

# Winter Jung(정겨울)

Seoul, South Korea ▪ me@winterjung.dev ▪ github.com/winterjung

Backend Engineer at Karrot (40M MAU) with 8+ years of experience in high-throughput systems, cost optimization, and recommendation infrastructure. I also build and operate LLM pipelines in production — from classification and evaluation to cost-efficient serving — leveraging AI and code agents as practical tools to ship faster and better.

## Karrot

2024.04–Present, Seoul

### Software Engineer

2024.04–Present

- **Saved ~\$1M+/year** by redesigning the feature logging pipeline (Kafka → Go consumer → GCS → BigQuery batch load) handling 3GB/s of recommendation event traffic, scaling to 4.5x growth with stable operation.
- Built and operate a production LLM pipeline (**2T+ tokens processed**) that classifies marketplace listings into **10,000+ category taxonomy** with attribute extraction. **Reduced LLM costs by 90%** through embedding-based pre-filtering, prompt caching, and hybrid retrieval (BM25 + semantic search). Designed LLM-as-a-Judge evaluation framework for continuous quality monitoring.
- Operated the feature store and recommendation engine powering Karrot's home feed — serving **10M+ daily requests at p999 <30ms**. Led set type support enabling carousel-based recommendations.
- Speaker at [GopherCon Korea 2024: "Deterministic Testing in Go"](#)

## Banksalad

2018.11–2023.09, Seoul

### Senior Software Engineer, Server Chapter Lead

2022.01–2023.09

- Contributed 15+ packages **as the first committer** to an in-house Go library used by 50+ repositories.
- Over 3 years, developed a gRPC project generator, an IDL repository (a company-wide API interface repo based on Protocol Buffers), and a DDL repository (a DB schema repo).
- **Reviewed 10K+ GitHub PRs** across 120+ repositories. Consistently **ranking #1** in code reviews annually. Reviewed technical specifications for 20+ server engineers, ensuring code quality and optimal implementation.

- **Reduced p99 latency by 90%**, queries by 95% and RDS CPU peak by 15%p for the core service that performed 260M queries/day by adding Redis and in-memory cache layer.
- **Reduced monthly costs by 80%** by migrating a single MySQL table with 70TiB and 26 billion rows into AWS Athena.
- Over 2 years, **shut down 16 legacy services** and removed 250+ legacy API endpoints.
- Planned and operated the server chapter onboarding process for over two years.

## Software Engineer

2018.11–2021.12

- **Participated as a tech lead** in the MyData server development project, handling **100M daily traffic from 250+ financial companies**. Managed performance improvements and issue resolutions for over 2 years.
  - **Migrated 130+ self-hosted Kubernetes services** to AWS EKS with **no downtime** to comply with MyData security requirements.
  - Decomposed 200+ monolithic legacy APIs built on Python, MongoDB into **15 microservices based on Golang, MySQL, and gRPC with dual-write and shadowing**.
  - Increased the financial card company integration by 318% and the Korea-government tax service integration by 868% by developing the year-end tax reconciliation algorithm.
  - Built a new ChatOps CI/CD pipeline, **reducing deployment steps by 80%** and **increasing production deployments by 400% per week**.
  - Built a zero-downtime deployment system based on VMs and L4 switches in a private cloud IDC using Packer and Ansible.
  - Established a company-wide Python style guide.
- 

## Extracurricular

- [Karrot tech blog posting "Lessons from Spending 2 Trillion Tokens on Category Classification"](#)
  - [Banksalad tech blog posting "Production ready gRPC in Golang"](#)
  - [Banksalad tech blog posting "Banksalad Golang coding convention"](#)
  - [2019 PyCon Korea speech "Why is Python 3.7 fastest"](#)
- 

## Education

Hongik Univ.

2014–2024, Seoul

Bachelor of Science in Urban Engineering