Ui Hyun Cho

Experienced in Large Language Model Product Developments

Seoul, Republic of Korea
(+82) 10-4886-1937
cuihyun12@naver.com

Education

Yonsei University, Seoul — *Fourth Year*

March 2019 - Feb 2026

Bachelor of Engineering, Industrial Engineering (1st)

Bachelor of Statistics, Applied statistics (2nd)

GPA: 3.94/4.30 (15th out of 106)

Relevant Courses: A+ in IC-PBL Projects, Generative Models, Reinforcement Learning, Text Mining, Mathematical Statistics, SW Programming, etc.

Experience

Yonsei University, Seoul — *Part Time Instructor*

June 2024 - Present

Instructor of Online Course "Prompt Engineering and LangChain"

- Consists of 8 lectures on prompt engineering
- Available to all students at course website, LearnUs
- Provides lecture notes and codes with annotations
- Syllabus: https://tinyurl.com/yeyk6e2x

Samsung Electronics Device Solution, Hwaseong — *Intern*

Sep 2024 - Dec 2024

Internship at Software Development Team Solution P.E. Team

- Contributed to the development of intellectual assets LLM Chatbot
- Implemented Retrieval Pipeline to existing chatbot system, reducing the amount of false generation by 45%
- Used Programs: tmux, huggingface TEI, vllm, meilisearch, mongoDB, APISIX, grafana, prometheus

Research Interests

Generative Models

Large Language Models

Reinforcement Learning

Awards

Academic Scholarship
- freshmen, 1st & 2nd semester
- sophomore, 2nd semester
- junior, 1st & 2nd semester
- senior, 1st & 2nd semester

2023 Big Data Competition
- Grand Prize
- Proposed marketing solution tackling declining sales of Real Estate Enterprise based on statistical insights derived from survival data analysis -
Article Link: https://tinyurl.com/4ttd7n86
Used Skills :
1. Survival Analysis
2. Data Analysis
3. Data Preprocessing
4. Web Scraping

2021 Nexon root-i Creative Platform -
Excellence Prize

2020 Social Impact Championship -
Grand Prize
- Article Link: https://tinyurl.com/5n7efa7z

WeSome, Seoul — *Developer*

Apr 2024 - Dec 2024

Participant of Pre-Startup Packages Program

- Responsible for development and distribution of LLM induced sentiment tagger and synthetic diaries
- In charge of linking model API to mobile application
- Accumulated $32,500 amount of funding from investments

Research Experience

Agent-Integrated Denoising Experts : — *First Author*

- Proposed how an LLM-driven agentic framework (A-IDE) enhances LDCT denoising by routing scans to anatomy-specialized experts
- Paper under Review
- Paper Link: https://arxiv.org/pdf/2503.16780

Anatomical Prior Driven LDCT Denoising: — *First Author*

- Proposed BioAtt, a novel framework that enhances LDCT denoising by guiding attention with anatomical priors without increasing model complexity.
- Paper under Review
- Paper Link: https://arxiv.org/pdf/2504.01662

Development of a Depression Prevention Platform using Multi-modal Emotion Recognition AI Technology: — *Second Author*

- Proposed a multi-modal emotion recognition model that boosts accuracy to 88.0% by fusing text and speech features.
- Annual Conference of KIPS Short Paper
- Paper Link:https://tinyurl.com/3ytuxbkk

Projects

RAG Optimization using PPO-Clip — *Team Leader*

Apr 2024 - Jun 2024

Reinforcement Learning experiments and presentation

- aimed to enhance efficiency of the RAG pipeline by applying PPO-Clip reinforcement learning algorithm.
- involves clustering and optimization to reference specific clusters using a gating mechanism controlled by PPO-Clip
- Project Report Link: https://tinyurl.com/mpuembfr
- Presentation Link: https://tinyurl.com/2p8y5y3f

Languages

Native in Korean Language

Fluent in English

- OPIC: AL
- TOEFL IBT: 112 (expires at Nov 2024)
- TOEIC: 975 (expires at Dec 2024)

Comento, Industry Academic Project — *Head Manager*

Oct 2023 - Jan 2024

On Development of resume builder assistant LLM pipeline

- Head Student Manager of total 14 participating students
- Improved LLM output alignment by RAG pipeline from exemplary resumes provided by the company
- Contents under license

Rotary Embedded Diffusion Transformers — *Team Leader*

Nov 2023 - Dec 2023

Generative Models experiments and presentation

- aimed to enhance efficiency of the Diffusion Transformers by applying the Rotary Embedding encoding algorithm.
- involves replacing absolute positional embedding codes into rotary embeddings along with test results
- Presentation Link: https://tinyurl.com/3watupt6

Denoising Diffusion Probabilistic Models — *Team Leader*

Sep 2023 - Oct 2023

Generative Models paper applications and presentation

- Reviewed the content of popular diffusion paper
- Presentation Link: https://tinyurl.com/ms3auarv

Naver Financial, Industry Academic Project — *Assistant*

Apr 2023 - Jun 2023

On Development of Text2SQL LLM pipeline

- Contributed to the pipeline by attempting several few-shot prompt engineering techniques
- Contents under license