



Job Title: AI Engineer (LLM foundations tools) – Falcon LLM Team

Location: Technology Innovation Institute, Abu Dhabi, UAE

Technology Innovation Institute (TII) is a publicly funded research institute, based in Abu Dhabi, United Arab Emirates. It is home to a diverse community of leading scientists, engineers, mathematicians, and researchers from across the globe, transforming problems and roadblocks into pioneering research and technology prototypes that help move society ahead.

Artificial Intelligence Cross-Center Unit

The Artificial Intelligence Cross-Center Unit is the machine learning powerhouse of TII, working in close collaboration with our other research centers to harness the full benefits of AI across our projects – and drive innovation from new computing paradigms, designing and delivering new AI methodologies, technologies, solutions, and systems that address challenging issues across multiple sectors of the economy – from technology to healthcare, cybersecurity, and government, among others.

We incorporate core elements of intelligence (perception, sensing, planning, and language) in the ideation, design, and prototyping of next-generation systems with human-like intelligence. We build advanced AI computing and scalable AI-based software stacks and hardware systems to deliver significant enhancements in systems infrastructure. Our AI researchers, scientists, and engineers collaborate to ensure innovative outcomes, from AI theory to AI technologies towards better intelligence.

Falcon LLM Team

The Falcon LLM team at the Technology Innovation Institute (TII) is at the forefront of developing cutting-edge generative AI and language models. Our Falcon models have garnered significant open-source adoption, and we are committed to pushing the boundaries of AI performance, alignment, and safety. Join our dynamic team to advance the capabilities of our foundational models and make impactful contributions to the AI community.

Role Overview:

We are seeking a talented AI Engineer to join our LLM Foundational Team. In this role, you will be responsible for maintaining and developing critical tools that support our large language model infrastructure across multiple domains including training, post-training, evaluation, and ecosystem integration, while contributing into the next iteration of Falcon LLM series.

Key Responsibilities:

- Develop and maintain robust training pipelines for large language models, ensuring efficient and scalable distributed training workflows
- Design and implement post-training tools including fine-tuning frameworks, parameter-efficient adaptation methods, and model optimization utilities
- Create comprehensive evaluation frameworks to assess model performance across various benchmarks, tasks, and quality metrics
- Build and maintain integration tools that enable seamless deployment of LLMs within our broader product ecosystem and infrastructure
- Monitor and optimize the performance of existing ML tools, identifying bottlenecks and implementing improvements
- Collaborate with research teams to translate cutting-edge LLM techniques into production-ready tooling
- Document tooling architectures, usage guidelines, and maintenance procedures for cross-team collaboration
- Troubleshoot complex issues across the entire LLM lifecycle from training to production deployment
- Stay current with industry developments in LLM research and engineering practices to continuously improve our toolset

Minimum Qualifications:

- Master's or Ph.D. in Computer Science, AI, ML, or a related technical field.
- Strong proficiency in Python, with proven experience in scalable software development.
- Hands-on experience with distributed training frameworks such as Megatron-LM, DeepSpeed, FSDP, etc.
- Experience with large-scale data processing and optimizing model throughput and memory efficiency in deployment.
- Strong analytical and problem-solving skills with a drive to understand and improve the underlying mechanics of model performance.

Preferred Qualifications:

- Experience with low-level optimization: CUDA, Triton kernels, low-precision training and inference acceleration.
- Demonstrated contributions to open-source projects related to ML, deep learning, or systems optimization.
- Experience with Kubernetes for container orchestration and deployment



- Deep curiosity and ownership mindset—willing to dig into metrics, logs, and performance reports to drive improvements.

What we offer:

- **Competitive Benefits:** Enjoy competitive compensation, access to state-of-the-art computational resources, and the chance to work with some of the brightest minds in the AI field. Our collaborative and inclusive work culture is centered on innovation and personal growth.
- **Mentorship and Project Involvement:** Benefit from close mentoring and active participation in exciting AI projects that will help you grow your skills.
- **Equal Opportunity:** We are committed to creating a diverse and inclusive workplace. TII values diversity and does not discriminate based on race, religion, gender, age, national origin, sexual orientation, marital status, veteran status, or disability.