




Varun Gumma









SCAI Center Fellow | Microsoft Research

@ varun230999@gmail.com  github.com/VarunGumma  Google Scholar
 twitter.com/VarunGumma23  varungumma.github.io

Education

June 2023 Aug 2021	Indian Institute of Technology (IIT), Madras M.Tech. in Computer Science & Engineering CGPA: 9.62/10 Dept Rank: 3 Teaching Exp: Pattern Recognition & Machine Learning, Linear Algebra & Random Processes, Fundamentals of Deep Learning	Chennai, India
Jun 2021 Aug 2017	Birla Institute of Technology and Science (BITS), Pilani B.E. in Computer Science & Engineering, with Minor in Physics CGPA: 9.14/10 Teaching Exp: Introductory Physics, Computer Programming, Discrete mathematics, Data Structures & Algorithms, Foundations of Data Science, Machine Learning	Hyderabad, India

Experience

Present July 2023	Microsoft Research  SCAI Center Fellow Advisor: <i>Dr. Kalika Bali</i> Working on multi-dimensional evaluation of Large Language Models and building Machine Translation systems for low-resource languages.	Bangalore, India
July 2023 May 2023	Microsoft Research  Research Intern Advisor: <i>Dr. Sunayana Sitaram</i> Worked on multi-dimensional modular debiasing of Pre-trained Language Models (BERT/Roberta).	Bangalore, India
May 2023 Aug 2022	Indian Institute of Technology, Madras AI4Bharat  Graduate Research Advisors: <i>Dr. Raj Dabre, Prof. Mitesh Khapra</i> Working on understanding the effects of Knowledge Distillation on Multilingual Neural Machine Translation models, specifically for Indian languages.	Chennai, India
Jul 2022 Jun 2022	National University of Singapore  Visiting Research Scholar Advisor: <i>Dr. Aditya Karanam</i> Worked on developing Machine Learning and Deep Learning models to extract <i>suggestions</i> and identify <i>suggested-features</i> from noisy user comments using Name-Entity-Recognition.	Kent Ridge, Singapore
May 2021 Jan 2021	Birla Institute of Technology and Sciences, Pilani  Undergraduate Research Advisor: <i>Prof. N.L. Bhanu Murthy</i> Worked on developing a Deep Learning model to automatically generate comments for Java code snippets.	Hyderabad, India
May 2021 Jan 2021	Birla Institute of Technology and Sciences, Pilani  Undergraduate Research Advisor: <i>Prof. Barsha Mitra</i> Worked on analyzing the performance of various Machine Learning models on ABAC datasets for <i>Policy Generalization & Augmentation</i> .	Hyderabad, India
May 2020 Jan 2020	Birla Institute of Technology and Sciences, Pilani  Undergraduate Research Advisor: <i>Prof. N.L. Bhanu Murthy</i> Worked on fine-tuning BERT for Name-Entity-Recognition in Telugu.	Hyderabad, India
July 2019 May 2019	Defense Research & Development Organization-Research Centre Imarat  Student Intern Worked on building an Object Detection Model to identify cars in a given image.	Hyderabad, India

- [A.1] **Towards Inducing Document-Level Abilities in Standard Multilingual Neural Machine Translation Models**
[Varun Gumma](#)^{*}, Pranjal A. Chitale^{*}, Kalika Bali
Arxiv e-Print [Arxiv]
- [A.2] **HEALTH-PARIKSHA: Assessing RAG Models for Health Chatbots in Real-World Multilingual Settings**
[Varun Gumma](#), Ananditha Raghunath, Mohit Jain[†], Sunayana Sitaram[†]
Arxiv e-Print [Arxiv]
- [W.2] **Contamination Report for Multilingual Benchmarks**
 Sanchit Ahuja^{*}, [Varun Gumma](#)^{*}, Sunayana Sitaram
Evaluating Evaluations: Examining Best Practices for Measuring Broader Impacts of Generative AI [EvalEval '24]
- [C.8] **PARIKSHA: A Large-Scale Investigation of Human-LLM Evaluator Agreement on Multilingual and Multi-Cultural Data**
 Ishaan Watts, [Varun Gumma](#), Aditya Yadavalli, Vivek Seshadri, Manohar Swaminathan, Sunayana Sitaram
2024 Conference on Empirical Methods in Natural Language Processing [EMNLP '24]
- [C.7] **Akal Badi ya Bias: An Exploratory Study of Gender Bias in Hindi Language Technology**
 Rishav Hada, Safiya Husain, [Varun Gumma](#), Harshita Diddee, Aditya Yadavalli, Agrima Seth, Nidhi Kulkarni, Ujwal Gadiraju, Aditya Vashistha, Vivek Seshadri, Kalika Bali
7th Annual ACM Conference on Fairness, Accountability, and Transparency [FAccT '24, **Best Paper**]
- [C.6] **METAL: Towards Multilingual Meta-Evaluation**
 Rishav Hada^{*}, [Varun Gumma](#)^{*}, Mohamed Ahmed, Kalika Bali, Sunayana Sitaram
Findings of the Association for Computational Linguistics: NAACL 2024 [NAACL '24]
- [C.5] **MEGAVERSE: Benchmarking Large Language Models Across Languages, Modalities, Models and Tasks**
 Sanchit Ahuja, Divyanshu Aggarwal, [Varun Gumma](#), Ishaan Watts, Ashutosh Sathe, Millicent Ochieng, Rishav Hada, Prachi Jain, Mohamed Ahmed, Kalika Bali, Sunayana Sitaram
2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT) [NAACL '24]
- [W.1] **MunTTS: A Text-to-Speech System for Mundari**
[Varun Gumma](#), Rishav Hada, Aditya Yadavalli, Pamir Gogoi, Ishani Mondal, Vivek Seshadri, Kalika Bali
7th Workshop on Computational Methods for Endangered Languages [ComputEL'24]
- [C.4] **MAFIA: Multi-Adapter Fused Inclusive Language Models**
 Prachi Jain^{*}, Ashutosh Sathe^{*}, [Varun Gumma](#), Kabir Ahuja, Sunayana Sitaram
18th Conference of the European Chapter of the Association for Computational Linguistics [EACL'24]
- [C.3] **Are Large Language Model-based Evaluators the Solution to Scaling Up Multilingual Evaluation?**
 Rishav Hada, [Varun Gumma](#), Adrian de Wynter, Harshita Diddee, Mohamed Ahmed, Monojit Choudhury, Kalika Bali, Sunayana Sitaram
Findings of the Association for Computational Linguistics: EACL 2024 [EACL'24]
- [J.1] **IndicTrans2: Towards High-Quality and Accessible Machine Translation Models for all 22 Scheduled Indian Languages**
 Jay Gala^{*}, Pranjal A. Chitale^{*}, Raghavan AK, [Varun Gumma](#), Sumanth Doddapaneni, Aswanth Kumar, Janki Nawale, Anupama Sujatha, Ratish Puduppully, Vivek Raghavan, Pratyush Kumar, Mitesh M. Khapra, Raj Dabre, Anoop Kunchukuttan
Transactions of Machine Learning Research [TMLR'23]
- [C.2] **An Empirical Study of Leveraging Knowledge Distillation for Compressing Multilingual Neural Machine Translation Models**
[Varun Gumma](#), Raj Dabre, Pratyush Kumar
24th Conference of the European Association of Machine Translation [EAMT'23]
- [C.1] **PAMMELA: Policy Administration Methodology using Machine Learning**
[Varun Gumma](#), Barsha Mitra, Soumyadeep Dey, Pratik Shashikantbhai Patel^{*}, Sourabh Suman^{*}, Saptarshi Das, Jaideep Vaidya
19th International Conference on Security and Cryptography - SECRIPT [SECRIPT'22]

Relevant Coursework

- M.Tech.** Advanced Data Structures & Algorithms, Pattern Recognition & Machine Learning, Fundamentals of Deep Learning, Natural Language Processing, Reinforcement Learning, Linear Programming & Combinatorial Optimization
- B.E.** Software Engineering, Foundations of Data Science, Machine Learning, Deep Learning, Quantum Information and Computing

Skills

- Programming Languages:** Python, Java, C++, C, SQL, \LaTeX
- Libraries & Frameworks:** HuggingFace Transformers, Fairseq, PyTorch-Lightning, PyTorch, TensorFlow, Scikit-Learn, Weights & Biases, Unix, GIT

Academic Service

Peer Reviewer EMNLP'23, EAACL'24, NAACL'24, ICLR'25, ACL Rolling Review




Awards and Achievements

IIT Madras STAR TA, 2021-2023 For outstanding contributions as a Graduate Teaching Assistant

GATE CS&IT, 2021 Secured an All India Rank of 159 with a score of 816/1000

BITS Merit Scholarship, 2018 For meritorious academic performance in the year 2017-2018

References

- > Dr. Kalika Bali *Senior Principal Researcher, Microsoft Research, India* 
- > Dr. Sunayana Sitaram *Principal Researcher, Microsoft Research, India* 
- > Dr. Raj Dabre *Senior Researcher, NICT; Adjunct Faculty, IIT Madras* 
- > Prof. Mitesh Khapra *Associate Professor, IIT Madras, Chennai* 