Vansh Kapoor

📱 +1 412-606-5875 • ✉ vanshk@cs.cmu.edu • 🌐 vansh28kapoor.github.io

Research Interests

Reinforcement Learning & Applications in LLM Reasoning/RLHF/Robotics, Generative Modeling (Diffusion, Stochastic Interpolants), Multi-Agent Learning & Game Theory

Education

Carnegie Mellon University (CMU)   Pittsburgh, PA
Master of Science in Machine Learning | GPA: **4.22/4**   Dec'25
- Key Courses: Deep RL (A+), Generative AI (A+), Statistics (A+), ML Optimization (A+), Adv ML (A+)
- Teaching Assistant for *Deep Reinforcement Learning*; delivered recitations, and mentored 8+ team projects

Indian Institute of Technology Bombay (IITB)   Mumbai, India
Bachelor of Technology in Electrical Engineering with Honors   May'24
GPA: **9.76/10** *(Department Rank* 6th *amongst 200+ students)*   *Honors GPA:* **10/10**
- Received *Undergraduate Research Award* for outstanding research conducted in Partially Observable MDPs
- Awarded AP Grade (*Course Topper* among 200+) in *Reinforcement Learning*, Adv ML, Adv Deep Learning

Publications

- V. Kapoor, J. Nair. "MDPs with State Sensing Costs." *Under review at AISTATS'26*   (Paper 🔗)
- V. Kapoor, A. Gupta, H. Chen, A. Beniwal, J. Huang, A. Kumar. "TRIM: Hybrid Inference via Targeted Stepwise Routing in Multi-Step Reasoning Tasks" *Under review at ICLR'26*   (Paper 🔗)

Research & Industry Experience

TRIM: Hybrid Inference via Targeted Stepwise Routing in Reasoning Tasks 🔗 *(May'25 - Oct'25)*
Guide: Prof. Aviral Kumar, Anurag Beniwal   *Amazon AI Intern*
- Introduced TRIM, a step-level LLM routing framework for reasoning tasks (math, code) that prevents cascading failures by routing only critical steps to stronger models, while delegating routine continuations to smaller ones.
- Designed routing policies ranging from threshold-based heuristics to **RL-trained and POMDP-based controllers** that leverage stepwise rewards and uncertainty estimates to make budget-aware, step-level intervention decisions.
- Demonstrated **5x higher cost efficiency** than SOTA routing approaches on MATH-500, with RL- and POMDP-based variants matching the performance of a large expert model while using up to 80% fewer expensive tokens.

Partially Observable MDPs With State Sensing Cost 🔗 *(Jan'23 - Jan'25)*
Guide: Prof. Jayakrishnan Nair   *Research Assistant*
- Conducted research on a special class of Partially Observable MDPs (POMDPs) with state sensing cost, where the agent remains uncertain about its state, unless a **fixed cost is paid to reveal the state**.
- Designed the Selective Policy Improvement (SPI) algorithm to compute near optimal policies **outperforming state-of-the-art POMDP solvers** across all sensing costs with sub-optimality guarantees.
- Derived lower bounds on optimal value function, enabling **computation of suboptimality gaps** for arbitrary policies and established a **quantitative threshold on sensing cost** below which always sensing is provably optimal.

Early Rumor Detection on Large-Scale Social Networks 🔗 *(Jun'23 - Jul'24)*
Guide: Prof. Nikhil Karamchandani, Dr. Manish Jain   *Google DeepMind Collaborator*
- Developed online learning algorithms, in collaboration with Google, for detecting bot activity on social platforms such as YouTube by **leveraging crowd-sourced reliability signals**, achieving mitigation of rumor cascades.
- Designed scalable approximations enabling our algorithms to **operate efficiently on massive, adversarial graphs**.
- Established performance guarantees and demonstrated superior detection performance over SOTA greedy baselines.

Google | *Silicon Engineering Intern* 🔗   *(May'23 - Jul'23)*
- Optimized **design verification** process by **15%** with toggle coverage analysis using Python-based automation.
- Developed **automated checkers** for data retention flops in low-power mode applications, streamlining verification.

Selected ML Projects

Imitation Learning from Denoising Diffusion Models ↗ Sep'24 - Nov'24

○ Developed **transformer-based diffusion policies** for imitation learning in BipedalWalker OpenAI Gym environment, achieving **98% of expert performance** and benchmarking against algorithms like DAgger and Behavior Cloning.

Text-to-Image Diffusion Models with Enhanced Semantic Understanding ↗ Jan'24 - May'24

○ Devised SUR (Semantic Understanding & Reasoning) architecture utilizing **LLAMA-based prompt enrichment.**

○ Boosted **multi-modal** visual question answering accuracy (counting/color/action) by **20%** over baseline CLIP.

MuZero VS Actor-Critic & DQN for Atari Games ↗ Aug'24 - Dec'24

○ Implemented MuZero using Upper-Confidence Tree Search (**UCT**) and evaluated its performance against model-free algorithms: Actor-Critic & DQN, demonstrating **4x higher sample efficiency** for maximum reward on CartPole.

ToyGPT: Miniature GPT for Arithmetic Addition ↗ Sep'24 - Nov'24

○ Trained small GPTs for 2-digit addition task evaluating **Positional Encoding variants** and Layer-Norm architectures.

Deep Recurrent Q-Learning for Partially Observable MDPs ↗ Aug'23 - Dec'23

○ Implemented RL-LSTM-Q network, integrating **Transfer Learning** & LSTM for playing **flickering Atari games.**

RL in Billiards and Football Half-field Offense ↗ Jul'23 - Dec'23

○ Implemented **Monte-Carlo Tree Search (MCTS)** for potting balls in minimal attempts (< 10) for noisy billiards.

○ Executed MDP Planning to devise an optimal half-field football offense strategy using value and policy iteration.

Teaching & Mentoring Experience

Graduate Teaching Assistant

Deep Reinforcement Learning (CMU Fall'25)   Instructor: *Prof. Aviral Kumar*

○ Assisted in creating solutions and grading exams/assignments for 200+students, while mentoring 8+ team projects.

Error Correcting Codes (IITB Fall'24)   Instructor: *Prof. Nikhil Karamchandani*

○ Assisted the instructor in creating solutions and grading course assignments and quizzes for a batch of 40+ students.

FIITJEE Brand Ambassador   *(Oct'20 - Present)*

○ Mentored senior high school students through personalized sessions, seminars, and workshops, guiding them in selecting engineering disciplines by clarifying future career paths and their prospects.

Skills

ML Frameworks: PyTorch, JAX, TensorFlow applied to RL, NLP, and CV research and course projects
Programming Languages: Python, Julia, C++, C, MATLAB

Courses Undertaken

CMU...

Reinforcement Learning: Deep Reinforcement Learning, Algorithmic Foundations of Interactive Learning
ML & AI: Generative AI, Deep Learning for Vision, ML Optimization, Statistics, ML in Practice, Adv ML

IIT Bombay..

Reinforcement Learning: Stochastic Optimization, Foundations of Intelligent & Learning Agents, Online Learning & Decision Making, Control Theory
Machine Learning: Probabilistic Graphical Models, Advanced Deep Learning, Image Processing
Applied Probability: Advanced Probability and Random Processes, Advanced Concentration Inequalities, Markov Chains and Queuing Systems, Algorithmic Game Theory
Mathematics and Physics: Calculus I & II (Differential and Integral Calculus), Linear Algebra, Complex Analysis, Differential Equations I & II, Number Theory & Cryptography, Quantum Physics, Electromagnetism