Cheul Young Park

Seoul, Korea | cheulyop@gmail.com | cheulyop | cheulyop | cheul.dev | 0000-0003-0414-272X

Education

Korea Advanced Institute of Science and Technology — Daejeon, Korea
Master's in Knowledge Service Engineering — *Aug 2018 — Feb 2021*
- Thesis: *"Towards Adaptive Sampling of Emotions in the Wild with Active Learning"*

Carnegie Mellon University — Pittsburgh, USA
Bachelor of Science in Economics — *Aug 2010 — May 2017*
- Science and Humanities Scholar, Additional Major in Human-Computer Interaction, Minor in Physics

Work Experience

Software Engineer — Jun 2023 — Present
Upstage AI — *Seoul, Korea*
- Automated qualitative evaluation for LLMs: built Slack bot with interactive UI for evaluation orchestration, integrated AWS ECS/Fargate for scalable evaluation runs, implemented PostgreSQL-backed scoring system across multiple benchmarks; enabled team-wide one-click model assessment beyond traditional metrics
- Developed Human Feedback System (HFS): built Streamlit-based annotation platform supporting multiple task types, implemented role-based access control, bulk task assignment system, and real-time feedback collection; enabled systematic collection of human annotations to improve the Retrieval-Augmented Generation (RAG) performance of an LLM
- Built chat backend for virtual K-pop idol *MAVE*: architected FastAPI microservices, implemented DynamoDB-based user memory for personalization, deployed LLMs fine-tuned for persona-driven dialogue; served 3,000+ users and 50,000+ interactions during 3-week global beta
- Automated an insurance-evaluation pipeline for Hanwha Life using RAG to streamline underwriting assessments

Data Manager — Apr 2022 — Jun 2023
Upstage AI — *Seoul, Korea*
- Developed Python library to automate OCR data management: implemented data format conversions between OCR and key-value extraction, built dataset versioning with automatic train/val/test splits, created validation pipeline for annotation quality, handled multi-page document processing; standardized data workflow across teams
- Led a data standardization committee that defined and implemented policies for OCR & key-value extraction data formats to ensure data consistency and streamline pipeline workflows

AI Researcher — Feb 2021 — Mar 2022
Silvia Health — *Seoul, Korea*
- Developed voice conversion framework using AutoVC and attention mechanisms to extract speaker-independent embeddings for cognitive impairment detection
- Built and deployed ML backend API for mobile app, implementing real-time speech analysis pipeline with acoustic feature extraction and cognitive scoring

Selected Publications

[1] Cha, I., Oh, J., Park, C. Y., Han, J., Lee, H. *"Unlocking the tacit knowledge of data work in machine learning"* **CHI EA** (2023) [abs]

[2] Heo, D., Park, C. Y., Cheun, J., Ko, M. J. *"Separating content from speaker identity in speech for the assessment of cognitive impairments"* **arXiv preprint** (2022) [abs]

[3] Park, C. Y. et al. *"K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations"* **Scientific Data, Nature** 7.1 (2020) [abs, doi]

Skills

- **Programming Languages**: Python, JavaScript/TypeScript, SQL, Bash, HTML/CSS, Go
- **Backend/Infrastructure**: Docker, Kubernetes (EKS), AWS, GCP, BentoML, Streamlit
- **Tools & Frameworks**: PyTorch, Git, UNIX, React, Tailwind CSS, FastAPI, PostgreSQL