

Nayoung Kim

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

RESEARCH INTERESTS

My research focuses on advancing Trustworthy and Truthful AI by developing *fair, robust, and accurate* NLP and large language models, with expertise in social media data mining and experience in several research projects.

EDUCATION

Arizona State University

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

WORK EXPERIENCE

AMD

Aug – Dec 2024, May – Aug 2025

AI/ML Intern

Austin, TX

- Apply machine learning, large language models, and advanced retrieval-augmented generation (RAG) techniques to improve AI tool truthfulness, enhance the quality of LLM outputs, and mitigate hallucinations.
- Integrate user feedback to enhance the performance of a multi-agent conversational AI pipeline.

DHS-CAOE

May 2022 – Aug 2024

Graduate Research Assistant

Tempe, AZ

- Develop and implement NLP models for topic modeling and text summarization using BERT and Llama-2-7b.
- Partner with an interdisciplinary team to design a trustworthy AI-enabled decision support system (AI-DSS) leveraging GPT-4 for intelligence analysis.
- Design and management of an interactive data analysis and visualization dashboard using NodeJS and Flask.

ONR

Jan 2021 – Aug 2022

Graduate Research Assistant

Tempe, AZ

- Research on integrating COVID-19-related online and offline data using topic modeling methods.
- Analysis of 2 million COVID-19-related tweets, focusing on sentiment analysis and stance detection.

Mathpresso

Jan – May 2021

Graduate Research Assistant

Tempe, AZ

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

SELECTED PUBLICATIONS [[Google Scholar](#)]

PADTHAI-MM: A Principled Approach for the Design of Trustworthy, Human-Centered AI systems using the MAST Methodology

AI Magazine'25

Myke C. Cohen, **Nayoung Kim**, Yang Ba, Anna Pan, Shawaiz Bhatti, Pouria Salehi, James Sung, Erik Blasch, Michelle V. Mancenido, Erin K. Chiou

Robust Stance Detection: Understanding Public Perceptions in Social Media

ASONAM'24

Nayoung Kim, David Mosallanezhad, Lu Cheng, Michelle V. Mancenido, Huan Liu

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

Bayesian Learning based Uncertainty Measurement for Hallucination Mitigation **2025**
 ▪ The ongoing project focused on mitigating hallucinations in generated text from large language models.
 ▪ Leverages Bayesian Neural Networks and Parameter-efficient tuning methods to measure epistemic uncertainty.

MEGAWATT: MAST for Evaluating Generative AI in Worker-Automation Team Tasks **2024 - 2025**
 ▪ 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).
 ▪ Conducted human subject studies to assess the suitability of both standard and improved outputs, including evaluating correct rejections of model outputs.

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)

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, and 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