

# Satwik Bhattamishra

PhD student, University of Oxford

[satwikb.com](https://satwikb.com) [satwik55@gmail.com](mailto:satwik55@gmail.com) [github.com/satwik77](https://github.com/satwik77) [Google Scholar](https://scholar.google.com/citations?user=...)

## Education

<b>Present</b> <b>Oct 2021</b>	<b>University of Oxford</b> Ph.D. in Computer Science • Google DeepMind Scholarship Advisors: <a href="#">Prof. Phil Blunsom</a> and <a href="#">Prof. Varun Kanade</a>	<b>Oxford, UK</b>
<b>May 2019</b> <b>Aug 2014</b>	<b>Birla Institute of Technology and Science Pilani</b> B.E. (Hons.), Computer Science and Int. M.Sc. (Hons.), Biological Science	<b>Pilani, India</b>

## Industry Experience

<b>Mar 2025</b> <b>Dec 2024</b>	<b>Google</b> <i>Student Researcher / Cloud AI Research Team</i> 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	<b>Sunnyvale, US</b>
<b>Oct 2024</b> <b>June 2024</b>	<b>Cohere</b> <i>Intern of Technical Staff / Foundations team</i> Pretrained LLMs with <i>hybrid</i> architectures involving SSMS and Transformers. Pretrained 7B models, optimized inference strategies, & integrated $\mu$ P for hyperparameter transfer. Tools: JAX, TPUs	<b>London, UK</b>
<b>Oct 2023</b> <b>June 2023</b>	<i>Intern of Technical Staff / Foundations team</i> 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	
<b>July 2021</b> <b>July 2019</b>	<b>Microsoft Research</b> <i>Research Fellow / Advisors: <a href="#">Dr. Navin Goyal</a>, <a href="#">Dr. Monojit Choudhury</a></i> 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	<b>Bangalore, India</b>

## Publications

<b>Automata Learning and Identification of the Support of Language Models</b> <a href="#">Satwik Bhattamishra</a> , Michael Hahn, Varun Kanade 2026 Under Review	[paper] [Preprint]
<b>Benefits and Limitations of Communication in Multi-Agent Reasoning</b> Michael Rizvi-Martel, <a href="#">Satwik Bhattamishra</a> , Neil Rathi, Guillaume Rabusseau, Michael Hahn 2026 Under Review	[paper] [Preprint]
<b>The Transformer Cookbook</b> A. Yang, C. Watson, A. Xue, <a href="#">Satwik Bhattamishra</a> , J. Llerena, W. Merrill, E. Ferreira, A. Svete, D. Chiang 2026 Transactions on Machine Learning Research	[paper] [TMLR '26]
<b>A Formal Framework for Understanding Length Generalization in Transformers</b> X. Huang, A. Yang, <a href="#">Satwik Bhattamishra</a> , Y. Sarrof, A. Krebs, H. Zhou, P. Nakkiran, M. Hahn 2025 International Conference on Learning Representations	[paper] [ICLR '25]
<b>Separations in the Representational Capabilities of Transformers and Recurrent Architectures</b> <a href="#">Satwik Bhattamishra</a> , Michael Hahn, Phil Blunsom, Varun Kanade 2024 Conference on Neural Information Processing Systems	[paper] [NeurIPS '24]
<b>Understanding In-Context Learning in Transformers by Learning to Learn Discrete Functions</b> <a href="#">Satwik Bhattamishra</a> , Arkil Patel, Phil Blunsom, Varun Kanade 2024 International Conference on Learning Representations [Oral, Top 1.2%]	[paper] [ICLR '24]
<b>DynaQuant: Compressing Deep Learning Training Checkpoints via Dynamic Quantization</b> A. Agrawal, S. Reddy, <a href="#">Satwik Bhattamishra</a> , V. Sarath Nookala, V. Vashishth, K. Rong, A. Tumanov 2024 ACM Symposium on Cloud Computing	[paper] [SoCC '24]
<b>Simplicity Bias in Transformers and their Ability to Learn Sparse Boolean Functions</b> <a href="#">Satwik Bhattamishra</a> , Arkil Patel, Varun Kanade, Phil Blunsom 2023 Annual Meeting of the Association for Computational Linguistics	[paper] [ACL '23]

<b>MAGNIFICO: Evaluating the In-Context Learning Ability of Large Language Models to Generalize to Novel Interpretations</b>	<a href="#">[paper]</a>
Arkil Patel, <a href="#">Satwik Bhattamishra</a> , Siva Reddy, Dzmitry Bahdanau	
2023 Conference on Empirical Methods in Natural Language Processing <a href="#">[Oral]</a>	[EMNLP '23]
<b>Revisiting the Compositional Generalization Abilities of Neural Sequence Models</b>	<a href="#">[paper]</a>
Arkil Patel, <a href="#">Satwik Bhattamishra</a> , Phil Blunsom, Navin Goyal	
2022 Annual Meeting of the Association for Computational Linguistics	[ACL '22]
<b>Are NLP Models really able to Solve Simple Math Word Problems?</b>	<a href="#">[paper]</a>
Arkil Patel, <a href="#">Satwik Bhattamishra</a> , Navin Goyal	
2021 Conference of the North American Chapter of the Association for Computational Linguistics	[NAACL '21]
<b>On the Ability and Limitations of Transformers to Recognize Formal Languages</b>	<a href="#">[paper]</a>
Satwik Bhattamishra, Kabir Ahuja, Navin Goyal	
2020 Conference on Empirical Methods in Natural Language Processing	[EMNLP '20]
<b>On the Practical Ability of Recurrent Neural Networks to Recognize Hierarchical Languages</b>	<a href="#">[paper]</a>
<a href="#">Satwik Bhattamishra</a> , Kabir Ahuja, Navin Goyal	
2020 International Conference on Computational Linguistics <a href="#">[Best Short Paper Award]</a>	[COLING '20]
<b>On the Computational Power of Transformers and Its Implication in Sequence Modeling</b>	<a href="#">[paper]</a>
<a href="#">Satwik Bhattamishra</a> , Arkil Patel, Navin Goyal	
2020 Conference on Computational Natural Language Learning	[CoNLL '20]
<b>Unsung Challenges of Building Language Technologies for Low Resource Language Communities</b>	<a href="#">[paper]</a>
P. Joshi, C. Barnes, S. Santy, S. Khanuja, S. Shah, A. Srinivasan, <a href="#">Satwik Bhattamishra</a> , S. Sitaram, M. Choudhury, K. Bali	
16 <sup>th</sup> International Conference on Natural Language Processing	[ICON '19]
<b>Submodular Optimization-based Diverse Paraphrasing and its Effectiveness in Data Augmentation</b>	<a href="#">[paper]</a>
Ashutosh Kumar*, <a href="#">Satwik Bhattamishra</a> *, Manik Bhandari, Partha Talukdar (* = Equal Contribution)	
2019 Conference of the North American Chapter of the Association for Computational Linguistics <a href="#">[Oral]</a>	[NAACL '19]

## Selected Dev Projects

<b>Student Developer</b>	<b>Google Summer of Code (GSoC) 2016</b>	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. <a href="#">[Project Page]</a>		
<b>Machine Learning Contests</b>	<a href="#">[Kaggle Profile]</a>	
Kaggle Level: Competitions Expert. <b>Silver</b> medal in Kaggle Santander Value Prediction Challenge, Rank: Pvt. 185 <sup>th</sup>   Pub. 189 <sup>th</sup> /4484. <b>Bronze</b> medal in Kaggle Instacart Competition, Rank: Pvt. 195 <sup>th</sup>   Pub. 74 <sup>th</sup> /2623. Qualified for <b>Zonal</b> Round in India Hacks Machine Learning Competition by HackerEarth, Rank: 29 <sup>th</sup> /860.		
<b>Review Miner</b>	<b>Microsoft Code.Fun.Do Hackathon</b>	<a href="#">[GitHub]</a>
Developed a cross-platform application which analyses reviews from commercial websites and provides insights about products based on keyword extraction and sentiment analysis. <b>Winner</b> of Hackathon at BITS Pilani, Rank 1 <sup>st</sup> /90+ teams.		

## Selected Talks

<b>Representational Capabilities of Transformers and Recurrent Architectures</b>	<a href="#">[Video]</a>
<ul style="list-style-type: none"> <li>Charles University, Czech Republic [10/25]   Saarland University, Germany [07/25]</li> </ul>	
<b>Language Modelling with Recurrent and State Space Architectures</b>	<a href="#">[Video]</a>
<ul style="list-style-type: none"> <li>Georgia Tech (SysML Guest Lecture) [11/24]</li> </ul>	
<b>On the Ability of Neural Sequence Models to Recognize Formal Languages</b>	<a href="#">[Slides]</a>
<ul style="list-style-type: none"> <li>DeepMind NLP Reading Group [03/22]   <a href="#">FLaNN</a> [02/22]   MALL Lab, IISc [12/20]</li> </ul>	

## Teaching and Services

<b>Teaching Assistant</b> ×2	<b>Computational Learning Theory, Oxford</b>	Fall 22 and Fall 23
<b>Teaching Assistant</b>	<b>Neural Networks and Fuzzy Logic, BITS Pilani</b>	Jan'18 - May'18
<b>Reviewer</b>	ICML 2024, 2023; NeurIPS 2023, 2022; ACL 2023, 2022; ICLR 2022; EMNLP 2022, 2021, 2020; NAACL 2021; ACL Rolling Review	