Vigneswar Sundaramurthy

908-297-2557 | Philadelphia, PA | vigneswar.drexel@gmail.com | Linkedin | Portfolio

EDUCATION

Drexel University — **Philadelphia, PA**
Master of Science in Computer Engineering — *September 2022 - June 2024*
Honors: Magna Cum Laude

Post Graduation program in Machine learning and Artificial intelligence — **The University of Texas Austin, Online**
Masters in Artificial Intelligence and Machine learning — *March 2020 - September 2021*

Bachelor of Technologyin Electronics and Communications Engineering — **Kerala Technical University**
Bachelors in Electronics and Communications Engineering — *June 2015 - December 2019*

SKILLS

- **Frameworks/ Platforms:** Amazon Web Services(AWS), Microsoft Azure, Google Cloud Platform (GCP), Vertex AI, Amazon Sagemaker, LangChain, LlamaIndex, TensorFlow, PyTorch, Caffe, JAX, Keras, OpenCV, Scikit-learn, Windows, Linux, Jupyter, JupyterLab, Visual Studio Code, Databricks, Neo4j, Docker, Kubernetes, Tesseract, Pandas, Numpy, Scikit, scipy, xgboost
- **Specializations:** Artificial Intelligence, Deep Learning, Statistical Modelling, Data Analytics, Natural Language Processing, Recommendation Systems, Graph Databases, Vector Databases(Pinecone, Elasticsearch), Web Scraping, APIs, Prompt Engineering, Reinforcement Learning, Computer Vision
- **Programming Languages:** Python, C, C++, MATLAB, PySpark, SQL, R, JavaScript, HTML, CSS, Node.js, Go, HTML, CSS, React

CERTIFICATIONS

- Microsoft Certified AI Engineer Associate — *September 2024*
  Validates expertise in using Microsoft Azure AI to build and manage AI solutions that leverage machine learning and cognitive services.
- Databricks Accredited Generative AI Fundamentals — *October 2024*
  Validates expertise in using Databricks to build and manage Gen AI solutions that leverage machine learning and cognitive services.
- PCEP – Certified Entry-Level Python Programmer — *November 2024*
  Certified Python Programmer accredited by the Python Institute, demonstrating expertise in Python programming and a commitment to professional development.

WORK EXPERIENCE

Prompt Owl — **Atlanta, GA**
AI Engineer — *Present*

- Developed a Retrieval-Augmented Generation (RAG) system using a **multi-query approach**, improving response accuracy by 20% through fusing multiple responses from Large Language Models (LLMs).
- Created a **scalable document** repository that processes and makes over 100,000 documents queryable in real time, ensuring seamless and precise information retrieval.
- Implemented a **web scraping** feature that allows users to input URLs, scrape data from 50+ websites, and query the database, enriching information with over 10,000 new data points.
- Improved data accessibility by integrating **LLMs with web-scraped data**, increasing query resolution time by 30% and enhancing user engagement.
- Deployed the application using **AWS Lambda**, enabling serverless architecture for efficient, scalable, and cost-effective operations.
- Deployed the application in a Docker container on a private server running at Private Server Deployment , enabling robust, scalable, and cost-effective operations.

Tools: **LangGraph**, **LangChain**, LangSmith, **Neo4j, CrewAI, KnowlegeGraph, LlamaIndex,** Python, BeautifulSoup, GPT-4, TensorFlow, Neo4j, Elasticsearch, Docker, Google Cloud Platform (GCP), Streamlit(for deployment)

Zoetis — **Parsippany, NJ**
Data Science Intern — *June 2023 - September 2023*

- Engineered and delivered a **Virtual sensor for Zoetis, enabling precise measurement of Lactate Levels in Bio Reactors** project completed ahead of schedule, **within a 12-week timeframe**, ensuring timely product launch and market advantage
- Implemented machine learning techniques, achieving a **99% predictive accuracy** using Random Forests, SVM, CNN, and Gradient Boosting.These optimizations significantly enhanced real-time cell health monitoring
- Synchronized and analyzed **unstructured data** from 20+ sources, resulting in a 15% increase in cost-effectiveness, saving approximately $25,000. Managed insights from over 1,000,000 data points

Tools: **Databricks**, Data Analysis, Statistical Modelling, Feature Engineering, **Git (Version Control)**, Data Integration, SQL Server Management Studio, Azure Blob Storage, Azure SQL Database, Shiny, Posit

Tata Consultancy Services — **Bangalore, India**
Systems Engineer — *January 2020 - September 2022*

- Streamlined defect detection & analysis through a machine learning model, reducing issues 25% in half a year as per SCRUM Agile Project Management
- Diagnosed and optimized In-vehicle networks (CAN/LIN/Ethernet) for rigorous physical layer tests for seamless operation and compliance

- Integrated machine learning into CAN testing slashed test time by 30%, boosted coverage by 15%, and reduced manual errors by 20%, saving $50,000 annually

Tools : **TensorFlow**, MATLAB, **Python** ,**SQL** , **PySpark (data processing)**, C

PROJECTS

Patent NLP **Philadelphia, PA**
Project Link: https://github.com/vigneswar96/Patent_NLP *November 2023 - July 2024*

- Designed and implemented a sophisticated information retrieval system using semantic search techniques, improving patent similarity matching by 40% compared to traditional keyword-based methods.
- **Designed and implemented a multi-input transformer**-based model leveraging BERT embeddings for NLP tasks, **integrating multiple text sources** through concatenation and pooling. Achieved efficient feature extraction using **GlobalAveragePooling** and optimized performance with dense layers and dropout, managing 918,529 trainable parameters for improved prediction accuracy.
- Utilized transfer learning techniques to fine-tune **pre-trained** language models on a corpus of over **1 million patent documents**, achieving a 95% accuracy in patent classification tasks.

Tools: BERT, **PyTorch**, Hugging Face Transformers, spaCy, NLTK, Elasticsearch, **Flask**, Docker, Google Cloud Platform(GCP)

Reinforcement Learning: Training Atari Game Players with TensorFlow Agents **Philadelphia, PA**
Project Link: https://github.com/vigneswar96/appliedML/tree/main/Tensorflow_agents/Notebooks *April 2023 - May 2023*

- Optimized Atari games using TensorFlow agents, achieving peak performance with the highest score
- Executed diverse game levels utilizing **TF-Agent**, encompassing skill levels from Poor to Intermediate and High, and systematically preserved the outcomes

Tools: Q-learning, SARSA, Deep Q Networks, Policy Gradient Methods, **OpenAI Gym**, RLlib, Google Cloud Platform (GCP)

Integrating Vision and Generative Models **Philadelphia, PA**
Project Link: https://github.com/vigneswar96/appliedML/tree/main/Generative_AI/Notebooks *March 2023 - April 2023*

- Instituted a Real-time image processing with **Stable diffusion** for seamless subject replacement, ensuring precision and efficiency
- Developed **custom datasets** by curating and annotating thousands of images for real-time object detection and subject replacement, ensuring high-quality training data for YOLOv5 and YOLOv8 models.
- Implemented **data augmentation techniques** such as rotation, scaling, and flipping to increase dataset diversity and improve model generalization, resulting in more robust image recognition performance.
- Utilized segmentation tools like **Segment Anything Model (SAM)** to automate object labeling, significantly reducing manual effort and enhancing dataset accuracy by 30%.

Tools: YOLOv5, YOLOv8, **S**egment **A**nything **M**odel (SAM), PyTorch, OpenCV

AI Powered content retrieval system **Philadelphia, PA**
Project Link: https://github.com/vigneswar96/AI-Powered-Content-Retrieval- *June 2024 - August 2024*

- Implemented a AI Powered Content Retrieval system. This could convert all the documents loaded into a **vector database** and then could be queried for analysis.
- Developed a cutting-edge **Retrieval-Augmented Generation (RAG)** system (using Langchain) that transforms unstructured documents into high-dimensional vectors, enabling efficient, high-precision information retrieval across large document collections.

Tools: Langchain, Ollama, Llama 3.3, Groq, Linux, Langsmith (Tracing)

Research Publications

- Published IEEE Journal Paper titled: "Machine Learning Approach for Clustering of Countries to identify the best Strategies to combat Covid 19". Presented in IEEE Toronto 2021 International Conference
  Project Link: https://ieeexplore.ieee.org/abstract/document/9422621 *September 2020 - November 2020*
- Published "Detecting Key Soccer match events to create highlights using Computer Vision" in **arxiv**
  Project Link: https://arxiv.org/abs/2204.02573 *June 2021 - September 2021*