Byeonghyeon Kim

CAREER PORTFOLIO

AI Engineer · LLM Systems

On-Prem LLM Infra · Access-Controlled Memory · Security-grounded

Gyeonggi-do, South Korea | bk@bkan.dev | (+82) 10-4676-5884 | GitHub | Portfolio | Threads

Career Summary

Period	Organization / Type	Role · Scope
2026	**Enterprise AI Systems (Freelance)**	AI Engineer · Technical Lead - on-prem LLM infrastructure · enterprise AI memory · database design · ADR governance
2025 - Present	**Solopreneur**	Founder / Engineer - planned, built, and launched AI products (OurFolio · ASHD v2 · BluePy · BK-HQ)
2025.06 - 2026.01	**FastCampus AI Bootcamp (14th cohort)**	Completed - 5 applied ML competitions · MLOps · multi-agent project
2021 - 2025	**Franchise Business Owner/Operator**	Owner/Operator - POS-driven demand forecasting and inventory optimization; prepared for AI career transition
2018.02 - 2021.02	**CMT EYES**	Security Monitoring Analyst - SOC / CERT incident response

Enterprise AI Systems - Freelance

2026

Designed and built an end-to-end AI environment for an enterprise that could not send data to external clouds, spanning LLM infrastructure, data models, enterprise knowledge management, and development processes. Also led technical governance for a small engineering team.

On-Prem LLM Multi-Agent Infrastructure

Role Sole owner for architecture, implementation, and operations · **Tech** Self-Hosted LLM · Tailscale · GGUF/MLX · Multi-Agent · Discord Bot

- **Context** Security requirements prohibited external cloud LLM APIs, requiring fully internal model-serving infrastructure
- **Action** Assigned workloads across **15 Mac Studios**, connected them through an isolated Tailscale network, self-hosted LLMs without public exposure, and provided a Discord bot as the single entry point
- **Action** Designed a **local LLM portfolio per device**, assigning GLM and Qwen models using quantified license, RAM, quantization (GGUF/MLX), and KV-cache headroom constraints; cross-checked official power data and corrected errors in the hardware sheet
- **Action** Built a **2-agent handoff** pipeline in which execution and review agents automatically exchanged work
- **Result** Delivered an always-on cluster with **automatic restart recovery**; enterprise memory and agent systems ran on this stack

Integrated Call-Center and Customer Database Schema

Role Sole owner for data modeling · **Tech** PostgreSQL · SQL-first Migration (goose) · ETL

- **Action** Analyzed legacy ETL and designed migration for 94K customers, 86K addresses, and 5.7K consultations; identified IVR/inbound duplication in 32K raw CDR rows and **normalized them into 23K actual calls**
- **Action** Designed a **27-table normalized schema** integrating call-center and customer data, including natural-key strategy, soft deletion, slowly changing dimensions (SCD), and explicit invariants
- **Action** Evaluated migration engines (rejected Atlas, dbmate, sqitch, and Flyway; **selected goose**), externally validated pgvector/halfvec capacity limits, and established a SQL-first SSOT plus a **schema-invariant checklist** as documentation-level TDD
- **Result** **Passed 2 external reviews** and adversarial code review

Engineering Governance · Technical Lead

Role Primary engineering administrator · decision-record owner · **Tech** GitHub Monorepo · ADR/MADR · Git Workflow

- **Action** Set up the GitHub organization, monorepo, and branching strategy (main/dev/feat) for a 3-person team; introduced and maintained an **ADR (technical) + Decision (non-technical) record system** (ADR-001-008)
- **Action** Designed work allocation across sales, advertising, and CRM tracks; led onboarding; and established **worktree isolation and Safe Staging rules** for multi-session AI collaboration
- **Result** Delivered a transition PR covering **202 files / +37,000 lines** plus a separate summary for a nontechnical team lead; documented infrastructure-related data loss neutrally and proposed recurrence-prevention rules

Enterprise AI Systems - Freelance (Continued) 2026

Team and Enterprise AI Memory System

Role Open-source customization and system architecture · **Tech** Hindsight · Graphify · Access Control

- **Context** Agent memory accumulated personal work records, requiring both privacy isolation and complete retention
- **Action** Customized open-source agent memory **Hindsight** with per-employee personal and team memory banks; controlled Recall and Reflect permissions by security tier so personal banks remained **visible only to their owners**
- **Action** Measured **211 production-like records in a 1-week trial**, quantified recall-based retrieval limits (**15.6% recall**), adopted direct time-filtered queries, and disabled the Reflect summary layer to prevent information contamination
- **Action** Built an **automated synthesis pipeline** for daily and weekly reports; measured quality with evals (**0% hallucination rate**) and applied executive-value curation rules
- **Result** Delivered an **enterprise knowledge-management foundation** that retained records without violating access boundaries, with a 3-stage personal → team → enterprise promotion model

Korean Embedding and Reranker 3-Way A/B Experiment

Role Sole owner for experiment design and execution · **Tech** MTEB-ko · FAISS · sentence-transformers · statistical testing

- **Context** A 1.38-point gap on the Korean embedding leaderboard was within noise, so benchmarks alone could not justify replacing the RAG model
- **Action** Designed an arctic-ko vs KURE-v1 vs BGE-m3-ko **3-way A/B experiment** with a pre-specified decision rule (Bonferroni α=0.025, MRR), 90 queries, 6 strategies, and self-retrieval integrity checks
- **Result** No pair reached statistical significance, and the sample size was insufficient; reported the study as **underpowered** and retained the model due to no evidence for replacement. Validated the decision with a **three-way cross-check**: leaderboard results, community signals, and internal experiments

Codebase Knowledge Graph

Role Sole owner for implementation and validation · **Tech** Graphify · Knowledge Graph

- **Action** Structured the codebase as a knowledge graph with **1,158 nodes / 1,797 relationships / 73 communities** and integrated it with team automation tools
- **Action** Audited the full pipeline and **found and fixed 4 graph-integrity bugs**: hyperedge merge contamination, phantom cross-project edges, noisy-node pruning, and stale state; then finalized metrics

Unified Admin Dashboard (Demo)

Role Sole developer · **Tech** Spec-Driven Dev (PRD/SDD/TDD) · Accessibility

- **Action** Built an admin system deriving **13 modules** from one module registry, including accessibility, dark mode, and responsive layouts
- **Action** Audited all flows, identified **12 UX/accessibility gaps, and shipped fixes for 10**, including unified search, WAI-ARIA keyboard navigation, dark-mode chart synchronization, and self-hosted fonts
- **Result** Used spec-driven development to reach **81 passing tests, up from 54**, with 0 regressions and 0 console errors across 13 routes

Korean Document Pipeline · DOCX Track-Changes Library

Role Sole developer · **Tech** Python · OOXML(ECMA-376) · pandoc · MinerU VLM · typst

- **Context** Executive reporting relied on HWP/DOCX/PDF, but no tool could ingest tracked changes and comments into the LLM pipeline
- **Action** Built a **pure-XML library** for DOCX tracked changes/comments, removing the LibreOffice dependency and using pandoc round trips as the TDD baseline through v1-v4; 48 tests
- **Action** Validated MinerU VLM on Korean PDFs with typst, then **packaged the workflow as a Claude Code skill**
- **Result** A/B pilot measurements showed a **55% reduction** in task time (67 sec → 31 sec) and half as many tool calls

Additional Work

- **Action** Conducted full-stack research for meeting recording → diarization → minutes automation (4-cycle deep dive); built and delivered a financial-data validation MCP server (6 tools), **explicitly labeling unverified results**; maintained the internal web platform

Solopreneur - AI Products

2025 - Present

Owned the full lifecycle from product planning and development through deployment and operating decisions. Launched 2+ product MVPs.

OurFolio - Evidence-Based Portfolio Platform GitHub ↗

ourfolio.bkan.dev/demo · Public Showcase

Role Solo product development (planning, engineering, deployment) · **Tech** Next.js 16 · Neon · Drizzle ORM

- **Action** Built a platform for developers to substantiate work with evidence, featuring server-rendered public profiles, one-click GitHub repository import, and dynamic OG/SEO generation
- **Action** Implemented a **2-stage, zero-downtime expand-contract schema migration**, test-enforced Sentry PII scrubbing, and policyVersion-based consent history; operated **7 CI checks** spanning Lighthouse, coverage, and migrations
- **Result** Independently built and deployed the MVP through a validated **v1**. After assessing market viability, ended monetization and converted it into a **public showcase with a read-only live demo**, demonstrating disciplined product sunset decisions

ASHD v2 - OCR+LLM Receipt and Warranty Management GitHub ↗

ashd.bkan.dev · MVP Launch → Public Snapshot

Role Solo developer · **Tech** Next.js 15 · Supabase (Auth/RLS) · OCR · LLM · CI

- **Action** Used OCR plus rule-based extraction and an LLM to structure receipts and warranties and send email/Telegram expiration alerts. Operated a production-grade service through payment-integration experiments and automated solo operations with **15 CI workflows** including evals, automerge, and P0 alerts
- **Result** Managed extraction accuracy quantitatively on the **KORIE benchmark (314 examples)**: field-level F1 improved for store **9% → 100%**, purchase date 89% → 99%, and amount 98% → 100%; operated an **eval regression gate** that blocked PRs below seeded gold-set thresholds

BluePy 2.0 - Open-Source Infrastructure Security Scanner GitHub ↗

v0.1 alpha · MIT

Role Sole developer (ground-up rebuild) · **Tech** Python 3.12 · Clean Architecture · CLI

- **Context** A 2017 KISA legacy security scanner included validators that returned false PASS results, making its output untrustworthy
- **Action** Redesigned it in Python 3.12 with Clean Architecture; exposed only checks with validated causality by default (Linux permissions; macOS SIP, FileVault, and automatic updates), removed false-PASS validators, and isolated unverified legacy rules as experimental
- **Action** Built an asyncssh-based **remote SSH scanning** architecture for Linux/macOS with PASS, FAIL, and MANUAL outcomes; added a real-SSH Docker E2E fixture and **546 tests**
- **Result** Released a CLI-first scanner with JSON/table output and strict SSH host-key verification as a **transparent v0.1 open-source project**, translating security operations experience into code-reliability standards

BK-HQ - Solo AI Company Operating System

In active development and operation

Role Architecture and operations · **Tech** Multi-Agent Orchestration · Automation Hooks

- **Action** Built an orchestration platform enabling one CEO to operate **15 AI agents** across 4 teams (product, legal, marketing, engineering), with 31 skills, 19 automation hooks, and 5 crew workflows
- **Result** Integrated 5 Smart Skill Chains, social automation, Agent Dashboard, and a compound knowledge system (25+ domains, 100+ documents) into one operating system, validating a **solo operating model**

Security Monitoring & CERT - CMT EYES

2018.02 - 2021.02

Security Monitoring & Incident Response

Role Security Monitoring Analyst (SOC/CERT) · **Tech** FortiGate/UTM · FortiSIEM · WAF/IPS · Wireshark

- **Scope** Remotely monitored and analyzed **20+ Fortinet-focused WAF/IPS/UTM/SIEM appliances** for **50+ clients**, including night/weekend shift rotation and the escalation process
- **Action** Handled **1,000+ events per month** across the full CERT lifecycle: detection → log/packet analysis → blocking → policy improvement → reporting
- **Action** Tuned SIEM correlation rules and custom IPS/WAF signatures; developed new detection scripts covering diverse security events
- **Action** Analyzed intrusions end to end - reconstructed timelines from logs and packets, identified attack paths, and wrote incident reports
- **Action** Ran emergency DDoS response (range blocking, traffic analysis) and triaged KISA/government threat advisories into client-impact assessments and urgent guidance
- **Action** Participated in **penetration-testing and compromise-analysis projects** for major enterprise clients; ran phishing simulation exercises
- **Result** Improved false/true-positive classification accuracy by **50%+** through policy and detection-rule tuning
- **Result** Automated repetitive reports with scripts, reducing per-event reporting time from **30 min → 5 min**

Security Engineering & Client Support

- **Action** Onboarded new clients into monitoring - initial appliance policy setup, baseline tuning, and firewall policy cleanup and migration
- **Action** Engineered client security appliances and operated Linux/Unix servers
- **Action** Supported an on-site **10G network expansion** for a major education client whose mock-exam traffic surges caused recurring annual outages - hands-on network installation work
- **Action** Supported the shift to SSL-VPN-based remote-work security during COVID-19
- **Action** Authored monthly and quarterly security reports and delivered regular client briefings
- **Action** Led new-hire training and owned the onboarding curriculum

ML Competitions & Projects - FastCampus AI Bootcamp (14th Cohort)

2025.06 - 2026.01

Completed competition-driven applied ML work. Metrics below show improvement over each competition baseline.

Toss Ads CTR	Built a CTR model on large-scale ad logs; improved LogLoss/AP through sequence features, downsampling, and tuning; placed in the **Top 10%**
Science QA IR	Improved RAG retrieval and ranking with Solar embeddings, rank-graph refactoring, and prompt tuning; MAP/MRR **+107%**
Document Images	Classified 17 types of scanned documents; backbone experiments, TTA, augmentation, and post-processing improved F1 **+470%**
Dialogue Summarization	Summarized everyday Korean dialogue; QLoRA fine-tuning, augmentation, and post-processing improved ROUGE **+30%**
Real-Estate Prediction	Improved RMSE by **58%**

Marketing AI Multi-Agent System

- **Action** Orchestrated **6 agents** to generate trend, copy, segment, review, competitor, and strategy reports, replacing repetitive research

Movie Rating Prediction MLOps

- **Action** Built a TMDB API-based rating prediction service with an **E2E automated** data collection → training → deployment → monitoring pipeline (CI/CD)

LLM-Based RAG Training Content

- **Action** Created training materials for practitioners adopting RAG for internal document search, covering BM25, vector, and hybrid comparisons, query expansion, and hands-on Hit@k/MRR evaluation

Portfolio Evidence

Project evidence, including screenshots, code, and live demos, is available at bkan.dev. See the 1-page resume at bkan.dev/resume-en.html.