

Junghyun Min

Seongnam, Korea | jmin10@jhu.edu | [Google Scholar](#) | <http://www.aatlantise.science/>

EDUCATION

Johns Hopkins University

Master of Arts in Cognitive Science

Baltimore, MD

Aug 2019 – Dec 2020

- Research track; concentration in computational approaches to linguistics; 3.91 GPA

Bachelor of Science in Physics with a second major in Mathematics

Sep 2014 – Dec 2017

- Dean's List Spring 2015, Fall 2017; General Honors; KGSA Baseball Captain; 3.53 GPA

Wolfram Summer School

Waltham, MA

Science Track

Jun 2016 – Jul 2016

- Courses on computational topics such as deep learning, image processing, optimization in Wolfram language

PUBLICATIONS & PROJECTS

- 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](#).
- Master's thesis: **Junghyun Min**. 2020. [The roots and effects of heuristics in natural language inference and question answering models](#).
- Project with forus.ai: Technical lead at [ai.ly](#), a GPT-2 based AI lyricist that reflects the user's preferences. 50k visits over 3 months of live service. Its first release: [hip-hop song](#)
- Computer vision project at Wolfram: **Junghyun Min** and Giorgia Fortuna. 2017. [What on Earth is this Map?](#)

WORK & RESEARCH EXPERIENCE

NCSOFT

Seongnam, Korea

Research Engineer, NLP Center

Jan 2021 – Current

Past projects and assignments

- Embedding concatenation yields accurate chunking model processing 10k requests per sec on < 4GB VRAM.
- Ten naïve rules applied on public dataset significantly improves resulting information extraction performance.
- Simple TF-IDF based importance metric effectively ranks events within a specified temporal window.
- Lead mentor for the Center's Language AI Global Summer Internship program with 20+ interns and mentors.

Current projects

- Engineer prompts to generate high-quality training data for low resource languages with large generative LMs.
- Explore large autoregressive generative language models' understanding of syntactic and semantic structure.

Johns Hopkins University

Baltimore, MD

Graduate Research Assistant, Computation and Psycholinguistics Laboratory

Jul 2019 – Oct 2020

- Authored ACL, BlackboxNLP publications that combine for 207 citations. Presented at ACL and NYAS NDS.
- BERT fine-tuned on MNLI is unstable and vulnerable to syntactic heuristics ([McCoy, Min, Linzen 2020](#)).
- Adversarial data augmentation via syntactic manipulation of training set data significantly increases robustness to augmentation-like examples and general syntactic sensitivity too ([Min, McCoy, Das, Pitler, Linzen 2020](#)).
- Heuristics likely arise from both the pre-training and the fine-tuning dataset. Currently popular fine-tuning and evaluation paradigm has drawbacks that can be patched with longer fine-tuning on unbiased datasets, multi-seed out-of-distribution evaluation, and syntactic adversarial augmentation ([Master's thesis](#)).

Harford Community College

Bel Air, MD

Research Associate, Department of Analytics and Planning

Mar 2019 – Jul 2019

- Improved student retention, success prediction by developing machine learning based predictive models.
- Increased data request processing volume by 20%, by automating edit checks and recurring data requests.

Research Assistant, Department of Analytics and Planning

Apr 2018 – Mar 2019

- Self-taught SAS, SQL, and SPSS to query and respond to internal requests, external compliance reports.
- Facilitated data-driven engagement for non-technical departments by overhauling enrollment data reports.

SKILLS & INTERESTS

Natural & Computer Language

Korean, English, German, Mandarin Chinese, Python, SAS, SQL, Wolfram language, R, Unix

Interests

Geography, cartography; baseball analytics, sabermetrics; urban planning, public transport; low-cost travel.