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 · Google DeepMind Scholarship
Advisors: Prof. Phil Blunsom and Prof. Varun Kanade

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

Industry Experience

Mar 2025 – Dec 2024 | **Google** — **Sunnyvale, US**
Student Researcher | Cloud AI Research Team
Explored 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. Tools: PyTorch, GPUs

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

Oct 2023 – June 2023
Intern of Technical Staff | Foundations team
Pretrained and evaluated LLMs with state-space and long-convolutional architectures. Pretrained 7B models and conducted ablations to evaluate the efficacy of architectural components. Tools: JAX, TPUs

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

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 the 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]
Ashutosh Kumar*, Satwik Bhattamishra*, 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

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

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

Selected Dev Projects

Student Developer **Google Summer of Code (GSoC) 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]

Machine Learning Contests **[Kaggle Profile]**

Kaggle Level: Competitions Expert. **Silver** medal in Kaggle Santander 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 1st/90+ teams.

Selected Talks

Representational Capabilities of Transformers and Recurrent Architectures [Video]
- Charles University, Czech Republic [10/25] | Saarland University, Germany [07/25]

Language Modelling with Recurrent and State Space Architectures [Video]
- Georgia Tech (SysML Guest Lecture) [11/24]

On the Ability of Neural Sequence Models to Recognize Formal Languages [Slides]
- DeepMind NLP Reading Group [03/22] | FLaNN [02/22] | MALL Lab, IISc [12/20]

Teaching and Services

Teaching Assistant ×2 **Computational Learning Theory, Oxford** Fall 22 and Fall 23

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

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