

Simran KHANUJA

[in linkedin.com/in/simran-khanuja-6b80b6144](https://www.linkedin.com/in/simran-khanuja-6b80b6144) @ khanuja.simran7@gmail.com
📍 Bangalore, India 📄 github.com/SimranKhanuja 🌐 simran-khanuja.github.io/

EDUCATION

Aug' 2020 | Birla Institute of Technology and Science, Pilani, GOA, India
Aug' 2015 | B.E. (Honors) Computer Science; M.Sc.(Hons.) Economics, CGPA : 8.81/10.00

PUBLICATIONS

- **MuRIL : Multilingual Representations for Indian Languages**
Simran Khanuja, Diksha Bansal, Sarvesh Mehtani, Savya Khosla, Atreyee Dey, Balaji Gopalan, Dilip Kumar Margam, Pooja Aggarwal, Rajiv Teja Nagipogu, Shachi Dave, Shruti Gupta, Subhash Chandra Bose Gali, Vish Subramanian, Partha Talukdar
[pdf](#) | [tfhub](#) | [huggingface](#)
- **GLUECoS : An Evaluation Benchmark for Code-Switched NLP**
2020 Annual Conference of the Association for Computational Linguistics (ACL '20)
Simran Khanuja, Sandipan Dandapat, Anirudh Srinivasan, Sunayana Sitaram, Monojit Choudhury
[pdf](#) | [code](#) | [website](#)
- **A New Dataset for Natural Language Inference from Code-mixed Conversations**
CALCS Workshop : International Conference on Language Resources and Evaluation, 2020 (LREC '20)
Simran Khanuja, Sandipan Dandapat, Sunayana Sitaram, Monojit Choudhury
[pdf](#) | [data](#)
- **Unsung Challenges of Building and Deploying Language Technologies for Low Resource Language Communities**
2019 International Conference on Natural Language Processing (ICON'19)
Pratik Joshi, Christain Barnes, Sebastin Santy, [Simran Khanuja](#), Sanket Shah, Anirudh Srinivasan, Satwik Bhattamishra, Sunayana Sitaram, Monojit Choudhury and Kalika Bali
[pdf](#)
- **Dependency Parser for Bengali-English Code-Mixed Data enhanced with a Synthetic Treebank**
18th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2019)
Urmi Ghosh, Dipti Misra Sharma, and [Simran Khanuja](#)
[pdf](#) | [code](#)

EXPERIENCE

- | | |
|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Present | Google Research, BANGALORE, India |
| Aug' 2020 | <i>Pre-Doctoral Researcher</i> <i>Advisors : Dr. Partha Talukdar</i> <ul style="list-style-type: none">➤ Working on neural semantic parsing for the Google Assistant with a special focus on Indian languages.➤ Experimented with merging multiple pre-trained language models in a teacher-student knowledge distillation framework. [Under Review]➤ Built a multilingual model specifically focused on Indian languages named MuRIL, which is now open-sourced. (paper tfhub huggingface) |
| Jul' 2020 | Microsoft Research, BANGALORE, India |
| Jul' 2019 | <i>Research Intern</i> <i>Advisors : Dr. Sunayana Sitaram, Dr. Monojit Choudhury</i> <ul style="list-style-type: none">➤ Worked on experimenting with fine-tuning strategies for contextualized multilingual models using the XTREME benchmark as a test-bed.➤ Experimented with cross lingual word embeddings and multilingual generalized language models on a variety of downstream NLP tasks for code-mixed data. Eventually built a benchmark for the evaluation of models/methods that process code-mixed data, which is now open-sourced. (code website) [ACL '20]➤ Proposed and oversaw the creation of a new dataset for the task of conversation entailment in code-mixed data, which is now open-sourced. (data) [CALCS@LREC '20] |

Aug' 2018	MT-NLP Lab, IIIT HYDERABAD, India
May 2018	Summer Research Intern Advisors : Dr. Dipti Misra Sharma [Github Link] <ul style="list-style-type: none"> ➤ Worked on generating valid Hindi-English code-mixed data to improve the accuracy of a code-mixed language model. Similar methodology of generation implemented for Bengali-English. [TLT@ SyntaxFest '19] ➤ Mainly used rule-based approaches wherein we chunk parallel sentences consistently and perform a post interleaving of the chunks based on head matching. ➤ Tools worked with include the Stanford Parser, LTRC Hindi Parser and the GIZA++ word alignment tool.
Nov' 2018	BITS Pilani, K.K. BIRLA GOA CAMPUS, India
Aug 2018	Undergraduate Research Project Supervisor : Dr. Sreejith V. [Github Link] [Report Link] <ul style="list-style-type: none"> ➤ Implemented an Emotion Recognition System to recognize the emotional state of a patient. The final state is a weighted average of several parameters including facial expressions, speech signals and physiological signals including heart rate and breathing rate.
Jul' 2017	Liveweaver India Pvt. Ltd., PUNE, India
May 2017	Summer Intern NLP for Voice Recognition <ul style="list-style-type: none"> ➤ Performed basic NLP tasks including, morphological segmentation, coreference resolution, sentiment analysis and named entity recognition. Also worked with Elasticsearch for clustering, indexing paragraphs and calculating scores of similarity between the paragraphs.

SKILLS

Languages	Python, C++, Java, HTML
Frameworks	PyTorch, NLTK
Tools	Visual Studio, Git, GIZA++, Stanford Parser, Elasticsearch
Courses	Machine Learning, Information Retrieval, Neural Networks, Data Structures and Algorithms, Design and Analysis of Algorithms, Object Oriented Programming, Probability and Statistics, Applied Econometrics, Mathematics and Statistics



HONORS AND AWARDS

2013	Secured an All India (National) Rank 1 in the 10th board examination, ICSE
2016	Received the institute's merit scholarship for the first semester

POSITIONS OF RESPONSIBILITY

- Teaching Assistant for the course, "Financial Management".
- Teaching Assistant for the course, "Applied Econometrics".
- Core Member and Vocalist, Music Society BITS Goa.
- Marketing Manager as part of DOSM for college fests in 2016.