Transforming Surgical Interventions with Embodied Intelligence for Ultrasound Robotics

Huan Xu[1]*, Jinlin Wu[1,2]*, Guanglin Cao[1,2], Zhen Chen[1✉], Zhen Lei[1,2,3], Hongbin Liu[1,2]

[1] Centre for Artificial Intelligence and Robotics (CAIR), HKISI-CAS
[2] MAIS, Institute of Automation, Chinese Academy of Sciences
[3] School of Artificial Intelligence, University of Chinese Academy of Sciences

Abstract. Ultrasonography has revolutionized non-invasive diagnostic methodologies, significantly enhancing patient outcomes across various medical domains. Despite its advancements, integrating ultrasound technology with robotic systems for automated scans presents challenges, including limited command understanding and dynamic execution capabilities. To address these challenges, this paper introduces a novel Ultrasound Embodied Intelligence system that synergistically combines ultrasound robots with large language models (LLMs) and domain-specific knowledge augmentation, enhancing ultrasound robots' intelligence and operational efficiency. Our approach employs a dual strategy: firstly, integrating LLMs with ultrasound robots to interpret doctors' verbal instructions into precise motion planning through a comprehensive understanding of ultrasound domain knowledge, including APIs and operational manuals; secondly, incorporating a dynamic execution mechanism, allowing for real-time adjustments to scanning plans based on patient movements or procedural errors. We demonstrate the effectiveness of our system through extensive experiments, including ablation studies and comparisons across various models, showcasing significant improvements in executing medical procedures from verbal commands. Our findings suggest that the proposed system improves the efficiency and quality of ultrasound scans and paves the way for further advancements in autonomous medical scanning technologies, with the potential to transform non-invasive diagnostics and streamline medical workflows. The source code is available at https://github.com/seanxuu/EmbodiedUS.

Keywords: Embodied Intelligence · Large Language Model · Ultrasound Robotics

1 Introduction

Ultrasonography is a cornerstone in non-invasive diagnostics, revolutionizing early detection in various medical fields [3,21]. Its application spans numerous disciplines, significantly enhancing patient care and outcomes [14,20,16]. This

---

∗ Equal contribution. ✉ Corresponding author.

technology has transformed the diagnosis of conditions like fetal abnormalities [22], gallbladder stones [5], and cardiovascular diseases [17], offering a window into the body's internal structures and greatly improving early diagnosis and patient management [12].

Despite the technological advances in ultrasonography, integrating robotics to enhance scanning efficiency and quality presents unresolved challenges [10]. Innovations in ultrasound robotics, including tactile sensing [2], compliant force control [18], trajectory planning [24], and image processing [4], have enabled automated patient scans. Meanwhile, challenges remain in clinical application: **1)** Instruction logic understanding, the logic in natural language instructions is difficult for robots to understand because there is no contextual information and clinical domain knowledge, and the information extracted directly from natural instructions is not enough to explain the logic [13,27]. **2)** Dynamic execution, there have been many studies on path generation for autonomous ultrasound scanning, including offline scan path generation [15,11], and online scan path generation [9]. However, it is still a challenge to adjust the scan plan in real time after execution errors.

In this work, we introduce an ultrasound embodied intelligence system that combines ultrasound robots with large language models (LLMs) to enhance their clinical performance. Our system uses LLMs to understand doctors' intentions and improve motion planning accuracy. To ensure reliable workflow and mitigate errors caused by misinterpretations, we have enriched LLMs with ultrasound-specific knowledge, such as APIs and robot handbooks. A specialized embedding model embeds the most relevant execution APIs and operational advice, aligning robot actions with doctors' intentions. This technique improves the capability of ultrasound robots to fulfill clinical demands and provides a more precise and efficient solution for doctors. Additionally, we have developed a dynamic execution mechanism inspired by the ReAct framework [26]. This system allows medical staff to verbally command robots, which then interpret the commands into precise scanning paths, minimizing the need for manual adjustments. The mechanism works through a thought-action-observation cycle, which continuously engages with the robot's APIs to execute commands seamlessly. Our system showcases the potential of LLMs in revolutionizing robotic precision and autonomy in the healthcare industry.

2   Methodology

2.1   Process Formulation

To comprehensively understand our system processes, we string together the individual methods as a formulation that serves as the basis of our methodology. This formulation guides the creation of our algorithms and their implementation. For any given ultrasound scan task, the process can be defined as follows:

$$C = R_n(R_{n-1}(\ldots R_2(R_1(A(U(D)))) \ldots)), \tag{1}$$
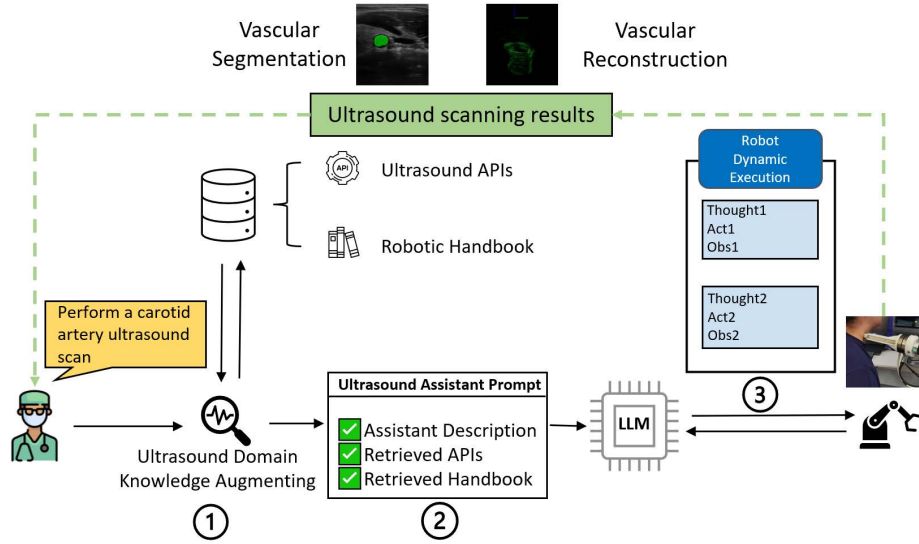
where

Fig. 1: The proposed system framework. Our Embodied Intelligence system interprets and executes medical procedures through verbal commands. This system has three components: a foundational large language model for command interpretation, the Ultrasound Domain Knowledge Augmenting technique for enhanced contextual understanding, and Robot Dynamic Execution for converting instructions into robotic actions.

- D represents the Doctor's Instructions.
- U denotes the process of Ultrasound Domain Knowledge Augmenting.
- A is the Assemble Ultrasound Assistant Prompt with retrieved APIs and retrieved robot handbook.
- R_i signifies the ith iteration of interaction with the robot through Robot Dynamic Execution, for $i = 1, 2, \ldots, n$.
- C is the Task Completion, indicating the end of the process after n iterations of dynamic execution.

2.2   Ultrasound Domain Knowledge Augmenting

Domain Knowledge Search The core of our approach utilizes a similarity search algorithm to tap into the ultrasound domain knowledge database, leveraging cosine similarity to match user queries with relevant knowledge. This method transforms user queries and knowledge base entries into vector representations within a d-dimensional space via an embedding model. In particular, we find the entry with the highest *cosine similarity* to the query vector, as follows:

$$\mathcal{S}(\mathbf{A}, \mathbf{B}) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}, \tag{2}$$

> **Role**: As a system proficient in sequential problem solving, you can leverage your expertise in ultrasound technology-related APIs to respond to user requests through the API, step by step.
> **Instruction**: You can access specific API information in this dialogue. Assess the necessity of invoking an API to address the user's issue. If an API call is warranted, provide the request in JSON format, including `api_name` and `parameters` fields. Enclose the API call request with `<|sot|>` and `<|eot|>` markers. Based on the API call outcome, craft an appropriate response. If an API call is not required, directly furnish the relevant response.
> **APIs List**: Note that only one API may be invoked per interaction. Below is the list of accessible APIs: `<api_list>`

Fig. 2: This figure presents an Ultrasound Assistant Prompt, detailing its role as an entity skilled in sequential problem-solving and its ability to respond to user queries via APIs, focusing on ultrasound technology-related APIs. The prompt outlines instructions for assessing the need for API calls to address user issues, including the format for such requests, and lists the available APIs for use.

where $\mathbf{A}$ and $\mathbf{B}$ are the representations of user queries and knowledge base entries, respectively.

Ultrasound APIs Retrieval. To streamline the selection of ultrasound APIs by LLMs, we refine the Ultrasound APIs Retrieval (UAR) method. This method relies on a dataset where each API is paired with a narrative describing its use context, improving tool selection for ultrasound scanning. The dataset is structured as $\mathcal{D} = \{(T_1, U_1), (T_2, U_2), \ldots, (T_n, U_n)\}$, where T represents tools and U represents usage of tools, facilitating accurate tool identification based on scenario-specific requirements.

Robotic Handbook Retrieval. At the same time, how LLMs discern the correct order and logical sequence of these API calls remains a huge hurdle. We present the Robotic Handbook Retrieval (RHR) method to address this issue. This approach enriches LLMs's context with a procedural knowledge base accessible through vectorized input queries. The core of this method is a similarity search. We systematically paired instructions with the handbook to guarantee the identification of relevant instructions during the similarity search, followed by the extraction of corresponding handbooks.

2.3   Ultrasound Assistant Prompt

Facing the challenge of commands lacking context, we enhance model comprehension and intent accuracy through structured prompts and added context. This approach ensures commands are interpreted precisely, aligning results with user expectations. Additionally, prompts are integrated with an execution session, allowing for specific output structures to trigger various APIs. The Fig. 2 illustrates our approach succinctly.

APIs Name: Image_Seg

APIs Description: The APIs allow for segmentation of the scan results, segmenting the patient's artery.

APIs Parameters:
– position (float, float): Specify the location of the artery.
– threshold (float): Specify the threshold of the segmentation area.

Robot Handbook Description: The carotid artery ultrasound process involves initializing the depth camera, displaying the artery model, activating the robotic system, segmenting the scan image, and finally generating and printing the report.

Fig. 3: Overview of Ultrasound API Functionality and Robotic Procedure: This figure presents a detailed depiction of the Image_Seg API capabilities for artery segmentation in scan results, alongside a step-by-step guide to the carotid artery ultrasound process facilitated by a robotic system.

Dynamic Execution Cycle Description:
The cycle consists of the following steps:

1. Observe the environment and current state to gather observations.
2. Think about the observations along with robot handbooks to make decisions.
3. Act based on the analysis to perform optimized actions.
4. Update the execution environment and state based on the actions taken.
5. Repeat the process until the task is complete or an error threshold is exceeded.

Fig. 4: This figure outlines the cyclical process utilized in robotic systems for dynamic execution, comprising steps such as observation, thought, action, environment updating, and repetition until task completion or error threshold breach.

2.4 Robot Dynamic Execution

Building upon the inspiration drawn from the ReAct framework mentioned in the introduction, we introduce a dynamic execution mechanism for robotic systems. This mechanism operates through a cyclical process comprising three main steps: Observation, Thought, and Action. This operational cycle aims to minimize errors and optimize task execution by continuously adapting to real-time feedback. The process is detailed in Fig. 4.

3 Experiments

3.1 Experimental Setup

Robot System. The robotic arm used in our system is the RM65-B model from RealMan-Robotics. We employ the H20 ultrasound machine from Angell. The depth camera is RealSense from Intel. Additionally, the force sensor integrated
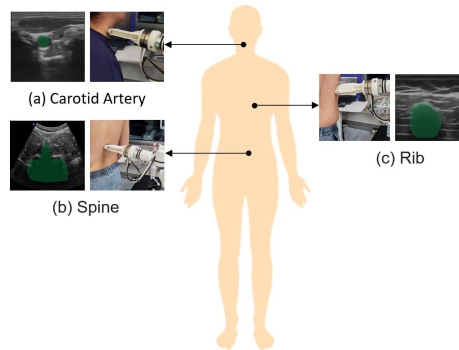
Fig. 5: Illustration of the ultrasound scanning and subsequent image segmentation of (a) carotid artery, (b) spine, and (c) rib, as conducted in our experiments.

Table 1: Comparison of embedded model training results demonstrating the superior performance of our domain adapted-model on different modules

Module	Model	Recall@1	Recall@3	Recall@10
Ultrasound APIs Retrieval	bge-large-en-v1.5	0.82	0.94	0.97
	Ours	**0.86**	**0.96**	**0.99**
Robotic Handbook Retrieval	bge-large-en-v1.5	0.76	0.95	0.97
	Ours	**0.88**	**0.97**	**0.98**

into our setup is M3815B from Sunrise. The control algorithms for the robot system are written by ourselves.

Models Configuration. To evaluate the effectiveness of our proposed enhancements on system performance, we used the foundational model GPT4-Turbo [19], with initial parameters set to *Temperature = 0.7* and *Top P = 0.95*. For embedding, we used the domain-adapted bge-large-en-v1.5 [25] model alongside FAISS [6] for efficient data embedding storage and vector search operations. The performance of the domain-adapted model is compared with the performance of the original model, as shown in Table 1.

Datasets and Preprocessing. We conducted experiments with a synthetic dataset, generating 522 instances for the Robotic Handbook and 622 for the Ultrasound APIs. This dataset, designed to mirror the complexities of ultrasound scans and API calls, allowed for comprehensive evaluation across different scenarios. This dataset is also used to train the embedding model.

Experimental Framework and Metrics. Our experimental framework was carefully designed to assess the impact of each augmentation introduced. We ensured the robustness and reliability of our findings by replicating each experimental step twenty times. Various models' performance within our system was compared using defined metrics. Furthermore, we performed tests on multiple parts of the human body, showcasing the practical applicability of our approach, with visual results presented in Fig. 5.

Table 2: Ablation and Different Models Study on APIs Execution Success Rates

Type	Module	First Step (%)	Overall (%)
Module Ablation	LLMs + UAR	35	0
	LLMs + UAR + RHR	**100**	**80**
Different Models	Qwen1.5-1.8B-Chat [1]	30	0
	Qwen1.5-14B-Chat [1]	45	10
	Llama2-7B [23]	50	10
	Llama2-13B [23]	65	10
	Mistral-7B-v0.1 [7]	65	20
	Mixtral-8x7B-Instruct-v0.1	70	45

3.2 Results and Analysis

In our experiments, we conducted ablation studies on different modules and explored the performance of various models, complemented by case studies to illustrate their practical impacts. The experimental data from the first two experiments are shown in Table 2.

Effectiveness of Modules. The ablation study results reveal the stepwise performance boost of foundational LLMs with each added module. Initially, LLMs without modifications fail to successfully execute the first API call due to a lack of specific API knowledge after receiving natural language instructions. The introduction of the UAR module significantly improves performance, achieving a 35% success rate in the initial evaluation phase. This suggests that even a basic list of APIs enables LLMs to understand better and initiate API calls, leveraging their inherent natural language processing abilities. Adding the RHR module further enhances performance across all evaluation phases, indicating its role in improving task initiation and maintaining performance growth despite diminishing returns in later stages due to task complexity. This improvement highlights the value of structured guidance in API selection, demonstrating LLMs' ability to utilize structured information for task-specific enhancements.

Effectiveness of Different LLMs. In our experimental analysis, we evaluated the performance of various LLMs augmented with ultrasound domain knowledge to determine their effectiveness in controlling ultrasound robotics through natural language instructions. The models under consideration exhibited varying success in executing the initial API calls and achieving overall task completion.

The performance comparison across models reveals significant variability. Notably, the Mixtral-8x7B-Instruct-v0.1 [8] model outperforms others in both the initial step success rate (70%) and overall task completion rate (45%). This suggests that larger model sizes and specific domain knowledge integration play crucial roles in enhancing task-specific performance.

Cases Study. In our experiments, while different models demonstrated varied capabilities within this system, we identified several common failure patterns. We used the same command for different LLMs: *Perform a carotid artery ultrasound scan*, given or without UAR and RHR, and got the following results. The example is shown in Fig. 6.
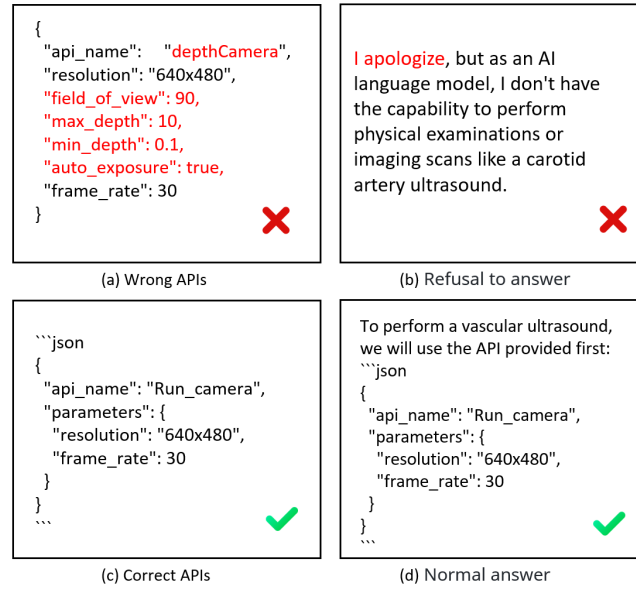
Fig. 6: Case comparisons. (a) Wrong API: LLMs produce incorrect API information due to their absence of prior knowledge, with both the name and parameters of this API being a misconception propagated by the model. (b) Refusal to answer: LLMs, owing to their deficiency in contextual understanding, often decline to provide responses for precise operational directives. We hypothesize that this reluctance is related to the fact that these models are aligned with human preferences at the time of training. In contrast, (c) and (d) are the correct execution.

4   Conclusion

In this study, we introduce an innovative Embodied Intelligence system designed to enhance ultrasound robotics by integrating advanced LLMs and domain-specific knowledge. This system, adept at interpreting verbal instructions for ultrasound scans, features three key modules that boost its responsiveness and accuracy. Our results show the system's ability to accurately carry out medical procedures from verbal commands, representing a significant step towards fully autonomous medical scans. This work highlights the transformative potential of embodied intelligence in non-invasive diagnostics and paves the way for further research to broaden its healthcare applications, with the ultimate goal of streamlining medical workflows.

Disclosure of Interests. The authors declare no competing interests.

References

1. Bai, J., Bai, S., Chu, Y., Cui, Z., Dang, K., Deng, X., Fan, Y., Ge, W., Han, Y., Huang, F., Hui, B., Ji, L., Li, M., Lin, J., Lin, R., Liu, D., Liu, G., Lu, C., Lu, K., Ma, J., Men, R., Ren, X., Ren, X., Tan, C., Tan, S., Tu, J., Wang, P., Wang, S., Wang, W., Wu, S., Xu, B., Xu, J., Yang, A., Yang, H., Yang, J., Yang, S., Yao, Y., Yu, B., Yuan, H., Yuan, Z., Zhang, J., Zhang, X., Zhang, Y., Zhang, Z., Zhou, C., Zhou, J., Zhou, X., Zhu, T.: Qwen technical report. arXiv preprint arXiv:2309.16609 (2023)
2. Cao, G., Chen, M., Hu, J., Liu, H.: An ultra-fast intrinsic contact sensing method for medical instruments with arbitrary shape. IEEE Robotics and Automation Letters (2023)
3. Chan, V., Perlas, A.: Basics of ultrasound imaging. Atlas of ultrasound-guided procedures in interventional pain management pp. 13–19 (2011)
4. Chen, G., Li, L., Zhang, J., Dai, Y.: Rethinking the unpretentious u-net for medical ultrasound image segmentation. Pattern Recognition p. 109728 (2023)
5. Cooperberg, P., Burhenne, H.: Real-time ultrasonography: diagnostic technique of choice in calculous gallbladder disease. New England Journal of Medicine **302**(23), 1277–1279 (1980)
6. Douze, M., Guzhva, A., Deng, C., Johnson, J., Szilvasy, G., Mazaré, P.E., Lomeli, M., Hosseini, L., Jégou, H.: The faiss library (2024)
7. Jiang, A.Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D.S., de las Casas, D., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Lavaud, L.R., Lachaux, M.A., Stock, P., Scao, T.L., Lavril, T., Wang, T., Lacroix, T., Sayed, W.E.: Mistral 7b (2023)
8. Jiang, A.Q., Sablayrolles, A., Roux, A., Mensch, A., Savary, B., Bamford, C., Chaplot, D.S., de las Casas, D., Hanna, E.B., Bressand, F., Lengyel, G., Bour, G., Lample, G., Lavaud, L.R., Saulnier, L., Lachaux, M.A., Stock, P., Subramanian, S., Yang, S., Antoniak, S., Scao, T.L., Gervet, T., Lavril, T., Wang, T., Lacroix, T., Sayed, W.E.: Mixtral of experts (2024)
9. Jiang, Z., Li, Z., Grimm, M., Zhou, M., Esposito, M., Wein, W., Stechele, W., Wendler, T., Navab, N.: Autonomous robotic screening of tubular structures based only on real-time ultrasound imaging feedback. IEEE Transactions on Industrial Electronics **69**, 7064–7075 (2020)
10. Jiang, Z., Salcudean, S.E., Navab, N.: Robotic ultrasound imaging: State-of-the-art and future perspectives. Medical image analysis p. 102878 (2023)
11. Jiang, Z., Wang, H., Li, Z., Grimm, M., Zhou, M., Eck, U., Brecht, S.V., Lueth, T.C., Wendler, T., Navab, N.: Motion-aware robotic 3d ultrasound. 2021 IEEE International Conference on Robotics and Automation (ICRA) pp. 12494–12500 (2021)
12. Kasoju, N., Remya, N., Sasi, R., Sujesh, S., Soman, B., Kesavadas, C., Muraleedharan, C., Varma, P.H., Behari, S.: Digital health: trends, opportunities and challenges in medical devices, pharma and bio-technology. CSI Transactions on ICT pp. 1–20 (2023)
13. Knepper, R.A., Tellex, S., Li, A., Roy, N., Rus, D.: Recovering from failure by asking for help. Autonomous Robots **39**, 347–362 (2015)
14. Mayo, P.H., Copetti, R., Feller-Kopman, D.J., Mathis, G., Maury, É., Maury, É., Mongodi, S., Mojoli, F., Volpicelli, G., Zanobetti, M.: Thoracic ultrasonography: a narrative review. Intensive Care Medicine **45**, 1200 – 1211 (2019)

15. Merouche, S., Allard, L., Montagnon, E., Soulez, G., Bigras, P., Cloutier, G.: A robotic ultrasound scanner for automatic vessel tracking and three-dimensional reconstruction of b-mode images. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control **63**, 35–46 (2016)

16. Moore, C.L., Copel, J.: Point-of-care ultrasonography. The New England journal of medicine **364 8**, 749–57 (2011)

17. Nezu, T., Hosomi, N.: Usefulness of carotid ultrasonography for risk stratification of cerebral and cardiovascular disease. Journal of atherosclerosis and thrombosis **27**(10), 1023–1035 (2020)

18. Ning, G., Chen, J., Zhang, X., Liao, H.: Force-guided autonomous robotic ultrasound scanning control method for soft uncertain environment. International Journal of Computer Assisted Radiology and Surgery **16**(12), 2189–2199 (2021)

19. OpenAI, :, Achiam, J., Adler, S., et al., S.A.: Gpt-4 technical report (2023)

20. Robba, C., Goffi, A., Geeraerts, T., Cardim, D., Via, G., Czosnyka, M., Park, S., Sarwal, A., Padayachy, L., Rasulo, F.A., Citerio, G.: Brain ultrasonography: methodology, basic and advanced principles and clinical applications. a narrative review. Intensive Care Medicine **45**, 913–927 (2019)

21. Shung, K.K.: Diagnostic ultrasound: Past, present, and future. J Med Biol Eng **31**(6), 371–4 (2011)

22. Sonek, J.: First trimester ultrasonography in screening and detection of fetal anomalies. In: American Journal of Medical Genetics Part C: Seminars in Medical Genetics. vol. 145, pp. 45–61. Wiley Online Library (2007)

23. Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C.C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., Fuller, B., Gao, C., Goswami, V., Goyal, N., Hartshorn, A., Hosseini, S., Hou, R., Inan, H., Kardas, M., Kerkez, V., Khabsa, M., Kloumann, I., Korenev, A., Koura, P.S., Lachaux, M.A., Lavril, T., Lee, J., Liskovich, D., Lu, Y., Mao, Y., Martinet, X., Mihaylov, T., Mishra, P., Molybog, I., Nie, Y., Poulton, A., Reizenstein, J., Rungta, R., Saladi, K., Schelten, A., Silva, R., Smith, E.M., Subramanian, R., Tan, X.E., Tang, B., Taylor, R., Williams, A., Kuan, J.X., Xu, P., Yan, Z., Zarov, I., Zhang, Y., Fan, A., Kambadur, M., Narang, S., Rodriguez, A., Stojnic, R., Edunov, S., Scialom, T.: Llama 2: Open foundation and fine-tuned chat models (2023)

24. Wang, Z., Zhao, B., Zhang, P., Yao, L., Wang, Q., Li, B., Meng, M.Q.H., Hu, Y.: Full-coverage path planning and stable interaction control for automated robotic breast ultrasound scanning. IEEE Transactions on Industrial Electronics **70**(7), 7051–7061 (2022)

25. Xiao, S., Liu, Z., Zhang, P., Muennighoff, N.: C-pack: Packaged resources to advance general chinese embedding (2023)

26. Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., Cao, Y.: React: Synergizing reasoning and acting in language models. arXiv preprint arXiv:2210.03629 (2022)

27. Zampogiannis, K., Yang, Y., Fermüller, C., Aloimonos, Y.: Learning the spatial semantics of manipulation actions through preposition grounding. 2015 IEEE International Conference on Robotics and Automation (ICRA) pp. 1389–1396 (2015)