A Supplementary

A.1 Video of stability testing

The video of stability testing is available in the link[1].

A.2 Video of solid-state workflow testing

The video of the solid-state workflow testing is available in the link[2]. The key steps where LIRA performs vision-based localization, inspection, and reasoning throughout the workflow are summarized in Table S1, with corresponding timestamps indicating when each function is executed in the video.

Time	LIRA vision-based functions
0 min 53 sec	**Localization** at the ChemSpeed station to unload the 8-hole white rack to the robot base.
2 min 36 sec	**Reasoning** the placement of 8-hole white rack on the robot base.
7 min 14 sec	Localization at the preparation station to load the 8-hole white rack.
8 min 00 sec	**Reasoning** the placement of 8-hole white rack on the preparation station.
8 min 33 sec	**Localization** at the preparation station to load the PXRD plate.
9 min 20 sec	**Reasoning** the placement of the PXRD plate on the preparation station.
9 min 48 sec	**Localization** at the preparation station to unload the 8-hole white rack to the robot base.
10 min 26 sec	**Reasoning** the placement of the 8-hole white rack on the robot base.
10 min 46 sec	**Localization** at the preparation station to unload the PXRD plate to the robot base.
11 min 26 sec	**Reasoning** the placement of the PXRD plate on the robot base.
14 min 18 sec	**Localization** at the PXRD station to open the doors.
15 min 54 sec	**Reasoning** the opening states of doors.
16 min 48 sec	**Localization** inside the PXRD station to load the PXRD plate.
18 min 26 sec	**Reasoning** the placement of the plate inside the PXRD station.
19 min 21 sec	**Localization** at the PXRD station to close the doors.
20 min 49 sec	**Reasoning** the opening states of doors.
21 min 22 sec	**Localization** at the PXRD station to open the doors.
22 min 59 sec	**Reasoning** the opening states of doors.
23 min 47 sec	**Localization** inside the PXRD station to unload the PXRD plate.
25 min 44 sec	**Reasoning** the placement of the PXRD plate on the robot base.
26 min 27 sec	**Localization** at the PXRD station to close the doors.
27 min 59 sec	**Reasoning** the closing states of doors.

Table S1. Timeline of LIRA vision-based functions during the solid-state workflow testing.

A.3 Video of error recovery

The video of error recovery process is available in the link[3]. Figure S1 illustrates the error recovery process demonstrated in Video 3, where the robot detects and corrects a translational offset in object placement. At step ①, the robot captures an image from the designated inspection pose. The image is then sent to LIRA, which returns an inspection and reasoning result (step ②). This result is processed by the InspectionHandler function, which determines the appropriate corrective action.

Based on the reasoning result, the robot identifies the offset direction. For example, in Figure S1, the PXRD rack is detected as misaligned toward the back direction. The robot then moves to a predefined position, aligning itself for corrective action (step ③). If the error is classified as recoverable, the robot executes a controlled push to reposition the rack correctly (step ④).

[1] https://youtu.be/ErtmiBhSD6Q
[2] https://youtu.be/9KjuhlKK-ag
[3] https://youtu.be/7gfVNBQCjw8

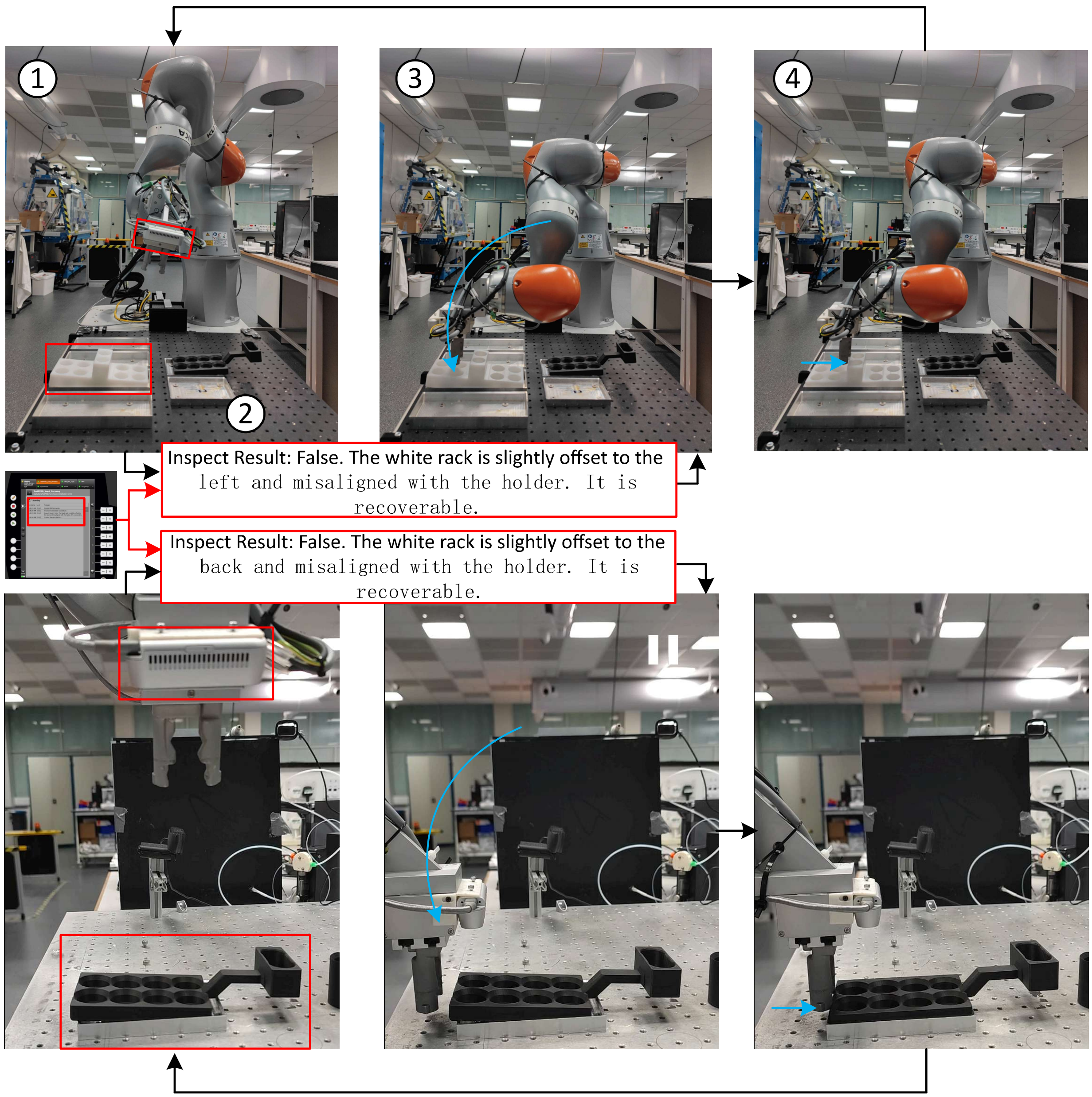


Figure S1. **Illustration of the error recovery process demonstrated in Video 3, where the robot reasons about a translational offset and executes corrective manipulation.**

A.4 Fine-Tuning Dataset and Prompt Dictionary

A total of 336 images were used to fine-tune the VLM, while 60 images were allocated for evaluation during training. An additional 66 unseen images were reserved for testing the fine-tuned VLM. Some images of door manipulation failures were reused, as opening and closing failures exhibit the same characteristics. The dataset is available in `dataset` folder through the link[4].

Table S2 presents examples of failure cases from the image dataset and prompts used to fine-tune the VLM for inspection and reasoning tasks. For PXRD Plate and 8-hole Rack, the dataset includes reasoning results of recoverable placement errors, where the robot attempts to correct the misalignment based on reasoning feedback. These errors involve translational offsets that can be adjusted through predefined robotic actions. In contrast, for other targets such as the LCMS rack, NMR rack, and PXRD door, once a failure is detected, the system pauses the workflow and wait for human intervention, as these misalignments—especially rotational errors and complex occlusions—are not recoverable by the robot. The table categorizes failure types based on inspection targets and their respective locations, with structured prompts used during model inference. This structured dataset enables the VLM to accurately detect placement failures and determine appropriate responses for automated correction or operator intervention.

A.5 Testing of fine-tuned VLM

To evaluate the performance of the fine-tuned VLM before applying it to the robot, we conducted inference testing on 66 unseen images. The evaluation code is available in the link[4].

The evaluation was performed as follows:

1. **Model Setup**: The fine-tuned VLM was loaded onto an NVIDIA GPU for inference.
2. **Dataset Preparation**: A separate test dataset containing 66 images was used, with each sample including:
- An input image of a placement scenario.
- A structured inspection prompt.
- A ground truth response.

3. **Inference and Comparison**: - The model processed the input images and prompts, generating textual responses. - The **generated response** was compared against the ground truth using a string similarity metric. - Responses were marked correct if their similarity score exceeded a threshold of 0.9.

4. **Performance Metrics**: - The model achieved an overall success rate of 100%, correctly classifying all 66 samples.

Users can re-run the inference testing by following the provided instructions and dataset structure in the link[4]. Our testing result for the fine-tuned VLM is shown in `evaluation_results.json`

A.6 Inspection Handler

The Inspection Handler is a critical component of the LIRA framework, responsible for processing vision-based inspection and reasoning results from the VLM. This module determines whether the robotic manipulator should proceed with its workflow, initiate an error recovery action, or pause for human intervention.

Both the video[3] and Figure S2 illustrate the error recovery process executed by the robot using the Inspection Handler. When the robot performs an inspection, it sends a natural language prompt to LIRA, which returns an evaluation result. If the response is `"True"`, the robot continues its workflow. If the response is `"False"`, the system analyzes the reasoning output to determine whether the error is recoverable.

For recoverable errors, such as translational placement offsets, the handler extracts directional information (e.g., left, right, forward, or backward) and triggers predefined correction movements using the robotic arm. The correction process involves moving to a predefined pose and applying a controlled push to realign the misplaced object. After that, the robot will repeat the inspection.

For non-recoverable errors, such as excessive misalignment or rotational offsets, the handler pauses the workflow and awaits human intervention. The system logs the error and notifies the operator for manual correction.

The Java implementation (InspectionHandler.java[4]) demonstrates how these components interact to enable closed-loop error recovery.

A.7 Reproduce

To reproduce LIRA, please follow the steps of the repository[4].

B Materials & Correspondence

[4]https://github.com/cooper-group-uol-robotics/LIRA.git

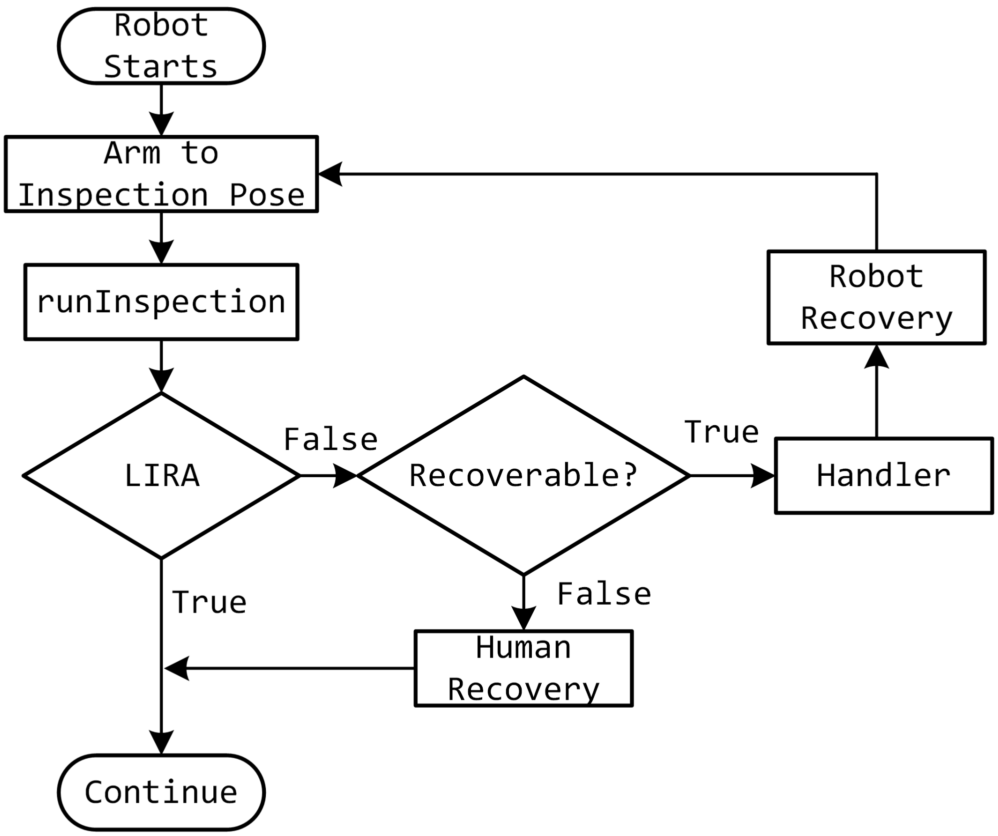


Figure S2. **Flowchart of the recovering procedure**

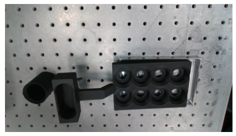
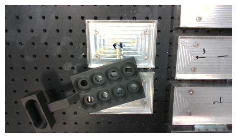
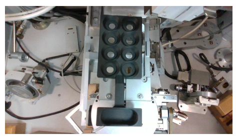
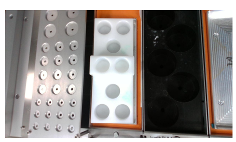
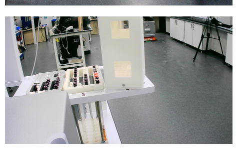
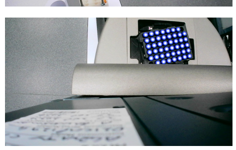
Inspect Targets	Location	Image	Prompt
PXRD Plate	Preparation Station		Q: "Has the black rack been placed onto the holder properly?" A: "False. The black rack is slightly offset to the right and misaligned with the holder. It is recoverable."
PXRD Plate	Robot base		Q: "Has the black rack been placed onto the holder properly?" A: "False. The black rack is slightly rotated and misaligned with the holder. It's unrecoverable."
PXRD Plate	PXRD		Q: "Has the black rack been placed onto the holder properly?" A: "False. The black rack is slightly offset to the left and misaligned with the holder. It is recoverable."
8-hole Rack	Preparation Station		Q: "Has the 8-hole white rack been placed onto the holder properly?" A: "False. The white rack is slightly offset to the left and misaligned with the holder. It is recoverable."
8-hole Rack	Robot base		Q: "Has the 8-hole white rack been placed onto the holder properly?" A: "False. The white rack is slightly offset to the left and misaligned with the holder. It is recoverable."
8-hole Rack	Chemspeed		Q: "Has the 8-hole white rack been placed onto the holder properly?" A: "False. The white rack is slightly offset to the left and misaligned with the holder. It is recoverable."
Door	PXRD		Q: "Has the door been closed properly?" A: "False."
NMR Rack	NMR		Q: "Has the NMR rack been placed onto the holder properly?" A: "False. The black rack is slightly rotated and misaligned with the holder. It's unrecoverable."
LCMS Rack	LCMS		Q: "Has the white long rack been placed onto the holder properly?" A: "False."

Table S2. Examples of failure cases and the reasoning outcomes of LIRA