

RAUNIT KOHLI

raunitkohli.com [774-262-9233](tel:774-262-9233) raunitkohli@gmail.com [raunit-kohli](https://www.linkedin.com/in/raunit-kohli)

WORK EXPERIENCE

Rocket Software

Boston, MA

Machine Learning Software Engineer III

09/2024 – Present

- Design and implement novel Large Language Model Citation Generator using multi-Gaussian clustering algorithms and dynamic content similarity techniques, yielding 80% more relevant document references for production-grade RAG model.
- Deploy RESTful API endpoint for user query feedback in parallel collaboration with cross-functional teams against time-critical deadline, allowing for immediate customer feedback necessary for creating gold-standard datasets.
- Perform experimental process of fine-tuning, testing, and evaluating multiple transformer-based sentence embedding models with custom tokenizers to balance keyword relevance and semantic similarity for improved document reranking.
- Develop proof-of-concept few-shot LLM prompting framework to translate natural language queries into executable SQL, integrating regex-based syntax validation, enhancing table-based document retrieval and question answering.
- Augment AI Code Assistant tools to create a novel LLM-powered SAS to Python IT chain management code translation tool. Build and integrate ReAct agent for SAS code analysis and equivalent Python architecture generation.
- Utilize AWS CloudWatch and develop multi-purpose scripts to parse and analyze on-going trial logs, enabling proactive troubleshooting and creating consistently responsive data to customer requests, leading to higher user satisfaction.

Robotics Using Bayesian Inference Lab - UC San Diego

San Diego, CA

Machine Learning Research Consultant (Published)

09/2024 – Present

- Developed analysis pipeline and co-authored IEEE ICDL 2025 publication on explore-exploit dynamics during rat-robot social interactions, characterizing homebase-derived proxemic features for pro-social behavior.
- Led technical engineering of network synchronized multi-modal Embedded-System aimed to host AI-powered classroom.

Staff Machine Learning Research Engineer (Published)

01/2021 – 09/2024

- Spearheaded implementation of Reinforcement Learning algorithms in PyTorch for robotic PiRat with ROS utilizing extensive neural and positional data from 2500+ rat experiments to mimic and predict multi-agent interactions.
- Developed signal processing applications for live and post analysis, leveraging advanced filtering techniques to enhance identification of stress-inducing behaviors from neural and physiological signals (HRV, PPG) of project subjects.
- Built real-time Homebase Detection Algorithm with convolution feature extractor for novel experiment apparatus.

IBM (International Business Machines)

San Jose, CA

Automation and Testing Developer Intern

06/2023 – 09/2023

- Scripted in Ruby to automate API endpoint testing for entire IBM Aspera on Cloud Workflows App designed to automate workflows of global content transfer and exchange across on-premises and multi-cloud environments.
- Developed and documented comprehensive automated regression framework with 780+ API tests covering all endpoints.
- Identified 15+ critical issues and bugs on live production server, communicated with developers to determine ideal expected behavior, and raised issues in project management tools to expertly report relevant failure details.

HIGHLIGHTED PROJECTS

VLM-Guided Hierarchical RL in Habitat-Lab (WPI RL Course, 2024) – Hierarchical PPO/SAC for Fetch in Habitat-Lab with a VLM high-level planner for navigation/manipulation from language; [details](#).

Bio-Influenced Regulation Framework (RUBI Lab, 2023–2024) – Simulated rat-robot regulation using RL/control ideas on 2500+ neural + pose trials to study multi-agent behavior; [details](#).

Rocket.Build: LLM-Powered Documentation Search (Rocket Software, 2024) – RAG prototype using OpenSearch + reranking to improve internal doc/ticket relevance for support workflows; [details](#).

Rocket AI Enablement Lecture Series (Rocket Software, 2024) – Internal LLM/GenAI lectures + notebooks on prompting, evaluation, and Python API usage, adopted by engineering teams; [details](#).

TECHNICAL SKILLS

Languages: Python, Ruby, Java, C++, SQL, Unix; **Agile:** Atlassian (Jira, Confluence, Bitbucket), Git(hub), M365 Suite
Skills: Policy-Optimization Algorithms, NLPs, Data Visualization, API Development, Neural Networks, Data Structures, ML Classification/Regression, Markov Decision Processes, Kubernetes Deployment, AWS Suite Supervision, Signal Processing

EDUCATION

University of California – San Diego

09/2020 – 12/2023

Bachelor of Science, Machine Learning and Neural Computation

Major GPA: 3.91/4.0

Minor: Computer Science; **Program of Concentration:** Philosophy

Awards: President & Recruitment Advisor – Phi Sigma Pi National Honor Fraternity; Provost Honors (Warren College)