

# SAKSHI

(413) 725-8679 | [ssakshi@umd.edu](mailto:ssakshi@umd.edu) | <https://www.linkedin.com/in/sakshi113/> | [S Sakshi - Google Scholar](#) | [Website](#)

## EDUCATION

<b>University of Maryland, College Park, MD, USA</b> <i>Degree: Ph.D. in Computer Science</i>	Aug 2024 – Present
<b>University of Massachusetts Amherst, MA, USA</b> <i>Degree: Master of Science in Computer Science</i> <i>Coursework: Graph Modelling, Machine Learning, Natural Language Processing, Information Retrieval, Algorithms for Data Science</i>	Sep 2022 – May 2024 GPA : <b>3.71 / 4.00</b>
<b>Siddaganga Institute of Technology, Tumkur, India</b> <i>Degree: B.Tech in Computer Science and Engineering</i> <i>Coursework: Data Structures &amp; Algorithms, Computer Network, Digital System Design</i>	Aug 2016 – Jun 2020 GPA : <b>9.05 / 10.00</b>

## Papers and Patents

- [GAMA: A Large Audio-Language Model with Advanced Audio Understanding and Complex Reasoning Abilities](#) **EMNLP 2024 (Under Review)**; [Media Coverage 1](#) [Media Coverage 2](#)
- [EH-MAM: Easy-to-Hard Masked Acoustic Modeling for Self-Supervised Speech Representation Learning](#) **EMNLP 2024 (Under Review)**
- [Do Vision-Language Models Understand Compound Nouns?](#) **NAACL 2024**
- [CompA: Addressing the Gap in Compositional Reasoning in Audio-Language Models](#) **ICLR 2024**
- [ASPIRE: Language-Guided Augmentation for Robust Image Classification](#) **ACL 2024**
- [ABEX: Data Augmentation for Low-Resource NLU via Expanding Abstract Descriptions](#) **ACL 2024**
- [DALE: Generative Data Augmentation for Low-Resource Legal NLP](#) **EMNLP 2023**
- [M-MELD: A Multilingual Multi-Party Dataset for Emotion Recognition in Conversations](#) **arXiv**
- [DeToxy: A Large-Scale Multimodal Dataset for Toxicity Classification in Spoken Utterances](#) **InterSpeech 2022**
- [Multivariate Time-series Unsupervised Anomaly Detection in 5G Networks](#) **US Patent with Cisco Systems**

## TECHNICAL SKILLS

**Programming / Scripting Languages:** Python, Java, C/C++, SQL, HTML, CSS, Angular, RESTful API  
**Frameworks:** PySpark, Hadoop, PyTorch, Keras, Tensorflow, Spark, Flask, Django, Keras, Numpy, Pandas  
**Tools / Platforms:** Apache Airflow, Dataswarm, Git, Microsoft Azure, Databricks, Apache Oozie, Linux, Tableau, Docker, AWS EC2, JIRA

## WORK EXPERIENCE

<b>LightBeam.ai, San Jose, USA</b> <i>Machine Learning Intern</i>	May 2024 - Aug 2024
<ul style="list-style-type: none"><li>• Contributed to the development of Spectra, an advanced data protection engine. My role involved leveraging AI and machine learning to link fragmented data and create detailed identity graphs for customers, employees, and others. This helped in assessing risks and <b>preventing data leaks</b>.</li><li>• Utilized <b>generative AI</b> technologies to enhance data cataloging, control, and compliance processes, ensuring robust data security and privacy measures.</li></ul>	
<b>UMass NLP Lab, Amherst, USA</b> <i>NLP Research Intern</i>	May 2023 - Aug 2023
<ul style="list-style-type: none"><li>• Worked on the International Hate Observatory Project focused on the <b>detection of hate speech</b> in YouTube videos by implementing Natural Language Processing (<b>NLP</b>) and <b>audio processing</b> techniques in <b>PyTorch</b>.</li><li>• Collaborated with a multidisciplinary team to design and optimize scalable solutions for processing <b>multimodal data</b> and applying state-of-the-art techniques to identify and classify instances of hate speech.</li></ul>	
<b>Cisco Systems, Bangalore, India</b> <i>Software Engineer II</i>	Aug 2020 - Jul 2022
<ul style="list-style-type: none"><li>• Developed and implemented an AI-based network security system for a telecom customer, leveraging advanced AI techniques to detect and <b>flag network anomalies</b> while contributing to Cisco's AI R&amp;D division, patented under US Patents.</li><li>• Implemented MPLS backbone services for Bharti Airtel and leveraging Cisco's <b>Network Service Orchestration (NSO)</b> tool, the performance of their data transmission across the globe was improved by ~10%.</li><li>• Worked on automating Cisco's Data Center Network Manager(<b>DCNM</b>) APIs modules using <b>Ansible</b> and <b>Python</b>, increasing the operational efficiency of DCNM by 30%.</li></ul>	
<b>Cisco Systems, Bangalore, India</b> <i>Software Engineering Intern</i>	Jan 2020 - Jun 2020
<ul style="list-style-type: none"><li>• Devised an end-to-end framework to gather network configurations using <b>MEAN STACK</b>, directly integrating them into the network service orchestration pipeline, resulting in a 30% increase in operational efficiency.</li></ul>	

## PROJECTS

---

### Data Augmentation for Low Resource Legal NER | [GitHub](#)

- Used PyTorch and **HuggingFace** libraries for generating augmentations with **BART** on low-resource legal data for better neural network learning of NER and passing the augmentation to **fine-tune legalBERT**.
- Used a novel **keyword extraction algorithm** to pre-train BART that improves performance by 17% over baseline.

### Road Symbols Detection | *Smart India Hackathon Project, aided by Govt. of India*

- Built an End-to-End system using **Convolutional neural networks** and IoT to classify extracted frames in a real-time app for tracking road symbols from a moving vehicle to display all the required information related to it to help the drivers follow traffic rules and reduce the possibility of accidents.

## ACHIEVEMENTS AND EXTRACURRICULAR ACTIVITIES

---

- Recognized by Cisco CX CTO and higher management on multiple occasions for my research and innovation initiatives, 2020.
- KPIT Innovation Award in a national-level hackathon, SMART INDIA HACKATHON 2018.
- Received token of appreciation for volunteering in organizing numerous vocational teaching activities for helping underprivileged children, Lok Kala Parishad, in the year 2017.