Mingi Shin

mingi0116s@gmail.com | github.com/mingi-sid | Updated on 2024-11-01

EXPERIENCE

Assistant Research Engineer   July 2024 – Present
Metaverse Lab., Nara Information   *Seoul, South Korea*

- Managing a Python package for RAG (document parsing and vector search)

Student Research Assistant   September 2021 – August 2023
Data Science Group, Institute for Basic Science (IBS)   *Daejeon, South Korea*

- Detecting and analyzing hate speech and harassment in crowd-sourced chatbot data using large language models (LLMs)
- Finding sentiment change of chatbot users between pre- and post- COVID-19 pandemic
- Developing a neural topic model to analyze online contents using LLM and VAE

Intern   June 2021 – February 2023
Simsimi inc.   *Seoul, South Korea*

- Increasing the performance of autonomous hate speech detection from user-created contents
- Research of pre-/post-pandemic emotional change from user utterances, focussing on depression and COVID-19 shutdown.
- Detecting and removing malicious contents in database

EDUCATION

Korea Advanced Institute of Science and Technology (KAIST)   Daejeon, South Korea
Master of Science   *September 2021 – August 2023*

- Thesis: "Toward Chatbot Safety: Context-Aware Offensive Language Detection in Chatbot-Human Conversation Data"

Korea Advanced Institute of Science and Technology (KAIST)   Daejeon, South Korea
Bachelor of Science   *March 2014 – August 2021*

- Double major in Mechanical Engineering

PROJECTS

CALIOPER | *Python, Pytorch, Transformers*   September 2022 – March 2024

- Annotated dataset and model code for "Context-Aware Offensive Language Detection in Human-Chatbot Conversations"
- I designed and managed the dataset annotation process.
- I also designed and implemented a neural net model for detecting offensive language in a conversation context, achieving the best performance in offensive chat detection benchmarks.

UTopic | *Python, Pytorch, Transformers*   August 2022 – May 2023

- VAE-based topic model from "Unified Neural Topic Model via Contrastive Learning and Term Weighting"
- I was in charge of implementing the model, running experiments, finding weaknesses, and suggesting better design. UTopic combines the two types of topic model schemes, achieving favorable results among SotA topic models.

PUBLICATIONS – INTERNATIONAL

Detecting Offensive Language in an Open Chatbot Platform   May 2024
LREC-COLING 2024   *Conf.@Torino, Italia*

- Hyeonho Song, Jisu Hong, Chani Jung, Hyojin Chin, **Mingi Shin**, Yubin Choi, Junghoi Choi, and Meeyoung Cha

Context-Aware Offensive Language Detection in Human-Chatbot Conversations   February 2024
IEEE International Conference on Big Data and Smart Computing (BigComp2024)   *Conf.@Bangkok, Thailand*

- **Mingi Shin**, Hyojin Chin, Hyeonho Song, Yubin Choi, Junghoi Choi, and Meeyoung Cha

The Potential of Chatbots for Emotional Support and Promoting Mental Well-Being in Different Cultures: Mixed Methods Study   August 2023
Journal of Medical Internet Research (IF 7.08 / SCIE)   *Journal*

- Hyojin Chin, Hyeonho Song, G. Baek, **Mingi Shin**, Chani Jung, Meeyoung Cha, Jeonghoi Choi, and Chiyoung Cha

Unified Neural Topic Model via Contrastive Learning and Term Weighting   May 2023
European Chapter of ACL (EACL2023)   *Conf.@Dubrovnik, Croatia*

- Sungwon Han, **Mingi Shin**, Sungkyu Park, Changwook Jung, and Meeyoung Cha
- Citations: 9

User-Chatbot Conversations During the COVID-19 Pandemic: Study Based on Topic Modeling and Sentiment Analysis   January 2023
Journal of Medical Internet Research (IF 7.08 / SCIE)   *Journal*

- Hyojin Chin, Gabriel Lima, **Mingi Shin**, Assem Zhunis, Chiyoung Cha, Junghoi Choi, and Meeyoung Cha
- Citations: 34

A risk communication event detection model via contrastive learning   December 2020
3rd NLP4IF Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda   *Conf.@Barcelona, Spain*

- **Mingi Shin**, Sungwon Han, Sungkyu Park, and Meeyoung Cha

Publications – Domestic

Unique Characteristics of Human-Chatbot Conversations and Their Potential for Mental Health Support
KCC 2022

June 2022
Conf.@South Korea

- Assem Zhunis, Gabriel Lima, Hyojin Chin, **Mingi Shin** (presenter), Jeonghoi Choi, Chiyoung Cha, and Meeyoung Cha

Hate speech detection in chatbot data using KoELECTRA
HCLT 2021

October 2021
Conf.@South Korea

- **Mingi Shin**, Hyojin Chin, Hyeonho Song, Jeonghoi Choi, Hyeonseung Lim, and Meeyoung Cha

Identifying risk communication trend using an event detection model based on contrastive learning
KSC 2021

December 2020
Conf.@South Korea

- **Mingi Shin**, Sungwon Han, Sungkyu Park, and Meeyoung Cha

Technical Skills & Experiences

- **Languages**: Korean (Native), English (Intermediate), Japanese (Intermediate, JLPT N2)

- **Programming Languages**: Python, R, C++, Javascript/Typescript

- **Frameworks/Libraries**: Pytorch, Transformers, Scikit-Learn, Tensorflow, React, Docker

- **Developer Tools**: Git, Jira