Last updated in 2026.03.20

Suhwan Choi

Seoul, South Korea | milkclouds00@gmail.com | Suhwan Choi | suhwanchoi.me | milkclouds

Summary

I am an undergraduate student majoring in Physics and Computer Science at Seoul National University. My main area of interest is approximating and imitating human behavior and intelligence in multmodal modalities, utilizing end-to-end architectures and scalable training suites.

Experience

Principal Researcher *Feb 2024 – Present*
Maum.ai

- Founded autonomous robotics research division as the first researcher, leading strategic decisions and team expansion to 10 researchers.
- Contributed as first author to majority of research projects in robotic navigation and embodied ai.
- Led CORE: Slurm-based DGX Cluster construction project (96 H100 GPUs, 12 nodes). [Blog]
- Implemented company-wide Notion workspace enhancing productivity, streamlining workflows. [Template]

Machine Learning Engineer Intern *July 2023 – Jan 2024*
Hyperconnect

- Worked on diffusion-based personalized profile image generation for real-world applications. The output images were conditioned not only on prompts but also on the face in the input image.

Publications

vla-eval: A Unified Evaluation Harness for Vision-Language-Action Models [Code] — 1st author
Suhwan Choi, Yunsung Lee, Yubeen Park, Chris Dongjoo Kim, Ranjay Krishna, Dieter Fox, Youngjae Yu
arXiv 2026

D2E: Scaling Vision-Action Pretraining on Desktop Data for Transfer to Embodied AI [Website] — 1st author
Suhwan Choi*, Jaeyoon Jung*, Haebin Seong*, Minchan Kim, Minyeong Kim, Yongjun Cho, Yoonshik Kim, Yubeen Park, Youngjae Yu†, Yunsung Lee†
ICLR 2026

CostNav: A Navigation Benchmark for Real-World Economic-Cost Evaluation of Physical AI Agents [Website] — 9th author
Haebin Seong*, Sungmin Kim*, Yongjun Cho*, ..., **Suhwan Choi**, ..., Youngjae Yu, Yunsung Lee (23 authors)
arXiv 2025

Revisiting Residual Connections: Orthogonal Updates for Stable and Efficient Deep Networks — 4th author
Giyeong Oh, Woohyun Cho, Siyeol Kim, **Suhwan Choi**, Youngjae Yu†
NeurIPS 2025

CANVAS: Commonsense-Aware Navigation System for Intuitive Human-Robot Interaction [Website] — 1st author
Suhwan Choi*, Yongjun Cho*, Minchan Kim*, Jaeyoon Jung*, Myunchul Joe, Yubeen Park, Minseo Kim, Sungwoong Kim, Sungjae Lee, Hwiseong Park, Jiwan Chung, Youngjae Yu†
ICRA 2025
Workshop on Open-World Agents at NeurIPS 2024 **(Outstanding paper Awards, 3%)**

ESREAL: Exploiting Semantic Reconstruction to Mitigate Hallucinations in Vision-Language Models — 4th author
Minchan Kim*, Minyeong Kim*, Junik Bae*, **Suhwan Choi**, Sungkyung Kim, Buru Chang†
ECCV 2024

Education

Seoul National University *2021 – Present*
B.S in Physics & Computer Science and Engineering

Awards

QHack Coding Challenge [link] *2023 and 2024*

- Ranked **4th/793** teams in 2023, Ranked **3rd/618** teams in 2024.
- Contest implementing quantum algorithms, quantum machine learning, quantum chemistry, and brain-teasing puzzles.

2023 Quantum Hackathon [link] *2023*

- **1st place, Minister of Science and ICT Award**
- Utilizing symmetry to solve variational quantum algorithm (quantum machine learning) efficiently.

NAVER CLOVA AI RUSH 2022 [link] *July 2022 – Sept 2022*

- **3rd place** on Landmark Detection (3,000,000 KRW)
- **2nd place** on Shopping User Embedding Extraction, Classification (7,000,000 KRW)

Activities

Open Source Projects

- **allenai/vla-evaluation-harness**: Primary author. Unified evaluation framework for VLA models across robot simulation benchmarks with 47x throughput via batch parallel evaluation, Docker-isolated benchmarks, and the largest unified VLA leaderboard (500+ models × 17 benchmarks).
- **MilkClouds/awesome-vla-study**: Curated reading list on VLA models – from diffusion/flow matching foundations through robot foundation model architectures to data scaling, RL fine-tuning, and world models.
- **open-world-agents**: Core contributor and maintainer with 180+ merged PRs. Built comprehensive multimodal desktop agent framework including optimized data collection tool (ocap), standardized efficient data format (OWAMcap), dataset visualizer, multimedia data management/processing pipelines, agent training and also Python packaging, CI/CD infrastructure.
- **open-world-agents/MediaRef**: Pydantic media reference for images and video frames with lazy loading and optimized batch video decoding.
- **MilkClouds/vla0-trl**: Unofficial reimplementation of VLA-0 using TRL's SFTTrainer. While common VLA codebases are over 10,000 lines, vla0-trl contains only 1,200 lines total. Gets 92% on LIBERO by just fine-tuning Qwen2.5-VL to predict actions as text, without any custom architecture.
- **MilkClouds/smon**: Real-time Slurm cluster monitoring TUI. Visualizes GPU/CPU/memory allocation with job-level drill-down.

Open Source Contributions

- **linkedin/Liger-Kernel** (★6k): Added InternVL3, SmolVLM2 support (**#878**, **#919**).
- **huggingface** (★155k+): Fixed multimodal processor bugs (**transformers#41864**, **#41865**, **#41871**), added ImageProcessor API (**#39603**), PeftModel warning (**#32085**), SDXL batch fix (**diffusers#5807**).
- **open-mmlab** (★40k+): Fixed distributed training deadlocks (**mmdetection#9476**, **mmsegmentation#2398**)

Deepest: SNU Deep Learning Society [Website] *Sept 2022 – Present*

- Participated in research projects: neural ODE study, RL-based game solving, parallel computing, causal inference, voice generation

Skills

Programming Languages: Highly proficient in Python / Experience with: C++, Java, HTML/CSS/JS, Golang, Rust, ...

ML/AI: PyTorch, HuggingFace (Transformers, Diffusers, LeRobot, ...), open-mmlab libraries, ROS

Infrastructure & DevOps: Linux, Slurm, Docker, Ansible, Ceph, GitHub Actions CI/CD, HPC infrastructure

Languages: Korean (Native), English