ADITYA RATAN JANNALI

+19167767065 | jannali.a@northeastern.edu | LinkedIn | GitHub

EDUCATION

Northeastern University (GPA: 4.00) Boston, MA
MS in Artificial Intelligence, September 2023 – December 2025
Courses: Machine Learning, Foundations of Artificial Intelligence, Programming Design Paradigms, Algorithms, NLP, MLOps, Masters Project (Kolmogorov Arnold Networks & Reinforcement Learning), AI for HCI (Audit), Stats for CS (Audit)

Vellore Institute of Technology (GPA: 9.11) Chennai, India
Bachelor of Technology in Electronics and Communications Engineering July 2017 – June 2021
Courses: Statistics, Calculus, Linear Algebra, Probability Theory, Data Structures, Machine Learning, Digital Image Processing.

EXPERIENCE

Institute for Experiential AI
Research Assistant *July 2025 – present. Boston, MA*

- Working with Dr. Agata Lapedriza Garcia's lab as on-campus RA to test Vision Language Models (VLMs) Theory of Mind capability.

Data Scientist Co-op, Generative AI *Jan 2025 – Aug 2025. Portland, ME*

- Developed and deployed a serverless QA RAG system on AWS, implementing end-to-end testing on multi-format document data for a Reference Management Service. The system achieved a retrieval latency of ~100ms.
- Worked on legal document analyses and classification for a Legal Consultancy firm. Project involved building custom ML Classification models to identify best model for the task.

Amazon
Application Engineer III *August 2021 – August 2023. Chennai, India*

- Directed the ticket-reduction initiative spanning multiple DevOps-supported SDE teams, achieving a 35% drop in support tickets and reclaiming 20+ engineering hours each month through targeted automation.
- Designed and built end-to-end ETLM pipelines and monitoring plugins using Python, PostgreSQL, and AWS, delivering dashboards that provided leadership with real-time insights into pipeline health, usage, and Redshift cost metrics—enabling data-driven load assessment and resource optimization across teams.

Software Support Engineer II Intern *January 2021 – July 2021. Chennai, India*

- Designed, developed, and deployed a utility that collects service level raw data and distributes aggregated metrics to relevant service owners and stakeholders. This utility streamlines data management, enhances information accessibility, and facilitates decision making for key personnel.
- Assumed the role of release engineer, taking ownership of production deployment for multiple pipelines while ensuring their smooth operation.

Antpod
System Development Engineer Intern *April 2020 – December 2020. Chennai, India*

- Involved in Research and Development of a proof of concept for an unmanned vehicle in 'Land Stress Identification and Remote Sensing'.
- Prototyped an algorithm using Deep Fully Connected Convolution Network using Keras and TensorFlow to segment images from the dataset and perform classification to identify plant disease.

SKILLS

- **Programming Languages:** Python, Java, C++, SQL, JavaScript, HTML/CSS, Shell/Bash
- **Machine Learning & AI:** PyTorch, TensorFlow, Keras, Scikit-learn, Transformers, Hugging Face, OpenAI Gymnasium, Reinforcement Learning (Q-Learning, DDQN), Deep Learning (CNNs), Transfer Learning, Fine-tuning, Kolmogorov Arnold Networks (KANs)
- **NLP & Vision:** GPT-2, ALBERT, DeBERTa, Qwen, Gemma, Llava, Llama, RAG (Retrieval-Augmented Generation), OpenCV, Medical Image Analysis, Image Segmentation, Document Classification
- **MLOps & Data Engineering:** TensorFlow Data Validation (TFDV), DVC (Data Version Control), Model Versioning, CI/CD Pipelines, NumPy, Pandas, PostgreSQL, ETL/ETLM Pipelines, Data Preprocessing, Monitoring & Automation, Dashboard Development
- **Cloud & DevOps:** AWS (EC2, S3, Lambda, SageMaker, Redshift, DynamoDB, QuickSight, Athena, IAM, KMS, VPC, ECS), Google Cloud Platform (GCP), Serverless Architecture, Git, Release Engineering, End-to-End Testing, Container Orchestration
- **Software Engineering:** MVC Architecture, Design Patterns, Unit/Integration Testing, JSwing, Team Leadership

CERTIFICATIONS

- (CITI) Conflict of Interest, Human Subject Research, Responsible Conduct of Research for Engineers.
- (Coursera) Machine Learning, Deep learning using TF – CNN and NLP, Natural Language Processing – Classification And Vector Spaces, Probabilistic Models, Sequential Models, Digital Image Processing.

PROJECTS

1. **Disease Prediction & Medical RAG system:** [GitHub], Team Lead, Sept 2025 – present
   Leading a team to develop an end-to-end MLOps pipeline for disease prediction from radiological scan images with an integrated RAG system for medical report analysis and diagnosis assistance. Implementing data versioning with DVC, data validation using TFDV, and deploying on Google Cloud Platform.
   Skills: Google Cloud Platform**,** TensorFlow, TFDV, DVC, GCP, Medical Image Analysis, RAG, MLOps, Team Leadership

2. **Research Project: Evaluating KANs for Reinforcement Learning Applications,** Sept 2024 – December 2024
   Implemented a DDQN using KANs under the guidance of Dr. Raj Venkat to evaluate performance on Atari game environments in OpenAI Gymnasium. This project is a study of comparative analysis between KAN-based and traditional MLP-based reinforcement learning architectures and training algorithms. Investigating KANs' potential as an alternative to Multilayer Perceptron models in Deep RL frameworks.
   Skills: PyTorch, Reinforcement Learning, Neural Network Design, Experimental Analysis

3. **Language Model Interpretability** [GitHub]**,** Oct 2024 – Dec 2024
   Fine-tuned **GPT-2 Small** on a **6-label emotion dataset** for sequence classification. Masked attention heads and replaced key tokens to analyze **attention and token importance and interpretability of predictions**.
   Skills: Transformers, NLP, Explainable AI, Model Evaluation, Attention Analysis

4. **RL Tic-tac-toe** [GitHub], May 2024 – June 2024
   Implemented Q-Learning to train two agents with different reward schemes to play against each other. Each valid board configuration is a state in this environment which is modelled as a 9-digit ternary sequence eliminating the need to precompute the observation space.
   Skills: Reinforcement Learning, Python, NumPy, OpenAI Gym, Environment Design

5. **LLM generated vs Human text classification** [GitHub], April 2024 – April 2024
   Finetuned and evaluated two transformer models, ALBERT and DeBERTa-XS on DAIGT Proper Train dataset to detect LLM generated essays. Achieved an accuracy of 96.57% for ALBERT, and 99.43% for DeBERTa.
   Skills: NLP, Transfer Learning, Transformer Fine-Tuning, Data Preprocessing, Model Evaluation

6. **Image Processor** [GitHub], November 2023 – December 2023
   Designed and MVC, developed and wrote unit and integration tests for an image processing application that performs various image manipulations on custom images. The controller uses command design pattern to listen and process user interaction, and JSwing for the view.
   Skills: Java, Software Design Patterns, GUI Development, Testing & QA