Tania Chakraborty

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

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.
- **Splits! A Flexible Dataset and Evaluation Framework for Sociocultural Linguistic Investigation** Eylon Caplan, Tania Chakraborty, Dan Goldwasser. *Preprint, under review.*
- **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.*

WORK EXPERIENCE

Software Engineer, Google TV *Google, Mountain View, CA* *Jun 2020 – Jul 2022*
- Designed and developed back-end infrastructure, including an improved, maintainable, and reliable notification system.
- Contributed to the launch of Google TV on iOS, ensuring cross-platform feature parity.
- Participated in on-call rotations, troubleshooting and resolving production incidents.

Engineering Resident, Languages & Linguistics *Google, Mountain View, CA* *Jan 2020 – Jun 2020*
- Built large-scale crawler and data processing pipelines to mine pronunciation and linguistic data from Wikipedia.
- Automated generation of language rules (G2P mappings), reducing manual effort by 100% when resources existed.
- Produced structured linguistic datasets to improve multilingual and low-resource language applications.
- Delivered production-ready pipelines for internal use in multilingual linguistic data processing.

Engineering Resident, Android Security & Privacy *Google, Mountain View, CA* *Jul 2019 – Jan 2020*
- Developed nightly monitoring framework for Android app safety system, improving error detection within the team.
- Built internal dashboard for real-time tracking of app safety metrics.

Software Engineer Intern *Oracle, Bangalore, India* *May 2018 – Jul 2018*
- Developed rule-based automatic ID verification system for Oracle's Intelligent Document Verification platform.
- Built flexible parser to handle multiple ID formats (e.g., driver licenses), improving system adaptability.

RESEARCH EXPERIENCE

PhD Student *Purdue University, West Lafayette, IN* *Aug 2022 – Present*
- **Research Focus:** Intersection of AI/NLP, computational social science and reasoning.
- Currently working on multilingual and multicultural evaluation frameworks for LLMs, addressing cross-linguistic and cross-cultural variation in NLP.
- Developed multimodal NLP methods, including applying vision-language models for sociocultural communication.

SWE Rotation *Google, Mountain View* *Jan 2020 – Jun 2020*
- **Research Focus:** Low-resource language technologies in the context of Google's *Next Billion Users (NBU)* initiative.
- Investigated data-driven approaches to expand linguistic coverage for underrepresented languages.
- Collaborated with linguists to design scalable methods for multilingual knowledge extraction.
- Work contributed to subsequent research on low-resource pronunciation datasets (arXiv 2021).

TEACHING EXPERIENCE

Course Developer / Mentor *AI Forge, Purdue University* *May 2025 – Present*
- Designed generative AI–based course project for senior undergraduates and early graduate students.
- Facilitated class projects and mentored individual AI projects.

Graduate Teaching Assistant *C Programming for Engineers, Purdue* *Jan 2023 – May 2025*
- Conducted programming labs and office hours for students.
- Mentored undergraduate TAs in lab instruction.

Undergraduate Teaching Assistant *C Programming for Engineers, Purdue* *Jan 2015 – May 2019*
- Conducted labs and undergraduate TA office hours.

SKILLS

Technical: Python, C++, Java, C, PyTorch, Hugging Face Transformers, NLTK, Pandas, NumPy, G2P modeling
Research: LLMs, VLMs, AI Agent systems, multilingual LLMs, multicultural LLMs
Languages: English (fluent), Bengali (fluent), Hindi (fluent)