# Nayoung Kim

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

## RESEARCH INTERESTS

My research interest mainly lies within trustworthiness in **Machine Learning (ML)** and **Natural Language Processing (NLP)** algorithms and their applications, including bias mitigation and domain generalization.

## 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**

**Korea University** **2017 – 2019**
*MSc, Computer Science & Engineering* *Seoul, South Korea*
- **3.7/4.0** GPA

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

## TECHNICAL SKILLS

Data analysis using Python, PyTorch, Tensorflow, Keras, Numpy, and Scikit – SQL – Web Servers – AWS – Google Cloud Platform

## WORK EXPERIENCE

**DHS-CAOE** **May 2022 – Present**
*Graduate Research Assistant* *Tempe, AZ*
- Created and managed a comprehensive interactive dashboard for data analysis and visualization using NodeJS and Flask
- Built and implemented BERT-based topic modeling and text summarization models
- Conducted research on designing a trustworthy AI-enabled decision support system (AI-DSS)

**ONR** **Jan 2021 – Aug 2022**
*Graduate Research Assistant*
- Conducted research on connecting COVID-19-related online data to offline data using topic modeling methods
- Conducted a comprehensive analysis of 2 million COVID-19-related tweets, focusing on sentiment analysis and stance detection

**Mathpresso** **Jan 2021 – May 2021**
*Research Assistant*
- Led a project to automatically classify image-based mathematical problems based on their difficulty levels
- Implemented LaTeX format mathematical formula embeddings using Tangent-S and static word embeddings

**Korea University** **2016**
*Undergrad Researcher*
- Collected and annotated biomedical reference dataset for training an RNN-based relation extraction model

## SELECTED PROJECTS

**Fair Toxicity Detection with Masking Spurious Artifacts** **2023**
- Detected spurious artifacts from the top-N important words for toxicity detection and hate speech detection using a language model

- Trained a model to learn fairness and mitigate bias using reinforcement learning (RL)

**Automated Evaluation of Machine-generated Summaries using RLHF** **2023**

- Trained a Transformer-based classifier to evaluate a document-summary pair through multi-class classification and reinforcement learning with handcrafted human preferences dataset
- Conducted expert evaluations on the output scores to validate the effectiveness of the proposed learning method

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

- Developed a novel AI design framework, addressing the challenge of designing trustworthy AI systems
- Demonstrated the effectiveness of the framework through the development of the AI-enabled d ㅋ ecision support system, with the framework positively impacting trust perceptions among users
- Conducted association analysis between participants' ratings and trust-impacting information, providing a theoretical basis for the framework's effectiveness in enhancing AI system trustworthiness

**READIT: REporting Assistant for Defense and Intelligence Tasks** **2022**

- Trained and developed a text summarization system for use in intelligence analysis, utilizing Transformer-based models
- Implemented a user-friendly web interface for the text summarization system using NodeJS and the Google Cloud Platform, allowing analysts to easily access summarized reports, enhancing their workflow and productivity

**Facewise: An AI-based Face ID Verification System** **2022**

- Engineered a robust and accurate face ID verification system employing Convolutional Neural Networks (CNN) and ResNet, ensuring a reliable and efficient means of identity authentication in security screening scenarios.
- Implemented advanced facial recognition algorithms and fine-tuned model parameters to optimize the system's performance, thus enhancing the overall security and user experience.

**Interpreting Text Classifiers with Counterfactual Explanation** **2021**

- Completed as the final project for CSE 472 (Social Media Mining)
- Implemented counterfactual models for a multi-layer neural network used in text classification

**Fake/Bot Detection in Twitter Activity** **2021**

- Completed as the final project for CSE 573 (Semantic Web Mining)
- Discovered undisclosed bots and identified top manipulators through community detection

**Biomedical Entity Relation Extraction** **2017**

- Extracted Biomedical entities and identify their relation existence
- Utilized the Comparative Toxicogenomics Database (CTD) dataset, which provides chemical-gene, chemical-disease, and gene-disease relation data collections through distant supervision due to the lack of training data
- Implemented and trained a tree-RNN based model, SPINN, in conjunction with a word-character embedding model (CharWNN)

## PUBLICATION & PRESENTATION (Nayoung Kim - Google Scholar)

**Fair Hate Speech Detection via Mitigating Spurious Artifacts** - *Under Review*
**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** - *Under Review*
**Nayoung Kim**, Myke C. Cohen, Yang Ba, Anna Pan, Shawaiz Bhatti, Pouria Salehi, James Sung, Erik Blasch, Michelle V. Mancenido, Erin K. Chiou

**STANCE-C^3: Domain-adaptive Cross-target Stance Detection via Contrastive Learning and Counterfactual Generation** - *Under Review*
**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

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

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

## AWARDS

**Conference Scholarship** **2022**
SBP-BRiMS 2022

**Fulton Scholarship** **2021**
**Ira A. Fulton Schools of Engineering, Arizona State University**
Offered in recognition of academic achievements

**General Scholarship** **2017**
**College of Information, Korea University**
Offered in recognition of extraordinary academic achievements

**Work-Study Scholarships** **2015**
**College of Information, Korea University**
Offered in recognition of extraordinary academic achievements

**Academic Excellence Scholarships** **2013**
**College of Information & Communication, Korea University**
Offered to top 6% freshmen in the College of Information & Communication

## EXTRACURRICULAR ACTIVITIES

| | |
|---|---|
| **Program Committee (PC) member of ASONAM 2023 conference** | **2023** |
| **Program Committee (PC) member of 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** |