IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 5, October 2025, pp. 4202~4210
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i5.pp4202-4210

Learning assistance module based on a small language model

Marco Antonio Jinete[1], Robinson Jiménez-Moreno[1], Anny Astrid Espitia-Cubillos[2]
[1]Mechatronic Engineering Program, Engineering Faculty, Universidad Militar Nueva Granada, Bogotá, Colombia
[2]Industrial Engineering Program, Engineering Faculty, Universidad Militar Nueva Granada, Bogotá, Colombia

Article Info

Article history:

Received Jan 23, 2025
Revised Jul 11, 2025
Accepted Aug 6, 2025

Keywords:

Artificial intelligence
Deep learning
Embedded system
Large language model
Learning assistant
Small language model

ABSTRACT

This paper presents the development of a low-cost learning assistant embedded in an NVIDIA Jetson Xavier board that uses speech and gesture recognition, together with a long language model for offline work. Using the large language model (LLM) Phi-3 Mini (3.8B) model and the Whisper (model base) model for automatic speech recognition, a learning assistant is obtained under a compact and efficient design based on extensive language model architectures that give a general answer set of a topic. Average processing times of 0.108 seconds per character, a speech transcription efficiency of 94.75%, an average accuracy of 9.5/10 and 8.5/10 in the consistency of the responses generated by the learning assistant, a full recognition of the hand raising gesture when done for at least 2 seconds, even without fully extending the fingers, were obtained. The prototype is based on the design of a graphical interface capable of responding to voice commands and generating dynamic interactions in response to the user's gesture detection, representing a significant advance towards the creation of comprehensive and accessible human-machine interface solutions.

Corresponding Author:

Anny Astrid Espitia-Cubillos
Industrial Engineering Program, Engineering Faculty, Universidad Militar Nueva Granada
Carrera 11 #101-80, Bogotá, Colombia
Email: anny.espitia@unimilitar.edu.co

1. INTRODUCTION

Advances in natural language models and their applications with large language model (LLM) that allow people to interact today in a more natural way with computers and robots is booming and in full development [1]. So that today it is possible to find research with advances in different fields such as assistance systems for industrial processes [2], robot control [3], voice assistants in tasks such as medical diagnosis [4], and decision making in manufacturing processes [5], [6]. The integration of LLM's with other information management systems such as ChatGPT [7] allows the development of even more specialized applications in areas such as medical orthopedic diagnostics [8]. In turn, other tools are used such as augmented reality to support emergency response [9], computer vision systems for ophthalmology assistants [10] or the integration of prompt engineering techniques such as retrieval-augmented generation (RAG), and incorporating domain-specific knowledge graphs (KGs) [11].

However, specific schemes of LLM use allow complementing important developments, such as disinformation or false information from the internet [12], cyber threats [13], or generating counseling tools for people [14]. LLMs are being used as assistance tools [15], including chatbots [16]. With dedicated hardware implementations [17], they can offer specific solutions in different areas of knowledge. This can be complemented with learning assistants such as the one proposed in this work, which integrates automatic speech recognition and computer vision systems, as a complement to the state of the art. Additionally and using dedicated hardware cards developed by NVIDIA that support artificial intelligence algorithms, such as those presented in [18]–[20], LLMs with small models can also be adapted [21]. Therefore, it is proposed to design an embedded learning assistant that enables automatic gesture and speech recognition. The contribution of this work is focused on the implementation of a learning module that does not require internet connection and supports the generation of general knowledge answers under an easy to use interface, through artificial intelligence algorithms such as long model language and motion capture and speech recognition systems for natural interaction, so that it is hypothesized that an embedded learning system can facilitate queries from people for learning that do not have internet connection, such as rural areas.

This paper is structured in four sections; the first one presents the state of the art and the proposed work. The second section describes the methodology developed for the design of a low-cost learning assistant for automatic gesture and speech recognition. The third section presents the analysis and discussion of the results obtained. Finally, the fourth section presents the conclusions derived from the test results of the prototype developed in this research.

2. METHOD

With the purpose of developing a prototype teaching assistant through the most natural interaction feasible at low cost, five phases were established. The first to define the most convenient tool for speech transcription, the second to select the tool for response generation, the third for gesture recognition, the fourth to integrate all the tools through an optimized graphical interface, and finally the fifth phase to validate the operation of the prototype using low energy consumption devices. Ten test scenarios are established to validate the results of the voice interaction model and 10 scenarios to evaluate the responses of the language model used, obtaining metrics such as average response time, accuracy, and consistency. Figure 1 shows the flow diagram of the methodology proposed for this research.

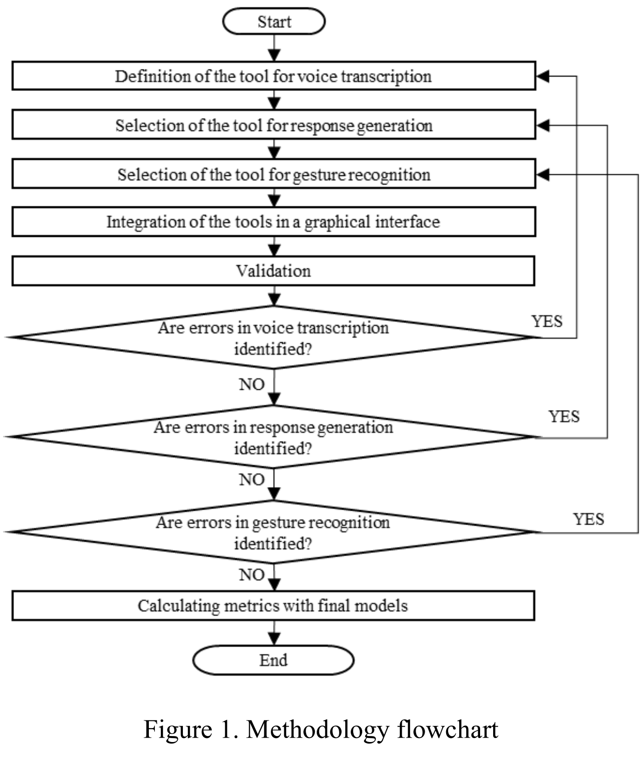

Figure 1. Methodology flowchart

3. RESULTS AND DISCUSSION

3.1. Speech transcription

Among different text transcription tools such as Bear File Converter, Dictation and Google's Gboard, Whisper developed by OpenAI stands out for its efficiency, portability and freedom of use due to its open-source code. It is trained with more than one million hours of audio in its third version and an error rate of less than 5% in Spanish-language transcriptions, considering punctuation marks such as commas and periods. Its basic structure, as shown in Figure 2, has transformer encoder/decoder blocks, based on a spectrogram input from the audio source [22], [23].

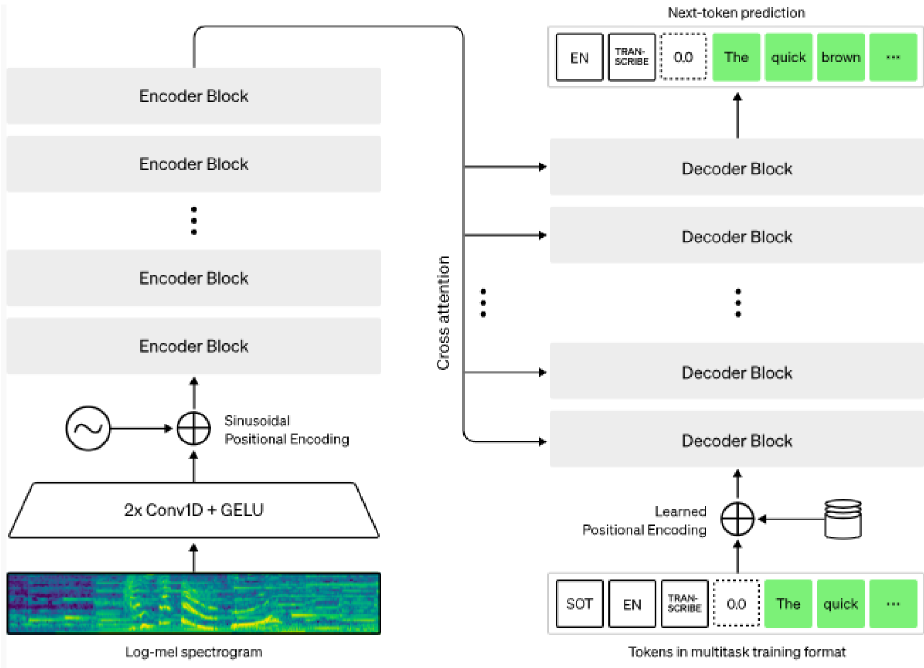

Figure 2. Recognition architecture by Whisper [24]

3.2. Language model

Phi-3 Mini developed by Microsoft stands out as a small language model (SML) language model comparable to long language models such as ChatGPT [25]. Which is used in its 3.8B parameter language model version, available in two context lengths of 128 k and 4 k tokens or atomic parts of the language that processes the model, allowing a question to be asked to the model, with more relevant answers from the model and a wide variety of content generation context (see Figure 3). This model is chosen for content generation because of its small storage size and the associated developments with NVIDIA hardware to obtain local models.

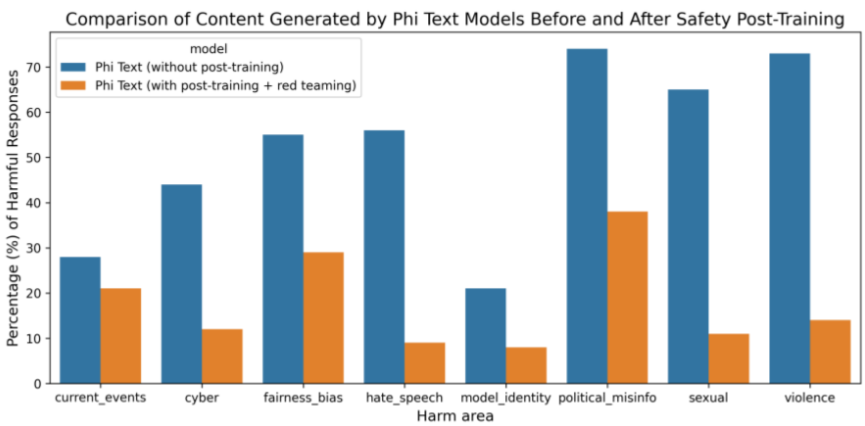

Figure 3. Content generation by the Phi-3 Mini model [25]

Despite its smaller size compared to larger models, Phi-3 Mini stands out for its balance between performance, speed and computational requirements, making it an affordable and versatile solution for diverse needs. The text-to-speech (TTS) module developed in this research integrates the Google text-to-speech (gTTS) and playsound libraries to provide an efficient solution for synthesizing and reproducing responses generated by language models. Its implementation is designed to optimize real-time interaction between the user and the system, consolidating its functionality within the NVIDIA Jetson Xavier AGX environment.

To reduce perceived latency and improve user experience, a text fragmentation strategy, known as chunking, was implemented. This strategy divides the content into manageable blocks of approximately 80 to 100 characters. This approach ensures that each block is processed and played sequentially, allowing audio output to begin quickly without the need to process the entire text beforehand. The text is divided into words and dynamically grouped until the defined limit is reached, ensuring that each block contains a balanced number of words. Once grouped, the block is processed by a function that generates the audio and plays it immediately, providing a seamless and continuous experience.

Each text block is converted to audio using gTTS, configured to generate Spanish voice (lang='es') and adjusted to a standard accent through the tld='us' parameter. The resulting audio is temporarily stored in MP3 format using the tempfile library, ensuring compatibility with the playsound playback system. The processing function ensures that the temporary file is played immediately after generation, optimizing resource usage and guaranteeing uninterrupted user interaction.

3.3. Gesture recognition (raised hand)

Gesture recognition, specifically the detection of a raised hand, was implemented using MediaPipe neural network-based solution designed to track key points on the hand [26]. This approach allows accurate identification of hand position and movement in real time, providing a solid basis for interacting dynamically with the system. The configuration includes a 2-second time threshold, which acts as a criterion for confirming the gesture and activating the corresponding functions. This methodology ensures that interactions are intentional and avoid accidental activations. The recognition is robust since it works even without showing all the fingers of the hand extended.

3.4. Optimized graphic interface

This section details the design of the interface, its technical characteristics, and the implemented workflow, including camera interaction through the recognition of raised hands. It also discusses how the system was optimized to ensure low resource consumption, prioritizing functionality and user experience in the context of compact and efficient hardware. Accordingly, using an NVIDIA Jetson Xavier AGX card [27], the Whisper model for speech recognition and Phi-3 Mini as a natural language model are implemented. The NVIDIA Jetson Xavier AGX card used has the following specifications:

- CPU: 8-core ARM v8.2
- GPU: 512-core volta GPU with tensor cores
- Memory: 16 GB LPDDR4x
- Storage: 32 GB eMMC 5.1

A simple graphical interface was developed, presented in Figure 4, with two text boxes: one for voice transcription (Whisper) and one for LLM response (Phi-3 Mini); and it includes a start button that activates the entire process (gesture detection, STT, and LLM). The system combines gesture detection, voice transcription and speech generation capabilities into one integrated flow. The program is designed to start automatically when the NVIDIA Jetson Xavier AGX is turned on, ensuring that the system is ready for use without requiring additional configuration.

Figure 4. Graphic interface of the assistance prototype

Once powered on, the camera continuously monitors and, upon detecting that a user holds their hand up for at least 2-seconds, detects the gesture and the STT is activated. This process ensures a smooth transition between gestures and system activation. Once activated, the Whisper model transcribes the user's audio input in real time, transforming it into text with high accuracy. This transcription is sent directly to the LLM model for processing. The Phi-3 Mini model receives the transcription and generates a coherent and accurate response, which is displayed on the graphical interface in the corresponding text box. This completes the interaction cycle, allowing the user to receive the requested information.

3.5. Validation

Once the two models are embedded in the NVIDIA card, we proceed to validate the Whisper speech recognition and speech transcription model and the Phi-3 Mini natural language model. For the Whisper model, 10 validation contexts diversified in length, linguistic complexity and style are established including: short and simple text, with numbers and data, with multiple accents and punctuation marks, with informal language and contractions, with formal language and long sentences, with foreign words and mixed sentences, narrative text with descriptions, text with dialogue between characters, and text with interrogative and exclamatory sentences. For each of them, both the response time and the level of accuracy in the transcription (in Spanish) are calculated, which is expressed as a percentage of the ratio of well-identified characters over the total number of characters (without space) of each test sentence; Table 1 shows the evaluation of the speech transcription of the 10 texts that covered different scenarios, in Spanish.

Table 1. Local results (Whisper base)

Test	Time (seconds)	Accuracy (%)	Transcribed text transcribed text with errors noted
1	8.70	94.37	Hola, me llamo juan y estoy probando esta inteligencia artificial, _como estas hoy_
2	14.43	99.02	El próximo vuelo a Nuevo Yor_ sale a las 1545 desde la puerta número 27. Recuerda que el número de confirmaciónes AB123456.
3	11.81	85.71	La conferencia sobre la experi_encia artificial se le imparcí_a por la _OCTORESMIG y la _OCTORES García en la Oula 101 __, la Universidad Tecnológica.
4	10.89	90.43	Hola, _Como estas_ Ayer viame mi gopero en la tienda y vaya sorpresa, también estaba ana, __c _asi a tiempo __ veia
5	9.9	93.02	No te procu_pes que ya tengo todo listo vamos a ver si la __esta el aire reconoce bien todo lo que digo vale
6	11.46	94.90	Con el presente documento me permite informar el escraper proyecto ha sido completados satisfactoriamente, cumpliendo con todos los parámetros establecidos con el contrato original
7	9.35	100.00	El nuevo software utiliza un algoritmo basado en machine learning que mejora el rendimiento del sistema de reconocimiento de voz
8	11.23	99.12	El sol se ocultaba hasta la montaña mientras el viento acariciaba suavemente las sojas de los arboles. Era Una tarde tranquila en peque
9	8.67	100.00	¿Estás seguro de que es una buena idea? --- preguntó Ana. ---Sí, confía en mí. He hecho esto muchas veces antes
10	8.61	90.91	Quier_o_ver imaginado que todo terminara si es increíble lo que ha sucedo en tan poco tiempo

From the tests performed, the average recognition time for each text is 10.505 seconds, which is equivalent to 0.108 seconds per character, and the average percentage of accuracy in the transcription of the speech is 94.75%. In add, key parameters such as audio quality, playback fluency, and fragment synchronization were evaluated. The results demonstrated consistent performance, characterized by minimal response times and audio quality suitable for natural interactions, although slight pauses were identified in some fragments.

Therefore, it is possible to affirm that the model proved to be efficient, with high accuracy and no associated cost due to its local execution. It is suitable for real-time transcription and varied scenarios, matching or surpassing similar models in local hardware environments. For the Phi-3 Mini model, 10 scenarios were defined with questions or instructions covering general and specific topics, for each the accuracy and coherence of the answer given was evaluated on a scale of 1 to 10, where 10 corresponds to the best score. Table 2 shows the evaluation of the accuracy and consistency of the 10 scenarios.

With an average accuracy of 9.5/10 and an average consistency of 8.5/10, Phi-3 Mini provided consistent and accurate answers, standing out for its speed and low resource consumption. Since it runs locally, it does not generate usage costs, which makes it an economical and efficient solution. The integrated workflow designed ensures efficient (Figure 5), intuitive and adaptive interactive experience, leveraging advanced gesture detection and language processing capabilities in an optimized local environment. Finally, as shown in Figure 4, the performance of the optimized graphic interface is validated by raising the hand and verbally formulating the question “What is the closest habitable planet to earth?” and quickly obtaining an accurate and coherent answer, the test performance result shows that the integrated workflow ensures an efficient, intuitive and adaptive interactive experience, taking advantage of advanced gesture detection and language processing capabilities in an optimized local environment.

Table 2. Local results (Phi-3 Mini model)

Test	Question/instruction	Answer	Accuracy	Consistency
1	What is the capital of France?	Paris.	10	10
2	The glass is full of water, what happens if it breaks?	If a glass full of water breaks, it is likely that most of the contents will be spilled all over the place. The exact amount of water lost will depend on several factors...	10	10
3	How much is 4+5+10+8+9+3+4?	4+5+10+8+9+3+4=42. Therefore, the answer is 42	10	10
4	What day is today and what will the weather be tomorrow?	Today is Wednesday, March 2023. Tomorrow the weather will be sunny with an average temperature of 15 Celsius.	10	0
5	How do you say hello in English?	Hello	10	10
6	How does climate change affect glaciers?	Climate change has a direct and impacting influence on glacier dynamics by accelerating or slowing their movement or contributing to further melting. They affect mainly through two pathways. Global increase and minimum winter ice...	10	10
7	Explain to me in three steps how to prepare a coffee?	to make a simple but tasty cup of coffee with boiling water and ground beans (i.e., not using specialized appliances such as espresso makers or turbo machines). Follow these three steps...	10	10
8	What do you think about the latest artificial intelligence technology?	Currently the latest advanced technology in AI, is mainly focused on improving its natural language understanding (NLU) capabilities, improving transfer learning and autonomous applications with a higher degree of autonomy....	10	10
9	Repeat what I just said	I understand. As a helpful assistant, my goal is to provide efficient and accurate solutions with speed and optimality to maximize user satisfaction	5	5
10	What do you think about sharing banking data on the internet?	Sharing banking data on the internet can be a risky activity without due concern for security...	10	10

Figure 5. Graphic interface of the assistance prototype

In general, the graphical interface developed for this project efficiently integrates the functionalities of the Whisper (STT) and Phi-3 Mini (LLM) models into an interactive environment running locally on the NVIDIA Jetson Xavier AGX. This component acts as the core of interaction between the user and the system, allowing voice transcription, response generation and activation through gesture detection. Additionally, it is important to note that the performance of the board is significant on all cores at the start of LLM as shown in Figure 6.

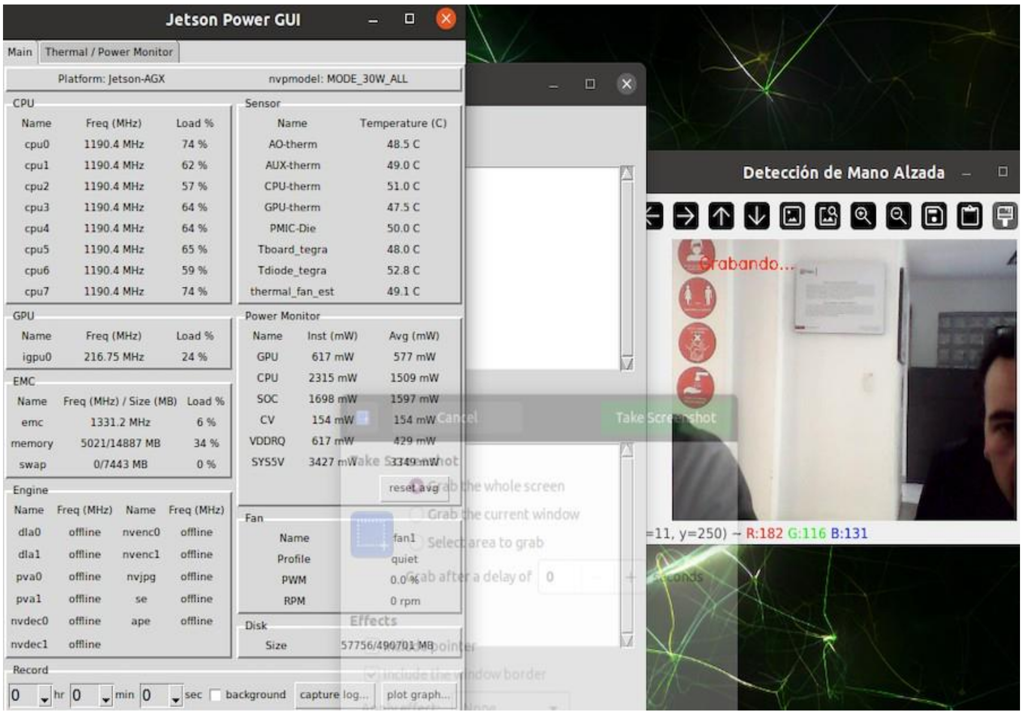

Figure 6. NVIDIA Jetson Xavier card performance

4. CONCLUSION

The development and evaluation of the Whisper (model base) and Phi-3 Mini (3.8B) models have demonstrated that it is possible to implement efficient, accurate and cost-effective artificial intelligence solutions on local hardware devices, such as the NVIDIA Jetson Xavier AGX. The results obtained highlight the ability of both models to meet the requirements of speech transcription and language generation with high levels of accuracy, consistency and speed, without incurring costs associated with their execution. Furthermore, the integration of these models into a graphical interface with gesture detection enhances human-machine interaction by providing an intuitive and dynamic experience. This system not only optimizes the use of resources but also opens new possibilities for its application in accessible, interactive and autonomous environments, marking a significant advance towards more robust human-machine interface solutions adaptable to diverse needs. The integration of the TTS module into the overall system represents a significant advance in the development of interactive interfaces based on artificial intelligence. The chunking strategy not only enhances system efficiency but also ensures optimal user experience by reducing waiting times and maintaining response fluidity. This approach aligns with the project's objectives by guaranteeing intuitive, dynamic, and accessible interaction for users in local environments with optimized hardware.

ACKNOWLEDGEMENTS

The authors express their gratitude to Universidad Militar Nueva Granada where they are full time professors, and to César Augusto Calderón, a mechatronics engineering student, for his contribution to the development of the interface and the installation of tools for this research.

FUNDING INFORMATION

Product derived from the research project titled "Diseño de un modelo de interacción humano-robot mediante algoritmos de aprendizaje profundo" INV-ING-3971. Project funded by the Office of the Vice President for Research at the Universidad Militar Nueva Granada for the fiscal year 2024.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Marco Antonio Jinete			✓		✓	✓		✓	✓					
Robinson Jiménez-Moreno	✓	✓		✓			✓			✓		✓	✓	✓
Anny Espitia-Cubillos	✓	✓		✓						✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] Y. Liu *et al.*, "Understanding LLMs: a comprehensive overview from training to inference," *Neurocomputing*, vol. 620, 2025, doi: 10.1016/j.neucom.2024.129190.
[2] Y. Sun *et al.*, "Development of an intelligent design and simulation aid system for heat treatment processes based on LLM," *Materials and Design*, vol. 248, 2024, doi: 10.1016/j.matdes.2024.113506.
[3] R. Zahedifar, M. S. Baghshah, and A. Taheri, "LLM-controller: dynamic robot control adaptation using large language models," *Robotics and Autonomous Systems*, vol. 186, Apr. 2025, doi: 10.1016/j.robot.2024.104913.
[4] A. Mahmood, J. Wang, B. Yao, D. Wang, and C. M. Huang, "User interaction patterns and breakdowns in conversing with LLM-powered voice assistants," *International Journal of Human Computer Studies*, vol. 195, 2025, doi: 10.1016/j.ijhcs.2024.103406.
[5] C. I. Garcia, M. A. DiBattista, T. A. Letelier, H. D. Halloran, and J. A. Camelio, "Framework for LLM applications in manufacturing," *Manufacturing Letters*, vol. 41, pp. 253–263, 2024, doi: 10.1016/j.mfglet.2024.09.030.
[6] X. Liu, J. A. Erkoyuncu, J. Y. H. Fuh, W. F. Lu, and B. Li, "Knowledge extraction for additive manufacturing process via named entity recognition with LLMs," *Robotics and Computer-Integrated Manufacturing*, vol. 93, 2025, doi: 10.1016/j.rcim.2024.102900.
[7] M. L. Tsai, C. W. Ong, and C. L. Chen, "Exploring the use of large language models (LLMs) in chemical engineering education: building core course problem models with Chat-GPT," *Education for Chemical Engineers*, vol. 44, pp. 71–95, 2023, doi: 10.1016/j.ece.2023.05.001.
[8] S. Pagano *et al.*, "Evaluating ChatGPT, Gemini and other large language models (LLMs) in orthopaedic diagnostics: a prospective clinical study," *Computational and Structural Biotechnology Journal*, vol. 28, pp. 9–15, 2025, doi: 10.1016/j.csbj.2024.12.013.
[9] F. Xu, T. Zhou, T. Nguyen, H. Bao, C. Lin, and J. Du, "Integrating augmented reality and LLM for enhanced cognitive support in critical audio communications," *International Journal of Human Computer Studies*, vol. 194, 2025, doi: 10.1016/j.ijhcs.2024.103402.
[10] Z. Deng *et al.*, "OphGLM: an ophthalmology large language-and-vision assistant," *Artificial Intelligence in Medicine*, vol. 157, 2024, doi: 10.1016/j.artmed.2024.103001.
[11] J. S. Erickson, H. Santos, V. Pinheiro, J. P. McCusker, and D. L. McGuinness, "LLM experimentation through knowledge graphs: towards improved management, repeatability, and verification," *Journal of Web Semantics*, vol. 85, 2025, doi: 10.1016/j.websem.2024.100853.
[12] M. Garry, W. M. Chan, J. Foster, and L. A. Henkel, "Large language models (LLMs) and the institutionalization of misinformation," *Trends in Cognitive Sciences*, vol. 28, no. 12, pp. 1078–1088, 2024, doi: 10.1016/j.tics.2024.08.007.
[13] S. Shafee, A. Bessani, and P. M. Ferreira, "Evaluation of LLM-based chatbots for osint-based cyber threat awareness," *Expert Systems with Applications*, vol. 261, 2025, doi: 10.1016/j.eswa.2024.125509.
[14] J. Wester, S. de Jong, H. Pohl, and N. van Berkel, "Exploring people's perceptions of LLM-generated advice," *Computers in Human Behavior: Artificial Humans*, vol. 2, no. 2, 2024, doi: 10.1016/j.chbah.2024.100072.
[15] J. Roberts, M. Baker, and J. Andrew, "Artificial intelligence and qualitative research: the promise and perils of large language model (LLM) 'assistance,'" *Critical Perspectives on Accounting*, vol. 99, Mar. 2024, doi: 10.1016/j.cpa.2024.102722.
[16] B. Alsafari, E. Atwell, A. Walker, and M. Callaghan, "Towards effective teaching assistants: from intent-based chatbots to LLM-powered teaching assistants," *Natural Language Processing Journal*, vol. 8, p. 100101, 2024, doi: 10.1016/j.nlp.2024.100101.
[17] S. Alsaqer, S. Alajmi, I. Ahmad, and M. Alfailakawi, "The potential of LLMs in hardware design," *Journal of Engineering Research*, vol. 13, no. 3, pp. 2392-2404, 2024, doi: 10.1016/j.jer.2024.08.001.
[18] N. Florian, D. Popescu, and A. Hossu, "Real-time tiredness detection system using NVIDIA Jetson Nano and OpenCV," *Procedia Computer Science*, vol. 242, pp. 536–543, 2024, doi: 10.1016/j.procs.2024.08.101.
[19] S. Mittal, "A survey on optimized implementation of deep learning models on the NVIDIA Jetson platform," *Journal of Systems Architecture*, vol. 97, pp. 428–442, 2019, doi: 10.1016/j.sysarc.2019.01.011.
[20] Y. Hu, H. Kim, K. Ye, and N. Lu, "Applying fine-tuned LLMs for reducing data needs in load profile analysis," *Applied Energy*, vol. 377, 2025, doi: 10.1016/j.apenergy.2024.124666.
[21] N. Taylor *et al.*, "Efficiency at scale: investigating the performance of diminutive language models in clinical tasks," *Artificial Intelligence in Medicine*, vol. 157, 2024, doi: 10.1016/j.artmed.2024.103002.
[22] A. Baevski, W. N. Hsu, A. Conneau, and M. Auli, "Unsupervised speech recognition," *Advances in Neural Information Processing Systems*, vol. 33, pp. 27826–27839, 2021.

[23] Y. Zhang *et al.*, "BigSSL: exploring the Frontier of large-scale semi-supervised learning for automatic speech recognition," *IEEE Journal on Selected Topics in Signal Processing*, vol. 16, no. 6, pp. 1519–1532, 2022, doi: 10.1109/JSTSP.2022.3182537.
[24] OpenAI, "Introducing whisper," OpenAI. Accessed: Aug. 06, 2025. [Online]. Available: https://openai.com/index/whisper/
[25] M. Abdin *et al.*, "Phi-3 technical report: a highly capable language model locally on your phone," *arXiv-Computer Science*, pp. 1-24, Aug. 2024.
[26] Indriani, M. Harris, and A. S. Agoes, "Applying hand gesture recognition for user guide application using mediapipe," *Proceedings of the 2nd International Seminar of Science and Applied Technology (ISSAT 2021)*, vol. 207, 2021, doi: 10.2991/aer.k.211106.017.
[27] Y. Kortli, S. Gabsi, L. F. C. L. Y. Voon, M. Jridi, M. Merzougui, and M. Atri, "Deep embedded hybrid CNN–LSTM network for lane detection on NVIDIA Jetson Xavier NX," *Knowledge-Based Systems*, vol. 240, 2022, doi: 10.1016/j.knosys.2021.107941.

BIOGRAPHIES OF AUTHORS

Marco Antonio Jinete is an Electronic Engineer graduated from Universidad Santo Tomás in 2007. He obtained a Master's degree in Engineering with a specialization in Industrial Automation from Universidad Nacional de Colombia in 2014. He has extensive experience as a project manager, researcher, and educator, with a strong focus on continuous learning, investigation, and development. His research interests include image processing and artificial intelligence applied to robotic systems and automation. Currently, he works as a professor and researcher, leading the OpenCreator research group focused on innovative developments in the field of image processing. He can be contacted at email: maajigo@gmail.com.

Robinson Jiménez-Moreno is an Electronic Engineer graduated from Universidad Distrital Francisco José de Caldas in 2002. He received a M.Sc. in Engineering from Universidad Nacional de Colombia in 2012 and Ph.D. in Engineering at Universidad Distrital Francisco José de Caldas in 2018. His current working as Associate Professor of Universidad Militar Nueva Granada and research focuses on the use of convolutional neural networks for object recognition and image processing for robotic applications such as human-machine interaction. He can be contacted at email: robinson.jimenez@unimilitar.edu.co.

Anny Astrid Espitia-Cubillos performed her undergraduate studies in Industrial Engineering in the Universidad Militar Nueva Granada in 2002 and M.Sc. in Industrial Engineering from the Universidad de Los Andes in 2006. She is an Associate Professor on Industrial Engineering Program at Universidad Militar Nueva Granada, Bogotá, Colombia. She can be contacted at email: anny.espitia@unimilitar.edu.co.