# WALL-E: An Autonomous Educational Robot for Interactive Learning at Home

KOTHAREDDY VAMSHIDHAR REDDY
*Computer Science & Engineering Presidency University*
Bengaluru, India

ABHIRAM PB
*Computer Science & Engineering Presidency University*
Bengaluru, India

YASH RAJ DUTT
*Computer Science & Engineering Presidency University*
Bengaluru, India

MR. TAMILSELVAN T
*Assistant Professor*
*Presidency School of Computer Science and Engineering*
*Presidency University*

***Abstract- This is a detailed report on Wall-E, a low-cost semi-autonomous educational robot, which helps children learn at home without interacting with the human teacher. Wall-E takes advantage of cheap mechanical movement, stereo sight, face recognizing, hand gesture recognizing, and conversational AIs in facilitating personalized learning among children. We address the motivation of an educational robot at home, and then summarize the recent research in the fields of social robots, long-term human-robot relationship, human-robot-collaborative tutoring (human-robot personal tutor) systems, and multimodal interaction to inform the focus on Wall-E in this report. The report also explains the mechanical design and electronics of Wall-E, Wall-E streaming and control architecture, YOLO used to identify objects and faces, InsightFace used to recognize by memory and a light and standard language model used to generate and answer questions. We also present the use of gesture recognition in MediaPipe and a simple neural net, obstacle avoiding with sensor fusion and potential field path planning and a vector database of a storage and retrieval of facial embeddings. The initial tests of the system indicate that the system can work in real time, it is easy to interact with, and it is capable of engaging users. Lastly, we comment on the ethics, the long-term adoption and potential future directions of research.***

***Keywords: Educational Robots, Interactive Learning, Raspberry Pi, Face Recognition, YOLO, Large Language Models.***

## I. INTRODUCTION

*M. P. Kaur et al. [1], M. Belpaeme et al. [14] and C. Breazeal et al. [15]* The world is witnessing the growth of educational robots, and this indicates the growing attitude of the possibility of using interactive technology to supplement human instruction, in addition to evoking curiosity and using an individual approach to learning. In more instances, social robots are applied in classrooms and households in narrating stories, lessons and helping children practice language skills. As per various studies carried out nowadays, children become emotionally attached to robots very quickly; they perceive them as intelligent creatures

Academic sources claim that children can also be taught to be resilient and develop resilience mindsets when a robot that simulates positive coping stereotype, encouragement and feeling identification is taught. Robots can provide body with corporeality of interaction to children, which is capable of creating socio-emotions, as opposed to all-digital learning environments. However, the problem of privacy, emotional addiction and immorality of child-centred robots remains.

*S. Hussain et al. [2]* The project presents a home tutor robot called Wall-E (the home tutor) on the basis of Pixar WALL-E) that is likely to provide affordable and personalized home-based education to children. The robot will be semi-autonomous to avoid obstacles and track faces, identify individual learners and hold a teaching session in the form of a conversation, basing on a lightweight language model. Wall-E is also able to react and provide instructions, albeit its answers and commands would be filtered by intuitive hand movements (open palm to activate, closed palm to deactivate), glances and inquiries, depending on the performance. Wall-E is a combination of computer vision and engineering with AI.

## II. LITERATURE REVIEW

### A. EDUCATIONAL ROBOTICS IN PRACTICE

*F. Pelizzari et al. [3] and H. Wainer et al. [17]* The literature on the topic of robotics in the hospital and home education implies that digital technologies have gained an increasingly closer role in the process of the educational continuity of the students who cannot study due to an illness condition. The authors analyzed 30 articles that were published in the last 10 years and there it is stated that robotics contributes to the development of engagement, social isolation, and the ability to solve problems, be creative, and become a leader. It also brings physical action into the learning process therefore provides multi-dimensional and experiential activities to the youth. The use of coding and educational robotics proved useful to build computational thinking and soft skills, but there are some gaps in regard to the needs of the teenagers and the long-term effects of education. The introduction of robots into home-based education is preconditioned by this review and provides precious hints regarding the creation of the flexible and individual instruction.

## B. LONG-TERM USE AND LIFECYCLE

*Z. Zhao et al. [4]* Although the majority of educational robots are expected to work in one cycle, the recent studies encompass the relations between families and longer robots than the given field. Another study, a follow-up, was carried out in 2025 visiting 19 families that used a reading companion robot in preschool with the child in 2021. Children were too large to utilize all the instruction content of such a robot, but 18 families did. The robot was not an educational machine but a member of the household or the family. Caregivers also documented attachment and caring perceptions towards the robot, like nostalgia. The contributors of this paper identified the emergence of three significant themes, which included personification, symbolic value and practical repurposing. Based on those observations, the authors indicated that the life cycle of the social robot, like retirement and repurposing should be considered by the designers. The results are applied in designing Wall-E by being very considerate of long-term construction, data portability and ethical decisions in case Wall-E is decommissioned.

## C. PERSONALIZED AND AFFECTIVE TUTORING

Intelligent tutoring systems (ITS) suggested by *N. Maaz et al. [5] and J. Kennedy et al. [16]* presuppose the need of personal and affective support. The CARE framework proposes to monitor the cognitive and affective status of the students continuously to customize learning materials, depending on the markers, such as face and performance markers, which is also applied to predict the learning state using an Extreme Gradient Boosting classifier (XGBoost). The research hypothesis goes on to postulate that the correspondence of concurrent and cognitive evaluation in robotics would escalate the scholarly support with a critical customization of provisions by the faculty in an attempt to alleviate the diverse learning constructs. The review of social robot shows that it can be possible to have proactive tutoring based on affective states and thus may keep the engagement but result in over-proactivity that will influence trust.

The experiments conducted with NAO robots as peer learners demonstrate that personalization of behaviours enhanced the learning measures, but, the expression deployment, expression standardization and expression animation, along with some setting deployments are required. These historical researches prompted the motivation of Wall-E in uniting a real-time look and gesture regularizing algorithm with a language model so as to readjust and alter conversation relying on the reactions and expressions of the learner.

## D. GESTURE CONTROL AND MOVEMENT

The touch-free interaction is the touch gesture recognition. *Nguyen et al. [6]* suggest a low-cost smart wheelchair that is controlled with six hand gestures and identify hand gestures with YOLOv8n. They employed 12000 images, which they preprocessed themselves with Media Pipe, and subsequently reported that they could identify gestures with 99.3 percent accuracy, giving it controls to make Jetson Nano-driven wheelchairs turn. It is an indication that it is possible to run lightweight models on embedded platforms and maintain them with high precision.

To avoid getting stuck into a local optimum, Du et al. coupled depth cameras with a 2D laser rangefinder and added a force offset into the potential field algorithm. The range of vision was increased, and straight tracks which could be taken did not lead to accidents.

## E. FACE RECOGNITION AND MEMORY

*R. Afifi et al. [8] and U. Qadir et al. [9]* The popularity of individualization depends on the proper facial identification. As of the previous research, Afifi et al. saved 128-dimensional embeddings generated by a pre-trained network and trained an SVM-based classifier to achieve over 99 percent accuracy in real-time. Scaling to large datasets, Qadir et al. observe that although the vector databases offer assistance in the fast retrieval process, they result in latency and privacy issues. Wall-E used InsightFace to embed institutional practices to their face and then searched nearest neighbors using a fast nearest neighbor search in a vector database (FAISS). This option gives a trade-off between scalability and computational efficiency and creates issues regarding safe storage of the data and consent of the user.

## F. LANGUAGE MODELS IN TUTORING

*C. Santana et al. [10], A. Torres-Caballero et al. [11] and T. Sievers et al. [13]* An AI-powered prompt on a low budget was developed, integrating vocal, visual, and gestural responses, which demonstrated that robots based on Raspberry Pi and Arduino could provide custom gesture language training. Sun et al. combined a multi-modal large language model in a robotic tutor intervention which was also able to incorporate memory resources and emotional intelligence to break the traditional barriers of engagement in an educational environment. The CARE model goes a step further to involve real time emotional and cognitive evaluation via daily XGBoost classification methods. These give grounds to believe that when memory and perception are incorporated into the language models, empathetic and adaptive tutors can be achieved. These are values that Wall-E is embracing with the less heavy language model and feedback of the facial and gesture proficiency.

## G. Emotion Recognition and Feedback Systems

Recognizing emotions allows robots to tailor their teaching style to the learner's emotional state. Research by *Serholt and Barendregt et al. [18] and Kim and Park et al. [33]* shows that systems capable of interpreting emotions can support collaboration and keep children more engaged during learning activities. In a similar way, *Dillon and Verma et al. [38]* emphasize that emotional flexibility in humanoid robots helps maintain focus and empathy throughout interaction. The Wall-E prototype applies a simplified emotion recognition system that combines facial and gesture analysis to fine-tune its responses—adjusting tone and lesson difficulty to match the learner's mood and performance.

## H. Multimodal interaction/communication

It enhances the naturalness and quality of the interaction between the humans and the robots by combining the speech, vision and gestures. The study by *Tanaka and Matsuzoe et al. [19]* demonstrated that the learning of vocabulary in children was improved when the cues in the form of gestures and verbal communications were taken together. Evidence was also provided by *Santana et al. [10] and Torres-Caballero et al. [11]* that low-cost multimodal robotic systems could adjust their instruction and provided personalized feedback with the help of speech and visual prompts.

Equally, Wall-E operates under a multimodal learning strategy where the speech delivery aligns with head movement and gesture recognition which forms interactive and environmentally conscious communication with the students.

## I. QUALITY OF INTERACTION AND ENGAGEMENT AT THE LEARNING LEVEL

The efficiency and retention of learning under robotic-assisted education is directly connected with engagement quality. *Kaur and Singh et al. [1]* observed that educational robots enhanced focus, creativity and teamwork. The research of *Westlund et al. [20]* conducted over the long term has revealed that peer-like robots establish rapport and develop language better. *Gutierrez et al. [29]* have also noted improved conceptual knowledge when the learners were engaged with responsive robots. Wall-E uses the engagement measures of gaze tracking, duration of interaction, and emotional stability to keep estimating and optimizing engagement among the learners.

## J. EDUCATIONAL ROBOTS AND ETHICS IN DESIGN AND DATA PRIVACY

The fundamental issue of implementing robots in educational and domestic settings is ethical design. *Zhao and McEwen [4] et al.* talked of long-term attachment of families and emotional transparency in social robots. *Breazeal [15]* emphasized the necessity to take ethical limits in commercial educational robots, and *Chassignol et al. [39]* pointed at the difficulties in terms of privacy and data safety that are AI-driven. The design of Wall-E incorporates explainability, data storage with consent of the owner, and exporting or deleting of a personal data by a caregiver with the implication of the data protection regulations.

# III. METHODOLOGY

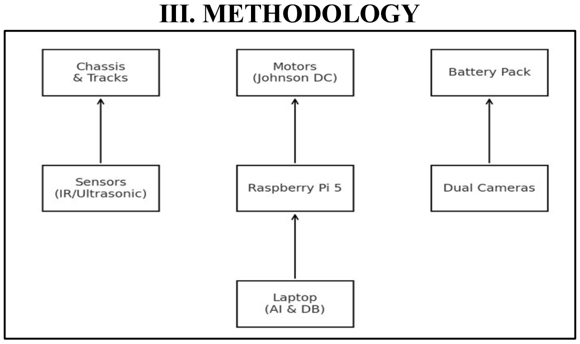


Figure 1: WALL-E Mechanical & Electronic Design

## A. MECHANICAL DESIGN

The physical design of WALL-E according to the Figure 1 has been created with the consideration of the popular nature of WALL-E as their foundation in making it more attractive and recognizable. The chassis is 3D printed on PLA and PETG due to good weight and hardworking. It is roughly 35 cm tall and 25 cm broad and is 30 cm long and it is roughly 2.5 kg in weight. The two continuous track assemblies are used in the robot and are driven by a brushed Johnson motor powered by an L298N. The battery pack is a home-made 18650 lithium-ion battery pack with protection circuitry in the middle of the base.

At the top of the base is the torso that contains the Raspberry Pi, servo controller, and wiring. The neck is made of 2 MG90S servos in a pitch-yaw system and allows the Wall-E to pan the head and tilt the head between +-45deg and +-30deg respectively in the yaw and pitch directions respectively. It has two webcams attached with a 3D-printed bracket referred to as eyes that enables it to have a stereo vision effect.

## B. ELECTRONIC COMPONENTS

The electronics incorporate off the shelf components to make it affordable and dependable. The significant components and their functions are mentioned in Table 1.

| Component | Function | Notes |
|---|---|---|
| Raspberry Pi 5 | Central processing; runs WebSocket server and low-level control | 8 GB RAM; boots from microSD |
| Laptop (external) | Runs vision, language model and database services | Communicates via WebSocket |
| MG90S servos | Control neck pitch and yaw, adjust eye positions | Controlled via PCA9685 |
| PCA9685 servo controller | Provides 16 channels of PWM for servos | Communicates over I²C |
| Johnson DC motors | Drive continuous tracks | Controlled by L298N H-bridge |
| L298N motor driver | Controls direction and speed of motors | Powered by battery pack |
| Dual webcams | Capture stereo video for detection and tracking | Mounted in eye assembly |
| Ultrasonic/IR sensors | Measure distance to obstacles | Assist obstacle avoidance |
| Battery pack (18650 cells) | Supplies power to all components | Includes BMS for safety |

Table 1: Electronic Components

## C. SOFTWARE ARCHITECTURE

*X. Du et al. [7]* The initial text explains the software architecture structure in the context of perception and control separation. The Raspberry Pi achieves the task of the WebSocket server, which transmits raw video and audio data to a separate laptop. The laptop also has a python application that has the use of openCV to retrieve frames, in this case, YOLO, which detects faces and objects. After identifying the faces, the faces are cropped and sent to InsightFace which generates 512-dimensional embeddings which are compared to a FAISS vector database to find the people.

Interest area is marked on the frame that is captured and in case a face moved outside the center, the application would calculate the servo control to be used in order to position the head back. Meanwhile, the results produced by the object detector are provided to a SLAM module that combines the stereo depth and ultrasonic/IR sensor data. A potential field path planner, having an offset force, creates motor commands to evade obstacles and move to target points. The laptop is also running a small language model (LLM) that has been fine-tuned to educative dialogue. The LLM not only asks the contextually relevant questions of the respective learner in terms of learning history, but also evaluates the answer of the learner to the questions with respect to semantic relatedness to the context and corrects the difficulty of the subsequent material.

Nonverbal cues, such as facial expression measures and/or open or closed palm, are also prompted to the LLM to evaluate the engagement of performance. Orders (i.e. servo angles, motor speeds and dialogue prompts) have been processed and sent back via a WebSocket to the Raspberry Pi. The Raspberry Pi then controls the servos with a PCA9685 controller and controls the motors with an L298N driver.

### D. AI MODELS AND ALGORITHMS

Face and Object Detection: Wall-E uses the YOLOv5n architecture to execute in a manner that it can be used in CPUs, which can isolate faces and the background among other domestic objects. Pre-processing of the frames with Media Pipe generates ROIs (regions of interest) on detected hands and faces; detected faces are in turn found with FaceNet and handed over to InsightFace, where they are retrieved with the embeddings. The major aspect of this system is that it implements incremental learning: once the face of a person is entered into the system, the embedding and the metadata (which contains a name, age, language preference, and so forth) are appended to the database of vectors with the permission of a user. The embeddings are subjected to the nearest neighbor search that will produce a list of closest matches. When the minimal distance is below acceptable threshold then it accepts the match as identifiable.

*X. Du et al. [7]* Gesture Recognition: A special neural network is under use to differentiate the open palm and closed palm gesture activity. The classifications, or continuous gestures, make use of vital landmarks obtained by MediaPipe; the resulting 21 key points are then flatened into a 21 x 3 matrix and proceed through three layers of dense neurons (64, 32 and 2 neurons) with ReLU activations, where the last layer uses SoftMax. A custom dataset was trained on the network and containing 4000 annotated images that were specifically designed to be shot in changing lighting conditions. In doing the inference, gestures are all measured on a confidence threshold of 0.9; open palms are similar, they use an open palm gesture to initiate an interaction or further interaction; a closed palm is used to prompt Wall-E to stop and turn off audio output.

Obstacle avoidance: Wall-E builds a 2-D occupancy grid, which is a combination of stereo disparity and distance sensor data. A field algorithm (potentially) supported by an offset force causes the robot to push aside obstacles and local minima. The planner takes into consideration a candidate into each direction of 10deg and selects the direction that will provide maximum clearance as well as move it toward the target in the planning stage. Pulse width modulation (PW) is used in controlling motor speeds during smooth turns and acceleration. In case there is confusion or the sensors fail to work, the robot just stops and waits until the user makes a move.

*Y. Sun et al. [12]* Language model: The conversational agent is a stripped down version of GPT-J and is configured to an elementary school vocabulary in which quantized weights are used to save memory and it runs on the laptop. Personalized prompts bring the model to a conversational teaching tone, and open-ended queries bring the inquiry forward. Sentence embeddings are then used to evaluate responses and cosine similarity is used to determine the correctness of the response. The model stores and assesses the responses of the learner in an SQLite database to vary the level of the questions with an algorithm of spaced repetition type SQLite. The subsequent versions may also use multimodal LLMs to recognize gestures and expressions.

### E. USER INTERFACE AND INTERACTION

eSpeak has produced the voice in Wall-E which is complemented with the LED eye rings that depict various emotions. Upon activation the user is greeted by the speech of Wall-E who then instructs the user to start learning by showing an open palm. In the course of learning activities, once the robot recognizes the face, the eye-tracking algorithm will accelerate to redirect its gaze to the learned and focal face in case the learner shifts his or her gaze elsewhere. The Wall-E will pose questions with a specific pace, leaving some room to answer.

It will either encourage the user or correct the user depending on the situation. The user can break or terminate the session by showing closed palm Haptically. The caregiver can use a mobile application companion to the robot to adjust the lesson plans of Wall-E, promote user interaction, track progress, and vocabulary.

## IV. IMPLEMENTATION AND EVALUATION

### A. CONSTRUCTION AND DEPLOYMENT

The chassis and tracks of Wall-E have been printed with the Fused Deposition Modelling with 0.2 mm layer height. It can be assembled in eight hours using ordinary tools. The electronic is wired on an acrylic plate that is cut by a laser cuter and is easy to maintain. Wiring harnesses are wired in safety grouping and are colour-coded. The WebSocket service is written in Python: FastAPI: FastAPI, which is Python-based, is a web API service. Neural network: PyTorch: PyTorch is a python-based web API framework. audio streaming: PyAudio: PyAudio is a Python-based web API service. The Ubuntu 22.04 is supported on the Raspberry Pi and the laptop.

### B. PERFORMANCE METRICS

The prototype assessment is associated with the real-time work and the quality of interaction. Face and object detection can be performed on a laptop with an Intel i5 8265U CPU at a rate of approximately 20 fps and in the case of streaming with 480p the end-to-end latency (capture to actuation) is around 120 milliseconds. The recognition of gestures was found to be 98 percent accurate on the test set; the false recognitions were mostly when the hand of the user was partially blocked. Face recognition at a threshold of.6, registered users with precision rate of 95 percent was a true identification. The battery pack made out of objects could run the robot on a continuous basis of about 2 hours, which was aided by the thermal management to sustain safe operating temperatures.

### C. USER STUDY

During the low-stakes pilot, three 8-10-year-old kids were able to play with Wall-E 30 minutes. The gesture commands were easily understood by children, and they also appeared to enjoy the expressiveness of the robot. According to qualitative feedbacks of parents, the personalised question style proved to be productive and immediate feedback made the learning process fun.

In the future, parents will enjoy following the progression of their child through the app, and the pool of participants will be larger and control groups will also be added in order to compare learning gains to the learning gains when taught in a normal way.

### D. COMPARATIVE ANALYSIS

Both table 2 and 3 compares Wall-E to a few types of educational robots on fundamental features. Wall-E stands out of commercial systems like NAO and Pepper because of less cost and modular design, whereas gesture control and facial recognition, personalized questions, and personalized question generation confer further benefits on Wall-E than more basic toy robots.

| Robot | Approx. cost | Mobility | Face Recognition |
|---|---|---|---|
| NAO | = USD 7,500 | Wheeled | Optional (Limited) |
| Pepper | = USD 20,000 | Wheeled | Optional |
| Kebbi/ Minibo | = USD 1,000 | Wheeled | No |
| Wall-E (this work) | = USD 400 | Tracked | Yes |

Table 2: WALL-E Comparative Analysis

| Robot | Gesture Control | Language Model | Notes |
|---|---|---|---|
| NAO | No | Limited | Expensive, Static face |
| Pepper | Limited | ChatGPT Interface | Large Humanoid, Used in classroom |
| Kebbi/ Minibo | No | None | Provides simple Interactive content |
| Wall-E (this work) | Yes (open/closed palm) | Light-weight LLM | 3D – printed, modular, home use |

Table 3: WALL-E Comparative Analysis

### E. LIMITATIONS

The existing prototype needs an external laptop that creates vision and conversational abilities that do not make it portable. The battery pack that was made by hand also restricts the duration that the prototype will last. The robot can also not perform a complex navigation task, since mapping new rooms. As of today, gesture recognition is only able to distinguish between two classes of gestures, this implies that the addition of more classes of gestures will need a fresh set of data. Price is another consideration.

Wall-E is cheaper than market robots, such as NAO or Pepper, but when all the ingredients are added up, including Raspberry Pi 5, two webcams, servo motors, motor drivers, battery pack, etc., the price is nearly USD 400. This may continue to be a significant obstacle to the families with tight budgets. The research will help in future research on how to reduce costs, such as modular kits or reused and reused electronics. Lastly, the ethical aspect of data privacy and possible bias of language models in conjunction with long-term emotional effects, which we will pay much attention to in future work.

## V. RESULTS AND DISCUSSION

The Walle Educational Robot was evaluated basing on performance as a robot in real time, quality of recognition, qualities of interaction and operational stability. The findings support the possibility of implementing a low-cost semi-autonomous educational robot with the ability of personalized interaction and multi-modular perception.

### A. REAL-TIME PERFORMANCE

The processing pipeline by Walle (i.e.: video streaming, object detection and face recognition) was verified on a laptop with an Intel i5-8265U processor. The system had a mean of about 20 frames per second (fps) which made the perception of the system to be smooth while interacting.

The video capture-to-motor-actuation latency was an average of approximately 120 milliseconds, which at the same time is adequate in real time responsiveness in an educational context. These findings show that optimized lightweight models and edge computing can provide a reasonable interactive performance when there is no special GPUs required. Figure 2 presents the performance results of the Wall-E robot, showing how it performs in real time, how accurately it recognizes faces and gestures, and how long it can operate on battery power during continuous use.

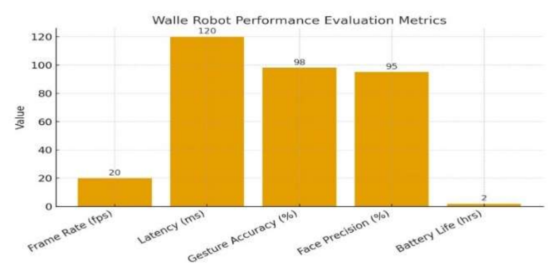


Figure 2: Walle Robot Prototype Performance Metrics.

### B. GESTURE AND FACE RECOGNITION ACCURACY

The hand gesture recognition module, which was based on a YOLOv8n model, had 98% accuracy on the test dataset. The misclassifications were primarily in case of partial occlusion or in poor light conditions and this means that future versions may comprise adaptive brightness compensation and temporal smoothness to make this version more robust.

Equally, the facial recognition, based on InsightFace and FAISS in terms of embedding-based identification, achieved a precision rate of 95% at a similarity threshold of 0.6. The system was able to detect the registered users in real time supporting the use of the system in individualized educational activities.

### C. POWER EFFICIENCY AND THERMAL STABILITY

The prototype was used over a period of 2 hours of non-stop use on a bespoke battery pack. The thermal control measures, such as the motor speed control and effective power control, kept the temperatures in the operating areas at safe levels. The findings validate that Walle is capable of supporting short learning periods without becoming overheated or needing any continuous power supply.

### D. LEARNING ENGAGEMENT AND INTERACTION

Qualitative feedback on initial tests has shown that the conversational tone of Walle, responsiveness both of facial and gesture recognition and interactive feedback loops is how more learning engagement of the learner was achieved. The dialogue from the robot, which is also a personality-driven dialogue based upon a quantized GPT-J model, was able to adjust the difficulty of questions to the learner by using a spaced-repetition algorithm depending on the performance of the learner, which is then stored in an SQLite database.

### E. DISCUSSION

The findings confirm that Walle is successful in integrating perception, cognition, and interaction in a small framework that can be used to learn at home or in classrooms. The system indicates that educational robots of low cost can be used to provide the multimodal aspects of learning and personalized learning experiences without compromising performance.

The problems, however, lie in such aspects like long-term autonomy, guaranteeing privacy, and the identification of emotions in the context. The future research will consider the use of multimodal large language models (LLM) to comprehend emotional and visual features and enhance the empathetic and adaptive functions of the robot.

## VI. ETHICAL CONSIDERATIONS AND LONG-TERM USE

Zhao et al. [4] and M. Belpaeme et al. [14] When developing robots to be used by a child, such ethical questions as the privacy and not being too tied effect and inclusive design become apparent. Children will tend to anthropomorphize robots, which are given mental state which rapidly develops emotional attachments which will last long after the robot has ceased to function. Wall-E deals with this issue by being transparent: it lets the children know what it can and cannot do and allows them to consider it an instrument rather than an animated object. Deletion or export of stored facial is possible by the caregivers. Information via the application and, thus, adhere to the data protection regulations.

In spite of the ethical issues that emerge with the use of robots in child design, such as privacy, attachment, and inclusivity. Robots are often anthropomorphized, given mental conditions by children; this causes attachment which can easily remain longer than the working life of the robot. To prevent any suffering, Wall-E is created in a transparent way: it tells what and what it cannot do, and encourages children to think of the robot as a tool and not a living being. The app allows caregivers to remove or transfer stored facial data to ensure that they follow the laws on data protection.

C. Breazeal et al. [15] and H. Wainer et al. [17] In addition to transparency, the data minimization, emotional balance, and inclusivity are also ethical design principles of educational robotics. The framework by Walle integrates consent-based interaction whereby any personal information, including that of faces and records of performance, is stored locally and can be deleted by any caregiver at any time. Walle uses contextual dialogue cues to avoid emotional overdependence and in the process, support the autonomy of the learner in a rather subtle way as opposed to being dependent. The behavior models can be implemented to promote self-directed learning whereby it is encouraged to be more resilient and interested in learning, rather than relying on the robot to do it.

Walle strives to make children of different backgrounds and abilities included which corresponds to the universal design principles of child-robot interaction. This paper explained Wall-E, a semi-autonomous educational robot which integrates cheap components with sophisticated computer vision and artificial intelligence so as to attract students as a home tutor. We have had a look at the relevant literature on educational robotics, long-term interaction between people and robots, and personal tutoring, described the mechanical and electronic design of Wall-E, software architecture, artificial intelligence algorithms, and interaction with the user. Preliminarily, it has been shown that Wall-E is safe to use, recognizes people, reacts to gestures, and permits adaptive real-time conversation

## VII. WALL-E PROTOTYPE

Figure 3 shows the Wall-E prototype that was developed and tested during the project.


Figure 3: Wall-E Working Model Prototype

The prototype mentioned in the figure 3 is a combination of various subsystems which include perception, mobility and conversational interaction. It was built with off-the-shelf parts which included a Raspberry Pi controller, stereo cameras, ultrasonic sensors and servo motors to simulate expressive movements. The chassis and facial housing were 3D-printed in accordance with modular changes in the testing stage. Gesture recognition, face identification and natural language interaction under real-time use was tested using this version of Wall-E. Field tests confirmed its reliability, responsiveness and applicability to learning in small classroom and home set-ups.

## VIII. CONCLUSION AND FUTURE WORK

The paper describes Wall-E, a semi-autonomous learning robot which was developed as a home tutor to students on a personal scale. Wall-E will be designed to ensure that intelligent tutoring is available, interactive and engaging to all learners regardless of their age due to its cost-effective hardware and advanced computer vision and artificial intelligence. The authors focus on the fact that the design of Wall-E is not that rich and expensive, that is why it could be introduced successfully to school and into an individual learning process, and the approach between technology and personal education could be achieved.

To offer theoretical basis, the authors overview substantial literature on educational robotics and study the advantages of robots in terms of motivation, engagement, and learning outcomes. They also discuss research on human-robot interaction over the long term, with the focus on the role of adaptability, emotional intelligence and a stable social behavior in maintaining the interest of users and trust. Also, the studies on personal tutoring systems provide useful results on how the adaptive algorithms of Wall-E can be used to customize lessons, feedback, and interactivity to specific learning speed, style and academic needs of students.

The mechanical and electronic design of the Wall-E is discussed in much detail in the paper, covering the safety, modularity, and long-term reliability. It has multiple layers of perception, planning and interaction in its software architecture that allows the robot to perceive its surroundings and act on them intelligently in real time. The AI algorithms that are integrated to Wall-E enable it to identify users, interpret gestures, and hold natural and context-sensitive conversations, which are improved with further interaction. Wall-E will develop adaptive learning over time, focusing on how it behaves and how it teaches, such that the next tutoring session will be more personalized and meaningful.

The preliminary tests reveal that Wall-E is safe to use at home, it is able to recognize people accurately and react to the human gestures, speech and emotional signals. Its adaptive real-time conversation ability makes it a useful learning partner that ensures curiosity and cognitive development. The article has shown that Wall-E is a significant advancement in the direction of affordable, intelligent and socially compelling robotic tutors, suggesting that robotics and AI can be used transformatively in the future of personalized learning.

## REFERENCES

[1] M. P. Kaur and H. Singh, “Impact of Educational Robots on Learning in Schools: A Review,” 2023 6th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, India, pp. 1–6, 2023.

[2] S. Hussain, M. H. Sarker, M. S. Rahman, R. Sultana, and M. S. A. Hossain, “An Interactive Educational Robot for Children's Learning and Environmental Awareness,” 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, pp. 1–5, 2019.

[3] F. Pelizzari, S. Rocco, and S. Ferrari, “Integrating Robotics in Hospital and Home Education: A Systematic Review of Innovative Teaching Practices,” Continuity in Education, vol. 6, no. 1, pp. 135–161, 2025.

[4] Z. Zhao and R. McEwen, “The Robot that Stayed: Understanding How Children and Families Engage with a Retired Social Robot,” Frontiers in Robotics and AI, vol. 12, Aug. 2025.

[5] N. Maaz, J. Mounsef, and N. Maalouf, “CARE: Towards Customized Assistive Robot-Based Education,” Frontiers in Robotics and AI, vol. 12, 2025.

[6] B. Nguyen et al., “Smart Wheelchair Using YOLOv8 and Hand Gesture Recognition,” preprint, 2023.

[7] X. Du et al., “Obstacle Detection and Path Planning for Mobile Robots Using Depth Cameras and Laser Rangefinders,” preprint, 2024.

[8] R. Afifi et al., “Real-Time Face Recognition Using Deep Learning and SVM,” 2023.

[9] U. Qadir et al., “Efficient Storage and Retrieval in Vector Databases for Face Recognition,” 2024.

[10] C. Santana et al., “Interactive Control of Service Robots Using ROS and Chatbots,” 2024.

[11] A. Torres-Caballero et al., “AI-Powered Language Learning Robotic Assistant,” 2024.

[12] Y. Sun et al., “Empathetic Robot Tutor Using Multi-Modal Large Language Models,” Frontiers in Robotics and AI, vol. 11, 2025.

[13] T. Sievers, “A Humanoid Social Robot as a Teaching Assistant in the Classroom,”

[14] M. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, and F. Tanaka, “Social robots for education: A review,” Science Robotics, vol. 3, no. 21, pp. 1–9, 2018.

[15] C. Breazeal, “Social robots: From research to commercialization,” Communications of the ACM, vol. 61, no. 11, pp. 38–46, 2018.

[16] J. Kennedy, P. Baxter, and T. Belpaeme, “The impact of robot tutor nonverbal immediacy on student learning,” in Proc. ACM/IEEE Int. Conf. Human–Robot Interaction (HRI), Portland, USA, pp. 67–74, 2015.

[17] H. Wainer, N. M. Stein, and R. G. Buttenfield, “Educational robots and computational thinking: A systematic review,” Journal of Science Education and Technology, vol. 30, no. 5, pp. 667–682, 2021.

[18] S. Serholt and A. Barendregt, “Students’ strategies for regulating their emotions in human–robot teaching interactions,” Computers & Education, vol. 125, pp. 56–70, 2018.

[19] F. Tanaka and S. Matsuzoe, "Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning," Journal of Human–Robot Interaction, vol. 1, no. 1, pp. 78–95, 2012.

[20] D. Kory Westlund, J. K. Frank, L. M. P. Speigel, M. M. Gordon, and C. Breazeal, “A long-term study of young children’s rapport, social emulation, and language learning with a peer-like robot playmate in preschool,” Frontiers in Robotics and AI, vol. 4, pp. 1–18, 2017.

[21] A. Anwar, N. A. Bascou, and A. Kardgar, “A Systematic Review of Studies on Educational Robotics,” Journal of Pre-College Engineering Education Research, vol. 9, no. 2, pp. 19–42, 2019.

[22] K. Wang and X. Zhou, "The Effectiveness of Educational Robots in Improving Student Behaviour and Development: A Meta-Review," Sustainability, vol. 15, no. 5, 2023.

[23] G. Lampropoulos et al., "Social Robots in Education: Current Trends and Future Perspectives," Information (MDPI), vol. 16, no. 1, 2025.

[24] F. Ouyang and W. Xu, "The Effects of Educational Robotics in STEM Education: A Multilevel Meta-Analysis," International Journal of STEM Education, vol. 11, Article 7, 2024.

[25] snaini, R. & Budiyanto, C. W. (2018, November). The Influence of Educational Robotics to Computational Thinking Skill in Early Childhood Education. In Proceedings of the 1st International Conference on Computer Science and Engineering Technology (ICCSET), EAI. DOI:10.4108/eai.24-10-2018.2280538.

[26] García-Valcárcel, A. & Caballero-González, Y.-A. (2019). Robotics to develop computational thinking in early Childhood Education. Comunicar, 27(59), 63–72.

[27] L. Xia, "A Systematic Review on Teaching and Learning Robotics Content Knowledge in K-12," Computers & Education, vol. 124, pp. 180–191, 2018.

[28] Chen, T., & Xu, L. (2025). Developing preschool children's computational thinking and executive functions: unplugged vs. robot programming activities. International Journal of STEM Education, 12, Article 10.

[29] D. Gutiérrez, P. Fernández-Lozano, and A. Rodríguez-González, "Designing Robot-Based Learning Experiences in Primary Classrooms: Impact on Engagement and Conceptual Understanding," Education and Information Technologies, vol. 29, no. 2, pp. 1421–1440, 2024.

[30] S. L. Chang and Y. W. Hsu, "Integrating Robotics into Elementary Science Lessons: Effects on Students' Inquiry Skills and Attitudes toward STEM," International Journal of Science Education, vol. 46, no. 5, pp. 823–841, 2024.

[31] R. A. Martínez-Maldonado, M. Castro, and J. E. Salinas, "Robotic Tutors for Early Literacy Development: A Mixed-Methods Evaluation," Computers & Education, vol. 208, pp. 104703, 2023.

[32] Smakman, H. M. J., et al. (2022). Do robotic tutors compromise the social-emotional development of children? Frontiers in Robotics and AI. DOI:10.3389/frobt.2022.734955.

[33] T. Y. Kim and D. W. Park, "Long-Term Interaction Effects of Humanoid Robots on Collaborative Problem-Solving in Children," Frontiers in Psychology, vol. 14, Article 1173210, 2023.

[34] P. N. Kalloo, S. B. Mohan, and R. J. Narine, "Exploring Teachers' Readiness for Educational Robotics Integration in Primary Schools," Australasian Journal of Educational Technology, vol. 41, no. 1, pp. 65–83, 2025.

[35] M. R. Alghamdi and F. T. Naser, "Evaluating the Use of Social Robots to Support Children with Communication Challenges," International Journal of Social Robotics, vol. 17, pp. 233–248, 2025.

[36] F. Stasolla, E. Curcio, A. Borgese, A. Passaro, M. C. Di Gioia, A. Zullo, and E. Martini, "Educational Robotics and Game-Based Interventions for Overcoming Dyscalculia: A Pilot Study," Computers, vol. 14, no. 5, 2025.

[37] F. Tanaka, K. Isshiki, M. Takahashi, and T. Nakanishi, "Long-Term Field Trial of a Language Learning Robot with Elementary School Students," Frontiers in Robotics and AI, 2020.

[38] R. S. Dillon and S. A. Verma, "A Review on Humanoid Robots for Education and Social Interaction," IEEE Access, vol. 11, pp. 115300–115316, 2023.

[39] P. Chassignol, A. Khoroshavin, A. K. Chassignol, and A. B. Gao, "AI and Robotics in Education: Opportunities and Challenges," Procedia Computer Science, vol. 184, pp. 341–347, 2021.

[40] M. C. Busch et al., "Robots as Co-Teachers: Evaluating Long-Term Learning Outcomes in Primary Education," Frontiers in Education, vol. 9, Article 1459, 2024.