RISHABH JAIN

(646)484-0525 rj2790@columbia.edu r-rishabh-j.github.io linkedin.com/in/rishabhj11 github.com/r-rishabh-j

EDUCATION

Columbia University New York Aug 2025 - Dec 2026
Master of Science (MS) Computer Science (Machine Learning) *New York, NY*
- Machine Learning, High Performance ML, Natural Language Processing, Computational Aspects of Robotics

Indian Institute of Technology (IIT) Ropar Jul 2019 - May 2023
B.Tech (Honors) Computer Science and Engineering, Concentration in AI *Ropar, IND*

PROFESSIONAL EXPERIENCE

Software Engineer, Arista Networks Jul 2023 - Jun 2025
C, C++, Python *Bengaluru, India*
- Engineered low-latency data and control plane components in C++ for the EOS software forwarding engine, enhancing throughput and efficiency for high-volume packet processing and telemetry across 16 core repositories
- Architected scalable state management agents and synchronization modules in C++ for ICMP, TCP, IPFIX and GRE protocol stacks to support up to 90 million entries in the EOS concurrent packet flow hash table
- Re-wrote network switch stress test libraries with multiprocessing in Python, achieving a 6× gain in eval throughput
- Led the creation of an automated build orchestration tool that resolves dependency graphs for upstream AlmaLinux RPMs, streamlining workflows for 15+ teams during a company-wide transition from Perforce to Git

Edison AI Intern, General Electric Healthcare Jun 2022 - Jul 2022
PyTorch, FastAPI, PostgreSQL, Docker *Bengaluru, India*
- Built an end-to-end spatio-temporal patient tracking pipeline using YOLOv5 and PostgreSQL, allowing for the continuous storage and retrieval of patient location data. Deployed it through containerized APIs using FastAPI
- Ablated YOLOv5 to develop a lightweight model, and fine-tuned it with a 5GB GPU on 30,000 self-annotated images

RESEARCH WORK AND PROJECTS

NFR Benchmarking for IT Automation Agents in IBM ITBench Oct 2025 - ongoing
AI Agents, CrewAI, Langfuse | github.com/ITBench-NFR *IBM Research, Columbia University*
- Co-developed a non-functional requirements evaluation framework extending ITBench, defining a comprehensive two-level taxonomy for agent-specific requirements (cost efficiency, reliability, observability) and instrumenting SRE, CISO and Mini-SWE agents with Langfuse, and vLLM for granular telemetry
- Conducted comparative evaluations across ReAct and Plan&Execute architectures on 15 SRE incidents and 3 CISO scenarios using Gemini and Qwen LLMs, revealing Plan&Execute agents achieved up to 15x higher Prompt-to-Completion Ratio and significantly lower latency than ReAct

Aligning LLMs for Speculative Decoding via Task-Adaptive Knowledge Distillation Oct 2025 - Dec 2025
PyTorch, Speculative Decoding, Knowledge Distillation | github.com/r-rishabh-j/distillSpec *Columbia University*
- Implemented speculative decoding with a draft model supporting prompt batching and non-uniform acceptance lengths
- Performed white-box, token-level On-Policy Knowledge Distillation to align low-cost draft models from Qwen3, SmolLM families with larger target models, effectively mitigating exposure bias to accelerate speculative generation
- Benchmarked token and sequence level acceptance rates over Forward KL, Reverse KL and JSD divergence objectives, achieving a 5% increase in token acceptance rate after just 1 epoch of distillation on GSM8k and 4% on CNN-DM

Viewpoint-Invariant Robot Manipulation via 3D Geometric Priors Oct 2025 - Dec 2025
PyTorch, MuJoCo, Gymnasium | github.com/r-rishabh-j/3DEgoACT *Columbia University*
- Modified ACT to take PointNet-encoded 3D priors as input tokens along with egocentric 2D features to mitigate inference-time covariate shift from view-point perturbations in imitation learning based policies
- Performed ablations to demonstrate that egocentric cues are crucial alongside allocentric 3D features for contact-rich tasks
- Demonstrated zero-shot generalization to novel viewpoints, achieving a 70% success rate in scenarios where standard imitation learning failed by effectively decoupling global geometric structure from view-dependent appearance

Video Transformer Based Multi-view Body Language and Behaviour Recognition May 2023 - Oct 2023
PyTorch, Deep Learning, Computer Vision *Monash University, IIT Ropar*
- Built a multi-view feature-fusion pipeline with a finetuned VideoSwin transformer backbone for multi-label classification
- Placed 2nd in the ACM MultiMedia 2023 Bodily Behaviour Recognition Grand Challenge | GitHub: MAGIC-TBR
- Published papers at ACM MultiMedia 2023 and IEEE Transactions on Affective Computing

TECHNICAL SKILLS

Languages: C, C++, Python, Java
Math & AI: NumPy, OpenCV, PyTorch, Gym, MuJoCo
Database & Backend: PostgreSQL, PostGIS, FastAPI
Tools: Git, Perforce, Bash, Docker, HuggingFace, vLLM