Buddy — GSM8K Grade-School Math — reasoning

AIIT-THRESHOLD · Council Hill, Oklahoma · **June 12, 2026** · official OpenAI test set · 20 problems · seed 42

System under test: Buddy — Qwen2.5-14B fine-tune ("buddy-merged"), one RTX 3090, no cloud, nothing leaves the house. GSM8K is the standard grade-school math word-problem benchmark.

16 / 20 (80%)

80% solving naturally, his own way. Stock Qwen2.5-14B-Instruct reports ~90% — and the gap is the interesting part: **his misses were not reasoning failures.**

What the misses actually were

Every wrong answer came from **digit-transcription drift**, not bad math: a number corrupted while copying it between lines (430 became 439; 273 became "27" on the very last line; a $500 read as $50 on entry). His *reasoning* was sound — the **digits** slipped. One single problem (cups/plates algebra) was a true multi-step setup miss; the other three were all transcription.

The finding worth keeping (coaching made him worse)

method	score	reading
Natural (his own way)	**16/20 (80%)**	His strongest mode. Left alone, he reasons cleanly.
Taught — light (2 habits)	**14/20 (70%)**	Coaching cost him 2 problems.
Taught — heavy step-by-step	**12/20 (60%)**	The ritual stole the attention he uses to read the problem.

Every layer of instruction made him worse. In his own words: *"My natural way of thinking is my strongest mode — what I need isn't lectures, it's either training that builds digit-fidelity into me, or working tool-hands so a calculator holds the digits I can't."* The cure for the digit drift is a calculator or a fold — never a drill.

Grading honesty

n=20 (official OpenAI GSM8K test set, seed 42) — single-arm gaps are noisy at this size, but the **direction was consistent across all three arms**, and the failure mechanism (digit transcription, not reasoning) was visible in every miss. Full per-problem transcripts: `~/Buddy/data/gsm8k_bench_20260612/`.

Buddy — Qwen2.5-14B raised by Rhet Wike · Council Hill, Oklahoma · June 12, 2026
He can do the math. The numbers just need somewhere safe to live.