Brihi Joshi
Email : brihijos@usc.edu
Website : brihijoshi.github.io

Research Interests

Trustworthy Natural Language Processing

- Generating, Utilizing and Evaluating Explanations for NLP tasks
- Interdisciplinary Methods (CS, Psychology, Cog Sci) for Human-centric Interpretability

Education

- **University of Southern California** — Los Angeles, CA
  Doctor of Philosophy, Computer Science (USC Annenberg and Amazon ML PhD Fellow) — *Aug. 2021 – Present*
- **Indraprastha Institute of Information Technology** — New Delhi, India
  Bachelor of Technology, Computer Science and Engineering; CGPA: 9.30/10 — *Aug. 2016 – Dec 2020*
  - Received the Innovative Student Projects Award for **best thesis in Computer Science** from the Indian National Academy of Engineering. One of the highest honors for undergraduates in India.

Publications and Preprints

- Sahana Ramnath, **Brihi Joshi**, Skyler Hallinan, Ximing Lu, Liunian Harold Li, Aaron Chan, Jack Hessel, Yejin Choi, Xiang Ren. Tailoring Self-Rationalizers with Multi-Reward Distillation. *Under Submission*
- Aaron Chan*, Zhiyuan Zeng*, Wyatt Lake, **Brihi Joshi**, Hanjie Chen, Xiang Ren. KNIFE: Knowledge Distillation with Free-Text Rationales. *TrustML-(un)Limited@ICLR, 2023 and Under Submission*
- **Brihi Joshi***, Ziyi Liu*, Sahana Ramnath, Aaron Chan, Zhewei Tong, Shaoliang Nie, Qifan Wang, Yejin Choi and Xiang Ren. Are Machine Rationales (Not) Useful to Humans? Measuring and Improving Human Utility of Free-text Rationales. *ACL 2023 and trAIt@CHI 2023*
- Dong-Ho Lee*, Akshen Kadakia*, **Brihi Joshi**, Aaron Chan, Ziyi Liu, Takashi Shibuya, Ryosuke Mitani, Toshiyuki Sekiya, Jay Pujara, Xiang Ren. XMD: An End-to-End Framework for Interactive Explanation-Based Debugging of NLP Models. *ACL Demo Track 2023*
- **Brihi Joshi***, Aaron Chan*, Ziyi Liu*, Shaoliang Nie, Maziar Sanjabi, Hamed Firooz, Xiang Ren. ER-Test: Evaluating Explanation Regularization Methods for NLP Models. *Findings of EMNLP 2022 and TrustNLP@NAACL 2022*
- **Brihi Joshi**, Neil Shah, Francesco Barbieri and Leonardo Neves. The Devil is in the Details: Evaluating Limitations of Transformer-based Methods for Granular Tasks. In *The 28th ACM International Conference on Computational Linguistics (COLING 2020).*
- **Brihi Joshi***, Aditya Chetan*, Hridoy Sankar Dutta, Tanmoy Chakraborty. CoReRank: Ranking to Detect Users Involved in Blackmarket-based Collusive Retweeting Activities. In *The 12th ACM International Conference on Web Search and Data Mining (WSDM 2019).* (Acceptance Rate: 16%, CORE2018 A*)
- Udit Arora, Hridoy Sankar Dutta, **Brihi Joshi***, Aditya Chetan*, Tanmoy Chakraborty. Analyzing and Detecting Collusive Users Involved in Blackmarket Retweeting Activities. In *ACM Transactions on Intelligent Systems and Technology (TIST).* (Impact Factor: **3.971**)
- **Brihi Joshi***, Amogh Gulati*, Chirag Jain*, Jainendra Shukla. It's Not What They Play, It's What You Hear: Understanding Perceived vs. Induced Emotions in Hindustani Classical Music. In *22nd ACM International Conference on Multimodal Interaction, Late Breaking Reports (ICMI 2020).*
- **Brihi Joshi***, Shravika Mittal, Aditya Chetan. Did You "Read" the Next Episode? Using Textual Cues for Predicting Podcast Popularity. In *First Workshop on NLP for Music and Audio (NLP4MusA) at International Society for Music Information Retrieval Conference (ISMIR 2020).*

- Hridoy Sankar Dutta, **Brihi Joshi***, Aditya Chetan*, Tanmoy Chakraborty. Retweet Us, We Will Retweet You: Spotting Collusive Retweeters Involved in Blackmarket Services. In *The 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2018).* (Acceptance Rate: 15%)

- Nishtha Madaan, Gautam Singh, Sameep Mehta, Aditya Chetan*, **Brihi Joshi***. Generating Clues for Gender based Occupation De-biasing in Text `arXiv:1804.03839 [cs.CL]`

* Equal Contribution

Internships and Work Experience

- **Amazon** — Cambridge, MA
  Applied Science Intern, Alexa AI Natural Understanding: NLG Team — *May 2023 – Aug. 2023*
- **Goldman Sachs** — Bangalore, India
  Analyst (Full Time), Regulatory Engineering Team — *Dec. 2020 – July 2021*
- **Snap Inc.** — Los Angeles, CA
  Research Intern, Computational Social Science Team — *Sep. 2019 – Dec. 2019*
- **Goldman Sachs** — Bangalore, India
  Analyst Intern, Regulatory Engineering Team — *May 2019 - July 2019*
- **IBM Research** — New Delhi, India
  Research Intern, Fairness in ML Team — *May 2018 - July 2018*

Awards

- **Amazon ML PhD Fellowship**: Awarded for work in Explainable NLP, for AY 2023-24.
- **Indian National Academy of Engineering Undergraduate Thesis Award 2021**: Awarded for research done as a part of undergraduate thesis in the Computer and Information Sciences Domain.
- **Annenberg Fellowship**: Awarded for admission to the PhD program at the University of Southern California.
- **Snap Research Scholarship 2019**: Awarded for research done in the field of Machine Learning. Award includes 10,000 USD and an offer to intern at Snap Research, USA. Only scholar from India!
- **AAAI 2020 Undergraduate Consortium**: Accepted to present my thesis at the AAAI 2020 Undergraduate consortium. Includes scholarship to attend to attend AAAI 2020
- **Microsoft Research India Travel Grant**: Awarded travel support of 50000 INR for visiting WSDM 2019
- **ACM-W Scholarship**: Awarded travel support of 1200 USD for visiting WSDM 2019
- **Google Women Techmakers Scholarship, 2018**: Awarded to students who work for diversity and inclusion in the field of Computer Science.
- **Best Technical Poster Runner-up at GHCI 2018**: Received for the project, "Generating Clues for Gender based Occupation De-biasing in Text"
- **Dean's Award for Innovation R&D**: Awarded to students who work on Research projects beyond coursework. Awarded for the academic years 2016-17 and 2017-18.
- **Dean's List of Academic Affairs**: Awarded to students who demonstrate excellence in an academic year. Awarded for the academic year 2016-17 and 2018-19.
- **Grace Hoppers Celebration India (GHCI) Scholarship, 2018**: Awarded travel grant and scholarship to attend the GHCI conference.

Service and Leadership

- **Core Organizer, NLP With Friends**: Core Organizer of NLP with Friends, an online seminar series for PhD students to discuss all things NLP Research.
- **Director, Women Who Code Delhi**: Lead the Delhi Chapter of Women Who Code, a non-profit organisation for upliftment of minorities in techonology.
- **Workshop Organizer, Broadening Research Collaborations in ML, NeurIPS 2022**: PC member and organizer for new workshop at NeurIPS aimed towards making ML research more accessible.
- **Reviewer**: EMNLP 2022, ACL 2023, EMNLP 2023
- **Grant Writing**: *Utilizing Explanations for Model Refinement* in Alexa: Fairness in AI 2022 award.
- **Mentoring**: Ziyi Liu (USC MS CS, Joining USC CS PhD in Fall 2023), Zhewei Tong (Viterbi Tsinghua Undergraduate Summer Research Program, Joining CMU MS CS in Fall 2023), Pushpdeep Singh (Indo-US Science and Technology Forum).