Sam Neubauer
Raleigh, NC
samneubauer@gmail.com
(410) 925-6431

August 25, 2025

Hiring Committee
Lenovo — Advanced AI Technology Center (AAITC)
Morrisville, NC

Dear Lenovo Team,

I’m excited to apply for the Advisory Engineer, AI Model Evaluation role. Lenovo’s push to advance practical enterprise-grade AI, and to do it at scale through initiatives like the AAITC, aligns with my background building and evaluating production LLM systems and my passion for rigorous, science-driven testing. Your focus on assessing performance, robustness, and safety for LLMs, LVMs, and multimodal models is exactly where I operate day-to-day.

Over the last several years, I’ve led deployments of RAG pipelines, agentic workflows, and model-in-the-loop applications in regulated contexts (healthcare and insurance). My work pairs hands-on Python/SQL engineering with evaluation design: golden-set curation, adversarial test generation, hallucination/groundedness scoring, latency and cost telemetry, and guardrail/prompt-policy enforcement. I’ve implemented retrieval evaluators (answer faithfulness, context recall/precision), red-team harnesses, and CI/CD gates that block promotion when eval KPIs regress. These are the same levers a world-class evaluation team must pull to ensure safe, reliable user experiences at scale.

I’m particularly drawn to Lenovo’s hybrid AI strategy and ecosystem building — turnkey enterprise solutions, validated designs with AI Innovators partners, and customer-centric briefings that translate research into outcomes. Those signals tell me this evaluation role isn’t an academic exercise; it’s a platform function that de-risks deployments for real customers. I’d be eager to help formalize model evaluation playbooks that fit Lenovo’s portfolio — from data-center to edge — so product teams can ship with confidence.

What I bring to AAITC for model evaluation

- End-to-end eval design: Build truth sets; define task taxonomies; wire up model/adapter variants; automate hallucination, safety, and robustness suites (prompt attacks, distribution shift, low-resource, multilingual).

- RAG/agent evaluators: Grounded QA scoring, context attribution, retrieval quality, tool-use success rates, multi-hop reasoning checks; drift alerts on embeddings and content stores.

- Operational excellence: CI/CD for models and prompts; canarying and A/Bs; eval dashboards with SLOs/SLAs for quality, latency, and cost; post-incident reviews with reproducible traces.

- Cloud fluency: Experience standing up pipelines on Azure and AWS (containers, registries, orchestration), integrating eval gates into deployment workflows to keep teams shipping safely.

Lenovo's remit spans frontier research to real-world delivery, and your teams ship — across PCs, servers, and services. That breadth is rare and exactly where evaluation matters most: consistent, comparable, and trustworthy metrics that travel from lab to field. I'm a quick study who actively tests new LLMs and agent frameworks and then translates that learning into practical guardrails and scorecards teams actually use.

Thank you for considering my application. I'd welcome the chance to discuss how I can help AAITC institutionalize evaluation — so every Lenovo AI experience is measurably robust, safe, and ready for customers.

Sincerely,

Sam Neubauer