

ROHIT SONKER

+1-412-284-8493 | rsonker@andrew.cmu.edu | [LinkedIn](#) | [Google Scholar](#)

My research focuses on **LLM agentic decision making for sequential long horizon tasks**, spanning bayesian optimization and reinforcement learning and multi-modal learning. Prior to my PhD, I spent 4+ years in industry developing and deploying production LLM systems, including RAG pipelines and agents used by enterprises in clinical trial optimization and legal tech.

EDUCATION

Carnegie Mellon University – PhD in Robotics : 4.0/4.0 (Jan 2024 – Jun 2027)
Research Interests – AI for Science, Decision making with LLMs, Multi-modal Learning, Reinforcement Learning

Indian Institute of Technology (IIT) Kanpur (Jul 2014 – Jun 2019)
Dual Degree (Bachelors + Masters) in Mechanical Engineering : 9.3/10

WORK EXPERIENCE

Researcher (Jan 2024 –Present)
Auton Lab, Carnegie Mellon University (Advisor – Prof. Jeff Schneider)

- Designed **LLM-based Bayesian Optimization** framework, utilizing LLM priors and language feedback for adaptive decision-making
- Developed **adaptive offline RL algorithm** from **fusion world model**, latent context conditioned policies adapt to changes in plasma dynamics. Experimentally validated at the DIII-D Tokamak Fusion Facility.
- Developed **multi-timescale bayesian optimization** algorithm for sequential decision-making under sparse data and expensive evaluations, tested on nuclear fusion tokamak improving stability by 117%
- Built **multimodal language-timeseries model** using **Q-Former** adapter and **LoRA** finetuning for cross modal reasoning and question-ans
- Trained a **diffusion** model for **text conditioned image generation**, systematically comparing UNet and transformer backbones for scaling behavior, data efficiency, and generation quality

Tech Lead Manager (Machine Learning) (Aug 2023 – Jan 2024)
CS DISCO – Legal Tech Company

- **Managed cross-functional teams** to design LLM agents for **conversations, summarization, and document tagging** to improve information retrieval from legal data, led to first paid customers for DISCO AI services
- Improved **RAG knowledge retrieval chatbot** with **conversational memory** and **pronoun coreference resolution** for long context, accurate multi-turn retrieval over legal documents, leading to 18% improvement in accuracy and task completion

Senior Data Scientist (Jul 2019 – Jul 2023)
PricewaterhouseCoopers (PwC) US Advisory – Pharma and Life Sciences Division

- Led development of **Large Language Model (LLM)** system to clinical trial documents from trial parameters, compared **prompting** GPT3.5 vs **fine tuning** T5, BERT, GPT models. Deployed in client infra securing projects worth **\$2M+**
- **Architected** an **AWS platform** for healthcare team – created serverless **ETL pipeline (Spark, S3, RDS)**, development env (**Sagemaker**), deployment env (**EC2**) with third party integrations – unified **10+ projects and data sources**, trained over **50+ team members**.
- Developed a **NLP nearest neighbour** algorithm for **entity matching**. Productionized as an end-to-end AWS pipeline with **CI/CD** workflows, unit tests, support for automated scheduled jobs and API service for on-demand processing.

PUBLICATIONS

-
- [1] TG, R Sonker, et al. Large Language Model-based Bayesian Optimization for Tokamak Stabilization, [NeurIPS 2025, ML4PS Workshop](#)
 - [2] R Sonker et al. Offline Reinforcement Learning for Rotation Profile Control in Tokamaks, [L4DC 2026](#)
 - [3] R Sonker et al. Multi-Timescale Dynamics Model Bayesian Optimization for Plasma Stabilization in Tokamaks, [ICML 2025](#)
 - [4] VS, DB, R Sonker, et al. Hidden Parameter Recurrent State Space Models For Changing Dynamics Scenarios, [ICLR , 2022](#)
 - [5] R Sonker et al. Adding Terrain Height to Improve Model Learning for Path Tracking on Uneven Terrain by a Four Wheel Robot, [IEEE Robotics and Automation Letters, Jan. 2021](#)
 - [6] R Sonker et al. Techniques for Medical Concept Identification from Multi-Modal Images, [CEUR Proc & CLEF Conference 2020 \(Oral\)](#)

SKILLS

Programming Languages: C/C++, Python, R, SQL, Spark
ML, AI: PyTorch, Diffusion Models, LLM post-training, Finetuning, RLHF, Vector DBs (FAISS, Pinecone), Langchain
MLOps & Cloud: MLflow, DVC, Docker, Kubernetes, Retrieval Systems,
Data Analytics & Visualization: Pandas, Numpy, Tableau, PowerBI, RShiny, Streamlit, Excel, Powerpoint