Heungsub Lee

Contact
heungsub@subl.ee

Web Sites
subl.ee, github.com/sublee, linkedin.com/in/sublee

Interest

Software architectures, distributed systems, cost optimization, developer experience, and the open source culture.

Skills

Programming Languages
Go, Python, TypeScript, Bash

Service Development
Linux, AWS, K8s, Pulumi, React, gRPC, ZeroMQ, NoSQL, Concurrent programming, Testing

ML Engineering
PyTorch, NVIDIA Nsight Systems, NCCL

Work Experience

Software Engineer
Global AI Platform, Sep 2023 – Present
Developing Aster, a planning-oriented AI agent service.

Software Engineering Manager
NAVER, Aug 2020 – Jul 2023
Supervised MLOps platforms by leading 25 software engineers to optimize the inference performance and productivity of HyperCLOVA, an LLM specializing in Korean culture.

Developed NSMLv2, a large-scale ML research platform in CLOVA. Designed its multitenant architecture based on economics to share GPU clusters for HPC among diverse organizations with complex desires while utilizing GPU resources efficiently.

Software Engineer
Kakao Brain, Dec 2018 – Aug 2020
Developed and published a pipeline parallelism library named torchgpipe in open source.

Developed a serverless training framework and a distributed hyperparameter search platform for an AutoML service.

Game Server Engineer
NEXON, Mar 2011 – Dec 2018
Developed cloud-based distributed MMORPG servers for Durango using pub/sub communication over the spatial grid system. Achieved up to 70k concurrent users per game world.

Developed online racing game servers and matchmaking for KartRider Dash and KartRider Coin Rush.

Back-end Web Developer
nPine, Dec 2008 – Feb 2011
Developed web services selling stock images.

Open Source Experience

torchgpipe, Feb 2019 – Apr 2020
Implemented GPipe, a multi-GPU pipeline parallelism technique for training giant models, as a PyTorch library with optimization for CUDA, the autograd engine, and long skip connections. This project has become a part of PyTorch as Pipe APIs

Hangulize, Oct 2010 – Present
Invented a Hangul transcription algorithm and served as a web tool at zero cost. Many professional Korean translators use this tool to translate undocumented proper nouns.

TrueSkill, Jan 2012 – Dec 2015
Implemented TrueSkill™, the rating algorithm for Xbox Live, as a Python library. This project was introduced in PyData Berlin 2019.

Contributions

- For PyTorch, fixed potential GPU memory violation (#27371); deprecated an inconsistent API (#21006, #25985).
- For Flask, fixed a bug to generate a URL with a subdomain (#108).
- For jQuery 1.4.3, fixed a bug on content negotiation in Ajax requests.

Publications

- B. Kim et al., "What changes can large-scale language models bring? Intensive study on HyperCLOVA: Billions-scale Korean generative pretrained Transformers," arXiv:2109.04650, Sep 2021.
- C. Kim*, Heungsub Lee* et al., "torchgpipe: On-the-fly pipeline parallelism for training giant models," arXiv:2004.09910, Apr 2020.

*Contributed equally

Public Speeches

- "NSML, the hyper-scale ML training platform," KRnet, Jun 2022.
- "Remake of Hangulize," Golang Korea Meetup, Aug 2018.
- "Profiling," PyCon Korea, Aug 2015.
- "The server architecture of Durango," NDC, 2014, 2016, and 2018.

Languages

- Korean — Native
- English — Conversant in reading and writing

Education

Computer Software, Kwangwoon University, 2008, Completed the first year only.

(Last updated on Jun 17, 2025)