

Anushree Hede

Email: anushree.hede@gmail.com | LinkedIn: www.linkedin.com/in/anushree-hede | Work authorization: US Citizen

EDUCATION

University of Pennsylvania (Philadelphia, PA)

May 2021

Master of Science and Engineering, Computer and Information Science

GPA: 3.8 / 4

Courses: Deep Learning, Computational Linguistics, Software Systems, Internet & Web Systems, Reasoning for NLU

Birla Institute of Technology and Science, Pilani (Hyderabad, India)

July 2019

Bachelor of Engineering (Hons.), Computer Science

CGPA: 9.1 / 10

Courses: Data Structures, Algorithms, Information Retrieval, Databases, Data Mining, Probability, Operating Systems

TECHNICAL SKILLS

Languages: Python, C++, Java | **ML/NLP libraries:** PyTorch, Scikit-learn, HuggingFace, NLTK | **Databases:** MySQL, Pandas, S3 | **Web:** JavaScript, HTML, CSS | **Frameworks:** Hadoop, EMR, Spark | **Tools:** Git, Linux

SELECTED PROJECTS

Master's Thesis - advised by Dr. Ani Nenkova and Dr. Byron C. Wallace [\[pdf\]](#)

Jan. 2021 - May 2021

- Analyzed words from the Civil Comments dataset that triggered high toxicity scores from Jigsaw's Perspective API
- Trained a BERT-based toxicity model that reduces the error on person identity/group words from above word list

PennSearch - Mini Search Engine with distributed Web Crawler and Indexer

Mar. 2021 - May 2021

- Developed a Mercator-style crawler built on a simple version of Apache Storm and crawled 200k documents
- Implemented a distributed Indexer and PageRank model using AWS EMR, used S3 and RDS for data storage
- Hosted search engine on EC2 which ranked relevant documents by weighting TFIDF similarities and PageRank

PennCloud - Distributed Cloud platform with Mail (SMTP) and File Storage Services

Oct. 2020 - Dec. 2020

- Developed a cloud platform in C++, using a key-value store in backend and Protobuf to pass messages over TCP
- Built backend servers supported sequential consistency, primary-based replication, fault tolerance and recovery
- Built frontend servers supported HTTP requests, cookies, load balancing and fault tolerance

Explainability for Multiple-Choice Science Question-Answering

Mar. 2020 - May 2020

- Selected the AristoRoBERTa model for multiple-choice QA; which takes in the question, answer option, and a set of supporting context sentences for each answer option; and predicts the correct answer choice for the question
- Performed data-driven probing on the model to find context sentences that best explain the predicted answer choice

EXPERIENCE

Graduate Research Assistant

Sep. 2019 - Dec. 2020

Advised by Dr. Ani Nenkova and Dr. Diana C. Mutz at University of Pennsylvania (Philadelphia, PA)

- Curated a dataset of video clips and transcript segments of American news shows with annotations of incivility
- Examined the suitability of a popular toxicity model (Jigsaw's Perspective API) to identify incivility in news shows
- Demonstrated that high Perspective scores are spuriously correlated with presence of non-offensive error words

Research Intern

Jan. 2019 - June 2019

Bosch Research and Technology Center (Bangalore, India)

- Developed a word-embedding-based trend detection algorithm for time-stamped automobile consumer complaints
- Demonstrated PoC by comparing results with a topic modelling algorithm (online-LDA), using the bumpiness metric

PUBLICATIONS

From Toxicity in Online Comments to Incivility in American News: Proceed with Caution

[\[pdf\]](#)

Anushree Hede, Oshin Agarwal, Linda Lu, Diana C. Mutz and Ani Nenkova

European Chapter of the Association for Computational Linguistics (EACL) 2021

TEACHING ASSISTANTSHIP

Deep Learning for Data Science

Jan. 2021 - May 2021

- Prepared teaching materials and homework for Recurrent Neural Networks and NLP topics on Google Colab
- Mentored a pod of 9 students for 5 hours/week with deep learning concepts and the final project