

HWIJEEN AHN

hahn2@cs.cmu.edu | +1 412-613-9903 | [hwiyeen.github.io](https://github.com/hwiyeen) | www.linkedin.com/in/hwiyeenahn/

EDUCATION

Carnegie Mellon University

Pittsburgh, PA

MASTER IN LANGUAGE TECHNOLOGIES — *Advisor: [Eduard Hovy](#)*

Aug. 2024

6-MONTH VISITING STUDENT — *Funded by Korean government*

Feb. 2020

Sogang University

Seoul, Korea

M.S. IN COMPUTER SCIENCE AND ENGINEERING — *Advisor: [Jungyun Seo](#)*

Feb. 2021

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

Feb. 2018

2-year leave of absence as military interpreter

PROFESSIONAL EXPERIENCE

Apple

Seattle, WA

MACHINE LEARNING RESEARCH ENGINEER INTERN

May. 2023 - Aug. 2023

- Developed generative model that efficiently and accurately edits texts using large language model
- Built prototype that allow users to easily try out new input experience
- *Mentor: [Aaron Jaech](#)*

NAVER

Seongnam, Korea

MACHINE LEARNING RESEARCH ENGINEER

Jan. 2021 - May. 2022

- Designed and implemented framework for parameter-efficient finetuning of GPT3 using model parallelism
- Researched effect of pretraining corpora on in-context learning and coauthored NAACL 2022 paper [\[pdf\]](#)
- Developed in-house python library for pretraining and distributing diverse language models [\[talk\]](#)

PUBLICATIONS

Seongjin Shin*, Sang-Woo Lee*, Hwiyeen Ahn, Sungdong Kim, Hyoungho 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

Grammar error correction

Jul. 2020 - Sep. 2020

NAVER, AI RUSH COMPETITION

Seongnam, Korea

- Proposed method to generate diverse synthetic data using self-training and back translation to increase BLEU score by 5.26 and won competition with \$10,000 prize [\[interview\]](#)

Quantifying cross-cultural similarity

Dec. 2019 - Jul. 2020

CARNEGIE MELLON UNIVERSITY, LANGUAGE TECHNOLOGIES INSTITUTE

Pittsburgh, PA

- Proposed cultural similarity features for cross-lingual transfer that can be used to rank transferable languages and published EACL 2021 paper [\[pdf\]](#)

Multilingual offensive language detection

Dec. 2019 - Feb. 2020

CARNEGIE MELLON UNIVERSITY, LANGUAGE TECHNOLOGIES INSTITUTE

Pittsburgh, PA

- Devised cross-lingual data selection method for offensive language detection and achieved competitive results in SemEval 2020 including 1st place in Arabic [\[pdf\]](#)

Dependency Parsing

Jul. 2019 - Sep. 2019

NATIONAL INSTITUTE OF KOREAN LANGUAGE, NLP COMPETITION

Seoul, Korea

- Incorporated graph based and transition based parser and examined the role of morphemes, accomplishing SOTA in Korean dependency parsing and won gold prize [\[patent\]](#)

SKILLS / OPEN SOURCE CONTRIBUTION

Programming languages: Python, C

ML Ecosystem: PyTorch, Transformers([\[merged PR1\]](#), [\[2\]](#), [\[3\]](#)), Datasets, MegatronLM([\[merged PR\]](#)), Pytorch-Lightning, Numpy, Pandas, Scikit-learn, PySpark

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