

HOJUN CHOI

AI Researcher, KIM JAECHUL GRADUATE SCHOOL OF ARTIFICIAL INTELLIGENCE, KAIST

85 Hoegi-ro, Dongdaemun-gu, Seoul 02455, South Korea

☎ (+82) 10-7185-1250 | ✉ eric970412@gmail.com | 🏠 <https://github.com/hchoi256> | 🔗 <https://www.linkedin.com/in/hojun-choi-2b10b11a0/>

Research Interests

Computer Vision

Machine Learning

Model Compression

Hyper-scale AI

Tech Stack

Programming Languages

Python | Java | C | R | C++ | C# | SQL

Frameworks

PyTorch | TensorFlow | Apache | .NET | Android | Flask | Node.js

Education

The Kim Jaechul Graduate School of AI at Korea Advanced Institute of Science and Technology

South Korea

M.S; Major in ARTIFICIAL INTELLIGENCE;

Jan. 2024 - May. 2025

- GPA: 3.85/4.0;

University of Wisconsin-Madison

Madison, WI

B.S; Major in COMPUTER SCIENCES; Major in DATA SCIENCE;

Sep. 2020 - Dec. 2022

- GPA: 3.863/4.0; Dean's List 2020-2022;

University at Albany-SUNY

Albany, NY

Attended; Major in COMPUTER SCIENCES;

Jan. 2017 - Dec. 2017

- GPA: 3.9/4.0; Dean's List 2017;

Publication

CONFERENCES

- C1 S. Lee, **H. Choi**, J Choi, J Lee, S Oh, S Ko, and I Shin. "Explore, Select, Derive, and Recall: Augmenting LLM with Human-like Memory for Mobile Task Automation," The International Conference on Mobile Systems, Applications, and Services (**MobiSys**) 2024 Submission.
- C2 S. Park, **H. Choi**, and U Kang. "Accurate Retraining-free Pruning for Pretrained Encoder-based Language Models," The International Conference on Learning Representations (**ICLR**) 2024 Poster.

Honors and Awards

2022	Top 10% in Predicting Autonomous Sensor Antenna Performance, LG Research AI Hackathon	South Korea
2022	Nominated for Developing User-friendly Online History Search Chrome Extension, Hatathon	Madison, WI
2022	Dean's List for All Semesters Attended in Recognition of Exceptional Academic Excellence	Madison, WI
2017	Dean's List for All Semesters Attended in Recognition of Exceptional Academic Excellence	Albany, NY

Professional Experience

FLUIZ

Seoul, South Korea

AI Intern Researcher, Advisor: Prof. Insik Shin

Sep. 2023 - Feb. 2024

- Cooperated to develop a research proposal for an Autonomous Agent based on open-source VLM.
- Spearheaded the development of an efficient tree-based memory structure for representing complex app architectures using Selenium and ChatGPT API for Web/Mobile Robotic Process Automation (RPA) technology.

- Led the development of UI Mirroring technology for interacting with data from various web pages within the LS Cable & System enterprise ERP platform, employing image rendering and coordinate calculation methods instead of iframes, and devising optimization techniques for real-time data video rendering.
- Contributed to the MobileGPT research project using ChatGPT API by proposing an efficient memory structure and developing Random Explore technology for automating Android mobile app processes, leading to a published paper.
- Research Assistant in KTP601: AI Convergence Group.
 - Team A: Assisted in the development of a network facilitating the integration of Vector DBMS and LangChain-powered applications, enabling the generation of arXiv paper summaries.
 - Team B: Assisted in building an end-to-end model architecture for effective resizing and positional adjustments of images, leveraging LLMs, a SoTA object detector of Grounding DINO, and Stable Diffusion.

Seoul National University

Seoul, South Korea

AI Researcher, Advisor: Prof. U Kang

Jan. 2023 - July. 2023

- Specialized in AI model compression, particularly in quantization research; meanwhile, presenting a paper as a poster at ICLR 2024, proposing an offline pruning technique leveraging pre-trained model knowledge to achieve performance comparable to fine-tuning without the need for additional costly training.
- Contributed to a survey paper on lightweight Transformer-based models, focusing on the Quantization domain within model compression techniques, including pruning, KD, etc.
- Completed several graduate-level courses with an A grade or higher:
 - Fundamentals of Mathematics and Statistics for Machine Learning and Data Science.
 - Advanced High-Performance Computing: involved catching on to basics of parallelization like loop-carried dependence and implementing various parallelization algorithms like strip mining using CUDA and OpenMP on a cluster server based on Slurm. The final project was to optimize the GFLOPS of a Transformer-based French-English Translation model.

University of Wisconsin-Madison

Madison, WI

Undergraduate AI Research Intern, Advisor: Prof. Ran

Sep. 2022 - Dec. 2022

- Developed a SoTA object detection architecture with DETR-based models and evaluated their performance within the context of autonomous driving vehicles using a custom CARLA environment.
- Presented a regular meeting on reading papers submitted to 1st-tier conferences such as CVPR related to research topics on Object Detection and Semantic Segmentation.

University of Wisconsin-Madison

Madison, WI

Academic Mentor, Employer: Maisee Her

Sep. 2021 - Dec. 2022

- Actively engaged in serving the UW community and dedicated to developing intercultural leadership and communication skills.
- Diligently mentored my mentees through academic counseling, resource sharing, and a sincere dedication to their success.
- Enhanced proficiency in crucial AI domains including Linear Algebra, Artificial Intelligence, Human-Computer Interaction, and C Programming.

Teaching

RESEARCH ASSISTANT, HYBRID, SEPTEMBER 2023 – DECEMBER 2023

Fall 2023 KT-AI Convergence Education Program, Research Assistant

PRIVATE TUTOR, REMOTE, OCTOBER 2022 – PRESENT

Fall 2023	English Speaking Tutoring, Advanced-level
Fall 2023	Mathematical Statistics with Applications, graduate-level
Fall 2023	Linear Algebra, undergraduate-level
Summer 2023	Mathematical Statistics with Applications, graduate-level
Spring 2023	Mathematical Statistics with Applications, graduate-level
Spring 2023	English Speaking Tutoring, Intermediate-level
Fall 2022	Linear Algebra, undergraduate-level

ACADEMIC MENTOR, UW-MADISON, MADISON, WI, SEPTEMBER 2021 – DECEMBER 2022

Fall 2022	Introduction to Human Computer Interactions (CS570), undergraduate-level
Fall 2022	Introduction to Artificial Intelligence (CS540), undergraduate-level

Spring 2022	Elementary Matrix and Linear Algebra (CS340), undergraduate-level
Spring 2022	Introduction to Computer Systems (CS354), undergraduate-level
Spring 2022	Programming III (CS400), undergraduate-level
Fall 2021	Elementary Matrix and Linear Algebra (CS340), undergraduate-level
Fall 2021	Introduction to Computer Engineering (CS252), undergraduate-level

Professional Activities

COMMUNITY SERVICE

English Translation, NAVER Knowledge iN

Online Freelance English Translation Expert

Remote

May. 2022 - Present

CERTIFICATIONS

NLP Development Program, Seoul ICT Innovation Square

Trainee

Remote

July. 2022 - Sep. 2022

Machine Learning A-Z: Hands-on Python And R, Udemy

Trainee

Remote

Aug. 2022 - Aug. 2022

Machine Learning Practical Development: 8 Practical Projects, Udemy

Trainee

Remote

Aug. 2022 - Aug. 2022

LG Aimers, LG AI Research

Trainee

Remote

July. 2022 - Aug. 2022

AI Engineering Program, Seongnam Industry

Trainee

Remote

July. 2022 - Aug. 2022

Intensive English Language Program (IELP), University at Albany-SUNY

Trainee

Albany, NY

ORGANIZATIONS

CAVH Graduate Student Research Group, University of Wisconsin-Madison

Member

Madison, WI

Computer Science Undergraduate Research Group, University of Wisconsin-Madison

Member

Madison, WI

Student Software Development Club, University of Wisconsin-Madison

Member

Madison, WI

References

References are provided upon request.