Last updated in December 2024

Rohin Manvi

rohinm@cs.stanford.edu (818) 913-5171 rohinmanvi.github.io rohinmanvi rohinmanvi

Education

Stanford University *Sept 2020 – June 2025*
Bachelor of Science (BS) in Computer Science

- Concurrent enrollment in Coterminal program for MS in Computer Science since Sept 2023 (not pursuing completion)
- **Selected Coursework:** Large Language Models from Scratch, Deep Reinforcement Learning, Decision Making Under Uncertainty, Natural Language Processing, Computer Vision, Machine Learning with Graphs, Artificial Intelligence Principles

Publications and Preprints

Adaptive Inference-Time Compute: LLMs Can Predict if They Can Do Better, Even Mid-Generation Paper
Rohin Manvi, Anikait Singh, Stefano Ermon
Under Review at International Conference on Learning Representations (ICLR), 2025

Large Language Models are Geographically Biased Paper
Rohin Manvi, Samar Khanna, Marshall Burke, David Lobell, Stefano Ermon
International Conference on Machine Learning (ICML), 2024

GeoLLM: Extracting Geospatial Knowledge from Large Language Models Paper
Rohin Manvi, Samar Khanna, Gengchen Mai, Marshall Burke, David Lobell, Stefano Ermon
International Conference on Learning Representations (ICLR), 2024

Research Experience

Ermon Group @ Stanford (PI: Prof. Stefano Ermon) *Apr 2024 – present*
Research Assistant

- Trained LLMs to self-evaluate inexpensively with next-token prediction utilizing existing KV cache
- Developed capability-aware self-evaluations to adaptively allocate samples and select the best one
- Developed mid-generation self-evaluations to stop the generation of unpromising samples early
- Expanded the action space of LLMs with text editing or copying through special tokens

Sustain Lab @ Stanford (PIs: Prof. Stefano Ermon, Prof. David Lobell) *June 2023 – Mar 2024*
Research Assistant

- Developed fine-tuning and prompt generation method for geospatial predictions with precise coordinates
- Implemented various LLMs and baselines including GPT-3.5, LLaMa 2, RoBERTa, GPT-2, XGBoost, KNN
- Uncovered a range of sensitive geographic biases in LLMs using zero-shot predictions and a novel bias score
- Tested time-series forecasting methods and multimodal geospatial predictions with LLMs

Stanford Intelligent Systems Laboratory (PI: Prof. Mykel Kochenderfer) *Jan 2023 – June 2023*
Research Assistant

- Leveraged the common-sense abilities of LLMs to enhance decision-making in autonomous driving
- Enabled LLMs to guide autonomous vehicles using a hierarchical control scheme
- Trained LLMs to generate precise textual waypoints that serve as inputs for a low-level controller
- Designed and implemented state estimation using sensor fusion with lidar and 4 cameras

Industry Experience

Lacework *Mountain View, CA*
Software Engineer Intern *June 2023 – Aug 2023*

- Boosted anomaly detection by 3% by adding a K8s audit logs preprocessing layer for transformer model
- Enhanced Kubernetes Dossier by developing data pipelines that use Lacework's K8s node and cluster collectors
- Increased query speed through a data aggregation optimization layer, reducing data volume by 100x
- Streamlined data extraction from Snowflake via refined queries for K8s resources, including nodes, pods, etc.

Meta *New York, NY*
Software Engineer Intern *June 2022 – Sept 2022*

- Designed a new platform on Facebook Watch to boost user engagement, ultimately increasing new follows by 1% while accumulating 6M+ impressions
- Built custom UI components using Hack/PHP to minimize client-side latency and improve user interactivity
- Implemented heavily-optimized Python data pipeline to identify users with the most interest in a given topic
- Wrote affinity inference algorithm and custom page sorting logic to find users who would enjoy this feature

Technical Skills

Machine Learning: Python, PyTorch, Pandas, Numpy, Jupyter/Colab

High Performance Compute (HPC): Slurm, Distributed Training and Inference, DeepSpeed, GPU clusters

Cloud: Kubernetes, Docker, AWS, GCP