

Rajeev Atla

AI/ML Engineer applying cutting-edge techniques to build secure, scalable, complex systems

US Citizen | [732-209-3995](tel:732-209-3995) | rajeev.atla@gmail.com | github.com/RajeevAtla | linkedin.com/in/rajeev-atla | rajeevatla.com

EDUCATION

Rutgers University - School of Engineering <i>Master of Science in Computer Engineering (Specialization in Machine Learning)</i>	Sep 2025 — Dec 2026 New Brunswick, NJ
Rutgers University - School of Engineering <i>Bachelor of Science (Triple Major) in Computer Engineering, Computer Science, and Data Science</i>	Sep 2021 — May 2025 New Brunswick, NJ

Recipient of the Eleanor and Samuel Sneath Endowed Merit Scholarship for Engineering Students

Coursework: Reinforcement Learning, Multimodal AI, High Performance/Distributed Computing

Coursework: AI, Distributed Deep Learning, Data Science, Statistical Learning, Computer Vision, Information and Network Security

SKILLS

- Programming Languages:** Python, R, SQL, Java, C/C++/CUDA, JavaScript/TypeScript, Rust, Bash
- AI/ML:** NumPy, PyTorch, JAX, TensorFlow, Keras, Pandas, Scikit-Learn, OpenAI API, LangChain/LangGraph, DSPy, RAG, HuggingFace (Transformers, Tokenizers, Datasets, Diffusers), vLLM, pgvector, Pydantic, FastAPI, NLTK, spaCy
- Data Visualization:** Matplotlib, Seaborn, Plotly, Tableau
- Cloud & DevOps:** AWS, Microsoft Azure, OCI (Oracle Cloud Infrastructure), GCP, GitHub Actions, Docker, Kubernetes
- Tools & Databases:** Jupyter Notebooks, Apache Kafka, Git, Linux (Ubuntu), PostgreSQL, MongoDB, Jira, Codex, Claude Code

CERTIFICATIONS

- AWS:** [Certified Cloud Practitioner](#), [Certified Machine Learning Specialist](#), [Certified AI Practitioner](#)
- Oracle (OCI):** [AI Foundations Associate](#), [Generative AI Professional](#), [Data Science Professional](#), [Vector AI Search Professional](#)

WORK EXPERIENCE

AI Engineering Intern Atlait Inc.	May 2024 — Sep 2024 Remote
<ul style="list-style-type: none">Developed a Python-SQL compression script for form data, reducing storage costs by 7% for enterprise clientsAccelerated mean response time by 96 milliseconds by integrating PyTorch inference models into Kafka microservicesCreated a > 1TB RAG system, utilizing A/B testing to evaluate and optimize AI-powered search accuracyOptimized CI/CD pipeline to speed up build times by 13% in an Agile environment, ensuring efficient development cycles	

PROJECTS

dexMCP	https://bit.ly/dexmcp
<ul style="list-style-type: none">Engineered Model Context Protocol (MCP) server exposing 5+ reusable tools and 5+ Pydantic modelsImplemented parameter validation across 20+ typed fields and 100% of tool inputsBuilt asynchronous clients using DSPy and LangChain to auto-discover tools and execute multi-step requests	
DocuMint	https://bit.ly/DocuMint
<ul style="list-style-type: none">Built a 5-agent LangGraph + Gemini API doc-modernizer with Gradio, achieved 90%+ modernization coverage on sample docs, cut manual edit time by 50% with a 4-tab UX, hardened with 8 deterministic pytest cases and network-safe skipsAuthored a modular multi-agent system with structured prompts and severity-prioritized research, lifting modernization accuracy by 35% and trimming LLM API spend by 20%	
SuperconGAN	https://bit.ly/3z7JaqZ
<ul style="list-style-type: none">Built a PyTorch-based GAN to create synthetic superconductivity data of various materials, enhancing generative AI applicationsExtracted and processed 80,000+ dataset entries from the UCI ML Repository using Pandas efficientlyReleased Python package on PyPI, achieving over 80,000 downloads and widespread adoptionAuthored LaTeX paper detailing experimental design and validation methodology for 500,000+ data points	
Cityscape Mobile Tour App (Won 2nd Overall at HackExeter 2021)	https://bit.ly/3OZjJ07
<ul style="list-style-type: none">Wrote controllers and models for MongoDB using Mongoose ODM to store 30+ kB of geographic data in NoSQL schemaBuilt mobile user interface allowing users to search, review, rank, and explore 100+ tours using Flutter/DartConstructed REST API using Express.js and nodemon to increase development velocity by 20% with hot-reloading	