

# Megh Vipul Thakkar

Linkedin: [megh-thakkar](#)

Github: [Megh-Thakkar](#)

Email: [megh.1211@gmail.com](mailto:megh.1211@gmail.com) Phone: +91-9829799877

---

- 2023 - Present**     **MILA - Quebec AI Institute and Université de Montréal, Montreal, Canada.**  
Master's of Science in Computer Science with focus on Artificial Intelligence.
- 2016 - 2020**     **Birla Institute of Technology and Science (BITS), Pilani, Pilani Campus, India.**  
Bachelor of Engineering in Computer Science. **CGPA:** 9.23/10.

## SELECTED PUBLICATIONS ([Google Scholar](#))

- EMNLP '23**     Self-Influence Guided Data Reweighting for Language Model Pre-training. [*Accepted*]  
[Megh Thakkar](#), Sriram Ganapathy, Shikhar Vashishth, Tolga Bolukbasi, Sarath Chandar, Partha Talukdar.
- ACL '23**     Towards Robust Low-Resource Fine-Tuning with Multi-View Compressed Representations. [[paper](#)]  
Linlin Liu\*, Xingxuan Li\*, [Megh Thakkar](#), Xin Li, Shafiq Joty, Luo Si, Lidong Bing.
- ACL '23**     Randomized Smoothing with Masked Inference for Adversarially Robust Text Classifications. [[paper](#)]  
Han Cheol Moon, Shafiq Joty, Ruochen Zhao, [Megh Thakkar](#), Xu Chi.
- EMNLP '22**     Tweet Based Reach Aware Temporal Attention Network for NFT Valuation. [[paper](#)]  
Ramit Sawhney\*, [Megh Thakkar](#)\*, Ritesh Soun\*, et al.
- ACL '22**     Chart-to-Text: A Large-Scale Benchmark for Chart Summarization. [[paper](#)]  
Shankar Kantharaj\*, Rixie Tiffany Leong\*, Xiang Lin\*, Ahmed Masry\*, [Megh Thakkar](#)\*, Enamul Hoque, Shafiq Joty.
- ACL '22**     DMix: Adaptive Distance-aware Interpolative Mixup. [[paper](#)]  
Ramit Sawhney\*, [Megh Thakkar](#)\*, Shrey Pandit\*, Ritesh Soun, Di Jin, Diyi Yang, Lucie Flek.
- EMNLP '21**     HYPMIX: Hyperbolic Interpolative Data Augmentation. [[paper](#)]  
Ramit Sawhney\*, [Megh Thakkar](#)\*, Shivam Agarwal, Di Jin, Diyi Yang, Lucie Flek.

## EXPERIENCE

- Aug '23 - Present**     **Expert LLMs in Multi-agent Systems.**  
Research Assistant: [MILA - Quebec AI Institute](#), Montreal, Canada  
*Advisor:* [Dr. Sarath Chandar](#)
- Benchmarking the capabilities of LLMs when trained on different preferential datasets on diverse evaluation tasks.
  - Exploring methods to effectively combine multiple domain-expert LLMs to solve complex downstream tasks.
- Jul '22 - Jul '23**     **Self-influence Guided Pre-training of Language Models.**  
Pre-doctoral Researcher: [Google Research](#), Bangalore, India  
*Advisor:* [Dr. Partha Talukdar](#)
- Formulated a novel online sample reweighting scheme for language model pre-training using self-influence scores.
  - Scaled the strategy to 1B model pre-training, obtaining significant improvements in low resource environments.
- Feb '22 - Jul '22**     **Robust Training of Language Models.**  
Research Assistant: [Alibaba-NTU Singapore Joint Research Institute](#), Singapore  
*Advisors:* [Dr. Shafiq Joty](#), [Dr. Lidong Bing](#) (Alibaba DAMO, Singapore)
- Explored robust fine-tuning methods for language models by using autoencoder adapters within transformers.
  - Worked on adversarially robust training of language models using smoothed perturbations for text classification.
- Jun '21 - Feb '22**     **Neural Voice Cloning and Tabular Question Answering.**  
Machine Learning Engineer: [Avaamo](#), Bangalore, India
- Led the work on voice cloning from low-resource recordings using state-of-the-art synthesizers and vocoders.
  - Worked on semantic parsing and question answering over tabular data.
- Dec '20 - Dec '21**     **Chart-based Text Generation and Graph-to-text Open Domain Question Answering.**  
Research Assistant: [NTU-NLP Group](#), Nanyang Technological University (NTU), Singapore  
*Advisors:* [Dr. Shafiq Joty](#), [Dr. Enamul Hoque](#) (York University, Canada)
- Designed a multi-modal network using image encoders and text decoders adapted for summarization of charts.
  - Explored graph network architectures for leveraging knowledge bases in conjunction with text for open domain QA.

**Dec '20 - Aug '21 Online Text Stream Modeling and Interpolative Data Augmentation.**  
Research Assistant: [Multimodal Digital Media Analysis Lab \(MIDAS\)](#), IIT Delhi, India  
*Advisors:* [Dr. Di Jin](#) (Amazon Alexa AI, Sunnyvale), [Dr. Rajiv Ratn Shah](#)

- Introduced a novel model, dataset and modality agnostic interpolative regularization method using the Riemannian manifold for effective representation of complex latent embeddings for text, speech, and vision.
- Extended the work to a distance-aware strategy for more efficient interpolations as compared to vanilla mixup.

**Aug '19 - Dec '19 Speech Emotion Detection.**  
Research Intern: [Speech and Language Laboratory](#), Nanyang Technological University (NTU), Singapore  
*Advisors:* [Dr. Yashvardhan Sharma](#) (BITS, Pilani), [Dr. Chng Eng-Siong](#)

- Built a resource efficient CNN based model for speech emotion detection and deployed it to production.

**May '19 - Jul '19 Language Model for Query Auto-completion.**  
DAAD-WISE Scholar: [Language Technology Group \(LT\)](#), University of Hamburg, Germany  
*Advisor:* [Dr. Chris Biemann](#)

- Developed a bi-LSTM based character-level language model coupled with word embeddings such as GloVe, ELMo, and Flair, as well as a combination of FastText and sent2vec.

**May '18 - Aug '18 BioJS Webapp Backend.**  
Summer Intern: Google Summer of Code, 2018 ([Project Link](#)), India-Australia-United Kingdom

- Led the development of a new website for BioJS, a community-based project compiling JavaScript widgets and modular components to visualize and process biological data using web technologies.

## SKILLS

**Languages:** Python, C, Java, C++, SQL, Scala

**Libraries and Frameworks:** PyTorch, pytorch-geometric, JAX, Keras, sklearn, NLTK, networkx, Django, REST

## ACHIEVEMENTS

**2023** Awarded the MILA scholarship of CAD 27,000/year to pursue graduate studies.  
**2023** Awarded the UdeM-Bourse d'exemption grant of CAD 10,046 to pursue graduate studies.  
**2019** Awarded the DAAD-WISE scholarship to pursue summer research in Germany and the mitacs Globalink scholarship to pursue research in Canada.  
**2016-2018** Awarded the Institute Merit Scholarship given to the top 3 percentile of students across all the departments.  
**2016** Secured All India Rank 105 out of about 1.2 million candidates in the JEE Mains Entrance Examination.