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 across academia and industry with interest in hierarchical structure prediction, representation learning, and multimodality and multilinguality in language models. Incorporated hierarchical information to improve generative and discriminative AI solutions for financial documents at NCSOFT; developed low-resource language models, datasets, and NLU benchmarks at the University of Toronto. Authored impactful research publications at ACL, BlackboxNLP, RepL4NLP. Eager to leverage this extensive research and engineering background to advance language AI initiatives both in theory and in practice.

WORK EXPERIENCE

University of Toronto, Toronto, ON
Visiting Researcher, Computer Science — May 2025 – Aug 2025
- Investigate effects of transfer learning for low-resource NLU; open-source produced models, recipes, and benchmarks.
- Supervise student researchers on topics in low-resource languages, machine translation, and cross-lingual transfer.

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.
- Proposed punctuation restoration as pre-training objective, improving 6 NLU systems including OpenIE, NER, SRL.
- Extract financial / biochemical entities and relations for downstream tasks like market sensing, drug discovery.

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 recurring data validation and reports in SAS, SQL, leading to 20% increase in request processing volume.

EDUCATION

Georgetown University, Washington, DC
Ph.D. Linguistics. Advisor: Ethan Wilcox. — Exp. 2029

Johns Hopkins University, Baltimore, MD
M.A. Cognitive Science. Advisor: Tal Linzen. — Dec 2020
B.S. Physics, secondary major in Mathematics. — Dec 2017

SELECTED PUBLICATIONS

- **Junghyun Min**, Minho Lee, Woochul Lee, Yeonsoo Lee. RepL4NLP at NAACL 2025. Punctuation Restoration Improves Structure Understanding without Supervision. Tech blog (Korean).
- **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.

SELECTED 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.

TEACHING

Co-supervisor, CSC494 Computer Science Project, University of Toronto — Fall 2025
Teaching assistant, LING1000 Intro to Language, Georgetown University
Co-supervisor, Fields Undergraduate Summer Research Program, Fields Institute — Summer 2025
Teaching assistant, 110.302 Differential Equations, Johns Hopkins University — Fall 2017

SKILLS

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