Simran Khanuja

PhD in Computer Science | Carnegie Mellon University

simran-khanuja.github.io @ khanuja.simran7@gmail.com github.com/simran-khanuja
Google Scholar twitter.com/simi_97k

Research Summary

I work on multimodal AI and alignment, with research spanning evaluation, post-training, data, interpretability, and image generation. I'm most drawn to problems with the potential for real-world impact, where I can add value to society and people's lives.

Education

- Aug 2022	**Carnegie Mellon University** PhD in Computer Science (NLP), *QPA: 4.03/4.33* *Advisor: Prof. Graham Neubig*	**Pittsburgh, Pennsylvania**
Aug 2020 Aug 2015	**Birla Institute of Technology and Science, Pilani** B.E. (Honors) Computer Science; M.Sc.(Hons.) Economics, *CGPA: 8.81/10.00*	**Goa, India**

Experience

Aug 2025 May 2025	**Google DeepMind** [🌐] *Student Researcher, Personalization Team* | *Hosts: Dr. Lun Wang, Dr. John Lambert* Project: Steering LLMs for Culturally Localized Generation	**MTV, California**
Aug 2022 Aug 2020	**Google Research** [🌐] *Pre-Doctoral Researcher* | *Advisors: Dr. Partha Talukdar, Dr. Sebastian Ruder, Dr. Alexis Conneau* Projects: MuRIL, MergeDistill, Multilingual NSP for Assistant, FLEURS	**Bangalore, India**
Jul 2020 Jul 2019	**Microsoft Research** [🌐] *Research Intern (Bachelor Thesis)* | *Advisors: Dr. Sunayana Sitaram, Dr. Monojit Choudhury* Projects: Code-Mixed NLU and NLI (*GLUECoS*), Adapting TULR for Code-Mixing	**Bangalore, India**

Awards, Fellowships, Grants, Selected Talks

Heidelberg Laureate Forum (HLF) 2026 [🌐] Selected as a Young Researcher among 200 others worldwide

Jane Street Fellowship - Honorable Mention [🌐]

BITS 30 Under 30 - Research Leaders [🌐]

MIT EECS Rising Star 2025 [🌐]

Invited Keynote: CEGIS @ ICCV 2025 [🌐] Evaluating Cultural Diversity in Generative AI

Rising Star in AI, AI for Science Symposium 2025, University of Michigan [🌐]

Imminent Translated Research Grant Recipient, 2025 [🌐] For Human-AI Platform on Image Transcreation

Invited Keynote: DemoDiv @ CVPR 2025 [🌐] Data collection and Evaluation efforts for Image Transcreation

Invited Keynote: AmericasNLP @ NAACL 2024 [🌐] Introducing task and Evaluation efforts for Image Transcreation

Best Paper, EMNLP 2024 [🌐] Introducing the new task of Image Transcreation for localizing multimodal content

CMU Waibel Presidential Fellowship, 2024-25

Best Paper, SLT 2022 [🌐] Introducing the largest multilingual speech-text dataset, FLEURS

National ICSE Topper, 2013 Rank 1 in the country for board examinations

Selected Publications

[15] **T-DIG: Text-Conditioned Diversity for Image Generation** [PDF | website | code | demo]
Simran Khanuja, Rohit Gandikota, Gaurav Parmar, Graham Neubig
Under review **[Preprint '26]**

[14] **Steering LLMs for Culturally Localized Generation** [PDF]
Simran Khanuja, Hongbin Liu, Shujian Zhang, John Lambert, Mingqing Chen, Rajiv Mathews, Lun Wang
Conference on Language Modeling **[COLM '26]**

[13] **CAIRE: Cultural Attribution of Images by Retrieval-Augmented Evaluation** [PDF | code]
Arnav Yayavaram*, Siddharth Yayavaram*, Simran Khanuja*, Michael Saxon, Graham Neubig
European Chapter of the Association for Computational Linguistics (also at CEGIS@ICCV '25) **[EACL '26]**

[12] **Grounding Multilingual Multimodal LLMs With Cultural Knowledge** [PDF | code]
Jean De Dieu Nyandwi, Yueqi Song, Simran Khanuja, Graham Neubig
Conference on Empirical Methods in Natural Language Processing **[EMNLP '25]**

[11] **Towards Automatic Evaluation for Image Transcreation** [PDF]
Simran Khanuja*, Vivek Iyer*, Claire He, Graham Neubig
Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics **[NAACL '25]**

[10] **Pangea: A Fully Open Multilingual Multimodal LLM for 39 Languages** [web | PDF | code | models]
Xiang Yue*, Yueqi Song*, Akari Asai, ... Simran Khanuja, ... Graham Neubig
International Conference on Learning Representations **[ICLR '25]**

[9] **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]**

[8] **[Best Paper] 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]**

[7] **[Best Paper Runner-Up] HILITE: Human-in-the-loop Interactive Tool for Image Editing**
Arya Pasumarthi, Armaan Sharma, ..., Diyi Yang, Graham Neubig, Simran Khanuja
2024 IEEE International Conference on Big Data (Undergraduate Symposium) **[IEEE BigData '24]**

[6] **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]**

[5] **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]**

[4] **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]**

[3] **[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]**

[2] **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]**

[1] **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]**

Academic Service

Lecturer	CMU-11737 (Multilingual NLP) on Image-Text Modeling for Multilingual NLP (slides)
Workshop Co-organizer	GenAI4World at COLM '26; Visual Persuasion at ECCV '26; MAPS at CVPR '26; Multilingual, Multicultural Evaluation at EACL '25
Reviewer	EMNLP'25, NAACL'25, ACL'25, EMNLP '23, ACL'23, MRL@EMNLP'21, TALLIP'20
Sub-Reviewer	EMNLP'21, EMNLP'20, ACL'20

Volunteer Service

AI For Education In engagement with Inspiring Teachers to use my research to create educational content for children in Ghana. Also in engagement with Linguistics Justice League to help localize stories for children for their multilingual story app in 108 languages, using my research.

Talks and Interviews

CEGIS @ ICCV 2025 Keynote: *Cultural Inclusivity in Multimodal AI* [slides] Oct 2025
AI for Science Symposium, U. Michigan Rising Star in AI: *Research Overview* [slides | video] Oct 2025
DemoDiv @ CVPR 2025 Keynote: *Image Transcreation for Cultural Diversity* [slides | video] Jun 2025
Google Invited talk: *Image Transcreation for Cultural Relevance* [slides] Jan 2025
Microsoft Invited talk: *Image Transcreation for Cultural Relevance* [slides] Jan 2025
IISc Invited talk: *Image Transcreation for Cultural Relevance* [slides] Dec 2024
U. Edinburgh Invited talk: *Image Transcreation for Cultural Relevance* [slides] Sep 2024
Pinterest Invited talk: *Image Transcreation for Cultural Relevance* [slides] Sep 2024
AmericasNLP @ NAACL 2024 Keynote: *Cultural NLP & Image Transcreation* [slides | video] Jun 2024
Google Research India *Decode with Google* [video] Aug 2022
CVIT Summer School, IIIT Hyderabad Tutorial: *An Introduction to (Modern) TensorFlow* [slides] Aug 2021
Rotaract Club, BITS Hyderabad *Journey into Research* [blog] Jan 2021
ICSE National Topper [India Times | Times of India | Indian Express] May 2013

Teaching and Leadership

Multimodal Machine Learning (Graduate Course), CMU *Teaching Assistant* Fall '25
› TA for 11-777 Multimodal Machine Learning at CMU. Grade assignments and mentor course projects.

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.

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.

Fireside Chat with Jeff Dean *Host* Sep '20
› Hosted a Fireside Chat with Jeff Dean on his virtual Google India visit!

Student Mentorship

› Parin Pengpun [Undergraduate, CMU]
› 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]