Junghyun Min

jm3743@georgetown.edu | Google Scholar | Personal website | Washington, DC | (443) 414-4405

SUMMARY

Computational linguistics PhD student with comprehensive expertise in NLP and language data science across academia and industry. Developed structure-aware generative and discriminative AI solutions for financial documents at NCSOFT, leverage multi-modal LLMs as models of language and linguistic theory at Georgetown University, with research interest in human and computational representations of structure in language. Authored three impactful research publications focused on language model robustness and reliability at ACL, BlackboxNLP, and Repl4NLP. Eager to leverage this extensive research and engineering background to advance cutting-edge AI initiatives both in theory and in practice.

WORK EXPERIENCE

NCSOFT
Seongnam, Korea
NLP Engineer, Financial Language Understanding — Jan 2021 – Apr 2024

- Developed and served fast, light, and accurate chunking system by concatenating word embeddings. Implemented asynchronously with FastAPI in Python. API processes 10k requests per second on <4GB VRAM at 96% acc.
- Designed a granularity control algorithm for financial information extraction, improving performance by 10 points.
- Proposed punctuation restoration as pre-training objective, improving LLM performance across 13 NLU datasets.
- Led mentorships for the Center's 2022 Language AI Global Summer Internship program with 10+ interns.

Harford Community College
Bel Air, MD
Data Analyst, Analytics & Planning — Apr 2018 – Jul 2019

- Distilled expert insight in student retention and success in machine learning models with ~80% accuracy in Wolfram.
- Automated edit checks and recurring data reports in SAS, SQL, leading to 20% increase in request processing volume.

EDUCATION

Georgetown University
Washington, DC
Computational Linguistics Doctor of Philosophy. — Exp. 2029

Johns Hopkins University
Baltimore, MD
Cognitive Science Master of Arts. Focus on Computational approaches to linguistics. — Dec 2020

Johns Hopkins University
Baltimore, MD
Physics Bachelor of Science. Mathematics, secondary major. Quantum mechanics track. — Dec 2017

PUBLICATIONS

- **Junghyun Min**, R. Thomas McCoy, Dipanjan Das, Emily Pitler, and Tal Linzen. ACL 2020. Syntactic data augmentation increases robustness to inference heuristics.
- R. Thomas McCoy, **Junghyun Min**, and Tal Linzen. BlackboxNLP at EMNLP 2020. BERTs of a feather do not generalize together: Large variability in generalization across models with similar test set performance.
- **Junghyun Min**, Minho Lee, Woochul Lee, Yeonsoo Lee. Repl4NLP at NAACL 2025. Punctuation restoration improves structure understanding without supervision. Poster. Tech blog (Korean).

PROJECTS

- Lead engineer for ai.ly, a GPT-2 based, personalized AI lyricist. 50k visits over 3 months of service. Hip-hop sample.
- Technical lead for wecommit's prototype genDOC, an LLM-powered document automation solution for startups.
- WIP: Measuring the redundancy between syntactic structure and prosodic design by leveraging information theory.
- WIP: The first dense Korean Universal Dependencies dataset, with case and adposition and supersense annotation.

SKILLS

Programming Languages: Python (proficient), Java, JavaScript, C++, R, Unix shell, SAS, SQL, Wolfram.
Tools and Libraries: PyTorch, TensorFlow, transformers. Flask, FastAPI, async. OpenAI, GCP, Docker, vLLM, Hydra.
Natural Languages: Korean (Standard, Busan), English, German, Chinese (Mandarin)