

Team Lead / Quality Calibrator — SolumnAI

Experience: 4–7 years · **Type:** Full-time · **Work week:** 6 days

About Solumn AI

Solumn AI builds the safety and alignment layer for frontier AI systems.

We work with leading AI labs to make advanced models deployment-ready, across the post-training and assurance stack: red teaming, vulnerability discovery, safety evaluations, agent-behavior testing, and the datasets and frameworks that make it measurable. If capability asks "what can the model do?", we ask "what happens when it fails, and how do we prevent it?"

Why this matters: Every technology cycle begins with capability and matures through trust. The internet, cloud, mobile and social all ran the same arc, with security and governance becoming indispensable only after the first wave of scale. AI is crossing that threshold now, and the assurance layer is becoming foundational. We're building the firm that does it to a benchmark standard.

The role

You will set the quality bar for everything we ship. You author tasks yourself, lead QC, and run the operations around it — reviewers, guidelines, calibration and delivery.

What you'll do

- Author RLE tasks and reference solutions to set the gold standard for the team.
- Lead QC on task packages: instructions, grading scripts, environment definitions, reference solutions.
- Write and maintain guidelines, rubrics and edge-case playbooks.
- Run calibration: measure reviewer agreement, run calibration rounds, fix drift fast.
- Manage QC operations: reviewer allocation, throughput, turnaround, and escalations.
- Own quality metrics and client-facing quality reports; turn client feedback into guideline updates.
- Hire, onboard and coach LLM Trainers and reviewers.

What we're looking for

- 4–7 years in LLM training data, evaluation or QA, with at least 1–2 years leading a team.

- Strong code-reading in Python (plus one of JS/TS or Java); comfortable with Docker and test harnesses.
- Hands-on experience authoring or reviewing SFT/RLHF/RL data for coding or agentic models.
- Proven record building QC processes, rubrics and calibration loops at scale.
- Clear writer who can turn a vague client spec into precise guidelines.
- Strong pedigree: a top-tier institute, or a top AI lab or leading data/eval company.

Nice to have

- Security background (OWASP, CWE, secure code review).
- Experience with inter-annotator agreement metrics.
- Delivered data programs for frontier labs.