Mehar Bhatia

Ph.D. Student in Computer Science | McGill University & MILA

mehar.bhatia@mila.quebec | meharbhatia.github.io/ | github.com/meharbhatia | bhatia_mehar | Google Scholar

Research Focus

My research centers on **socio-technical alignment** in AI, aiming to develop language and multimodal models that are **safe, inclusive, and culturally sensitive**. I focus on incorporating diverse values and **pluralistic preferences**, addressing the limitations of existing alignment techniques that often overlook cultural nuances and **value pluralism**. To ensure these models align with the needs of a global user base, my research also involves designing **evaluation frameworks** that assess real-world impact, prioritize transparency, and promote robustness in alignment practices.

Education

Sept 2024 - Present | **McGill University | MILA Quebec AI Institute** — **Montreal, Canada**
Doctor of Philosophy (Ph.D.) in Computer Science
Advisors: Prof. Siva Reddy & Prof. Vered Shwartz (UBC)

Sept 2022 - Aug 2024 | **University of British Columbia | Vector Institute** — **Vancouver, Canada**
Master of Science (MS) in Computer Science, *GPA: 4.0*
Advisor: Prof. Vered Shwartz

July 2016 - June 2020 | **Shiv Nadar University** — **Greater Noida, India**
B.Tech in Computer Science and Engineering, Minor in Mathematics, *CGPA: 8.15/10*

Experience

Sept 2024 - Present | **McGill NLP Lab [🌐] & MILA Quebec AI Institute [🌐]** — **Montreal, Canada**
Graduate Student Researcher with Prof. Siva Reddy
- Currently assessing how varying components in LLM alignment impact model behaviour on downstream tasks, focusing on performance differences across cultural and demographic groups.

Jan 2023 - Aug 2024 | **UBC NLP Lab [🌐] and Vector Institute of AI [🌐]** — **Vancouver, Canada**
Graduate Student Researcher with Prof. Vered Shwartz
- Developed **Geo-Diverse COMET** model for generating culturally relevant commonsense inferences.
- Created a multicultural image-text **CulturalSnap** dataset (∼75K images) for analyzing cultural understanding in vision-language (VL) models using contrastive learning and designed **two benchmarks** for cultural retrieval and visual grounding, revealing gaps in VL models' cultural awareness.
- In collaboration with UWashington & AI2, developed large-scale **CulturalBench** to evaluate and track LLMs' cultural knowledge and performance across varied cultural contexts.

July 2021 - July 2022 | **NeuralSpace [🌐]** — **Remote / London, UK**
Applied Research Scientist | Mentors: Felix Laumann (CEO) and Ayushman Dash (CTO)
- Worked on understanding the disparity between research and deployment for low-resource languages and developing technologies to mitigate this gap (*'Language Technologies for All'*).
- Incorporated pipelines for joint multiple intent detection and slot filling using contrastive learning. Benchmarked accuracy improvement of **30%** when compared to HF models for Indic languages.

Aug 2021 - Nov 2021 | **Deep Cognition and Language Research (DeCLaRe) Lab | SUTD [🌐]** — **Remote / Singapore**
Research Associate | Advisors: Prof. Soujanya Poria and Deepanway Ghoshal (Ph.D.)
- Worked on unsupervised causal commonsense extraction from short story dataset, GLUCOSE using GNNs.
- *Commonsense-aware NLP* *Graph Neural Networks* *Dialogue Systems* *NLG*

June 2020 - June 2021 | **IIIT Delhi | Multimodal Digital Media Analysis Lab [🌐]** — **New Delhi, India**
Research Assistant | Advisor: Prof. Rajiv Ratn Shah
- Devised approaches for attributing the prediction of neural AES models to its input features.
- Responsible for developing ASR for Japanese-accented speech using self-supervised methods. Deployed and used by **300,000** middle-school students.
- Identified cognitive theories on bilingual language acquisition using visual lip-reading models.
- *Interpretability* *Speech Processing* *Vision*

Jan 2020 | Research Intern | Advisors: Prof. Rajiv Ratn Shah and Prof. Junyi Jessi Li
- Proposed a model-agnostic adversarial suite to evaluate the robustness of Automatic Essay Scoring (AES) systems, and presented associated metrics to test natural language understanding capabilities.
- Implemented a pipeline to identify bias, disparate error rates with respect to different groups for AES.
- *Question Answering* *Adversarial Robustness* *NLU* *Fairness*

June 2019 - Aug 2019 | **IIIT Delhi | Multimodal Digital Media Analysis Lab** [🌐] **New Delhi, India**
Research Intern | Advisors: Prof. Rajiv Ratn Shah and Dr. Debanjan Mahata
- Proposed a novel solution for Automatic Knowledge Graph Construction from unstructured text across multiple domains. **Awarded Honorable Mention Prize** and **Travel Grant** at ICDM 2019 in Beijing, China.
- *Knowledge Graphs* *Event Coreference Resolution*

May 2019 - July 2019 | **Languages Technologies Research Centre | NLP and MT Lab, IIIT-H** [🌐] **Hyderabad, India**
Summer Research Intern | Advisors: Prof. Dipti Misra Sharma and Prof. Manish Shrivastava
- Identified linguistic factors crucial for the performance of Statistical and Neural Machine Translations and implemented seminal papers in Neural MT for English – Hindi translations.
- *Neural Machine Translation* *Indic Languages*

May 2018 - July 2018 | **Software Engineering Research Centre | IIIT-H** [🌐] **Hyderabad, India**
Summer Research Intern | Advisor: Prof. Y Raghu Reddy
- Surveyed, implemented and analyzed the performance of state-of-the-art algorithms and developed an end-to-end semi-automated ontology enrichment pipeline using a sequential deep learning model.
- *Ontologies* *NLP* *Machine Learning*

Publications

S=In Submission, C=Conference, W=Workshop, P=Poster/Demo

[C.5] **From Local Concepts to Universals: Evaluating the Multicultural Understanding of Vision-Language Models** [🔗]
Mehar Bhatia, Sahithya Ravi, Aditya Chinchure, Eunjeong Hwang, Vered Shwartz
The 2024 Conference on Empirical Methods in Natural Language Processing [**EMNLP'24, Main**]

[S.1] **CulturalTeaming: AI-Assisted Interactive Red-Teaming for Challenging LLMs' (Lack of) Multicultural Knowledge** [🔗]
Yu Ying Chiu, Liwei Jiang, Maria Antoniak, Chan Young Park, Shuyue Stella Li, Mehar Bhatia, Sahithya Ravi, Yulia Tsvetkov, Vered Shwartz, Yejin Choi
[**In Submission**]

[C.4] **GD-COMET: A Geo-Diverse Commonsense Inference Model** [🔗]
Mehar Bhatia, Vered Shwartz
The 2023 Conference on Empirical Methods in Natural Language Processing [**EMNLP'23, Main**]

[C.3] **One To Rule Them All: Towards Joint Indic Language Hate Speech Detection** [🔗]
Mehar Bhatia, Tenzin Singhay Bhotia, Akshat Agarwal, Kumar Shridhar, Felix Laumann, Ayushman Dash
Forum for Information Retrieval Evaluation [**FIRE'21**]

[W.2] **Neural Automatic Scoring Systems are both Overstable and Oversensitive: Feature Interpretation** [🔗]
Mehar Bhatia, Yaman Kumar Singla, Rajiv Ratn Shah
Women in Machine Learning Workshop at 2020 Conference on Neural Information Processing Systems [**WiML@NeurIPS'20**]

[C.2] **Evaluation Toolkit For Robustness Testing Of Automatic Essay Scoring Systems** [🔗]
Anubha Kabra*, Mehar Bhatia*, Yaman Kumar Singla*, Junyi Jessy Li, Di Jin, Rajiv Ratn Shah (* = Equal Contribution)
ACM International Conference on Data Science & Management of Data [**CODS-COMAD'21**]

[C.1] **A survey on Ontology Enrichment Text** [🔗]
Vivek Iyer, Lalit Mohan, Mehar Bhatia, Y Raghu Reddy
16th International Conference on Natural Language Processing, Hyderabad, India [**ICON'19**]

[W.1] **OntoEnricher: A novel iterative pipeline to enrich seed ontology from text corpus**
Vivek Iyer, Mehar Bhatia, Lalit Mohan, Y Raghu Reddy
34th ACM Symposium on Applied Computing (SAC), Cyprus, Europe [**SAC'18**]

Honours and Awards

Awarded BPOC Graduate Excellence Award (amount CAD$1650) University of British Columbia, Vancouver, April 2024
Awarded Vector Research Grant (amount CAD$4000) Vector Institute for Artificial Intelligence, Toronto, 2024
Selected for 2024 CRA-WP Grad Cohort for Women (Funded) Minneapolis, Minnesota USA, April 2024
Awarded Cohere For AI Research Grant (amount USD$500 credits) Cohere AI Toronto, Sept 2023
Granted International Tuition Award (amount CAD$3200) University of British Columbia, Vancouver Sept 2023
Awarded Vector Research Grant (amount CAD$4000) Vector Institute for Artificial Intelligence, Toronto, 2023
Granted International Tuition Award (amount CAD$3200) University of British Columbia, Vancouver Sept 2022
Awarded for Top 5 Best BTech Thesis Projects in Computer Science Department Shiv Nadar University, 2020
Travel Grant | ICDM Conference 2019 To present my work at ICDM Conference, Beijing, China
Complete Tuition Fee Wavier for Undergraduate Program Shiv Nadar University (2016-2020)

Media

La Presse, Canada (April 2024) Imagine the future of AI on the beach [Article: 🌐]
Global News, Canada (Jan 2024) Solving Diversity Issues in Artificial Intelligence Models [Live TV Interview: 🌐]
UBC News (Jan 2024) ChatGPT has read almost the whole internet. That hasn't solved its diversity issues [Article: 🌐]
Radio-Canada (Jan 2024) Même après avoir lu presque tout le web, l'intelligence artificielle a des préjugés [Article: 🌐]
L'Obs (Jan 2024) ChatGPT, une intelligence artificielle pleine de biais (mais ça se soigne) [Article: 🌐]
The Conversation Canada (Feb 2024) AI needs to be trained on culturally diverse datasets to avoid bias) [Article: 🌐]

Technical Skills

Programming Languages Python Java R C JavaScript SQL HTML/CSS
Technologies PyTorch Tensorflow Keras NLTK MySQL MongoDB LaTex Git

Talks

UBC CS MSc Presentation (Aug 2024) Exploring Cultural Competence in Language and Multimodal Models [Slides: 🌐]
Guest Lecturer at UBC's CPSC 532V (Grad) (Feb 2024) Exploring Cross-Cultural Phenomena in NLP [Slides: 🌐]

Teaching

Teaching Assistant (April-May 2023, April 2024) Prompt Engineering Labs conducted by Vector Institute, Toronto with 120 participants
Graduate Teaching Assistant (Sept-Dec 2022) Course UBC CPSC 436N - NLP by Prof. Vered Shwartz [🌐]
Advisor at Linguistics Justice League Preserving low-resource languages and support those language learners. [🌐]

Conferences

Attended EMNLP'24 EMNLP'23 ACL'23 NAACL'21 EMNLP'21
Summer School Participated in CIFAR Deep Learning & Reinforcement Learning (DLRL) Summer School 2023
Invitational Workshop Bellairs Invitational Workshop on Contemporary, Foreseeable and Catastrophic Risks of Large Language Models held in Barbados, April 2024