Dr. Lakshay Arora

(613)-413-0202 | lakshayarora2701@gmail.com | https://www.linkedin.com/in/lakshay-arora-satellite- sentient/
| http://lakshay-arora.com/ | **Toronto, Ontario**

Professional Summary:

- Ph.D. with 5+ years of experience applying AI, and machine learning to real-world challenges multiple domains. My work spans deep learning, generative AI, and optimization under uncertainty—translating research into impactful, data-driven solutions. I enjoy building intelligent systems that are both innovative and practical.

- Expert in Generative AI, specializing in fine-tuning Large Language Models (LLMs) and architecting advanced Retrieval-Augmented Generation (RAG) and agentic RAG solutions. Highly proficient in building autonomous, multi-agent workflows using modern frameworks like LangChain, LangGraph, and CrewAI to solve complex, real-world problems.

- Pioneer in designing fully autonomous, agentic AI systems using Lang Graph, CrewAI, and Model Context Protocols (MCPs). Excels at engineering multi-agent workflows that perform complex reasoning, planning, and dynamic tool use. Skilled in building robust, self-correcting agents capable of orchestrating APIs, databases, code execution, and external tools (powered by MCP standardization) to solve dynamic, multi-step problems without human intervention

- Technically proficient in Python and its scientific stack (Pandas, NumPy), with deep expertise in deep learning frameworks (PyTorch, TensorFlow, Hugging Face) and developing serverless, event-driven architectures to support high-throughput data ingestion and model serving.

- Skilled in AIOps and cloud-native AI deployment, with a strong track record of orchestrating containerized microservices (Docker, FastAPI) and implementing robust CI/CD pipelines. Experienced in automated model deployment, real-time inference, and full AI/ML lifecycle management on GCP (Vertex AI) and AWS (SageMaker).

Skills Summary:

Programming Language and libraries	Python, Julia, R, Pandas, NumPy
Deep learning /AI Frameworks	PyTorch, TensorFlow, Keras, Hugging Face Transformers
Agentic AI Frameworks	LangChain, LangGraph, CrewAI, AutoGen, Pydantic, Langflow, MCP
RAG Architectures	Retrieval-augmented generation (RAG), Agentic RAG, (CRAG)
LLM Finetuning techniques	PEFT (LoRA, QLoRA), Reinforcement Learning from Human Feedback (RLHF)
Web/API frameworks	FastAPI
Prompt Engineering techniques	Chain of Thought, Few-Shot, Self-Consistency, Reason + Act (React)
GCP	Vertex AI, AI Platform, Model Garden, Cloud Functions, Firestore
AWS	SageMaker, Bedrock, Lambda, DynamoDB, API Gateway, S3
LLM APIs	OpenAI, Anthropic, Gemini, Perplexity, Grok (X), Groq
Databases	MySQL, PostgreSQL, MongoDB, Cassandra
CI/CD , Version Control & Dev Tools	*GitHub Actions*, Git versioning, API development and integration
Data Reporting & Analytics	Power BI
Model Optimization & AI Safety	NVIDIA NeMo Guardrails, Quantization, pruning, distillation
Machine Learning Algorithms	Regression, Classification, Clustering Algorithms using scikit-learn
	Time series forecasting (ARIMA, SARIMA, SARIMAX)
Natural Language Processing (NLP)	NLTK, Text Blob, VADER, Genism, Testacy
GPU / Acceleration	NVIDIA CUDA / cuDNN

Employment History:

ML Engineer   **Nov 2022 – Feb 2026**
Benchsci, Toronto, ON

Job Description:
- **Built robust RAG-based applications:** architected pipelines that chunked and vectorized large-scale domain data using embedding models (e.g., text-embedding-ada-002), optimizing vector database retrieval quality through semantic splitting and hybrid chunking strategies
- **Developed agentic RAG systems via LangGraph orchestration:** engineered intelligent multi-agent workflows capable of dynamic retrieval, reasoning, and tool invocation, leveraging LangGraph to seamlessly integrate retrieval- and agent-based modules.
- **Engineered autonomous workflows with human-in-the-loop using MCP integration:** built multi-agent pipelines that integrate external APIs, databases, and tools via Model Context Protocol (MCP), enabling secure tool discovery, dynamic execution, and human oversight for complex task orchestration

Technical Contributions:
- **Scalable LLM Deployment:** Dockerized and containerized Hugging Face Llama 3.1 405B LLM instruct via FastAPI, PyTorch and Transformers; engineered an asynchronous, 8-bit quantized inference pipeline with KV-cache reuse to cut latency by 60% (20s → 8s per 100 tokens) with GPU acceleration using CUDA for improved performance.

- **LangGraph AI Agent Ecosystem:** Built full-stack, multi-tool agents for code execution, database orchestration, web scraping, and dynamic BI visualizations; implemented custom callbacks and Agentic Langgraph State management with human feedback.

- **CrewAI Multi-Agent Framework:** Engineered an autonomous, agentic workflow with CrewAI and a bespoke task-manager agent; optimized inter-agent messaging to accelerate PoC development cycles from days to hours.

- **LangGraph Orchestrated Agentic RAG-Powered ServiceNow AI Assistant:** Engineered an end-to-end agentic RAG solution in Azure ML Studio with LangChain and LangGraph, Azure Cosmos DB and custom vector schemas using text-embedding-ada-002; integrated ServiceNow REST APIs (MCP Standardization) for dynamic ticketing, incident resolution, and workflow automation with high-precision (metrics BERTScore, BLEU), context-aware enterprise support.

Tech Stacks: LangGraph, CrewAI, LangChain, Hugging Face, PyTorch, Transformers, quantization, KV-cache, MCP (Model Context Protocol), Vector DBs, embeddings, Azure ML, Multi-agent crews, React, human-in-loop

Applied ML Engineer   **Sept 2020 – Dec 2025**
Spacecraft Robotics and Control Lab, Ottawa, Canada
Job Description:

- Engineered predictive guidance models under uncertainty: applied Koopman-based expectation methods to model system variability and uncertain initial conditions, achieving **84% precision improvement** in trajectory planning.
- Optimized large-scale trajectory simulations: architected scalable pipelines that reduced runtime by **70%**, leveraging parallelized Julia/Python backends for high-fidelity uncertainty propagation.
- Deployed ML-optimized guidance algorithms on physical spacecraft proximity operations testbeds (SPOT): bridged the sim-to-real gap by validating autonomous docking trajectories on air-bearing spacecraft platforms.
- Integrated algorithms into multi-platform toolchains: collaborated with researchers and engineers to embed Julia, MATLAB/Simulink, and PyTorch-based ML modules directly into experimental spacecraft guidance software, enabling reproducible real-world trials.

Technical Contributions:

- **Uncertainty-Aware Optimization:** Implemented Koopman-based optimization for robust trajectory selection; outperformed Monte Carlo baselines in accuracy and runtime efficiency across thousands of test cases.

- **Hybrid ML-Physics Simulation:** Combined PyTorch deep learning surrogates with nonlinear spacecraft dynamics, accelerating guidance evaluation while preserving physical interpretability.

- **Hardware-in-the-Loop Validation**: Engineered pipelines to run ML-generated trajectories on SPOT's experimental

granite-bed testbed (10 kg spacecraft platforms), enabling live thruster variability testing.

- **Scalable Experimentation Stack:** Built modular simulation workflows in Julia/Python with automated result logging, uncertainty quantification, and visualization dashboards for reproducibility and cross-team collaboration.

Tech Stack: Julia, Python, PyTorch, MATLAB, Simulink

Data Analyst / Data Scientist   **Sept 2016 – July 2017**
Albatronix Bengaluru, India

Job Description:

- Designed and implemented advanced analytics solutions: developed regression and predictive models to generate actionable business recommendations, optimizing sales incentive planning, production forecasting, and financial sensitivity analysis by customer segment.
- Built scalable data pipelines and ETL processes: automated ingestion, cleaning, and normalization of structured/unstructured datasets using Python (Pandas, NumPy, SciPy) and SQL Server, enabling reproducible model training workflows.
- Architected computer vision–enabled data solutions: recommended and prototyped embedded image processing system designs for multi-camera sensor applications, integrating vision-based insights into broader analytic frameworks.
- Delivered BI insights to stakeholders: built Tableau dashboards and Power BI visualizations to surface KPIs, performance metrics, and forecasting outputs, streamlining decision-making across sales and operations.

Technical Contributions:

- **Predictive Modelling & Statistical Analysis**: Applied multivariate regression, correlation, ANOVA, factor analysis, and hypothesis testing to optimize customer segmentation models and production efficiency forecasts.
- **Data Infrastructure & SQL Optimization:** Maintained complex stored procedures, functions, and queries in SQL Server 2008 R2; improved query performance and reporting pipelines for high-volume transactional data.
- **End-to-End ML Pipeline Development**: Engineered workflows in Python and Django to preprocess, normalize, and evaluate predictive models; boosted model accuracy and reduced manual intervention through automated validation.
- **BI & Visualization Layer**: Designed cross-functional dashboards and reports in Tableau/Power BI with embedded statistical outputs, enabling executives to track real-time sales trends and incentive performance.

Tech Stack: Python, scikit-learn, Pandas, NumPy, SciPy, Matplotlib, Seaborn, SQL Server 2008R2, SQL Profiler, Tableau, Power BI, Django

Publications:
- **Arora,L., Ulrich, S.** *Koopman Expectation-Based Optimization for Offline Spacecraft Rendezvous Guidance Under Parametric Uncertainties,* Journal of Guidance, Control, and Dynamics (JGCD), 2026**.**

- **Arora,L., Ulrich, S.** *Koopman Expectation-Based Guidance for Spacecraft Rendezvous and Proximity Operations under Uncertainties.* 35th AAS/AIAA Space Flight Mechanics Meeting, 2025.

- **Arora,L. Dutta,A** *Reinforcement Learning for Sequential Low-Thrust Orbit Raising Problem.* AIAA SciTech Forum, 2020.

- **Dutta,A.,Arora,L.** *Objective Function Weight Selection for Sequential Low-Thrust Orbit-Raising Optimization Problem.* 29th AAS/AIAA Space Flight Mechanics Meeting, 2019.

Education:

Doctor of Philosophy in Aerospace Engineering   **SEPT 2020 - DEC 2025**
Carleton University, Ottawa, Ontario, Canada

Master of Science in Aerospace Engineering   **AUG 2017 - FEB 2020**
Wichita State University, Wichita, Kansas, USA

Bachelor of Technology in Aeronautical Engineering   **SEPT 2013 – MAY 2017**
Manipal Institute of Technology, Udupi, Karnataka, India