Youngmin Ko

Seoul, South Korea | rhdudals0505@naver.com | github.com/mandu5 | mandu05.com

Education

Pennsylvania State University — Aug 2026
B.S. in Information Sciences and Technology — GPA: 3.68 / 4.00 (Major: 3.87)

Work Experience

AI Forward Deployed Engineer — Jun 2026 – Present
KRAFTON, AX FDE Team — Seoul, South Korea

- Built AI-enabled meeting and customer-support workflows for game publishing operations, connecting product interfaces with enterprise systems.
- Designed grounded outputs and user authorization, with approval-bound actions and recoverable execution.
- Jointly improved customer-feedback analysis and finance integrations with business owners, using evidence-based evaluation.
- Supported business developers with architecture reviews, platform migration and deployment.

Research Engineer Intern — Jan 2026 – Feb 2026
Hanwha Aerospace — Pangyo R&D Center

- Designed and built, alone in eight weeks, a system that detects communication anomalies between unmanned ground vehicles and their control station, turning an operator's gut call into measurable signals such as latency, jitter, and packet loss.
- With no labeled data, rules decide first and Isolation Forest joins in, so detection works from day one; each alert shows which signals moved and by how much, so operators can check it.

Research

Replay Strategy Selection in Continual Learning (sole author) — 2026

- What Ten Seeds Can and Cannot Resolve About Replay Strategy Selection Under Distribution Shift. Accepted at a NeurIPS 2026 workshop (CL4FMAgents).

Verifying Claims in ML-Engineering Agent Runs — 2026

- Matched Twins Show What an Evidence Contract Cannot Carry to a Verifier of ML-Experiment Claims. Preprint, Zenodo, Nov 2026 (doi:10.5281/zenodo.23933494). Paired runs that differ only in a label leak: re-execution separates them, the evidence a verifier reads barely does (best field AUC 0.67).

Evaluation of Time-Series Anomaly Detection Metrics (sole author) — Oct 2025 – Jun 2026

- Showed that the often-cited 31% rank disagreement between metrics sits below its 50% chance level; where both metrics clearly separate two models it is 2% (25 models, 180 series).
- Held my own earlier metric to the same test and reported its limits. tsad-eval-site.onrender.com

Selected Projects

sessionreel: Claude Code sessions as short videos | open source, PyPI — Sep 2026

- Turns a long agent session into a short video. A reviewer agent run over 29 real sessions found three kinds of false success, each checked against the logs and fixed. github.com/mandu5/sessionreel

Honors and Talks

Invited Panelist, Koderunner 2026 (KakaoBank): putting AI agents to work inside a company — Aug 2026
Dean's List, College of Information Sciences and Technology, Pennsylvania State University — 2025–2026
Richard and Claudia Fischer Open Doors Scholarship ($3,000, merit-based) — Summer 2026
LG Aimers: 7th cohort, top 4.2% of ∼1,500 teams; 5th cohort, top 5.5% (product defect detection)
Google DeepMind Hackathon Seoul: EquiLaw, 20th of 1,500 teams — Feb 2026

Technical Skills

AI & Agents: Claude API, Claude Code, Codex, MCP, RAG, PII de-identification, QLoRA
Engineering: Python, TypeScript, SQL, FastAPI, Next.js, PostgreSQL, Docker, Kubernetes, GitHub Actions
Evaluation & ML: golden-set design, regression gates, bootstrap evaluation, scikit-learn, PyTorch, XGBoost