

Mohnish Sai Prasad

+1-831-869-4373 | mohnish2401@gmail.com | [linkedin.com/in/mohnish7](https://www.linkedin.com/in/mohnish7) | github.com/mohnish-7

Summary

Recently graduated student seeking full-time ML or SWE roles (available May 2026). Skilled in AI agents, RAG, cloud (AWS), Linux systems, Apache Spark, API design, and building scalable, automated systems.

Skills

- **Programming Languages & Technologies:** Python, JavaScript, C++, HTML, CSS, SQL
- **Machine Learning & AI:** LangGraph, LangChain, LlamaIndex, Spark MLlib, TensorFlow, PyTorch, Scikit-Learn, MLflow
- **Cloud & Distributed:** Databricks, Apache Spark, PySpark, AWS (Lambda, EC2, S3, SQS, DynamoDB), Amazon Connect
- **Frameworks & Tools:** FastAPI, Flask, Vue.js, MySQL, PostgreSQL, ChromaDB, Docker, jQuery
- **Systems & Platforms:** RHEL, IBM Spectrum LSF, Linux

Education

University of California, Santa Cruz Sep 2023 - Mar 2026
Master of Science, Computer Science and Engineering
GPA: 4.0 / 4.0
Coursework: Machine Learning, Generative AI, Data Structures and Algorithms, Foundations of Deep Learning, Crowdsourcing

Madras Institute of Technology, Anna University Aug 2019 - July 2023
Bachelor of Engineering, Electronics and Communication Engineering
GPA: 9.15 / 10.0
Coursework: Data Structures and Algorithms, Software Engineering, AI and ML, Internet of Things

Work Experience

University of California Santa Cruz Sept 2023 - Mar 2026
Graduate Teaching Assistant Santa Cruz, CA

- Served as a TA for CSE 16 Applied Discrete Mathematics for Fall 23, Winter 24, Spring 24, Spring 25, Fall 25, and Winter 26; CSE 101 Data Structures and Algorithms for Fall 24; and CSE 30 Abstractions in Python for Winter 25
- Led weekly discussion sections and held office hours providing detailed explanations about the concepts covered in the course
- Provided personalized support to enhance students' learning
- Consistently achieved positive feedback from students with regards to my teaching and assistance with the course every quarter

MaxLinear, Inc. July 2025 - Sept 2025
AI Intern Carlsbad, CA

- Automated technical spec lookups by developing a RAG-based IEEE 802.11 query tool using Copilot Studio and LlamaIndex.
- Extracted complex semiconductor data by building a Dockerized LlamaParse POC on LSF nodes for table and figure analysis.
- Accelerated standard cell analysis by engineering an AI agent to parse Liberty (.lib) files and automate property comparisons.
- Optimized engineering workflows by implementing context-based query routing for tool documentation via Copilot Studio Topics.
- Achieved a **4x** performance gain by re-architecting a Go-based redirection app to Python/Flask on RHEL.
- **Tech Stack:** Python, LlamaIndex, LlamaParse, Copilot Studio, IBM Spectrum LSF, Linux, Docker, Flask, Apache Server

Sandeza Inc Aug 2024 - Dec 2024
ML Intern Alpharetta, GA

- Reduced human workload by **50%** by developing an LLM agent for Amazon Connect using AWS Kinesis and OpenAI Whisper.
- Expected to boost daily customer resolutions by **15%** through a GPT-4o-mini conversation summarizer deployed via FastAPI.
- Improved customer satisfaction scores and minimized wait times by engineering a GPT-4o-mini powered IVR system.
- Enabled the deployment of advanced AI features by refactoring legacy JavaScript API layers into Python.
- **Tech Stack:** Python, JavaScript, FastAPI, AWS S3, DynamoDB, AWS Lambda, AWS Kinesis, GPT-4o

Projects

- **BioAgent: Autonomous Metabolic Research Assistant** Mar 2026 - Apr 2026
 - Automated RNA-seq analysis and protein profiling by engineering a LangGraph agent integrating a Python REPL and the Human Protein Atlas (HPA) API as custom LangChain tools.
 - Minimized API latency by implementing multithreading to handle batch tool requests and enhanced literature search precision by implementing PubMed [TIAB] query logic.
 - Deployed a FastAPI backend with a frontend UI designed using Lovable.
 - **Tech Stack:** Python, LangGraph, LangChain, Multi-threading, FastAPI, Pandas, Biopython, OpenAI, HPA API

- **GPU Model Classification using PySpark and MLlib** Feb 2026 - Mar 2026
 - Built a PySpark ML pipeline for multi-class classification using MLflow for experiment tracking; applied correlation analysis and class weighting to handle heavy label imbalance.
 - Achieved a 0.92 F1-score with Random Forest, a 20% lift over Logistic Regression benchmarks.
 - **Tech Stack:** PySpark (MLlib), MLflow, Python, SQL, Databricks.
- **News Agent: AI Agent to Search the Latest News** Feb 2025 - Mar 2025
 - Developed an AI news agent using LangChain and Anthropic Claude to search and summarize real-time news.
 - Integrated tools like AskNewsSearch, YahooFinanceNewsTool, WikidataQueryRun, and DuckDuckGoSearchRun for multi-source retrieval.
 - Designed an agentic workflow for dynamic tool use and context-aware responses.
 - **Tech Stack:** Python, LangChain, LangChain Community Tools, Claude Sonnet 3.5, Web Search APIs.
- **Enhanced Educational Video Lecture Summarization** Jan 2024 - Mar 2024
 - Explored two summary enhancement methods: a RAG system with LlamaIndex + Llama 3, and a few-shot prompt-based GPT-3.5 model using additional reading material.
 - The RAG system produced limited improvements, while the few-shot approach generated more informative summaries.
 - A user survey showed 64% of participants preferred the enriched summaries (with reading material) over those without it.
 - **Tech Stack:** RAG, ChromaDB, Llama 3, GPT-3.5, LlamaIndex, and Python.
- **Forest Fire and Human Detection from Images for Alert Systems** Sept 2022 - Jan 2023
 - Vision Transformers were fine-tuned using PyTorch to detect wildfires from videos with 98.9% accuracy.
 - Used YOLO algorithm to detect the presence of humans inside a wildfire and sent their coordinates to the fire department.
 - **Tech Stack:** Vision Transformers, PyTorch, YOLO, and Python.

Publications

- **A Survey on Human-AI Teaming with Large Pre-Trained Models**, *arXiv*, 2024. ([Link](#)) *Explores the impact of large pre-trained models in enhancing Human-AI collaboration.*
- **Secured IoT Framework For Soil Moisture Detection**, *IEEE Xplore*, 2023. ([Link](#)) *Developed a secure IoT framework for real-time soil monitoring using Raspberry Pi, MQTT, and InfluxDB, enabling data-driven agricultural decision-making.*
- **Vision Transformer based Forest Fire Detection for Smart Alert Systems**, *IEEE Xplore*, 2022. ([Link](#)) *Developed a ViT-based intelligent forest fire detection system with automated alerts and human detection, achieving 98% accuracy.*
- **CISKGG: a collective intelligence strategy for heterogeneous knowledge graph generation for criminology as a domain of choice**, *IEEE Xplore*, 2022. ([Link](#)) *Deep learning-based framework for automated knowledge graph generation in criminology, integrating ontology alignment, topic modeling, and entity enrichment to achieve 97% precision.*
- **DKMI: Diversification of Web Image Search Using Knowledge Centric Machine Intelligence**, *Springer*, 2022. ([Link](#)) *Introduces the DKMI framework for improved web image recommendation using knowledge-based techniques.*
- **Deep Learning based Forest Fire Detection and Alert System**, *IEEE Xplore*, 2022. ([Link](#)) *Proposes a CNN-based model for fire detection with real-time alerts via Raspberry Pi; contributed a custom dataset of 2500 images per class.*

Additional

- Work Authorization: Authorized to work in the US with **OPT + STEM OPT**
- Start Date: May 18th, 2026 for FTE
- Relocation: Open to relocate to anywhere in the US