WorldLoop

Sample report

Putting an earphone in its charger in the Living Room

Eval report | 1 task | 1 place | Living Room

This is the report a lab gets after the eval runs. It shows the result, every failure with a clip in a full eval, how the result was judged, and the world files.

The result

62 / 100

Your policy passed 62 of 100 runs of putting the earphone in its charger.

A pass means the earphone ended in the dock, upright, inside the limits. A fail means it did not.

Reached the earphone: 98

Grasped the earphone: 81

Lifted the earphone: 70

Placed in the dock: 62

Passed: 62

Failed: 38

Every run starts different: earphone position, tilt, and clutter in the room. That is what the eval varies.

Failures

38 runs failed, in 4 ways. Each row names what broke and shows the clip in a full eval.

1. Missed grasp - 14 runs

The gripper closed 2 cm short of the earphone edge, on the smooth side.

Try next: move the approach 2 cm lower and slow the last 10 cm.

Clip: not recorded in this demo. Still frame of the world, not a clip of this failure.

2. Slipped on lift - 11 runs

The earphone left the table, then slipped as the arm rotated.

Try next: add 0.4 s of settle time before the lift.

Clip: not recorded in this demo. Still frame of the world, not a clip of this failure.

3. Timed out - 7 runs

The policy spent 60 seconds searching for an earphone it had already passed.

Try next: cap the search sweep at 20 seconds.

Clip: not recorded in this demo. Still frame of the world, not a clip of this failure.

4. Placed outside the dock - 6 runs

The drop point drifted 4 cm past the dock edge on hard variations.

Try next: re-check the dock center after each variation.

Clip: not recorded in this demo. Still frame of the world, not a clip of this failure.

The judge

After each run, the judge checks the earphone against the dock: sideways, front to back, height, and tilt. The dock target is x 0.620 m, y 0.080 m, z 0.808 m.

Good placement: PASS

Sideways: off 0.000 m, limit 0.020 m, PASS

Front to back: off 0.000 m, limit 0.010 m, PASS

Height: off 0.000 m, limit 0.010 m, PASS

Tilt: off 0.0 deg, limit 15.0 deg, PASS

Verdict: PASS, all 4 checks inside their limits.

Bad placement: FAIL

Sideways: off 0.055 m, limit 0.020 m, FAIL

Front to back: off 0.000 m, limit 0.010 m, PASS

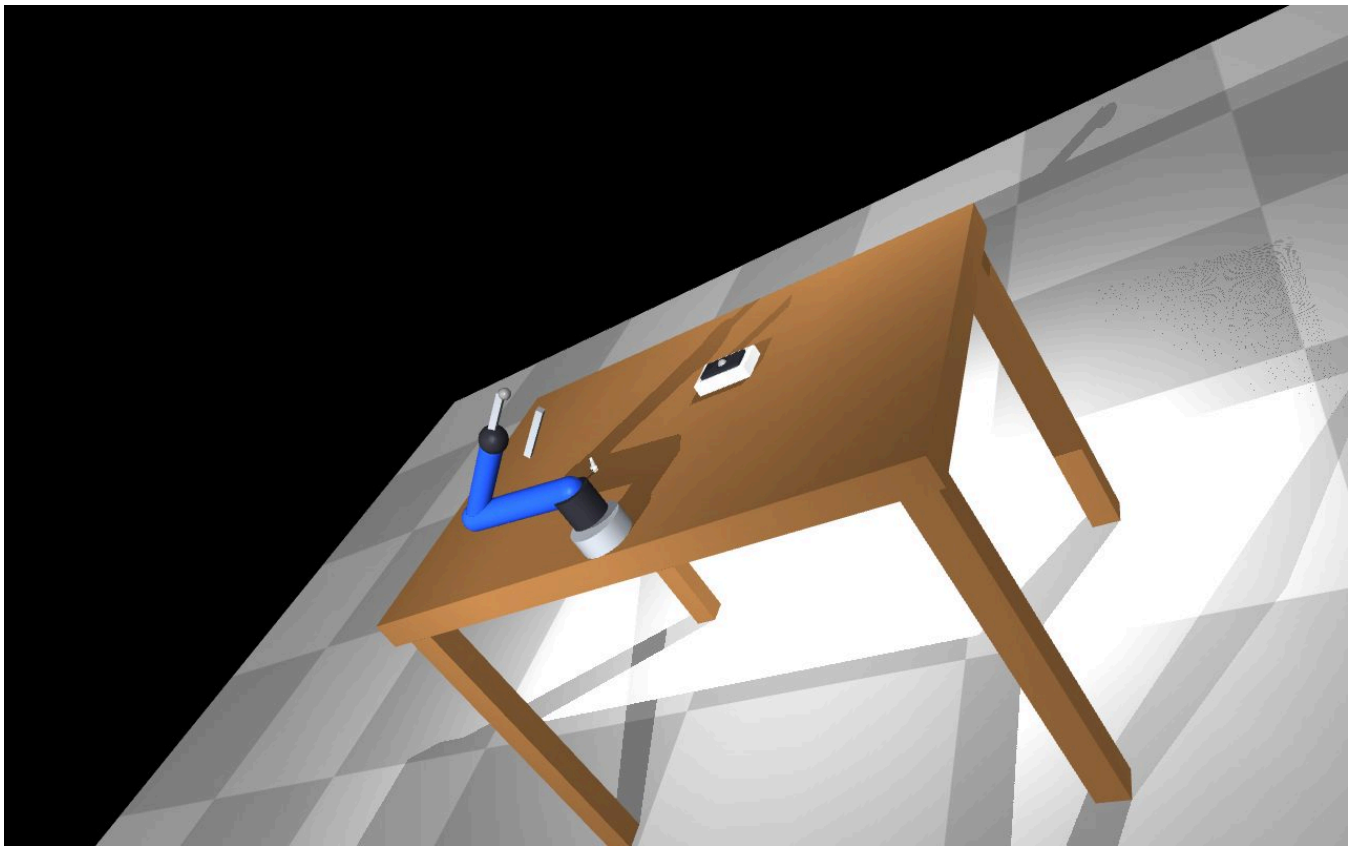
Height: off 0.002 m, limit 0.010 m, PASS

Tilt: off 40.0 deg, limit 15.0 deg, FAIL

Verdict: FAIL, 2 of 4 checks outside their limits.

The judge works: it passes good placement and fails bad placement. Run it yourself with the files below.

The world



The world. Table, charger dock, earphone, robot arm. Open it in MuJoCo with the world.xml file below.

Files in this pack

report.pdf - This report as a PDF
world.xml - The world, 6.7 KB
world-preview.png - The world, one frame, 161.5 KB
success-test.py - The judge, 6.1 KB
forced-pass.txt - Good placement result, 0.6 KB
forced-fail.txt - Bad placement result, 0.6 KB

How the eval runs

1. The sim loads the world. MuJoCo opens world.xml: the room, the table, the dock, the earphone, the arm.

2. The robot runs the task, 100 times, in a full eval. Each run starts different: earphone position, tilt, clutter in the room.

3. The judge checks the result. After each run, success-test.py checks the earphone against the dock.

4. The report counts it up. Pass rate, failures by type, and a clip behind each failure. In a full eval, that report is this page.

What happens next

Top failure: Missed grasp, 14 runs. The gripper closed 2 cm short of the earphone edge, on the smooth side.

Next test to run: the earphone starts 2 cm farther from the gripper, 20 runs. The gripper approaches 2 cm lower and slows in the last 10 cm.

Honest note. This is a sample report. No 100-run eval has run yet, so the score (62 of 100) and the failure counts (14, 11, 7, 6) are illustrative, taken from the sample in the upload flow. The judge results are real: good placement passes 4 of 4 and bad placement fails 2 of 4, run by hand in MuJoCo 4.0.1 and plain Python. The clips will be real recordings once the eval runner is built.