

# Harit Vishwakarma

Senior Postdoctoral Research Fellow, University of Oxford

[harit7.github.io](https://harit7.github.io) | [harit.vishwakarma@gmail.com](mailto:harit.vishwakarma@gmail.com) | [LinkedIn](#) | [Google Scholar](#)

## Research Experience and Interests

- Foundations of machine learning, artificial intelligence with focus on **data efficiency** and **reliability**.
- **Data efficiency:** self-training, active learning, weak supervision, RLHF, DPO.
- **Reliability:** uncertainty quantification, OOD robustness, factuality, safe anytime valid inference.
- Improving **reliability** and accuracy of LLMs with **test-time compute** and **statistical inference**.
- Equally enjoy **applied/empirical research**, **building systems**, and **theoretical analysis**.

## Education

2019 – 2025 **University of Wisconsin-Madison, WI, USA.**

Ph.D. in Computer Science CGPA: 3.9/4.0.

Advisors: *Prof. Frederic Sala & Prof. Ramya Korlakai Vinayak*

Thesis: Tackling the Data Bottleneck and Inference Challenges in Machine Learning.

2014 – 2016 **Indian Institute of Science, Bangalore, KA, India.**

M.E. in Computer Science CGPA: 7.1/8.0 Class Rank: 3/50.

Advisor: *Prof. Chiranjib Bhattacharyya*.

Thesis: Discovering Groups of Correlated Event Streams from Multi-Dimensional Point Process Data.

2008 – 2012 **Shri G.S. Institute of Technology & Science, Indore, MP, India.**

B.E. in Computer Science CGPA: 76%.

## Selected Papers (Full list at the end)

ICML WS '25 **Time To Impeach LLM-as-a-Judge: Programs are the Future of Evaluation**

Tzu-Heng Huang, Harit Vishwakarma, Fred Sala

*ICML 2025 Workshop on Programmatic Representations for Agent Learning*

In Submission **Is Conformal Factuality Robust to Distractors?**

Daisuke Yamada, Yi Chen, Harit Vishwakarma, Ramya K. Vinayak

*Under Review, 2025.*

ICML '25 **Prune 'n Predict: Optimizing LLM Decision-making with Conformal Prediction**

Harit Vishwakarma, A. Mishler, T. Cook, N. Dalmasso, N. Raman, S. Ganesh

*International Conference on Machine Learning (ICML), 2025.*

AISTATS '24 **Taming False Positives in Out-of-Distribution Detection with Human Feedback**

Harit Vishwakarma, Heguang Lin, Ramya Korlakai Vinayak

*International Conference on Artificial Intelligence and Statistics (AISTATS), 2024.*

NeurIPS '23 **Promises and Pitfalls of Threshold-based Auto-labeling**

Harit Vishwakarma, Heguang Lin, Fred Sala, Ramya Korlakai Vinayak

*Neural Information Processing Systems (NeurIPS), 2023, (Spotlight).*

NeurIPS '22 **Lifting Weak Supervision to Structured Prediction**

Harit Vishwakarma, Nick Roberts, Fred Sala

*Neural Information Processing Systems (NeurIPS), 2022.*

ACM-HT '18 **Know Thy Neighbors, and More! Studying the Role of Context in Entity Recommendation**

Sumit Bhatia, Harit Vishwakarma

*ACM Conference on HyperText and Social Media (HT), 2018, (Best Paper Nominee).*

## Employment

- Nov 2025 **University of Oxford** *Senior Postdoctoral Research Fellow*, Oxford, UK.  
○ Working on fundamental aspects of LLM reliability with Prof. Yee Whye Teh.
- Jul-Oct 2025 **Snorkel AI** *Research Intern*, San Francisco, CA, U.S.  
○ Fine-tuning LLMs for complex reasoning tasks and understanding data scaling laws.  
○ Automating benchmark design using LLM as optimizer.  
○ Skills: Large Language Models (LLMs), Fine-tuning, RLHF/RLVR, trl and vllm.
- Summer 2024 **JPMorgan AI Research**, *Research Intern*, New York City, NY, U.S.  
○ Improved uncertainty quantification and LLM inference in decision-making tasks such as tool usage, MCQs, etc. using conformal prediction and scaling test-time compute.  
○ Presented this work at NeurIPS '24 (Stats for LLM workshop) and submitted to ICLR '25.  
○ Skills: Deep Learning, PyTorch, Statistical Inference, Large Language Models (LLMs), NLP.
- Summer 2021 **Amazon Alexa**, *Applied Scientist Intern*, Seattle, U.S./Remote.  
○ Developed a new method for learning entity embeddings based on multi-view representation learning.  
○ The new embeddings improved performance on entity matching tasks on a collection of songs.  
○ Showed that language models such as BERT give unreliable embeddings e.g. for dates. My proposed method overcomes these issues as it designs embeddings for each data type or attribute.  
○ Skills: Deep Learning, PyTorch, Data Analysis, NLP.
- 2016 – 2019 **IBM Research**, *Research Engineer*, Bangalore, India.  
○ Improved performance on contextual entity retrieval by combining knowledge graph and text information in a principled manner that (ACM HT 2018).  
○ Worked on quantum embeddings for knowledge graphs and showed their effectiveness in reasoning tasks (NeurIPS 2019).  
○ In addition to publishing, integrated these works into the internal neuro-symbolic reasoning system.  
○ Skills: Deep Learning, PyTorch, Large-scale Graph Processing, Graph Databases, Natural Language Processing.
- Summer 2015 **Flipkart**, *Research Intern*, Bangalore, India.  
○ Analyzed users' sessions on e-commerce and built a purchase prediction model using session features.  
○ Modeled the transaction data using the Hawkes Process to learn the interaction among product categories and using this built a recommendation system, improving average precision and recall by 46%.  
○ Skills: Data Science, Generative models, Predictive modeling of real-world data, Python, Apache Spark.
- 2012–2014 **Ittiam Systems**, *Senior Software Engineer*, Bangalore, India.  
○ Developed cloud-based video transcoding and live streaming service (FarmOTT).  
○ Led the development of an efficient media transcoding engine with in-house and open-source AV codecs.  
○ The product was showcased at the NAB (National Association of Broadcasters) show.  
○ Skills: Java and C Programming, Software Engineering, Amazon Web Services, Distributed Systems.

## Awards/Achievements

- 2024 Among the 4 nominees for Google Ph.D. fellowship from UW-Madison.
- 2023 Paper recognized as spotlight at NeurIPS '23, (top 3.06%).
- 2022 Top reviewer for NeurIPS '22, (top, 8%) .
- 2020 Paper recognized as spotlight at NeurIPS '20, (top 2.96%).
- 2023 NeurIPS Scholar Award for years 2019, 2022 and 2023.
- 2018 Best paper nominee in ACM HyperText.
- 2018 ACM HyperText Ted Nelson Newcomer Award. Awarded to the best paper by new authors.
- 2014 All India Rank 155 (top 0.1%) in GATE – national level exam for grad schools in India.
- 2008 State Rank 113 (top 0.1%) in state engineering entrance test (MP-PET).

## Skills / Languages / Libraries

- Languages Python, Java, C/C++, Javascript, SQL, Shell (bash) Script.
- Libraries PyTorch, Apache Spark, Sklearn (scikit-learn), numpy, pandas, JAX, Tensorflow, Huggingface.
- Databases/OS MySQL, MongoDB, Neo4J, Unix/Linux.
- Scalability Distributed computing with PyTorch, Apache Spark, Amazon Web Services, Cloud Computing.
- Fundamentals Data Structures and Algorithms, Probability & Statistics, Linear Algebra, Optimization.

## Selected Courses

UW-Madison Mathematical Foundations of Machine Learning (CS 761), Theoretical Machine Learning (CS 861), Non-Linear Optimization I (CS 726), Big Data (CS 744), Topics in Deep Learning (CS 839).

## Talks

- Aug, 2025 **Google DeepMind, Seattle.** *Towards Reliable AI Without Breaking the Bank.*
- Jan, 2025 **Microsoft Research, Vancouver.** *Towards Reliable AI: From Data to Deployment.*
- Nov, 2024 **AmFam and Data Science Institute, Madison.** *Improved Confidence Functions for Auto-labeling.*
- Nov, 2024 **SILO Seminar, UW-Madison.** *Improved Confidence Functions for Auto-labeling.*
- Aug, 2024 **JPMorgan AI Research, NY.** *Prune 'n Predict to Improve Decision-Making with LLMs.*
- Mar, 2024 **IFDS Seminar, UW-Madison.** *Improved Confidence Functions for Auto-labeling.*
- Nov, 2023 **MLOPT Seminar, UW-Madison.** *Promises and Pitfalls of Threshold-based Auto-labeling.*
- Feb, 2023 **IFDS Seminar, UW-Madison.** *Promises and Pitfalls of Threshold-based Auto-labeling.*
- Oct, 2023 **IFDS Seminar, UW-Madison.** *Human-in-the-Loop OOD Detection with FPR Control.*

## Service and Organization

- 2025 Co-organizing a workshop on reliability of LLMs at AAAI 2026.
- 2021 – Now Served as reviewer for NeurIPS, ICML, ICLR, AISTATS, AAAI, TMLR, DMLR.
- 2023 Organized a reading group on ML theory.
- 2016 Organized machine learning competition during CSA Open days at IISc.
- 2012 Organized coding competitions in undergraduate techfest.

## Mentoring

- 2024-25 **Daisuke Yamada**, CS PhD Student, UW-Madison.
  - Research on OOD robustness, safe anytime valid inference, conformal factuality.
  - 2 successful submissions to top conferences.
- 2023-25 **Yi Chen**, ECE PhD Student, UW-Madison.
  - Research on auto-labeling, semi-supervised learning and conformal factuality.
  - 3 successful paper submissions, 1 publication.
- 2022-23 **Tzu-Heng Huang**, CS PhD Student, UW-Madison.
  - Research on parameter markets, game theory, and optimization.
  - 1 publication in NeurIPS.
- 2022-23 **Heguang Lin**, CS & Math Major Student, UW-Madison → CS Masters, UPenn .
  - Research on active learning, auto-labeling, and OOD robustness.
  - 2 publications in NeurIPS and AISTATS.
- 2023-24 **Sui Jiet Tay**, CS Major Student, UW-Madison → MS at NYU Courant.
  - Research on auto-labeling and semi-supervised learning.
  - 1 publication in NeurIPS and 1 successful submission.
- 2023-24 **Srinath Namburi**, CS Masters Student, UW-Madison→ G.E. AI Research.
  - Research on auto-labeling and semi-supervised learning.
  - 1 publication in NeurIPS and 1 successful submission.

## Teaching

- Spring 2022 **Mathematical Foundations of Machine Learning (CS 761)**  
Role: Lead Teaching Assistant. Instructor: Prof. Rob Nowak.
- Fall 2021 **Machine Learning (CS 760)**  
Role: Teaching Assistant. Instructor: Prof. Fred Sala.
- Fall 2020 **Java Programming (CS 400)**  
Role: Lead Teaching Assistant. Instructor: Prof. Florian Heimerl.
- Spring 2020 **Java Programming (CS 300)**  
Role: Teaching Assistant. Instructor: Prof. Gary Dahl.

## Full List of Papers

- Manuscript '25 **Measuring the Effectiveness of RLVR in Low Data and Compute Regimes**  
J Bauer, T Walshe, D Pham, H Vishwakarma, A Parchami, F Sala  
*Under Review, 2025.*
- [arXiv '25](#) **Automating Benchmark Design**  
A Dsouza, H Vishwakarma, Z Qi, J Bauer, D Pham, T Walshe, A Parchami, F Sala, P Varma  
*Under Review, 2025.*
- [ICML WS '25](#) **Time To Impeach LLM-as-a-Judge: Programs are the Future of Evaluation**  
Tzu-Heng Huang, Harit Vishwakarma, Fred Sala  
*ICML 2025 Workshop on Programmatic Representations for Agent Learning*
- Manuscript '25 **Is Conformal Factuality Robust to Distractors?**  
Daisuke Yamada, Yi Chen, Harit Vishwakarma, Ramya K. Vinayak  
*Under Review, 2025.*
- [arXiv '25](#) **Adaptive Scoring and Thresholding with Human Feedback for Robust OOD Detection**  
Daisuke Yamada, Harit Vishwakarma, Ramya K. Vinayak  
*Under Review, 2025.*
- [ICML '25](#) **Prune 'n Predict: Optimizing LLM Decision-making with Conformal Prediction**  
Harit Vishwakarma, A. Mishler, T. Cook, N. Dalmasso, N. Raman, S. Ganesh  
*International Conference on Machine Learning (ICML), 2025.*
- [ICML '25](#) **Rethinking Confidence Scores and Thresholds in Pseudolabeling-based SSL**  
Harit Vishwakarma\*, Yi Chen\*, Srinath Namburi, Sui J. Tay, Ramya Vinayak, Fred Sala  
*International Conference on Machine Learning (ICML), 2025.*
- [NeurIPS '24](#) **Pearls from Pebbles: Improved Confidence Functions for Auto-labeling**  
Harit Vishwakarma, Yi Chen, Sui Jiet Tay, Srinath Namburi, Fred Sala, Ramya K. Vinayak  
*Neural Information Processing Systems (NeurIPS), 2024.*
- [NeurIPS '24](#) **OTTER: Effortless Label Distribution Adaptation of Zero-shot Models**  
Changho Shin, Jitian Zhao, Sonia Cromp, Harit Vishwakarma, Fred Sala  
*Neural Information Processing Systems (NeurIPS), 2024.*
- [AISTATS '24](#) **Taming False Positives in Out-of-Distribution Detection with Human Feedback**  
Harit Vishwakarma, Heguang Lin, Ramya Korlakai Vinayak  
*International Conference on Artificial Intelligence and Statistics (AISTATS), 2024.*
- [NeurIPS '23](#) **Promises and Pitfalls of Threshold-based Auto-labeling**  
Harit Vishwakarma, Heguang Lin, Fred Sala, Ramya Korlakai Vinayak  
*Neural Information Processing Systems (NeurIPS), 2023, (Spotlight).*
- [NeurIPS '23](#) **Train 'n Trade: Foundations of Parameter Markets**  
Tzu-Heng Huang, Harit Vishwakarma, Fred Sala  
*Neural Information Processing Systems (NeurIPS), 2023.*
- [ICLR WS '23](#) **ScriptoriumWS: A Code Generation Assistant for Weak Supervision**  
T. Huang, C. Cao, S. Schoenberg, H. Vishwakarma, N. Roberts, F. Sala  
*Workshop on Deep Learning for Code (DL4C), ICLR, 2023.*
- [NeurIPS '22](#) **Lifting Weak Supervision to Structured Prediction**  
Harit Vishwakarma, Nick Roberts, Fred Sala  
*Neural Information Processing Systems (NeurIPS), 2022.*
- [ICLR '22](#) **Universalizing Weak Supervision**  
Changho Shin, Winfred Li, Harit Vishwakarma, Nick Roberts, Fred Sala  
*International Conference on Learning Representations (ICLR), 2022.*
- [NeurIPS '20](#) **Optimal Lottery Tickets via Subset-Sum: Logarithmic Over-param. is Sufficient**  
Ankit Pensia, Shashank Rajput, Alliot Nagle, Harit Vishwakarma, Dimitris Papailiopoulos  
*Neural Information Processing Systems (NeurIPS), 2020, (Spotlight).*

- NeurIPS '20 **Attack of the Tails: Yes, you Really Can Backdoor Federated Learning**  
H. Wang, K. Sreenivasan, S. Rajput, H. Vishwakarma, S. Agarwal, J. Sohn, K. Lee, D. Papailiopoulos  
*Neural Information Processing Systems (NeurIPS), 2020.*
- NeurIPS '19 **Quantum Embedding of Knowledge for Reasoning**  
D. Garg, S. Ikkbal, S. K Srivastava, H. Vishwakarma, H. Karnam, L. V. Subramaniam  
*Neural Information Processing Systems (NeurIPS), 2019.*
- ACM-HT '18 **Know Thy Neighbors, and More! Studying the Role of Context in Entity Recommendation**  
Sumit Bhatia, Harit Vishwakarma  
*ACM Conference on HyperText and Social Media (HT), 2018, (Best Paper Nominee).*
- D4GX '17 **An End-To-End Machine Learning Pipeline That Ensures Fairness Policies**  
S. Shaikh, H. Vishwakarma, S. Mehta, K. R. Varshney, K. N. Ramamurthy, D. Wei  
*Bloomberg Data for Goods Exchange (D4GX), 2017.*