

Sumin Yu

ysmssoomin@snu.ac.kr | +82-10-5584-4883

Education

Seoul National University

Mar. 2023 - Present

M.S. in Electrical and Computer Engineering
Advised by Prof. Taesup Moon

Seoul National University

Mar. 2019 - Feb. 2023

B.S. in Electrical and Computer Engineering (GPA: 3.95/4.3)
Advised by Prof. Taesup Moon
Thesis: "Achieving Rawls' Fairness for MRI-based Alzheimer's Disease Prediction"

Research Interest

Safety and trustworthiness in AI; algorithmic fairness, bias, harmful content in generative models, privacy risks; AI governance and accountability.

Publications

(*Marks first-author contribution)

Sumin Yu*, Hyunwoong Bae*, Taesup Moon. "Revisiting VLLM Safety Evaluation: Disentangling Benign Grounding from True Safety Failures in VLLMs." Submitted to the Workshop on Safe and Trustworthy Multimodal AI Systems at International Conference on Computer Vision (ICCV), 2025.

Sumin Yu*, Taesup Moon. "SP-Guard: Selective Prompt-adaptive Guidance for Safe Text-to-Image Generation." the European Workshop on Trustworthy AI at European Conference on Artificial Intelligence (ECAI), 2025.

Sangwon Jung*, **Sumin Yu***, Sanghyuk Chun, Taesup Moon. "Do Counterfactually Fair Image Classifiers Satisfy Group Fairness?—A Theoretical and Empirical Study." Neural Information Processing Systems (NeurIPS), 2024. [paper](#)

Experience

Consultant, SNU AI Ethics Guideline Task Force

Aug. 2025 - Present

Seoul National University AI Policy Initiative (SAPI)

Supervised by Prof. Sangchul Park. Compiled and analyzed domestic and international university guidelines on generative AI; proposed directions and specific recommendations for Seoul National University's policy.

SNU Tomorrow's Edge Membership (STEM), Honor Society

Sep. 2022 - Aug. 2023

Selected based on academic excellence and interview performance. Participated in research seminars and discussions across diverse engineering fields. Provided mentorship to high school students on studying engineering and pursuing research.

Research Intern, M.IN.D Lab, Seoul National University

Sep. 2022 - Feb. 2023

Supervised by Prof. Taesup Moon & Dr. Sangwon Jung. Investigated fairness issues in deep learning classifiers for Alzheimer's diagnosis; developed methods to improve worst-case group accuracy while maintaining overall diagnostic performance.

Student Society for Social Responsibility, SNUSR

Sep. 2019 - Aug. 2020

Member (Sep. 2019 – Aug. 2020); President (Jan. 2020 – Aug. 2020). Collaborated across departments to design social contribution programs; as team leader, directed and supervised all project teams.

Honors & Awards

SNU Summa Cum Laude

Feb. 2023

Dept. of Electrical and Computer Engineering, Seoul National University

Full Scholarship

Mar. 2019 – Feb. 2023

Provided by the Cho Chang-seok Cultural Foundation