Shreya Havaldar

PhD Candidate | Researching LLM Alignment & Evaluation

shreyahavaldar@gmail.com | shreyahavaldar.com

Education

University of Pennsylvania August 2021 - May 2026
PhD Computer and Information Science
Advisors: Lyle Ungar, Eric Wong

University of Southern California August 2017 - May 2021
B.S. Computer Science, B.S. Applied & Computational Mathematics
Advisor: Morteza Dehghani
Summa cum laude

Research Statement

I study how to make LLMs adapt to the social and cultural contexts in which they are used. My work has developed evaluation frameworks that measure how models represent and respond to context-dependent phenomena, such as emotion, linguistic style, implied language, and sociocultural norms. I have also developed alignment methods that improve model behavior, such as preserving speaker intent in translation and improving the safety of mental health advice.

Awards & Honors

- MIT Rising Stars in EECS (2025)
- NSF Graduate Research Fellowship (2023)
- Best Paper Award at WASSA @ ACL (2023)
- Area Chair Award at AACL (2023)
- USC Computer Science Outstanding Student Award (2021)
- USC Computer Science Outstanding Service Award (2021)
- USC Presidential Scholarship (∼$30,000/year)

Selected Publications

** indicates equal contribution; full list on **Google Scholar**.*

Havaldar, S., Rai, S., Cho, Y.M., & Ungar, L. (2025). Culturally-Aware Conversations: A Framework & Benchmark for LLMs. *4th Workshop on Bridging Human-Computer Interaction and Natural Language Processing (EMNLP)*.

Brown, D., Balehannina, P., Jin, H., **Havaldar, S.**, Hassani, H., & Wong, E. (2025). Adaptively Profiling Models with Task Elicitation. *EMNLP*.

Havaldar, S., Alvari, H., Palowitch, J., Hosseini, M. J., Buthpitiya, S., & Fabrikant, A. (2025). Entailed Between the Lines: Incorporating Implication into NLI. *ACL*.

Havaldar, S.*, Stein, A.*, Wong, E., & Ungar, L. (2025). Towards Style Alignment in Cross-Cultural Translation. *ACL*.

Jin, H.*, **Havaldar, S.***, Kim, C.*, Xue, A.*, You, W.*, et al. (2025). The FIX Benchmark: Extracting Features Interpretable to eXperts. *DMLR*.

Havaldar, S., Giorgi, S., Talhelm, T., Guntuku, S. C., & Ungar, L. (2024). Building Knowledge-Guided Lexica to Model Cultural Variation. *NAACL*.

Havaldar, S., Pressimone, M., Wong, E., Ungar, L. (2023). Comparing Styles across Languages. *EMNLP*.

Havaldar, S., Rai, S., Singhal, B., Liu, L., Guntuku, S. C., & Ungar, L. (2023). Multilingual Language Models are not Multicultural: A Case Study in Emotion. *13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis (ACL)*. **Best Paper Award**.

Lyu, Q.*, **Havaldar, S.***, Stein, A.*, Zhang, L., Rao, D., Wong, E., Apidianaki, M., & Callison-Burch, C. (2023). Faithful Chain-of-Thought Reasoning. *AACL*. **Area Chair Award**.

Industry Experience

Spotify (New York City, NY) Summer 2025
Research Scientist Intern

- Modeled user signals from interactive comments/feedback on podcast episodes; analyzed how user subjectivity influences interaction on Spotify's platform.

Google DeepMind (Mountainview, CA) May - November 2024
Student Researcher

- Benchmarked LLM performance in implied language understanding; analyzed how conversational and sociocultural norms influence implication.
- Developed a large-scale dataset to benchmark cultural understanding of various Gemini models (internal project).

Microsoft (Bellevue, WA | Remote) Summer 2020, Summer 2021
Software Engineering & Data Science Intern

- Configured a topic extraction pipeline to include database ingestion to include 10,000+ tenant-level features for the Microsoft Search, Assistance, and Intelligence (MSAI) team.
- Led research and experimentation for a proposed topic classification model using various supervised learning and feature selection methods.

Teaching Experience

CS1900: Introduction to Designing LLM Agents January 2026 – present
Creator and Lecturer ***(Course Website)***
UPenn Department of Computer Science

- Created and taught a new undergraduate course at Penn on methods for interacting with LLMs and building LLM agents.

CS3990: Mathematics of Machine Learning
CS5300: Computational Linguistics August 2022 – May 2023
Teaching Assistant
UPenn Department of Computer Science

- Assisted with curriculum development and testing, and held office hours to support students with theoretical concepts and assignments.

CS170: Discrete Methods in Computer Science
CS270: Introduction to Algorithms Fall 2018 – May 2021
Teaching Assistant
USC Department of Computer Science

- Designed and led weekly discussion sections, and held office hours to support students with theoretical concepts and assignments.

Leadership

Women in Machine Learning (WiML) @ UPenn Spring 2024 - present
President, Founder

- Started a chapter of WiML at UPenn to create and foster a community for underrepresented genders in ML; organizing meetups and events for PhD students, postdocs, and faculty.

AthenaHacks Spring 2019 - May 2021
Logistics Organizer

- Worked on the Logistics committee to plan Southern California's largest all-female hackathon and empower high school and undergraduate women in the technology field.

USC Undergraduate Center for AI in Society (CAIS++) Fall 2019 - May 2021
Curriculum Lead

- Held weekly ML lessons for a cohort of 8 undergraduate students, focused on the fundamentals of applied and theoretical machine learning.