

# Simran Khanuja

## PhD in Computer Science | Carnegie Mellon University

🌐 [simran-khanuja.github.io](https://simran-khanuja.github.io) @ [khanuja.simran7@gmail.com](mailto:khanuja.simran7@gmail.com) 🌐 [github.com/simran-khanuja](https://github.com/simran-khanuja)  
🎓 Google Scholar 🐦 [twitter.com/simi\\_97k](https://twitter.com/simi_97k)

## Education

-	<b>Carnegie Mellon University</b>	Pittsburgh, Pennsylvania
Aug 2022	PhD in Computer Science (NLP), QPA: 4.16/4.00 Advisor: Prof. Graham Neubig	
Aug 2020	<b>Birla Institute of Technology and Science, Pilani</b>	Goa, India
Aug 2015	B.E. (Honors) Computer Science; M.Sc.(Hons.) Economics, CGPA: 8.81/10.00	

## Experience

Aug 2022	<b>Google Research   Natural Language Understanding</b> [🌐]	Bangalore, India
Aug 2020	Pre-Doctoral Researcher   Advisor: Dr. Partha Talukdar Collaborators: Dr. Sebastian Ruder, Dr. Alexis Conneau Projects: Multilingual Representations for Indian Languages (MuRIL), Merging Pre-trained Language Models using Distillation (MergeDistill), Multilingual Neural Semantic Parsing for Google Assistant (mNSP), Cross-Lingual Text Speech Embeddings (XTeSE)	
Jul 2020	<b>Microsoft Research   Project Mélange</b> [🌐]	Bangalore, India
Jul 2019	Research Intern (Bachelor Thesis)   Advisors: Dr. Sunayana Sitaram, Dr. Monojit Choudhury Projects: General Language Understanding and Evaluation for Code-Switching (GLUECoS), Code-Mixed Natural Language Inference, Adapting TULR for Code-Mixing	

## Selected Publications

- [14] **Pangea: A Fully Open Multilingual Multimodal LLM for 39 Languages** [web | PDF | code | models]  
Xiang Yue\*, Yueqi Song\*, Akari Asai, Seungone Kim, Jean de Dieu Nyandwi, [Simran Khanuja](#), Anjali Kantharuban, Lintang Sutawika, Sathyanarayanan Ramamoorthy, Graham Neubig  
[Under Review]
- [13] **NaturalBench: Evaluating Vision-Language Models on Natural Adversarial Samples** [PDF]  
Baiqi Li, Zhiqiu Lin, Wenxuan Peng, Jean de Dieu Nyandwi, Daniel Jiang, Zixian Ma, [Simran Khanuja](#), Ranjay Krishna, Graham Neubig, Deva Ramanan  
Conference on Neural Information Processing Systems [NeurIPS '24]
- [12] **An image speaks a thousand words but can everyone listen? On image transcreation for cultural relevance** [web | PDF | code | talk]  
[Simran Khanuja](#), Sathyanarayanan Ramamoorthy, Yueqi Song, Graham Neubig  
Conference on Empirical Methods in Natural Language Processing [EMNLP '24]
- [11] **DeMuX: Data-efficient Multilingual Learning** [PDF | code]  
[Simran Khanuja](#), Srinivas Gowriraj, Lucio Dery, Graham Neubig  
Conference of the North American Chapter of the Association for Computational Linguistics [NAACL '24]
- [10] **GlobalBench: A Benchmark for Global Progress in Natural Language Processing** [PDF]  
Yueqi Song, Catherine Cui, [Simran Khanuja](#), Pengfei Liu, ..., Graham Neubig  
Conference on Empirical Methods in Natural Language Processing [EMNLP '23]
- [9] **Multi-lingual and Multi-cultural Figurative Language Understanding** [PDF | code]  
Anubha Kabra\*, Emmy Liu\*, [Simran Khanuja](#)\*, Alham Fikri Aji, ..., Graham Neubig  
Annual Conference of the Association for Computational Linguistics [ACL '23 Findings]
- [8] **Evaluating Inclusivity, Equity, and Accessibility of NLP Technology: A Case Study for Indian Languages** [PDF]  
[Simran Khanuja](#)\*, Sebastian Ruder\*, Partha Talukdar  
European Chapter of the Association for Computational Linguistics [C3NLP | SIGTYP | EACL '23]

- [7] **[Best Paper] FLEURS: Few-Shot Learning Evaluation of Universal Representations of Speech** [PDF]  
Alexis Conneau\*, Min Ma\*, [Simran Khanuja\\*](#), Yu Zhang, ..., Ankur Bapna  
*IEEE Spoken Language Technology Workshop* [SLT 2022]
- [6] **MuRIL: Multilingual Representations for Indian Languages** [PDF | large | base]  
[Simran Khanuja](#), Diksha Bansal, Sarvesh Mehtani, Savya Khosla, ..., Partha Talukdar  
*Media: Economic Times | Indian Express | Google AI Blog* [Technical Report: Mar '21]
- [5] **MergeDistill: Merging Pre-trained Language Models using Distillation** [PDF]  
[Simran Khanuja](#), Melvin Johnson, Partha Talukdar  
*Annual Conference of the Association for Computational Linguistics (Virtual)* [Findings of ACL'21]
- [4] **GLUECoS: An Evaluation Benchmark for Code-Switched NLP** [PDF | code | website]  
[Simran Khanuja](#), Sandipan Dandapat, Anirudh Srinivasan, Sunayana Sitaram, Monojit Choudhury  
*Annual Conference of the Association for Computational Linguistics (Virtual)* [ACL'20]

## Other Publications

---

- [3] **A New Dataset for Natural Language Inference from Code-mixed Conversations** [PDF | data]  
[Simran Khanuja](#), Sandipan Dandapat, Sunayana Sitaram, Monojit Choudhury  
*International Conference on Language Resources and Evaluation* [CALCS, LREC'20]
- [2] **Unsung Challenges of Building and Deploying Language Technologies for LRL Communities** [PDF]  
Pratik Joshi, Christain Barnes, Sebastin Santy, [Simran Khanuja](#), Sanket Shah, Anirudh Srinivasan, Satwik Bhat-  
tamishra, Sunayana Sitaram, Monojit Choudhury, Kalika Bali  
*International Conference on Natural Language Processing, Hyderabad, India* [ICON'19]
- [1] **Dependency Parser for Bengali-English Code-Mixed Data enhanced with a Synthetic Treebank** [PDF | code]  
Urmi Ghosh, [Simran Khanuja](#), Dipti Misra Sharma  
*International Workshop on Treebanks and Linguistic Theories* [TLT, SyntaxFest'19]

## Academic Service

---

<b>Lecturer</b>	CMU-11737 (Multilingual NLP) on Image-Text Modeling for Multilingual NLP ( <a href="#">slides</a> )
<b>Reviewer</b>	NAACL'24, EMNLP '23, ACL'23, MRL@EMNLP'21, TALLIP'20
<b>Sub-Reviewer</b>	EMNLP'21, EMNLP'20, ACL'20

## Volunteer Service

---

<b>AI For Education</b>	In engagement with <a href="#">Inspiring Teachers</a> to use my research to create educational content for children in Ghana. Also in engagement with <a href="#">Linguistics Justice League</a> to help localize stories for children for their multilingual story app in 108 languages, using my research.
-------------------------	--

## Talks and Interviews

---

<b>“Invited Keynote: Image Transcreation for Cultural Relevance”</b>	
> <a href="#">Slides</a>   <a href="#">Video</a>	
> <a href="#">AmericasNLP</a> , NAACL '24	July 2024
<b>“Image Transcreation for Cultural Relevance”</b>	
> <a href="#">University of Edinburgh</a>	March 2024
<b>“Multimodality for Multilingual NLP: The need, an application, and open questions”</b>	
> <a href="#">Slides</a>	
> <a href="#">Google Research</a>   <a href="#">Microsoft Research</a>   <a href="#">Indian Institute of Science (IISc)</a>   <a href="#">Microsoft IDC</a>	January 2024
<b>“Decode with Google”</b>	
> <a href="#">Speaker List</a>   <a href="#">Talk (2:28:00 onwards)</a> (registration required) <a href="#">🔗</a>	August 2022
<b>“An Introduction to (Modern) TensorFlow”</b>	
> <a href="#">CVIT Summer School</a> , IIT Hyderabad <a href="#">🔗</a>	August 2021 (Remote)

## “Journey into Research”

- › Rotaract Club, BITS Hyderabad [📍]
- › Google Research, India

January 2021 (Remote)  
December 2020 (Remote)

## “ICSE National Topper”

- › India Times | Times of India | Indian Express

May 2013 (Pune, India)

## Teaching and Leadership

---

**Multilingual NLP (Graduate Course), CMU** *Teaching Assistant* Fall '23

- › TA for 11737-Multilingual NLP at CMU, taught by Dr. Lei Li. Design assignments and guide course projects for teams.

**Student Research Symposium, CMU** *Co-Organizer* Aug '23

- › Co-organizing the student research symposium at CMU, a LTI-wide event for students to share their research.

**LTI LLM Hackathon, CMU** *Co-Organizer* Apr '23

- › Co-organizing the LTI LLM Hackathon at CMU!

**AI Undergrad Mentorship Program, CMU** *Mentor* Spring '23

- › Mentor to undergraduate students starting out with NLP research at CMU.

**Coffee Club Mentorship Program, Google** *Co-Lead* July '21 - July '22

- › Co-Lead of the Coffee Club program at Google India.
- › Enabling women employees to seek mentorship from senior executives across Google.

**An Introduction to (Modern) TensorFlow** *Co-Instructor* Aug '21

- › Conducted a hands-on Tensorflow tutorial session attended by 100+ members from Academia.

**BITS Alumni Mentorship Program** *Mentor* Aug '20 - Aug '21

- › Mentorship for undergraduate students looking to secure research theses opportunities. My mentee secured a year-long thesis at Microsoft Research, India!

**IKDD NLP Session** *Host* Aug '21

- › Hosted the NLP networking session at IKDD 2021 where Dr. Monojit Choudhury was our guest speaker!

**Fireside Chat with Jeff Dean** *Host* Sep '20

- › Hosted a Fireside Chat with Jeff Dean on his virtual Google India visit!

**Music Society, BITS Goa** *Core Member* Aug '17 - Aug '18

- › Managed events of the Music Society, BITS Goa as one of the five core members in the organizational team and gave performances as a vocalist.

## Student Mentorship

---

- › OpenNLP Labs Summer Cohort [Miscellaneous]
- › Arnav Yayavaram [BITS Pilani, Hyderabad]
- › Siddharth Yayavaram [BITS Pilani, Hyderabad]
- › Yueqi Kaylin Song [Undergraduate, CMU]
- › Sathyanarayanan Ramamoorthy [MIIS, CMU]
- › Srinivas Gowriraj [MIIS, CMU]
- › Helen Wang [Undergraduate, CMU]