

Research Interests

My research centers on developing **Multi-Modal Foundational Models** for specialized domains, such as healthcare and biology, by addressing associated challenges:

- **Data Synthesis:** Using **LLMs** to generate synthetic data, with meta-learning-based feedback to optimize data generation for specialized tasks.
- **Continual Learning:** Enabling **LLMs** to adapt to new data streams while retaining prior knowledge, ensuring reliable performance in dynamic environments.
- **Security and Reliability:** Developing watermarking techniques for **LLMs** to reliably distinguish between human-written and machine-generated content, crucial for ensuring data integrity in specialized domains.

Education

2020 - Present **University of California San Diego (UC San Diego), CA.**

PhD, Electrical and Computer Engineering, GPA: 3.96

Advisor: Prof. Pengtao Xie

2016 - 2020 **Indian Institute of Technology Hyderabad (IIT H), India.**

Bachelor of Technology; Major: Electrical Engineering, Minor: Computer Science and Engineering, GPA: 9.68 (out of 10)

◦ **Academic Excellence Awardee** in academic terms 2016-2017, 2018-2019.

◦ **Second Highest CGPA** in B.Tech Program across all departments (240 students).

Work Experience

June–Sept 2023 **Research Intern, Apple**, Cupertino, CA, USA.

- Explored **multi-modal LLMs**, developing innovative methods to enhance user interactions and optimize user experience on iPhones.

June–Sept 2022 **Research Intern, Tencent AI Lab**, Bellevue, WA, USA.

- Enhanced **BERT** model performance by ~10% on extremely low-resource datasets through data augmentation, employing a task-dependent similarity matrix, with both this matrix and model weights optimized via a *meta-learning-based* approach.

May–Aug 2019 **Research Intern, Texas A&M University**, College Station, TX, USA.

Halliburton Engineering Global Program Scholar

- Developed a control algorithm for self-driving cars to navigate crowded environments by leveraging a pedestrian behavior model built with *Inverse Reinforcement Learning*.

May–Aug 2018 **Research Intern, L.V Prasad Eye Institute - Research Center**, Hyderabad, Telangana, India.

- Developed deep learning models for segmenting the optic disc and hard exudates in fundus images, and the choroid layer in OCT images. Built a web portal for doctors to access and utilize these tools.

Foundational Model Initiatives

Foundational Model for DNA 3-D Structure Prediction.

Developing a **foundational model** to predict missing gene coordinates, improving the accuracy of reconstructing the 3-D structure of DNA by analyzing both sequence data and folding patterns.

Skills

- **Machine Learning:** Hugging Face, PyTorch, TensorFlow, spaCy, NLTK, Scikit-Learn, pre-training and fine-tuning **LLMs**, Proximal Policy Optimization (PPO).
- **Programming Languages:** Python, C++, MATLAB, Java, SQL, Bash/Unix, Git.
- **Software Packages:** OpenCV, Jupyter, Pandas, Keras, Numpy, Matplotlib, seaborn.

Selected Publications

ICML, 2024 [Token-Specific Watermarking with Enhanced Detectability and Semantic Coherence for Large Language Models](#)

Sai Ashish Somayajula*, Mingjia Huo*, Youwei Liang, Ruisi Zhang, Farinaz Koushanfar, and Pengtao Xie

NAACL, 2024 [Generalizable and Stable Finetuning of Pretrained Language Models on Low-Resource Texts](#)

Sai Ashish Somayajula, Youwei Liang, Abhishek Singh, Li Zhang, and Pengtao Xie

- NAACL, 2024 [AutoLoRA: Automatically Tuning Matrix Ranks in Low-Rank Adaptation Based on Meta Learning](#)
Ruiyi Zhang*, Rushi Qiang*, **Sai Ashish Somayajula**, and Pengtao Xie
- Scientific Reports, Nature, 2024 [Improving Long COVID-Related Text Classification: A Novel End-to-End Domain-Adaptive Paraphrasing Framework](#)
Sai Ashish Somayajula, Onkar Litake, Youwei Liang, Ramtin Hosseini, Shamim Nemati, David O. Wilson, Robert N. Weinreb, Atul Malhotra, and Pengtao Xie
- Scientific Reports, Nature, 2024 [Improving Image Classification of Gastrointestinal Endoscopy Using Curriculum Self-Supervised Learning](#)
Han Guo, **Sai Ashish Somayajula**, Ramtin Hosseini, and Pengtao Xie
- ACL, 2023 [Bi-level Finetuning with Task-dependent Similarity Structure for Low-resource Training](#)
Sai Ashish Somayajula, Lifeng Jin, Linfeng Song, Haitao Mi, and Dong Yu
- TACL, 2022 [A Multi-Level Optimization Framework for End-to-End Text Augmentation](#)
Sai Ashish Somayajula, Linfeng Song, and Pengtao Xie

Awards and Honors

- o Travel grant for ICML 2024, NAACL 2024.
- o Awarded the *Jacob School of Engineering Departmental Fellowship* for PhD studies at UC San Diego 2020.
- o Microsoft Azure Award winner at the [Engineering the Eye-2018 Hackathon](#) 2018.
- o Runners-up in the PwC Challenge on “Smart Cities” at [Megathon](#) 2017.
- o Bronze Medal in the National Science Olympiad 2012.
- o Secured a rank in the top 0.1% among 0.5 million students in the IIT-Joint Entrance Exam 2016.
- o Diploma in Indian Carnatic Music, 2016.

Leadership

- 2021 **ENLACE program**, *UC San Diego*.
Led a team of college and high school students on the project '*Deep Learning Algorithms for Disease Segmentation in Chest X-rays*'; managed data preprocessing, model development, and evaluation, while fostering collaboration and technical skills.
- 2020 **First-Year PhD Representative**, *ECE Graduate Student Council, UC San Diego*.
Elected to represent and advocate for the needs and concerns of first-year PhD students in the department.
- 2018 **Elektronika Club**, *Head, IIT H*.
Led the Electronics and Signal Processing Club, fostering innovation and critical thinking in electronics and signal processing among students.
- 2017 **College Counsel**, *Co-Founder*.
Launched a startup that guided top-ranking students from India's prestigious IIT entrance exam in selecting academic branches, using a personalized approach based on exam rank and personal career aspirations.
- 2018 **Marketing Team**, *Head, IIT H*.
Led efforts to engage investors in supporting student-led startup initiatives and entrepreneurial ventures.

Teaching, Mentoring, and Reviewer Experience

Teaching Assistant.

- Fall 2024 ECE 250- Random Processes (Graduate)
- WI 2024, 21 ECE 269- Linear Algebra and Application (Graduate)
- Spring 2023 ECE 208- Computational Evolutionary Biology (Graduate) – **Best TA 10/10 rating**
- Fall 2023 ECE 271A- Statistical Learning I (Graduate)
- WI 2022 ECE 285- Deep Generative Models (Graduate)
- Fall 2021 ECE 100- Linear Electronic Systems (Undergraduate)

Mentorship Experience, During my PhD, I mentored historically underrepresented and underprivileged students, guiding them to excel in research projects with a strong emphasis on skill development and professional growth.

- 2023-24 Mingjia Huo (PhD, UC San Diego)
- 2022-24 Han Guo (PhD, UC San Diego)
- 2023-24 Xin Gao (MS, UC San Diego)
- 2023-24 Bokai Hu (MS, UC San Diego)

2022-23 Onkar Litake (MS, UC San Diego → Machine Learning Engineer, Transcarent)

2022-23 Ankit Garg (MS, UC San Diego → Machine Learning Engineer, TikTok)

2021-23 Abhishek Singh (MS, UC San Diego → Machine Learning Engineer, Oracle)

Reviewer Experience.

NeurIPS (2022–2024), ACL-ARR (2023–2024), ICLR (2023–2024), ICML (2023–2024), AAAI (2022), ICASSP (2022)