

Nazneen Fatema Rajani

CONTACT INFORMATION	Hugging Face Palo Alto, CA <i>Email:</i> emailnazneen@gmail.com <i>Website:</i> https://www.nazneenrajani.com/ Google Scholar: https://scholar.google.com/citations?user=eIRG81YAAAAJ&hl=en
RESEARCH INTERESTS	AI Safety, Robustness, Evaluation, Interpretability, Factual consistency, Commonsense Reasoning
EDUCATION	University of Texas at Austin , Austin, TX, USA Ph.D., Computer Science, 2014 - 2018 <ul style="list-style-type: none">• Advisor: Ray Mooney• Thesis: Explainable Improved Ensembling for Natural Language and Vision University of Texas at Austin , Austin, TX, USA M.S., Computer Science, 2012 - 2014 <ul style="list-style-type: none">• Advisor: Jason Baldridge• Thesis: New topic detection using topical alignment Birla Institute of Technology and Science , Pilani, India MSc. (Tech.), Information Systems, 2007 - 2011 <ul style="list-style-type: none">• Thesis: Sentiment Analysis for Tweets• Advisor: Onkar Dabeer (TIFR, Mumbai)
PROFESSIONAL EXPERIENCE	Hugging Face April 2022 - Present Robustness Research Lead Building tools for better model and data understanding for LLMs Salesforce Research, USA Jan 2019 – March 2022 Senior Research Scientist Led a team of 4 RS with a focus on building robust NLG systems IBM Watson Research, USA May 2016 – August 2016 Watson Research Staff Intern Ensembling for entity linking in the medical domain eBay Research Labs, USA May 2014 – August 2014 <i>Research Scientist Intern</i> Identifying "Interestingness" using Pinterest text data Qualcomm Research, USA May 2013 – August 2013 <i>Research Intern</i> Motion detection and classification
PUBLICATIONS Most up-to-date list on Google Scholar	Anamaria Crisan, Margaret Drouhard, Jesse Vig, Nazneen Rajani . Interactive Model Cards: A Human-Centered Approach to Model Documentation. ACM FAccT'22. Benjamin Newman, Prafulla Kumar Choubey, Nazneen Fatema Rajani . P-Adapters: Robustly Extracting Factual Information from Language Models with Diverse Prompts. ICLR '22. Han Guo, Nazneen Fatema Rajani , Peter Hase, Mohit Bansal, Caiming Xiong. FastIF: Scalable Influence Functions for Efficient Model Interpretation and Debugging. EMNLP '21.

Ben Krause*, Akhilesh Gotmare*, Bryan McCann, Nitish Shirish Keskar, Shafiq Joty, Richard Socher, **Nazneen Fatema Rajani**. GeDi: Generative Discriminator Guided Sequence Generation. EMNLP '21 Findings.

Jesse Vig, Wojciech Kryscinski, Karan Goel, **Nazneen Fatema Rajani**. SummVis: Interactive Visual Analysis of Models, Data, and Evaluation for Text Summarization. ACL '21 demo.

Karan Goel, Laurel Orr, **Nazneen Fatema Rajani**, Jesse Vig, Christopher Ré. Goodwill Hunting: Analyzing and Repurposing Off-the-Shelf Named Entity Linking Systems. NAACL '21 industry track.

Karan Goel*, **Nazneen Fatema Rajani***, Jesse Vig, Zachary Taschdjian, Mohit Bansal, Christopher Ré. Robustness Gym: Unifying the NLP Evaluation Landscape. NAACL'21 demo.

Shiyang Li, Semih Yavuz, Kazuma Hashimoto, Jia Li, Tong Niu, **Nazneen Fatema Rajani**, Xifeng Yan, Yingbo Zhou, Caiming Xiong. CoCo: Controllable Counterfactuals for Evaluating Dialogue State Trackers. ICLR '20.

Jesse Vig, Ali Madani, Lav Varshney and **Nazneen Fatema Rajani**. BERTology Meets Biology: Interpreting Attention in Protein Language Models. ICLR '20.

Tianlu Wang, Xi Victoria Lin, **Nazneen Fatema Rajani**, Bryan McCann, Vicente Ordonez, Caiming Xiong. Double-Hard Debias: Tailoring Word Embeddings for Gender Bias Mitigation. ACL '20.

Nazneen Fatema Rajani*, Rui Zhang*, Yi Chern Tan, Stephan Zheng, Jeremy Weiss, Aadit Vyas, Abhijit Gupta, Caiming Xiong, Richard Socher, Dragomir Radev. ESPRIT: Explaining Solutions to Physical Reasoning Tasks. ACL '20. * Indicates equal contribution.

Jay DeYoung*, Sarthak Jain*, **Nazneen Fatema Rajani***, Eric Lehman, Caiming Xiong, Richard Socher, Byron Wallace. ERASER: A Benchmark to Evaluate Rationalized NLP Models. ACL '20. * Indicates equal contribution.

Nazneen Fatema Rajani, Bryan McCann, Caiming Xiong and Richard Socher. Explain Yourself! Leveraging Language Models for Commonsense Reasoning. ACL '19.

Explainable Improved Ensembling for Natural Language and Vision (**Ph.D. Thesis**)

Nazneen Fatema Rajani and Raymond J. Mooney. Ensembling Visual Explanations. *Book chapter* for Explainable and Interpretable Models in CV and ML. Published by Springer. November 2018. https://link.springer.com/chapter/10.1007/978-3-319-98131-4_7

Nazneen Fatema Rajani and Raymond J. Mooney. Stacking with Auxiliary Features for Visual Question Answering. NAACL '18.

Nazneen Fatema Rajani and Raymond J. Mooney. Ensembling Visual Explanations for VQA. In Proceedings of the NIPS 2017 workshop on Visually-Grounded Interaction and Language (ViGIL).

Nazneen Fatema Rajani and Raymond J. Mooney. Using Explanations to Improve Ensembling of Visual Question Answering Systems. In Proceedings of the IJCAI 2017 Workshop on Explainable Artificial Intelligence (XAI).

Nazneen Fatema Rajani, Mihaela Bornea and Ken Barker. Stacking With Auxiliary Features for Entity Linking in the Medical Domain. In Proceedings of the ACL 2017 workshop on BioNLP.

Nazneen Fatema Rajani and Raymond Mooney. Stacking with Auxiliary Features. IJCAI '17.

Nazneen Fatema Rajani and Raymond Mooney. Stacking with Auxiliary Features: Improved Ensembling for Natural Language and Vision. **PhD Proposal**. November 2016.

Nazneen Fatema Rajani and Raymond Mooney. Combining Supervised and Unsupervised Ensembles for Knowledge Base Population. EMNLP '16.

Nazneen Fatema Rajani, Vidhoon Vishwanathan, Yinon Bentor and Raymond Mooney. Stacking Ensembles of Information Extractors for Knowledge Base Population. ACL '15.

Nazneen Fatema Rajani, Kate McArdle and Inderjit Dhillon. Parallel k Nearest Neighbor Graph Construction Using Tree-Based Data Structures. KDD workshop 2015.

Nazneen Fatema Rajani, Khashayar Rohanimanesh, Eduardo Oliveira and Aamer Hydrie. Identifying Interestingness in Fashion E-commerce using Pinterest Data. KDD workshop 2015.

Nazneen Fatema Rajani, Kate McArdle, Jason Baldrige. Extracting Topics Based on Authors, Recipients and Content in Microblogs. ACM SIGIR '14.

Nazneen Fatema Rajani, Rajoshi Biswas, Gaurav Dar and Ramesha C. K. Solution to the Tic-Tac-Toe problem using Hamming Distance Approach in a Neural Network. ISMS 2011.

HONORS AND
AWARDS

VentureBeat finalist for AI Research awards, 2020
Microsoft Women's Hackathon winner, 2014
eBay Scholarship 2013
Google India Women in Engineering Award 2011

TALKS

Invited talks at **ICML '21 UDL** and **EMNLP '21 SustainNLP** workshops.
Explainable Physical Reasoning. Invited talk at **Yale CS Dept.** March 2020.
Commonsense Reasoning using Explanations. Invited talk at **Toronto Machine Learning Seminar (TMLS)**. November 2019.
Explainable AI and Trust. Part of Research keynote at **Salesforce Dreamforce 2019**.
Leveraging Explanations for Performance and Generalization in NLP and RL. **Forum for AI (FAI)** talk at UT Austin. October 2019.
"How XAI influences ethical policies?" Invited talk for **UNESCO** session at Indaba 2019. Nairobi, Kenya.
Supervised and unsupervised ensemble for cold start slot filling. Selected talk at NIST TAC Meeting, Gaithersburg, MD. November 2015.
Ensembling slot filler systems. Selected talk at DEFT PI Meeting, Boulder, CO. May 2015.

TEACHING
EXPERIENCE

Online live course with Corise

[Interpreting ML Models](#)

Jan 2023

University of Texas at Austin, Austin, TX, USA

Spring 2013

Teaching Assistant

Introduction to Programming

Fall 2012

Principles of Computer Systems

SERVICE

Area chair for NAACL '21, EMNLP '20, '21
Reviewer for ACL '20, EMNLP '19, ACL '19, NAACL '19, EMNLP '17, CONLL '17, NIPS '16, EMNLP '16, NAACL '16, AAAI '15
Standing reviewer for ARR, TACL
Reviewer for INFORMS Journal on Computing

LANGUAGES

Java, Python, Scala, MATLAB

TOOLKITS

Pytorch, Streamlit.io, TensorFlow, Theano, Caffe

MEDIA

Quanta magazine, VentureBeat, SiliconAngle, TechCrunch, Datanami, ZDNet, The Batch