

Jeonghun Kim

ML Engineer / AI Researcher

portfolio.jhkim3ispark.work

FURIOSA  Software Engineer, Kernel Programming

bandlagom0927@gmail.com

+82 10-9559-9471 · Seoul

github.com/thuygom



Cover Letter

Dear FuriosaAI Hiring Team,

I am applying for the Software Engineer, Kernel Programming position (Job ID 4005375201). I bring model-side implementation and evaluation experience to kernel development. Training task-specific models and operating GPU workloads has given me a practical understanding of model operations, supported inputs, and the need to check outputs alongside performance. Before ISPARK, I served one year as an information-systems NCO in the Republic of Korea Air Force (Dec 2021–Dec 2022), maintaining Linux servers through shell scripting and log analysis.

I developed business-card information extraction used in TieMing, reaching 94% accuracy across seven core fields in a workflow covering approximately 5,000 cards. To address correct text being assigned to the wrong fields, I trained a smaller student model through knowledge distillation and low-rank adaptation (LoRA). I separately targeted difficult cases: follow-up training and selective regeneration of malformed outputs raised field-value accuracy from 30.2% to 65.3%. The work involved choosing the training method, diagnosing output errors, and implementing the changes in the extraction workflow.

I take responsibility for model deployment as well as training, building vLLM and FastAPI inference services in Docker with Prometheus and Grafana monitoring. Watching service failures alongside GPU usage helps me distinguish application problems from resource constraints. I ran distributed training on two RTX 4090 GPUs with BF16 precision and checkpoint recovery, and now also use two A100 GPUs through a national computing-support program. I track experiments in Weights & Biases so model changes can be compared across runs.

At FuriosaAI, I can contribute realistic model workloads and numerical reference checks while developing accelerator-specific implementation skills. I would evaluate each operation against explicit input ranges and correctness criteria before comparing performance.

Sincerely,

Jeonghun Kim