

ADITYA RATAN JANNALI

+19167767065 | jannali.a@northeastern.edu | [LinkedIn](#) | [GitHub](#)

EDUCATION

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

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

EXPERIENCE

Northeastern University July 2025 – present
On campus Research Assistant Boston, MA

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

Institute of Experiential AI Jan 2025 – Aug 2025
Data Scientist Co-op, Generative AI Portland, ME

- Developed and deployed a serverless QA RAG (Retrieval-Augmented Generation) 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 August 2021 – August 2023
Application Engineer III Chennai, India

- Owned and Delivered automation process to reduce incoming tickets by 35%, saving 20hrs/month of engineering time.
- Took ownership of designing and building end-to-end plugins and ETLM pipelines utilizing Python, PostgreSQL, and AWS to generate insightful reports of pipeline health metrics, usage patterns, and cost utilization of Redshift clusters.
- Contributed to dashboard setup, providing leadership with real-time access to actionable insights during weekly and monthly business reviews, enabling data-driven decisions on load assessment and resource optimization. The dashboards feature easy onboarding and scaling to other orgs/ teams.

Software Support Engineer II Intern January 2021 – July 2021

- 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 April 2020 – December 2020
System Development Engineer Intern 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

Languages: Python, Java, C++, SQL, HTML, CSS, JS, shell.

Libraries: OpenCV, sklearn, pytorch, TensorFlow, gym, transformers, Numpy, Pandas.

Frameworks: git, huggingface, AWS – Redshift, IAM, S3, Secret Manager, IAM, EC2, ECS, VPC, KMS Sagemaker, Secret Manager, QuickSight, Athena, DynamoDB.

CERTIFICATIONS

- (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. **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.
2. **Language Model Interpretability** [\[GitHub\]](#) Oct 2024 – Dec 2024
A study where the attention head(s) of a finetuned GPT-2 small Sequence Classification transformer are masked completely to understand which region is responsible for driving the prediction of a certain class. The project also expands into replacing important tokens in the document before passing it to the model's layers.
3. **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.
4. **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.
5. **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.