Satwik **Bhattamishra**

PhD student, University of Oxford

satwikb.com | satwik55@gmail.com | github.com/satwik77 | Google Scholar

Education

Present – Oct 2021 | **University of Oxford** — **Oxford, UK**
Ph.D. in Computer Science
Advisors: Prof. Phil Blunsom and Prof. Varun Kanade
Google DeepMind Scholarship

May 2019 – Aug 2014 | **Birla Institute of Technology and Science Pilani** — **Pilani, India**
B.E. (Hons.), Computer Science and Int. M.Sc. (Hons), Biological Science

Research Interests

Sequence modelling Architectures (e.g., Transformers, SSMs, etc.); Pretraining LLMs; AI Safety and Verification

Industry Experience

Mar 2025 – Dec 2024 | **Google** — **Sunnyvale, US**
Student Researcher | Cloud AI Research Team
Worked on scaling test-time compute and training for LLM agents. Explored various decoding and self-verification strategies to improve per-step action prediction accuracy of agents.

Oct 2024 – June 2024 | **Cohere** — **London, UK**
Intern of Technical Staff | Foundations team
Worked on pretraining LLMs with *hybrid* architectures involving SSMs and Transformers. Pretrained 7B models, optimized inference strategies, & integrated μP for hyperparameter transfer.

Oct 2023 – June 2023
Intern of Technical Staff | Foundations team
Worked on pretraining and evaluating LLMs with linear recurrent and state-space architectures. Pretrained 7B SSMs on 700B tokens, and conducted ablations to evaluate the efficacy of architectural components.

July 2021 – July 2019 | **Microsoft Research** — **Bangalore, India**
Research Fellow | Advisors: Dr. Navin Goyal, Dr. Monojit Choudhury
Worked on analyzing the computational capabilities of Transformers and LSTMs to model various classes of formal languages. Also worked on compositional generalization and math word problem solving.

Jan 2019
Research Intern | Advisors: Dr. Navin Goyal, Dr. Monojit Choudhury
Worked on Semantic Parsing problems, particularly on NL-to-SQL problem. Explored the usage of graph neural networks to encode sentence as well as the table schema to improve generalization performance.

Publications

A Formal Framework for Understanding Length Generalization in Transformers [paper]
X. Huang, A. Yang, Satwik Bhattamishra, Y. Sarrof, A. Krebs, H. Zhou, P. Nakkiran, M. Hahn
2025 International Conference on Learning Representations `[ICLR'25]`

Separations in the Representational Capabilities of Transformers and Recurrent Architectures [paper]
Satwik Bhattamishra, Michael Hahn, Phil Blunsom, Varun Kanade
2024 Conference on Neural Information Processing Systems `[NeurIPS'24]`

Understanding In-Context Learning in Transformers by Learning to Learn Discrete Functions [paper]
Satwik Bhattamishra, Arkil Patel, Phil Blunsom, Varun Kanade
2024 International Conference on Learning Representations **[Oral, Top 1.2%]** `[ICLR'24]`

DynaQuant: Compressing Deep Learning Training Checkpoints via Dynamic Quantization [paper]
A. Agrawal, S. Reddy, Satwik Bhattamishra, V. Prabhakara Sarath Nookala, V. Vashishth, K. Rong, A. Tumanov
2024 ACM Symposium on Cloud Computing `[SoCC'24]`

MAGNIFICo: Evaluating the In-Context Learning Ability of Large Language Models to Generalize to Novel Interpretations [paper]
Arkil Patel, Satwik Bhattamishra, Siva Reddy, Dzmitry Bahdanau
2023 Conference on Empirical Methods in Natural Language Processing **[Oral]** `[EMNLP'23]`

Simplicity Bias in Transformers and their Ability to Learn Sparse Boolean Functions [paper]
Satwik Bhattamishra, Arkil Patel, Varun Kanade, Phil Blunsom
2023 Annual Meeting of the Association for Computational Linguistics `[ACL'23]`

Revisiting the Compositional Generalization Abilities of Neural Sequence Models [paper]
Arkil Patel, Satwik Bhattamishra, Phil Blunsom, Navin Goyal
2022 Annual Meeting of the Association for Computational Linguistics `[ACL'22]`

Are NLP Models really able to Solve Simple Math Word Problems? [paper]
Arkil Patel, Satwik Bhattamishra, Navin Goyal
2021 Conference of North American Chapter of the Association for Computational Linguistics `[NAACL'21]`

On the Ability and Limitations of Transformers to Recognize Formal Languages [paper]
Satwik Bhattamishra, Kabir Ahuja, Navin Goyal
2020 Conference on Empirical Methods in Natural Language Processing `[EMNLP'20]`

On the Practical Ability of Recurrent Neural Networks to Recognize Hierarchical Languages [paper]
Satwik Bhattamishra, Kabir Ahuja, Navin Goyal
2020 International Conference on Computational Linguistics **[Best Short Paper Award]** `[COLING'20]`

On the Computational Power of Transformers and Its Implication in Sequence Modeling [paper]
Satwik Bhattamishra, Arkil Patel, Navin Goyal
2020 Conference on Computational Natural Language Learning `[CoNLL'20]`

Unsung Challenges of Building and Deploying Language Technologies for Low Resource Language Communities
P. Joshi, C. Barnes, S. Santy, S. Khanuja, S. Shah, A. Srinivasan, Satwik Bhattamishra, S. Sitaram, M. Choudhury, K. Bali
16th International Conference on Natural Language Processing `[ICON'19]`

Submodular Optimization-based Diverse Paraphrasing and its Effectiveness in Data Augmentation [paper]
Satwik Bhattamishra*, Ashutosh Kumar*, Manik Bhandari, Partha Talukdar (* = Equal Contribution)
2019 Conference of the North American Chapter of the Association for Computational Linguistics **[Oral]** `[NAACL'19]`

Preprints and Upcoming Works

Automata Learning and Identification of the Support of Language Models [paper]
Satwik Bhattamishra, Michael Hahn, Varun Kanade
Under Review

Benefits and Limitations of Communication in Multi-Agent Reasoning [paper]
Michael Rizvi-Martel, Satwik Bhattamishra, Neil Rathi, Guillaume Rabusseau, Michael Hahn
Under Review

The Transformer Cookbook [paper]
A. Yang, C. Watson, A. Xue, Satwik Bhattamishra, J. Llarena, W. Merrill, E. Ferreira, A. Svete, D. Chiang
Under Review

Selected Dev Projects

Student Developer **Google Summer of Code 2016** May'16 - Aug'16

Worked on the Phenopacket scraper project, which extracts information from texts scraped from life sciences websites, analyzes them, and generates a phenopacket based on the correct ontology references. [Project Page]

LibNMF **[Github]**

An easy-to-use Python library with implementations of a set of tested optimization and regularization methods of Non-Negative Matrix Factorization (NMF). Implemented Algorithms include graph regularized NMF, probabilistic NMF, a first-order primal-dual algorithm ...etc.

pyDPP **[Github]**

Developed a Python package available in pip (Python packaging index) with modules for sampling from Determinantal Point Processes (DPP). Contains implementations of algorithms to sample from DPPs that encourage diversity in the selection of a subset of points from a grounded superset.

Machine Learning Contests **[Kaggle Profile]**

Kaggle Level : Competitions Expert. **Silver** medal in Kaggle Satander Value Prediction Challenge, Rank: Pvt. 185th | Pub. 189th/4484. **Bronze** medal in Kaggle Instacart Competition, Rank: Pvt. 195th | Pub. 74th/2623. Qualified for **Zonal** Round in India Hacks Machine Learning Competition by Hackerearth, Rank: 29th/860.

Review Miner | Microsoft Code.Fun.Do Hackathon **[Github]**

Developed a cross-platform application which analyses reviews from commercial websites and provides insights about products based on keyword extraction and sentiment analysis. **Winner** of Hackathon at BITS Pilani, Rank 1^{st}/90+.

Talks

Representational Capabilities of Transformers and Recurrent Architectures [Video]

- Charles University, Czech Republic 10/2025
- Saarland University, Germany 07/2025

Language Modelling with Recurrent and State Space Architectures [Video]

- Georgia Tech (SysML Guest Lecture) 11/2024

Simplicity Bias in Transformers and their Ability to Learn Sparse Boolean Functions [Video]

- Formal Language and Neural Network Seminars (FLaNN) 11/2023

On the Ability of Neural Sequence Models to Recognize Formal Languages [Slides]

- Google DeepMind NLP Reading Group 03/2022
- Formal Language and Neural Network Seminars (FLaNN) 02/2022
- MALL Lab, Indian Institute of Science 12/2020

Teaching and Services

Teaching Assistant $\times 2$ **Computational Learning Theory, Oxford** Fall 22 and Fall 23

Responsible for taking six classes on learning theory concepts and problem-solving as well as for marking the answer sheets of students. Topics include PAC-learning, SQ-learning, VC dimension, Rademacher complexity, etc.

Teaching Assistant **Neural Networks and Fuzzy Logic, BITS Pilani** Jan'18 - May'18

Responsible for designing and evaluating 3 programming tests and a course project. Conducted tutorials on python and numpy for over 70 students.

Reviewer **ICML 2024, 2023, Neurips 2023, 2022, ACL 2023, 2022, ICLR 2022, EMNLP 2022, 2021, 2020, NAACL 2021, ACL Rolling Review**