Nayoung Kim

699 S Mill Ave, Room 561BA
Tempe AZ 85281
(+1) 480-669-1468
https://nayoungkim94.github.io
nkim48@asu.edu

To Whom It May Concern,

Thank you for considering me for the research internship position for the Summer of 2024. I am a fourth-year Ph.D. student at Arizona State University, pursuing my Ph.D. in Computer Science at the Data Mining and Machine Learning (DMML) lab under the guidance of Dr. Huan Liu and Dr. Michelle Mancenido. My research interest primarily focuses on establishing trustworthiness in Machine Learning and Natural Language Processing algorithms, with a specific emphasis on human-AI teaming, addressing model robustness and algorithmic fairness.

Currently, I have been leading studies on cutting-edge fair and social bias mitigated NLP models and LLMs. One aspect of my work involves building a bias-mitigated LLM-based NLP model leveraging parameter-efficient methods. Our goal is to effectively mitigate several social biases (e.g., gender, race, intersectionality) in a pre-trained model with fewer parameters. Additionally, I am involved in an interesting project focused on building an improved GPT-4-based decision support system enriched with trustworthiness knowledge to assist intelligence and analysis tasks. This project is a collaborative effort with experts from human systems engineering, statistics, and data science. My research experiences align closely with the position, as they emphasize the importance of human-AI collaboration and the ethical considerations of AI systems in supporting human tasks.

I have extensive experience in implementing language models for tasks such as text classification, summarization, text generation, enhancing model interpretability, and agent-based simulation modeling. I am proficient with Python-based machine learning frameworks, including PyTorch, Hugging Face and TRL (Transformer Reinforcement Learning) frameworks, and in tuning language models such as BERT, GPT, and Llama-2. I have explored and developed numerous deep learning methodologies for model training, encompassing Transformers, reinforcement learning (RL), human-in-the-loop approaches, counterfactual data augmentation (CDA), and parameter-efficient fine-tuning (PEFT). With the recent advancements in large language models (LLMs), I am also skilled in addressing LLM-related techniques for improving model capabilities for human interaction, such as retrieval-augmented generation (RAG) and in-context learning (ICL).

My experience also covers designing trustworthy AI systems where I've developed expertise in evaluating model trust and fairness from human and machine perspectives. This includes statistical analysis and conducting large-scale human subject studies. I've participated in projects aimed at improving models for tasks like stance detection and toxicity detection with the goal of building models that enhance trust among users in social media discourse on platforms like Twitter and Instagram.

My research interests have always been at the intersection of machine learning, ethics, and human-AI collaboration. With this internship, I aim to enhance my work on sociotechnical design, evaluation, and auditing methods to support robust and ethical human-AI teams from technical aspects. This experience is key to enriching my PhD research, allowing me to embed ethical considerations more effectively into my work. I'm particularly interested in applying large language models (LLMs) in ways that uphold ethical principles. This internship is not just a milestone for my academic growth but also crucial for my long-term goal of contributing to trustworthy AI. It emphasizes my dedication to raising ethical standards in my research and across the AI field, aiming to make a meaningful impact on society.

Thank you for your consideration.
Nayoung Kim