Embodied AI Robot Companion for Efficient Object Handling in Bimanual Teleoperation

Haolin Fei*, Songlin Ma*, Bo Xiao†, Allahyar Montazeri*, Elmira Yadollahi‡, Hak-Keung Lam§ and Ziwei Wang*
*School of Engineering
Lancaster University, Lancaster, LA1 4YW, UK
Email: z.wang82@lancaster.ac.uk
†Department of Mechanical Engineering, Imperial College London, SW7 2AZ, UK
‡Department of Computing and Communications, Lancaster University, Lancaster, LA1 4YW, UK
§Department of Engineering, King's College London, WC2R 2LS, UK

Abstract—**Bimanual teleoperation often entails complex object-handling tasks that place substantial demands on human operators. Although existing research suggests various robot assistance methods, effectively utilizing multimodal interaction to align human intent understanding with context-aware reasoning remains challenging. In this paper, we introduce bimanual teleoperation with a Large-Language-Model (LLM) based assistant (BTLA) framework that leverages the strengths of both human operators and AI-powered robot companions. Our framework employs an LLM to drive the robot companion, enabling it to assist the human operator with higher accuracy and providing real-time feedback in response to voice commands. We evaluate the effectiveness of our LLM-aided robot assistant through extensive experiments in bimanual collaborative handling tasks. Experimental results show that our framework improves task performance and reduces user mental load by 196.4% over solo teleoperation and 320.8% over dyadic teleoperation. These results highlight the potential of our approach to enhance human-robot collaboration for more efficient and intuitive bimanual teleoperation systems.**

I. Introduction

Teleoperation has become an increasingly important technique to facilitate robotic system control in inaccessible or dangerous environments while ensuring human safety [24, 16]. It has benefited various applications such as space rendezvous and docking [33], underwater operations [32], and remote surgery [5]. To address these scenarios, dual-arm robotic teleoperation has emerged as an effective tool for performing complex tasks that require higher dexterity and more degrees of freedom [5]. Compared to single-arm systems, dual-arm teleoperation offers greater maneuverability, enhanced stability, and the capability to perform asymmetric tasks [16, 35].

A single human controlling two robots with both hands (i.e., single teleoperation), and two humans each controlling a robot with their respective dominant hand (i.e., dyad teleoperation) are the widely deployed interactive patterns for dual-arm teleoperation. In terms of solo teleoperation, human control performance is highly sensitive to hardware design ergonomics, cognitive load, and task complexity [12]. The operator needs to simultaneously manage the motion and coordination of two robotic arms, which can lead to increased mental workload and reduced performance [3]. Regarding dyad teleoperation, human-human communication, synchronization, and control mechanism design are still challenging in ensuring intuitive collaboration and avoiding arbitration conflict among humans [11, 18, 22]. To this end, assistance for human operators, such as robotic partners, in dual-arm teleoperation is beneficial for providing sensory feedback, motor control and decision-making assistance as needed. With this shared mechanism, operators can focus on performing partial tasks while the assistive agent manages the remaining [15, 34]. However, current assistance systems tend to focus on autonomous robot assistants, which may overlook the human's intention and the intuitiveness of the system [8, 17, 27]. Despite extensive research efforts in the robotic community, robots still struggle to match human performance in understanding and adapting to intricate real-world scenarios, particularly in terms of perceiving and responding to human intentions.

To address the aforementioned challenges of bimanual teleoperation, we propose an intuitive approach that integrates human teleoperation with an LLM as an assistance system controlling another robot. This approach leverages the strengths of both human operators and robotic systems, enabling each to focus on tasks where they excel. The LLM-based robot companion processes numerical data more efficiently and accurately than humans, while the human operator handles complex environmental information, such as visual cues and high-level decision-making. We set the scene in a bimanual handling task and conducted experiments with the robot skill set focused on collaborative handling tasks with limited autonomy at three different levels. The proposed system aims to balance the strengths of both human operators and robotic systems, measuring both objective performance metrics and subjective user experience ratings in tasks that are difficult for humans to perform through teleoperation but relatively easy for robots, covering coarse approaching, precise numerical computations, rapid processing of low-level state information, and execution of well-defined, repetitive actions.

The main contributions of our work are summarized as follows:

1) An embodied AI collaborated bimanual teleoperation approach, entailing seamless human-robot interaction

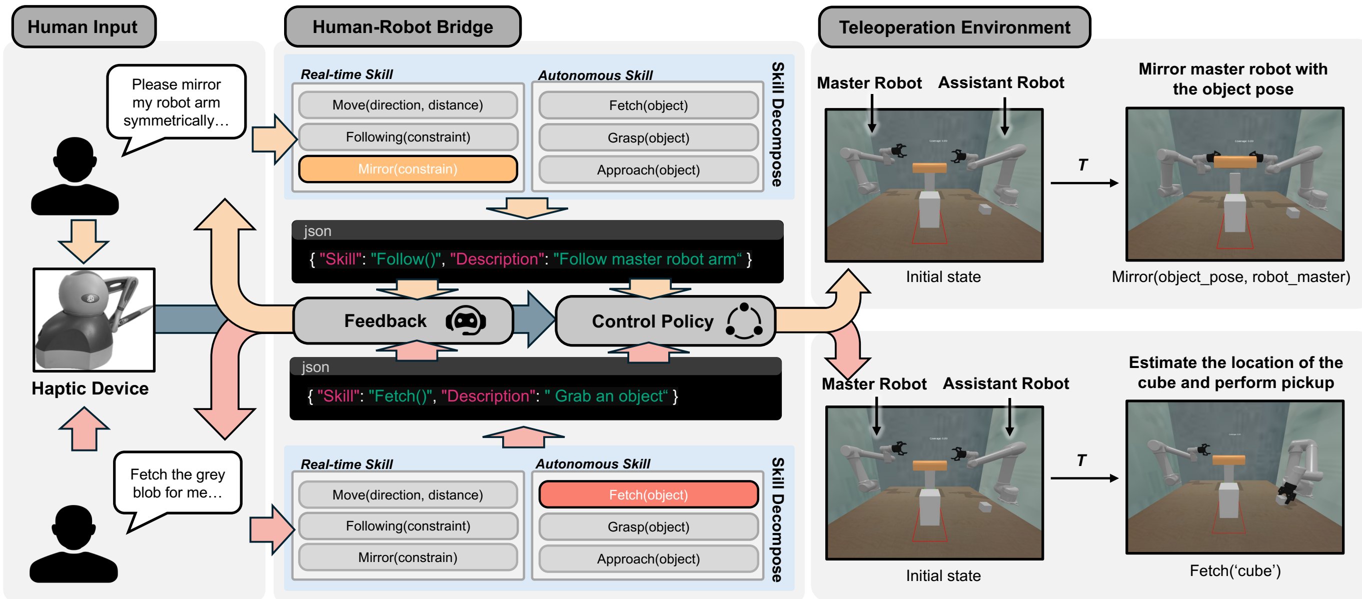

Fig. 1. Schematic diagram of the proposed BTLA method.

through the integration of LLM.

2) A comprehensive taxonomy of bimanual teleoperation tasks based on the level of autonomy and decision-making authority granted to the robotic assistant.
3) Experimental validation of the BTLA system in bimanual teleoperation scenarios, demonstrating improved efficiency, accuracy, and user satisfaction compared to other teleoperation approaches.

II. Related Works

The related work on the topics of bimanual teleoperation and LLM-based robotics has direct implications for our research and are presented.

Dual-arm teleoperation architecture can be broadly categorized into two main categories: single-person bimanual (SPB) teleoperation and dual-human, dual-arm (dyadic) teleoperation. In SPB teleoperation, one human operator controls both robotic arms simultaneously. This control paradigm often leads to a high mental workload for the operator, as they must manage the coordination and motion of two robotic arms in real time [28]. Two routes are widely adopted to overcome the above obstacle from the perspective of humans and robots: (i) developing more intuitive control interfaces, and (ii) designing control assistance algorithms. Intuitive human-machine interfaces aim to provide operators with natural sensations and user-friendly means of controlling dual-arm robots. Various interface technologies have been proposed, such as gesture-based interfaces [5], virtual reality-based interfaces [10], and haptic devices [26], reducing the cognitive burden associated with traditional control methods. Additionally, haptic feedback algorithms [30, 6, 37] have been proposed to provide force feedback to the operator, enhancing their situational awareness and control precision. Control assistance algorithms, on the other hand, focus on developing intelligent strategies to assist the operator in managing the dual-arm system, including mapping strategies that translate human input into efficient and coordinated robot motions. Shared control approaches [16, 20, 31, 18] have been introduced to combine human input with autonomous robot behaviors, assisting the operator in dual-arm manipulation tasks.

Dyadic teleoperation involves two human operators collaboratively controlling two robotic arms, leveraging the expertise and cognitive capabilities of multiple operators to tackle complex tasks. In[25], a model for interaction force computation in dyadic cooperative object manipulation tasks was proposed. In [7], the performance of human-human dyads and individuals in a teleoperation environment was compared, revealing that collaboration does not always lead to improved performance, especially in teleoperation tasks. Furthermore, in [19], researchers investigated the role allocation strategy for a leader–follower relationship in human dyads during collaborative tasks, providing insights into the dynamics of human-human collaboration in teleoperation scenarios.

LLM-based methods have shown promising results in enhancing interactive capabilities of robotic systems [36, 9, 29]. These methods leverage the strong understanding of the real world inherent in LLMs to perform high-level planning using image cues. The planned tasks are then executed by calling upon lower-level knowledge bases for automation [14, 36, 21], allowing for more flexibility and adaptation to handle various tasks and environments. However, these LLM-based methods may not be ideal for multi-contact teleoperation and physical interaction. Object grasping and manipulations in complicated or dynamic environments may be more suitable for human operators due to their intuitive understanding of the task and the ability to adapt quickly to minor variations [1, 4]. In such situations, the overhead of using an LLM for planning and automation may not justify the potential benefits. Instead of tasking the LLM with context understanding and decision-making, our approach leverages the human operator's expertise in these areas. We utilize the LLM as a human-robot interface, concentrating on its core strength of natural language processing to effectively convey human intentions. This enhances the responsiveness and intuitiveness of the entire system, while the human operator retains control over the planning and execution of the task.

III. Methodology

We first provide the formulation of the bimanual teleoperation problem in Section III-A. Subsequently, we present how BTLA utilizes LLM to assist humans in bimanual teleoperation tasks.

A. Problem Formulation

The main objective of the proposed method is to enable the assistant robot to generate human-desired behaviors that assist in performing tasks efficiently in dual-arm teleoperation scenarios, i.e., a human operator controls one robot arm (master robot) and an embodied AI system controls the other robot arm (assistant robot). The assistant robot receives natural language voice instructions L (e.g., *"help me push the green blob together"*) that specify the desired assistive behavior. These instructions can be long-horizon, context-aware, or ambiguously described (e.g., *"move a little bit upwards"*), requiring sophisticated contextual understanding. At any given time t, the assistant robot has access to the proprioceptive information from both the master robot ($\mathbf{s}_{master,t}$) and itself ($\mathbf{s}_{assistant,t}$). Additionally, the assistant robot can obtain environmental observations ($\mathbf{o}_{env,t}$) via its available sensors. Therefore, the problem formulation can be summarized as follows: given a natural language instruction l, the assistant robot's proprioceptive information $\mathbf{s}_{assistant,t}$, the master robot's proprioceptive information $\mathbf{s}_{master,t}$, human input $\mathbf{u}_t$, environment sensing information $\mathbf{z}_t$ at time t and environmental observations $\mathbf{o}_{env,t}$, the embodied AI system should generate a sequence of low-level skills from the skill base $\mathcal{S}$ and map them to a control policy π that enables the assistant robot to assist the human operator in performing the desired task effectively.

To this end, the assistant robot must decompose the high-level instruction l into a sequence of low-level skills selected from a predefined skill base $\mathcal{S}$. The chosen skills and their corresponding parameters are then mapped to a control policy π, represented by a skill function $BTLA(\cdot)$. The skill knowledge in the skill base $\mathcal{S}$ can be adapted to accommodate different task requirements. Therefore, the focus of our work is not on the acquisition of these skills but rather on the effective utilization of the available skills to assist the human operator.

B. Bimanual Teleoperation LLM Assistant

As shown in Fig. 1, the BTLA can be divided into three main components: the human operator, the human-robot interface, and the teleoperation environment. The human operator can concentrate on the current task by observing the environment via visual feedback, manipulating one robot arm with teleoperation devices, and soliciting support from the AI-assisted robot arm for collaborative task execution. We leverage LLM's natural language processing capabilities to enhance HRI, which differs significantly from button-based methods, as it allows for more nuanced and context-aware communication between the human operator and the robot assistant. Unlike a simple skill switcher, the LLM can interpret complex instructions, understand context, and provide feedback when needed. This flexibility enables the robot assistant to adapt to a wider range of scenarios and user needs.

Algorithm 1 Embodied AI-Assisted Robot Arm Control

```
Require: Initial skills base S with predefined skills, LLM
    initial language description l
Initialize t ← 0, skill ← None
while not finished do
    if voice_command received then
        skill ← LLM (voice_command)
        π ← BTLA (skill, skill_parameters, u_t, z_t)
        if skill is real-time then
            repeat
                Execute π
                t ← t + 1
            until voice_command to stop
        else if skill is autonomous then
            repeat
                Execute π
                t ← t + 1
            until skill is done
        end if
    end if
end while
```

The AI-assisted robot arm receives human language commands as input and identifies the most relevant skill from its skill database $\mathcal{S}$, along with the necessary task parameters. The selected skill, combined with environmental data from sensors (such as visual information), proprioceptive data, and human input, forms the control policy that guides the actions of the AI-assisted robot arm. Within this configuration, the human operator collaborates with the AI-assisted robot arm within the teleoperation environment to achieve the desired task with optimal efficiency and effectiveness. The human operator provides high-level guidance and control, while the AI-assisted robot arm contributes its capabilities and understanding of the context to support the human operator in achieving their objectives. The process flow of the embodied AI-assisted robot arm control is outlined in Algorithm 1, which describes how the system receives voice commands, interprets them using the LLM, and executes the corresponding skills.

The skill base is tailored to the bimanual handling tasks. Here, the emphasis lies less on autonomy and more on bolstering support for the human operator. In collaborative

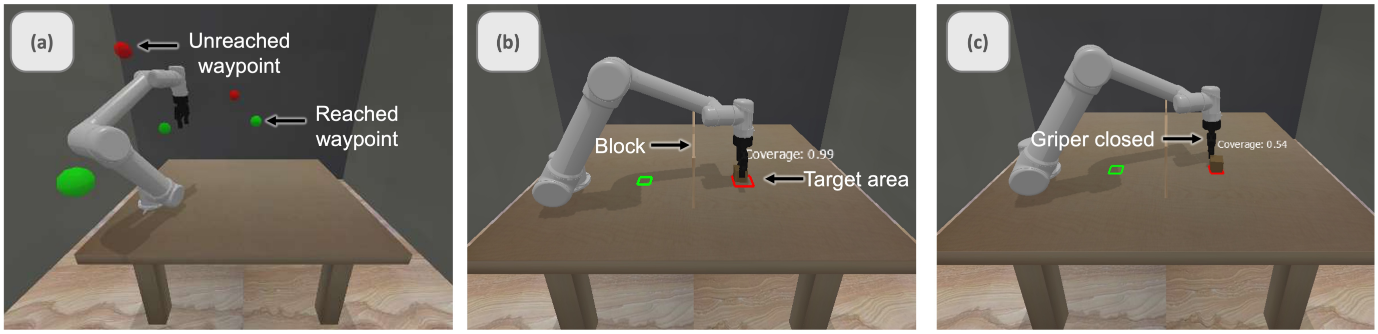

Fig. 2. Single arm training tasks: (a) target reaching, (b) pick-and-place, and (c) pushing.

tasks, decision-making remains the responsibility of the human operator, but the robot assumes a greater degree of control compared to human-dominated tasks. The skill base for collaborative tasks should prioritize working with humans and performing tasks simultaneously. The specific skills required by the robot are still determined by the human operator. Collaborative tasks typically have high spatial-temporal coherence, which can be better executed by the robot than by a human. These tasks may also involve smooth trajectory requirements, low autonomy needs, or easily planned actions. One example of a skill required in many industrial scenarios involving bimanual robot arms is symmetrical following [23]. This skill allows one robot to perform a task on one side of a structure (such as welding or holding inspection transmitters) while the other robot follows on the opposite side to carry out complementary tasks (such as providing additional views from inside or operating inspection receivers). Symmetrical following requires both robots to move with high consistency in terms of timing and location, which is less efficient and accurate when performed through SPB teleoperation or dyadic teleoperation.

Building upon the skill base and task categorization, our framework incorporates a mechanism to handle situations where the robot misunderstands a command or encounters singularities. To ensure safe and effective operation, the system implements a confirmation process before executing any task. When the robot receives a command, it first interprets the instruction and generates a plan for execution. Before proceeding, it communicates this plan back to the human operator for confirmation. This step allows the operator to verify that the robot has correctly understood the command and provides an opportunity to make corrections if needed. In cases where the robot encounters singularities or potential issues during task execution, it immediately halts the operation and seeks guidance from the human operator. This interactive loop between the human operator and the robot assistant ensures a robust and adaptable system that can recover from misunderstandings and navigate complex scenarios effectively, maintaining the balance between autonomous operation and human oversight.

IV. Experiment

The experiments aim to investigate the effectiveness of the proposed LLM-aided robot assistant framework in teleoperated bimanual handling tasks. We designed the experiments to answer the following key questions:

1) Does the proposed framework improve task performance compared to solo and dyadic teleoperation?
2) Does the proposed framework reduce the human operator's physical and mental workload in teleoperated bimanual handling tasks?

To address these questions, we conducted a series of experiments involving the manipulation and transportation of large, heavy objects using a bimanual robotic system. The experimental procedure, from operator training to performance assessment, is illustrated in Fig. 4.

A. Experiment Setup

1) Equipment and Software: Our experiments utilize two 3D systems Touch (previously Phantom Omni) haptic devices for interaction. The Pybullet API is employed for scene creation, robot arm control, and object visualization. To minimize uncontrolled variables that might influence the experiment results, we designed customized objects using Fusion 360 and converted them into URDF files. For realistic human voice interactions, we adopted the OpenAI Whisper model for speech-to-text and text-to-speech (TTS) tasks. We evaluated three language models: GPT-3.5-turbo, GPT-4, and Mistral-7B-OpenOrca (a locally running model using GPT4All [2]). Our tests indicated no significant differences in performance among these models. Consequently, we selected GPT-3.5-turbo for our experiments due to its lower cost and faster response time.

2) LLM Initial Prompt: To optimize the robot assistant's understanding of its role and objectives, we implement a set of predefined rules and instructions as an initial prompt for the LLM. This approach eliminates the need for users to communicate these demands before each interaction, enhancing operational efficiency. The initial prompt configures the LLM as an AI assistant designed to aid a robot arm in task execution. It instructs the LLM to generate scripts based on user's spoken commands, adhering to a specific JSON format:

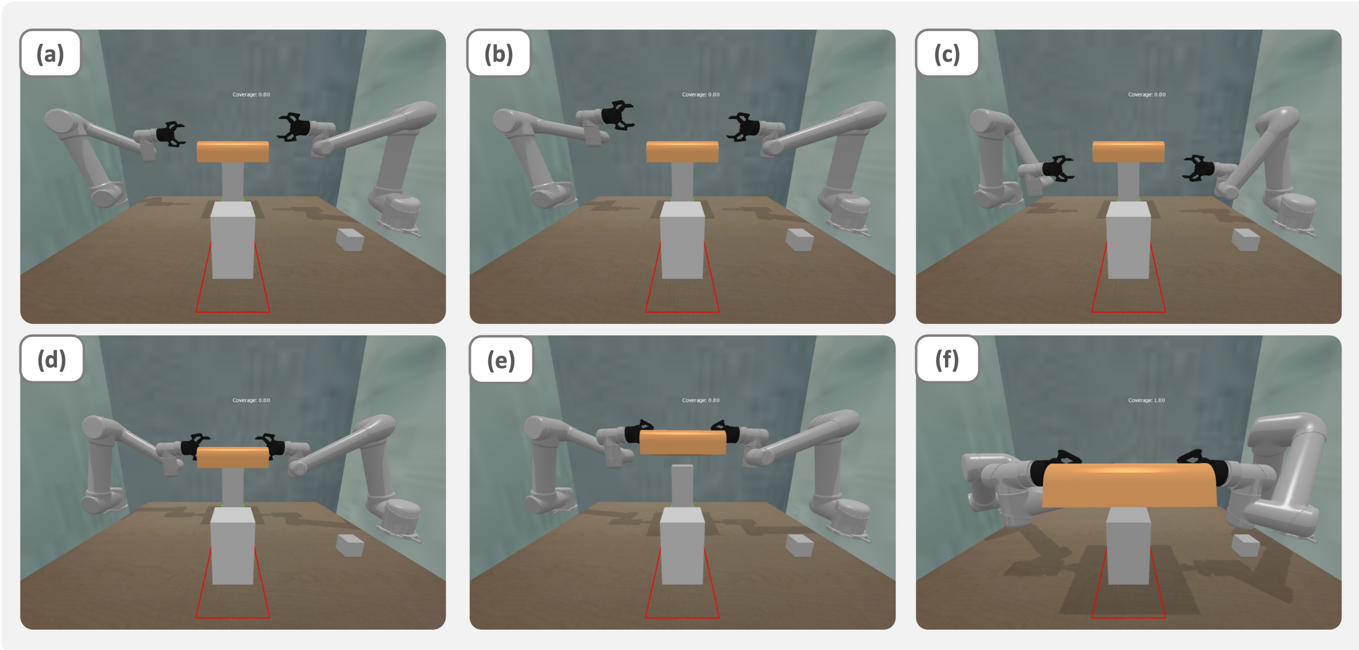

Fig. 3. Illustration of BTLA on object handling task: (a) Initial state with both robot arms at rest. (b) Left arm moves independently without following commands. (c) Right arm, controlled by the robot companion, exhibits symmetrical following behavior. (d) Both arms simultaneously move to the object pickup position. (e) The robot grippers grasp the object securely. (f) The human operator and the AI-powered robot companion collaborate to transfer the object to the designated location.

Script: "Skill": "Write the function here.", "Description": "Include a necessary description about this skill, as if you are talking to the user directly. Use 'you' to address the user." . The robot assistant is equipped with a comprehensive list of available skills from the skill database, enabling efficient matching of user commands with appropriate functions. The LLM is programmed to provide user feedback on its actions through the "Description" field in the JSON script. When a user's command corresponds to a known skill, the LLM generates the relevant script. In cases where no match is found, the assistant generates a script with an empty function and a description indicating that no action will be taken. This structured approach to the initial prompt ensures the LLM-aided robot assistant's ability to interpret user command and provide meaningful feedback which facilitates a more seamless and effective interaction between the human operator and the embodied AI system in bimanual handling tasks.

3) Skills: There are two types of skills, autonomous skills and real-time skills. Autonomous skills are executing actions in a series and exiting when the whole action is done, such as Handover() - Handover an object to the master arm; Approach() - Move the arm to approach an object (e.g. for listing objects together); Fetch() - Grab an object and bring it to the master arm. However, the Real-time Skills are continuous motions and exiting when the user gives the stop command, like Follow() - Follow the master robot arm (e.g. for pushing together); SymmetricalFollow() - Act a mirror behavior of the master robot arm; Move(distance, direction) - Move the arm (ask user for distance in meters and direction: "+x", "-x", "+y", "-y", "+z", "-z").

B. Participant Training

Training is essential to minimize differences in manipulation abilities among participants and to familiarize them with the haptic device and simulation platform. Prior to the experiments, each participant underwent four training sessions (one-hour intervals between sessions) for three tasks: target reaching, pick-and-place, and pushing as shown in Fig. 2. These tasks were progressively ordered from easy to difficult. In the target reaching task, the goal was to navigate to the red waypoints. The pick-and-place task required participants to use the gripper to grasp a square block and transport it to a target area while avoiding a vertical barrier. The pushing task involved pushing an object into a designated target area. Participants were required to complete the tasks within 4 minutes, and 3 minutes, respectively.

C. Experiments Implementation

The task is to grasp the object using two arms cooperatively and move the object to the appointed platform shown in Fig. 3. Fig. 3 (a) to Fig. 3 (d) shows the motion from the start position to the grasp position. Fig. 3 (e) to Fig. 3 (f) shows the motion to the appointed platform. There are three dimensions to assess the performance: the coverage rate, the average time cost and the number of repeated successes. The number of repeated successes will be recorded if the two used arms can: 1, lift the object simultaneously; 2, move gradually without falling

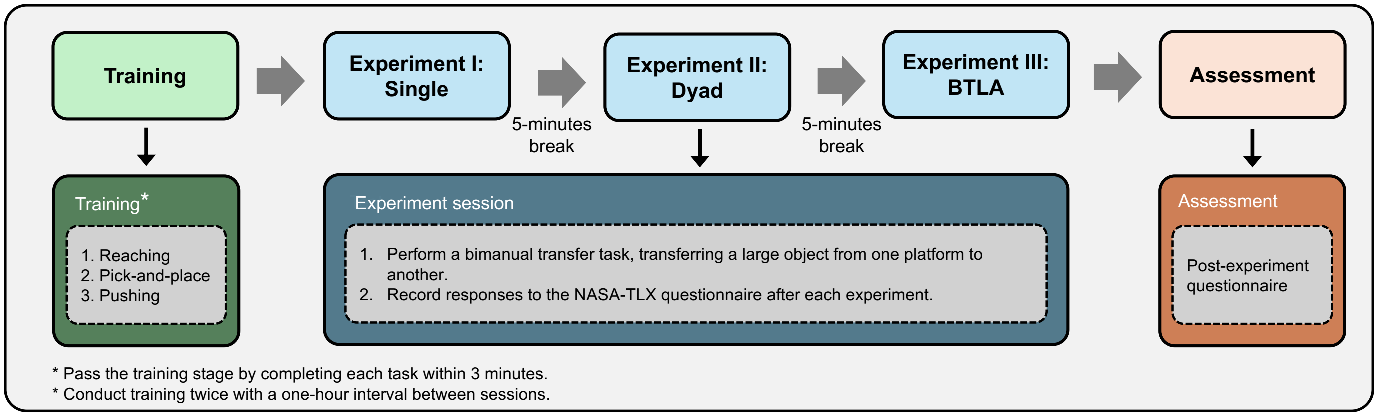

Fig. 4. Experiment flowchart.

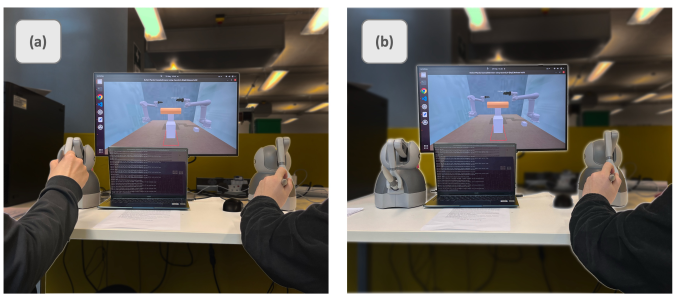

Fig. 5. Views of (a) dyadic teleoperation. (b) robot assisted teleoperation.

the object, 3, put the object steady on the specific platform and cover above 70% of the goal area. Three types of teleoperation are tested.

1) SPB Teleoperation: SPB teleoperation is executed as a comparison experiment to answer the first question. Each participant controls two robot arms using haptic devices to grab the object and move it to the goal area. After finishing the trial, record responses to the NASA-TLX questionnaire.

2) Dyadic Teleoperation: For the second question, dyadic teleoperation as another comparison experiment is implemented. The participant controls one robot arm and collaborates with a constant operator who controls another robot arm to achieve the same task. After finishing the trial, record responses to the NASA-TLX questionnaire.

3) BTLA Teleoperation: For our opposed method, participants control one robot arm and collaborate with an intelligent agent i.e., a robot assistant to finish the whole task. participants can tell the agent what they want it to do. Participants are known the skills that the agent has, but they don't know the special keywords/function names that installed in the system. Participants should use their judgements and natural language to ask the agent how it helps them to grab the object and move it to the goal area. After finishing the trial, the experience is recorded on the NASA-TLX questionnaire.

D. Assessment

The third question can be answered by using questionnaires. Then analyse the questionnaires that they filled out. To assess participants' perceptions of their experience using the Bimanual Teleoperation with LLM-based Assistance (BTLA) system under various assistance conditions, we administered two questionnaires. The first questionnaire, shown in Fig. 8, included nine scales measuring naturalness, satisfaction, perceived robot intelligence, and likeness. The second questionnaire was the NASA Task Load Index (NASA-TLX), which was completed by participants after finishing the assigned tasks, as shown in Fig. 7. The NASA-TLX is a widely used tool for evaluating the subjective workload experienced by users during the performance of a task [13].

V. Results and Discussions

A. Performance Metrics

To evaluate the effectiveness of the BTLA, we compared its performance with the Dyadic and SPB scenarios using three metrics: coverage, success rate, and task completion time, as shown in Fig. 6. The BTLA scenario demonstrated the highest mean coverage (0.861) and success rate (0.627) among the three scenarios, suggesting that the BTLA system is more effective in successfully completing tasks and covering a larger portion of the task space compared to the Dyadic and SPB scenarios. The Kruskal-Wallis test was performed to assess the statistical significance of the differences in coverage ($p = 0.003$) and success rate ($p = 0.004$) among the scenarios, and the results indicate the differences in these metrics among the scenarios are statistically significant. Although the BTLA scenario exhibited faster task completion times compared to the other scenarios, the differences were not statistically significant based on the Kruskal-Wallis test, which yielded a p-value of 0.117 for the time metric.

Furthermore, a correlation analysis was conducted to examine the relationship between coverage and success rate. The analysis revealed a strong positive correlation (0.708) between the two metrics, indicating that higher coverage is associated

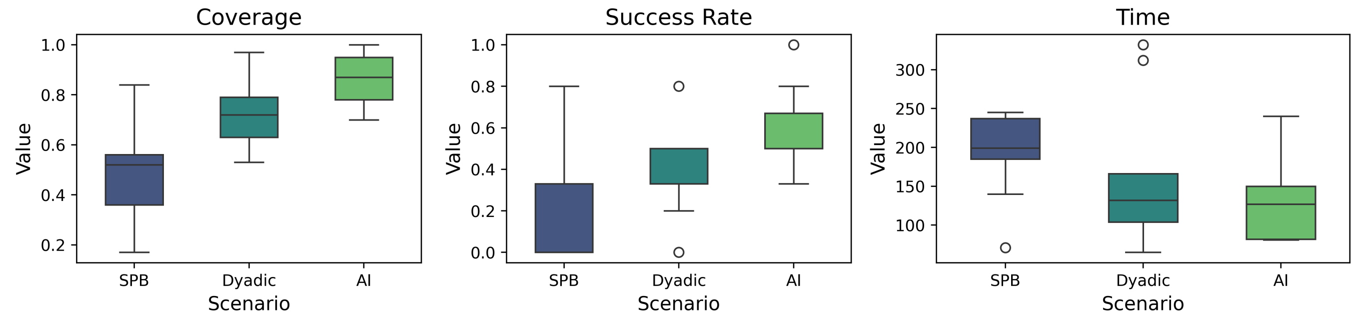

Fig. 6. Box plots for performance (a) coverage ($p < 0.05$) (b) success rate ($p < 0.05$) (c) time ($p = 0.117$) among all subjects for three experimental scenarios SPB, dyadic, and BTLA.

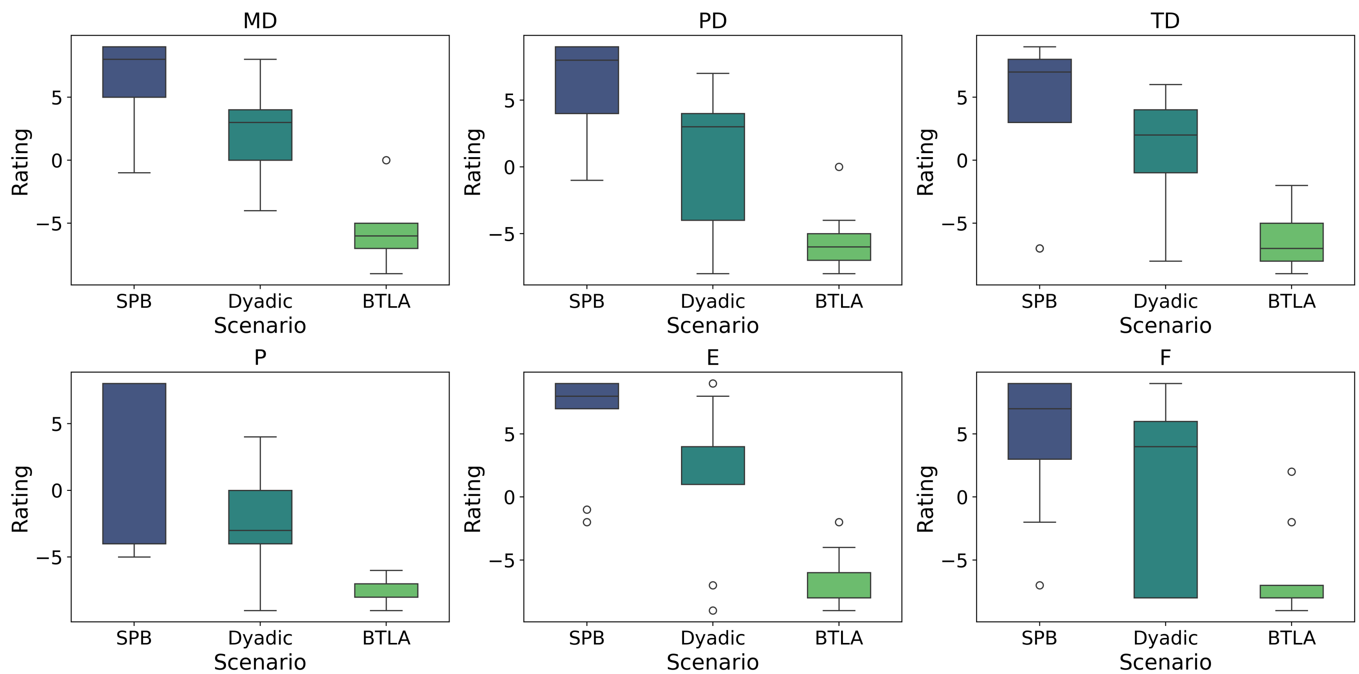

Fig. 7. Boxplots for NASA-TLX results among all subjects for three experimental scenarios SPB, dyadic, and BTLA, respectively. Rated aspects from NASA-TLX: mental demand (MD), physical demand (PD), temporal demand (TD), performance (P), effort (E), and frustration (F). (all metrics: $p < 0.05$).

Statement		Score									
How do you feel about the performance of the intelligent agent?	Unnatural	1	2	3	4	5	6	7	8	9	Natural
	Unpleasant	1	2	3	4	5	6	7	8	9	Pleasant
	Strange	1	2	3	4	5	6	7	8	9	Typical
	Dislikeable	1	2	3	4	5	6	7	8	9	Likeable

Fig. 8. Part of post experiment questionnaire examples.

with higher success rates. This finding suggests that the BTLA system's ability to cover a larger portion of the task space contributes to its higher success rates in completing tasks compared to the Dyadic and SPB scenarios.

TABLE I
DESCRIPTIVE STATISTICS FOR PERFORMANCE METRICS.

Scenario	Coverage Rate	Success Rate	Time (s)
SPB	0.489(±0.22)	0.184(±0.27)	190(±56)
Dyadic	0.721(±0.18)	0.369(±0.22)	163(±95)
BTLA	0.861(±0.11)	0.627(±0.20)	130(±55)

B. Subjective Assessment

For all NASA-TLX metrics (mental demand (MD), physical demand (PD), temporal demand (TD), performance (P), effort

(E), and frustration (F)), the BTLA scenario exhibited the most favorable ratings, with lower demands, effort, and frustration, as well as better perceived performance compared to the Dyadic and SPB scenarios as shown in Fig. 7. In contrast, the SPB scenario appeared to be the most challenging, with higher demands, effort, and frustration, and lower perceived performance. The Dyadic scenario fell between the BTLA and SPB scenarios, indicating moderate levels of demands, effort, frustration, and performance.

The Kruskal-Wallis test results revealed statistically significant differences among the three scenarios for all metrics, with the test statistics being 17.974 for MD ($p = 0.000$), 14.701 for PD ($p = 0.001$), 12.276 for TD ($p = 0.0002$), 15.723 for P ($p = 0.000$), 14.228 for E ($p = 0.0001$), and 11.018 for F ($p = 0.000$). The p-values for all metrics were less than 0.05, providing strong evidence against the null hypothesis of no difference among the scenarios.

We have also noticed that over 40% of participants reported that their performance was limited by the restricted 2D camera view. This limitation was due to either a loss of depth perception, making it difficult to discern spatial relationships, or because the images were partially obstructed.

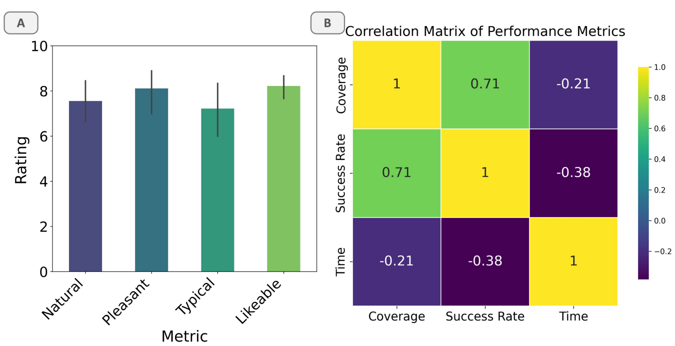

Fig. 9. (**A**) Likert Scale Ratings. (**B**) Correlation matrix of performance metrics.

VI. Conclusion

This paper presents a novel LLM-aided robot assistant framework for collaborative bimanual teleoperation that integrates a human operator with an embodied AI robot companion to enhance human-robot collaboration and reduce the operator's cognitive load. Experimental results demonstrate that the proposed framework outperforms solo and dyadic teleoperation in terms of coverage rate, success rate, and subjective workload assessment, highlighting its potential for improving task performance and user experience. The strong correlation between coverage rate and success rate suggests that the LLM-aided robot assistant's ability to enable the human to focus on more precise performance contributes to its higher success rates. These findings underscore the importance of leveraging embodied AI systems to create more efficient, intuitive, and adaptable teleoperation systems for bimanual handling tasks.

Acknowledgments

References

[1] Iretiayo Akinola, Jingxi Xu, Shuran Song, and Peter K Allen. Dynamic grasping with reachability and motion awareness. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9422–9429. IEEE, 2021.

[2] Yuvanesh Anand, Zach Nussbaum, Brandon Duderstadt, Benjamin Schmidt, and Andriy Mulyar. Gpt4all: Training an assistant-style chatbot with large scale data distillation from gpt-3.5-turbo. https://github.com/nomic-ai/gpt4all, 2023.

[3] Weibang Bai, Ziwei Wang, Qixin Cao, Hiroshi Yokoi, Masakatsu G. Fujie, Eric M. Yeatman, and Guang-Zhong Yang. Anthropomorphic dual-arm coordinated control for a single-port surgical robot based on dual-step optimization. *IEEE Transactions on Medical Robotics and Bionics*, 4(1):72–84, 2022.

[4] Ravi Balasubramanian, Ling Xu, Peter D Brook, Joshua R Smith, and Yoky Matsuoka. Human-guided grasp measures improve grasp robustness on physical robot. In *2010 IEEE International Conference on Robotics and Automation*, pages 2294–2301. IEEE, 2010.

[5] Jacob R Boehm, Nicholas P Fey, and Ann Majewicz Fey. Online recognition of bimanual coordination provides important context for movement data in bimanual teleoperated robots. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6248–6255. IEEE, 2021.

[6] Murat Cenk Çavusoglu, Alana Sherman, and Frank Tendick. Design of bilateral teleoperation controllers for haptic exploration and telemanipulation of soft environments. *IEEE Transactions on Robotics and Automation*, 18(4):641–647, 2002.

[7] Yuhang Che, Gabriel M Haro, and Allison M Okamura. Two is not always better than one: Effects of teleoperation and haptic coupling. In *2016 6th IEEE international conference on biomedical robotics and biomechatronics (BioRob)*, pages 1290–1295. IEEE, 2016.

[8] Janelle P. Clark, Gianluca Lentini, Federica Barontini, Manuel G. Catalano, Matteo Bianchi, and Marcia K. O'Malley. On the role of wearable haptics for force feedback in teleimpedance control for dual-arm robotic teleoperation. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 5187–5193, 2019. doi: 10.1109/ICRA.2019.8793652.

[9] Te Cui, Guangyan Chen, Tianxing Zhou, Zicai Peng, Mengxiao Hu, Haoyang Lu, Haizhou Li, Meiling Wang, Yi Yang, and Yufeng Yue. Human demonstrations are generalizable knowledge for robots. *arXiv preprint arXiv:2312.02419*, 2024.

[10] Alberto García, J Ernesto Solanes, Adolfo Muñoz, Luis Gracia, and Josep Tornero. Augmented reality-based interface for bimanual robot teleoperation. *Applied Sciences*, 12(9):4379, 2022.

[11] Ganesh Gowrishankar, Atsushi Takagi, Rieko Osu, Toshinori Yoshioka, Mitsuo Kawato, and Etienne Burdet. Two is better than one: Physical interactions improve motor performance in humans. *Scientific Reports*, 4(1):1–7, 2014.

[12] Yao Guo, Daniel Freer, Fani Deligianni, and Guang-Zhong Yang. Eye-tracking for performance evaluation and workload estimation in space telerobotic training. *IEEE Transactions on Human-Machine Systems*, 52(1): 1–11, 2022.

[13] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[14] Yingdong Hu, Fanqi Lin, Tong Zhang, Li Yi, and Yang Gao. Look before you leap: Unveiling the power of gpt-4v in robotic vision-language planning. *arXiv preprint arXiv:2311.17842*, 2023.

[15] Zhaoyang Jacopo Hu, Ziwei Wang, Yanpei Huang, Aran Sena, Ferdinando Rodriguez y Baena, and Etienne Burdet. Towards human-robot collaborative surgery: Trajectory and strategy learning in bimanual peg transfer. *IEEE Robotics and Automation Letters*, 8(8):4553–4560, 2023.

[16] Darong Huang, Bin Li, Yanan Li, and Chenguang Yang. Cooperative manipulation of deformable objects by single-leader–dual-follower teleoperation. *IEEE Transactions on Industrial Electronics*, 69(12):13162–13170, 2022. doi: 10.1109/TIE.2021.3139228.

[17] Yanpei Huang, Wenjie Lai, Lin Cao, Jiajun Liu, Xiaoguo Li, Etienne Burdet, and Soo Jay Phee. A three-limb teleoperated robotic system with foot control for flexible endoscopic surgery. *Annals of Biomedical Engineering*, 49:2282–2296, 2021.

[18] Zebin Huang, Ziwei Wang, Weibang Bai, Yanpei Huang, Lichao Sun, Bo Xiao, and Eric M Yeatman. A novel training and collaboration integrated framework for human–agent teleoperation. *Sensors*, 21(24):8341, 2021.

[19] Rebeka Kropivšek Leskovar, Jernej Čamernik, and Tadej Petrič. Leader–follower role allocation for physical collaboration in human dyads. *Applied Sciences*, 11(19): 8928, 2021.

[20] Marco Laghi, Michele Maimeri, Mathieu Marchand, Clara Leparoux, Manuel Catalano, Arash Ajoudani, and Antonio Bicchi. Shared-autonomy control for intuitive bimanual tele-manipulation. In *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, pages 1–9. IEEE, 2018.

[21] Shoujie Li, Ran Yu, Tong Wu, JunWen Zhong, Xiao-Ping Zhang, and Wenbo Ding. Growing from exploration: A self-exploring framework for robots based on foundation models. *arXiv preprint arXiv:2401.13462*, 2024.

[22] Yanan Li, Aran Sena, Ziwei Wang, Xueyan Xing, Jan Babič, Edwin van Asseldonk, and Etienne Burdet. A review on interaction control for contact robots through intent detection. *Progress in Biomedical Engineering*, 4 (3):032004, 2022.

[23] Eloise Matheson, Riccardo Minto, Emanuele GG Zampieri, Maurizio Faccio, and Giulio Rosati. Human–robot collaboration in manufacturing applications: A review. *Robotics*, 8(4):100, 2019.

[24] MD Moniruzzaman, Alexander Rassau, Douglas Chai, and Syed Mohammed Shamsul Islam. Teleoperation methods and enhancement techniques for mobile robots: A comprehensive survey. *Robotics and Autonomous Systems*, 150:103973, 2022. ISSN 0921-8890. doi: https://doi.org/10.1016/j.robot.2021.103973.

[25] Ehsan Noohi, Miloš Žefran, and James L Patton. A model for human–human collaborative object manipulation and its application to human–robot interaction. *IEEE transactions on robotics*, 32(4):880–896, 2016.

[26] Daniel Rakita, Bilge Mutlu, Michael Gleicher, and Laura M Hiatt. Shared control–based bimanual robot manipulation. *Science Robotics*, 4(30):eaaw0955, 2019.

[27] Aran Sena, Quentin Rouxel, Ekaterina Ivanova, Etienne Burdet, and Zhibin Li. Haptic bimanual system for teleoperation of time-delayed tasks. In *2021 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 1234–1239, 2021.

[28] Shiliang Shao, Ting Wang, Yongliang Wang, Yun Su, Chunhe Song, and Chen Yao. Research of hrv as a measure of mental workload in human and dual-arm robot interaction. *Electronics*, 9(12):2174, 2020.

[29] Ishika Singh, Valts Blukis, Arsalan Mousavian, Ankit Goyal, Danfei Xu, Jonathan Tremblay, Dieter Fox, Jesse Thomason, and Animesh Garg. Progprompt: Generating situated robot task plans using large language models. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11523–11530, 2023. doi: 10.1109/ICRA48891.2023.10161317.

[30] Servet Soyguder and Tayfun Abut. Haptic industrial robot control with variable time delayed bilateral teleoperation. *Industrial Robot: An International Journal*, 43(4):390–402, 2016.

[31] Da Sun, Qianfang Liao, and Amy Loutfi. Single master bimanual teleoperation system with efficient regulation. *IEEE Transactions on Robotics*, 36(4):1022–1037, 2020.

[32] Yanchao Sun, Xinyan Chen, Ziwei Wang, Hongde Qin, and Ruijie Jing. Adaptive interval type-2 fuzzy control for multi-legged underwater robot with input saturation and full-state constraints. *International Journal of Systems Science*, 54(15):2859–2874, 2023.

[33] Ziwei Wang, Hak-Keung Lam, Bo Xiao, Zhang Chen, Bin Liang, and Tao Zhang. Event-triggered prescribed-time fuzzy control for space teleoperation systems subject to multiple constraints and uncertainties. *IEEE Transactions on Fuzzy Systems*, 29(9):2785–2797, 2021.

[34] Ziwei Wang, Haolin Fei, Yanpei Huang, Quentin Rouxel, Bo Xiao, Zhibin Li, and Etienne Burdet. Learning to assist bimanual teleoperation using interval type-2 polynomial fuzzy inference. *IEEE Transactions on Cognitive and Developmental Systems*, 16(2):416–425, 2024.

[35] Qihan Wu, Meng Li, Xiaozhi Qi, Ying Hu, Bing Li,

and Jianwei Zhang. Coordinated control of a dual-arm robot for surgical instrument sorting tasks. *Robotics and Autonomous Systems*, 112:1–12, 2019. ISSN 0921-8890.

[36] Lihan Zha, Yuchen Cui, Li-Heng Lin, Minae Kwon, Montserrat Gonzalez Arenas, Andy Zeng, Fei Xia, and Dorsa Sadigh. Distilling and retrieving generalizable knowledge for robot manipulation via language corrections. *arXiv preprint arXiv:2311.10678*, 2023.

[37] Cheng Zhou, Longfei Zhao, Haitao Wang, Lipeng Chen, and Yu Zheng. A bilateral dual-arm teleoperation robot system with a unified control architecture. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pages 495–502. IEEE, 2021.