



# Jubileo: An Immersive Simulation Framework for Social Robot Design

Jair Augusto Bottega<sup>1</sup> · Victor Augusto Kich<sup>1</sup> · Junior Costa de Jesus<sup>2</sup> · Raul Steinmetz<sup>3</sup> ·  
Alisson Henrique Kolling<sup>4</sup> · Ricardo Bedin Grando<sup>4,5</sup> · Rodrigo da Silva Guerra<sup>4</sup> ·  
Daniel Fernando Tello Gamarra<sup>3</sup>

Received: 15 June 2023 / Accepted: 4 October 2023 / Published online: 7 December 2023  
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

## Abstract

This paper introduces Jubileo, an open-source simulated humanoid robot as a framework for the development of human-robot interaction applications. By leveraging the power of the Robot Operating System (ROS) and Unity in a virtual reality environment, this simulation establishes a strong connection to real robotics, faithfully replicating the robot's physical components down to its motors and enabling communication with servo-actuators to control both the animatronic face and the joints of a real humanoid robot. To validate the capabilities of the framework, we propose English teaching games that integrate Virtual Reality (VR), game-based Human-Robot Interaction (HRI), and advanced large language models such as Generative Pre-trained Transformer (GPT). These games aim to foster linguistic competence within dynamic and interactive virtual environments. The incorporation of large language models bolsters the robot's capability to generate human-like responses, thus facilitating a more realistic conversational experience. Moreover, the simulation framework reduces real-world testing risks and offers a cost-effective, efficient, and scalable platform for developing new HRI applications. The paper underscores the transformative potential of converging VR, large language models, and HRI, particularly in educational applications.

**Keywords** Human-robot interaction · Virtual reality · Learning game · English language learning · Large language models

## 1 Introduction

The increasing integration of robots into our everyday lives has spurred the growth of Human-Robot Interaction (HRI), a multidisciplinary field dedicated to the understanding, design, and evaluation of robotic systems intended for human use [1]. HRI encompasses several modes of communication, including voice, gestures, and body language. Social humanoid robots, which closely resemble human figures,

have a significant role in fostering direct human-robot social interaction. These robots, with their human-like appearance and capabilities, can engage in more natural and intuitive interactions with humans, thereby enhancing the effectiveness and user acceptance of robotic systems. This evolution in HRI has been supported by advances in various fields, all the way from designing social cues for collaborative factory robots [2, 3], up to designing social robots to help treat children with autism spectrum disorder [4], including the development of entire new fields such as Social Robotics [5], Android Science [6] and Cognitive Developmental Robotics (CDR) [7].

However, the process of creating real robots is time-consuming and costly, often restricting the ability to iterate variations in design and conduct extensive testing. The physical nature of real-world robot testing further limits the ability to explore different scenarios and interactions in real time. To overcome these difficulties, simulation has emerged as a valuable alternative that helps accelerate the design and development process, explore diverse design iterations, and extensively test the robot's performance in various scenarios, often faster than real time [8, 9]. Virtual Reality (VR),

✉ Jair Augusto Bottega  
jairaugustobottega@gmail.com

Daniel Fernando Tello Gamarra  
daniel.gamarra@ufsm.br

<sup>1</sup> Intelligent Robot Laboratory, University of Tsukuba,  
Tsukuba, Ibaraki, Japan

<sup>2</sup> Korea Advanced Institute of Science & Technology, Daejeon,  
South Korea

<sup>3</sup> Federal University of Santa Maria, Santa Maria, RS, Brazil

<sup>4</sup> Federal University of Rio Grande, Rio Grande, RS, Brazil

<sup>5</sup> Technological University of Uruguay, Rivera, Uruguay

with its capacity to deliver a highly immersive experience coupled with realistic visual cues, has emerged as another valuable asset, especially in Human-Robot Interaction (HRI) research [10, 11].

VR enables engaging, controlled, reproducible experiments, free from real-world variabilities. It provides a platform to easily parameterize and test variations, including the immersive human experience, allowing researchers to manipulate conditions and variables with ease. This flexibility facilitates a more comprehensive understanding of the potential outcomes and impacts of different robotic applications involving human interaction. Thus, VR serves as a powerful tool in HRI research, improving the efficiency, safety, and depth of robotics experimentation.

This paper introduces Jubileo, an open-source simulated humanoid robot framework that takes advantage of ROS, Unity, and virtual reality. In our previous works [12, 13], we introduced an open-source animatronic face and a VR simulation for conducting immersive experiments in the HRI field, effectively demonstrating the value of VR for testing simulated social robot applications. The current work builds upon this foundation, introducing enhancements to the existing VR framework and an innovative application for English language learning through game-based HRI, incorporating large language models into the robot's system. A brief introduction to the English language learning games can be seen in Fig. 1.

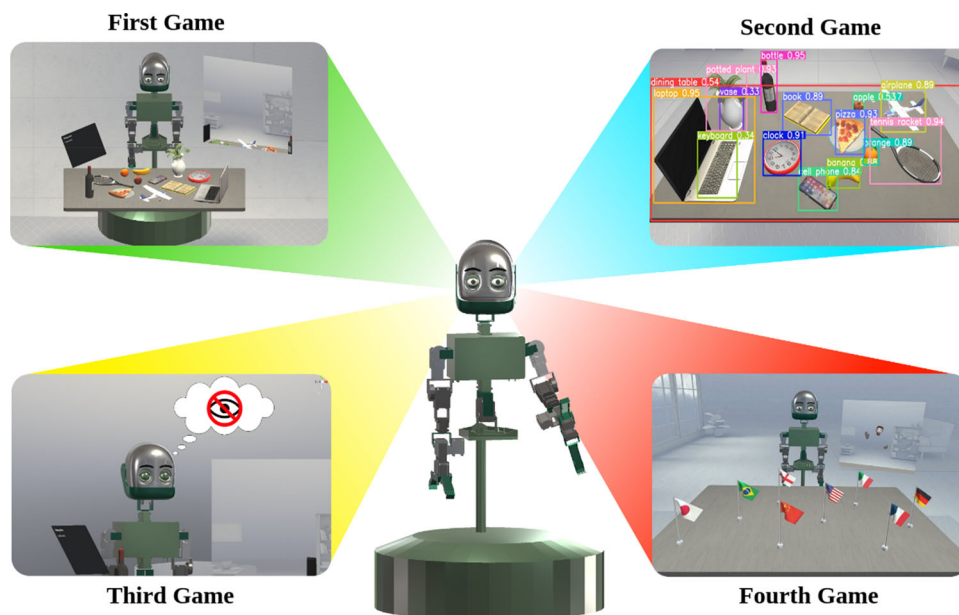
Large language models are a significant development in the field of artificial intelligence, designed to understand

and generate human-like text [14]. These sophisticated AI systems learn from vast amounts of textual data, allowing them to generate responses, complete tasks, and simulate conversation in a manner that is increasingly indistinguishable from human interaction [15, 16]. This integration further strengthens the robot's ability resulting in an immersive language learning experience.

The utilization of game-based environments for language education poses various challenges [17]. However, due to children's frequent interaction with such environments, incorporating gamified elements can be beneficial for language learning [18], despite the need for further evidence to support our proposed approach. The game incorporates our existing open-source animatronic face, which is designed to avoid the Uncanny Valley phenomenon [19] and encourage comfortable and receptive interactions.

Our immersive VR language learning game demonstrates the potential of integrating VR, HRI, and Large Language Models to create educational experiences that are engaging and enjoyable. While we acknowledge the extensive research and expertise in the field of language education, we believe the approach presented in this paper could offer an additional tool for educators, researchers, and developers.

The paper is organized as follows: Section 2 reviews relevant literature and prior research. The systems and methodologies employed to construct the comprehensive virtual reality platform, along with the conception and design of the language learning games, are detailed in Section 3. Section 4 presents an in-depth analysis and discussion of



**Fig. 1** The image features our virtual robot surrounded by four snapshots, each representing an English language learning game. The first game involves a user-led Q&A session, while the second game reverses

the roles, with the robot leading the Q&A. The third game focuses on the user's descriptive abilities, and the fourth game combines language learning with cultural exploration

the results, and the perspectives for VR and LLM applied to education. Lastly, Section 5 highlights the key contributions of this research and outlines potential directions for future work.

## 2 Related Works

Over the course of time, the continuous advancements in hardware and software technologies have significantly propelled the integration of robots within various industrial sectors. A notable recent development in this realm is the emergence of collaborative robots, commonly known as cobots, which have become a cornerstone of the Industry 4.0 paradigm [20]. Concurrently, the remarkable progress in machine learning techniques has precipitated a noteworthy trend, wherein robots are progressively permeating human daily routines and domestic environments. This phenomenon has spurred the emergence and evolution of the field of HRI, aiming to comprehend and enhance the interactions between humans and robots [21–23].

With the continuous evolution of research in the field of HRI, robots have undergone significant advancements, becoming increasingly intricate and human-like in their design and behavior. This progress is exemplified in studies conducted by Faraj et al. [24], where an adult-sized humanoid head that can emulate human facial expressions, head movements, and speech through the use of 25 artificial muscles was developed.

As robotics researchers strive to create robots that closely emulate human characteristics, they often encounter a concept known as the Uncanny Valley. The Uncanny Valley refers to a psychological phenomenon where humanoid robots or virtual characters that approach human likeness but do not quite achieve it can elicit feelings of unease or discomfort in human observers [19, 25, 26], shedding light on the factors that contribute to the Uncanny Valley effect and its implications for HRI. To overcome this challenge, many social robots are intentionally designed to have simplified or caricatured facial features.

Throughout the evolution of robotic systems, simulations have emerged as an integral and indispensable tool for scientists and researchers. Simulations provide means to develop, validate, and refine new functionalities and components without the inherent risks associated with hardware damage or potential harm to humans [27, 28]. They offer a controlled and safe environment to explore and experiment with various scenarios and configurations. One key advantage of simulations is their ability to facilitate rapid prototyping and iteration. It is often easier and more efficient to simulate and evaluate new behaviors or algorithms before implementing them in the physical world. Additionally, interactive

applications such as those involving human engagement, require simulations capable of running in real-time.

The utilization of robots in education has witnessed a significant upsurge in recent years [29]. Extensive research, such as the study conducted by Belpaeme et al. [30], has consistently demonstrated the profound impact of robots in fostering the development of cognitive and affective abilities, particularly when employed as tutors for children in focused learning tasks. Furthermore, the comprehensive study conducted by Toh et al. [31] delved extensively into the effects of robots on young children's education, specifically examining various domains including cognitive, language, conceptual, and social skills. Their research shed light on the wide-ranging benefits that robots can offer in enhancing these crucial areas of development in early education. Kim et al. [32] demonstrated a successful application of educational robots in the field of STEAM education. Their study showcased the effective utilization of a robot to enhance learning experiences in science, technology, engineering, arts, and mathematics. By integrating the robot into the educational framework, they were able to create an engaging and interactive environment that fostered holistic learning and development among students.

As thoroughly discussed by Zheng et al. [33], VR is an immersive and advanced human-computer interface that creates a realistic environment, offering users a simulated experience. Within this virtual world, users can navigate, explore, and interact from various perspectives. They can manipulate objects, observe the environment from different angles, and physically engage with it in various ways.

There have been significant advancements in the use of VR in game-based learning. For instance, research by Chen et al. [34] has demonstrated the potential of VR in creating immersive and engaging educational games. Their work focuses on the development of a VR-based language learning game, which shares similarities with our approach. Such research underscores the potential of VR in creating interactive and engaging learning environments, which is the core of our work.

Furthermore, Shim et al. [35] conducted a study that employed a VR approach to address learning difficulties among middle school students. The particular topic investigated in this study was the anatomical structure and functional processes of the eyes. Moreover, Bogusevschi et al. [36] conducted a study that investigated the efficacy of a 3D virtual reality approach for physics education, specifically emphasizing the topic of the water cycle.

In a groundbreaking endeavor, Shahab et al. [37] conducted a scientific investigation uniting VR technology and robotics. The study aimed to explore the efficacy of utilizing VR-simulated robots as educational tools for teaching music and providing educational support to children diagnosed with Autism Spectrum Disorder (ASD).

In recent years, the integration of Large Language Models (LLMs) such as ChatGPT into human-robot collaboration has shown promising results. A study conducted by Ye et al. [38] proposed RoboGPT, a system that integrates ChatGPT with robot control capabilities. The AI assistant in RoboGPT can understand the task context, procedure, and verbal commands, and interact with human operators to clarify commands or request more contextual information. This study demonstrated that the use of ChatGPT in robots significantly increased trust in human-robot collaboration. This increase in trust can be attributed to the robot's ability to communicate more effectively with humans. Furthermore, ChatGPT's ability to understand the nuances of human language and respond appropriately helps to build a more natural and intuitive human-robot interaction. This finding has significant implications for the development of human-robot collaboration systems, particularly in tasks that require complex interaction and understanding between humans and robots [38].

In another work, Christmann [39] designed an object guessing game in an HRI setup, with a real social robot, in a language learning scenario. To the best of our current knowledge, no research has been conducted thus far utilizing VR-simulated social robots as a means to teach foreign languages. With the current advancements in LLM-based chatbots, we believe the exploration of these kinds of applications will become essential for the development of more immersive educational experiences, making it a versatile tool for language learning across different age groups.

### 3 Material and Methods

This section delineates the methodology employed in our virtual reality-based language learning system. Beginning with a review of the foundational aspects of the Jubileo platform, we will detail the integration of state-of-the-art technologies like the YOLO (You Only Look Once) object detection system and the GPT-4 Large Language Model. We will then present a collection of interactive language learning games, designed to foster a dynamic and engaging learning environment. Each game emphasizes a different aspect of language acquisition, offering a comprehensive approach to enhancing English language proficiency.

#### 3.1 Framework and System Structures

In previous research [12, 13], we introduced Jubileo as a novel research platform. This framework brought together a variety of tools and technologies, designed with a focus on interaction and engagement. The robot's physical structure, made possible through 3D printing and controlled by servo

motors, was discussed in depth, but in this current work, we primarily utilized the VR representation of Jubileo.

At the heart of the system was the Robot Operating System (ROS) [40], a flexible framework for writing robot software. ROS facilitates the connection between multiple systems, and it was used to develop the robot system, which provides a powerful base of algorithms. The connections occur with a publisher-subscriber system, where nodes subscribe and publish on topics. In this manner, we enable capabilities such as facial and emotion recognition, object detection by color, visual tracking, speech-to-text integration, and more. It's worth noting that the Jubileo framework provided ROS structures are compatible with both real and simulated robots, which is crucial for the current VR-centric work.

The Unity Engine [41] is a tool that can be used for robotics simulation due to its 3D rendering, real-world physics simulation, and scripting API. These features allow developers to create realistic environments and interactions for robots. Unity's support for a wide range of hardware, including VR, further enhances its simulation capabilities. Moreover, its integration with robot operating systems like ROS allows for detailed and realistic simulations. In the case of Jubileo, the communication between Unity and ROS nodes was facilitated through the TCP/IP protocol, enabling real-time control of the robot within the Unity environment. Thus, Unity's flexibility and performance make it an ideal choice for various robotics applications, from industrial automation to autonomous driving. Furthermore, through the use of URDF open-source models, the simulated robot is not limited to Unity, and it can be extended to any supported URDF robotics simulation platform. Our previous work [12], provides a full description of the development of the Jubileo framework, highlighting the application of the ROS system and its communication with the VR subsystems provided by the Unity Engine.

By involving multiple senses, meaningful interactions can be achieved, as different sensory modalities contribute to a richer and more comprehensive engagement. In this manner, we provide applications of vision systems and other interacting ways, detailed in the work of [13], where experiments with the detection of multiple objects were conducted using mainly color detection. To improve those systems and integrate a real-time object detection system, we incorporated YOLO [42]. The strength of YOLO lies in its ability to detect objects in images with impressive speed and accuracy, which makes it ideal for applications needing real-time feedback. In our VR environment, YOLOv5x6 is used to identify and track the selected objects within the game. When the robot or the human user selects an object, YOLO recognizes the object from its pre-trained database and keeps track of it throughout the game session. This feature enables seamless and intuitive interaction with objects in the virtual environment, and



reinforces the realism of the gaming experience, making the learning process more engaging and immersive.

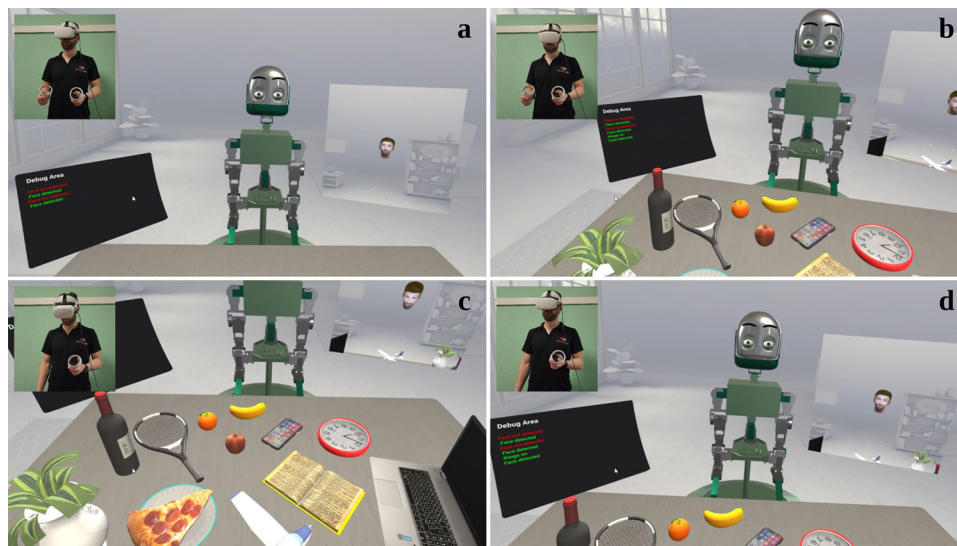
Regarding the LLM, we utilized GPT-4 [43], an advanced iteration of the Generative Pretrained Transformer models developed by OpenAI. The LLM is an artificial intelligence model that has been trained on an extensive corpus of text from the internet. As such, it is capable of understanding and generating human-like text, giving it the ability to engage in conversation, answer questions, and provide explanations in a way that is remarkably human-like. We integrated the GPT-4 model into the VR framework to serve as the “brain” of the virtual robot. When the robot receives a question or command from the user, it uses GPT-4 to understand the input and generate an appropriate response. This allows the robot to engage in meaningful, dynamic, and contextually aware conversations with the users. The incorporation of a large language model like GPT-4 thus greatly enhances the functionality of our system, making the VR language learning experience more interactive and effective.

To accomplish the task of translating user-spoken language into text inputs suitable for the LLM’s prompt, the Google Speech-to-Text API was employed [44]. This solution was chosen due to its notable robustness, cost-effectiveness, and swift processing capabilities. Additionally, for transforming the LLM’s generated text responses into audible output for the user, the Google Text-to-Speech API is used [45]. This API supports various audio formats and offers a variety of customization options. This flexibility makes it easier to adjust how the spoken output sounds, allowing for a more natural and personalized conversation experience for the user.

### 3.2 Learning Games

In the development of this language learning application, we conceptualized four distinct games that employ interaction with a social robot, along with a free conversation mode. These games were designed to fully exploit the functionalities offered by the framework, facilitating a rich and engaging interaction between the robot and the user. During the games, the robot displays facial expressions of joy, surprise, and sadness in response to the interaction’s outcomes. For instance, the robot assumes a sad expression when it loses a game, and exhibits surprise when the human participant makes a correct guess.

The first language learning game is called “Guess Game”. In this game, the virtual robot serves as the respondent while the human user assumes the role of the questioner. To begin, the robot selects an object from a predefined collection of twelve diverse items situated on a virtual table, using a uniform distribution. These objects, representative of everyday items, range in color, shape, texture, and function. Following the selection, the user is tasked to discern the identity of the chosen object through a series of inquiries, which the robot must respond to accurately and informatively. The questions, dictated by the user, are designed to be open-ended, facilitating a more comprehensive understanding of the English language as the user frames questions and processes responses. The goal is to promote active engagement and inquiry-based learning, by encouraging the user to formulate strategic questions in English. The interaction sequence of how the first game generally works can be seen in Fig. 2.



**Fig. 2** First game: Guess Game. (a) User initiates the game. (b) Objects are displayed, and the robot outlines the game rules. (c) User examines the objects and formulates questions for the robot. (d) Robot responds to the user’s inquiries

The second game, “Reverse Guess Game”, presents a dynamic shift in roles, designed to provide a different but equally engaging language learning experience. In this variant of the first game, the human participant selects one of the objects, and the robot takes on the task of guessing the selected item. Through an interactive and iterative process, the robot crafts open-ended questions, employing its language processing abilities to understand and interpret the human’s responses. The goal of the robot is to correctly identify the chosen object based on the human’s answers.

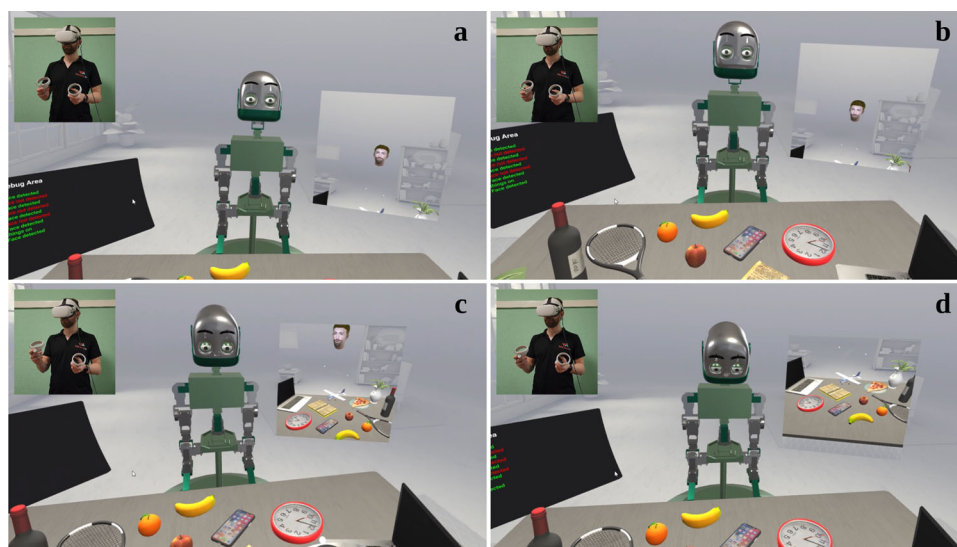
This inversion of roles provides a distinct linguistic focus. While the first game emphasizes speaking practice, the second game concentrates more on listening comprehension. The player listens to the robot’s inquiries and responds appropriately, fostering an understanding of question structure and improving language input processing. The interaction sequence of how the second game generally works can be seen in Fig. 3. It’s worth noting that Jubileo employs the YOLO (You Only Look Once) system when making a guess, adding a layer of interactive sophistication to the game dynamics.

The third game, “Closed Eyes Game”, engages the player in an exercise that tests their descriptive abilities. In this game, the robot “closes its eyes”, and a chosen object materializes before it. Unaware of the object’s identity, the robot relies solely on the user’s verbal descriptions to infer its characteristics. The human user is prompted to describe the object using the vocabulary at their disposal, practicing their ability to articulate thoughts in English. As the game progresses, the robot may interject with requests for additional information or clarification, thereby prompting the user to expand their descriptions and experiment with different descriptive

phrases. This interactivity encourages the user to think critically about the object, enhancing their ability to convey nuanced information. Once the robot has gathered sufficient information to make an educated guess, it identifies the object. Subsequently, the robot introduces more complex vocabulary and phrases related to the object, extending the user’s language repertoire. The interaction sequence of how the third game generally works can be seen in Fig. 4.

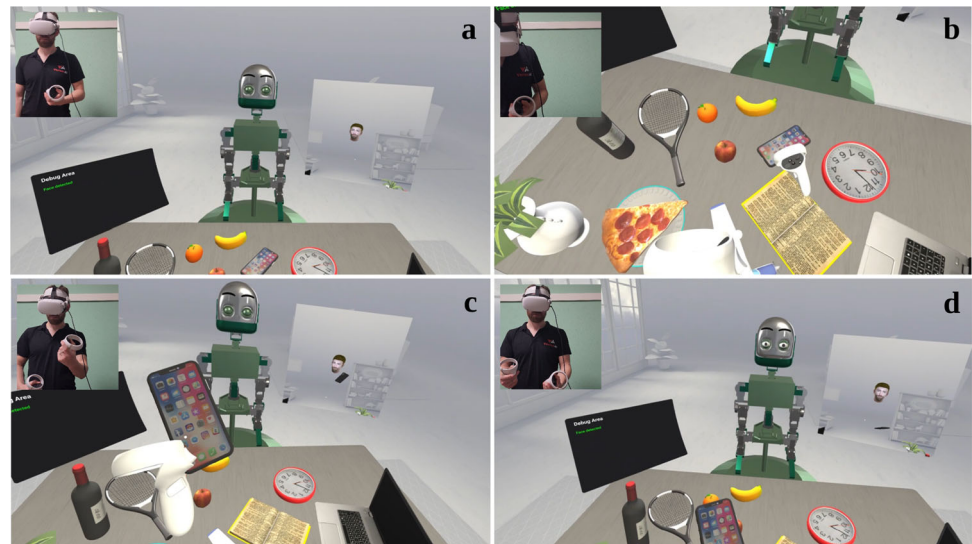
The fourth game, “Countries Game”, unfolds in the realm of cultural exploration and geographical knowledge, introducing an element of worldly curiosity into the language learning process. Here, the robot takes on the role of a travel guide, selecting a country at random and then providing intriguing snippets of information about that country. These clues can pertain to a wide array of cultural aspects, including but not limited to, traditional cuisine, population statistics, famous landmarks, historical facts, and unique customs. With each provided clue, the human user attempts to guess the country in question. The interaction sequence of how the fourth game generally works can be seen in Fig. 5.

As the game progresses, the robot gradually reveals more about the country in question, effectively guiding the user on a virtual tour. If the user’s guess is incorrect, the robot responds by divulging another fact about the country, adding to the user’s knowledge and further hinting at the country’s identity. This game not only expands the user’s English vocabulary, especially in the domains of culture and geography but also enhances their listening comprehension skills. The user is challenged to process the information provided by the robot, contextualize it, and make educated guesses. By intertwining language learning with cultural education, this



**Fig. 3** Second game: Reverse Guess Game. (a) Robot provides game instructions. (b) User answers the questions formulated by the robot. (c) Robot makes a guess and initiates the object detection system. (d) Robot observes the guessed object

**Fig. 4** Third game: Closed Eyes Game. (a) Robot provides game instructions and closes its eyes. (b) User selects an object for the robot to identify. (c) User describes the chosen object to the robot. (d) Robot reacts after making a correct guess about the object



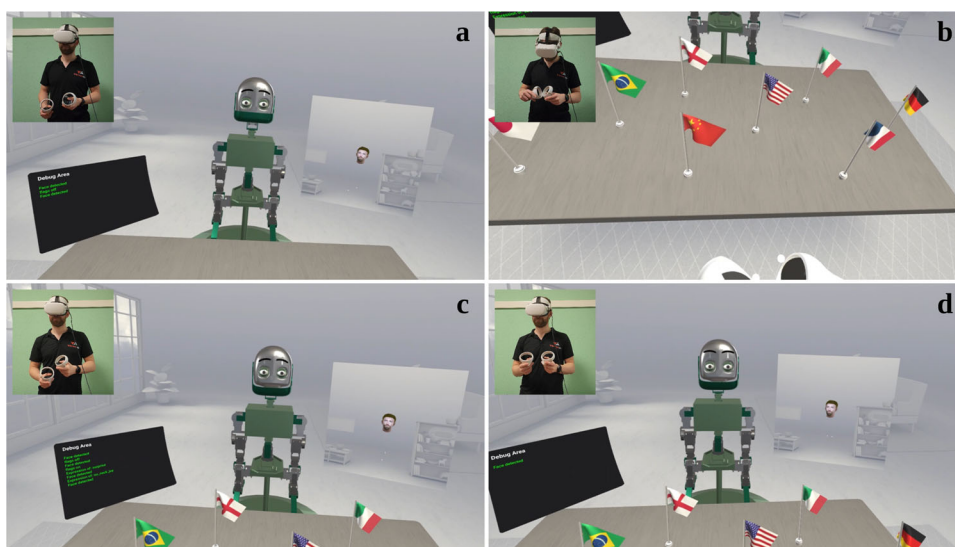
game enriches the user's understanding of the world while developing their English language proficiency.

In addition to the games themselves, our system includes an interface that enables free conversations and permits the user to choose and initiate games. To select a game, the user engages in a natural language conversation with the LLM agent. The user is free to express their game preference by simply requesting the agent to initiate the desired game. This design not only makes the game selection process smooth and intuitive but also enhances the overall interactivity of the system. It's worth noting that the user is not limited to game-related discussions with the LLM agent. The user may engage in varied conversations with the agent about a broad range of topics, thus making use of the LLM's

extensive language understanding and generation capabilities. This feature adds an additional layer of versatility to our system, fostering a more comprehensive and engaging language learning experience. The conversation can be initiated at the user's discretion, allowing them to converse and learn at their own pace before deciding to commence a game.

## 4 Results and Discussion

In the process of implementing our proposed methodology, numerous fascinating and noteworthy insights emerged that extend beyond the realm of English language acquisition. The results, albeit preliminary, reinforce the transformative



**Fig. 5** Fourth game: Countries Game. (a) User chooses to play the game. (b) VR environment adapts for the game. (c) Robot provides clues about the selected country. (d) Upon the user's correct guess, the robot reacts accordingly



potential of this innovative convergence of technologies, with significant implications for broader educational applications.

#### 4.1 Examining Human-Robot Game Interactions Outcomes

All the games proposed in our HRI setting underwent testing to assess the performance of the Jubileo robot. The test sessions can be viewed through the link provided in Section of Supplementary Material. The first game that was tested to analyze the interaction between the tester and the virtual robot was the Guess Game. “Guess Game” involves Jubileo acting as a respondent to the human user’s queries. It starts with the Jubileo choosing an object from a set, after which the human tries to figure out what the object is via querying. The Jubileo’s replies help the human construct a mental representation of the unknown target word. The interactions between the human and the robot are demonstrated below:

##### FIRST GAME: GUESS GAME

**Tester:** Okay, let’s play the first game.

**Jubileo:** Let’s play the game one. This game is called guess game. Can you see that there are some objects on the table? So, I will choose an object and you will make me questions in order to figure out what are the object. I will answer with no or yes and when you feel confident try to guess

**Tester:** Is it something edible?

**Jubileo:** Do you want to know if the object is safe to eat? The answer is yes

**Tester:** Is it a fruit?

**Jubileo:** Do you want to know if the object is a fruit? The answer is yes

**Tester:** Is it yellow?

**Jubileo:** Do you want to know if the object is yellow? The answer is yes

**Tester:** Is it a banana?

**Jubileo:** Do you want to know if the object is the banana? The answer is yes. You are correct. Congratulations, let’s play again? I already choose other object...

In the transcript of the interaction between Jubileo in the First Game, it is evident that the tester was able to interact with the robot seamlessly. The robot effectively communicated when the tester’s guess was correct, establishing a clear feedback mechanism while displaying a facial expression of surprise. This demonstrates the capability of our model to select an object in the environment and subsequently respond to the instructions and statements provided by the tester through the LLM.

The next is the “Reverse Guess Game”, where the human selects an object, and Jubileo must ask questions to determine what the chosen item might be. This game tests not just the Jubileo’s ability to generate contextually meaningful

questions, but also its ability to evaluate the gathered information in order to identify the object. Notably, Jubileo enhances the interactive dynamic by employing its object detection system; every time it forms a guess, it visually aligns its focus toward the object in question. This visual engagement simulates a more human-like interaction and offers an enriched HRI experience. The exchange between the human participant and Jubileo during the second game is outlined below:

##### SECOND GAME: REVERSE GUESS GAME

**Tester:** Let’s play the second game

**Jubileo:** Ready for game two? This time you will choose an object from the table and I will be the one asking the questions. I will try to guess the object based on your responses. Remember, you must answer with either yes or no. Let’s start when you’re ready.

**Tester:** Okay, I’m ready.

**Jubileo:** Is the object rectangular in shape?

**Tester:** No, it isn’t.

**Jubileo:** Is the object round?

**Tester:** No, it isn’t.

**Jubileo:** Is the object a toy?

**Tester:** Yes, it is.

**Jubileo:** Is the object a toy airplane?

**Tester:** Yes, you’re correct well done.

In the Second Game, the tester plays a crucial role in guiding Jubileo’s guessing process by answering its inquiries. The participant does not reveal the answer explicitly but confirms or negates the robot’s hypotheses through simple “yes” or “no” responses. This interaction requires Jubileo to generate effective questions and interpret the tester’s responses to identify the selected object. Consequently, the game emphasizes the robot’s ability to engage in a problem-solving approach that resembles complex human-to-human interactions.

The Third Game appears to involve engaging participants in an exercise designed to enhance their descriptive abilities. “Closed Eyes” is similar to the second game, but requires the human to describe something instead of just answering yes or no questions. In this game, the human picks up an object. Since the robot is with its eyes closed and possesses no prior knowledge of the object’s identity, it must rely exclusively upon the verbal descriptions provided by the human user to determine the object’s attributes. The interactions between the human and robot in the third game are demonstrated below:

##### THIRD GAME: CLOSED EYES GAME

**Tester:** Let’s play the game where you close your eyes.

**Jubileo:** Game Three is all about your description skills. I will close my eyes, and you will pick and describe an object to me. As I won’t be able to see it, try to provide as much detail as you can, and I will guess the object



based on your description. Are you ready to get started?

**Tester:** Yes, I'm ready.

**Jubileo:** I'm shutting my virtual eyes now. Go ahead and choose an object, then share some details about it.

**Tester:** It is small.

**Jubileo:** I need more information to make a guess. Can you provide any other details about the object?

**Tester:** It is black.

**Jubileo:** I believe the object is the black iPhone smartphone. Is that correct?

**Tester:** Yes, congratulations!

**Jubileo:** Yeah, I knew it. Choose another object!

The Third Game, where the robot closes its eyes, introduces a distinctive challenge for players. It compels the tester to employ creativity and descriptive language to assist the robot in identifying objects based solely on their physical properties. This activity not only evaluates the user's proficiency in expressing ideas clearly but also fosters critical thinking and attentiveness to detail. The demonstration showcases Jubileo's ability to successfully accomplish the task with the descriptions provided by the tester.

The last game is called "Countries Game", which focuses on cultural education. In this version, Jubileo takes on the role of a travel guide, discussing countries randomly selected from various regions around the world. The objective is to foster comprehensive language acquisition through simulated explorations and discoveries. Here, the robot serves as a virtual travel companion, providing insights into different nations. As each participant makes a guess about the specific location being discussed, the robot provides additional hints and background information about the chosen country. The interactions between the human and robot during the fourth game are demonstrated below:

#### **FOURTH GAME: COUNTRIES GAME**

**Tester:** Let's play Game Four.

**Jubileo:** Time for Game Four! In this game, I'll be your virtual travel guide. I'll pick a random country and provide you with some interesting facts about it. Your task is to guess the country based on these clues. It's a great way to learn more about the world while practicing your English. Let's get started.

**Tester:** Okay, you can start.

**Jubileo:** First curiosity: This country has a rich history of indigenous peoples and is home to the Inuit, First Nations, and Métis peoples.

**Tester:** Can you give me one more tip?

**Jubileo:** Second curiosity: It is home to many diverse cultures and has both English and French as official languages.

**Tester:** Is it Canada?

**Jubileo:** You are correct! Other curiosities about Canada: The Canadian Rockies are a stunning mountain range

that is popular for skiing and other outdoor activities. The iconic beaver is the national animal of this country.

In the Fourth Game, the robot takes on the role of a virtual travel companion, introducing the tester to a new country and presenting interesting facts for the tester to guess. This game serves as a valuable tool for expanding participants' knowledge and understanding of the world around them. To determine the correct answer, participants must rely on contextualized information and employ deductive reasoning. Jubileo actively promotes active listening and observation skills, enhancing conversational fluency throughout the interaction. Moreover, Jubileo provides additional hints and clues throughout the game upon the tester's request. The robot demonstrated its ability to sustain engaging conversations and provide assistance effectively during this game.

When the user can not provide the correct answer after five curiosities are given, Jubileo responds with encouragement and support. He acknowledges the fact that not everyone might know the right answer immediately and reassures the participant that there will be other chances to correctly guess future countries. He also checks if the student is prepared for another round of questioning before proceeding to select a different country. This approach helps to promote a positive environment where mistakes are accepted and seen as part of the learning process. The interactions between the human providing an incorrect answer and the robot are demonstrated below:

#### **CASE OF WRONG ANSWER**

**Tester:** I don't know the answer.

**Jubileo:** The country was Greece, but don't worry. You'll get the next one. Are you ready?

Lastly, the free conversation interface, as a complementary component to the games, offers an opportunity for users to engage in diverse and spontaneous discussions with the LLM agent. This unrestricted conversation enriches the user's language learning experience, exposing them to a wider range of conversational contexts and encouraging spontaneous language use. The exchange between the human participant and Jubileo during the second game is outlined below:

#### **FREE CONVERSATION**

**Tester:** Can you talk a little about Japanese culture?

**Jubileo:** Sure! Japanese culture is known for its unique traditions and customs. Some of the most popular aspects include anime, manga, and video games. The country is also famous for its cuisine, such as sushi and ramen. Japanese art, including calligraphy and origami, is highly regarded. Additionally, the country has a strong emphasis on respect and honor, which is reflected in its social and business practices. Is there anything specific you would like to know more about?

**Tester:** What is the capital of Japan?

**Jubileo:** The capital of Japan is Tokyo.

## 4.2 Discussion

These game-based approaches for language learning were created to demonstrate our framework bringing the social robot into an interactive and immersive education setting. We showed that in our tests, presented in Section 4.1, the participant is freed from the often rigid structure of traditional classroom learning and is able to explore language learning in a dynamic, self-paced manner. The immediate feedback and adaptive conversations facilitated by the integration of GPT-4 into the virtual robot provided a responsive learning environment, promoting active learning and user engagement. This suggests a profound shift from the conventional passive modes of learning, pushing towards a more learner-centric approach.

The use of VR and HRI in conjunction adds a new dimension to this methodology. The immersive VR environment simulates real-world scenarios, giving language learning context, relevance, and immediacy that traditional methods might lack. This approach could be adapted to various other learning scenarios, such as science experiments, historical explorations, or cultural studies, providing learners with firsthand virtual experiences to complement their theoretical knowledge.

Additionally, the deployment of a humanoid robot within the VR environment added a sense of social presence, fostering a unique interactive experience. The robot's human-like demeanor, powered by the large language model, created a "safe" space for users to practice their language skills without fear of judgment or criticism, an often-overlooked aspect of language learning. This indicates that such technology can play a significant role in reducing learner anxiety and promoting self-confidence, not only in language learning but also in broader educational contexts.

Moreover, the scalability and cost-effectiveness of this solution deserve attention. As the platform operates in a VR environment with a virtual robot, it circumvents the traditional barriers of physical resources, geographical location, and personal tutoring costs. This suggests a future where high-quality, personalized education could be widely accessible to learners regardless of their location or economic background. However, our framework being based on ROS and having realistic control and sensing interfaces, allows a seamless transition to a real social robot.

The novel application of a large language model like GPT-4 in an educational setting underscored its potential to foster human-like interactions. The language model's ability to generate coherent, contextually relevant responses can

offer learners a more realistic conversational experience. The scope of these language models could extend to various pedagogical settings, including personalized tutoring systems, scenario-based learning, and problem-solving exercises.

While the integration of GPT-4 in human-robot interactions and language learning offers significant advancements in generating human-like responses, it also presents ethical and practical challenges that warrant discussion. One notable issue is the phenomenon of "hallucinations," where the model may generate misleading or inaccurate information, a concern especially pertinent in educational settings. Additionally, the model may produce content that, although not explicitly censored, could be deemed inappropriate or sensitive, raising ethical questions about content filtering responsibilities. Therefore, a comprehensive assessment of these challenges is essential for the safe and effective application of GPT-4 in human-robot interactions and language learning.

Another pivotal concern centers around the integration of VR into English language learning. A noteworthy limitation is the absence of direct human interaction in VR technology, a fundamental component of traditional language teaching approaches. However, VR can serve as a valuable supplement to conventional methods or prove indispensable in situations where factors such as geographical constraints, financial considerations, or scheduling complexities render traditional English classes unfeasible. Other than that, VR remains somewhat inaccessible to the general public due to its frequently exorbitant costs. Nevertheless, there is a promising outlook for the future, with growing accessibility and awareness anticipated in the coming years. Additionally, VR technology can present technical issues that need to be addressed. These issues include unstable internet connections, large storage requirements, and audio difficulties with speech recognition, all of which can worsen the overall learning experience.

Finally, although this study focuses on English language learning, the methodology and results have broader implications. The integration of VR, HRI, and large language models represents a potent, scalable, and versatile tool for modern education. Further exploration and continued refinement of this approach can pave the way for new educational paradigms, extending the reach and improving the quality of learning experiences across various domains.

## 5 Conclusions

In this work, we have introduced a VR simulation framework for Human-Robot Interaction (HRI) research and development, extending our previous contributions. We

demonstrated how human-robot interactions can be tested and refined in a ROS-based simulation environment. This includes realistic control of servo-actuators and computer vision capabilities, thereby facilitating a smoother transition to more costly and time-consuming real robot implementations. To validate this tool, we presented four distinct yet complementary games aimed at interactive language practice through immersive human-robot interactions.

While we believe that the integration of technologies like GPT-4 and VR simulations signals a new era of interactive and immersive learning experiences, it is important to outline the steps to achieve this vision. Future work could focus on several key areas.

First, although facial expressions are currently tied to specific stages of the games, employing sentiment analysis could enable Jubileo to exhibit dynamic facial expressions or even head movements during free talk mode or in future applications. This would significantly enrich the social aspect of the robot.

Second, while the language games serve primarily as a demonstration of the general-purpose framework, implementing automatic pronunciation error detection could be a valuable addition. Jubileo could then offer real-time corrections, asking users for their preference in receiving such feedback.

Finally, the potential of integrating a large language model like GPT-4 in an educational setting is significant, particularly for fostering human-like interactions. However, this also raises ethical and practical challenges, such as content filtering and the risk of generating misleading or inaccurate information. Addressing these challenges is essential for the safe and effective application of such technologies. By focusing on these areas, we lay the groundwork for further refinement and expansion of this work in the future.

**Acknowledgements** The authors would like to thank the VersusAI team. This work was partly supported by the University of Tsukuba.

**Author Contributions** Jair Augusto Bottega conceived the research, writing of the article, designed and program the experiments, collected and processed the test data. Victor Augusto Kich write the article, program the experiments, collected and processed the test data. Junior Costa de Jesus write the article, program the experiments, collected and processed the test data. Raul Steinmetz write the article, program the experiments, collected and processed the test data. Alisson Henrique Kolling write the article, program the experiments, collected and processed the test data. Ricardo Bedin Grando write the article, collected and processed the test data. Rodrigo da Silva Guerra discussion and conception of the main ideas of the article, provided valuable comments. Daniel Fernando Tello Gamarra conceived the research, writing of the article and discussion of the main ideas of the article.

**Funding** Not applicable.

**Available of data and material** Available in GitHub <https://github.com/jajaguto/jubileo>.

## Declarations

**Ethical Approval** The article has the approval of all the authors.

**Consent to Participate** All the authors gave their consent to participate in this article

**Consent to Publish** The authors gave their authorization for the publishing of this article.

**Competing interest** There are not conflict of interest or competing interest

## References

1. Goodrich, M., Schultz, A.: Human-robot interaction: a survey. *Foundations and Trends in Human-Computer Interaction*. **1**, 203–275 (2007). <https://doi.org/10.1561/1100000005>
2. Terzioğlu, Y., Mutlu, B., Şhin, E.: Designing social cues for collaborative robots: the role of gaze and breathing in human-robot collaboration. In: *Proceedings of the 2020 ACM/IEEE International Conference on Human-robot Interaction*, pp. 343–357 (2020)
3. Cao, H.-L., Scholz, C., De Winter, J., Makrini, I.E., Vanderborght, B.: Investigating the role of multi-modal social cues in human-robot collaboration in industrial settings. *International Journal of Social Robotics*, 1–11 (2023)
4. Pennisi, P., Tonacci, A., Tartarisco, G., Billeci, L., Ruta, L., Gangemi, S., Pioggia, G.: Autism and social robotics: a systematic review. *Autism Res*. **9**(2), 165–183 (2016)
5. Breazeal, C., Dautenhahn, K., Kanda, T.: *Social robotics. Springer handbook of robotics. 1935–1972* (2016)
6. Ishiguro, H.: *Android science. Cognitive neuroscience robotics: a synthetic approaches to human understanding. 193–234* (2016)
7. Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., Ogino, M., Yoshida, C.: *Cognitive developmental robotics: A survey. IEEE Trans. Auton. Ment. Dev.* **1**(1), 12–34 (2009)
8. Akkaya, I., Andrychowicz, M., Chocie, M., Litwin, M., McGrew, B., Petron, A., Paino, A., Plappert, M., Powell, G., Ribas, R., et al.: Solving rubik's cube with a robot hand (2019). [arXiv:1910.07113](https://arxiv.org/abs/1910.07113)
9. Monteiro, F.F., Vieira, A.L.B., Teixeira, J.M.X.N., Teichrieb, V., et al.: Simulating real robots in virtual environments using nvidia's isaac sdk. In: *Anais Estendidos do XXI Simpósio de Realidade Virtual e Aumentada*. pp. 47–48 (2019). SBC
10. Tan, J.T.C., Inamura, T.: Sigverse-a cloud computing architecture simulation platform for social human-robot interaction. In: *IEEE International Conference on Robotics and Automation*. pp. 1310–1315 (2012). IEEE
11. Liu, O., Rakita, D., Mutlu, B., Gleicher, M.: Understanding human-robot interaction in virtual reality. In: *26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. pp. 751–757 (2017). IEEE
12. Bottega, J.A., Kich, V.A., Kolling, A.H., Dyonisio, J.D., Corçaque, P.L., Guerra, R.d.S., Gamarra, D.F.: Jubileo: An open-source robot and framework for research in human robot social interaction. In: *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*. pp. 149–154 (2022). IEEE
13. Bottega, J.A., Steinmetz, R., Kolling, A.H., Kich, V.A., De Jesus, J.C., Grando, R.B., Gamarra, D.F.T.: Virtual reality platform to develop and test applications on human robot social interaction. In: *2022 Latin American Robotics Symposium (LARS), 2022*

- Brazilian Symposium on Robotics (SBR), and 2022 Workshop on Robotics in Education (WRE), pp. 1–6 (2022). IEEE
14. Zhao, W.X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., et al.: A survey of large language models. (2023). [arXiv:2303.18223](https://arxiv.org/abs/2303.18223)
  15. Wang, B., Li, G., Li, Y.: Enabling conversational interaction with mobile ui using large language models. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. pp. 1–17 (2023)
  16. Ross, S.I., Martinez, F., Houde, S., Muller, M., Weisz, J.D.: The programmer's assistant: Conversational interaction with a large language model for software development. In: Proceedings of the 28th International Conference on Intelligent User Interfaces. pp. 491–514 (2023)
  17. Godwin-Jones, R.: Games in language learning: opportunities and challenges (2014)
  18. Klimova, B., Kacet, J.: Efficacy of computer games on language learning. Turkish Online Journal of Educational Technology-TOJET. **16**(4), 19–26 (2017)
  19. al, M.: The uncanny valley [from the field]. IEEE Robotics & Automation Magazine. **19**(2), 98–100 (2012)
  20. Sherwani, F., Asad, M.M., Ibrahim, B.S.K.K.: Collaborative robots and industrial revolution 4.0 (ir 4.0). In: 2020 International Conference on Emerging Trends in Smart Technologies (ICETST). pp. 1–5 (2020). IEEE
  21. Goodrich, M.A., Schultz, A.C., et al.: Human–robot interaction: a survey. Foundations and Trends® in Human–Computer Interaction. **1**(3), 203–275 (2008)
  22. Sheridan, T.B.: Human–robot interaction: status and challenges. Hum. Factors **58**(4), 525–532 (2016)
  23. Yanco, H.A., Drury, J.L.: A taxonomy for human-robot interaction. In: Proceedings of the AAAI Fall Symposium on Human-robot Interaction. pp. 111–119 (2002)
  24. Faraj, Z., Selamet, M., Morales, C., Torres, P., Hossain, M., Chen, B., Lipson, H.: Facially expressive humanoid robotic face. HardwareX. **9**, 00117 (2021)
  25. Walters, M.L., Syrdal, D.S., Dautenhahn, K., Te Boekhorst, R., Koay, K.L.: Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. Auton. Robot. **24**, 159–178 (2008)
  26. Zlotowski, J.A., Sumioka, H., Nishio, S., Glas, D.F., Bartneck, C., Ishiguro, H.: Persistence of the uncanny valley: the influence of repeated interactions and a robot's attitude on its perception. Front. Psychol. **6**, 883 (2015)
  27. Žlajpah, L.: Simulation in robotics. Math. Comput. Simul. **79**(4), 879–897 (2008)
  28. Datteri, E., Schiaffonati, V.: Robotic simulations, simulations of robots. Mind. Mach. **29**, 109–125 (2019)
  29. Nascimento, L.M.d., Neri, D.S., Ferreira, T.d.N., Pereira, F.d.A., Albuquerque, E.A.Y., Gonçalves, L.M.G., Sá, S.T.d.L.: Sbotics-gamified framework for educational robotics. Journal of Intelligent & Robotic Systems. **102**(1), 17 (2021)
  30. Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., Tanaka, F.: Social robots for education: a review. Sci. Rob. **3**(21) (2018)
  31. Toh, L.P.E., Causo, A., Tzuo, P.-W., Chen, I.-M., Yeo, S.H.: A review on the use of robots in education and young children. Journal of Educational Technology & Society. **19**(2), 148–163 (2016)
  32. Kim, J.-O., Kim, J.: Development and application of art based steam education program using educational robot. In: Robotic Systems: Concepts, Methodologies, Tools, and Applications. pp. 1675–1687. IGI Global, ??? (2020)
  33. Zheng, J., Chan, K., Gibson, I.: Virtual reality. Ieee. Potentials **17**(2), 20–23 (1998)
  34. Chen, Y.-L., Hsu, C.-C.: Self-regulated mobile game-based english learning in a virtual reality environment. Computers & Education. **154**, 103910 (2020)
  35. Shim, K.-C., Park, J.-S., Kim, H.-S., Kim, J.-H., Park, Y.-C., Ryu, H.-I.: Application of virtual reality technology in biology education. J. Biol. Educ. **37**(2), 71–74 (2003)
  36. Bogusevschi, D., Muntean, C., Muntean, G.-M.: Teaching and learning physics using 3d virtual learning environment: a case study of combined virtual reality and virtual laboratory in secondary school. J. Comput. Math. Sci. Teach. **39**(1), 5–18 (2020)
  37. Shahab, M., Taheri, A., Mokhtari, M., Shariati, A., Heidari, R., Meghdari, A., Alemi, M.: Utilizing social virtual reality robot (v2r) for music education to children with high-functioning autism. Education and Information Technologies, 1–25 (2022)
  38. Ye, Y., You, H., Du, J.: Improved trust in human-robot collaboration with chatgpt (2023). [arXiv:2304.12529](https://arxiv.org/abs/2304.12529)
  39. Christmann, G.H.G.: Um Jogo de Interação Humano-Robô para o Ensino de Língua Inglesa. Bachelor Thesis, UFSM. Advisor: Prof. Dr. Rodrigo S. Guerra (2019). <http://rodrigoguerra.com/wp-content/uploads/2019/02/tccrevisado.pdf>
  40. Quigley, M., Faust, J., Foote, T., Leibs, J., et al.: Ros: an open-source robot operating system (2009)
  41. Juliani, A., Berges, V.-P., Teng, E., Cohen, A., Harper, J., Elion, C., Goy, C., Gao, Y., Henry, H., Mattar, M., Lange, D.: Unity: A General Platform for Intelligent Agents (2020)
  42. Jocher, G., Stoken, A., Borovec, J., NanoCode012, ChristopherSTAN, Changyu, L., Laughing, tkianai, Hogan, A., lorenzomamma, yxNONG, AlexWang1900, Diaconu, L., Marc, wanghaoyang0106, ml5ah, Doug, Ingham, F., Frederik, Guilhen, Hatovix, Poznanski, J., Fang, J., Yu, L., changyu98, Wang, M., Gupta, N., Akhtar, O., PetrDvoracek, Rai, P.: ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements. Zenodo (2020). <https://doi.org/10.5281/zenodo.4154370>
  43. OpenAI: GPT-4 Technical Report (2023)
  44. Google: Google Cloud Speech-to-Text API. <https://cloud.google.com/speech-to-text> (2023)
  45. Google: Google Cloud Text-to-Speech API. <https://cloud.google.com/text-to-speech> (2023)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

**Jair Augusto Bottega** is pursuing his Master's in Computer Science at the University of Tsukuba. He earned his degree in Control and Automation Engineering from the Federal University of Santa Maria and is also a Computer Technician from the Federal Institute of Education, Science, and Technology. His research interests include applying large language models to social and mobile robots.

**Victor Augusto Kich** is pursuing his Master's in Computer Science at the University of Tsukuba. He obtained his degree in Computer Engineering from the Federal University of Santa Maria and is also a Computer Technician from the Federal Institute of Education, Science, and Technology. His areas of interest encompass autonomous systems and deep reinforcement learning.



**Junior Costa de Jesus** is a Ph.D. student in Computer Science at the Korea Advanced Institute of Science & Technology (KAIST). He holds a Master's in Computer Engineering from the Federal University of Rio Grande and a degree in Control and Automation Engineering from the Federal University of Santa Maria. His research interests lie in multi-agent systems and mobile robots.

**Raul Steinmetz** is a computer science graduate student at Universidade Federal de Santa Maria. Primarily focuses his research on applying deep learning to robotics, specifically deep reinforcement learning, and using computer vision for agricultural field monitoring through aerial images.

**Alisson Henrique Kolling** is a student of Master in Computer Engineering at Federal University of Rio Grande and Control and Automation Engineering at Federal University of Santa Maria. His interests are mobile robotics and deep reinforcement learning.

**Ricardo Bedin Grando** is a professor of Robotics at the Technological University of Uruguay. He is currently pursuing his Ph.D. in Computer Engineering at the Federal University of Rio Grande, where he also earned his Master's in Computer Engineering. Additionally, he holds a degree in Computer Engineering from the Federal University of Santa Maria. His research interests include mobile robotics and reinforcement learning.

**Dr. Rodrigo da Silva Guerra** is currently an Associate Professor at Centro de Ciências Computacionais, FURG/Brazil. He obtained his Bachelor's degree in Control and Automation Engineering from PUCRS/Brazil in 2001, followed by a Master's degree in Computer Vision from UFRGS/Brazil in 2004. In 2008, Dr. Guerra completed his Ph.D. in Robotics at Osaka University, after which he conducted research in Japan until 2011. He then pursued a postdoctoral position at UFRGS/Brazil until 2012 and served as an Assistant Professor at UFSM/Brazil until 2022. In 2016, he co-founded Qiron Robotics, a company focused on designing social robots for various applications. In 2019, Dr. Guerra was a Visiting Professor at the National Taiwan Normal University. He has expertise in developing intelligent robots with a focus on cognitive and social robots.

**Dr. Daniel Fernando Tello Gamarra** is graduated in Mechanical Engineering - Universidad Nacional Del Centro Del Perú (1998), and Master's in Electrical Engineering from the Federal University of Espírito Santo (2004), Master's in Informatics from Sussex University (2006) in England, Doctorate in Informatics from Scuola Superiore Sant'Anna in Pisa Italy. Sandwich Doctorate from Sussex University (2006). He has experience in robotics and has been developing research in the following areas: Robotics; Neural Networks; Machine Learning; Computational Vision; Artificial Intelligence; Intelligent Control, with emphasis on Engineering.