

Nayoung Claire Kim

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

RESEARCH INTERESTS

Advancing trustworthy and truthful AI through the development of fair, robust, and accurate NLP and large language models. Experienced in social media data mining, debiasing methods, retrieval-augmented generation, and AI evaluation frameworks.

EDUCATION

Arizona State University

PhD, Computer Science

2021 – Jul 2026

Tempe, AZ

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

Korea University

M.E., Computer Science & Engineering

2017 – 2019

Seoul, South Korea

Korea University

B.S., Computer Science & Engineering

2013 – 2017

Seoul, South Korea

WORK EXPERIENCE

AMAZON

Applied Scientist Intern

Sep – Dec 2025

Bellevue, WA

- Applied advanced ML to production sentiment forecasting system for community operations.

AMD

AI/ML Intern

Aug – Dec 2024, May – Aug 2025

Austin, TX

- Applied 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.
- Integrated user feedback to enhance the performance of a multi-agent conversational AI pipeline.
- Developed a synthetic QA generation pipeline using LLMs and built an LLM-as-a-judge system achieving strong agreement with domain experts for scalable evaluation.

DHS-CAOE (Center for Accelerating Operational Efficiency)

Graduate Research Assistant

May 2022 – Aug 2024

Tempe, AZ

- Designed NLP models for topic modeling and text summarization using BERT and Llama-2-7b.
- Collaborated with an interdisciplinary team to implement trustworthy AI-enabled decision support systems leveraging GPT-4.
- Developed an interactive data analysis and visualization dashboard using NodeJS and Flask.

ONR (Office of Naval Research Project)

Graduate Research Assistant

Jan 2021 – Aug 2022

Tempe, AZ

- Researched integration of online/offline COVID-19 datasets using topic modeling.
- Analyzed 2M tweets for sentiment and stance detection in pandemic-related discussions.

Mathpresso

Graduate Research Assistant

Jan – May 2021

Tempe, AZ

- Built a linear classifier for image-based math problems by difficulty level.
- Created LaTeX-based math 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 2026
▪ Applied Bayesian Neural Networks and parameter-efficient tuning to quantify epistemic uncertainty in generative models.

MEGAWATT: MAST for Evaluating Generative AI in Worker-Automation Team Tasks 2024 - 2025
▪ Applied MAST (AI trust assessment tool) to evaluate GPT-4 baseline performance in intelligence analysis.
▪ Enhanced responses via prompt engineering and RAG.
▪ Conducted human subject studies on response suitability and correctness in rejection detection.

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 & Deployment** (Git, Docker), **Collaboration** (Technical writing, project management, interdisciplinary research)

EXTRACURRICULAR ACTIVITIES

Reviewer at AMLC 2025 (Gen AI Evaluation Workshop) 2025
Invited Reviewer for NeurIPS 2025 (Reliable ML Workshop) 2025
Program Committee (PC) member of ASONAM 2024 conference 2024
Program Committee (PC) member of ASONAM, SBP-BRiMS 2023 conferences 2023
Invited Reviewer for EMNLP 2023 conference 2023
Reviewer at ECML-PKDD, ACM MultiMedia, ASONAM, and AAI conferences 2022
Volunteer at WSDM 2022 conference 2022
Reviewer at ASONAM, IEEE CogMI conferences 2021