AKSHITA JHA

+1-267-844-9798 | akshitajha@vt.edu | https://akshitajha.github.io/

EDUCATION

Virginia Tech, Washington DC-Baltimore Area *Fall 2019 - Present*
PhD Student, Dept. of Computer Science GPA: 4.0/4.0
Advisor: Dr. Chandan K. Reddy

IIIT-Hyderabad, Hyderabad, India *Aug 2016 - May 2018*
MS by Research, Computational Linguistics
Advisor: Dr. Radhika Mamidi

IIIT-Hyderabad, Hyderabad, India *Aug 2012 - May 2016*
Bachelor of Technology (Hons.), Computer Science GPA: 8.11/10.00

RESEARCH INTERESTS

Robustness, Interpretability, Fairness in Deep Learning, Natural Language Processing

PUBLICATIONS

- **Akshita Jha** and Chandan K. Reddy. *"CodeAttack: Code-based Adversarial Attacks for Pre-Trained Programming Language Models"*, 2022. Under Review.
- **Akshita Jha**, Adithya Samavedhi, Vineeth Mohan, and Chandan K. Reddy. *"Transformer based Models for Long Document Comparison: Challenges and Empirical Analysis"*, 2022. Under Review.
- **Akshita Jha**, Vineeth Mohan, Jaideep Chandrashekar, Adithya Samavedhi, and Chandan K. Reddy. *"Supervised Contrastive Learning for Interpretable Long-Form Document Matching"*. ACM Transactions on KDD, May, 2022.
- **Akshita Jha**, Bhanukiran Vinazamuri & Chandan K. Reddy. *"Fair Representation Learning using Interpolation Enabled Disentanglement"*. 2021
- [Spotlight Presentation] **Akshita Jha** and Radhika Mamidi. *"When does a Compliment become Sexist? Analysis and Classification of Ambivalent Sexism using Twitter Data"*. In the Proceedings of the Second Workshop on Natural Language Processing and Computational Social Science, ACL 2017.

RESEARCH PROJECTS

Adversarial Attacks on Pre-trained Programming Language Models

Collaborators: Chandan K. Reddy, Virginia Tech

- Build a simple yet effective blackbox attack model that uses code structure to generate imperceptible, effective, and minimally perturbed adversarial code samples to demonstrate the vulnerabilities of the stateof-the-art PL models to code-specific adversarial attacks. The attack is transferable to several code-code (translation and repair) and code-NL (summarization) tasks across different programming languages.

Supervised Constrastive Learning for Interpretable Document Matching

Collaborators: Vineeth Mohan, Jaideep Chandrashekhar, InterDigital; Chandan K. Reddy, Virginia Tech

- Built a supervised contrastive learning transformer-based framework to compute the (dis)similarity within and across different chunks and sections of long documents. The model uses custom positional embeddings to get meaningful and interpretable similarity scores at three different levels.

Disentangled Representations of Fairness
Collaborators: Chandan K. Reddy, Virginia Tech; Bhanukiran Vinzamuri, IBM Research

- Proposed a deep learning model that learnt latent representations to effectively disentangle factors of variation that help in identifying features for protected attributes. The model used interpolation along with adversarial autoencoder for better disentanglement.

Detecting Ambivalent Sexism from Twitter Data
Collaborators: Radhika Mamidi, IIIT-Hyderabad

- Created a novel dataset of 10,000 sexist tweets (Benevolent Sexist, Hostile Sexist and Others). Built a classifier using Support Vector Machines (SVM), Sequence-to-Sequence models, and FastText to classify tweets into one of the above three classes.

Measuring Benevolent Prejudice from Twitter Data
Collaborators: David Jurgens, University of Michigan, Ann Arbor

- Automatically generated words on a semantic axis in a high dimensional vector space using k-nearest neighbors, breadth-first search and dependency parsing. The aim was to identify implicit and explicit cross-cultural biases against different social and political groups in Twitter data by using inherent degrees of warmth and competence.

De-escalation of Hate-Speech on Reddit
Collaborators: Libby Hemphill, David Jurgens, University of Michigan, Ann Arbor

- Built a bot to automatically detect the escalation of hostility in a Reddit thread using a vocabulary of insults, part-of-speech tagger and dependency parser and automatically post a ‘de-escalating’ comment to prevent the conversation from going off the rails.

WORK EXPERIENCE

AI Lab, InterDigital, Palo Alto, CA | *Research Intern*	*May 2020 - Aug. 2020*
Google Summer of Code, Debian | *Intern*	*May 2015 - Aug. 2015*
The Linux Foundation, OpenDaylight | *Deb Maintainer*	*May 2016 - Apr 2017*
Intuit, Bangalore, India | *Software Engineer 1*	*Aug 2017 - May 2018*

AWARDS AND HONORS

- **CS Research Mentorship Program Scholar**, Google, 2021
- **Grace Hopper Celebration Scholarship**, 2020
- **Member of the Phi Kappa Phi Honor Society for Academic Excellence**

SKILLS

Languages (Proficient in): Python
Languages (Familiar with): C, C++, Java
Tools/Libraries: PyTorch, Keras, LaTeX, Bash Shell Scripting, Git

RELEVANT COURSEWORK

Deep Learning, Advanced Machine Learning, Topics in Human-computer Interaction, Natural Language Processing, Data Mining.