Rachit Bansal

Doctorate Student, **Harvard University**

rachitbansal.github.io @ rachitbansal@g.harvard.edu Google Scholar

Education

Harvard University, *Cambridge* — 08/2024 – Present
Ph.D. in Computer Science (ongoing)
- Advised by Prof. David Alvarez-Melis and Prof. Martin Wattenberg.
- Harvard Kempner Institute Graduate Fellow.

Delhi Technological University, *India* — 08/2018 – 07/2022
B.Tech. in Electrical Engineering
- Research Excellence Award 2022.
- Bachelor's thesis at the Technion, Israel (02/2022 – 07/2022).

Experience

Google DeepMind, *India* — 07/2022 – 07/2024
Pre-doctoral Researcher with Partha Talukdar and Prateek Jain

Technion, *Israel* — 09/2021 – 07/2022
Research Intern (Bachelor's Thesis) with Yonatan Belinkov

Adobe Research, *India* — 01/2021 – 09/2021
Research Intern with Balaji Krishnamurthy

Google Summer of Code, *Remote | University of Oxford* — 05/2020 – 01/2021
Contributor at CDLI with Jacob Dahl

Publications

[1] **LLM Augmented LLMs: Expanding Capabilities through Composition**
Rachit Bansal, Bidisha Samanta, Siddharth Dalmia, Nitish Gupta, Shikhar Vashishth, Sriram Ganapathy, Abhishek Bapna, Prateek Jain, Partha Talukdar
International Conference on Learning Representations **[ICLR 2024]**

[2] **Linear Connectivity Reveals Generalization Strategies**
Jeevesh Juneja, Rachit Bansal, Kyunghyun Cho, João Sedoc, Naomi Saphra
International Conference on Learning Representations **[ICLR 2023]**

[3] **Measures of Information Reflect Memorization Patterns**
Rachit Bansal, Danish Pruthi, Yonatan Belinkov
Conference on Neural Information Processing Systems **[NeurIPS 2022]**

[4] **Evaluating Explanations: How much do explanations from the teacher aid students?**
Danish Pruthi, Rachit Bansal, Bhuvan Dhingra, Livio Baldini Soares, Michael Collins, Zachary C. Lipton, Graham Neubig, William W. Cohen
Transactions of the Association for Computational Linguistics
Presented at the Annual Conference for the Association of Computation Linguistics **[TACL 2022]**

[5] **CoSe-Co: Text Conditioned Generative CommonSense Contextualizer**
Rachit Bansal, Milan Aggarwal, Sumit Bhatia, Jivat Kaur, Balaji Krishnamurthy
North American Chapter of the Association for Computational Linguistics **[NAACL 2022]**

[6] **LM-CORE: Language Models with Contextually Relevant External Knowledge**
Jivat Kaur, Sumit Bhatia, Milan Aggarwal, Rachit Bansal, Balaji Krishnamurthy
North American Chapter of the Association for Computational Linguistics (Findings) **[NAACL 2022]**

[7] **How Low is Too Low? A Computational Perspective on Extremely Low-Resource Languages**
Rachit Bansal, Himanshu Choudhary, Ravneet Punia, Niko Schenk, Jacob L Dahl, Émilie Pagé-Perron
Student Research Workshop (SRW) at **ACL** **[ACL SRW 2021]**

[8] **Combining exogenous and endogenous signals with a co-attention network for early fake news detection**
Rachit Bansal, William Scott, Nidhi Sultan, Tanmoy Chakraborty
Pacific-Asia Conference on Knowledge Discovery and Data Mining **[PAKDD 2021]**

Featured Academic Projects and Collaborations

Augmenting New Knowledge in Language Models through Composition 07/2022 – Present
w/ Partha Talukdar, Prateek Jain, Nitish Gupta, Sid Dalmia Google Research

- Worked as a part of a massive moonshot effort to create inclusive and equitable language representations.
- Led a large collaboration to introduce composition of language models as a paradigm to augment new knowledge.
- Proposed CALM: Using knowledge-specific models to augment new capabilities in a frozen language model. [**ICLR'24**]
- Working with Google DeepMind and the Bard team to test CALM for serving custom models to users.

Relationship between Information Distribution and Model Behavior 01/2022 – 07/2022
w/ Yonatan Belinkov, Danish Pruthi Technion

- Evaluating generalization of neural models is difficult: Requires creation of labeled out-of-distribution sets.
- Employed information-theoretic metrics to study the information distribution across neurons as an intrinsic metric.
- For the first time, showed that such intrinsic metrics strongly correlate with generalization behaviors of a model.
- Demonstrated the usefulness of the study for model selection. [**NeurIPS'22**]

Mode Connectivity in Loss Surfaces for Text Models 10/2021 – 10/2022
w/ Naomi Saphra, João Sedoc, Kyunghyun Cho New York University

- Analyzed linear model connectivity for multiple fine-tuned models from the same pre-trained language model.
- For the first time, observed clusters of models that lie in separate basins within the loss surface.
- Further observed that models belonging in the same cluster show identical generalization behaviors. [**ICLR'23**]
- Future work has utilized insights from our work for weight averaging and mechanistic interpretability. [◎]

Teacher-Student Paradigm to Evaluate Model Explanations 09/2020 – 12/2021
w/ Danish Pruthi, Bhuwan Dhingra, Zachary Lipton, Graham Neubig Carnegie Mellon University

- A number of model explainability approaches exist but no means to quantitatively evaluate and measure progress.
- Established a student-teacher communication paradigm for automatic evaluation of explanations. [**TACL'22**]

Grounding Language Models in Factual and Commonsense Knowledge 01/2021 – 09/2021
w/ Milan Aggarwal, Sumit Bhatia, Balaji Krishnamurthy Adobe Research

- Developed a framework to augment language model inputs with factual and commonsense knowledge on the fly.
- Demonstrated that our generic and efficient framework outperform large task-tuned models. [**NAACL'22**]

Neural Machine Translation for Sumerian 05/2020 – 01/2021
w/ Jacob Dahl, Émilie Pagé-Perron, Niko Schenk University of Oxford

- Sumerian is the earliest written language in Mesopotamia and perhaps the world—dating back to 4th millennium BC.
- Led this open-source initiative with CDLI to adapt modern NMT for extremely low-resource languages [**SRW, ACL'21**].
- Built an end-to-end information extraction pipeline for Sumerian widely used by Sumerian assyriologists today. [○]

Teaching and Featured Positions

Google Summer of Code, Cuneiform Digital Library Initiative (CDLI). *Mentor* Summer 2022

Reinforcement Learning, Coding Blocks. *Student Instructor w/ Prateek Narang* 2020

- Recorded 10-hours worth of lectures and held a number of live webinars. Collaborated with course mentors to build project ideas, assignments, and quizzes.

Foundations of Machine Learning & Deep Learning, Coding Blocks. *Teaching Assistant w/ Prateek Narang* 2019

- Conducted classes and doubt sessions for a batch of 60 senior undergraduate students from all across the country. Built course quizzes and programming assignments in collaboration with other TAs.

Reviewer: ICLR'24, NeurIPS'23, EMNLP'23, ACL'23, NeurIPS'22
Volunteer: NeurIPS'21, EMNLP'21, ICML'21, NAACL'21, NeurIPS'20, EMNLP'20, ICML'20, ACL'20

Featured Coursework

- **Mathematics:** Advanced Linear Algebra (2nd Sem., DTU; *University Rank-1*); MIT RES-6-012: Introduction to Probability, MIT OCW; Abstract Algebra, Group Theory, and Linear Algebra, IIT-KGP (NPTEL); Numerical and Engineering Optimization Methods (3rd Sem., DTU); Swarm and Evolutionary Optimization (7th Sem., DTU)
- **Machine Learning:** IFT 6760A: Matrix and tensor factorization techniques for machine learning, University of Montreal; MIT 18-065: Matrix Methods in Signal Processing, and Machine Learning, MIT OCW; Probabilistic Graphical Models Specialization, Stanford University; Bayesian Methods for Machine Learning, National Research University of Russia
- **Natural Language Processing:** CS11-737: Multilingual NLP, CMU; CS11-747: Neural Networks for NLP, CMU; Natural Language Processing (6th Sem., DTU)