# ACSO BahiaRT Team Description Paper RoboCup@Home 2025

Pedro H. O. dos Santos Gabrielle F. S. Carvalho
Vitória F. N. Matos Luis J. S. Junior Samuel J. Cesar
Pedro L. J. Santos Elias R. da Silva Filipe N. Silva
Kaique W. S. da Silva José Grimaldo S. Filho
Ivanoé J. Rodowanski Jorge A. P. Campos
Marco A. C. Simões Josemar R. de Souza
Ana Patrícia F. M. Mascarenhas

November 25, 2024

**Abstract.** This paper presents the BahiaRT team and describes an autonomous service robot named Bill and its capabilities, such as, navigation, manipulation, people and object recognition, human-robot interaction and decision making as well as It's hardware and software systems. Furthermore, the paper highlights BahiaRT research interests and scientific contributions. Complementary information about our tobot Bill and its code are available at: `https://gitlab.com/bahiart/athome/BahiaRT-atHome-2025`

## 1 Introduction

The Center of Computer Architecture, Intelligent Systems and Robotics (ACSO) at the State University of Bahia (UNEB) has been participating in RoboCup with the BahiaRT team since 2009 in leagues such as 2D Soccer Simulation, Mixed Reality, 3D Soccer Simulation, and @home.

The service robot proposal of BahiaRT for the RoboCup@Home league is called Bill (Bot Intelligent Large capacity Low cost). It was born in 2014 as a result of research projects in assistive robotics. Its main goal is to assist humans in common tasks. To do this it has some capabilities, such as, communicating with humans through natural language processing and signs, recognizing objects and faces, and navigating through unknown environments.

Over the years Bill has participated in both RoboCup@home competition and in the @Home Brazil Robotic Competition, securing notable achievements. In RoboCup@home league Bill got 13th place in 2015 and 21st place in 2016. In

RoboCup@Home Brazil, it got 2nd place in 2015, 3rd place in 2016, 3rd place in 2017, 6th place in 2023, and 7th place in 2024.

This paper presents the third generation of Bill, named **BILL Estranho**, and its main improvements. It includes re-engineering the hardware using new components, redefining the architecture, and changing the operating system to ROS2. As a consequence, this year, other functionalities, i.e.human-robot interaction, face and object recognition, navigation, and manipulation were the focus of update.

The remind of this paper is organized as follows: Section 2 introduces our research group interests and achievements. She section 3 represents the team's contribution to the league. Section 4 describes a task performed by Bill and Section 5 presents the conclusions and future work.

## 2 Advances in innovative technology and research interests

The main research interests of BahiaRT team involve robotics and artificial intelligence, specifically focusing on human-robot interaction, object recognition, and fault-tolerance of Bill's hardware.

Concerning human-robot interaction, in the last years we focused on researches to improve Bill's communication using Large Language Models. In addition, we have been working on providing interaction for hearing impairment people through signs languages. Moreover, interaction has been also improved by emotions capture, i.e., we can better fit the robot's answers according to the human voice intonation.

Related to object detection, we have been working on improving recognize accuracy. In this way, our strategy consists in integrate optical character recognition (OCR) and image recognition to better recognize the objects manipulated by Bill.

For fault-tolerance, Bill's current release adopted a method that we call Fox-Dog to control the left/right motor front and rear. So, if one microcontroller fails, the other takes over.

### 2.1 Navigation

Navigation is the keystone for efficient execution and environment interaction for robots. The components used by Bill for navigation are: encoders output, odometry, Slam_toolbox, behavior tree based navigation node called bt_navigator (ROS2), AMCL (ROS2), map_server (ROS2) and 360° laser scanner. Odometry module uses the he encoder data to estimate the movements of Bill in space. Further, the Behavior-Tree Navigator uses the odometry data to trace trajectory to a desired target. Once all data are published, the simultaneous mapping and localization using the AMCL is activated integrating the 360° laser scan data.

SLAM approach is responsible to map the environment and provide self-localization in this map. First, the incremental mapping package builds the map using Slam_toolbox [1]. Then, the grid map is generated by a LIDAR sensor, which is capable of capturing 2D environment information.

The next step is creating the path planning based on the occupancy grid map that is updated based on the navfn_planner. Then the shortest path to the goal is computed by means of the D* Lite algorithm to ensure obstacle avoidance over incremental mapping. The motion planning in charge of getting the path planning and relating linear and angular motion is triggered, which applies the kinematics control law, and sends a message to low-level control.

### 2.2 Vision

This module handles the reception, processing, and response to external stimuli through image capture. It consists of three main sub-modules: (i) object detection integrating image detection and Optical Character Recognition OCR; (ii) facial Recognition; and (iii) communication using sign language, witch represents one of the news team research project.

Concerning object detection, BILL uses YOLO (you only look once) version 8, a cutting-edge technology capable of real-time detection across a broad spectrum of objects. Training for BILL involved photos of distinct objects taken in various locations and orientations. We employed images containing varying numbers of objects per frame to assess the impact of parameters and train photo quantity on recognition accuracy. The labeling process was facilitated by RoboFlow software.

To improve the accuracy of our object detection system, we currently integrate Yolo with Tesseract [2], an Optical Character Recognition OCR. This strategy has already proven to be effective in sectors such as industry and traffic management, and we believe it could also contribute to service robot domain.

Inspired by BILL's emphasis on achieving high capacity at a low cost, we designed a two-stage system for efficient facial recognition. The first stage focuses on facial detection, a crucial step is isolate the face from the rest of the image. Here, we leveraged the power of the Haar-cascade algorithm [3]. This open-source technique, readily available within the OpenCV library [4], excels at swiftly pinpointing faces in an image. It achieves this by identifying specific rectangular patterns, essentially acting like a digital fingerprint scanner for facial features. By employing Haar-cascade, we can effectively isolate the face of interest for further processing. The second stage focuses on Dlib [5], which implements a sophisticated landmark detection technique. This technique goes beyond simply locating the face, it meticulously pinpoints 68 key coordinates across the face, creating a detailed map of its unique characteristics. This information, encompassing the relative positions and shapes of facial features like eyes, nose, and mouth, is then extracted and mapped to train the facial recognition model.

To gauge the system's effectiveness, we conducted recognition tests with a group of 55 students from Bahia State University. After a brief training period, the algorithm demonstrated the ability to identify specific individuals within a group setting.

The communication using signs language aims to provide robot-human interaction for hearing impairment people. This first release uses the Brazilian Sign

Language, named LIBRAS [6]. Figure 1 shows the communication process. The person does the sigh that is captured by Bills camera and processed by our solution, named RoboSign. It converts the sign to a text instruction and uses it as input to Bills natural language processing, just like the voice interaction.

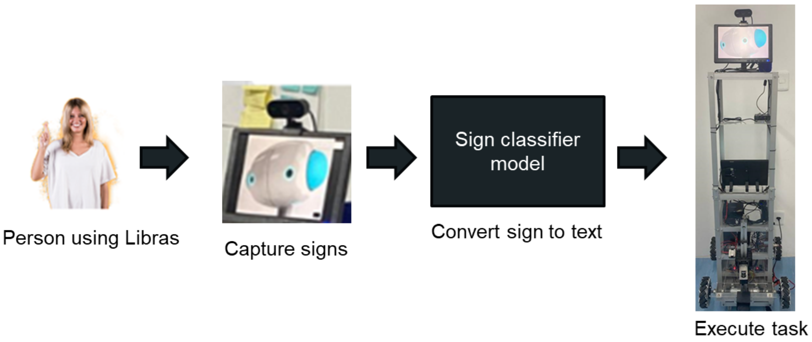

**Fig. 1.** Communication process using sign language.

The RoboSign solution uses the MediaPipe [7] landmark detection software to capture signs for dataset definition. Them, a recurrent neural network LSTM [8] uses this dataset in both training and validation. To assess the RoboSign accuracy a LIBRAS professor from our university interacted with Bill in real time. The results achieved 85% of accuracy in recognizing the signs demonstrating its potential to make Bill more accessible.

### 2.3 Speech Recognition and Voice

Voice is the most widely used form of human-machine interaction and we have adopted it as the main means of communication with BILL. Due to BILL's upgrade to ROS2, we are currently adopting Google's Speech Recognition software to enable voice interaction.

Our speech recognition system employs advanced machine learning techniques, specifically the GPT-2 [9] model for understanding and generating responses in natural language and the DistilBERT [10] for question-answer in specific context.

For our purposes, GPT-2 has been fine-tuned with a custom dataset consisting of various command phrases and natural language interactions relevant to BILL. Training of the GPT-2 model was conducted using TensorFlow[11] and PyTorch[12] libraries. Additionally, we use the Speech Recognition library to interface with Google's Speech Recognition API.

To train the DistilBERT algorithm we adopted the SQuAD [13] dataset from Stanford university and a dataframe divided into three columns, question, answer and context.

During the inference phase, the system captures voice commands through BILL's array of microphones. The captured audio is pre-processed to enhance clarity and reduce background noise. The processed audio is then transcribed into text using Google's Speech Recognition API. For general conversation, the transcribed text is fed into the fine-tuned GPT-2 model, which interprets it and generates an appropriate response. Similarly, for quiz tasks the transcribed text is fed into the DistilBERT algorithm.

Audio capture is a critical component of our speech recognition system. So, Bill is equipped with high-quality microphones strategically positioned to optimize sound capture. Audio processing includes noise reduction to filter background noise and improve signal-to-noise ratio, along with real-time processing to ensure audio capture and processing occur in real-time, enabling quick response to voice commands.

This comprehensive approach to speech recognition and voice interaction aims to create a robust and reliable system that enhances human-BILL interaction, making it more intuitive and efficient.

With the same purpose of improving communication between humans and robots, we have also been working on capturing emotions through voice. The research emphasizes the role of voice in expressing emotions and discusses practical applications in virtual assistants, voice control systems, and service robots. The integration of acoustic and lexical features offers innovative perspectives for developing more empathetic technologies. The research reviews related works, describes theoretical foundations, details the proposed approach, discusses experiments and results, and concludes by highlighting the importance of combining speech and text analysis for real-time emotion detection. It also suggests future studies on the impact of gender in emotion detection and proposes expanding datasets and exploring additional approaches to achieve more accurate emotion detection.

This research delves into the intricacies of emotion recognition in speech, emphasizing voice recognition in the context of Brazilian Portuguese. By integrating a convolutional neural network and a transformer model, the research achieved an accuracy rate of 87.5%. It underscores the significance of voice as a vital communication channel for conveying emotional nuances. The research highlights the challenges in detecting emotions in speech due to cultural and linguistic influences unique to the Brazilian context. The scarcity of Brazilian Portuguese data poses a barrier to developing culturally sensitive systems, underscoring the need for more data to enhance human-computer interaction and emotional well-being.

## 3 Contributions

Based on the characteristics of robots built by ACSO, used in competition over the last 18 years, and observations on equipment from other parts of the world, we had the honor to bring contributions in Bill to the league RoboCup@Home.

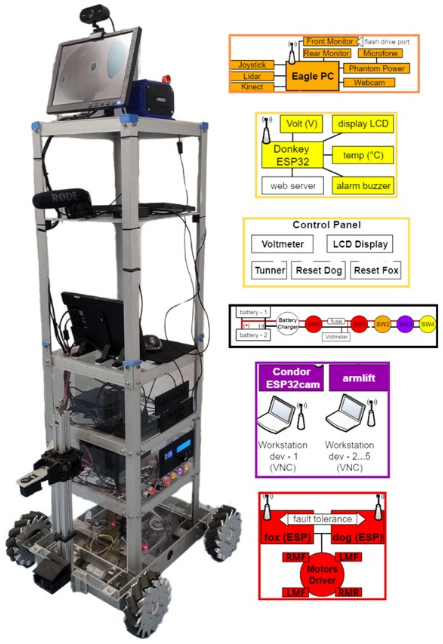

**Fig. 2.** BILL Estranho.

For better accessibility in human-robot communication, we bring the contribution of recognizing the Brazilian Sign Language (LIBRAS) through our RoboSign solution,allowing communication with hearing impairment people. RoboSign is avaliable at our GitHub repository and it can be used, studied and increased by other teams.

Another contribution regarding human-robot communication, our team brings the emotion detection by speech. This research aims to generate better answers according to the emotions detected in the human speech. It's code is also available at our GitHub repository.

Concerning object detection, we are currently working on the integration of YOLO and Tesseract for Optical Character Recognition and still have some results that indicates an improvement in detecting objects that have labels, e.g. supermarket products.

Finally, Bill Estranho electronics were improved with better protection for short circuits and higher power. Use of ROS2, micro-ROS (puts ROS2 onto microcontrollers) [14] and ESP32. Besiden, in an unprecedented way, the use of Fault Tolerance (Fox-Dog method) in Service Robots. The fault-tolerance Fox-Dog method is implemented using two ESP32 microcontrollers replicated with ROS2. They control the left/right motor front (LMF/RMF) and the left/right motor rear (LMR/RMR). So, if one microcontroller fails, the other takes over.

## 4 Example of Task Performed by Bill

The integration of Bill's diverse capabilities makes it possible to perform tasks to assist humans. For example, Bill may act as a hostess in a restaurant, interacting naturally with guests, seating them at a table and serving drinks.

In this task, the robot is initially positioned at the restaurant entrance. For each customer that arrives, Bill asks for their name, using his voice, and indicates them to a table. Afterwards, Bill goes to the table, recognizes the customer by face, and asks what drink he would like to serve. At a later time, it can also recognize the drink that each customer is consuming, using object recognition.

To perform this task, the following capabilities are required: natural language processing, to interact with customers using voice; people recognition, to serve the customer; object recognition, to recognize the drinks served to each customer; handling, to serve the requested drinks; and navigation, for moving the robot within the restaurant.

## 5 Conclusion

This paper introduced Bill, the service robot of BahiaRT team for the RoboCup@Home league. Bill's new improvements includes interaction using LLM, communication using sign languages, emotion detection, and the fault-tolerance system.

Related to LLM models, our current algorithm had a significant improvement in accuracy, showing that this technique can contribute a lot to human-robot communication.

Bill is also an inclusive robot as it can recognize LIBRAS signs. However, currently our dataset is limited to Bills task commands. So, we are working on increase it with new sign.

Concerning emotion detection, we initially made a specific quantity of progress. The main reason is because this subject has a wide and multidisciplinary field and much work are being done to achieve a large capacity of advance in this area.

Finally, the fault-tolerance system increases infallibility, giving an additional level of security to BILL's actions and performance tasks.

## References

1. Stefan Kohlbrecher, Oskar Von Stryk, Johannes Meyer, and Uwe Klingauf. A flexible and scalable slam system with full 3d motion estimation. In *Safety, Security, and Rescue Robotics (SSRR), 2011 IEEE International Symposium on*, pages 155–160. IEEE, 2011.
2. Tesseract. `https://github.com/tesseract-ocr/tesseract`. accessed: 202-06-25.
3. Haar. `https://docs.opencv.org/3.4/db/d28/tutorial_cascade_classifier.html`. accessed: 2023-07-21.
4. Opencv. `https://opencv.org/`. accessed: 2024-07-03.

5. Dlib algorithm. https://medium.com/brasil-ai/mapeamento-facial-landmarks-com-dlib-python-3a200bb35b87. accessed: 2023-07-21.
6. Linguagem brasilera de sinais.
7. Mediapipe. https://developers.google.com/mediapipe. Accessed: 2024-11-19.
8. Long short-term memory.
9. Chatgpt. https://chatgpt.com/. accessed: 2024-07-03.
10. Distilbert. https://huggingface.co/docs/transformers/model$_{d}oc/distilbert.Accessed: 2024-11-19$.
11. Tensorflow. https://www.tensorflow.org/?hl=pt-br. accessed: 2024-11-23.
12. Pytorch. https://pytorch.org/. accessed: 2024-11-23.
13. Squad dataset. https://rajpurkar.github.io/SQuAD-explorer/. Accessed: 2024-11-19.
14. micro-ros. https://micro.ros.org/. accessed: 2023-07-21.
15. V. Khandelwal. “the architecture and implementation of vgg16,”. https://pub.towardsai.net/the-architecture-and-implementation-of-vgg-16-b050e5a5920b. Accessed: 2021-08-25.
16. Roboflow. https://app.roboflow.com/. Accessed: 2024-06-19.
17. Ros2. https://docs.ros.org/en/humble/index.html. accessed: 2023-07-21.

Authors: Pedro H. O. dos Santos, Gabrielle F. S. Carvalho, Vitória F. N. Matos, Luis J. S. Junior, Samuel J. Cesar, Pedro L. J. Santos, Elias R. da Silva, Filipe N. Silva, Kaique W. S. da Silva, José Grimaldo S. Filho, Ivanoé J. Rodowanski, Jorge A. P. Campos, Marco A. C. Simões, Josemar R. de Souza, Ana Patrícia F. M. Mascarenhas

## Robot BILL Hardware Description

The robot is built based on performing house tasks. Specifications are as follows:

- Base:
  - Micro-controller

    ESP32-WROOM 38 PIN .
  - Sensors

    Encoder: B37Y3530-131EN LIDAR: lslidar Leishen N10.
  - Actuators

    4 Omnidirectional mecanum wheels. DC Motors . H-bridge L298 .
- Torso: Contain 4 shelves to support the hardware parts.
  - Emergency button.
- Arm: Mounted on torso.
  - 5 DoF TurtleBot Robotic Arm (built from a set of 5 Dynamixel AX-12+ model servo motors).
  - Controlled by Arduino MKR Zero board withRobotis Dynamixel Shield.
- Head: Screen with the robot face.
- Robot dimensions: Height – 1,20 m (max), Width – 0,70 m (max)
- Robot weight: 22,2 Kg.

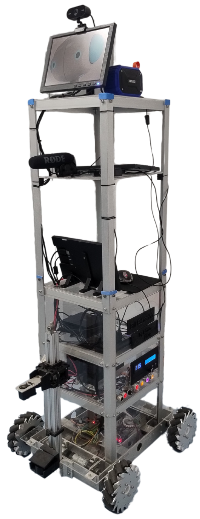

**Fig. 3.** Robot BILL

*Also, our robot incorporates the following devices:*

- Battery charge indicator
- Microphone: Shotgun Rode VideoMic Pro Compact
- Camera: C920 Logitech
- Sound Box: Bluetooth Portable Kimiso-112
- CPU: Beelink mini s (Intel 11th Gen N5095)
- Monitor: VBESTLIFE LCD 16:10 HD

Robot software and hardware specification sheet

## Robot's Software Description

*For our robot we are using the following software:*

- Platform: Robot Operating System (ROS) 2 - Humble
- Navigation: MicroROS, NAV2
- Face recognition: OpenCV, Dlib, Haar-cascade
- Speech recognition: Google Speech Recognition
- Speech generation: Google Text-to-Speech (gTTS)
- Object recognition: YOLOv8n
- Arms control: automatic controller
- Localization: SLAM
- Mapping: RVIZ 2
- Simulation environment: Gazebo
- People Tracking: OpenPose
- OS: Linux Ubuntu 22.04

## External Devices

*BILL relies on the following external hardware:*

- Keyboard
- Mouse
- Hub1: 7 Ports HUB
- Hub2: Exbom HUB 30A

## Cloud Services

*BILL connects the following cloud services:*

- API Google Text to String (gTTS);
- API Google Speech Recognition.