2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Self-Assessment of Robotic Laboratory and Equipment Readiness Using Large Language Models and Robotic Data Capture

Stefan Ilić and Josie Hughes

Abstract— **This study explores the potential of automating robotic laboratory readiness assessment by integrating Large Language Models (LLMs) with robotic data acquisition. It investigates the capability of LLMs to detect equipment motion and operational status using visual and auditory information. Despite the challenges LLMs face in spatial analysis, this study also investigates LLM grounding methods to ensure accurate workspace assessment. By inspecting a robotic cooking setup with camera-equipped robotic arm, LLMs can detect the motion of custom equipment via color-coded marks, and identify the operational status of kitchen appliances from a single image without any physical augmentations. Additionally, device operation perceived through the emission of loud noises can be assessed by post-processing sound recordings and analyzing loudness and sound frequency metrics presented in a visual plot form. For simple spatial tasks like saucepan positioning, LLM provides accurate assessments when grounded with a single image, while complex workspace safety assessment task requires extensive knowledge of past experiences. By reviewing status of each checklist item, the LLM can decide whether experiment needs to be halted or requires human intervention, offering a set of troubleshooting steps. These findings demonstrate feasibility of the self-assessment approach for robotic laboratory systems, paving the way for future deployments.**

I. INTRODUCTION

Robots are increasingly being deployed for lab science automation, with notables successes seen in chemistry [1], materials science [2] and fluid dynamics [3]. These complex systems integrate heterogeneous robots with diverse instruments and equipment, making the system sensitive and prone to failures and user setup errors. As such, the evaluation of equipment and workspace environment before and during operation is fundamental for intelligent robotic automation. It not only safeguards the personnel and equipment [4], but also guarantees that experiment will be performed under anticipated conditions. Traditionally, visual inspection is performed by a human before starting an experiment, manually checking against a detailed checklist. While comprehensive, this method is time-consuming and prone to errors due to human fatigue, potentially overlooking critical issues. These oversights can compromise safety, damage equipment, or yield inaccurate data, leading to false conclusions and waste of time and resources. As robots are deployed into lab and scientific environments, it is important that they can perform some 'self checks' and inspection protocols, to enhance efficiency and reduce risks, saving time and resources [5].

Robotic visual inspection is a growing field in both academia and industry, leading to widespread adoption of robots in field operations [6], [7]. However, currently, much of the research and industrial applications target environments with large footprint or inherent danger, such as in-

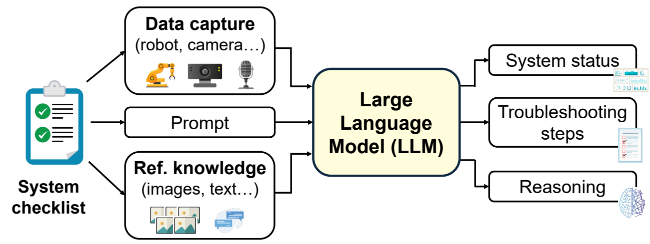

Fig. 1. Self-assessment process of the robotic laboratory system

dustrial plants or outdoor structures [8]. These applications typically involve legged robots or drones navigating predetermined paths to capture images of specific equipment or assess the environment more broadly, using human analysis or pre-trained vision models for data interpretation [9], [10]. The inspection effectiveness can be further enhanced by collecting sensor data from industrial systems, offering in-depth insights and enabling quick detection and correction of hazardous conditions [11]. By integrating external and internal inspection methods and executing them by robots, the system can conduct a thorough and robust self-assessment. Despite advancements in large-scale environments, the potential for robotic inspection in compact benchtop settings, such as those found in chemistry or food laboratories remains underexplored due to limited integration of robotic automation. Furthermore, the diversity of lab experimental setups requires generalizable and adaptable image data analysis that exceeds the capabilities of pre-trained vision models, necessitating an approach that can mimic human reasoning. This strategy would enable robotic visual inspection systems to evaluate unseen conditions and offer the flexibility necessary to accommodate any changes in the lab environment or inspection processes [12], [13]. Existing research in the lab automation domain primarily focuses on automated experimentation [14], [15]. However these systems still require human intervention for fault detection and correction.

To investigate the potential of robotic inspection and 'self-checks' in laboratory settings we propose combining robotic data capture with Large Language Models (LLMs) for data analysis. This combines the capabilities of the robots to systematically and efficiently capture visual and audio data, with the contextual prediction capabilities of LLMs [16], reducing the need for extensive prior knowledge. Our research aims to transform a traditional human-defined inspection checklist into a robotic framework that autonomously performs individual checklist tasks, evaluating both equipment and environmental conditions.

We test this method on a robotic cooking system with various failure modes. Our findings show that the robotic inspection system effectively identifies both anomalous and normal conditions with minimal reference data. Although the system's ability to evaluate the overall environment is limited in situations with sparse reference information, it demonstrates complete accuracy and robustness when provided with contextual data like past conditions and additional images. By assessing feedback from individual inspection tasks, the robotic system can autonomously decide whether an experiment can proceed, should be halted, or requires human intervention, along with set of troubleshooting steps.

In the remainder of the paper we introduce LLM analysis methods for captured data before providing details of the robotic and LLM implementation. Furthermore, we test performance of the robotic inspection system across several scenarios and conclude with a discussion on future steps.

II. METHODS

Checking the status of an experimental setup is crucial for laboratory experiments and this repetitive task is usually performed by humans. We propose system 'self-checks' by combining robotic automation for data capture with LLM data analysis to identify system status. Our approach uses a robot equipped with a camera and a microphone to collect data, which the LLM analyzes to determine system status, provide troubleshooting steps, and offer reasoning (Fig. 1).

A. Test Platform

The developed method is generalizable, however we deploy it for a robotic cooking setup that includes a commerical 6 Degrees of Freedom (DoF) robotic arm and a custom 3 DoF Cartesian arm as seen on Fig. 5 a). This setup requires checks for equipment functionality, presence of cooking ingredients and detection of unexpected or foreign objects.

B. Defining the checklist

We define each item of the checklist C_i by specifying it in the following format:

$$C_i = \text{Robot Data Capture, LLM Assessment Strategy}$$

where *Robot Data Capture* defines the position of the robotic arm in joint space, camera orientation, the type of recorded information (image or sound) and data post-processing method. Furthermore, *LLM Assessment Strategy* details the analysis method of captured data, including the type and quantity of the necessary grounding information.

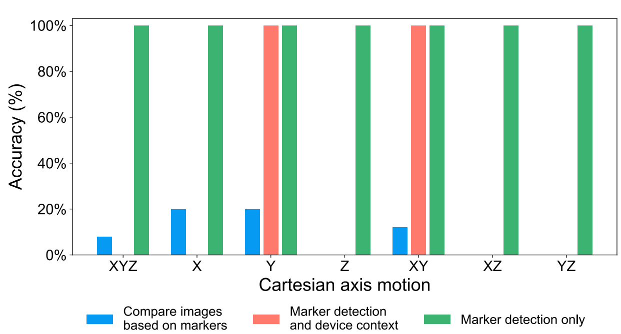

Fig. 2. Comparison of marker detection strategies across varied motions of the Cartesian platform.

C. Self-Checks

We classify the type of checklist tasks into different groups (i.e. malfunctioning equipment) and introduce the various robotic self-tests and how they can be made robust.

1) Detection of Equipment Malfunction:

a) Test for equipment motion: A primary failure mode of automation equipment is lack of motion, caused by hardware malfunctions, power outages or calibration discrepancies. To understand motion, LLMs typically analyze pairs of images taken before and after motion. This technique is effective for standard equipment like robot arms, with motion analysis task communicated to LLM through a textual prompt. However, for specialized robotic equipment like Cartesian arms, this approach is insufficient, as textual context does not accurately capture the device's operation as seen on Fig. 2. We propose an alternative approach using color-coded markers to highlight motion, where a missing marker of specific color indicates specific movements. This method not only improves accuracy and robustness but also requires analysis of just one image taken after the motion.

b) Test for equipment status: A wide range of devices display visual indicators that reflect their working conditions and settings (e.g. on/off buttons, LED screens). By interpreting these visual cues, LLM can determine equipment status. Although LLMs perform well at analyzing unedited camera images, their performance drops significantly when processing images crowded with various features and objects, typical in laboratory and industrial settings, and further degrades with images in incorrect orientation. To overcome these limitations, post-processing the original image, such as reorienting or cropping to emphasize key details is important. This step not only improves image understanding and reasoning but also reduces token usage and latency as seen on Fig. 3, a crucial metric for time intensive applications such as cooking. Analyzing full-resolution images with LLMs typically requires more tokens and takes longer than analyzing cropped images. If crucial details are divided between two smaller images, separate LLM queries for each image might be necessary to ensure accuracy, increasing token consumption and extending processing time. However, including all essential details into one cropped image, reduces latency due to a single LLM request, maintaining low token usage despite the larger image size. The efficiency can be further improved by adjusting the crop size or scaling the cropped image.

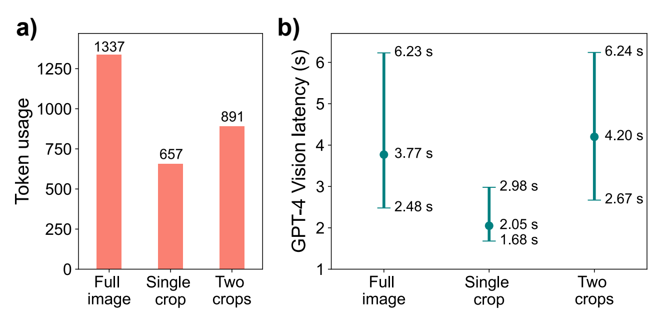

Fig. 3. LLM performance metrics for processing full or one or two cropped images. a) Token utilization, b) latency.

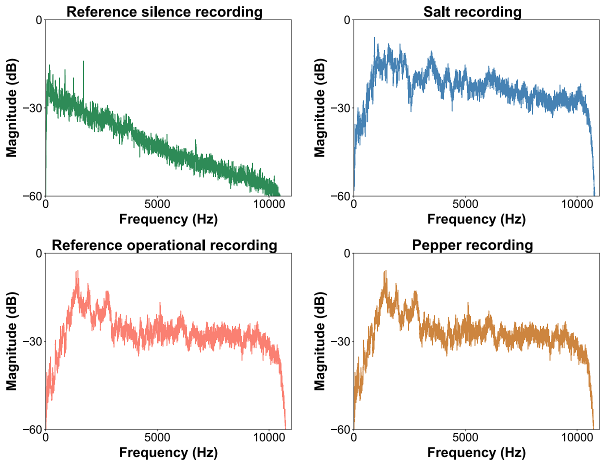

Fig. 4. Image provided to a LLM for analyzing a sound intensity across frequencies. Two latest and two reference recordings are provided.

c) Test for equipment noise: For devices like mixers and pumps, sound analysis is more effective than visual analysis for determining their status. Instead of developing a specific LLM for audio analysis, raw sound data can be converted into text or visual plots. Utilizing their image analysis capabilities, LLMs can examine and interpret plots, comparing reference data with latest information to evaluate equipment status, as shown in Fig. 4. In this instance, we assess operation of instruments by transforming sound recordings into frequency/magnitude plots. With this information, we can capture the unique sound signature of a specific device, enabling LLM to detect unfamiliar tones within certain frequencies, such as high-pitched sounds. Supplying the LLM with baseline silence recordings alongside the standard operating condition, provides necessary contextual insight. To enhance the LLM comprehension, both current and reference recordings are displayed as plots within a single image, using distinct colors for each plot.

D. Prior knowledge

Textual information effectively describes typical scenarios and identifies specific features within images, such as hob temperature (Fig. 5 g), but it cannot accurately identify object locations. By grounding LLM with imaging data, there is notable improvement in object localization tasks (Fig. 5 f).

E. Automated Data Augmentation

For tasks that evaluate spatial conditions, LLMs need extensive reference information to perform robustly, rather than minimal or no input. Providing a comprehensive background information of previous scenarios and maintaining awareness of recent developments enhances LLM reasoning. This strategy is particularly beneficial for tasks that exhibit considerable situational variability, such as workspace safety evaluation (Fig. 5 h). By leveraging historical data, LLMs can identify patterns and make informed decisions, thus boosting their adaptability in dynamic workspace settings.

F. Error handling

Our inspection method detects robotic system anomalies exclusively through external observations. By obtaining conclusions for each task in the checklist, LLM determines the overall status of the setup and provides troubleshooting steps if necessary. In case of major faults, like equipment failure, the experiment is stopped as mitigation of such condition requires significant amount of time. However, for minor anomalies, such as missing cooking ingredients, human intervention is requested before the experiment can proceed. The LLM can also offer set of troubleshooting steps by identifying root causes of problems using its general knowledge and checklist task conclusions.

III. EXPERIMENTAL SETUP

A. Experimental system description

The robotic cooking setup presented on Fig. 5 a) and b), features UR5 6 DoF robotic arm with mounted spatula. Additionally, a camera with integrated microphone is mounted atop a Dynamixel actuator on the cover of the 6th joint, enabling 360 degree rotation. This camera is utilized for inspecting the setup, while the robotic arm navigates the workspace through the set of waypoints. The rest of the cooking setup includes a custom 3 DoF Cartesian arm equipped with motorized spice dispensers and two syringe

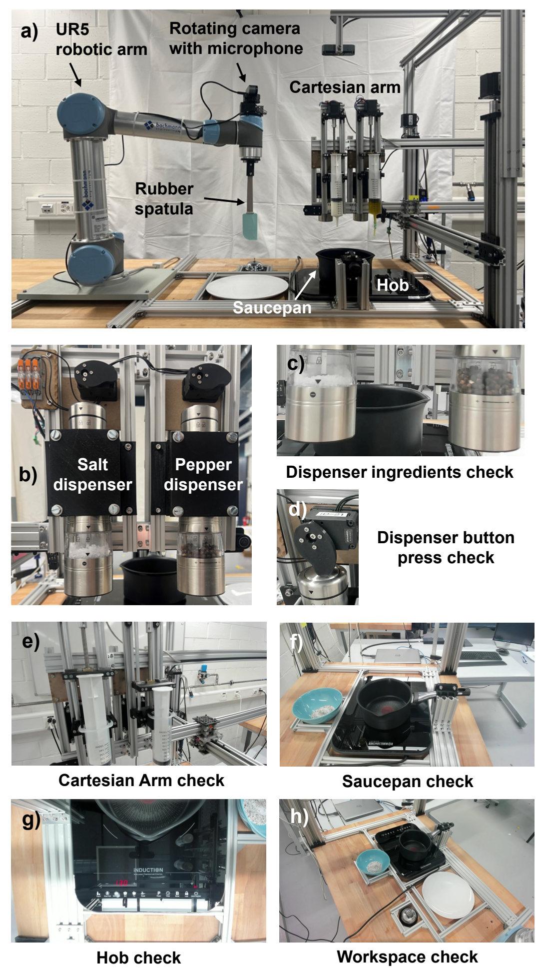

Fig. 5. Overview of the experimental setup and sample images for the LLM image analysis. a) Overall robotic cooking setup, b) spice dispensers, c) cropped image used for checking presence of ingredients in the spice dispensers, d) dispenser button press check, e) cartesian arm, f) saucepan, g) hob and h) workspace image check.

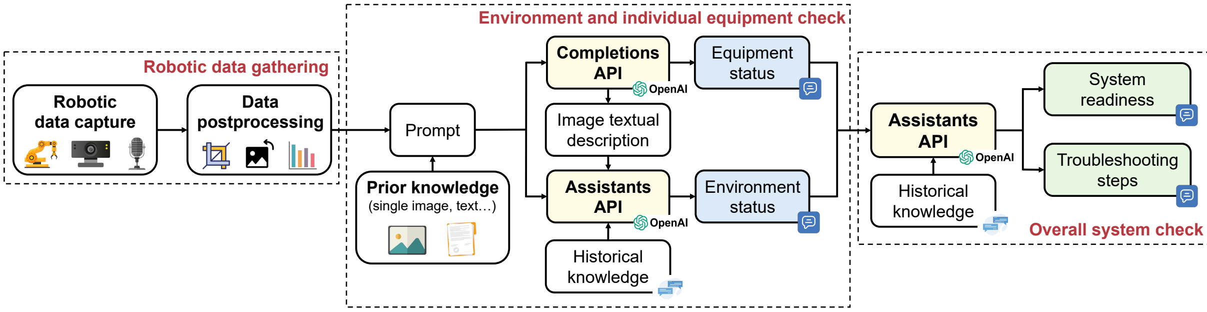

Fig. 6. Laboratory system self-assessment process including robotic data capture, individual task assessment and overall system readiness

pumps for precise ingredient handling. Actuators trigger the dispensers by pressing a button, and stepper motors drive the syringe pumps, ensuring accurate liquid dispensing. The cooking process takes place in a saucepan positioned on a induction hob, secured between aluminum profiles.

B. Implementation

1) Large Language Models: Captured data is analyzed using OpenAI GPT-4 model through Completions and Assistants API. Due to model limitations, image analysis is exclusively performed via Completions API, employing *gpt-4-1106-vision-preview* model [17]. For tasks requiring access to historical data, the Assistants API, utilizing *gpt-4-0125-preview* model is used. To ensure consistent output, the Completions API is configured by *temperature* parameter set to 0 and the fixed *seed* parameter [18].

2) Equipment operation - Image: Detecting operation of Cartesian platform (Fig. 5 e), induction hob (Fig. 5 g) and ingredient dispenser (Fig. 5 d) is performed using LLM vision capabilities, with context provided in textual form within a prompt. A single image, typically captured during operation or after a motion event is analyzed by the LLM. Specifically, for spice dispensers, visual inspection method is utilized as alternative approach.

3) Equipment operation - Sound: The operation of ingredient dispensers is primarily detected by analyzing 6-second sound recordings, as these devices emit distinct loud sound during operation. These recordings are post-processed for every inspection to extract overall loudness and sound frequency distribution. The loudness is quantified using the LUFS (Loudness Unit Full Scale) metric, which measures perceived loudness by averaging sound levels over time [19]. This metric enables LLM to identify loud sounds and is provided in text form. Furthermore, by analyzing sound intensity across frequencies, LLM can differentiate dispenser sounds from those of nearby equipment. This information is passed to the LLM in the visual format, as shown in Fig. 4.

4) Container substance level monitoring: Measuring ingredient quantities in spice dispensers presents a challenge for LLMs, allowing only detection of ingredient presence. Additionally, analyzing full images often leads to inconsistent outcomes, making it necessary to crop images to the area of interest (Fig. 5 c). With such focused input, minimal context within textual prompt suffices for the reliable detection of common substances such as salt and pepper.

5) Object placement evaluation: Given the constraints of LLMs in understanding spatial relationships, this task requires comparison of a recent image with a reference to accurately determine equipment status. In the analyzed workspace, the LLM is tasked with assessing two factors, the placement of the saucepan on the hob and the security of its handle. Accurately addressing both questions may be challenging for the LLM, raising the risk of overestimating potential hazards. However, within the context of the overall system evaluation, this concern is considered minor and can be easily addressed with minimal human intervention.

6) Workspace assessment: Before starting the experiment, the workspace must be inspected for unexpected objects that could cause experiment failure or damage equipment. This task poses a significant challenge for the LLMs, as it requires precise differentiation between safe and hazardous conditions, particularly in scenarios involving foreign objects. For accurate task execution, LLM requires historical context of past situations to make well-informed decisions. Currently, the Assistants API cannot directly analyze images. Therefore, the process begins with the Completions API, which interprets images and generates textual descriptions which are then forwarded to the Assistant for safety evaluation.

7) Overall system assessment: The evaluation of the overall system is carried out via Assistants API, which receives only the final status and reasoning for each checklist task. This approach prevents over-contextualization, allowing implementation of adaptive decision-making strategies to maximize system availability. For instance, upon detecting a fault, the LLM directs the robot to inspect faulty condition using an alternative method. The system is only declared unsafe if this secondary inspection also fails.

C. LLM Prompt Structure

For each interaction with the LLM via the Completions or Assistants API, the prompt is organized into four distinct sections: *Objective, Instructions, Important Notes and Output Indicator.* The *Objective* section outlines the goal of the task, while the *Instructions* section provides step-by-step guide for problem analysis. Final conclusion alongside reasoning is provided in format specified by the *Output Indicator* section. Additionally, the *Important Notes* section provides the LLM with additional task context. Typically, prompts are conveyed as a single message, but for tasks requiring image comparisons or extensive prompts, these sections may be delivered as separate messages to ensure better comprehension.

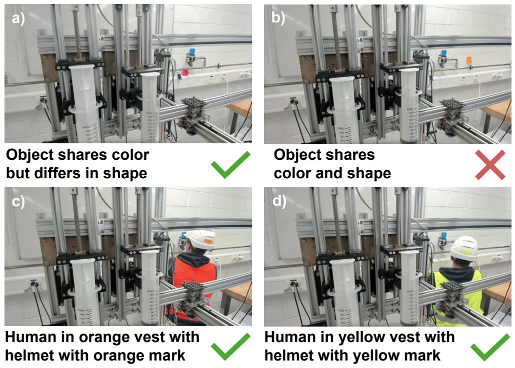

Fig. 7. Performance of LLM in detecting marks on a Cartesian arm across scenarios with objects a) similar in color but different in shape, b) similar in both color and shape, and c), d) featuring a human wearing a yellow or orange safety vest with matching color mark on the helmet.

IV. EXPERIMENTAL RESULTS

A. Detection of Cartesian Axis Motion

To detect Cartesian arm axis movements, the LLM relies on detection of color-coded markers. The performance was evaluated in a lab setting by assessing LLM's ability to detect markers on XYZ axes (Pass/Fail) and provide reasoning for its decisions. This strategy effectively distinguishes between similarly colored objects with different shapes as seen on Fig. 7 a). However, it faces difficulties when objects share both shape and color (Fig. 7 b). The LLM excels in scenarios where a human is present and is wearing a similarly colored vest and helmet with a matching color mark (Fig. 7 c, d).

B. Detection of Equipment through Sound

For devices which emit noise during operation, sound analysis is a preferred way to check functionality. Given the prevalence of background noise in the labs, it is crucial to evaluate LLM performance in both expected (e.g. mixer noise) and unexpected (e.g. bell sound) noise conditions. For each sound sample the LLM is tasked to detect operational status of the spice dispenser (Pass/Fail/External noise) and to provide brief reasoning. The results are given in Fig. 8.

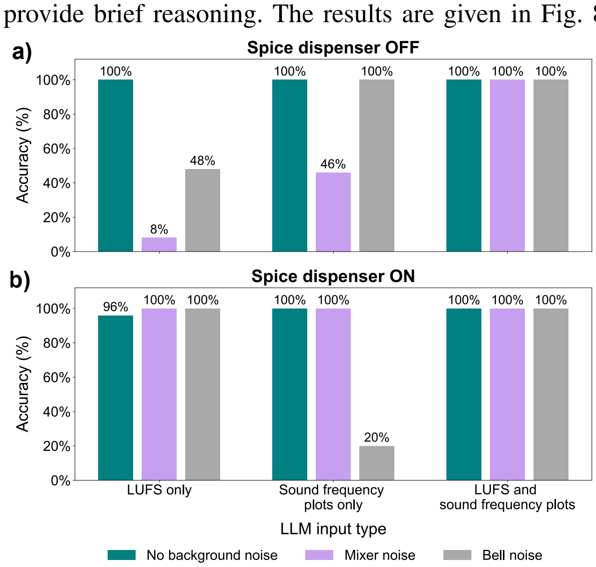

Fig. 8. Influence of various sound metrics on the ability of LLM to detect a) malfunctioning and b) operational spice dispensers.

Under ideal, noise-free conditions, the LLM accurately identifies dispenser status with only one or both sound metrics, as seen on Fig. 8 a) and b). In the presence of background noise, the sound emitted by the dispenser remains the most dominant, allowing reliable detection. In contrast, when the dispenser is faulty, background noise becomes a disruptive factor, impeding accurate detection in scenarios where one of the sound metrics is not considered.

C. Detection of Equipment Location

Checking the correct placement of equipment prior to experiments is crucial, but performing this task using LLMs is challenging due to their limitations. By grounding LLM with a single reference image that illustrates a safe condition, we task LLM to assess various saucepan positions in a workspace. For every inspection, the LLM accurately detects significant alterations from the reference image, but fails to identify hazardous conditions that resemble safe ones.

D. Workspace Safety Evaluation

Detecting presence and location of foreign objects, and assessing their risk to the safety of the overall system is a crucial task in system readiness assessment. By providing information regarding safety rules, past scenarios and images from different camera angles, we task LLM to classify workspace as safe or unsafe for experimentation.

Considering the diversity of workspace scenarios, the LLM underperforms when provided with no extensive context, often overestimating risk and providing inaccurate justifications in nearly half of the correct decisions, as shown in Fig. 9. With inclusion of historical data for both safe and hazardous conditions, the LLM overestimates safe situations but accurately justifies all correctly identified scenarios. Additionally, LLM image textual descriptions often misinterpret object sizes, for instance perceiving a pepper container for being as large as a milk carton. However, using images from three different angles mitigates these inaccuracies, thus enhancing decision-making and reasoning capabilities.

E. Full System Deployment

Using each of the demonstrated unit tests, a full system check can be performed. We simulate different errors, equipment faults, and also minor problems that can be resolved with quick human intervention. For four different scenarios,

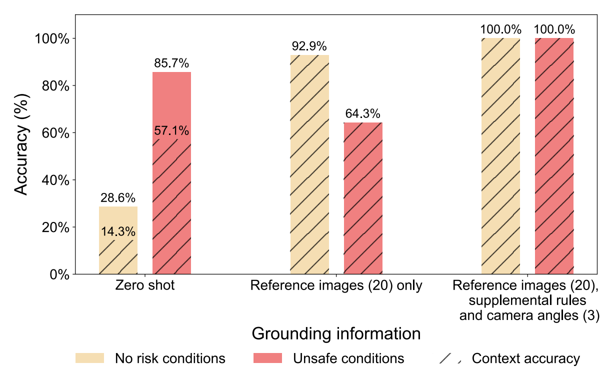

Fig. 9. Ability of the LLM to accurately assess workspace safety conditions with different levels of contextual information.

	Cartesian Arm	Saucepan	Hob	Environment	Dispenser (ingredients)	Dispenser (sound)	Dispenser (button)	Conclusion
System fully functional	✓	✓	✓	✓	✓	✓	N/A	Continue experiment
Equipment failure	✗	✓	✓	✓	✗	✓	N/A	Stop experiment
Human intervention	✓	✓	✓	✓	✗	✓	N/A	Human intervention
Optional check	✓	✓	✓	✓	✓	✗	✓	Continue experiment

✓ Equipment status PASS, verified by a human; ✗ Equipment status FAIL, verified by a human; N/A Equipment test not performed

Fig. 10. Overall system check with four distinct scenarios to evaluate LLM's equipment status assessment and reasoning.

a series of different checks, the status identified by the robot and the final conclusion given by the robot are shown in Fig. 10. The detection process accounts for downstream effects. For example, a failure in early checks, such as the Cartesian platform, also shows a failure in dispenser ingredient evaluation due to its unexpected position. Furthermore, the approach is flexible with ability to handle alternative assessments, such as directing robot to visually inspect dispensers, in case sound check fails.

V. DISCUSSIONS & CONCLUSIONS

In this work, we provide a method for combining robotic data capture and LLMs for self assessment of the readiness of robotic laboratory systems. This is applied to a robotic cooking setup that utilizes this method to analyze visual and auditory data captured by a camera-equipped robotic arm, ensuring accurate assessment of equipment and workspace readiness. Our findings show that color marker detection from an image is effective for detecting motion and operation of equipment, and correct device operation can be evaluated from images where no contextual information is given. For devices which signal their operation through sound, by using two separate metrics, robust sound detection can be performed even in presence of background noise. Regarding overall system health, the LLM reliably differentiates critical from minor issues, distinguishing between equipment malfunctions and minor faults that require quick human intervention to resume experimentation.

This approach could be extended to allow for continuous monitoring, providing more real-time insights, comparisons between scenarios and early detection of errors. Furthermore, captured data can be used for optimization of system parameters, allowing online adjustment of equipment for optimal performance. Finally, if the robot capabilities could be extended to incorporate failure recovery, this could enable self checking and self recovery from failure.

ACKNOWLEDGMENT

This project was supported by the EU's Horizon 2020 research and innovation program under the Marie Skłodowska Curie grant agreement No 945363.

REFERENCES

[1] P. Seavill, "Research with robotics," *Nature Synthesis*, vol. 2, no. 6, pp. 467–468, 2023.

[2] R. W. Epps and M. Abolhasani, "Modern nanoscience: Convergence of ai, robotics, and colloidal synthesis," *Applied Physics Reviews*, vol. 8, no. 4, 2021.

[3] T. Howison, J. Hughes, and F. Iida, "Large-scale automated investigation of free-falling paper shapes via iterative physical experimentation," *Nature Machine Intelligence*, vol. 2, no. 1, pp. 68–75, 2020.

[4] M. Costanzo, G. De Maria, G. Lettera, and C. Natale, "A multimodal approach to human safety in collaborative robotic workcells," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 2, pp. 1202–1216, 2021.

[5] I. Holland and J. A. Davies, "Automation in the life science research laboratory," *Frontiers in bioengineering and biotechnology*, vol. 8, p. 571777, 2020.

[6] C. Kanellakis, E. Fresk, S. S. Mansouri, D. Kominiak, and G. Nikolakopoulos, "Towards visual inspection of wind turbines: A case of visual data acquisition using autonomous aerial robots," *IEEE access*, vol. 8, pp. 181 650–181 661, 2020.

[7] A. Khan, C. Mineo, G. Dobie, C. Macleod, and G. Pierce, "Vision guided robotic inspection for parts in manufacturing and remanufacturing industry," *Journal of Remanufacturing*, vol. 11, no. 1, pp. 49–70, 2021.

[8] L. Yu, E. Yang, P. Ren, C. Luo, G. Dobie, D. Gu, and X. Yan, "Inspection robots in oil and gas industry: A review of current solutions and future trends," in *2019 25th International Conference on Automation and Computing (ICAC)*. IEEE, Sep. 2019.

[9] M. Stachowiak, W. Koperska, P. Stefaniak, A. Skoczylas, and S. Anufriiev, "Procedures of detecting damage to a conveyor belt with use of an inspection legged robot for deep mine infrastructure," *Minerals*, vol. 11, no. 10, p. 1040, 2021.

[10] J. J. Lin, A. Ibrahim, S. Sarwade, and M. Golparvar-Fard, "Bridge inspection with aerial robots: Automating the entire pipeline of visual data capture, 3D mapping, defect detection, analysis, and reporting," *J. Comput. Civ. Eng.*, vol. 35, no. 2, p. 04020064, Mar. 2021.

[11] M. Pech, J. Vrchota, and J. Bednář, "Predictive maintenance and intelligent sensors in smart factory," *Sensors*, vol. 21, no. 4, p. 1470, 2021.

[12] W. Chen, S. Hu, R. Talak, and L. Carlone, "Leveraging large language models for robot 3d scene understanding," *arXiv preprint arXiv:2209.05629*, 2022.

[13] S. Huang, Z. Jiang, H. Dong, Y. Qiao, P. Gao, and H. Li, "Instruct2act: Mapping multi-modality instructions to robotic actions with large language model," *arXiv preprint arXiv:2305.11176*, 2023.

[14] B. Burger, P. M. Maffettone, V. V. Gusev, C. M. Aitchison, Y. Bai, X. Wang, X. Li, B. M. Alston, B. Li, R. Clowes, N. Rankin, B. Harris, R. S. Sprick, and A. I. Cooper, "A mobile robotic chemist," *Nature*, vol. 583, no. 7815, pp. 237–241, Jul. 2020.

[15] T. Ha, D. Lee, Y. Kwon, M. S. Park, S. Lee, J. Jang, B. Choi, H. Jeon, J. Kim, H. Choi *et al.*, "Ai-driven robotic chemist for autonomous synthesis of organic molecules," *Science advances*, vol. 9, no. 44, p. eadj0461, 2023.

[16] F. Petroni, P. Lewis, A. Piktus, T. Rocktäschel, Y. Wu, A. H. Miller, and S. Riedel, "How context affects language models' factual predictions," *arXiv preprint arXiv:2005.04611*, 2020.

[17] OpenAI, "Gpt-4v(ision) system card," 2023.

[18] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.

[19] E. Skovenborg and T. Lund, "Loudness descriptors to characterize programs and music tracks," in *Audio Engineering Society Convention 125*. Audio Engineering Society, 2008.