

Sungmin Han

MLOps Lead at Riid
ML GDE (Google Developer Experts)
Golang Korea GDG (Google Developer Groups) Lead

Currently living in **Seoul, South Korea**
(Need support for Visa for overseas relocation)

+82 1036167981
kenneth@pigno.se

LinkedIn <https://www.linkedin.com/in/sungmin-han-768419133/>

GitHub <https://github.com/kennethanceyer>

Career

Riid - 2.6 years as MLOps Lead	2021. 01 ~ Current
Naver Clova - 2.5 years as Research Engineer	2018.07 ~ 2020. 12
IGAWorks - 2.8 years as Software Engineer	2015. 11 ~ 2018. 05
JisikLog - 6 months as Software Engineer	2015. 02 ~ 2015. 08

Tech Stack

Language	Python, Golang, Typescript, Kotlin, and C#
ML engineering	Kubeflow, BentoML, MLFlow, Feature Store, Pytorch, Tensorflow, and Keras
Data engineering	Spark, Kafka, Databricks, BigQuery, Athena, Redshift, Confluent, Airflow, and Fivetran Has experience in CDC-based pipelines and ETL/ELT and leads architecture design.
Backend engineering	GitHub (w/ Actions), Docker, CI/CD, Clean code (w/ Quality gate), Software architecture, and Testing (w/ Unit and Integrated test)
Cloud	AWS, GCP, and Azure Adopted and used all three CSPs in my work.
Infrastructure	Kubernetes, Helm chart, Terraform, Ansible, ArgoCD, APMs(Grafana, Prometheus)
Domain knowledge	Education, Chatbot (w/ NLU and NLP), and Dialog system (w/ Telephony network, SIP, SDP, and RTP)

Education

Bachelor	Soongsil University Media (Graphics Programming)
Audit	Korea Advanced Institute of Science & Technology (KAIST) Deep Learning course completion (AI502)

Language

English	Upper-intermediate
Korean	Native
Japanese	Pre-intermediate

Certifications

CCNP Routing (CISCO)	2015. 07
OCJP (Oracle)	2012. 10
CCNA (CISCO)	2012. 07

Work

Riiid	<p>[1] Managing the product team and the MLOps team.</p> <p>[2] Serving a total of 8+ educational models (Knowledge Tracing, Score prediction, and Recommender System) and supporting 3 B2C and 6 B2B businesses.</p> <p>[3] Creating and leading a product team for a generative service with an LLM.</p> <p>details:</p> <p>[a] Designing data structures and configuring CDC pipelines based on Databricks and Confluent.</p> <p>[b] Build and Operate an A100 SXM5 GPU Cluster.</p> <p>[c] Configure Airflow 2.4.x with Kubernetes Executor.</p> <p>[d] For Databricks, read through the entire Spark and DeltaLake configuration and Unity Catalog configuration.</p>
Naver Clova	<p>[1] Served as a core engineer during the initial team-building phase of the voiced chatbot system</p> <p>[2] Experienced the entire process from initial design to launch, and monetized the product by integrating it with a Japanese courier service.</p> <p>details:</p> <p>[a] Developed a dialogue engine connecting TTS, STT, and NLU.</p> <p>[b] Developed services for VoIP technologies including RTP and SIP/SDP in Go language and implement encoding algorithms such as G.711.</p>

[c] Design and operate a high-availability architecture that is structured to withstand 2,000 RPS or more.

[d] Participated in TensorRT and Pytorch serving optimization work.

[e] Leading Japanese engineers and collaborative projects with Japanese engineers.

IGAWorks

[1] Participated in major development from initial team building to launching.

[2] A BI solution that requires pivoting, aggregation, and visualization of the required data in the format requested by the UI, with approximately 100 million data sets per customer.

[3] Distributed processing for interactive queries and query builder development that changes depending on the database type are the main features.

[4] Required a lot of review on the distributed query engine (Impala with Kudu, Presto, Hive on Tez, Tajo).

[5] 2016, 2017 released to G-Star (Korea's largest game advertising partner B2B conference).