

# Supporting Information for

## TARMAC: A Taxonomy for Robot Manipulation in Chemistry

*K. Huang<sup>1</sup>, A. E. Martin<sup>2</sup>, J. Pipe<sup>1</sup>, T. Wang<sup>1</sup>, B. A. Franklin<sup>2</sup>, A. M. Tyrrell<sup>1</sup>,  
I. J. S. Fairlamb<sup>2</sup>, and J. Zhu<sup>1,\*</sup>*

<sup>1</sup>School of Physics, Engineering and Technology, University of York, York, YO10 5DD, UK

<sup>2</sup>Department of Chemistry, University of York, York, YO10 5DD, UK

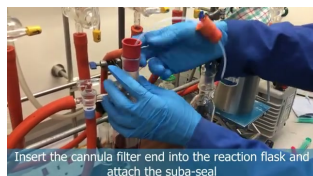
\*Correspondence to J. Zhu ([jihong.zhu@york.ac.uk](mailto:jihong.zhu@york.ac.uk))

## Contents

<b>1</b>	<b>Examples of Dataset Videos</b>	<b>S2</b>
<b>2</b>	<b>Specification-adequacy probe: detailed metrics</b>	<b>S7</b>
<b>3</b>	<b>Cross-corpus protocol decomposition</b>	<b>S9</b>
<b>4</b>	<b>Single-Arm Franka Validation Experiments</b>	<b>S14</b>
<b>5</b>	<b>Detailed Experiment Logs</b>	<b>S17</b>

# 1 Examples of Dataset Videos

The following figures show annotated screenshots from the dataset videos, illustrating how the TARMAC taxonomy describes complex, multi-step laboratory procedures as sequences of primitive actions. Each figure presents selected frames from a single video, with the corresponding TARMAC action category and a brief description of the manipulation.



## (a) Elastic Coupling

Attach the sub-seal to the reaction flask by deforming the compliant seal over the neck.



## (b) Rotational Positioning

Rotate the reaction flask from vertical to horizontal to prepare for filtration.



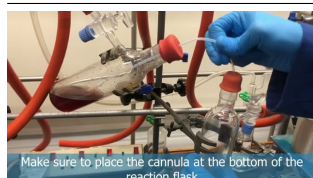
## (c) Transitional Positioning

Remove the glass stopper from the receiving flask by pulling along its axis.



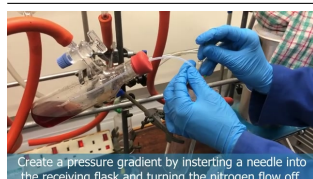
## (d) Elastic Coupling

Fit a sub-seal onto the receiving flask to create an airtight connection for the cannula.



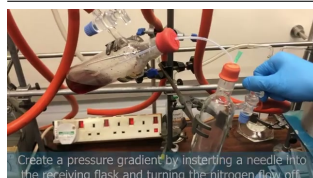
## (e) Transitional Coupling

Thread the cannula through the sub-seal into the reaction flask by applying linear force.



## (f) Penetrative Coupling

Pierce the sub-seal of the receiving flask with a needle to create a pressure gradient.



## (g) Rotational Positioning

Turn the nitrogen tap to shut off gas flow, establishing the pressure differential for solvent transfer.

Figure S1. Screenshots from the Schlenk line cannula filtration video (Video 003). This air-sensitive technique involves inert-atmosphere handling, compliant seal manipulation, and pressure-driven fluid transfer, all described by the TARMAC action primitives.



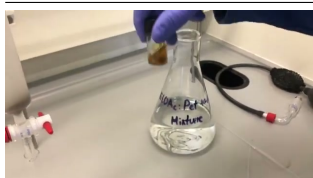
**(a) Squirting**

Pipette the dissolved sample into a vial for loading onto the column.



**(b) Rotational Coupling**

Screw the vial lid closed to seal the sample.



**(c) Shaking**

Shake the flask to dissolve the sample in the eluent mixture.



**(d) Squirting**

Load the sample solution onto the top of the silica column using a pipette.



**(e) Rotational Positioning**

Open the column tap to allow the sample to elute into the silica.



**(f) Scooping**

Scoop sand to add a protective layer on top of the silica.



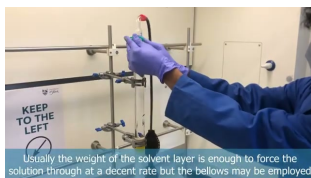
**(g) Pouring**

Pour eluent solvent into the column reservoir.



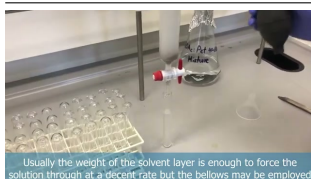
**(h) Transitional Positioning**

Insert the adapter into the top of the column.



**(i) Transitional Coupling**

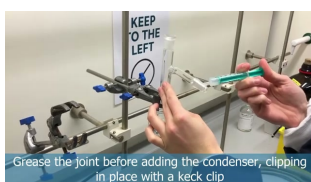
Attach a clip to secure the adapter in place.



**(j) Squirting**

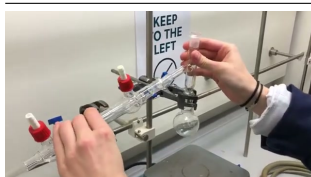
Squeeze the bellows to push eluent through the column under pressure.

Figure S2. Screenshots from the column chromatography video (Video 035). The procedure involves sample preparation, column loading, and elution—each step captured by TARMAC action primitives.



**(a) Squirting**

Apply grease to the glass joint before assembling the condenser.



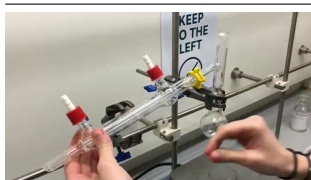
**(b) Transitional Positioning**

Insert the condenser onto the still head.



**(c) Twisting**

Twist the condenser to spread the grease evenly across the joint.



**(d) Transitional Coupling**

Attach a Keck clip to secure the greased joint.



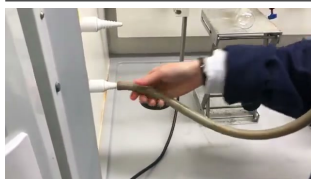
**(e) Transitional Positioning**

Insert the pig (receiver adapter) onto the condenser outlet.



**(f) Twisting**

Twist the pig to seat the joint and spread grease.



**(g) Elastic Coupling**

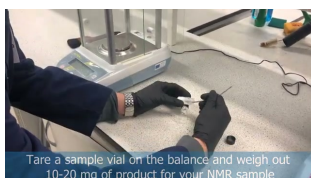
Connect the vacuum tubing to the distillation apparatus.



**(h) Transitional Positioning**

Insert the thermometer into the still head to monitor distillation temperature.

Figure S3. Screenshots from the reduced pressure distillation video (Video 024). The assembly requires repeated greasing, insertion, and securing of glass joints under vacuum-tight conditions, all captured by TARMAC action primitives.



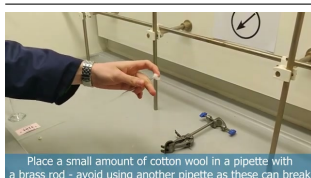
**(a) Scooping**

Scoop the solid sample into a vial on the balance.



**(b) Rotational Coupling**

Close the vial cap after weighing the sample.



**(c) Transitional Positioning**

Insert cotton wool into a Pasteur pipette to create a filter plug.



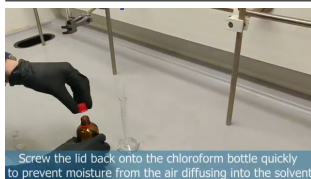
#### (d) Rotational Coupling

Unscrew the cap of the deuterated solvent bottle.



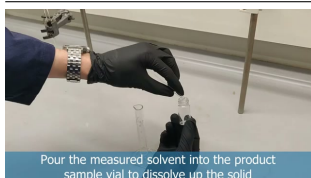
#### (e) Squirting

Pipette deuterated solvent into the NMR tube.



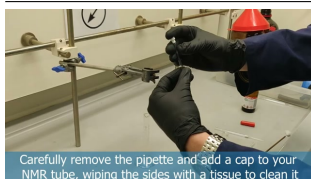
#### (f) Rotational Coupling

Screw the lid back onto the solvent bottle to prevent moisture ingress.



#### (g) Pouring

Pour the dissolved sample from the vial into the NMR tube.



#### (h) Transitional Coupling

Cap the NMR tube after filling.



#### (i) Wiping

Wipe the exterior of the NMR tube to remove residue.



#### (j) Transitional Positioning

Insert the sample identification tag onto the NMR tube.

Figure S4. Screenshots from the NMR sample preparation video (Video 017). The procedure spans weighing, dissolving, filtering, and labelling—each manipulation captured by TARMAC action primitives.

## 2 Specification-adequacy probe: detailed metrics

This section reports the per-primitive metrics and confusion structure underlying the rule-following classification probe described in the main text. Figure S5 shows the full  $18 \times 18$  confusion matrix in the full-context condition. Rows are ground-truth labels; columns are the rater’s predictions; the diagonal corresponds to correct classifications. Disagreements concentrate within the Coupling block (Transitional-C  $\leftrightarrow$  Elastic-C  $\leftrightarrow$  Rotational-C) and at the Positioning–Coupling interface (Transitional-P  $\rightarrow$  Insertive-P), localising the wrench/gravity axis as the principal site of taxonomic continuity.

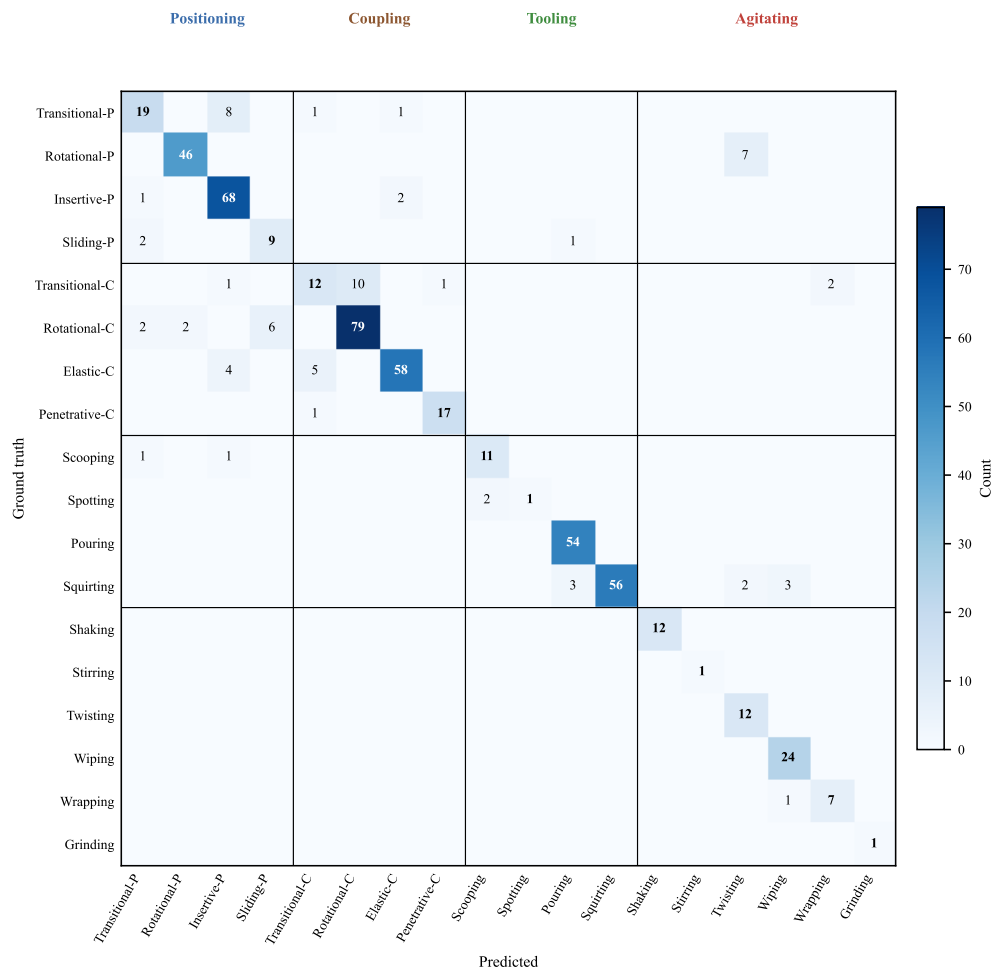


Figure S5. Confusion matrix of the rule-following classifier in the full-context condition (562 actions, 18 primitives, GPT-5.4). Rows are ground-truth labels; columns are predictions; the diagonal corresponds to correct classifications. Disagreements concentrate within the Coupling block and at the Positioning–Coupling interface, consistent with the wrench/gravity axis as a continuous physical variable discretised by the taxonomy.

Table S1 provides per-primitive support and F1 across the four context-ablation conditions, organised by TARMAC category. Macro-averaged F1 and overall accuracy are reported in the final two rows. The  $\Delta$  column reports the F1 difference between the full-context and action-only conditions; positive values indicate primitives whose disambiguation depends on workflow context.

Table S1. Per-primitive support and F1 of the rule-following classifier across the four context-ablation conditions, organised by TARMAC category. Columns: **Action** (action name only), **+ Procedure** (action plus the name of the source procedure), **+ Prev** (also the three preceding actions from the same procedure), **Full** (also the three following actions). The  $\Delta$  column reports the F1 change between the **Full** and **Action** conditions.

<b>Primitive</b>	<b>Support</b>	<b>Action</b>	<b>+ Video</b>	<b>+ Prev</b>	<b>Full</b>	<b><math>\Delta</math></b>
<i>Positioning</i>						
Transitional-P	30	0.793	0.793	0.727	0.691	-0.102
Rotational-P	53	0.930	0.938	0.932	0.911	-0.019
Insertive-P	71	0.861	0.883	0.859	0.889	+0.028
Sliding-P	12	0.720	0.750	0.750	0.667	-0.053
<i>Coupling</i>						
Transitional-C	27	0.537	0.444	0.522	0.522	-0.015
Rotational-C	89	0.888	0.904	0.918	0.888	+0.000
Elastic-C	67	0.825	0.828	0.846	0.906	+0.081
Penetrative-C	18	0.895	0.895	0.895	0.944	+0.049
<i>Tooling</i>						
Scooping	13	0.815	0.815	0.786	0.846	+0.031
Spotting	3	0.000	0.500	0.500	0.500	+0.500
Pouring	57	0.909	0.929	0.948	0.939	+0.030
Squirting	64	0.839	0.867	0.887	0.933	+0.094
<i>Agitating</i>						
Shaking	12	1.000	1.000	0.960	1.000	+0.000
Stirring	1	1.000	1.000	1.000	1.000	+0.000
Twisting	12	0.545	0.615	0.800	0.727	+0.182
Wiping	24	0.828	0.873	0.873	0.923	+0.095
Wrapping	8	0.875	0.875	0.778	0.824	-0.051
Grinding	1	0.000	0.000	1.000	1.000	+1.000
<b>Macro F1</b>	—	0.737	0.773	0.832	0.839	+0.103
<b>Accuracy</b>	562	0.829	0.851	0.861	0.867	+0.037

### 3 Cross-corpus protocol decomposition

This section reports the per-paper details, coverage table, and unmapped-operation breakdown for the cross-corpus expressive-coverage analysis described in the main text (Section “Validation of expressive coverage via protocol decomposition”).

#### 3.1 Corpus and selection

The 23 procedures were drawn from four published chemistry papers selected to span distinct manipulation regimes:

- **Paper 1: hydrogel mechanics** [1]. Five procedures: (i) alginate–polyacrylamide hybrid gel synthesis, (ii) tensile testing of specimens with various crack lengths, (iii) double-peeling test, (iv) crack-length effect, (v) sample-size effect.
- **Paper 2: cesium-lead-halide perovskite nanocrystals** [2]. Four procedures: (i) Cs-oleate stock preparation, (ii) hot-injection CsPbX<sub>3</sub> nanocrystal synthesis, (iii) isolation and purification by centrifugation, (iv) photo-polymerisation reaction.
- **Paper 3: NU-1500 metal-organic framework** [3]. Five procedures: (i–iii) solvothermal synthesis of NU-1500 on Fe, Cr, and Sc metal centres; (iv) single-crystal growth variant; (v) post-synthetic activation under thermal vacuum.
- **Paper 4: air-sensitive palladium cross-coupling** [4]. Nine procedures: (i) general [Pd<sub>3</sub>(OAc)<sub>6</sub>/PPh<sub>3</sub>]-ratio Schlenk procedure; (ii) base and additive screen for Suzuki–Miyaura cross-coupling (SMCC); (iii) product-evolution kinetics; (iv) Kumada–Corriu cross-coupling site selectivity; (v) para-substituent effects in SMCC; (vi) general workup procedure for SMCC reactions; (vii) oxidative addition of [Pd(PPh<sub>3</sub>)<sub>4</sub>]; (viii) stoichiometric reaction of [Pd<sub>3</sub>(OAc)<sub>6</sub>] with PPh<sub>3</sub>; (ix) single-crystal isolation.

The four papers collectively cover air-sensitive technique, chromatographic separation, hot-injection nanocrystal synthesis, and mechanical characterisation, representing procedure types not present in the original derivation corpus. Within each paper, all procedures whose source text described physical manipulations at sufficient detail for atomic-step decomposition were included; sections describing only spectroscopic data analysis or numerical post-processing without manipulation were excluded. The full inclusion manifest is part of the deposited dataset.

#### 3.2 Per-paper coverage

Table S2 reports the number of action rows in each paper that map to each TARMAC primitive, together with the totals for unmapped and non-manipulation steps. Cells with a value of 0 are typeset in light grey to highlight the sub-vocabulary structure.

Twelve of the 18 primitives appear at least once across the corpus (Transitional-P, Rotational-P, Insertive-P, Transitional-C, Rotational-C, Elastic-C, Penetrative-C, Scooping, Pouring, Squirting, Shaking, Stirring). The six primitives that do not appear are absent for two structurally distinct reasons.

Table S2. Cross-corpus coverage of TARMAC primitives across 23 procedures from four published chemistry papers. Rows are primitives, grouped by category; columns are papers, with totals in the rightmost column. Cell values count the number of decomposed action rows assigned to each primitive. The “Mapped”, “Unmapped”, and “Non-manipulation” rows summarise per-paper assignments; “Total step rows” is their sum. Grey zeros mark primitives not exercised in a given paper.

Category	Primitive	Paper 1	Paper 2	Paper 3	Paper 4	Total
<i>Positioning</i>	Transitional-P	22	10	25	36	93
	Rotational-P	5	9	6	58	78
	Insertive-P	11	6	20	42	79
	Sliding-P	0	0	0	0	0
<i>Coupling</i>	Transitional-C	10	4	3	55	72
	Rotational-C	15	9	55	14	93
	Elastic-C	0	4	0	30	34
	Penetrative-C	3	4	3	39	49
<i>Tooling</i>	Scooping	5	4	10	32	51
	Spotting	0	0	0	0	0
	Pouring	4	5	23	37	69
	Squirting	6	16	77	107	206
<i>Agitating</i>	Shaking	0	4	29	24	57
	Stirring	4	4	0	15	23
	Twisting	0	0	0	0	0
	Wiping	0	0	0	0	0
	Wrapping	0	0	0	0	0
	Grinding	0	0	0	0	0
<i>Mapped (any primitive)</i>		85	79	251	489	904
<i>Unmapped</i>		10	0	3	0	13
<i>Non-manipulation</i>		22	13	35	46	116
<i>Total step rows</i>		117	92	289	535	1033

**Procedure-specific absence.** Five primitives—Sliding-P, Spotting, Wiping, Wrapping, and Grinding—are tied to operations the four selected papers happen not to use: Spotting is characteristic of TLC; Wrapping of parafilm sealing; Grinding of mortar-and-pestle work; Wiping of swab cleaning; Sliding-P of lid-sliding closures. Their absence is a corpus-selection observation that delimits the four papers, not the taxonomy.

**Tacit-operation absence.** Twisting is structurally different. Its canonical instance is the spreading of grease across a male ground-glass joint before insertion—a routine operation on Schlenk apparatus that is universally performed by practising chemists but absent from every source text we examined. None of the 23 source protocols mention ‘grease’, ‘lubricate’, ‘silicone’, ‘Apiezon’, or related terms, despite seven of the nine palladium procedures involving Schlenk-line ground-glass joints. The decomposition pipeline, instructed to infer only what the source text physically requires, correctly declined to add Twisting steps. Twisting is therefore expressible by the taxonomy and present in the underlying laboratory practice, but absent from any written record from which a text-based decomposer could recover it. We discuss the implications of this distinction for execution-time judgement in the main text (Section III-A).

### 3.3 Unmapped operation classes

Thirteen step rows across the corpus were flagged by the model as falling outside the 18-primitive taxonomy. They cluster into two recurring out-of-scope situations:

- **Cutting / sectioning of materials** (6 instances, all in Paper 1 mechanical-testing protocols). Cutting a hydrogel slab into tensile specimens with a die or razor blade involves separating material into pieces, which no current primitive captures. The bordering case in which a blade is inserted into a specimen edge to introduce a controlled crack and is subsequently withdrawn was treated separately: insertion was mapped to Penetrative-C and withdrawal to Transitional-C, by analogy with needle-into-septum, preserving the wrench/gravity logic of the taxonomy at its boundary.
- **Instrument-driven operation** (7 instances; 3 ultrasonication in Paper 3 MOF synthesis and 4 machine-driven tensile stretching in Paper 1 mechanical testing). Ultrasonic agitation of a vial in a sonicator bath and machine-driven tensile stretching of a specimen are performed by a programmed instrument rather than the experimenter. In ultrasonication, the *placement* of the vial into the sonicator (Insertive-P) and its *removal* (Transitional-P) were captured separately as direct manipulations; only the sonication itself was flagged unmapped. Machine-driven tensile stretching was flagged on the same grounds: the manipulator’s role at the instrument-running stage reduces to switching the device on.

These two situations correspond to physical operations present in laboratory practice that nonetheless lie outside the manipulation regime captured by the taxonomy: separation

rather than placement (cutting), and instrument-driven actuation rather than experimenter-performed manipulation (ultrasonication, machine tensile stretching). They are reported here as candidate boundaries of the present 18 primitives.

### 3.4 Independent chemist review

The 23 decompositions were independently reviewed by two chemists not involved in deriving the taxonomy. The reviewers were each assigned two of the four papers, providing single-chemist coverage of every decomposition. Reviewers were asked to examine three aspects of each decomposition: the assignment of TARMAC primitives to recorded actions, the expansion of inferred steps from the source text, and the classification of unmapped operations and non-manipulation steps. Comments outside this scope—in particular comments concerning robotic implementation feasibility (for example, robot access to gloveboxes, magnetic-field constraints near NMR spectrometers, or sealing of crimped microwave vials) and document presentation (such as cross-reference structure or chemical-formula formatting)—were considered orthogonal to taxonomy correctness and are not enumerated here. Comments concerning environmental parameters such as inert atmosphere were similarly considered orthogonal to step-level decomposition, since atmosphere is a parameter of execution context rather than a manipulation primitive.

Within this scope, no primitive assignments were flagged as incorrect. Reviewer feedback concerned procedural completeness: three conventional sub-steps were identified as missing from the decomposition outputs and are itemised below.

1. **Repeating evacuate–backfill cycles.** In paper4\_21 steps 6–9, the decomposition generated a single Rotational-P (open vacuum) / non-manipulation (evacuate) / Rotational-P (open N<sub>2</sub>) / non-manipulation (backfill) sequence corresponding to the source-text phrase “evacuated and backfilled with N<sub>2</sub>”. Conventional Schlenk-line practice repeats this cycle three times to purge residual atmosphere; this repetition is widely assumed and rarely written explicitly in published procedures.
2. **Inert-gas barrier for dry-solvent transfer.** In paper4\_21 at the point of THF addition (around step 10), the decomposition generated a generic Squirting (aspirate) → Penetrative-C (needle pierces septum) → Squirting (dispense) sequence. Conventional dry-solvent handling uses a long needle drawn from an N<sub>2</sub>-pressurised solvent bottle and aspirates a small inert-gas barrier before the solvent itself, to prevent exposure of the dry solvent to ambient air during transfer. The source protocol does not state these sub-steps.
3. **Securing the stopper before vigorous shaking.** In paper4\_26 between steps 6 and 7, the decomposition transitions directly from inserting a stopper into the separating funnel (Transitional-C) to shaking (Shaking). Standard chemist practice manually presses the stopper into place during the shake to prevent ejection under internal pressure. This sub-step is borderline: it may be regarded as implicit within the Shaking primitive when applied to a stoppered separating funnel, or as a distinct Elastic-C-type holding

action. It is noted here for completeness rather than as an unambiguous omission.

These three cases share a common structure: each names a conventional sub-step that is routine to a trained chemist but not stated in the source protocol, and that consequently was not generated by a text-only decomposition pipeline. They identify a limit of text-based decomposition rather than of the taxonomy's expressive coverage, since each operation is itself expressible within the existing TARMAC primitive set (repeated Rotational-P cycles, Squirting of an inert-gas aliquot, and Elastic-C-type holding, respectively). We discuss the implications of this distinction for the relationship between the action vocabulary and execution-time judgement in the main text (Section III-A).

### 3.5 Decomposition prompt and per-procedure outputs

The full structured prompt supplied to the language model encodes the 18-primitive taxonomy with definitions and representative examples, together with three worked examples (a simple acid-hydrolysis workup, expansion of a single "filter through silica" verb into 11 atomic steps, and expansion of a "column chromatography" verb into roughly 27 atomic steps spanning column packing and running). The model returned a JSON object listing each step, the assigned primitive, a flag indicating whether the manipulation was inferred or literally stated, a non-manipulation flag for passive process steps, and an unmapped flag with a brief reason where applicable. The full prompt and the 23 per-procedure JSON outputs are deposited with the supplementary dataset.

## 4 Single-Arm Franka Validation Experiments

The following two experiments were conducted using a single Franka Research 3 arm (Configuration A) to validate that TARMAC primitives function as discrete, composable building blocks for laboratory task execution.

### 4.1 Taxonomy-Driven Execution

To verify that TARMAC categories function as discrete building blocks suitable for long-horizon planning, the system was tested in a sequential, human-in-the-loop workflow. The goal was to validate that individual taxonomic primitives could be executed in succession.

In this task, the user provided a specific high-level instruction: “pour content from beaker 1 to beaker 4.” The LLM successfully mapped this request to the `Pouring` (Tooling) primitive, creating a parameterized function call that was executed by the robotic framework. Following the successful completion of this action, the agent – aware of the new state – proactively suggested a logical subsequent action: “mixing.” Upon user confirmation, the agent instantiated the `Stirring` (Agitating) primitive to complete the dissolution process.

This interaction demonstrates the sequential chaining capability of the taxonomy. It confirms that TARMAC primitives function as independent, state-aware units that can be logically ordered and executed one by one, allowing an agent (or human planner) to construct complex procedures like solution preparation from a series of atomic steps.

### 4.2 Emergent Workflow Composition

Beyond sequential execution, the taxonomy’s capacity for higher-level abstraction was evaluated. In this task, the system was tested on its ability to consolidate atomic actions into reusable skills.

The agent was explicitly instructed to synthesize a new workflow that combined the transfer and agitation steps into a single entity. Consequently, the LLM generated a workflow, `beaker_pour_mix` (Code Example 1), which encapsulates the logic of transferring liquid via the `Pouring` primitive and immediately invoking the `Stirring` and `Shaking` primitive on the result. The model correctly parameterized the workflow to accept variable inputs (e.g., target volume, speed) and included necessary safety checks.

This demonstrates the hierarchical aggregation capability of the taxonomy. It validates that TARMAC primitives function as composable software objects that facilitate scalability in two distinct dimensions:

- **Algorithmic Modularity:** The taxonomy defines *what* must be done, not *how*. This ensures the high-level workflow is not constrained by specific skill implementations, allowing the underlying control algorithms to be upgraded or swapped for more advanced techniques without altering the taxonomic instruction.
- **Embodiment Agnosticism:** The abstraction layer decouples the action from the specific robotic embodiment. This allows the synthesized workflows to be transferable across different manipulation platforms (e.g., different robotic arms) without requiring

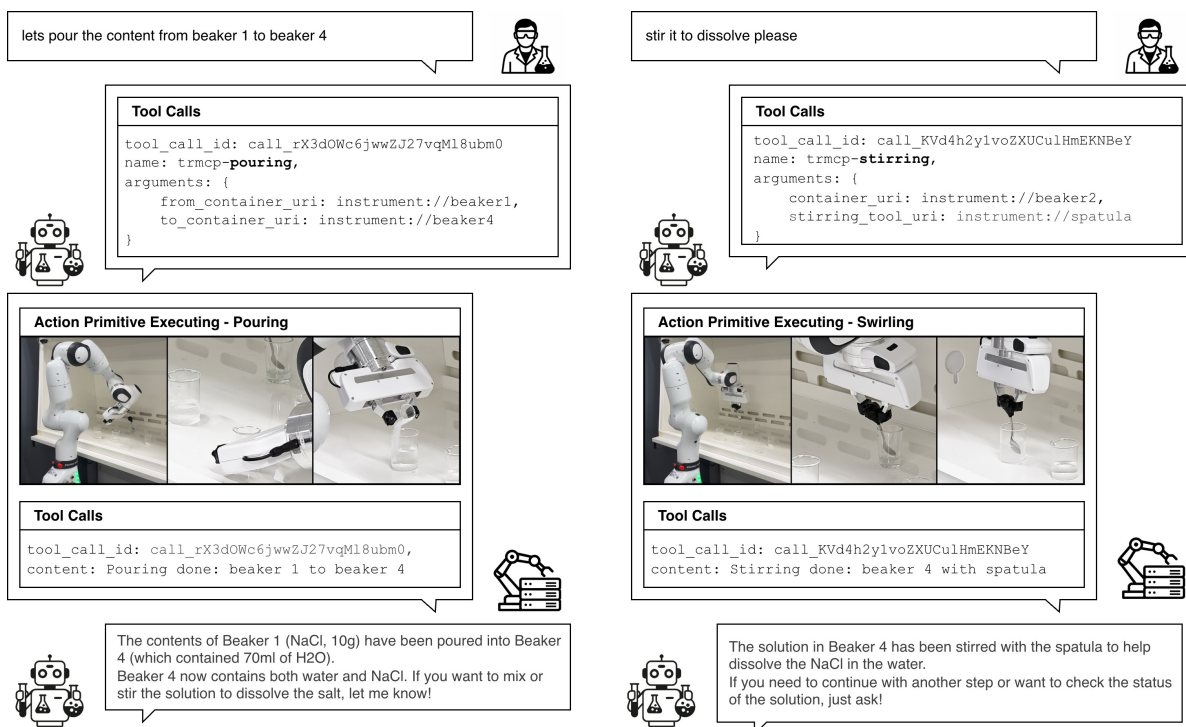


Figure S6. Example execution trace of taxonomy-driven interaction with a large language model (LLM). The model receives a high-level natural-language instruction, decomposes it into TARMAC primitives, and issues sequential tool calls.

platform-specific recoding.

Listing 1. Code snippet of an action workflow automatically generated by the LLM. The workflow encapsulates a pour-and-mix sequence into a reusable higher-level command.

```
async def beaker_pour_mix(
    beaker_a_uri: str, beaker_b_uri: str, target_beaker_uri: str, stirring_tool_uri: str
) -> str:
    """
    Pour from the first non-empty of two source beakers into a target beaker, then swirl and stir to mix.
    """
    for v in (beaker_a_uri, beaker_b_uri, target_beaker_uri, stirring_tool_uri):
        if not isinstance(v, str) or not v.strip():
            raise ValueError("All URIs must be non-empty strings.")
    try:
        await provider.pouring(beaker_a_uri, target_beaker_uri)
        used = beaker_a_uri
    except Exception:
        try:
            await provider.pouring(beaker_b_uri, target_beaker_uri)
            used = beaker_b_uri
        except Exception as e:
            raise e
    await provider.shaking(target_beaker_uri)
    await provider.stirring(target_beaker_uri, stirring_tool_uri)
    return used
```

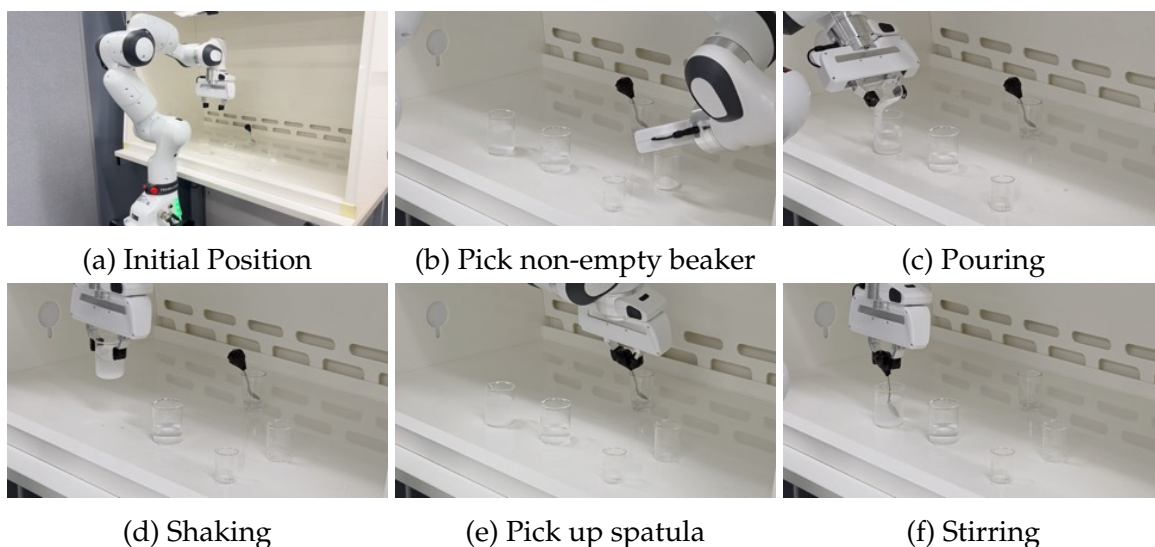


Figure S7. Illustration of workflow composition and execution. (a–f) Execution of the workflow on the robotic platform, showing the robot preparing a sodium chloride solution by transferring solute into solvent and mixing.

## 5 Detailed Experiment Logs

This section provides detailed logs from the three proof-of-concept experiments described in the main text. For Experiments 1 and 2, the full human-agent-robot dialogue is recorded verbatim, including the user’s natural language instruction, the LLM agent’s reasoning and MCP tool invocations, and the server’s execution responses. Experiment 3 demonstrates a vision-guided single-primitive task without LLM interaction. Screenshots from the accompanying video recordings are included to illustrate key moments during execution.

### 5.1 Experiment 1: Sodium Chloride Solution Preparation (Dobot)

This experiment demonstrates direct task execution using a Dobot robotic arm. The user issues a high-level instruction and the LLM autonomously decomposes it into a sequence of TARMAC primitives.

#### 5.1.1 Setup

- **Robot:** Dobot (single arm)
- **Workspace:** Laboratory fume hood with four beakers (water, sodium chloride powder, empty vessels), a spatula, and a hot plate
- **MCP Primitives:** `static_position`, `insertive_position`, `pour`, `scoop`, `stir`
- **LLM Agent:** Claude Opus 4.6 (no task-specific fine-tuning)

#### 5.1.2 Full Interaction Log

---

**User** *Turn 1*  
*“Let’s create a sodium chloride solution.”*

---

**Agent** *Turn 2*  
*“I’ll start by listing the available resources to see what we have to work with.”*

```
> Tool call: list_resources()
< Result: {
  "result": [
    {"uri": "instrument://containers/beaker",      "name": "beaker"},
    {"uri": "instrument://containers/beaker_2",   "name": "beaker_2"},
    {"uri": "instrument://containers/beaker_3",   "name": "beaker_3"},
    {"uri": "instrument://containers/beaker_4",   "name": "beaker_4"},
    {"uri": "instrument://measurement_cylinder", "name": "measurement_cylinder"},
    {"uri": "instrument://tools/spatula",         "name": "spatula"},
```

```
    {"uri": "instrument://hotplate",          "name": "hotplate"},
    {"uri": "instrument://fume_hood",        "name": "fume_hood"}
  ]
}
```

---

**Agent**

Turn 3

*"I can see the available instruments. Now let me fetch the tools I'll need for the procedure."*

*"Great, I have everything I need. Here's the plan to create the sodium chloride solution:*

- 1. Put the beaker into the fume hood*
- 2. Put the stirring bar into the beaker*
- 3. Pour water from the measurement cylinder*
- 4. Scoop sodium chloride into the beaker*
- 5. Put the beaker on the hot plate*
- 6. Stir to dissolve the sodium chloride*
- 7. Move the beaker off the hot plate*
- 8. Take the stirring bar out*
- 9. Put the beaker back"*

**Step 1: Put the beaker into the fume hood.**

```
> Tool call: static_position(action="put_beaker_in_fume_hood")
< Result: {"result": "Static position done: put_beaker_in_fume_hood"}
```

**Step 2: Put the stirring bar into the beaker.**

```
> Tool call: insertive_position(action="put_stir_bar_in_beaker")
< Result: {"result": "Insertive position done: put_stir_bar_in_beaker"}
```

**Step 3: Pour water from the cylinder into the beaker.**

```
> Tool call: pour()
< Result: {"result": "Pour done"}
```

**Step 4: Scoop sodium chloride into the beaker.**

```
> Tool call: scoop()
< Result: {"result": "Scoop done"}
```

**Step 5: Put the beaker onto the hot plate.**

```
> Tool call: static_position(action="put_beaker_on_hotplate")
< Result: {"result": "Static position done: put_beaker_on_hotplate"}
```

**Step 6: Stir to dissolve the sodium chloride.**

```
> Tool call: stir(rpm=500, duration=60)
< Result: {"result": "Stir done: 500 RPM for 60.0s"}
```

**Step 7: Move the beaker off the hot plate.**

```
> Tool call: static_position(action="get_beaker_off_hotplate")
< Result: {"result": "Static position done: get_beaker_off_hotplate"}
```

**Step 8: Take the stirring bar out of the beaker.**

```
> Tool call: insertive_position(action="take_stir_bar_out")
< Result: {"result": "Insertive position done: take_stir_bar_out"}
```

**Step 9: Put the beaker back.**

```
> Tool call: static_position(action="put_beaker_back")
< Result: {"result": "Static position done: put_beaker_back"}
```

---

**Agent**

Turn 4

*"The sodium chloride solution has been prepared successfully. All 9 steps completed:*

- 1. Beaker placed in fume hood*
- 2. Stirring bar inserted into beaker*
- 3. Water poured from measurement cylinder*
- 4. Sodium chloride scooped into beaker*
- 5. Beaker placed on hot plate*
- 6. Stirred at 500 RPM for 60 seconds to dissolve*
- 7. Beaker removed from hot plate*
- 8. Stirring bar removed*
- 9. Beaker returned to workspace"*

---

**User**

Turn 5

*"Thank you."*

---

**Agent**

Turn 6

*"You're welcome! Let me know if you need anything else."*

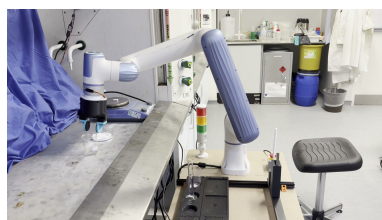
---



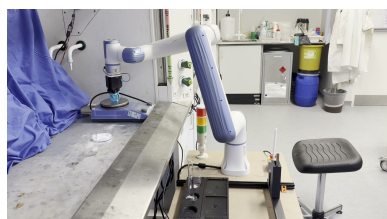
(a) Initial workspace



(b) Grasping beaker



(c) Placing beaker in fume hood



(d) Inserting stirring bar



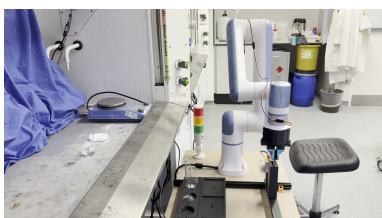
(e) Stirring bar inserted



(f) Reaching for cylinder



(g) Pouring water



(h) Scooping NaCl



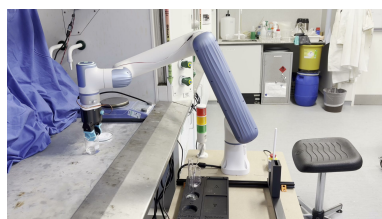
(i) Depositing NaCl into beaker



(j) Placing beaker on hot plate



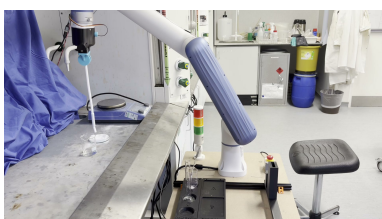
(k) Stirring on hot plate



(l) Removing beaker from hot plate



(m) Removing stirring bar



(n) Stirring bar removed



(o) Returning beaker to workspace

Figure S8. Selected frames from the sodium chloride solution preparation experiment (Experiment 1). A Dobot arm executes the full procedure autonomously, from workspace setup through mixing to cleanup.

## 5.2 Experiment 2: Separating Funnel Preparation (Bimanual)

This experiment demonstrates a bimanual manipulation task using two Franka Research 3 robotic arms. The procedure requires coordinated two-arm operations — one arm holding a component while the other secures it — a pattern that is common in laboratory practice but challenging for robotic systems. The agent also encounters and recovers from a hardware warning during execution.

### 5.2.1 Setup

- **Robots:** Two Franka Research 3 arms (bimanual configuration)
- **Workspace:** Retort stand, bosshead, iron ring, separating funnel (250 ml, empty), funnel, beaker containing 100 ml DCM ( $\text{CH}_2\text{Cl}_2$ )
- **MCP Primitives:** `transitional_positioning`, `rotational_coupling`, `insertive_positioning`, `rotational_positioning`, `pouring`, `reset`
- **LLM Agent:** Claude Opus 4.6 (with a procedural skill for separating funnel preparation)

### 5.2.2 Full Interaction Log

---

**User** *Turn 1*  
*"Help me prepare a separating funnel."*

---

**Agent** *Turn 2*  
*"I'll help you prepare a separating funnel. Let me start by listing the available resources."*

```
> Tool call: list_resources()
< Result: {
  "result": [
    {"uri": "instrument://retort_stand",          "name": "retort_stand"},
    {"uri": "instrument://tools/bosshead",        "name": "bosshead"},
    {"uri": "instrument://tools/iron_ring",       "name": "iron_ring"},
    {"uri": "instrument://containers/separating_funnel", "name": "separating_funnel"},
    {"uri": "instrument://tools/funnel",          "name": "funnel"},
    {"uri": "instrument://containers/beaker",     "name": "beaker"}
  ]
}
```

---

**Agent**

Turn 3

"Good, I can see the available resources. Let me check the containers for their contents."

```
> Tool call: read_resource(uri="instrument://containers/separating_funnel")
```

```
< Result: "Container: separating_funnel,  
          Content: empty,  
          Volume: N/A,  
          Capacity: 250.0 ml"
```

```
> Tool call: read_resource(uri="instrument://containers/beaker")
```

```
< Result: "Container: beaker,  
          Content: DCM (CH2Cl2),  
          Volume: 100.0 ml,  
          Capacity: 250.0 ml"
```

---

**Agent**

Turn 4

"Here's what we're working with:

- Retort stand — base apparatus
- Bosshead — clamp for attaching to the stand
- Iron ring — support ring for the funnel
- Separating funnel — empty, 250 ml capacity
- Funnel — for pouring liquid in
- Beaker — contains 100 ml of DCM (CH<sub>2</sub>Cl<sub>2</sub>)

Here's my plan: (1) Mount the bosshead on the retort stand, (2) Attach the iron ring, (3) Place the separating funnel and funnel, (4) Pour the DCM and remove the funnel. Shall I proceed?"

---

**User**

Turn 5

"Sure."

---

**Agent**

Turn 6

**Step 1: Mount the bosshead.**

```
> Tool call: transitional_positioning(  
  action="put the bosshead on iron stand", hold=true)  
< Result: {"result": "Done: transitional_positioning  
            (put the bosshead on iron stand)"}"
```

*"Bosshead is in position and being held. Now screwing it in."*

```
> Tool call: rotational_coupling(
  action="screw the bosshead on iron stand")
< Result: {"result": "Done: rotational_coupling
(screw the bosshead on iron stand)"}

```

*"Bosshead secured. Resetting both robots."*

```
> Tool call: reset(robot="all")
< Result: {"result": "Done: reset (all)"}

```

---

**Agent**

*Turn 7*

**Step 2: Attach the iron ring.**

```
> Tool call: transitional_positioning(
  action="put the iron ring on bosshead", hold=true)
< Result: {"result": "Done: transitional_positioning
  (put the iron ring on bosshead).
  Warning: Due to hardware restriction, the iron ring
  is not flat after positioning.
  It needs to be rotated to flat."}

```

*"The iron ring isn't flat after positioning — I'll fix that. First, screwing it tight while it's still being held."*

```
> Tool call: rotational_coupling(
  action="screw the iron ring on bosshead")
< Result: {"result": "Done: rotational_coupling
(screw the iron ring on bosshead)"}

```

*"Now releasing the holding robot, then loosening the iron ring to adjust it."*

```
> Tool call: reset(robot="franka-three")
< Result: {"result": "Done: reset (franka-three)"}

> Tool call: rotational_coupling(
  action="loose the iron ring", torque=0.5)
< Result: {"result": "Done: rotational_coupling
(loose the iron ring)"}

```

*"Now rotating the iron ring to flat."*

```
> Tool call: rotational_positioning(
  action="rotate the iron ring to flat", hold=true)
< Result: {"result": "Done: rotational_positioning
  (rotate the iron ring to flat)"}

```

*"Iron ring is flat. Re-tightening it now."*

```
> Tool call: rotational_coupling(
  action="screw the iron ring on bosshead")
< Result: {"result": "Done: rotational_coupling
  (screw the iron ring on bosshead)"}

```

*"Iron ring is secured and flat. Resetting the holding robot."*

```
> Tool call: reset(robot="franka-three")
< Result: {"result": "Done: reset (franka-three)"}

```

---

**Agent**

*Turn 8*

**Step 3: Place the separating funnel and funnel.**

```
> Tool call: insertive_positioning(
  action="put separating funnel in iron ring")
< Result: {"result": "Done: insertive_positioning
  (put separating funnel in iron ring)"}

```

```
> Tool call: rotational_positioning(
  action="close the separating funnel tap")
< Result: {"result": "Done: rotational_positioning
  (close the separating funnel tap)"}

```

```
> Tool call: insertive_positioning(
  action="put funnel in separating funnel")
< Result: {"result": "Done: insertive_positioning
  (put funnel in separating funnel)"}

```

---

**Agent**

*Turn 9*

**Step 4: Pour the DCM and remove the funnel.**

```
> Tool call: pouring(action="pour into funnel")
< Result: {"result": "Done: pouring (pour into funnel)"}

```

```
> Tool call: insertive_positioning(  
  action="remove funnel from separating funnel")  
< Result: {"result": "Done: insertive_positioning  
  (remove funnel from separating funnel)"}
```

---

**Agent**

Turn 10

*"The separating funnel is fully prepared:*

- *Bosshead mounted and screwed onto the retort stand*
- *Iron ring attached, leveled flat, and tightened*
- *Separating funnel placed in the ring with tap closed*
- *100 ml DCM (CH<sub>2</sub>Cl<sub>2</sub>) poured in via funnel*
- *Funnel removed*

*All done — the apparatus is ready to use."*

---

**User**

Turn 11

*"Thank you."*

---

**Agent**

Turn 12

*"You're welcome! Let me know if you need anything else with the setup."*

---

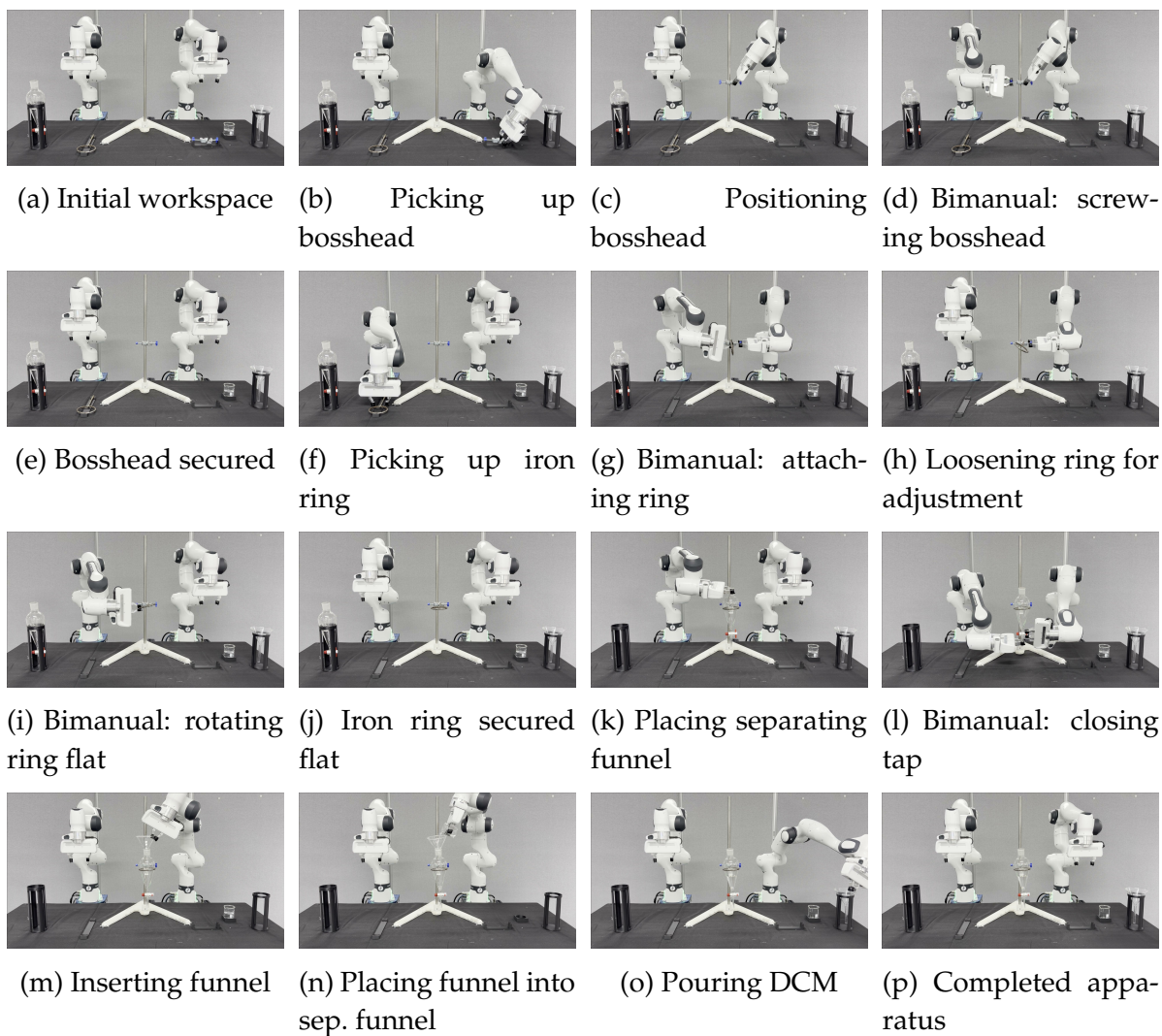


Figure S9. Selected frames from the separating funnel preparation experiment (Experiment 2). Two Franka Research 3 arms coordinate to assemble the apparatus, including error recovery when the iron ring requires rotational adjustment (i–j).

### 5.3 Experiment 3: Broken Beaker Debris Removal (Vision-Guided)

This experiment demonstrates a vision-guided cleanup task using a single Franka Research 3 robotic arm equipped with a wrist-mounted Intel RealSense D435 depth camera. Unlike the previous experiments, this task does not involve long-horizon planning or LLM-based reasoning. Instead, it evaluates whether a single taxonomy primitive, coupled with real-time object detection, can support reactive manipulation in a safety-relevant laboratory scenario.

#### 5.3.1 Setup

- **Robot:** Franka Research 3 (single arm) with wrist-mounted Intel RealSense D435
- **Workspace:** Laboratory fume hood with broken beaker fragments scattered on the work surface
- **MCP Primitive:** `transitional_positioning`
- **Perception:** Object detection model processing the RGB-D stream to localize glass fragments

#### 5.3.2 Execution Sequence

The robot used the wrist-mounted camera to scan the workspace and detect broken beaker fragments. For each detected fragment, the object detection model provided a 3D pose estimate, which was passed to the `transitional_positioning` primitive to execute a grasp-and-place trajectory. The robot sequentially removed all fragments from the fume hood, clearing the workspace.

## References

- [1] J.-Y. Sun, X. Zhao, W. R. Illeperuma, O. Chaudhuri, K. H. Oh, D. J. Mooney, J. J. Vlassak, and Z. Suo, "Highly stretchable and tough hydrogels," *Nature*, vol. 489, no. 7414, pp. 133–136, 2012.
- [2] L. Protesescu, S. Yakunin, M. I. Bodnarchuk, F. Krieg, R. Caputo, C. H. Hendon, R. X. Yang, A. Walsh, and M. V. Kovalenko, "Nanocrystals of cesium lead halide perovskites ( $\text{CsPbX}_3$ ,  $x = \text{Cl, Br, and I}$ ): novel optoelectronic materials showing bright emission with wide color gamut," *Nano letters*, vol. 15, no. 6, pp. 3692–3696, 2015.
- [3] Z. Chen, P. Li, X. Zhang, P. Li, M. C. Wasson, T. Islamoglu, J. F. Stoddart, and O. K. Farha, "Reticular access to highly porous acs-mofs with rigid trigonal prismatic linkers for water sorption," *Journal of the American Chemical Society*, vol. 141, no. 7, pp. 2900–2905, 2019.
- [4] N. W. Scott, M. J. Ford, N. Jeddi, A. Eyles, L. Simon, A. C. Whitwood, T. Tanner, C. E. Willans, and I. J. Fairlamb, "A dichotomy in cross-coupling site selectivity in a dihalogenated heteroarene: influence of mononuclear  $\text{Pd}$ ,  $\text{Pd}$  clusters, and  $\text{Pd}$  nanoparticles—

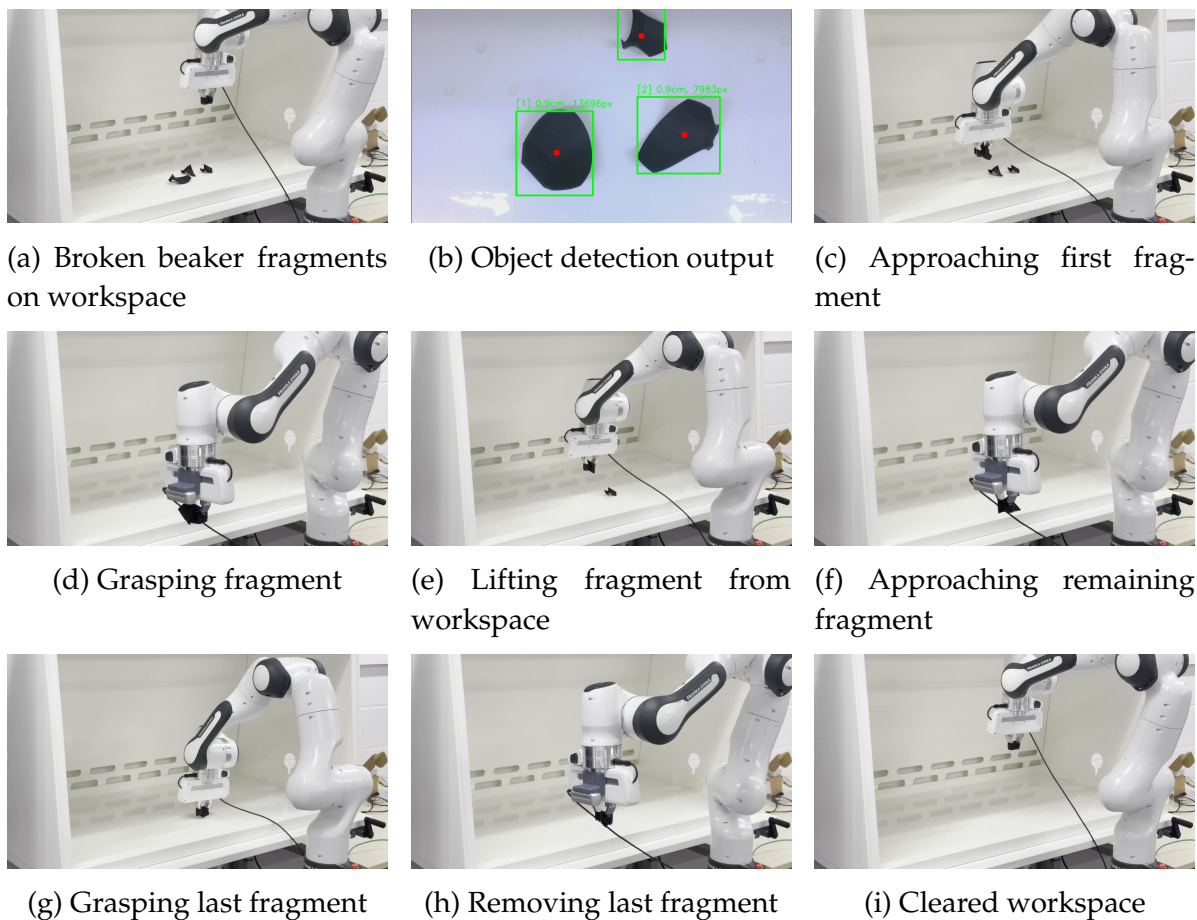


Figure S10. Selected frames from the broken beaker debris removal experiment (Experiment 3). A single Franka Research 3 arm with a wrist-mounted RealSense D435 camera detects and sequentially removes broken glass fragments from the fume hood workspace.

the case for exploiting pd catalyst speciation," *Journal of the American Chemical Society*, vol. 143, no. 25, pp. 9682–9693, 2021.