Mukund Srinath

State College, PA
mukund.srina@gmail.com — mukundsrinath.github.io

Education

Pennsylvania State University *Expected: May, 2024*
PhD, Informatics GPA: 3.85
Co-supervised: C. Lee Giles and Shomir Wilson

Pennsylvania State University *August, 2017 - May, 2019*
MS, Information Sciences and Technology GPA: 3.88

R. V. College of Engineering, India *August, 2011 - May, 2015*
BE, Electrical and Electronics Engineering GPA: 8.94

Work Experience

Trustpage Detroit, MI
Machine Learning Intern *May, 2022 - August, 2022*

- Designed evaluation metrics to track performance of company-wide suite of machine learning models.
- Implemented and **deployed** a question similarity model on GCP, **improving performance by 12%**.
- Applied feature engineering and synthetic data generation techniques on a task with few training samples.
- Conveyed findings to non-technical stakeholders, aligning with strategic business objectives.

ClearAvenue Columbia, MD
Software Engineering Intern *May, 2018 - August, 2018*

- Designed a proof-of-concept to improve medical device security in software defined networks.
- Implemented **unsupervised anomaly detection** methods to detect potential network security threats.

McFadyen Digital Bangalore, India
Software Engineer *July, 2015 - July, 2017*

- Implemented a full-life-cycle webapp of a human resource information system supporting ∼500 users.
- Developed backend and frontend online retail features for search, payment and checkout functionality.

Research Experience

Human Language Technologies Lab Pennsylvania State University
Research Assistant *August, 2019 - Present*

- Applying information extraction and information retrieval techniques to analyze privacy policies at scale.
- Developed a speciality crawler and collected the largest corpus of privacy policies containing **3.4M policies**.
- Designed, created and maintain, the **world's first privacy policy search engine**, including, creating custom ranking function, and statistical analysis of results and user engagement. PrivaSeer Search Engine
- Achieved state of the art results on classification, named entity recognition (NER) and question answering tasks on privacy policies by creating PrivBERT, a **privacy language model**. HuggingFace

Intelligent Information Systems Lab Pennsylvania State University
Research Assistant *August, 2019 - Present*

- Applying responsible AI principles in **analyzing and mitigating biases** in machine learning models.
- Designed recommendations for ethical, transparent, and socially responsible sentiment analysis.
- Designed **bias detection techniques** in generative large language models (LLMs).

PIKE Group Pennsylvania State University
Researcher *August, 2017 - July, 2019*

- Designed a novel technique to select best answers from a list of non-fact-based questions by generating questions.
- Implemented a convolutional neural network (CNN) based algorithm using TensorFlow to rank answers achieving a precision@1 score of 0.452 and mean reciprocal rank of 0.59, thereby setting a baseline for a novel dataset.

Awards and Honors

- **Best Student Paper**, @DocEng'23
- **Best Short Paper**, @TrustNLP Workshop (ACL'23)
- **Best Student Paper**, @PETS'23
- **Semi-finalist** in Nittany AI Challenge, 2023 for ClassCollab an adaptive learning system that supports socially-mediated peer learning, in low-resource learning communities.
- **Semi-finalist** in Nittany AI Challenge, 2023 for ChronoNews.ai an application that creates a timeline of news articles.
- **2nd place** in HackPSU, 2019 for *Beyond Tweet*, an application that combines geospatial census data with tweet sentiment to help make market predictions.
- **2nd place and $20,000 in funding** in Nittany AI Challenge, 2018 for *ProFound: A professor search engine*.
- **3rd place** in HackPSU, 2018 for *WeatherOrNot*, a travel application to recommend vacation destinations based on travel prices, and weather preferences.
- **1st place and honorable mention by IBM** at HackPSU, 2017 for *FindViser* an application that suggests research advisers to graduate students.

Additional Projects

News Timeline Summarization

- Designed a tool that captures and organizes news in the form of a timeline for improved sense-making.
- Developed the document clustering, timeline creation, and subjectivity detection modules using PyTorch.

Personal Information Type Extraction

- Designed a novel scalable technique to extract the types of personal information collected from users from over 1.4 million privacy policies. Trained a BERT based model to identify valid personal information types given inputs of noun phrases and their POS tags.

ProFound: Professor Search Engine

- Designed an end-to-end system to track professors based on research topics and interests. Crawled, indexed and ranked ∼500 academic websites.

Skills

Machine Learning Skills: sklearn, Tensorflow, PyTorch, R. Experience with multi-GPU processing.

Development Skills: Python, Java, C++, ElasticSearch, Django, SQL. Experience with Git and multiple OS.

Selected Publications

Mukund Srinath, Lee Matheson, Pranav Venkit, Gabriela Fortuna, Florian Schaub, Lee Giles, Shomir Wilson
Privacy Now or Never: Large-Scale Extraction and Analysis of Dates in Privacy Policy Text
[**Best Paper**] The 23rd ACM Symposium on Document Engineering (**DocEng, 2023**)

Mukund Srinath, Soundarya Sundareswara, C. Lee Giles, Shomir Wilosn
Privacy Lost and Found: An Investigation at Scale of Web Privacy Policy Availability
The 23rd ACM Symposium on Document Engineering (**DocEng, 2023**)

Pranav Venkit, **Mukund Srinath**, Shomir Wilson
Automated Ableism: An Exploration of Explicit Disability Biases in Sentiment and Toxicity Analysis Models
[**Best Paper**] 3rd Workshop on Trustworthy Natural Language Processing (**ACL, 2023**)

Abraham Mhaidli, Selin Fidan, An Doan, **Mukund Srinath**, Lee Matheson, Shomir Wilson, Florian Schaub
Researchers' Experiences in Analyzing Privacy Policies: Challenges and Opportunities
[**Best Paper**] Proceedings on Privacy Enhancing Technologies (**PETS, 2023**)

Pranav Venkit, **Mukund Srinath**, Shomir Wilson
A Study of Implicit Language Model Bias Against People With Disabilities
29th International Conference on Computational Linguistics (**COLING, 2022**)

Mukund Srinath, Shomir Wilson, C. Lee Giles.
Privacy at Scale: Introducing the PrivaSeer Corpus of Web Privacy Policies.
59th Annual Meeting of the Association for Computational Linguistics (**ACL, 2021**)

Mukund Srinath, Soundarya Sundareswara, Shomir Wilson, C. Lee Giles.
PrivaSeer: A Privacy Policy Search Engine.
International Conference on Web Engineering (**ICWE, 2021**)

News and Media Mentions

The Hill: Common AI language models show bias against people with disabilities: study 2022
El Pais: Latin texts and unaffordable lengths: the revelations of a privacy policy search engine 2021
Penn State News: Search engine could help researchers scour internet for privacy documents 2021
The Collegian: HackPSU winning website aims to combat challenge of finding an adviser 2018
Penn State News: 2018 Nittany AI Challenge winners provide EdTech solutions, receive $50,000 2018