

# 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 (Large Language Model) 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 both 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 fields [that has potential interdisciplinary collaboration] that need reliable AI, which is suitable for rigorous standards.

## Education

### Korea Advanced Institute of Science and Technology

Master's degree, Artificial Intelligence

Advisor: *Professor. Edward Choi*

Research Area: Federated Learning, Natural Language Processing

Thesis: Towards the Practical Utility of Federated Learning in the Medical Domain

South Korea

September 2020 - August 2022

### Yonsei University

Bachelor's degree, Computer Science

**Magna cum 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)

South Korea

March 2014 - August 2020

## 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.** South Korea  
**NLP Research Engineer** November 2022 – Present
- Investigate methods to reduce the parameter size of LLMs, such as pruning and quantization, to meet GPU requirements without significant performance loss.
  - Instruct-tune LLMs, such as LLaMA, and developing prompting strategies for in-game applications.
    - For more details,  PPD,  KORani,  AutoEvalGPT
- NHN Cloud** South Korea  
**AI Researcher** October 2022 - November 2022
- Designed tutorials for benchmarking Korean Language Models.
- Korea Advanced Institute of Science and Technology** South Korea  
**Graduate Student Researcher** September 2020 - August 2022
- Researched Federated Learning and Natural Language Processing under the guidance of Professor Edward Choi
- Teaching Assistant** September 2020 - December 2021
- Served as a teaching assistant for courses titled "Machine Learning for Healthcare ", and "Programming for AI"instructed by Professor Edward Choi.
  - Led tutorials, managed class assignments, and demonstrated key coding skills for up to 100 students at a time.

## 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 September 2020 - August 2022
  - Full Scholarship at Yonsei March 2017 - August 2020

## Skills & Interests

### Technical Skills

- **Programming Languages:** Python, C/C++
- **Machine Learning Frameworks:** PyTorch, TensorFlow
- **Typesetting:** L<sup>A</sup>T<sub>E</sub>X
- **Tools:** Git, Linux