

HWIJEEN AHN

hahn2@cs.cmu.edu | 412-613-9903 | [hwiyeen.github.io](https://github.com/hwiyeen) | www.linkedin.com/in/hwiyeenahn/

EDUCATION

Carnegie Mellon University

MASTER IN LANGUAGE TECHNOLOGIES — *Advisor: Eduard Hovy*

Pittsburgh, PA

Aug. 2024

Sogang University

M.S. IN COMPUTER SCIENCE AND ENGINEERING — *CGPA: 4.38 / 4.5*

6-month visiting program to CMU LTI

B.A. IN AMERICAN CULTURE, PSYCHOLOGY — *Summa Cum Laude*

2-year leave of absence as military interpreter

Seoul, Korea

Feb. 2021

Feb. 2018

EXPERIENCE

NAVER, CLOVA

NLP RESEARCH ENGINEER

Seongnam, Korea

Jan. 2021 - May. 2022

- Implemented in-house library for parameter-efficient transfer learning of Korean GPT3 using model parallelism
- Researched effect of pretraining corpora on in-context learning of GPT-3 and coauthored NAACL 2022 paper
- Developed in-house python library for pretraining and distributing diverse language models

PUBLICATIONS

Seongjin Shin*, Sang-Woo Lee*, Hwiyeen Ahn, Sungdong Kim, HyungSeok Kim, Boseop Kim, Kyunghyun Cho, Gichang Lee, Woomyoung Park, Jung-Woo Ha, Nako Sung. On the Effect of Pretraining Corpora on In-context Few-shot Learning by a Large-scale Language Model. *NAACL 2022*. [\[pdf\]](#)

Hwiyeen Ahn*, Jimin Sun*, Chan Young Park*, Yulia Tsvetkov, David R. Mortensen. Cross-Cultural Similarity Features for Cross-Lingual Transfer Learning of Pragmatically Motivated Tasks. *EACL 2021* [\[pdf\]](#)

Hwiyeen Ahn*, Jimin Sun*, Chan Young Park*, Jungyun Seo. NLPDove at SemEval-2020 Task 12: Improving Offensive Language Detection with Cross-lingual Transfer. *SemEval 2020*. [\[pdf\]](#)

Hwiyeen Ahn, Minyoung Seo, Chanmin Park, Juae Kim, Jungyun Seo. Extensive Use of Morpheme Features in Korean Dependency Parsing. In *IEEE BigComp 2019*.

PROJECTS

Measuring the impact of AI on U.S economy

CARNEGIE MELLON UNIVERSITY, LANGUAGE TECHNOLOGIES INSTITUTE

Aug. 2022 - Current

Pittsburgh, PA

- Developing BERT-based classifier tailored for patent data classification, discovering AI-related patents
- Extracting AI experts from publication dataset with LDA and studying their career path

Grammar error correction

NAVER, CLOVA

Jul. 2020 - Sep. 2020

Seongnam, Korea

- Proposed method to generate diverse synthetic data using self-training and back translation to increase BLEU score by 5.26 and won the NAVER AI RUSH competition with \$10,000 prize

Quantifying cross-cultural similarity

CARNEGIE MELLON UNIVERSITY, LANGUAGE TECHNOLOGIES INSTITUTE

Dec. 2019 - Jul. 2020

Pittsburgh, PA

- Proposed cultural similarity features for cross-lingual transfer that can be used to rank transferable languages and published EACL 2021 paper

Multilingual offensive language detection

CARNEGIE MELLON UNIVERSITY, LANGUAGE TECHNOLOGIES INSTITUTE

Dec. 2019 - Feb. 2020

Pittsburgh, PA

- Devised cross-lingual data selection method for offensive language detection and achieved competitive results in SemEval 2020 including 1st place in Arabic

Dependency Parsing

SOGANG UNIVERSITY, NATURAL LANGUAGE PROCESSING LAB

Jul. 2019 - Sep. 2019

Seoul, Korea

- Incorporated graph based and transition based parser and examined the role of morphemes, accomplishing SOTA in Korean dependency parsing and applied for patent

SKILLS

Programming languages: Python, C

ML libraries: PyTorch, Transformers([\[merged PR1\]](#), [\[merged PR2\]](#)), Datasets, MegatronLM([\[merged PR\]](#)), Pytorch-Lightning, Numpy, Pandas, Scikit-learn, PySpark

Dev environments: Git, Docker, Vim, Tmux, Shell, Linux, Databricks