Junghyun Min

jmin10@jhu.edu | Google Scholar | https://www.aatlantise.science/

Summary

Experienced computational linguist with a robust engineering background developed over multiple professional and academic roles in both industry and research. Excelled in designing efficient systems, including a high-throughput chunking mechanism and innovative algorithms for open information extraction. Demonstrated success in leading and mentoring teams while contributing to significant advancements reported in top-tier publications. Passionate about leveraging expertise in computational linguistics to drive innovative solutions and achieve project goals effectively.

Education

Georgetown University
Doctor of Philosophy, Linguistics. Computational linguistics concentration. Aug 2024 - Present

Johns Hopkins University
Master of Arts, Cognitive Science. Computational approaches to linguistics focus. Aug 2019 - Dec 2020

Johns Hopkins University
Bachelor of Science, Physics. Second major in Mathematics. Early graduation. Sep 2014 - Dec 2017

Publications

- Tech blog: Junghyun Min. March 2024. Punctuation restoration improves structure understanding without supervision (Korean). Abstract.
- ACL 2020 paper: Junghyun Min, R. Thomas McCoy, Dipanjan Das, Emily Pitler, and Tal Linzen. 2020. Syntactic data augmentation increases robustness to inference heuristics.
- 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.
- Project with forus.ai: Technical lead at ai.ly, a GPT-2 based AI lyricist that reflects the user's personality. 50k visits over 3 months of live service. Its first release: a hip-hop song.

Work Experience

Georgetown University Washington, DC
Graduate Research Assistant, Department of Linguistics Aug 2024 - Present

- Re-annotate the Korean semantic network of adposition and case supersense (k-SNACS) dataset to improve quality and Universal Dependencies compatibility, enhancing the dataset's utility for multilingual, parallel corpus studies.
- Provide detailed, multi-layer annotations to expand the Georgetown University Multilayer (GUM) corpus, increasing its size and genre diversity for effective linguistic analysis and downstream natural language processing applications.

NCSOFT Seongnam, Korea
NLP Engineer, NLP Center Jan 2021 - Apr 2024

- Developed and served fast (10 reqs/ms) and accurate chunking on < 4GB VRAM by concatenating embeddings.
- Designed an effective (+10%p) granularity control algorithm in open information extraction with syntax parsing depth.
- Led mentorships for the Center's 2022 Language AI Global Summer Internship program with 10+ interns.
- Proposed punctuation restoration as an unsupervised objective to improve performance across 7 structure-related tasks.
- Expanded entity, relation extraction to facilitate downstream tasks in financial market sensing, drug discovery.

Johns Hopkins University Baltimore, MD
Graduate Research Assistant, Computation and Psycholinguistics Laboratory Jul 2019 - Oct 2020

- Presented research publications at ACL as first author; BlackboxNLP; and NYAS NLP, Dialog, and Speech.
- Identified instability and vulnerability in BERT-based MNLI systems, highlighting areas for improvement in NLU.
- Implemented adversarial data augmentation through simple syntactic manipulation, significantly enhancing model robustness to inference heuristics, stability across initialization, and generalized syntactic sensitivity.
- Analyzed the emergence of shallow heuristics at all stages of training, proposed solutions to NLP paradigm drawbacks

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

- Distilled expert insight in student retention, student success prediction to explainable ML models with ~80% accuracy.
- Automated edit checks, recurring data requests and reports, resulting in 20% increase in request processing volume.

Skills

Natural & Computer Language: Korean, English, German, Mandarin Chinese, Python, Wolfram, R, Unix shell
Interests: Geography and maps, baseball analytics and sabermetrics, public transport and low-cost travel