

Tae Hwan Jung

AI Engineer · Founding Engineer @ Clober

nlkey2022@gmail.com | github.com/graykode | [Google Scholar](https://scholar.google.com/) | graykode.github.io

SUMMARY

어떤 도메인이든 빠른 learning curve로 그 분야의 탑 티어가 될 수 있습니다.

“AI knows everything” 시대에 단순히 “정보를 많이 아는 것”은 더 이상 중요하지 않다고 생각합니다. 특히 연구 조직이 아닌, 프로덕트가 더 중요한 스타트업에서 풀어야 하는 문제는 대부분 본인에게 익숙한 도메인에서 시작되지 않습니다. 따라서 저는 본인이 잘하는 문제를 푸는 것보다, **시장에서 꼭 풀어야 하는 문제를 능동적으로 찾아 푸는 방향성**이 앞으로의 엔지니어에게 필요하다고 생각합니다. 이는 문제에 접근하는 방식, 유지보수를 고려한 아키텍처 설계, 그리고 프로덕트를 매니징하는 전 과정을 아우르는 역량입니다.

저는 AI 엔지니어로 커리어를 시작했습니다. 이후 블록체인 필드에서 풀고 싶은 문제를 발견했고, AI에서 했던 것과는 전혀 다른 분야를 zero base부터 시작했습니다. 4년 동안 **훌륭한 팀원들**과 함께 3-4개월 주기로 프로덕트를 출시하며 6-7번의 프로덕트 사이클을 경험했습니다. 이 과정에서 엔지니어로서는 방향성은 맞지만 아직 **개척되지 않은 문제들을 깊이 고민**했고, 만든 프로덕트를 시장에서 성공시키기 위한 다양한 시도를 팀원들과 함께 해보았습니다.

창업을 경험하기 이전과 이후의 저는 비교할 수 없을 만큼 달라졌습니다. 엔지니어로서의 주체성뿐 아니라, 문제를 풀어나가는 과정에서의 복잡한 인간관계와 낙오의 경험을 통해 훨씬 깊게 성장할 수 있었습니다. 이 과정에서 가장 깊이 배운 덕목은 **윤리성과 책임감**입니다.

제가 풀고자 하는 문제가 있는 곳이라면, 다음 커리어의 도메인은 상관없습니다. 하지만 엔지니어로서는 어떤 분야에서든 AI조차 풀지 못하는 어려운 문제를 프로젝트 관점에서 풀어나가고 싶고, 팀원으로서 팀에 큰 위기가 닥쳤을 때 책임감 있게 그 무게를 질 수 있는 사람이 되고 싶습니다.

SKILLS

AI Skills

- LLM Distributed Pre-training
- Knowledge Distillation & Compression
- Programming NLP
- Document AI (LayoutLM, OCR)
- PyTorch / Python

Blockchain Skills

- Financial Modeling & DeFi Protocol Design
- Solidity Development & Gas Optimization
- Web3 Full-stack (FE / BE)
- On-chain Data Indexing Pipeline
- Data-driven Market Making & Arbitrage
- TypeScript (main) / Rust

EDUCATION

B.S. in Computer Engineering

Kyung Hee University (incl. two years of military service)

2015 – 2022

Republic of Korea

EXPERIENCE

Founding Engineer, Core Protocol

Clober (Merekat)

Apr 2022 – Present

Gangnam-gu, Seoul

Clober is a DeFi protocol company — “Building DeFi primitives that matter” — building an order-book DEX on-chain. The core problem we worked on was bringing a decentralized order-book exchange fully on-chain and shipping it as an optimized, usable product. Over four years of founding-stage work I helped take Clober from zero to one across the full product surface — protocol internals, backend, and frontend.

Beyond the engineering itself, what four years of founding-stage work at Clober gave me: hands-on zero-to-one product experience (planning, UI/UX, execution, launch); comprehensive domain knowledge in finance; and practical experience in DeFi architecture design and financial modeling. These are the parts I carry forward the most.

Clober v1 — full-stack DEX

2022.04 – 2023.03 | clober.io

- Built Clober v1 across the full stack (Smart contracts + Backend + Frontend).
- On the contract side, implemented gas-optimized on-chain order matching using segmented segment trees and octopus heaps.
- Write-up: [ethresear.ch post](#) · [recap thread](#) · [gas benchmarks](#)

DEX aggregator on Polygon zkEVM

2023.03 – 2023.06

- Built a DEX aggregator for Polygon zkEVM in Rust, solo development, to support Clober’s growth on the chain.

Coupon Finance — fixed-rate DeFi lending

2023.06 – 2024.06 | coupon.finance

- Built the frontend and backend of Coupon Finance, a sister protocol built to drive Clober’s growth.
- Coupon Finance is a fixed-rate DeFi lending protocol that tokenizes interest rates as tradable “coupons” on an on-chain order book — collapsing the deposit–borrow APY spread down to the order book’s bid–ask spread, so both lenders and borrowers get tighter rates than on traditional money markets.

ETH/USDC vault strategy

2024.06 – Present | base.clober.io

- Responsible for the development and management of Clober’s ETH/USDC vault strategy, end-to-end.
- Improved market-making strategy using Markov modeling.
- Built the data pipeline powering strategy backtesting and simulation.
- Designed a real-time on-chain event tracking pipeline (Ponder-style indexing → Grafana dashboards) for live strategy monitoring and data analysis.
- Optimized DEX ↔ Clober arbitrage to reduce market-maker losses.
- Launches: Monad mainnet (2025.11), Base chain vault (2026.01).

AI Research Engineer (Full-time)

Upstage AI

Jul 2021 – Apr 2022

Yongin, Republic of Korea

Upstage AI is a Korean AI company focused on enterprise AI. I joined as a research engineer on the Document AI team during its early phase, working on the foundational models that later became part of Upstage's document-understanding product line.

LayoutLM training and fine-tuning

- Trained and fine-tuned LayoutLM — a multimodal transformer that jointly models page layout and text — for Korean document understanding.

OCR labeling dataset

- Built the labeling dataset used to train the document understanding models.

Document parser

- Built a parser for document processing.

AI Engineer (Contract)

Brunel AI

Sep 2020 – Jun 2021

Mapo-gu, Republic of Korea

Brunel AI is a startup that builds AI-powered search products for patents. This was a formative experience in thinking about AI from a product perspective — the models I trained shipped into the actual product and came back with real user feedback.

- Helped scope research problems into AI tasks that could be realized as real product features (e.g., deciding which data to collect for long-term value, and framing user problems as concrete ML tasks).
- Improved search quality by developing a model that recommends relevant patents from a user's input query.
- Built the supporting infrastructure for the data pipeline and model serving.

AI Research Internship

NAVER Clova AI

Dec 2019 – Jun 2020

Seongnam, Republic of Korea

NAVER Clova AI is one of Korea's leading AI research organizations. I worked in the LaRva team on Korean language modeling, which is where I built deep practical experience with large-scale training on distributed GPU clusters. In six months I was able to contribute to several tasks at once, thanks to the team's resources and mentorship.

Pretraining BERT / RoBERTa on distributed GPU cluster

- Trained on up to 64 V100 GPUs with large-scale corpora in a distributed setup.
- Ran hundreds of pre-training experiments and learned how to avoid CPU/GPU bottlenecks, use mixed precision effectively, and manage large training datasets.

Efficient and lightweight pretrained language models

- Studied how to apply Lample's Product Key Memory (PKM) to pretrained LMs while avoiding catastrophic drift during Masked-LM pretraining and downstream fine-tuning. The work was accepted to Findings of EMNLP 2020.
- Applied knowledge distillation to build a lightweight Korean RoBERTa.

KorQuAD 2.0 leaderboard — 1st place (F1/EM 83.54 / 66.95)

- KorQuAD 2.0 is a reading-comprehension dataset in the spirit of Google's Natural Questions: the model has to find the correct answer span inside a full Wikipedia page, where the answer can be a table, a list, or a paragraph — much harder than single-paragraph QA.

Platform is a startup that builds products that recommend emoji based on chat text. This was my first industry role and where I first learned how an engineering team collaborates day to day.

- Annotated and cleaned a raw chat corpus for training.
- Built a sentiment-analysis-based emoji recommendation model.

PUBLICATIONS

Large Product Key Memory for Pretrained Language Models

Gyuwan Kim*, Tae Hwan Jung* (*equal contribution)

Findings of EMNLP 2020 · [PDF](#) · [Code](#)

CommitBERT: Commit Message Generation Using Pre-Trained Programming Language Model

Tae Hwan Jung

ACL NLP4Prog Workshop 2021 · [PDF](#) · [Code](#)

SELECTED PROJECTS

nlp-tutorial — Natural Language Processing Tutorial for Deep Learning Researchers	★ 14,882
nlp-roadmap — Mind map and keyword roadmap for NLP learners	★ 3,264
distribution-is-all-you-need — Tutorial on basic probability distributions in deep learning	★ 1,635
gpt-2-Pytorch — Minimal PyTorch implementation of OpenAI's GPT-2	★ 1,013
xlnet-Pytorch — Simple XLNet implementation with a PyTorch wrapper	★ 581
commit-autosuggestions — AI that automatically recommends commit messages (CommitBERT)	★ 389
ai-docstring — VS Code extension — automatic Python docstring generation via NLP	★ 313
matorage — Object storage manager for tensors (PyTorch, TensorFlow, Keras)	★ 73
clober-dex/core — On-chain central limit order book engine (Solidity)	★ 11
clober-dex/v2-sdk — TypeScript SDK for Clober DEX v2	★ 9