wi-Fi, Bluetooth	Wi-Fi, Ethernet
Price	Low	High

capability with its quad-core ARM Cortex-A57 processor and graphical processing unit (GPU). Jetson Nano can provide additional processing power for preprocessing the data, controlling inputs and outputs, and possibly executing some lightweight AI algorithms locally, even if the external AI model will handle the heavy computation. Being a fully-fledged Linux computer, and compared to the ESP32-CAM, it offers better capability for voice control and complicated object recognition applications. However, the Jetson Nano's higher price may be taken into account, particularly if the project does not make full use of the extra computing power it provides.

4 Methodology and Experimental Work

The block diagram of the algorithm is depicted in Fig. 1. It can be seen that the software runs in a continuous loop that passively listens and temporary transcribes every word it recognizes. A constantly active microphone allows the program to be sufficient and excludes the need for a button for activation, which facilitates usage and optimizes user interaction. The listening loop is part of the voice control module, which contains four submodules, namely Google Cloud SDK, PyAudio, gSTT (Google Speech-To-Text), gTTS (Google Text-To-Speech).

The device requires two additional sensors, a microphone and speakers. The captured audio data streams to the Google Speech-To-Text API for real-time transcription, converting spoken words into text. Then the transcribed text is being received and processed by applying noise reduction and filtering to improve accuracy. This step is necessary from using this mobile platform in high-noise scenarios. The transcribed text is then analyzed to identify keywords or commands. Recognized commands map to specific actions and

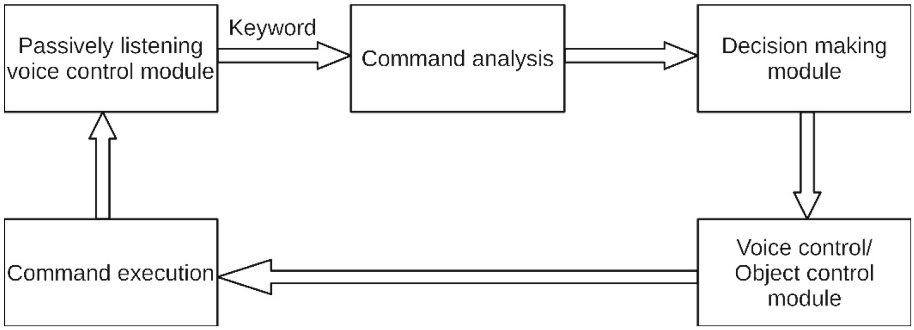

Fig. 1. Block diagram of the algorithm

functions within the application. This is explained better in the decision-making module, which is responsible for command responses and system monitoring. After the command executes the responses are being generated using the gTTS library, converting them into audio files. This allows the program to provide spoken feedback and the exclusion of a monitor connected to the platform itself. The generated audio responses play back through the connected speaker, hoping the user is nearby to hear the response.

The object detection module obtains video input from a camera connected directly to the Jetson Nano. The board is equipped with an algorithm that resizes, normalizes pixel values and coverts color spaces. This step is needed due to the possibility of differences in the frame-by-frame capture or addition or change of hardware. That secures that the software is compatible with any and all video-capturing devices. The video-stream is send to a pre-trained object detection model, optimized for the mobile platform. The model is applied to every processed frame and its bounding box in the video stream. Afterwards the post-processing begins with non-maximum suppression (NMS) to remove overlapping bounding boxes and retain only the most confident detections for each object. To ensure that the detection is correct, a visualization in the form of red border with text square with detected object is added. The output displays the annotated video stream with the detected objects in real-time.

The algorithm work over photos is given in the following figures. The detection of a dog in full size and by only seeing its nose is depicted in Fig. 2.

The object detection algorithm used recognizes faint detailing as can be noticed in Fig. 2 (a) and (b) that is the same dog. However, in Fig. 2 (b) the dog is not in full size. Despite that fact the algorithm recognized the head as a head of a dog and to the command “Show me the dog” it correctly outlined it.

The same algorithm is applied but with two different commands – “Show me the cat” and “Show me the bag” as shown in Fig. 3.

(a) (b)

Fig. 2. Detection of a dog "seeing" it in full size (a) and recognizing it by its nose (b)

(a) (b)

Fig. 3. Response to two different commands on one and the same "Show me the cat" (a), and "Show me the bag" (b)

It can be seen that the algorithm does not generalize on the same picture. Because it uses reinforcement learning, it would be very easy to confuse it and teach it that every time this picture is being viewed, one asks to see the cat. An interesting decision can be observed in Fig. 3 (b) as the algorithm had recognized the treat pouch (the bag behind the cat) as a bag (there can be seen two bags in this figure). This shows the level of generalization reached because the treat pouch, as the name suggests, is often not labelled as a bag. That little detail often is overseen but for some of the future uses of this algorithm is very important.

The object tracking is a separate module due to its usage. An API stream has been created that is optimized for viewing on a second device. That second tracking algorithm uses the detected module and extends the detection over the movement of the object by predicting where the object is moving whether there is any movement. This fills the previously removed overlapping boxes and allows for seamless tracking of the object. A loop has been created that reads the next frame from the video stream while showing the current one to the user. The user then can interact with the video by pausing and resuming the video stream. This can allow for human analysis of a frame or a video stream and adjustments to the object detection algorithm.

The decision-making module can initializes the state of the platform such as location, orientation and available sensor information. It contains the setup for the communication channels for receiving and answering commands via the voice control, object detection and tracking, and AI connections. The multi-threading of this module allows simultaneous voice control monitoring, object tracking and streaming which can be beneficial for the user. The decision-making module is equipped with dangerous situations pre-trained module, which allows it to notify the user or immediately stop a functionality. It also contains all the sensor information with the undesirable states of the sensors and the ability to software restart or inform of hardware change need.

The testing of every module is quite simple. Each module is tested on accuracy, speed, robustness and two module specific criteria. The results are summarized in Table 3 below.

The testing process included variety of places, including tormenting the pets of one of the researchers. The testing methodology is the following: the platform is placed in different environments and is set to log progress. Accuracy testing was implemented with commands "Correct" and "Wrong" and the statistic was calculated accordingly to in the percentage of "Correct" command given per the amount tested. The logs when the function tested executes. The robustness was tested with results reviewed by monitoring the error handling logs. The voice-control was tested by repeating existing and non-existing commands with various English accents. The purpose of the tester is to confuse the platform as much as possible. The object detection module received a diverse dataset of images and videos containing various objects. The algorithm logs and borders as many objects as possible. The object tracking was tested on real and non-real scenarios, including single-object tracking, occlusions, and abrupt changes in object motion. For the decision-making module, similarly to the object tracking, a range of scenarios were created, including hardware disappearance, faulty cables and weather conditions.

Table 3. Accuracy, speed, and robustness test results.

Module/ Criteria	Voice Control	Object Detection	Object Tracking	Decision-making
Accuracy	95%	92%	90%	85%
Speed	1.5 s	30 ms/frame	N/A	100 ms
Robustness	High robustness demonstrated with commands in different English accents	High robustness demonstrated across different lighting conditions	High robustness demonstrated under varying lighting conditions, background, clutter, and occlusions	High robustness demonstrated across a range of input data distributions and environmental conditions Able to maintain consistent decision quality under noise uncertainty, and adverse inputs
Special criteria #1	Error handling: Clear feedback provided for unrecognized commands	Detection range: Reliable detection range of 2 to 20 m depending on object size and environmental factors	Latency: 100 ms Real-time tracking achieved with minimal delay	Adaptability: Demonstrated ability to adapt and improve decision quality overtime through continuous learning from new data
Special criteria #2	Integration: Platform responded promptly without any noticeable delays	False Positives/Negatives: Low false positive rate (<5%) observed Occasional false negatives (<10%)	Long-term Tracking: Sustained accurate tracking over extended periods without noticeable drift or loss Occasional re-initialization required to correct drift or maintain track in complex scenes	Resource Efficiency: Efficient utilization of computational resources with minimal memory footprint and low processing overheat

5 Conclusions

A voice-controlled autonomous mobile platform with integrated camera sensor based on state-of-the-art NVIDIA Jetson Nano single board Linux computer was discussed in the paper. The research was headed towards improving user-to-machine interaction by establishing a connection with the Vertex AI via the provided application-programming interface. The elaborated voice control mechanism was using natural language processing techniques to convert spoken instructions into executable commands. The proposed idea has been verified in real environment by conducting some experiments that demonstrated very good accuracy and speed results. It is the authors' hope that this work can serve as a foundation for future improvements leading to easier navigation, decision-making in complex environments, better timing and reliable response in human-robot mutual co-existence and communication. The results achieved, as well as the planned future work, may be particularly useful for people with visual impairments or limited mobility.

Acknowledgments. This research is supported by a 2024 grant issued by The Central Fund for Strategic Development of the New Bulgarian University.

References

1. Abioye, S., Oyedele, L., Akanbi, L., et al.: Artificial intelligence in the construction industry: a review of present status, opportunities and future challenges. J.Building Eng. **44**(103299) (2021). https://doi.org/10.1016/j.jobe.2021.103299
2. Adi, P., Wahyu, Y.: Performance evaluation of ESP32 camera face recognition for various projects. Iota **02**(01) (2022). https://doi.org/10.31763/iota.v2i1.512
3. Alrousan, Q., Alzu'bi, H., Tasky, T., Varasquim, J.: LiDAR-based fail-safe emergency maneuver for autonomous vehicles. SAE Technical Paper (2023). https://doi.org/10.4271/2023-01-0578
4. Arachchi, H., Samarasinghe, G.: Impact of embedded AI mobile smart speech recognition on consumer attitudes towards AI and purchase intention across Generations X and Y. Emerald European J. Management Studies (2023). https://doi.org/10.1108/EJMS-03-2023-0019
5. Baby, C., Khan, F., Swathi, J.: Home automation using IoT and a chatbot using natural language processing. Innovations in Power and Advanced Computing Technologies (i-PACT), pp.1–6 (2017). https://doi.org/10.1109/IPACT.2017.8245185
6. Berg, G.: Image Classification with Machine Learning as a Service - A Comparison Between Azure, SageMaker, and Vertex AI (Dissertation) (2022). https://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-113829
7. Cao, L., Zhu, X.: An autonomous service mobile robot for indoor environments. International Symposium on Parameterized and Exact Computation (2020). https://doi.org/10.1109/IPEC49694.2020.9115180
8. https://www.zdnet.com/home-and-office/smart-home/best-home-automation-system/
9. Janíček, M., Ružarovský, R. Velíšek, K, Holubek, R.: Analysis of voice control of a collaborative robot. J. Physics: Conference Series, 1781(012025). International Conference on Applied Sciences (ICAS) (2020). https://doi.org/10.1088/1742-6596/1781/1/012025
10. Kalpana, S., Rajagopalan, S., Ranjith R., Gomathi, R.: Voice recognition based multi robot for blind people using lidar sensor. International Conference on System, Computation, Automation and Networking (ICSCAN), pp.1–6 (2020). https://doi.org/10.1109/ICSCAN49426.2020.9262365
11. Klančar, G., Seder, M., Blažič, S.: Advanced sensors technologies applied in mobile robot. Sensors **23**(2958), (2023). https://doi.org/10.3390/s23062958
12. Li, S., Liu, Y., Chen, Y., et al: Voice interaction recognition design in real-life scenario mobile robot applications. Appl. Sci. **13**(3359) (2023). https://doi.org/10.3390/app13053359

13. Liu, M., Hou, Z., Sun, Z., et al.: Campus guide: a lidar-based mobile robot. European Conference on Mobile Robots (ECMR), pp.1–6 (2019). https://doi.org/10.1109/ECMR.2019.8870916
14. Luo, J., Zhou, X., Zeng, C., et al.: Robotics perception and control: key technologies and applications. Micromachines **15**(4), 5312024 (2024). https://doi.org/10.3390/mi15040531
15. Mai, N., Bien, D.: Vision and voice recognition for 6 dof industrial robot model based on a multi-layered artificial intelligence network. Research Developments in Science and Technology **4** (2022). https://doi.org/10.9734/bpi/rdst/v4/2275B
16. Makubhai, S., Pathak, G., Chandre, P.: Exploring the trade-offs between blackbox and explainable AI: a comparative study. In: 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), pp. 1–8 (2023). https://doi.org/10.1109/ICCUBEA58933.2023.10391996
17. Masina, F., Orso, V., Pluchino, P., et al.: Investigating the accessibility of voice assistants with impaired users: mixed methods study. J. Medical Internet Res. **22**(9) (2020). https://doi.org/10.2196/18431
18. Rani, P., Bakthakumar, J., Kumaar, B., et al.: Voice controlled home automation system using natural language processing (NLP) and Internet of Things (IoT). International Conference on Science Technology Engineering & Management (ICONSTEM), pp. 368–373 (2017). https://doi.org/10.1109/ICONSTEM.2017.8261311
19. Salvini, P., Paez-Granados, D., Billard, A.: Safety concerns emerging from robots navigating in crowded pedestrian areas. International Journal of Social Robotics **14**, 441–462 (2022). https://doi.org/10.1007/s12369-021-00796-4
20. Sharan, S., Quoc, T., Nauth, P., et al.: Implementation and testing of voice control in a mobile robot for navigation. IEEE/ASME International Conference on Advanced Intelligent Mechatronics (2019). https://doi.org/10.1109/AIM.2019.8868892
21. Sun, J., Zhao, J., Hu, X., et al.: Autonomous navigation system of indoor mobile robots using 2D lidar. Mathematics **11**(1455) (2023). https://doi.org/10.3390/math11061455
22. Umar, A., Giwa, M., Kassim, A., et al.: Development of an android-based, voice controlled autonomous robotic vehicle. Eng. Proc. **58**(1), 48 (2023). https://doi.org/10.3390/ecsa-10-16026
23. Valladares, S., Toscano, M., Tufiño, R., et al.: Performance evaluation of the nvidia jetson nano through a real-time machine learning application. Springer, Cham. Advances in Intelligent Systems and Computing, **1322** (2021). https://doi.org/10.1007/978-3-030-68017-6_51
24. Xiang, Y., Yang, X., Yang, W., et al.: Localization and mapping algorithm for the indoor mobile robot based on LIDAR. IOP Conference Series: Materials Science and Engineering **831**(012021) (2020). https://doi.org/10.1088/1757-899X/831/1/012021
25. Yiheng, L., Tianle, H., Siyuan, M., et al.: Summary of ChatGPT-related research and perspective towards the future of large language models. Meta-Radiology 100017 (2023). https://doi.org/10.1016/j.metrad.2023.100017
26. Zhang, J., Singh, S.: Loam: lidar odometry and mapping in realtime. Robotics: Science and Systems **2**, 9 (2014). https://doi.org/10.15607/RSS.2014.X.007