Seongjun Yang

wns7169@gmail.com • Google_Scholar • SeongjunYang • @seongjun_yang

Research Interests

Methods for Ensuring Trustworthy AI

- Creating novel methods to ensure users safely achieve intended outcomes with A.I systems like LLM agents under specific goals and constraints. Studying risks such as threats and privacy leaks in A.I systems when handling text and non-text data, and proposing empirically and theoretically robust methods to mitigate the vulnerabilities.

Safety Standards for AI Systems

- Developing an evaluation framework in order to address risks and social impact in generative AI systems, contributing to AI community advancement and policy recommendations. Evaluation for designing accountable AI systems that focus on obtaining high confidence and legibility in the decision-making process. For example, decision-making processes within the legal and healthcare field (that has potential interdisciplinary collaboration) needs reliable AI that is suitable for rigorous standards.

Education

Korea Advanced Institute of Science and Technology — Daejeon, South Korea
Master's degree, Artificial Intelligence — 2020.09 - 2022.08
Advisor: *Prof. Edward Choi*
Thesis: Towards the Practical Utility of Federated Learning in the Medical Domain
Research Area: Federated Learning, Natural Language Processing

Yonsei University — Seoul, South Korea
Bachelor's degree, Computer Science — 2014.03 - 2020.08
Magna cumme laude in Dept. of Computer Science
2 years of absence due to obligatory military service (2015 - 2016)
Club Activities

- IronBats (Engineering College Baseball Club)
- Yupalaw (Yonsei University Department of Public Administration Law Society)
- YCC (Yonsei University Computer Game Club)

Publications

Seongjun Yang*, Gibbeum Lee*, Jaewoong Cho, Dimitris Papailiopoulos, and Kangwook Lee, Predictive Pipelined Decoding: A Compute-Latency Trade-off for Exact LLM Decoding, TMLR 2024

Seongjun Yang*, Hyeonji Hwang*, Daeyoung Kim, Radhika Dua, Jong-Yeup Kim, Eunho Yang, and Edward Choi, Towards the Practical Utility of Federated Learning in the Medical Domain, CHIL 2023

Radhika Dua, **Seongjun Yang**, Yixuan Li, and Edward Choi, Task Agnostic and Post-hoc Unseen Distribution Detection, WACV 2023

Gyubok Lee, Hyeonji Hwang, Seongsu Bae, Yeonsu Kwon, Woncheol Shin, **Seongjun Yang**, Minjoon Seo, Jong-Yeup Kim, and Edward Choi, EHRSQL: A Practical Text-to-SQL Benchmark for Electronic Health Records, NeurIPS 2022 Datasets and Benchmarks

Junu Kim, Kyunghoon Hur, **Seongjun Yang**, and Edward Choi, Universal EHR Federated Learning Framework, In Extended abstract in ML4H 2022

Gyubok Lee*, **Seongjun Yang***, and Edward Choi, Improving Lexically Constrained Neural Machine Translation with Source-Conditioned Masked Span Prediction, ACL 2021 (Short)

Employment History

KRAFTON Inc. Seoul, South Korea
NLP Research Engineer 2022.11 –

- Researching methods to reduce the parameter size of LLMs, such as pruning and quantization, to meet GPU requirements without significant performance loss. Instruct-tuning LLMs, such as LLaMA, and developing prompting strategies for in-game applications.
- For more details, PPD, KORani, AutoEvalGPT

NHN Cloud Seongnam, South Korea
AI Researcher 2022.10 - 2022.11

- Employed as an AI researcher at NHN Cloud. Duties included designing tutorials for benchmarking Korean Language Models.

Korea Advanced Institute of Science and Technology Daejeon, South Korea
Graduate Student Researcher 2020.09 - 2022.08

- Selected to be a Graduate Student Researcher at the Graduate School of AI, Korea Advanced Institute of Science and Technology (KAIST) under the guidance of Prof. Edward Choi. Research areas included Federated Learning and Natural Language Processing.
- **Teaching Assistant** Served as a teaching assistant for courses tiled "Machine Learning for Healthcare ", and "Programming for AI"instructed by Prof. Edward Choi. Duties included leading tutorials for groups of up to 100 students, demonstrating key coding skills, and managing class assignments.

Miscellaneous Experience

Awards and Achievements

Department Prize for Outstanding Student Performance 2019, 2020

- Awarded for achieving grades within the top 3% at Yonsei University.

Graduation Capstone Design 2019

- 3rd Award in graduation capstone design program at Yonsei University.

Industrial Design Competition 2018

- Proposed a system for road damage management
- 3rd Award (Hosted by South Korean Ministry of Trade, Industry and Energy)

Grant

- Nation Scholarship at KAIST 2020.09 - 2022.08
- Full Scholarship at Yonsei 2017.03 - 2020.08

Skills & Interests

Technical Skills

- **Programming Languages**: Python, C/C++
- **Machine Learning Frameworks**: PyTorch, TensorFlow
- **Typesetting**: LaTeX
- **Tools**: Git, Linux

Personal Interests

- **Baseball**: Can play as a shortstop
- **Running**: Learned marathon running from my father, who was a marathon runner in middle and high school
- **Sitcom Enthusiast**: Enjoy watching sitcoms such as "The Big Bang Theory".