# Junghyun Min

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

## EDUCATION

**Johns Hopkins University**                                                                                                          **Baltimore, MD**
*Master of Arts in Cognitive Science*                                                                                        *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

**NCSOFT**
- Preprint: **Junghyun Min**, Minho Lee, Woochul Lee, Yeonsoo Lee. Feb 2024. Punctuation restoration improves structure understanding without supervision. Relevant blog post (Korean).
- Preprint: Minho Lee, **Junghyun Min**, Woochul Lee, Yeonsoo Lee. Feb 2024. Structured language generation model for robust structure prediction.

**Johns Hopkins University**
- 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.

**Other**
- 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

## WORK & RESEARCH EXPERIENCE

**NCSOFT**                                                                                                                                **Seongnam, Korea**
*Research Engineer, NLP Center*                                                                                              *Jan 2021 – Current*
- Concatenate embeddings to develop and serve fast (10 reqs / ms) and accurate chunking on < 4GB VRAM.
- Syntax parsing depth is a useful heuristic granularity control in elements of open information extraction.
- Lead mentorships for the Center's Language AI Global Summer Internship program with 10+ interns.
- Punctuation restoration is a simple unsupervised training objective that improves structure understanding.
- Extract financial / biochemical entities and relations for downstream tasks like market sensing, drug discovery.

**Johns Hopkins University**                                                                                                      **Baltimore, MD**
*Graduate Research Assistant, Computation and Psycholinguistics Laboratory*                    *Jul 2019 – Oct 2020*
- Authored ACL, BlackboxNLP publications that combine for 256 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.