# Jeonghun Kim

Machine Learning Engineer / AI Researcher
OpenAI · Applied AI Engineer
Seoul, South Korea · +82 10-9559-9471 · bandlagom0927@gmail.com
portfolio.jhkim3ispark.work / github.com/thuygom

## Profile

I build AI applications around retrieval, SQL execution, and speech processing. At ISPark, I implemented model evaluation, tenant access control, and job recovery alongside domain-model fine-tuning and inference APIs.

## Experience & Education

**ISPark R&D Center · AI Researcher** Aug 2025–Present

Work-study intern Aug 2025–Feb 2026; full-time since Mar 2026

One of two AI developers on TieMing. Responsible for OCR/KIE, web search, speech, CRM Text-to-SQL, and voice-assistant integrations. Demonstrated at AI Summit Seoul 2026; Android beta and Play Store registration in preparation. portfolio-tieming.jhkim3ispark.work

**Hansung University · B.S., Computer Engineering** Feb 2026

Big Data major · GPA 3.95/4.5 · IRIS Lab undergraduate researcher, 2024–2026

## Applied AI Systems

**CRM Text-to-SQL · Retrieval and execution** Jun 2026–Present

- Reconciled database descriptions and relationships with PostgreSQL metadata; combined lexical, dense, sparse, and catalog retrieval by rank to select a question-specific schema.
- Checked single SELECT/CTE statements, table/column allowlists, and LIMIT with SQLGlot before read-only execution with timeouts.
- Designed and validated PostgreSQL row-level security (RLS) to restrict CRM queries to each user's permitted rows.

portfolio-t2s.jhkim3ispark.work

**iScribe / Real-time STT** Mar 2026–Present

- Selected Community-1 after a four-meeting AMI comparison: **weighted DER 28.72% → 26.92%**, with zero collar and overlapping speech included.
- Fixed synchronization and duplicated transcripts in a separate five-speaker, speaker-enrolled streaming regression case: **WER 62.77% → 34.06%**, DER 48.58% → 30.23% under loose scoring.
- Separated live sessions from long-running jobs; built PostgreSQL-ledger recovery, GPU advisory-lock admission, and Keycloak/OIDC tenant access control.

iscribe.jhkim3ispark.work

Demo login: demo / Password: demo

## Model Training & Evaluation

**GomOCR · Vision-language fine-tuning** Apr 2026–Present

- Split 2,906 reviewed cards into 2,034 training, 581 validation, and 291 fixed-test examples; checked ID and exact image-hash overlap.
- Froze the Qianfan-OCR vision tower; fine-tuned the language model and projector with NF4 QLoRA, BF16, two RTX 4090 DDP replicas, and effective batch 16.
- Fine-tuning plus decoding calibration raised **field-value accuracy from 30.2% to 65.3%** and **list micro-F1 from 54.5% to 80.2%** versus the untuned model on 291 fixed-test cards.
- Parse-only retries reduced errors from **27 to 1** with the same adapter. Chose decoding on validation; test had been observed earlier. Address F1 and record exactness regressed slightly.

portfolio-gomocr.jhkim3ispark.work

**Text-to-SQL · Separate model-training experiment**

- Fine-tuned Distil-Qwen3-4B with NF4 QLoRA on 2,172 examples; achieved **85.1% SQL token exact match on 241 same-template synthetic validation examples**.
- On a separate 100-question real evaluation: **45% strict result match**, 95/100 pipeline completion, and 0.767 mean value-Jaccard. Kept the adapter experiment separate from CRM service inference.

## Research & Patent

**Five papers, four as first author.** MRO, IJIBC 17(3), 144–154 (2025), studied multimodal fusion-path selection through NAS. IN-MRO (KSC 2025) extended this to implicit NAS policy learning. Other papers: AURA, SCOPE (first author), and SCOPE-X (co-author).

**Inventor, Korean patent application 10-2026-0119188.** Database-description generation and hallucination-suppression preprocessing for Text-to-SQL; filed June 30, 2026, with examination requested.

## Deployment & Ongoing Training

- Built vLLM/FastAPI inference APIs and Docker deployments through Cloudflare Tunnel. Monitored service, container, and GPU state with Prometheus/Grafana.
- Currently training an LLM on two NIPA-supported A100 GPUs, tracking runs with Weights & Biases, and storing weights in private Hugging Face repositories.

## Additional Background

**Certifications:** SW Development Level 5; Industrial Engineer Information Processing; Linux Master Level 2.

**Military:** ROK Air Force information systems; 21 months enlisted, then professional NCO, Dec 2021–Dec 2022. Linux server administration.