

EXPERIENCE

- **Toss Bank** Seoul, South Korea
Data Scientist *Jan 2026 - Present*
 - **On-premise AI Agent System:** Building an on-premise opencode agent system for internal AI-assisted development workflows in a banking environment.
 - **AI-based Authentication:** Developing authentication models and pipelines for face verification and ID card verification.
 - **Document Information Extraction:** Developing document AI systems to extract structured information from banking documents.
- **SuperbAI** Seoul, South Korea
Machine Learning Engineer *Sep 2021 - Jan 2026*
 - **[Lead] Vision Foundation Model:** Spearheaded a multi-modal vision foundation model that accepts text and semantic prompts for object detection and segmentation; led two ML engineers through planning, dataset construction, training, API deployment, and AWS Marketplace release.
 - **Dataset and Evaluation:** Prepared a 1.1M-image in-house dataset and used automated model-assisted verification to improve labeling efficiency; won 2nd place in IOD and 4th place in FSOD at CVPR 2025 challenges.
 - **[Lead] ML Training & Serving Platform:** Designed AWS Batch based training infrastructure and optimized production inference with TensorRT and Triton Inference Server, improving throughput by 5x over pure PyTorch serving.
 - **Scale:** Generated approximately 1,100 machine learning models and 60 customer endpoints in one year across customer tenants.
 - **Efficient Training:** Applied parameter- and memory-efficient methods for large models, reducing GPU memory usage by 65.6% and training time by 44.3% while preserving model performance.
 - **Vision Language Model:** Developed a flexible pipeline for applying LoRA, Adapter, and LST methods to reduce computational cost and participated in the CVPRW 2023 NICE challenge.
 - **Web ML:** Developed browser, hybrid, and cloud-based interactive segmentation tools using RepViT-SAM, FocalClick, SAM, and SAM2; increased segmentation labeling speed by 25x and built video tracking with fine-tuned personalized SAM.
 - **Transfer Learning:** Developed pretraining and parameter-efficient transfer learning algorithms using DETR architectures and CBNetV2; built multi-domain datasets such as night and aerial imagery with box and polygon annotations.
 - **Uncertainty Estimation:** Implemented Monte Carlo dropout for keypoint and segmentation models, reducing human inspection effort to 30% by ranking predictions with top uncertainty scores.
- **Kakao Enterprise** Pangyo, South Korea
Machine Learning Engineer Intern *Feb 2021 - Aug 2021*
 - **AutoML:** Fixed the scaling issue of AutoGluon's NeuralNetFastAI algorithm and developed a Flask-based training/inference AutoML framework with a simple front-end.
- **Team ISLAND** Seoul, South Korea
Chief Technology Officer *Mar 2019 - Oct 2020*
 - **Startup:** Created an MZ video production platform based on scribble effects; led development, investor relations, and product management with a five-person team across Android, Unity, and deep learning.
 - **Fundraising:** Prepared investor pitch materials and helped raise approximately \$290K through government subsidies and seed funding.
 - **Mobile ML:** Developed a lightweight 3D human pose estimation model for mobile devices and converted PyTorch models to TensorFlow Lite.
- **Seerslab Inc.** Seoul, South Korea
Software Engineer Intern *Jul 2018 - Aug 2018*
 - **Annotation Tool:** Developed a facial landmark annotation tool using D-lib, Python, and PyQt.

OPEN SOURCE

- **Hugging Face Transformers - Segment Anything 2**: Led the integration of SAM2 into Transformers across model architecture, image/video processing, conversion scripts, documentation, and integration tests. The PR was merged on 2025-08-13 after a long community review cycle and is credited in the official [SAM2 documentation](#).
- **Hugging Face Transformers - Molmo2**: Opened Molmo2 support for Transformers, including model integration work for a multimodal VLM family, and published [danelcsb/Molmo2-4B](#) on the Hugging Face Hub.
- **RT-DETR**: Implemented RT-DETR support in Transformers.
- **ViTPose**: Implemented ViTPose support in Transformers.
- **Training and Distributed Fixes**: Fixed DETA auxiliary training loss issue and improved gathering behavior in multi-GPU training pipelines; additional contributions are available in [Transformers pull requests](#).

PAPERS

- **ZERO: Multi-modal Prompt-based Visual Grounding**: arXiv - Jul 6, 2025
- **MobileHumanPose: Toward real-time 3D human pose estimation in mobile devices**: IEEE/CVF CVPR Workshop - Jun 19, 2021

PRESENTATIONS

- **Building Vision Foundation Model in Startup**: Superb AI First Online Seminar - Oct 29, 2025
- **Automation in AutoLabeling**: MLOps World: Machine Learning in Production Bay Area - Jun 19, 2022

PROJECTS

- **Stable Diffusion**: Generated personalized portraits and customized styles by fine-tuning diffusion weights with LoRA.
- **iOS Inspection App**: Built an in-house iOS data inspection function for project managers and distributed it through TestFlight.
- **Binaural Sound**: Generated a human evaluation metric for binaural sound generation.

EDUCATION

- **KAIST (Korea Advanced Institute of Science and Technology)** Daejeon, South Korea
Professional Master in Electrical Engineering; GPA: 3.55/4.3 *Feb. 2020 – Aug. 2021*
- **POSTECH (Pohang University of Science and Technology)** Pohang, South Korea
Bachelor in Electrical Engineering; GPA: 3.4/4.3 *Mar. 2015 – Feb. 2020*
- **UIUC (University of Illinois Urbana-Champaign)** Illinois, United States
Exchange Student; GPA: 3.6/4.0 *Jan. 2018 – Dec. 2018*
- **Sejong Science High School** Seoul, South Korea
One year early graduation *Mar. 2013 – Feb. 2015*

PROGRAMMING SKILLS

- **Professional**: Python, PyTorch, AWS, Transformers **Experienced**: TensorFlow, Swift, LangChain, Langfuse