Tania Chakraborty

Email: tania.rini@gmail.com | Website: taniaisarini.github.io | GitHub: @taniaisarini | Phone: (765) 637-1149

Seeking Machine Learning and NLP Internships

EDUCATION

Ph.D. in Natural Language Processing *Expected May 2027*
Purdue University, West Lafayette, IN Department of Computer Science
MS in Computer Science *May 2024*
Purdue University, West Lafayette, IN Department of Computer Science
B.Sc. in Electrical and Computer Engineering *May 2019*
Purdue University, West Lafayette, IN School of Engineering and Honors College

PUBLICATIONS

- **VIBE: Can a VLM Read the Room?** Tania Chakraborty, Eylon Caplan, Dan Goldwasser. *Findings of EMNLP 2025*, Suzhou, China. (research at Purdue)
- **Splits! A Flexible Dataset and Evaluation Framework for Sociocultural Linguistic Investigation** Eylon Caplan, Tania Chakraborty, Dan Goldwasser. *Preprint, under review.* (research at Purdue)
- **A Large Scale Low-Resource Pronunciation Data Set Mined From Wikipedia** Tania Chakraborty, Manasa Prasad, Theresa Breiner, Sandy Ritchie, Daan van Esch. *arXiv preprint, 2021.* (research at Google)

SKILLS

AI/ML: Deep Learning Methods, LLMs, VLMs, multimodal NLP, multilingual models, AI agents, reasoning, computational social science, data pipelines, G2P modeling, model fine-tuning, prompt engineering
Programming: Python, C++, Java, C, JavaScript, PyTorch, Hugging Face, Pandas, NumPy, SQL
Select Classes: Natural Language Processing, Linear Algebra, Algorithms, Statistical Machine Learning
Languages: English (fluent), Bengali (fluent), Hindi (fluent)

INDUSTRY WORK EXPERIENCE

Google, Mountain View, CA | Software Engineer (Google TV) *Jun 2020 – Jul 2022*
- Designed, deployed and maintained back-end infrastructure for Play Movies & TV and Google TV.
- Designed and launched new Google TV notification system.
- Contributed to launch of Google TV on iOS, ensuring cross-platform feature parity.
- Debugged and resolved production issues through on-call rotations; communicated findings to cross-functional teams.

Google, Mountain View, CA | Engineering Resident (Languages & Linguistics) *Jan 2020 – Jun 2020*
- Built large-scale data processing pipelines to extract pronunciation and linguistic patterns from Wikipedia.
- Automated rule generation (grapheme-to-phoneme mappings), reducing manual linguistic workload.
- Produced dataset to improve multilingual and low-resource AI applications, enabling ML model training and evaluation.

Google, Mountain View, CA | Engineering Resident, Android Security & Privacy *Jul 2019 – Jan 2020*
- Developed monitoring framework and tracking dashboard for Android app safety system.

Oracle, Bangalore, India | Software Engineer Intern (Oracle Banking Platform) *May 2018 – Jul 2018*
- Developed rule-based ID verification system for Oracle's Intelligent Document Verification platform.
- Built a flexible parser for multiple ID formats, increasing adaptability and reducing manual effort.

RESEARCH EXPERIENCE

PhD Student, NLP Purdue University, West Lafayette, IN *August 2023 – Present*
- Currently working on multimodal NLP methods with a focus on vision. Involved in several projects to probe and improve VLM performance on general reasoning as well as social reasoning tasks.
- Also working on multilingual and multicultural abilities in LLMs and VLMs.
- Currently a mentor for a project to apply SOTA AI systems to wildlife conservation efforts.

SWE Rotation, Next Billion Users Google, Mountain View, CA *Jan 2020 – Jun 2020*
- Researched and prototyped low-resource language technologies, creating pipelines for multilingual data extraction.
- Collaborated with linguists and engineers to expand coverage for underrepresented languages.
- Contributed the *Large-Scale Low-Resource Pronunciation Dataset* (arXiv 2021).

TEACHING AND LEADERSHIP

Graduate Teaching Assistant | Purdue University *January 2023 – Present*
- (AI Forge) Currently Head GTA for AI Forge, Purdue's new Generative AI initiative.
- (AI Forge) Designed a multimodal AI research project for students to apply SOTA AI systems to conservation efforts.
- (C Programming) GTA for C programming class, conducted labs and office hours.

Undergraduate Teaching Assistant | C Programming, Purdue University *January 2016 – May 2019*
- Conducted labs and office hours for students.