



Shaz Furniturewala

✉ mohammadshaz529@gmail.com |  linkedin.com/in/shazfurniturewala |  Google Scholar

NLP, Mechanistic Interpretability, Fairness in Language Models

EDUCATION

- **Birla Institute of Technology and Science, Pilani** Rajasthan, India
Bachelor of Engineering - Computer Science; GPA: 8.91/10 *July 2020 - July 2024*

RESEARCH EXPERIENCE

- **National University of Singapore** | *Research Intern* Aug 2023 – Jul 2024
 - **Human Alignment in Conversational LLMs (WASSA @ ACL 2024):**
Constructed parallel human-LLM conversation corpora and developed new stylistic and psychometric alignment metrics. Analyzed decoding strategies (beam size, nucleus sampling) and identified parameter regimes that improve conversational human-likeness.
 - **Fairness and Trust in LLMs (EMNLP 2024):**
Designed a structured self-critique prompting framework to reduce demographic bias in LLM outputs without model parameter access. Demonstrated fairer generations across multiple datasets while maintaining task performance.
 - **Modeling Emotions via Cognitive Appraisal Theory (ACL 2024 Findings):**
Implemented multi-task models that incorporate cognitive appraisal dimensions as auxiliary supervision to improve emotion and post-consumption behavior prediction. Showed improved prediction quality over text-only baselines.
- **Adobe MDSR Lab** | *Machine Learning Research Intern* May 2023 – Aug 2023
 - **Counterfactually Aware Fair Text Generation (AAAI 2024):**
Co-developed CAFIE, a plug-and-play inference-time framework that incorporates counterfactual demographic prompts to equalize next-token distributions across groups. Achieved fairer and more equitable text generation without model retraining.

PUBLICATIONS

- **Shaz Furniturewala** and Arkaitz Zubiaga. Towards Inclusive Toxic Content Moderation: Addressing Vulnerabilities to Adversarial Attacks in Toxicity Classifiers Tackling LLM-generated Content. *Under Review at ICWSM, 2026*
- **Shaz Furniturewala**, Surgan Jandial, Abhinav Java, Pragyan Banerjee, Simra Shahid, Sumit Bhatia, and Kokil Jaidka. Thinking Fair and Slow: On the Efficacy of Structured Prompts for Debiasing Language Models. *EMNLP, 2024*
- Pragyan Banerjee, Abhinav Java, Surgan Jandial, Simra Shahid, **Shaz Furniturewala**, Balaji Krishnamurthy, and Sumit Bhatia. All Should Be Equal in the Eyes of LMs: Counterfactually Aware Fair Text Generation. *AAAI, 2024*
- Gerard Yeo, **Shaz Furniturewala**, and Kokil Jaidka. Beyond Text: Leveraging Multi-Task Learning and Cognitive Appraisal Theory for Post-Purchase Intention Analysis. *ACL Findings, 2024*
- **Shaz Furniturewala** and Kokil Jaidka. Impact of Decoding Methods on Human Alignment of Conversational LLMs. *WASSA, ACL, 2024*
- Anshika Gupta, **Shaz Furniturewala**, Vijay Kumari, and Yashvardhan Sharma. Rhetorical Role Labelling of Legal Documents using Transformers and Graph Neural Networks. *ACL, 2023*
- **Shaz Furniturewala**, Racchit Jain, Vijay Kumari, and Yashvardhan Sharma. Legal Text Classification and Summarization using Transformers and Joint Text Features. *FIRE (Working Notes), 2021*

INDUSTRY EXPERIENCE

- **AlphaSense** | *AI Research Engineer* July 2024 - Present
 - **Generative Search Explanations:**
Integrated natural language explanations for answers generated by all the LLM-powered generative search tools at AlphaSense.
 - **GenAI Integration:**
Prompt engineered and deployed a variety of LLM flows extracting insights and generating summaries from financial documents. This work resulted in 4 types of summaries being generated for every earnings and conference call transcript available on AlphaSense.
 - **Company Industry Classifier:**
Built an end-to-end pipeline to identify the industry of a company from its homepage. Curated, scraped and cleaned data, and finetuned a multi-level transformer-based classifier for this task. Achieved 97% accuracy.
- **Clientell Technologies** | *Machine Learning Intern* May 2022 - Aug 2022
 - **Email Topic Categorization Model:**
Trained a model to classify client emails into over 30 business categories for sales and customer support.
 - **Meeting Email Detection Tool:**
Worked on a tool that utilized machine learning to detect emails related to meetings and assisted in scheduling them.

TEACHING EXPERIENCE

- **Graph Mining** | *BITS Pilani* Aug 2023 - Dec 2023
 - **Weekly Labs:**
Conducted weekly labs teaching 36 students to implement theoretical concepts taught in class using NetworkX.
 - **Assignments:**
Prepared and graded assignments and worksheets based on what was taught in the previous week's lab.
- **Neural Networks and Fuzzy Logic** | *BITS Pilani* Aug 2022 - Dec 2022
 - **Introductory Lecture:**
Conducted an introductory lecture for 117 students explaining fundamental concepts of neural networks such as forward and backward propagation.
 - **Assignments:**
Designed and graded 3 assignments for students in the format of Jupyter notebooks implementing various types of neural networks.
 - **Student Mentorship:**
Mentored a group of 6 students for their semester project that involved implementing a recent ML paper.

AWARDS AND DISTINCTIONS

- National Achiever in the National Science Talent Search Examination 2019 (All India Rank 77).
- JEE Main 2020 percentile: 99.82% (99.92% in Physics).
- All India Rank 3196/150838 in JEE Advanced 2020 (99.979 percentile).
- One of 4 applicants out of over a 100 selected for the ICWSM Global Initiative 2025
- Zonal representative at the Indian National Olympiad in Informatics 2019

LEADERSHIP/EXTRACURRICULAR

- **Secretary, The Debating Society** | *BITS Pilani, India* May 2022 - May 2023
- **Chief Editor, Cactus Flower Magazine** | *BITS Pilani, India* June 2022 - June 2023
- **Election Commissioner, Society for Student Mess Services** | *BITS Pilani, India* Feb 2022 - Feb 2023