LLM Fine-tuning & Model Optimization Engineer

Location: San Francisco (US) or Berlin (GER)
Type: Permanent (Full-time)
Start Date: As soon as possible / by arrangement

ABOUT DAINA

DAiNA is a precision-oncology company focused on enabling personalized cancer treatment for individual patients. We combine comprehensive molecular tumor data including genomics, transcriptomics (bulk, single cell and spatial), proteomics and epigenetics, with AI-driven analysis. Our platform connects multi-omic profiling with functional ex-vivo tumor models and personalized liquid-biopsy monitoring, creating a continuous workflow from biopsy and treatment selection through to therapy monitoring and adaptation. We also operate GMP manufacturing to produce individualized N=1 therapeutics. In short, we help physicians make more informed, personalized treatment decisions based on high-dimensional molecular tumor data.

For more information, visit: www.daina.com

THE ROLE

As LLM Fine-tuning & Model Optimization Engineer, you will define how DAiNA adapts, evaluates and optimizes open-weight language models for biomedical and clinical use cases.

You will work on model selection, domain adaptation, fine-tuning, quantization, distillation and benchmarking. Your goal is to ensure that models are not only technically strong, but also reproducible, reliable and suitable for secure deployment in sensitive biomedical environments.

Your work will support DAiNA's AI-driven reporting, biomedical knowledge retrieval, molecular interpretation and internal decision-support workflows. A key focus is to build repeatable and auditable methods for model adaptation while respecting data privacy, data residency and compliance requirements.

WHAT YOU'LL DO

- Select and evaluate open-weight models for DAiNA-specific biomedical and clinical use cases.
- Run domain adaptation, full fine-tuning, quantization and distillation.
- Benchmark models for quality, factuality, safety, latency, cost and target-hardware performance.
- Build reproducible training, evaluation, versioning and packaging pipelines.
- Define clear criteria for when fine-tuning adds value versus retrieval, prompting or model routing.
- Document methods, datasets, model artifacts, evaluation results and limitations.
- Prepare optimized models and reusable adaptation workflows for secure operational deployment.

- Work closely with AI/ML, RAG, data engineering and platform teams to integrate optimized models into DAiNA workflows.

WHAT YOU BRING

- Hands-on experience with LLM fine-tuning
- Practical experience with open-weight model families such as Llama, Mistral, Qwen or comparable models.
- Good understanding of quantization, distillation, inference optimization and hardware-performance trade-offs.
- Experience with benchmarking, model evaluation and reproducible ML workflows.
- Understanding of model safety, reliability, factuality and domain adaptation.
- Ability to create methods and pipelines that other teams can understand, reproduce and operate safely.

NICE TO HAVE

- Experience with biomedical, scientific or clinical text corpora.
- Experience adapting models for healthcare, biotech, diagnostics, pharma or other regulated environments.
- Familiarity with evaluation for hallucination reduction, grounding, factuality and clinical/scientific reliability.
- Experience with secure cloud, isolated deployment environments or data-residency constraints.
- Familiarity with RAG systems, retrieval evaluation or biomedical knowledge bases.

WHY DAINA

- Employer contributions toward private health insurance or supplementary health coverage, as well as pension or retirement savings.
- Hybrid model with ~60% of working time **expected** on-site and the rest **remotely**, depending on team and business needs.
- Performance-based bonus opportunity, depending on company and individual performance.
- A personal development budget for conferences, courses, certifications, and training.
- The opportunity to work directly on real patient cases and help shape a first-in-class precision-oncology platform
- A proactive, collaborative team with fast decision-making and strong ownership of your domain.

HOW TO APPLY

Send your CV and a short note on why this role fits you to **recruiting@daina.com**, quoting “LLM Fine-tuning & Model Optimization Engineer” in the subject line. We review applications on a rolling basis and aim to reply quickly.

DAiNA is an equal-opportunity employer. We welcome applicants of every background and assess every candidate on merit, regardless of age, gender, ethnicity, religion, disability, sexual orientation or origin. If you need any adjustment to the process, let us know in your application.