Venkata Sai Saravan Pathapati

410-869-5848 saravan.pathapati@gmail.com linkedin.com/in/saravan-pathapati saravan-portfolio

Professional Summary

AI Engineer specializing in designing and delivering AI solutions using foundation models and large language models. Skilled in building high-performance backend architectures, integrating machine learning applications, and automating workflows using Python and FastAPI. Experienced in deploying optimized Dockerized microservices and collaborating with cross-functional teams to drive client success. Committed to continuous learning and innovation in AI technologies.

Technical Skills

Languages: `Python, Rust, C#, JavaScript, BASH, PowerShell, HTML5, CSS3`

AI/ML Frameworks: `TensorFlow, PyTorch, Keras, Hugging Face Transformers, OpenAI API, Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn, NLTK, SpaCy, Fastai, PyTorch Lightning, XGBoost, OpenCV, Dask`

Web Technologies: `FastAPI, Flask, RESTful APIs, OpenAPI 3.0, Swagger, GraphQL, React`

Cloud/DevOps: `Docker, Kubernetes, GCP, AWS, Azure, Git, Vertex AI, Chroma`

Databases: `SQL, NoSQL, MySQL`

Other: `Unix, Unity, State Machine, Trigger Systems`

Experience

UMBC DREAM Lab — Jan 2024 – Present

Guest Researcher — *Baltimore, USA*

- **Developed Proof of Concepts (POCs)** for an intelligent tutoring system using **foundation models** and **large language models (LLMs)** to validate and demonstrate AI solutions in adaptive learning.
- Leveraged **OpenAI embeddings**, **Hugging Face Transformers**, and **PyTorch Lightning** for semantic search and adaptive content delivery, enhancing the feasibility and effectiveness of AI solutions.
- Applied advanced **Natural Language Processing (NLP)** and **text analytics** techniques using **SpaCy** and **NLTK**, fine-tuning **LLMs** with **Fastai** for specific educational content, ensuring accurate, context-aware, and real-time tutoring responses that delight users.
- Implemented efficient model training and deployment pipelines using **PyTorch Lightning** and **Dask** for large-scale data processing.
- Optimized solution performance by monitoring and fine-tuning models, identifying opportunities to enhance efficiency, accuracy, and speed through algorithmic improvements and infrastructure optimization.
- Documented solution architectures, design decisions, and implementation details, contributing to internal knowledge sharing and mentoring new team members.
- Stayed up to date with the latest trends in AI, foundation models, and LLMs, evaluating emerging technologies like **OpenAI API** to assess their potential impact on solution design.

  Technologies Used: *Python, Rust, OpenAI Embeddings, Hugging Face Transformers, PyTorch Lightning, Fastai, NLP, SpaCy, NLTK, RAG, Vertex AI, Chroma, Dask, SQL, Machine Learning*

UNIIAN, DataNue — Sep 2024 – Dec 2024

Backend Developer — *Texas, USA*

- Developed **Python connectors** for a data migration automation tool, enabling seamless integration and data exchange between various systems through **RESTful APIs** in **FastAPI**.
- Engineered **multi-agent AI models** using **TensorFlow** and **PyTorch** to enhance automation capabilities and improve the efficiency of data migration processes.
- Implemented **multi-agent systems** to coordinate tasks and workflows within the data migration platform.
- Integrated **machine learning applications** into backend connectors to optimize data processing, utilizing **Dask** for handling large-scale data.
- Optimized **database queries** and backend code using **NumPy** and **Pandas** to boost API performance and ensure system scalability.
- Utilized **OpenAPI 3.0** and **Swagger** for API design and documentation, streamlining integration processes.
- Applied **Test-Driven Development (TDD)** to ensure code quality and robustness.
- Collaborated with cross-functional teams in a startup environment, adapting to both backend development and machine learning tasks as needed.

  Technologies Used: *Python, FastAPI, RESTful APIs, TensorFlow, PyTorch, Dask, NumPy, Pandas, Multi-Agent Systems, Machine Learning, OpenAPI 3.0, Swagger, SQL, Docker, Kubernetes, Git, TDD*

Dazarus Private Limited **Jul 2023 – Dec 2023**
Machine Learning Engineer Trainee *Bangalore, India*

- Developed and deployed **machine learning models** for predictive maintenance on **IoT vending machines**, enhancing operational efficiency.
- Analyzed large-scale **sensor data** using **anomaly detection** and **supervised learning algorithms** like **XGBoost** to mitigate risks.
- Preprocessed and transformed raw data using **Pandas** and **NumPy**, extracting key features for **model training**.
- Visualized data patterns using **Matplotlib** and **Seaborn** for insightful analysis.
- Collaborated with software engineers and project managers to deploy solutions, ensuring adherence to best practices.
- Utilized **predictive analytics** to identify critical faults, ensuring proactive maintenance.
  Technologies Used: *Python, TensorFlow, Scikit-learn, XGBoost, Pandas, NumPy, Matplotlib, Seaborn, Machine Learning, IoT, Anomaly Detection*

Education

University Of Maryland **Sep 2022 – Jun 2024**
Masters in Computer Science - GPA: 3.3 *Baltimore County, MD*

Vellore Institute Of Technology **Jun 2018 – May 2022**
Bachelor's of Technology in Computer Science - GPA: 8.42 *Vellore, IN*

Publications

Artificial Intelligence in Game Programming **Nov 2022**
ICT4SD 2022 International Conference, Springer Publications *DOI: https://doi.org/10.1007/978-981-19-5221-0_60*

- Showcased the application of AI-driven NPCs and interactive environments, highlighting the impact of imperfect NPC interactions on enhancing game realism and immersion.
- Designed and implemented **Character State Machines**, **Animation Transitions**, **Trigger Systems**, and **Audio Mixers** for dynamic, immersive gaming experiences.

Certifications

Google Cloud Professional Machine Learning Engineer *Issued: Oct 2024 | Expires: Oct 2026*
View Credly Badge