AI Engineer

Build rigorous evaluations and research environments to understand frontier AI model behavior.

About Solumn AI

Solumn AI works to understand and evaluate how frontier AI models behave in challenging, real-world environments. We develop sophisticated evaluation environments, research infrastructure, and rigorous methodologies to assess AI systems against frontier labs' safety and alignment standards.

Our goal is to translate AI safety research into practical assets—evaluations and training data—that help drive measurable improvements in model behavior.

The opportunity

We are looking for AI Engineers who combine engineering fundamentals with research judgment. You will investigate frontier-model behavior, formulate hypotheses, design meaningful experiments, and build environments, benchmarks, and datasets that produce reliable evidence. The work goes beyond implementing existing benchmarks or AI applications: you will help decide what to measure, how to test it rigorously, and how to turn open-ended questions into reproducible technical systems.

What you will work on

01 **Research and experimental design**

Translate safety questions into hypotheses, controlled experiments, evaluation methodologies, baselines, and measurement criteria.

02 **Evaluation environments and harnesses**

Build realistic agentic and tool-use environments; develop execution harnesses, trajectory collection, and scalable experimental workflows.

03 **Benchmarks, datasets and evaluation quality**

Create benchmark and evaluation datasets, robust grading and verification systems, and QA processes that detect shortcuts or misleading scores.

04 **Model behavior analysis**

Investigate safety-relevant behaviors and failure modes; distinguish model limitations from environment, rubric, or verifier defects.

05 **Research engineering standards**

Improve reproducibility, test coverage, documentation, internal tooling, and technical standards across research projects.

The candidate

What you bring

Engineering fundamentals
Strong Python and software engineering skills, including experience building and debugging APIs, pipelines, tools, or other complex technical systems.

Modern AI systems
Practical understanding of LLMs, transformers, inference, tool use, and agentic architectures. Familiarity with post-training or RL is valuable.

Research aptitude
Ability to critically read technical work, formulate hypotheses, design experiments, and interpret results. A publication record is not required.

Rigor and analytical judgment
Attention to evaluation validity, confounders, reproducibility, verifier robustness, and the causes of unexpected experimental outcomes.

Ownership and curiosity
Comfort with ambiguous problems, rapid learning, independent implementation, and a genuine interest in AI safety and alignment.

Particularly relevant experience

Experience in any of the following areas is advantageous, but not mandatory:

- LLM evaluations, benchmark development, or model behavior analysis
- Agentic systems, multi-agent environments, or tool-use frameworks
- AI safety, red teaming, adversarial testing, or security research
- ML research engineering, experimental infrastructure, or automated grading
- Dataset development, synthetic data, or training and post-training workflows

Why Solumn AI

Work on consequential questions about frontier AI behavior while building the environments, methods, and systems needed to answer them. You will take ownership across research and engineering, from an initial idea to a working evaluation and credible findings.

We value engineers who care as much about whether an evaluation is scientifically meaningful as whether the implementation works.