ROHIT SONKER

+1-412-284-8493| rohitsonker96@gmail.com | LinkedIn

WORK EXPERIENCE

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

- Developed **Bayesian optimization** algorithm to stabilize plasma confinement, deployed at DIII-D National Fusion Facility, achieved a **117% improvement** in instability avoidance
- Designed a **pipeline** for **data processing, training recurrent neural networks** to model plasma and **reinforcement learning** for plasma control. Optimized hyperparameters with Optuna, **deployed policy** at real tokamak facility (results under submission)
- Developing a **multi-modal language time-series model** using modality alignment alongside foundational time-series models and LLMs

Senior Machine Learning Scientist *(Aug 2023 – Jan 2024)*
CS DISCO – Legal Tech Company

- **Managed 3 member team** to design LLM based features like **conversations, summarization,** and **document tagging** to improve information retrieval from legal data, leading to first paid customers for DISCO AI services
- Added **conversational memory** and **pronoun co-reference resolution** to **retrieval augmented generation (RAG) chatbot**
- Conducted **A/B tests** to validate enhancements, led to a 20% reduction in erroneous responses and improved task completion rate

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

- Led a team to develop an application to generate to clinical trial text from input parameters using **Large Language Models (LLMs)**, compared **prompting** vs **fine tuning** strategies, deployed in AWS securing projects worth $2M+ with multiple clients
- **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**
- Created a **clinical trail design system** using **predictive models**, **simulations**, **LLM agents**, project used across multiple big pharma clients, yielding $500K+ in total revenue
- Developed a **nearest neighbour entity-matching** algorithm as an end-to-end AWS pipeline with **CI/CD** workflows, unit testing, **Step Functions**, **SageMaker** jobs for scheduled tasks and created **API service** for on-demand processing
- Devised a **ML model monitoring system** on AWS to detect drift through automated statistical testing, applied to 10+ live models
- Created a **nearest neighbour** algorithm to match medical concepts in **radiographic images** achieving SOTA results, gave oral presentation at CEUR conference

EDUCATION

Carnegie Mellon University – School Of Computer Science *(Jan 2024 – Nov 2025)*
Master of Science by Research – Robotics : 4.0/4.0
Course work: Reinforcement Learning, Language Models, Computer Vision, Generative AI, Deep Learning

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

SKILLS

Programming Languages: C/C++, Python, R, SQL, Spark
Machine Learning & AI: PyTorch, Keras, Scikit-Learn, XGBoost, FAISS, Pinecone, Langchain, Deep Learning (RNN, CNN, Transformers)
Data Analytics & Visualization: Pandas, Numpy, Tableau, PowerBI, RShiny, Streamlit, Excel, Powerpoint
MLOps & Cloud: MLflow, DVC, Docker, Kubernetes, AWS, GCP, GitHub Actions, Linux, FastAPI, Git

PUBLICATIONS

- **R. Sonker** et al. "Multi-Timescale Dynamics Model Bayesian Optimization for Plasma Stabilization in Tokamaks" *International Conference for Machine Learning (ICML) 2025*
- V Shaj, D Büchler, **R Sonker**, P Becker, G Neumann "Hidden Parameter Recurrent State Space Models For Changing Dynamics Scenarios", *International Conference for Learning Representations (ICLR) , 2022*
- **R. Sonker** et al. "Adding Terrain Height to Improve Model Learning for Path Tracking on Uneven Terrain by a Four Wheel Robot," in *IEEE Robotics and Automation Letters, Jan. 2021*
- **Rohit Sonker** et al. "Techniques for Medical Concept Identification from Multi-Modal Images", in *CEUR Workshop Proceedings and CLEF 2020 Conference Greece 2020 (Oral)*