

Training updates work. Learning improvement is still unproven.

Reusable research briefing | October 16, 2026

Last independently measured benchmark: 2/12, unchanged. No new 12-item score has been established. The Qwen comparison remains untested.

New measured evidence

Completed no-update control	18 objective evaluations and two generated responses; zero updates and unchanged parameters. Original evidence, cleanup and independent review closed.
New correct-label attempt*	One update changed 8,837 of 11,914 parameter values. The saved state remained finite, with its update counter advancing from 10,720 to 10,721.
What the update optimized*	Total objective fell from 7.80744 to 7.50161 . Correct-answer confidence fell slightly, from 99.91045% to 99.90639% . The answer stayed "a"; it was already correct before teaching.

*These are independently checked saved observations from a run whose overall runtime qualification failed. They are not a passing learning benchmark, new-association result or retention demonstration.

Why this matters for the 2/12 problem

The training path can apply a real update and save changed parameters. However, reducing the combined objective did not improve the correct-answer probability in this local example. The earlier completed control found that source/version and end-marker targets dominated the loss and could compete with the answer target.

This is a concrete repair hypothesis, not a proven explanation of the historical 2/12 score. The planned controls must be completed under a qualified owner before separately testing answer-versus-format weighting. Novel association, restart retention and independent 12-item scoring remain necessary.

Why the newest run was not accepted

A controller process appeared in one job census and disappeared before the next. The observer lacked retained evidence proving the same original process had exited. It correctly stopped qualification. Recorded query-handle closes completed; process absence alone cannot establish identity-safe cleanup. The external cleanup inventory was empty.

The consumed test and all failed originals are preserved. It will not be replayed. No startup RAM, process-family, numerical-worker or time limit was raised.

A focused route to stronger, reusable evidence

Next measured milestones

1. Reliable evidence custody	Retain original process identity and liveness evidence across process exits. Keep unknown, reused or escaped generations blocked. Review and qualify the corrected owner before another model test.
2. Learning and comparison	Finish matched controls from the same starting state; test the specific loss hypothesis separately. Report a higher score only after independent scoring. Run Qwen under an explicitly comparable, qualified profile.
3. Body and robotics	Preserve the measured skin result and qualify the remaining organ/runtime connections. Validate the NewBrain action interface before claiming live ROS 2 or robot control.

Other verified foundations and present limits

Skin: a separate five-pose control moved 512 sampled weighted skin vertices and measured exact restoration. This establishes commanded geometry, not passive recoil, touch learning or realistic tissue forces.

Robotics: the existing typed ROS 2 bridge is implemented; its dated repository report records standalone software results. A new adapter passed source review, with 18 written test methods still unrun. A live NewBrain action producer and real ROS execution remain unqualified.

Lungs and Patient Horizon: realistic coupled lung-brain physiology, organ validation and a demonstration video remain unrun. Patient-specific anatomy and disease simulations require validated physiology and medical imaging methods; the language-learning decoder is not a validated medical brain model.

Why additional resources and collaborators help

The newest attempt passed the **8 GiB available-RAM startup guard**; its available-memory samples stayed above **12.31 GiB**. Its rejection was a process-evidence problem. More memory or a larger GPU alone will not correct that logic or a competing training objective.

Additional RAM, an appropriate GPU, managed laboratory compute and a home research server would support larger comparisons, longer simulations and work outside campus hours. Capacity should be chosen with university engineers using measured model and simulation requirements. Secure home-to-lab job submission and evidence transfer would allow repeatable tests on more capable equipment.

Potential collaborators span AI and computer science, robotics, biomedical engineering, physiology and medical imaging. Early patient demonstrations should use synthetic records while researchers establish biological validity, data authorization and clinical-review requirements.

Evidence: completed control M1-001; consumed, failed-runtime attempt M1-002; skin control BODY-SKIN-590-CONTROL-001. Detailed originals and independent reviews are retained. Robotics overview: kiralabs.org/companion-robotics.html.