Nayoung Kim

nkim48@asu.edu | https://nayoungkim94.github.io | https://www.linkedin.com/in/NayoungKimASU/

RESEARCH INTERESTS

My research centers on ensuring the **trustworthiness** of **Machine Learning** and **Natural Language Processing** algorithms, with a particular emphasis on **Large Language Models**. I investigate key areas such as algorithmic fairness and domain robustness to improve the reliability and ethical impact of the technologies.

EDUCATION

Arizona State University **Spring 2021 – 2025**
PhD, Computer Science *Tempe, AZ*

- Data Mining & Machine Learning Lab (Advisor: Dr. Huan Liu)
- Funded by **DHS-CAOE** (Co-advisor: Dr. Michelle V. Mancenido)

Korea University **2017 – 2019**
MSc, Computer Science & Engineering *Seoul, South Korea*

Korea University **2013 – 2017**
BE, Computer Science & Engineering *Seoul, South Korea*

TECHNICAL SKILLS

Machine Learning & Deep Learning (PyTorch, TensorFlow, Transformers, OpenAI, LangChain, LlamaIndex, Retrieval-augmented generation, Prompt engineering, Reinforcement learning), **Data Analysis** (Numpy, Pandas, Matplotlib, SQL), **Web Development & Cloud** (Flask, Streamlit, AWS, GCP), **Version Control & Container Tools** (Git, Docker), **Collaboration & Communication** (Technical writing, project management, interdisciplinary teamwork)

WORK EXPERIENCE

AMD **August – December 2024**
Machine Learning Software Development Intern *Austin, TX*

- Applying machine learning, large language models, and advanced retrieval-augmented generation (RAG) techniques in software product lines.

DHS-CAOE **May 2022 – August 2024**
Graduate Research Assistant *Tempe, AZ*

- Developed and implemented NLP models for topic modeling and text summarization using BERT and Llama-2-7b.
- Partnered with an interdisciplinary team to design a trustworthy AI-enabled decision support system (AI-DSS) leveraging GPT-4 for intelligence analysis.
- Designed and managed an interactive data analysis and visualization dashboard using NodeJS and Flask.

ONR **Jan 2021 – Aug 2022**
Graduate Research Assistant *Tempe, AZ*

- Researched the integration of COVID-19-related online and offline data using topic modeling methods.
- Analyzed 2 million COVID-19-related tweets, focusing on sentiment analysis and stance detection.

Mathpresso **Jan 2021 – May 2021**
Research Assistant *Tempe, AZ*

- Led a project to automatically classify image-based mathematical problems by difficulty level.
- Implemented LaTeX format mathematical formula embeddings using Tangent-S and static word embeddings.

PUBLICATION & PRESENTATION (Nayoung Kim - Google Scholar)

Robust Stance Detection: Understanding Public Perceptions in Social Media ***ASONAM'24***
Nayoung Kim, David Mosallanezhad, Lu Cheng, Michelle V. Mancenido, Huan Liu

PADTHAI-MM: A Principled Approach for the Design of Trustworthy, Human-Centered AI systems using the MAST Methodology *AI Magazine'24*
Nayoung Kim, Myke C. Cohen, Yang Ba, Anna Pan, Shawaiz Bhatti, Pouria Salehi, James Sung, Erik Blasch, Michelle V. Mancenido, Erin K. Chiou

Evaluating Trustworthiness of AI-Enabled Decision Support Systems: Validation of the Multisource AI Scorecard Table (MAST) *JAIR'23*
Pouria Salehi, Yang Ba, **Nayoung Kim**, David Mosallanezhad, Anna Pan, Myke C. Cohen, Yixuan Wang, Jieqiong Zhao, Shawaiz Bhatti, Michelle V. Mancenido, Erin K. Chiou

Debiasing Word Embeddings with Nonlinear Geometry *COLING'22*
Lu Cheng, **Nayoung Kim**, Huan Liu

Bridge the Gap: the Commonality and Differences Between Online and Offline COVID-19 Data *SBP-BRiMS'22*
Nayoung Kim, David Mosallanezhad, Lu Cheng, Baoxin Li, Huan Liu

An Approach towards Cross-sentence Entity Relation Extraction regarding Encoders and Relation Representations *KCC'18*
Doyeong Hwang, **Nayoung Kim**, Sangrak Lim, Jaewoo Kang

SELECTED PROJECTS

Towards Fair Language Modeling via Parameter-Efficient Methods by Machine Feedback **2024**

- Ongoing project focused on mitigating social biases in large language models (e.g., T5, BERT, LLaMA 3) for toxicity and hate speech detection.
- Currently training large language models to learn fairness and reduce bias using reinforcement learning (RL) and parameter-efficient tuning methods (e.g., LoRA, P-tuning).

MEGAWATT: MAST for Evaluating Generative AI in Worker-Automation Team Tasks **2024**

- Applied MAST (AI trust assessment tool) to evaluate baseline performance, inform improvements, and guide the adoption of OpenAI's GPT-4 for intelligence analysis (IA) tasks.
- Enhanced GPT-4 response quality through prompt engineering and advanced retrieval-augmented generation (RAG) for general conversation and various NLP tasks (e.g., text summarization, entity recognition).
- Conducted human subject studies to assess the suitability of both standard and improved outputs, including evaluating correct rejections of model outputs.

PADTHAI-MM: A Principled Approach for Designing Trustworthy, Human-centered AI systems using the MAST Methodology **2023**

- Developed a novel AI design framework to enhance the trustworthiness of AI systems.
- Validated the framework by creating an AI-enabled decision support system, improving user trust perceptions.
- Analyzed participant ratings and trust-impacting factors, supporting the framework's effectiveness in boosting AI system trustworthiness.

EXTRACURRICULAR ACTIVITIES

Program Committee (PC) member of ASONAM 2024 conference **2024**

Program Committee (PC) member of ASONAM, SBP-BRiMS 2023 conference **2023**

Invited Reviewer for EMNLP 2023 conference **2023**

Reviewer at ECML-PKDD, ACM MultiMedia, ASONAM, AAAI conferences **2022**

Volunteer at WSDM 2022 conference **2022**

Reviewer at ASONAM, IEEE CogMI conferences **2021**

Volunteer at KDD 2021 conference **2021**

Teaching Assistant for CSE 205: Object-Oriented Programming and Data Structures **2021 – 2022**