Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 18 No. 23 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijim.v18i23.51419

PAPER

Future Prospects of Large Language Models: Enabling Natural Language Processing in Educational Robotics

S. Vinoth Kumar[1](✉), R. B. Saroo Raj[2], J. Praveenchandar[3], S. Vidhya[4], S. Karthick[1], R. Madhubala[5]

[1]Department of CSE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India

[2]Department of IT, Rajalakshmi Engineering College, Chennai, Tamil Nadu, India

[3]Department of Artificial Intelligence and Machine Learning, School of Computer Science and Technology, Karunya Institute of Technology and Sciences (Deemed to be University), Coimbatore, Tamil Nadu, India

[4]Department of Information Science and Engineering, The Oxford College of Engineering, Bangaluru, Karnataka, India

[5]Information Technology Department, University of Technology and Applied Sciences, Shinas, Oman

drsvinothkumar@veltech.edu.in

ABSTRACT

Large language models (LLMs) have recently shown considerable promise in educational robotics by offering generic knowledge necessary in situations when prior programming is not possible. In general, mobile education robots cannot perform tasks like navigation or localization unless they have a working knowledge of maps. In this letter, we tackle the issue of making LLMs more applicable in the field of mobile education robots by helping them to understand Space Graph, a text-based map description. This study, which focuses on LLMs, is divided into several sections. It explores basic natural language processing (NLP) techniques and highlights how they can help create smooth education discussions. Examining the development of LLMs inside NLP systems, the paper explores the benefits and implementation issues of important models utilized in the education sector. Applications useful in educational discussions are described in depth, ranging from patient-focused tools like diagnosis and treatment recommendations to systems that support education providers. We provide thorough instructions and real-world examples for quick engineering, making LLM-based educational robotics solutions more accessible to novices. We demonstrate how LLM-guided upgrades can be easily included in education robotics applications using tutorial-level examples and structured prompt creation. This survey provides a thorough review and helpful advice for leveraging language models in automation development, acting as a road map for researchers navigating the rapidly changing field of LLM-driven educational robotics.

KEYWORDS

large language models (LLM), path planning, natural language processing (NLP), healthcare, robotics, applications

1 INTRODUCTION

Natural language processing (NLP) has advanced recently, resulting in the creation of large language models (LLM) such as ChatGPT and strong language models such as the generative pretrained transformer (GPT) family. Numerous NLP applications,

Vinoth Kumar, S., Saroo Raj, R.B., Praveenchandar, J., Vidhya, S., Karthick, S., Madhubala, R. (2024). Future Prospects of Large Language Models: Enabling Natural Language Processing in Educational Robotics. *International Journal of Interactive Mobile Technologies (iJIM)*, 18(23), pp. 85–97. https://doi.org/10.3991/ijim.v18i23.51419

Article submitted 2024-08-27. Revision uploaded 2024-10-13. Final acceptance 2024-10-17.

such as question-answering, text summarization, and language translation, have seen exceptional performance from these pre-trained models on massive volumes of textual data. The ChatGPT model, in particular, has proven useful in several domains, such as writing, human-machine interaction, medical care, schooling, and science [1]. One of the most significant achievements in LLM research and development is the framework known as InstructGPT, which enables guidance in fine-tuning a language model that has already been trained using reinforced learning from user input. By utilizing user feedback, this approach makes an LLM extremely flexible and adaptive, allowing it to adjust to a broad range of NLP-related duties.

An important improvement over big linguistic models that only learned text corpus through unsupervised pre-training is that RLHF allows the system to align with human tastes and beliefs. Applications in a variety of fields, including learning, healthcare, human-machine interaction, healthcare, and scientific research, are prompted by these unparalleled NLP capabilities. Due to the enormous interest and curiosity that ChatGPT has garnered, more and more research and applications are being developed to fully utilize its tremendous potential.

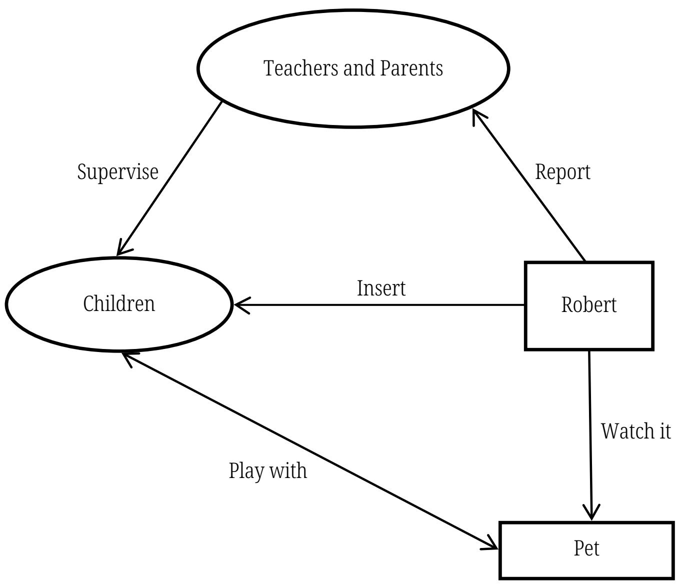

Fig. 1. An illustration of how an educational robot could be used to support preschoolers' learning

An increasing number of robotic toys and apps are being created and utilized not only in educational settings but also in playtime to enhance kids' social and moral growth, language, interaction, and cognitive abilities. Figure 1 is a typical conceptual diagram illustrating the usage of an educational robot as a companion to enhance preschoolers learning.

Moreover, deep learning is just one of the numerous machine learning-based approaches to problem-solving. In many diverse domains, including robots, healthcare, nature, production, weather observatories, understanding patterns, agriculture, business, economics, and education, artificial intelligence and machine learning can be applied. For instance [2], ML can be used to monitor and forecast climate change, population growth, urbanization, LULC changes, and resource quality. Meanwhile, ML can be used to monitor and ensure crop quality, harvest crops, assess the impact of climate change on cropping systems, identify plant diseases, assess the nutritional status of farms, monitor weeds, determine the appropriate amount and timing of herbicides and pesticides, and combat insect infestations.

As they are represented in the robot set of beliefs, semantics supplies a vital basis in establishing the impartiality of linguistic expressions. Additionally, certain situations must be considered in HRI. Certain circumstances necessitate precision, such as robotic rescue obligations [3], when a misunderstanding of signals is strictly forbidden. Recent studies on the comprehension of natural language instructions for robots working in controlled environments typically focus on specific portions of the English language. The adoption of the specific grammars that make up the allowable human language is the main tool used in the interpretation process. In this case, it permits a robotic platform to learn about routes and map navigation. Grammar-based approaches often result in highly robust stipulated systems that are also very restrictive and domain-specific. Wide coverage grammar development, however, can be fairly expensive and requires the setting up of academic profiles. On the contrary, during tasks that are less controlled and more complex, such as cleaning tasks, humans lack adherence to pre-established subsets of vocabulary. In order to effectively deal with the fluidity of natural language and utilize more generic procedures that may adjust the language models through observations, this calls for advanced understanding techniques. Robustness and domain integration are crucial in many NLP positions.

This final, feasible version of the project was just completed. This paper begins with an overview of LLM in instructional robotics using NLP techniques. The relevant prior data is examined in Section 2. Section 3 discusses the methods described in NLP for creating a typical robot for teaching purposes. Section 4 provides an experimental evaluation of NLP in educational robots, and Section 5 discusses possible future advancements for the work.

2 RELATED WORKS

Large language models have shown tremendous promise in achieving human-like intelligence in recent years, as seen by their noteworthy accomplishments. This capacity results from utilizing large-scale training datasets in addition to a significant number of model variables. Expanding on this potential, a burgeoning field uses LLMs as central regulators to build autonomous agents with decision-making abilities akin to those of humans [4]. In contrast to learning via reinforcement, LLM-based agents possess more extensive internal world knowledge, enabling them to take more educated actions even in the absence of domain-specific training. Furthermore, LLM-based agents can communicate with people through more adaptable and understandable natural language interactions.

An AI agent's brain functions as its central nervous system, just like a human's does. It retains important recollections, data, and expertise and analyzes information, makes decisions, plans, and reasons. It is the primary factor that determines the agent's capacity for intelligent behavior [5]. The perception module is then introduced. This module functions for an agent in a manner akin to that of human sensory organs. Its main purpose is to convert the agent's text-only perceptual environment into a multimedia space that incorporates a variety of sensory modalities, including text, sound, images, smell, taste, and others. The agent can now perceive details about the external environment more clearly thanks to this augmentation. The actions module, which allows an agent's action space to be expanded, is finally presented.

For the time being, we concentrate our story on NLP since that field has seen the greatest development of foundation models. Having said that, we view foundation models as a generic paradigm of AI rather than one that is in any way exclusive to NLP, much as neural networks were popularized in computer vision but exist outside of it [6]. By the end of 2018, a new seismic shift in the field of NLP was set to commence, ushering in the era of the foundational model. Technically speaking, scalability and transfer learning make foundation models possible. Today's AI systems are mostly driven by machine learning, which uses predictive models that are trained on past data to forecast future events.

Another approach to assessing the planning capabilities described in the literature is for the user to gradually engage with the LLM, asking it to highlight certain aspects of its plans each time, in the hopes that the LLM will eventually provide an actionable plan. Because the actual planning for such assessments is done by the people in the loop rather than the LLMs themselves [7], they are infamous for the Clever Hans impact. As a result, we divide our assessment into two modes: independent and supporting external planners and reasoning. Additionally, there have been initiatives that primarily relied on LLMs to "translate" natural language issue descriptions into formal specifications, which are subsequently given to competent outside designers.

They talked about ChatGPT's possible applications and its capacity to produce responses that resemble those of a human [8]. They look at how AI-driven chatbots and virtual assistants built on GPT models can help with research assignments, boost educational outcomes, and improve library services. They also include issues including biases, data privacy, and the requirement for ethical standards. All things considered, this study report emphasized the revolutionary potential of AI and GPT systems while stressing the significance of human control and responsible implementation. They discussed the effects of such models on teaching and learning, highlighting both the advantages and drawbacks of employing them.

While examples from diverse institutions indicate the usefulness of analytics in producing positive results, the financial restrictions of academic institutions must be acknowledged. Recent improvements in generative AI systems offer considerable possibilities for widening the reach of statistics to a broader spectrum of institutions [9]. To overcome these financial limitations and redirect personnel time, the usage of cheap and flexible generative AI solutions built exclusively for educational establishments should be investigated. Financially strapped institutions can also make use of the revolutionary potential of educational analytics and learning to improve teaching methods, boost student performance, and increase organizational efficiency by making these tools more widely available.

Neural networks that have been trained on extraordinarily large volumes of natural language data can produce LLMs, which have up to 3 billion variables. Language-driven reasoning tasks and text generation are among the applications that can be made of these models due to their proficiency in interpreting, producing, and contextualizing human language [10]. Neural network-based models marked a major advancement in the development of LLMs in NLP, which had previously started with models based on rules and moved through statistical frameworks. The SA process improved comprehension of contextual subtleties in language patterns over extraordinarily lengthy periods by allowing emphasis on various portions of a long input stream simultaneously.

3 METHODS AND MATERIALS

3.1 Natural language processing variables

Natural language processing is divided into two categories: natural language generation (NLG) and understanding, which progresses the work of producing and comprehending text. NLP's broad classification is shown in Figure 2. This part aims to address NLG and natural language understanding (Linguistic) (NLU).

Natural language understanding. Natural language understanding allows machines to extract ideas, individuals, emotions, phrases, and other information from natural language. It is employed in customer service apps to comprehend issues that clients may report orally or in writing. The study of language meaning, language context, and language varieties is known as semantics. Thus, it's critical to comprehend the many key terms used in NLP as well as its numerous levels. Next, we go over a few terms that are frequently used at various NLP stages.

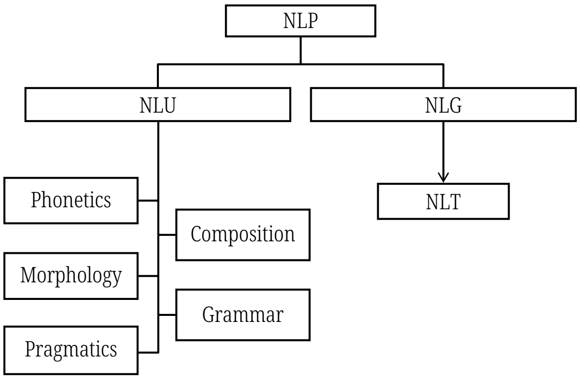

Fig. 2. General natural language processing classifications

- **Morphology:** The word's constituent parts stand for tiny meaning units or morphemes. Morphemes are the starting point for morphology, which is the nature of words. The word precancellation, for instance, can be semantically examined to reveal three distinct morphemes: the preposition pre, the initial cancella, and the endingtion. This is an example of a morphe. Humans can break down every unknown word into its constituent phrases to gain an understanding of its meaning, as morphemes always have the same meaning throughout all words. For instance, adding the suffixed indicates that the verb's action occurred in the recent past. A word's various grammatical categories—such as anxious, race, individual, attitude, component, certainty, and animacy—are altered as it acquires inflectional morphemes. The root park becomes parked when the inflectional morphemeed is added. The combination of derived morphemes with another word modifies its semantic meaning. For instance, the verb "normalize" becomes an adjective (typical) when the bound morpheme "-ize" is added to the base word "normalize."
- **Lexical:** Both humans and NLP algorithms decipher word meanings in lexical systems. Word-level knowledge is conferred by various processing kinds, the first of which assigns a part-of-speech tag to every word. Phrases that can function as many parts as possible of words are given the most likely part-of-speech tag in this processing, depending on the setting in which they appear. Single-meaning words can take the place of semantic maps at the lexical level. In actuality, the kind of representation used in the NLP systems differs depending on the linguistic theory that is applied. For instance, lemmas aim to produce the appropriate basic form—drive or derived—depending on the context in which it is employed, while stemming yields "drive" in the instance of token driven.
- **Grammar:** Following lexical-level PoS tagging, words are categorized into phrases, which are then grouped into clauses, and finally, at the grammatical level, phrases are concatenated to generate sentences. By dissecting the sentence's grammar, it highlights the proper construction of a sentence. A phrase that displays the structural dependencies between items is the result of this level. It is sometimes referred to as parsing, which identifies the sentences that have

greater significance than a single word. For readers who are currently employed in NLP and related subjects, the latter two objectives may act as an overview of the literature. They may also inspire readers to learn more about the topics covered in this paper. It should be noted that although literature surveys contain an extensive amount of work on NLP, there is still a dearth of work on regional languages, which could be the subject of future studies.

3.2 Approaches to human-robot communication

The advent of artificial intelligence has led to a redefining of HRI, which previously relied on conventional control methods. Its current research encompasses a wide range of topics, from fundamental methods to complex ethics, and as such, it finds applications across multiple domains. These fields include, but aren't limited to, human supervision and management of robots performing repetitive activities, robot work in hazardous and tight spaces, and robot socialization with humans to help people in need.

Human-robot interaction in person

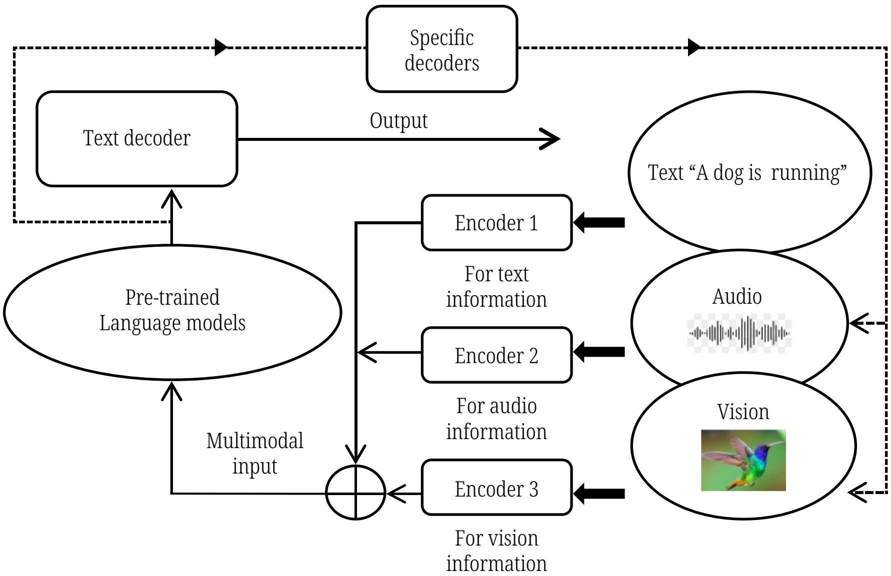

Fig. 3. Multi-modal queries are fed into the pre-trained model using the appropriate coders

When robotics was first developing, wearable devices that allowed humans and robots to sense force were commonly used to achieve conventional haptic human-robot interaction. These devices also led to the development of exoskeleton master arms, which are devices that can detect the direction and amount of force applied by the user, allowing for both free and restricted movement as well as the provision of input forces in a multimodal contact. Wearable technology is also necessary because direct physical interaction between humans and robots can overload haptics when such robots are employed for conversation and task execution.

Thus, a bracelet with vibratory motors is offered to do robot follower creation tasks guided by the trajectories of the human user, serving as a suitable illustration of this cleavage. This is why these gadgets have been transformed into virtual environments through the development of virtual reality (VR) technology. Several motion-generating techniques are demonstrated that enable a robot to navigate an obstacle-filled virtual environment and provide tactile feedback for human perception combined with the use of VR displays.

3.3 A robotic education system is being proposed

The intended ER system in this work aims to assist young children who are only beginning to learn the alphabet in writing the characters correctly. Singular value equalization is used to first improve the lighting of the recorded sequences. A multi-diagonal matrix filter is then used to detect the edges of the characters, and a part-based tree-structured algorithm is used to detect the characters. The general block diagram of the suggested method is shown in Figure 4.

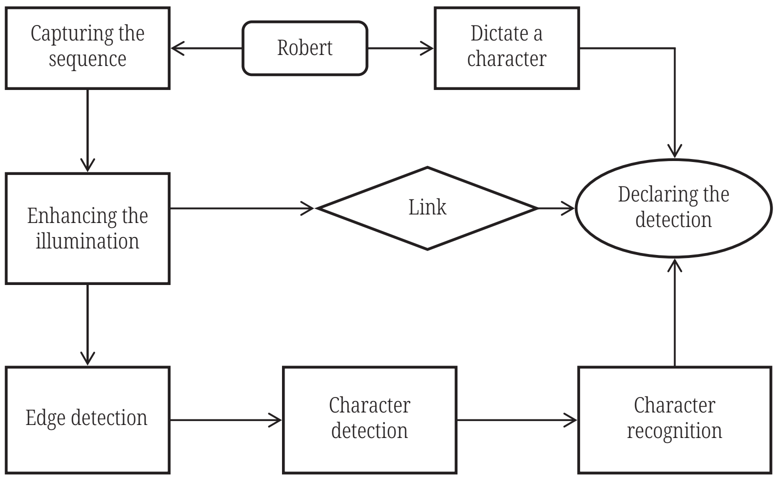

Fig. 4. Block schematic of the suggested robotic teaching system

Singular value equalization-based illumination enhancement is being used to improve the illumination of each frame in the sequence. The augmented frame will be transformed into binary form, and a multi-diagonal matrix filter will be used to extract the character edges. Part-based tree-structured letter detection uses the extracted edges as an input. Every character is depicted by a tree that consists of nodes and their topological relationships. The histogram of oriented gradients (HOG) is employed as a character descriptor during the detecting step. The character that was dictated by the robot earlier will be used to verify the recognized character. The robot will congratulate the student if they are both the same; if not, it will urge the student to give it another go. Should the robot be unable to identify the character, it will prompt the student to rewrite the alphabet.

Robots are employed in education in a manner akin to the contact-free notion. This indicates that a robot can assist with learning remotely even when it is not physically present with students. Your study used a real humanoid robot that was used in a classroom to help the teacher teach English. A robotic handwriting recognition system operates on its own. When youngsters communicate closely with a robot, they eventually reach a point where they want to write quickly to satisfy the robot's demands or clearly enough for the robot to grasp what they are writing.

4 IMPLEMENTATION AND EXPERIMENTAL RESULTS

Results for the topographical test that agree with the ground truth are considered successful. A successful completion of the hierarchical test is defined as determining the right structure. Answers are first graded using ChatGPT-3.5, and a human verifier checks the results. Based on each model's success rate (which ranges from 0 to 1) on the test information sets, each model's performance is assessed.

It's crucial to remember that ChatGPT-4 has a high success rate, but all of its answers are lengthy. Even when the prompt specifies brief outputs, such as "only output the room figures," this exuberance continues. Although this talkativeness might be appropriate in human-to-human interactions, it makes it difficult for conventional machines to make use of the LLMs' responses. Even yet, if the room numbers agreed with the real information, we considered those responses to be accurate.

Table 1. The comparison of the success rate of ChatGPT-3.5 and 4 on different prompt levels and OSMAG variants

Task Description Level	Original osmAG		osmAG Variant 1		osmAG Variant 2	
	Chat GTP3.5	Chat GTP4	Chat GTP3.5	Chat GTP4	Chat GTP3.5	Chat GTP4
Level 1	0.55	0.86	0.51	0.96	0.70	0.96
Level 2	0.43	0.88	0.46	0.98	0.71	0.96
Level 3	0.50	0.88	0.58	0.97	0.70	0.97

As indicated in Section 4, we have been testing the osmAG topology comprehension performance of ChatGPT-3.5 and ChatGPT-4.0 (gpt-4-0125-preview) utilizing three levels of prompts. The results are shown in Table 1, where it is not always the case that better performance results from providing an instance in the prompt—as one may have predicted.

Nevertheless, as Level 3 is marginally superior, we select it as our prompt in the training dataset for topological tasks. On ChatGPT-3.5 and ChatGPT-4, we compared the original osmAG with two of its variations that had varying levels of prompting. Based on the findings compiled in Table 1, osmAG Variant 2 performs better than the others on each model. As a result, we have decided to adopt this version in our fine-tuning dataset as the recommended map format for learning management systems.

Table 2. A maximum detail of fine-tuning

Hyperparameter	Fine-Tune LLaMA2-7B on Topological Task	Fine-Tune LLaMA2-13B on Topological Task	Fine-Tune LLaMA2-7B on Hierarchical Task	Fine-Tune LLaMA2-13B on Hierarchical Task
Type of GPUs	NVIDIA A40	NVIDIA A100	NVIDIA A40	NVIDIA A100
Number of GPUs	5	5	5	5
Training Time	6.7 hours	5.6 hours	3 hours	1.6 hours
Epochs	2.6	2.6	7.77	7.77
Size of Training Dataset	12521	12521	1057	1057
Size of Adapter	16.9MB	26.3MB	16.9MB	26.3MB

We evaluate the topology and hierarchy map understanding capabilities of the LLaMA2-7B and LLaMA2-13B models using Dataset 1–5, which we created using alert Level 3 and osmAG Variant 2. For comparison, we also use the APIs of ChatGPT-3.5 and ChatGPT-4 on these five datasets. The results of topological comprehension tasks without fine-tuning are displayed in Table 2, where the average success rate on both LLaMA2 models is only 0.1, far from being useful in real-world scenarios. The estimated success rate of ChatGPT-3.5 is 0.5, making it unsuitable for real-world use.

The results of the hierarchy experiment are likewise displayed in Table 3. The LLaMA2-13B models have a 0.55 successful rate, which is significantly better than the 0.19 successful rate of LLaMA2-7B, but it remains insufficient for practical application. With an achievement rate of 0.66, ChatGPT-3.5 beats LLaMA2-13B, but it is still not good enough for widespread use. On both tasks, ChatGPT-4 has an excellent success percentage.

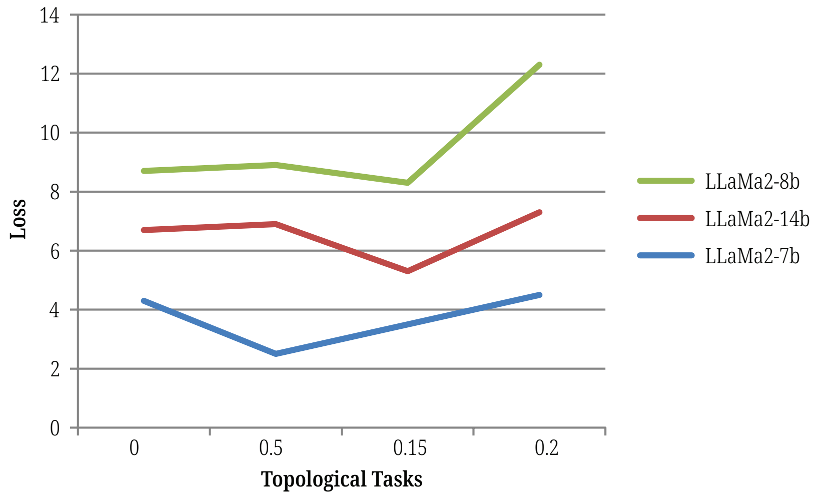

Fig. 5. Handling Topological and Hierarchical Tasks: Optimizing LLaMA2-8B and LLaMA2-14B

Our foundation model is the publicly accessible LLaMA2 model from Meta, which is based on an amplifier architecture that solely includes decoders. Additive low-rank matrices are introduced to the neural network layers by LoRA, which updates these matrices during fine-tuning while maintaining the original weight matrix's frozen state. This method significantly reduces the number of parameters that need to be learned. With substantially less processing power, LoRA makes it possible to effectively train LLMs by lowering the total number of trainable parameters. We use the open-source repository to fine-tune the LLaMA2 framework, focusing on dataset 2 for topological tasks and dataset 4 for hierarchical tasks. Deep Speed and the Zero Redundancy Optimizer (ZeRO) were used to optimize this procedure in order to effectively control memory utilization. We choose to use a cosine learning rate planner, and the learning rate is set to 5e-5. Table 3 provides specific fine-tuning settings, and Figure 5 shows the tuning procedure.

Table 3. A comparison of the LLMs success rates

	LLaMA2-7B	LLaMA2-13B	Fine-Tuned LLaMA2-7B	Fine-Tuned LLaMA2-13B	ChatGPT-3.5	ChatGPT-4.0
Dataset 1 (T)	0.11	0.13	0.99 (0.79)	0.98 (0.92)	0.55	0.100
Dataset 2 (T)	0.06	0.667	0.94 (0.61)	0.95 (0.95)	0.51	0.89
Dataset 3 (T)	0.12	0.15	0.89 (0.76)	0.97 (0.93)	0.54	0.97
Dataset 4 (T)	0.20	0.56	1.0 (0.99)	0.99 (0.99)	0.67	0.100
Dataset 5 (T)	0.96	0.96	0.96	0.96	2	2

When the layout is utilized for fine-tuning, the LLaMA2-8B model performs better than 0.9 on test data from Dataset 1–2; however, when unseen layouts are employed, its performance degrades to 0.79.

On the other hand, the LLaMA2-14B model shows a 0.87 success rate, indicating stronger generalization to novel map designs. We use a fixed question in our initial training dataset. In contrast, we use a number of random prompts during testing in addition to the question used for fine-tuning to evaluate the capacity of the refined model to generalize across a range of variables. The fine-tuned LLaMA2-7B model's performance declines, whereas the fine-tuned LLaMA2-13B model's efficiency stays consistently excellent, according to the results shown in Table 3 in brackets. This implies that the refined LLaMA2-13B model, which exhibits significant generalization to previously unidentified prompts, is better for tasks involving unexpected human contact. On the other hand, the optimized LLaMA2-7B model is a good choice in situations when the prompt is produced by a static programmer, particularly in cases where computing power is limited. The LLaMA2-8B and LLaMA2-14B systems both attain a success rate of 1 in hierarchy tasks after fine-tuning. They are very useful because they also generalize effectively on unseen cues.

5 CONCLUSIONS

An overview of the incorporation of LLMs into diverse robotic systems and tasks is presented in this work. According to our investigation, LLMs have remarkable reasoning, language comprehension, and multimodal processing skills that can greatly improve robots' comprehension of directions, surroundings, and necessary actions. The purpose of this letter is to describe osmAG, a flexible map representation for future LLM robot systems that is meant to be comprehensible by human beings, compatible with conventional robotic algorithms, and interpreted by LLMs. We give datasets to assess the model's understanding of osmAG and osmAG variations to enhance performance for proprietary models such as ChatGPT.

In the case of open-source models such as LLaMA2, we offer fine-tuned adaptations of the LLaMA2 models for thorough testing in addition to datasets and techniques for creating datasets to refine the predictions. We understand that the length of actual pathways is important in real-world robotic applications. By removing metrics from the map, we are unable to ensure the best result.

We extensively review the literature on LLM-based autonomous agents in the present review. We present and discuss these investigations from three perspectives: the design, implementation, and assessment of the agents. We offer a thorough taxonomy for each of these elements to help make connections between the current research, including a summary of the key methods and their evolutionary histories. We expect quick advancements in the integration of language models with robotics and simulation-based learning as they continue to amass substantial grounded knowledge from multimodal information. This could make it possible for intelligent robots to be developed intuitively and validated completely in simulation utilizing simulation-to-real approaches prior to deployment. Such advancements have the potential to significantly improve and change the way we design, test, and implement intelligent robotic systems.

In the present study, we have developed a novel educational robotic system that teaches beginning students how to write the alphabet correctly. The system makes use of sophisticated computer vision algorithms to identify written characters, and because of its design and functionality, students' learning abilities are enhanced.

All things considered, the synergistic fusion of robotics and NLP is a promising field full of chances and difficulties that demand further multidisciplinary study in the future.

6 REFERENCES

[1] Y. Liu *et al.*, "Summary of ChatGPT-related research and perspective towards the future of large language models," *Meta-Radiology*, vol. 1, no. 2, p. 100017, 2023. https://doi.org/10.1016/j.metrad.2023.100017

[2] N. Kumari and S. Pandey, "Application of artificial intelligence in environmental sustainability and climate change," in *Visualization Techniques for Climate Change with Machine Learning and Artificial Intelligence*, pp. 293–316, 2023. https://doi.org/10.1016/B978-0-323-99714-0.00018-2

[3] J. Wang *et al.*, "Large language models for robotics: Opportunities, challenges, and perspectives," *arXiv preprint arXiv:2401.04334*, 2024.

[4] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: State of the art, current trends and challenges," *Multimedia Tools and Applications*, vol. 82, pp. 3713–3744, 2023. https://doi.org/10.1007/s11042-022-13428-4

[5] Z. Xi *et al.*, "The rise and potential of large language model-based agents: A survey," *arXiv preprint arXiv:2309.07864*, 2023.

[6] R. Bommasani *et al.*, "On the opportunities and risks of foundation models," *arXiv preprint arXiv:2108.07258*, 2021.

[7] K. Valmeekam, M. Marquez, S. Sreedharan, and S. Kambhampati, "On the planning abilities of large language models: A critical investigation," *Advances in Neural Information Processing Systems*, vol. 36, pp. 75993–76005, 2023.

[8] G. Yenduri *et al.*, "GPT (generative pre-trained transformer)—A comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions," *IEEE Access*, vol. 12, pp. 54608–54649, 2024. https://doi.org/10.1109/ACCESS.2024.3389497

[9] J. Hutson and D. Plate, "Enhancing institutional assessment and reporting through conversational technologies: Exploring the potential of AI-powered tools and natural language processing," *Journal of Artificial Intelligence and Robotics*, vol. 1, no. 1, pp. 11–22, 2023. https://doi.org/10.59232/AIR-V1I1P102

[10] F. Zeng, W. Gan, Y. Wang, N. Liu, and P. S. Yu, "Large language models for robotics: A survey," *arXiv preprint arXiv:2311.07226*, 2023.

[11] F. Xie and S. Schwertfeger, "Empowering robotics with large language models: osmAG map comprehension with LLMs," *arXiv preprint arXiv:2403.08228*, 2024.

[12] C. Zhang, J. Chen, J. Li, Y. Peng, and Z. Mao, "Large language models for human-robot interaction: A review," *Biomimetic Intelligence and Robotics*, vol. 3, no. 4, p. 100131, 2023. https://doi.org/10.1016/j.birob.2023.100131

[13] B. Zhang and H. Soh, "Large language models as zero-shot human models for human-robot interaction," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 7961–7968. https://doi.org/10.1109/IROS55552.2023.10341488

[14] C. Y. Kim, C. P. Lee, and B. Mutlu, "Understanding Large-Language Model (LLM)-powered human-robot interaction," in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 2024, pp. 371–380. https://doi.org/10.1145/3610977.3634966

[15] X. Zhao, M. Li, C. Weber, M. B. Hafez, and S. Wermter, "Chat with the environment: Interactive multimodal perception using large language models," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 3590–3596. https://doi.org/10.1109/IROS55552.2023.10342363

[16] H. Altin, A. Aabloo, and G. Anbarjafari, "New era for educational robotics: Replacing teachers with a robotic system to teach alphabet writing," in *Proceedings of the 4th International Workshop Teaching Robotics, Teaching with Robotics & 5th International Conference Robotics in Education*, 2014, pp. 164–166.

[17] B. He, M. Xia, X. Yu, P. Jian, H. Meng, and Z. Chen, "An educational robot system of visual question answering for preschoolers," in *2017 2nd International Conference on Robotics and Automation Engineering (ICRAE)*, 2017, pp. 441–445. https://doi.org/10.1109/ICRAE.2017.8291426

7 AUTHORS

Dr. S. Vinoth Kumar is presently working as Professor in the Department of Computer Science and Engineering at Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai. He received his Bachelor of Engineering in Computer Science from Anna University, Master of Engineering in Computer Science from Anna University, Tirunelveli. He completed his PhD at Karpagam University, Coimbatore. His research interests include Image Mining, Data Mining, Networks and Cloud Computing. He published 30 papers in refereed international journals and presented 20 papers in national and international conferences. He is an active member of IAENG. He has visited various countries like South Korea, Malaysia and Singapore for Faculty Exchange Program on Academic Affairs (E-mail: drsvinothkumar@veltech.edu.in).

Dr. R. B. Saroo Raj is working as an Assistant Professor in the Department of IT, Rajalakshmi Engineering College, Chennai, Tamil Nadu, India. He obtained his Bachelor's degree in Information Technology (B.Tech IT) in 2009 from Adhiyamaan College of Engineering, Hosur, and later earned his Master's degree in Computer and Communication. His Persistent interest in research, particularly in the field of Trajectory Data Analysis, led his to pursue Doctoral research in Semantic Data Analysis in ITS at Sathyabama Institute of Science and Technology, Chennai, India, under the supervision of Dr. S. Prayla Shyry in the Department of Computer Science and Engineering. He has published research articles in peer-reviewed journals, including intelligent automation and soft Computing and Multimedia Tools and Applications (E-mail: sarooraj.rb@rajalakshmi.edu.in).

Dr. J. Praveenchandar, BE, ME, PhD is working as an Assistant Professor (S.G) in the Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore. He has as got more than 15 years of teaching and research experience in various reputed engineering colleges in Tamil Nadu in the field of Computer Science and Engineering, in specific Cloud computing and Big Data Analytics. He Completed his BE (CSE), ME (CSE), and PhD (ICE) at the Anna University, Chennai. He has published his research articles in many reputed SCI/Scopus/WoS indexing journals and presented his ideas in many International conferences. He has three International and Indian design patents and copyrights. He is the reviewer of many SCI indexed Journals. His research interests include Cloud computing and Big Data Analytics (E-mail: praveenchandar@karunya.edu).

S. Vidhya is working as Assistant Professor in the Department of Information Science and Engineering at The Oxford College of Engineering, Bangalore. She received her BE degree in Computer Science and Engineering from Velalar college of Engineering and Technology, Erode, India, ME degree in Computer Science and Engineering from Sona College of Technology, Salem, Currently, she is pursuing PhD at the Anna University, Chennai, India. She is interested in Data Science, AI, and ML.

She published 10 papers in refereed international journals and presented 6 papers in national and international conferences (E-mail: vidhyatoce@gmail.com).

Dr. S. Karthick is a highly accomplished and experienced academic with over 18 years of experience in the field of Computer Science & Engineering. His impressive profile includes an Assistant Professorship position in the Department of Computer Science & Engineering. He is a member of prestigious organizations such as ISTE and CSI. His research interests lie in the areas of Internet of Things, Wireless Sensor Networks, and Software Engineering. He has authored an impressive book and secured two granted Indian patents, in addition to publishing 20 research articles in reputed international peer-reviewed journals and 10 papers in international peer-reviewed conference proceedings. He has also contributed significantly to the academic community by organizing numerous workshops, FDPs, international conferences as an organizing chair. His expertise in the field has earned him recognition from various international journals such as IEEE Access, Springer, IGI Global, Inderscience, EAI Cloudcomp Journals, and Expert Systems, where he has served as a peer-reviewer (E-mail: drkarthicks@veltech.edu.in).

Dr. R. Madhubala currently serves as a Lecturer in the Department of IT at the University of Technology and Applied Science, Shinas in Oman. She earned her Doctor of Philosophy degree from Vels Institute of Science, Technology and Applied Science, India 2021. Her research interests include cloud computing, artificial intelligence, networks, IoT, service-oriented architecture, and data structures. She's been a teacher for twenty-four years. Her research has been presented at conferences, in book chapters, and in 13 international journals. She has been serving as reviewer for several reputed journals (E-mail: r.madhubala@utas.edu.om).