Shresth Verma

Phone: +1 857 799 0801
Email: sverma@g.harvard.edu
Website: vermashresth.github.io

Education

Harvard University 2023 - Present
PhD in Computer Science
▷ Advised by Prof. Milind Tambe

Indian Institute of Information Technology, Gwalior 2015 - 2020
B.Tech. + M.Tech. in Information Technology
▷ Advised by Prof. Joydip Dhar
▷ Received outstanding grades in Bachelor's and Master's theses.

Work Experience

The Alan Turing Institute – United Kingdom May 2024
Facilitator working with Adeline Pelletier, Carlos Mougan

Google DeepMind – India Jun 2021 - Aug 2023
Pre-doctoral researcher with Dr. Aparna Taneja

UnitedHealth Group – India Aug 2020 - May 2021
Data Scientist with Kishore V. Ayyadevara

Featured Publications

* = joint first author

[1] *DPO-PRO: Direct Preference Optimization with Preference Robustness*
Woo Kim C.*, **Verma S.***, Tec M.*, Tambe M.
Reliable Machine learning Workshop at NeurIPS 2025 **[ReliableML NeurIPS 2025]**

[2] *Navigating the Social Welfare Frontier: Portfolios for Multi-objective Reinforcement Learning*
Woo Kim C.*, Moondra J.*, **Verma S.,** Pollack M., Kong L., Tambe M., Gupta S.
International Conference on Machine Learning 2025 **[ICML 2025]**

[3] *Priority2Reward: Incorporating Healthworker Preferences for Resource Allocation Planning*
Verma S., Kong L., Boehmer N., Tambe M.
AAAI Conference on Artificial Intelligence 2025 **[AAAI 2025]**

[4] *Balancing Act: Prioritization Strategies for LLM-Designed Restless Bandit Rewards*
Verma S., Boehmer N., Kong L., Tambe M. ;
GenAI for Health Workshop at NeurIPS 2024 **[GenAI4Health NeurIPS 2024]**

[5] *Limited Resource Allocation in a Non-Markovian World*
Danassis P., **Verma S.,** Killian J., Taneja A., Tambe M.
International Joint Conference on Artificial Intelligence **[IJCAI 2023]**

[6] *Restless Multi-Armed Bandits for Maternal and Child Health: Results in Decision-Focused Learning*
Verma S., Mate A., Wang K., Taneja A., Tambe M.
International Conference on Autonomous Agents and Multiagent Systems 2023 **[AAMAS 2023]**

[7] *Scalable Decision Focused Learning in Restless Multi-Armed Bandits*
Wang K.*, **Verma S.***, Shah S., Mate A., Taneja A., Tambe M.
AAAI Conference on Artificial Intelligence 2023 **[AAAI 2023]**

[8] *Robust Planning over Restless Groups*
Killian J.*, Xu L.*, Biswas A.*, **Verma S.*** , Nair V., Rodriguez, P., Johnson-Yu S., Taneja A., Tambe M.
AAAI Conference on Artificial Intelligence 2023 **[AAAI 2023]**

[9] *Towards Sample-Efficient Learners in Population based Referential Games through Action Advising*
Verma S.
International Conference on Autonomous Agents and Multiagent Systems **[AAMAS 2021]**

Featured Projects and Collaborations

Goal Condition RL for Multi-Turn Dialogues — Harvard University
w/ Prof. Kiante Brantley — *August 2025 - Present*

- Developed LLM finetuning techniques for multi-turn dialogues where current LLMs fail to ask clarifying questions or engage in information gathering.
- Formalised this as a goal-conditioned RL problem and theoretically analysed the number of steps needed to reach a human satisfaction level in a conversation.

Aggregating Multi-User Preferences for LLM Alignment — Harvard University
w/ Prof. Niclas Boehmer, Prof. Swati Gupta — *April 2024 - Present*

- Designed Social Choice Language Model (SCLM), a framework to generate reward function code for reinforcement learning policies under conflicting preferences, improving social welfare utility by upto 15% [**GenAI4Health NeurIPS 2024**].
- Formulated Portfolios for LLM policies which can approximate the entire Pareto front of multi-dimensional human preferences with $>97\%$ accuracy through < 3 policies. This allows decision-makers to efficiently navigate tradeoffs [**ICML 2025**].

Code-Generation using LLMs and Robust Post-Training — Harvard University
w/ Dr. Mauricio Tec, Prof. Milind Tambe — *December 2024 - Present*

- Proposed Distributionally Robust Direct Preference Optimization (DPO-PRO) to reduce inference time costs from linear to constant in problem size through offline fine-tuning on noisy code preference datasets [**ReliableML NeurIPS 2025**].
- Developed an LLM-agent based on evolutionary search and self-reflection which converts user preferences into reward function code for reinforcement learning policies and achieves $5\times$ speedup over Decision Language Model [**AAAI 2025**].

Decision Focused Learning — Google DeepMind
w/ Prof. Kai Wang — *June. 2021 - August 2023*

- Developed end-to-end differentiable pipelines combining estimation and planning using Decision Focused Learning, improving performance by 31% in a real-world study [**TSRML NeurIPS 2022**, **AAAI 2023a**, **AAMAS 2023**]
- Formulated provably-fair methods to simultaneously learn and perform budget allocations in restless bandit problems. Our method also optimizes for equitable decision quality, outperforming baselines by 20-40% in fairness metrics [**UAI 2024**]

Restless Multi-Armed Bandits for Public Health — Google DeepMind
w/ Prof. Lily Xu, Prof. Arpita Biswas, Dr. Aparna Taneja — *June. 2021 - August 2023*

- For mobile health programs in India under limited budget, developed robust regret-minimizing policies to allocate resources. Resultant robust policies reduce minimax regret by up to 50% [**AAAI 2023b**].
- Conducted large-scale Randomized Control Trials (RCT) involving 100K+ beneficiaries to show improvement in engagement over the current standard of care by using Restless Multi-Armed Bandit (RMAB) models [**AAAI 2022** , **IAAI 2023**].
- Developed a Multi-Action Index for sequential resource planning based on time-series forecasting for non-markovian environments, thereby preventing 90.8% more engagement drops as compared to SoTA. [**IJCAI 2023** , **IAAI 2024**].

Hospital Readmission Risk Modeling — United Health Group
w/ Kishore V. Ayyadevara — *Aug 2020 - May 2021*

- Worked alongside Chief Medical Officer to model hospital readmission risk for 40M+ beneficiaries. Used high-dimensional ICD-10 embeddings to encode visit history and produced explainable predictions for caregivers.
- Enhanced real-time capabilities to track patient wellness journeys using the largest healthcare graph database in the world (10B+ nodes). Streamlined cross-team collaborations using GitOps-based workflows.

Learning to Communicate through Deep Multi-Agent Reinforcement Learning — Master's thesis
w/ Prof. Joydip Dhar — *July 2019 - July 2020*

- Demonstrated the emergence of language systems in speaker and listener agents for written, visual, and population-based referential games. [**AAAI 2020** , **LaReL ICML 2020** , **AAMAS 2021** , **LaReL NeurIPS 2022**]
- Developed an autonomously coordinated multi-agent model for watershed management optimized through inter-agent communication and intrinsic social-motivation rewards [**AASG AAMAS 2023**].

Awards, Grants & Honours

Accepted into Supervised Program for Alignment Research (SPAR) AI Safety Fellowship 2025
Student Travel Grant by Hudson River Trading for presenting research poster at ICML, Vancouver, CA 2025
Accepted into ERA AI Safety Fellowship, University of Cambridge, UK 2025
Accepted into Eastern European Machine Learning Summer School, Bosnia and Herzegovina 2025
Travel Grant for attending Data Study Group at The Alan Turing Institute, London, UK 2024
Accepted into Harvard's Technical AI Safety Fellowship 2024
Student Travel Grant for presenting research poster at AAAI, New York, US 2020
Student Travel Grant for attending IEEE High Performance Computing Conference, Hyderabad, IN 2019
Rotaract National Technical Quiz Pune, India - 1^{st} in India among $500+$ teams 2018
Ramanujan Mathematics Olympiad - 3^{rd} in State among $400+$ participants 2013
Regional Mathematics Olympiad - 22^{nd} in State among $2000+$ participants 2013
National Cyber Olympiad - 8^{th} in India among $30000+$ participants 2012
Qualified for Indian National Mathematics Olympiad thrice - Top $900/50000$ in India 2011-2013

Invited Talks

- Google Deepmind Game Theory Group, September 2024. *Prioritization Strategies for LLM Designed Reward Functions for Restless Bandits*
- Harvard SEAS AI for Social Good Seminar, February 2024. *Ensuring Group Fairness in Decision Focused Learning through RMABs*
- Chopal Seminar Series at Google Research India, April 2023. *A Review of Restless Multi-Armed Bandits for Mobile Health*
- Multi-Agent Systems for Social Impact Seminar at Google Research India, September 2021. *Robustness in Restless Multi-Armed Bandits*
- UHG-Optum India Data Science Seminar, February 2021. *Tutorial on Graph Neural Networks for Healthcare*
- Symposium on CyberPhysical Systems at Indian Institute of Science, July 2019. *Deep Reinforcement Learning for Damage Adaptation in Robotics*
- UHG-Optum India Data Science Seminar, June 2019. *Attention Mechanisms for Optical Character Recognition*
- Abhigyan Abhikaushalam Students Forum Seminar Series, January 2019. *An Introduction to Git and Open Source Software*

Academic Service

PC Member Autonomous Agents for Social Good Workshop at AAMAS 2024
Reviewer AAAI '26, IJCAI '25, AAMAS '25, IAAI '25, MINT NeurIPS '24, AASG AAMAS '24, TSRML NeurIPS '22
Teaching AI for Social Impact Course - Harvard Fall 2024, Open Source Software - IIITM Gwalior 2019

Selected Open-Source Projects

- Developed and open-sourced **Jupyter-Probe**, a library to monitor, declare, and manage resource usage on shared Jupyter environments. Published the library on PyPI software repository to be used as pip package.
- Contributor to **TensorForce**, a library for production-grade Reinforcement Learning (3K+ github stars). Implemented bindings with RL simulation environments such as Deepmind Pycolab and Unity ML Agents.
- Contributor to scientific python libraries for Astronomy - **AstroPy**, and Heliophysics - **SunPy**. Added features in the Time module for astronomical calculations which is at the core of the libraries' functionality.

Skills

Programming Languages and Tools Python, C++, C#, LaTeX, Git, Docker, Kubernetes
ML Frameworks and RL Environments PyTorch, TensorFlow, JAX, Stable Baselines3, RLlib, OpenAI Gym, MuJoCo
LLM Training and Experimentation Hugging Face Transformers, PEFT, LoRA, DeepSpeed, FSDP, W&B, Hydra