

Junghyun Min

jmin10@jhu.edu | [Google Scholar](#) | <https://www.aatlantise.science/> | (443) 414-4405

Summary

Computational linguistics PhD student with comprehensive expertise in NLP across academia and industry. Developed structure-aware language engineering solutions at NCSOFT, contributed to UD-compliant open-source datasets in Korean (k-SNACS) and English (GUM) at Georgetown University. Authored two impactful research publications focused on language model robustness and reliability, delivering insights at ACL and BlackboxNLP. Eager to leverage this extensive research and engineering background to advance cutting-edge AI initiatives both in theory and in practice.

Education

Georgetown University

Doctor of Philosophy, Linguistics. Concentration in Computational linguistics.

Aug 2024 – Exp. May 2029

Johns Hopkins University

Master of Arts, Cognitive Science. Focus on Computational approaches to linguistics.

Jul 2019 – Oct 2020

Johns Hopkins University

Bachelor of Science, Physics. Second major in Mathematics. Early graduation.

Aug 2014 – Dec 2017

Projects and Publications

- ACL 2020 paper, collaboration with Google: Junghyun Min, R. Thomas McCoy, Dipanjan Das, Emily Pitler, and Tal Linzen. 2020. [Syntactic data augmentation increases robustness to inference heuristics](#). TensorFlow and Python.
- BlackboxNLP 2020 paper: R. Thomas McCoy, Junghyun Min, and Tal Linzen. 2020. [BERTs of a feather do not generalize together: Large variability in generalization across models with similar test set performance](#).
- Tech blog: Junghyun Min. 2024. [Punctuation restoration improves structure understanding without supervision \(Korean\)](#). Architecture and train-test pipeline implementation in PyTorch and Lightning. [Abstract](#).
- Project with forus.ai: Technical lead at ai.ly, a GPT-2 based AI lyricist whose results reflect the user's personality. Accumulated 50k visits over 3 months of live service. Its first release: a [hip-hop song](#).

Work Experience

Georgetown University

Graduate Research Assistant, Department of Linguistics

Washington, DC

Aug 2024 - Present

- Expand the [Korean SNACS](#) dataset for UD. Granularize tokenization and complement neural analysis with Linguistics for quality assurance, enhancing dataset's utility for multilingual LLM, parallel corpus studies. Python implementation.
- Provide detailed, multi-layer annotations to add to the Georgetown University Multilayer (GUM) corpus, increasing its size and genre diversity. Leverage GUM for modeling information structure and common ground inventORIZATION.

NCSOFT

NLP Engineer, NLP Center

Seongnam, Korea

Jan 2021 - Apr 2024

- Developed and served fast (10 sentences/ms), light (< 4GB VRAM), and accurate (1p% within SOTA) chunking system by concatenating pre-transformer embeddings. Implemented asynchronously with FastAPI in Python.
- Designed an effective (+10%p) granularity control algorithm in open information extraction with syntax parsing depth.
- Proposed punctuation restoration as unsupervised objective that improves performance across 7 structure-related tasks.
- Led mentorships for the Center's 2022 Language AI Global Summer Internship program with 10+ interns.

Johns Hopkins University

Graduate Research Assistant, Computation and Psycholinguistics Laboratory

Baltimore, MD

Jul 2019 - Oct 2020

- Presented research publications at ACL as first author; BlackboxNLP; and NYAS NLP, Dialog, and Speech.
- Identified unstable vulnerabilities in BERT-based MNLI systems in TF, highlighting areas for improvement in NLU.
- Implemented [adversarial data augmentation](#) through simple syntactic manipulation, significantly enhancing robustness to inference heuristics, stability across initializations, and generalized syntactic sensitivity. Experiment run in PyTorch.

Harford Community College

Data Analyst, Analytics & Planning

Bel Air, MD

Apr 2018 – Jul 2019

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

Skills

Computer Language: Python (proficient), Java, JavaScript, C++, R, Unix shell, SAS, SQL, Wolfram.

Tools and Libraries: PyTorch, TensorFlow, Flask, FastAPI, async, OpenAI. GCP

Natural Language: Korean (Standard, Busan), English, German, Chinese (Mandarin)