Junghyun Min

jm3743@georgetown.edu | Google Scholar | Personal website | Washington, DC | (443) 414-4405

WORK EXPERIENCE

University of Toronto — Toronto, ON
Visiting Researcher, Computer Science — May 2025 – Aug 2025
- Investigate representation learning dynamics; open-source 2 pre-trained models and recipes, 7-task NLU benchmark.
- Supervise 5 student researchers on topics in machine translation, continual pre-training, and cross-lingual transfer.

NCSOFT — Seongnam, Korea
Natural Language Processing Engineer, Financial Language Understanding — Jan 2021 – Apr 2024
- Implemented and trained tokenizers for the Language Model Task Force, precursor to the 1.3-13B VARCO LLMs.
- Developed and served asynchronous Stanza-like NLP API, parsing 10k requests per sec on 4GB VRAM at 96% acc.
- Introduced punctuation restoration as pre-training objective (+11%p), format loss and forced decoding (+8%p) in fine-tuned relation, entity extraction models in financial and biochemical domains for taxonomy and ontology analysis.

Harford Community College — Bel Air, MD
Data Analyst, Analytics & Planning — Apr 2018 – Jul 2019
- Optimized databases, query pipelines, data science projects in Tableau, SAS. Increased processing volume by 20%.

SELECTED TECHNICAL PROJECTS

Co-lead, LLM legal interpretation.
- Implemented LLM legal interp. judgment extraction tool with vLLM. Oral presentation, NLLPW at EMNLP 2025.
- Showed even large (70B) and commercial LLMs are sensitive to surface form, unstably correlated to human judgment.

Lead, Multi-modal text-prosody model for entropy estimation.
- Developed 3-modal transformer architecture for syntactic structure prediction with multi-modal text and audio input.
- Measured entropy reduction in predicting syntax with modal features like text (-74%), duration (-3%), pause (-1.7%).

Lead, Visual grounding and word choice.
- Designed vision-language model pipeline to measure word choice variation by image input in 35k image-caption pairs.
- Proposed VLMs possess super-human sensitivity to visual features, bimodal distribution in preposition grounding.

Other engineering projects.
- **Lead engineer, ai.ly**. Fine-tuned and deployed a GPT-based AI lyricist that generates lyrics tailored to the user's preferences. Accumulated 50k+ visits over 3 months of service and cross-functional collaboration. Hip-hop sample.
- **Lead engineer, genDOC**. Built RAG-based agentic document automation software with LangChain on OpenAI API.

SELECTED PUBLICATIONS

- Minho Lee, **Junghyun Min**, Yerang Kim, Woochul Lee, Yeonsoo Lee. FrontierIR at AAAI 2026. Structured Language Generation Model: Loss Calibration and Formatted Decoding for Knowledge Retrieval and Robust Structure Prediction.
- **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. Collaboration with Google Research.

EDUCATION

Georgetown University — Washington, DC
Ph.D. Computer Science and 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. General Honors, early graduation. — Dec 2017

SKILLS

Programming Languages: Python (proficient), Java, JavaScript, C++, R, Unix shell, SAS, SQL, Wolfram.
Software Development: PyTorch, TensorFlow, transformers. Flask, FastAPI, async, GCP, Docker, Hydra, Git.
LLM Use: OpenAI, LangChain, RAG, prompt engineering. vLLM, quantization, distributed training, cloud compute.
Natural Languages: Korean, English (fluent), German, Chinese (intermediate).
Service: Reviwer; ACL Rolling Review, Language Resources and Evaluation, Machine Learning Engineering.