
Work Experience - 4+ years

Toyota Motors North America(Contract)

Plano- Texas, USA

Senior AI Engineer - Ascentt

July 2024 - Present

- Leading one GenAI product for Toyota Motors with a team of 5, incorporating AI-research, backend, and infra.
- Developed RAG product for Toyota's manufacturing, processed 5TB+ unstructured data for real-time query resolution which resulted in 25% increase in fault detection at Battery manufacturing plant.
- Developed core framework to streamline data ingestion for all AI based applications across Ascentt and Toyota.
- Developed knowledge graph for the 5TB+ motor manufacturing data to feed into RAG as a source of truth.

Prefab Innovations, Inc

California, USA

Machine Learning Software Engineer

Jan 2024 - July-2024

- Led the fine-tuning of multi-modality Layout-LLM models using PeFT methods in PyTorch following mixed precision.
- Developed ETL pipelines for embedding data into VectorDB, ensuring data integrity and accessibility for ML tasks.
- Utilized Python, PyTorch, Scikit-learn, MLFlow, and Airflow to build and deploy ML pipelines on AWS Sagemaker.
- Conducted exploratory data analysis (EDA) to identify key insights and inform model development.

Paytm

Delhi, India

Software Engineer

Sept 2019 - Aug 2021

- Designed and implemented Java-based REST APIs for runtime distributed caching reloads on production servers.
- Resolved critical production issues related to payment gateway response times, system overloads, and downtimes.
- Managed production releases on AWS EC2 using CI/CD pipelines.
- Collaborated with cross-functional teams (QA, Product) and assisted in A/B and user acceptance testing (UAT).

Zillious Solution Pvt Ltd

Delhi, India

Technical Associate

June 2018 - Sept 2019

- Developed Java-Spring-MVC based REST APIs with atomic SQL transactions for B2B tools.
- Collaborated on a microservice for a deep learning OCR parsing model, improving operational efficiency and accuracy.
- Worked with 0.5 million image data-sets to develop, implement data-driven solutions for expense management tool.

Zillious Solution Pvt Ltd

Delhi, India

Software Engineer Internship

June 2017 - Aug 2017

- Developed Java-Spring-MVC based Biometric attendance analysis REST APIs for ERP tool used in Zillious Solutions.
- Developed UI/UX for the ERP tool and a detailed leave tracking analysis dashboard for the tool.

Education

California State University - Fresno, CA USA

Master of Science (M.S.)

Aug 2021 - Dec 2023

Major in Computer Science

Sant Longowal Institute of Engineering & Technology, India

Bachelor of Engineering(B.E.)

Aug 2014 - May 2018

Major in Computer Science

Projects | [Github](#)

Thesis - M.S |Computer Science - CSU Fresno | ([Thesis-Link](#))

Topic: Blindfolded Spider-Man Optimization: A Single-point Metaheuristic Suitable for Continuous and Discrete Space

- Developed a new metaheuristic algorithm using memoization concept in higher dimensions.
- Researched on unbounded knapsack discrete problems and applied metaheuristics to achieve the most optimum result following several higher dimensional implementations ranging from 2-Dimension to 20-Dimensions.
- Data analysis on statistical significance tests like shapiro-wills, kruskal-wallis, converse tests with 0.05 p value.

Non-Convex Optimization Opensource Python Package | [pypi.org](#) | [GitHub](#)

- Python package **published on pypi.org** providing a collection of non-convex functions for testing and benchmarking optimization algorithms, written in Python and **install using pip** command

Technical Skills

Programming Languages & Frameworks: Python, Java, JavaScript, Spring-MVC, Flask, Data Structures**Software Machine Learning Engineering:** PyTorch, TensorFlow, Scikit-learn, Scipy, OpenCV, HuggingFace-Transformer, LLMs Fine Tuning, ETL ML Pipeline AWS-Bedrock, AWS Sagemaker, Opensearch, DynamoDB.**Development Practices & Tools:** Object-Oriented Programming, SDLC, Agile Development, TDD, CI/CD, JIRA, Docker, MLFlow, Apache Airflow, Kafka, RDBMS, VectorDB, Terraform