

Curriculum Vitae

Shreya Havaldar

shreyah@seas.upenn.edu | Philadelphia, PA | shreyahavaldar.com

EDUCATION

University of Pennsylvania

PhD Computer and Information Science

Advisors: Lyle Ungar, Eric Wong

August 2021 - May 2026 (anticipated)

University of Southern California

B.S. Computer Science

B.S. Applied & Computational Mathematics

Advisor: Morteza Dehghani

August 2017 - May 2021

RESEARCH STATEMENT

I specialize in cross-cultural NLP, focusing on quantifying and mitigating Anglocentric bias in modern LMs. I am broadly interested in helping LLMs understand subjective constructs that vary across people and cultures, such as linguistic style (e.g. emotion, politeness), implied language, and social norms. I am currently working on: (1) Quantifying how translation fails to preserve components of linguistic style, (2) Subjectivity in implication understanding, and (3) Benchmarking cultural awareness in LLMs.

AWARDS & HONORS

NSF Graduate Research Fellowship (2023)

Best Paper Award at WASSA @ ACL (2023)

Area Chair Award: Interpretability and Analysis of Models for NLP @ AACL (2023)

Summa Cum Laude (2021)

USC Computer Science Outstanding Student Award (2021)

USC Computer Science Outstanding Service Award (2021)

Cisco Engineering Excellence Scholarship (\$5,000/year)

USC Presidential Scholarship (~\$30,000/year)

National Merit Scholarship (\$1,000/year)

Brittingham Social Enterprise Lab Scholar (2020)

Grace Hopper Research Scholar (2018, 2019)

PREPRINTS (* indicates equal contribution)

Jin, H., **Havaldar, S.**, Kim, C., Xue, A., You, W., et al. (2024). The FIX Benchmark: Extracting Features Interpretable to eXperts. *Under Review*.

Rai, S., Zaveri, KZ., **Havaldar, S.**, Nema, S., L Ungar, L., & Guntuku, S.C. (2024) A Cross-Cultural Analysis of Social Norms in Bollywood and Hollywood Movies. *Under Review*.

Giorgi, S., **Havaldar, S.**, Ahmed, F., Akhtar, Z., Vaidya, S., Pan, G., Ungar, L., Schwartz, H. A., & Sedoc, J. (2024). Human-Centered Metrics for Dialog System Evaluation. *Under Review*.

SELECTED PUBLICATIONS (* indicates equal contribution)

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.*, You, W.*, Ungar, L., Wong, E., (2023). Visual Topics via Visual Vocabularies. *XAI in Action: Past, Present, and Future Applications (NeurIPS)*.

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. *ACL*. **Area Chair Award**.

Havaldar, S., Stein, A., Wong, E., & Ungar, L. (2023). TopEx: Topic-based Explanations for Model Comparison. *ICLR (Tiny Papers Track)*.

Mostafazadeh Davani, A., Atari, M., Kennedy, B., **Havaldar, S.**, & Dehghani, M. (2020). Hatred is in the Eye of the Annotator: Hate Speech Classifiers Learn Human-Like Social Stereotypes. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*.

Hoover, J., Portillo-Wightman, G., Yeh, L., **Havaldar, S.**, Davani, A. M., et al. (2020). Moral Foundations Twitter Corpus: A Collection of 35k Tweets Annotated for Moral Sentiment. *Social Psychological and Personality Science*.

TEACHING EXPERIENCE

UPenn Department of Computer Science August 2022 - Present

Teaching Assistant: CS5300 (Computational Linguistics), CS3990 (Mathematics of Machine Learning)

- Assisting with creating/testing curriculum, holding weekly office hours to help students with theoretical understanding and homework assignments

USC Department of Computer Science Fall 2018 - May 2021

Teaching Assistant: CS170 (Discrete Methods in Computer Science), CS270 (Intro to Algorithms)

- Designed and led weekly discussion sections, held office hours to help students with theoretical understanding and homework assignments

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.

USC Joint Educational Project Fall 2018 - Spring 2019

Undergraduate Teacher

- Designed and implemented a comprehensive curriculum to teach local LA elementary school students about introductory computer science concepts through MIT's Scratch

INDUSTRY EXPERIENCE

Google Deepmind (Mountainview, CA) May 2024 - present

Student Researcher

- Working on measuring LLM performance in implied language understanding + analyzing how personal/cultural subjectivity influences implication.
- Created the INLI benchmark that focuses on implied entailment in the conversational and situational domains (preprint under review).

Microsoft - MSAI (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 team.
- Led research and experimentation for a proposed topic classification model using various supervised learning methods and domain-specific feature selection.

Microsoft - Azure Identity (Redmond, WA) Summer 2019

Software Engineering Intern

- Worked on Microsoft Identity's mid-tier graph-based database system, added functionality to implement link counts and complex link querying capabilities.

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.

Association for the Advancement of Artificial Intelligence

Fall 2017 - Spring 2020

President, Event Coordinator

- Led the AAAI USC student chapter; planned AI-focused workshops, hackathons, and outreach events to encourage USC students from all backgrounds to explore AI
- Developed workshops on perceptrons, neural networks, and computer vision

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/university women in the technology field

Environmental Student Assembly

Fall 2019 - May 2021

Director of Technology

- Redesigned and maintained the ESA website, worked with USC's undergraduate student government to facilitate sustainability on campus
- Assisted with the organization of a campus-wide composting initiative

AAAI Elementary Student Outreach

Spring 2018 - Spring 2019

Program Director

- Worked with Dr. Sheila Tejada to lead AI outreach events for local children during the AAAI-18 and AAAI-19 international conferences

Society of Women Engineers

Fall 2017 - Spring 2019

High School Mentor

- Acted as a year-round mentor to three girls from a Los Angeles high school; provided guidance on standardized tests and the college application process as an engineering major