PRANAV SHRIDHAR KOWADKAR

pk.kowadkar@gmail.com | github.com/p-kowadkar | linkedin.com/in/pkowadkar | p-kowadkar.github.io

EXPERIENCE

NEW JERSEY INSTITUTE OF TECHNOLOGY, Newark, NJ | AI Engineer **Mar 2025 – Present**

- **[PyTorch, HuggingFace, LLMs]** Fine-tuned open-source LLMs (Llama 4 Scout, Llama 3.2 3B) for neuroimaging classification, achieving 12% higher accuracy than general-purpose LLMs on fMRI datasets.
- **[Python, Neural Networks, Tokenization]** Designed custom tokenization pipelines converting 90k+ fMRI time-series voxel signals into sequential embeddings compatible with transformer architectures.
- **[Knowledge Distillation, GPTQ/AWQ]** Implemented model distillation from GPT-3.5 to smaller variants, reducing inference costs by 40% while maintaining 95% of baseline performance.
- **[Attention Mechanisms, Clinical ML]** Integrated specialized attention mechanisms prioritizing clinically relevant brain regions, improving model interpretability for medical professionals.
- **[Quantization, Edge Computing]** Quantized fine-tuned models using GPTQ/AWQ techniques, reducing memory footprint by 60% for deployment on edge devices.

BAYER, Remote, NJ | Data Engineer **Sept 2023 – Dec 2023**

- **[Databricks, Spark, Snowflake]** Architected end-to-end data pipelines processing healthcare data to generate marketing personas, ensuring data accuracy and reliability.
- **[PowerBI, DAX, Analytics]** Developed dashboards visualizing 121+ consumer behavior metrics, improving campaign efficiency with insights that increased ROI by 22%.
- **[Azure Data Factory, ETL, Python]** Enhanced ETL processes for real-time point-of-sale data, optimizing advertising strategy and boosting campaign performance by 18%.
- **[Data Quality, Automation]** Created automated monitoring systems ensuring data integrity throughout processing pipelines while maintaining compliance with data privacy regulations.

DASSAULT SYSTÈMES, India | R&D Software Engineer **Apr 2020 - July 2022**

- **[C/C++, Memory Profiling]** Fixed 10+ critical memory issues in CATIA applications, enhancing system stability and reducing crash rates by 20% for enterprise clients.
- **[Linux, Cross-platform]** Led migration of CATIA applications to Linux environment by resolving 20+ platform-specific errors, cutting migration downtime by 40%.
- **[Python, Test Automation]** Spearheaded automation of CADAM testing, developing a comprehensive framework that reduced execution time by 87% (from 15 to 2 minutes).
- **[Python, Pandas, Matplotlib]** Automated team performance analytics with scripts processing weekly contribution data, eliminating 3+ hours/week manual work and enabling data-driven decisions.
- **[Apache, Dashboard Development]** Engineered self-service reporting solution transforming contributor achievements into interactive dashboards, adopted by 8+ departments.

VAANDU LLC, Newark, NJ | Programmer Analyst **Nov 2024 - Feb 2025**

- **[Oracle, PL/SQL, Unix]** Strengthened data pipelines using advanced stored procedures and scripting, boosting query speeds by 20% and ensuring smoother batch processes.
- **[Prometheus, Monitoring]** Implemented metrics collection for Oracle database monitoring with custom exporters capturing key performance indicators.
- **[Python, Code Quality]** Enhanced codebase through targeted refactoring, cutting review times by 15% and improving deployment stability.
- **[Grafana, Observability]** Developed comprehensive dashboards visualizing real-time pipeline health metrics, establishing alerting thresholds that reduced system outages by 45%.
- **[AWS Lambda, CloudWatch]** Built an observability framework providing end-to-end visibility into data flows, enabling rapid troubleshooting of performance bottlenecks.

PROJECTS

LLM-based Brain Region-to-Function Mapping

- **[PyTorch, Neuroimaging, Transformers]** Developed a specialized transformer architecture mapping fMRI signals to functional brain regions, achieving 89% classification accuracy.
- **[Transfer Learning, Fine-tuning]** Applied transfer learning from pre-trained vision transformers to neuroimaging domain, cutting training time by 40% compared to training from scratch.
- **[Data Augmentation, Python]** Implemented custom augmentation techniques for limited fMRI samples, effectively doubling the training dataset size without additional scans.
- **[Model Interpretability, Visualization]** Created interactive visualizations highlighting attention patterns across brain regions, providing neuroscientists with explainable AI insights.

Gravitational Wave Detection

- **[TensorFlow, Keras, Signal Processing]** Engineered deep learning model for gravitational wave detection in LIGO/VIRGO data, achieving 86.8% accuracy on test data.
- **[PySpark, Distributed Computing]** Processed 76GB+ of astronomical time-series data, overcoming memory limitations through distributed computing clusters.
- **[EfficientNetB7, RNN]** Implemented hybrid neural architecture combining convolutional and recurrent networks for precise classification of gravitational wave signals.
- **[Ensemble Methods, Plotly]** Created ensemble approach improving detection sensitivity by 22% for low signal-to-noise ratio events, with interactive visualization tools.

AI-powered Portfolio Chatbot

- **[LangChain, OpenAI, RAG]** Developed Netflix-inspired recommendation chatbot for portfolio site using retrieval-augmented generation with custom embeddings.
- **[FastAPI, React, Docker]** Built full-stack application with containerized microservices architecture, ensuring consistent development and deployment environments.
- **[Vector Database, Embeddings]** Implemented efficient similarity search using pre-computed embeddings, delivering sub-200ms response times for complex queries.
- **[GitHub Actions, CI/CD]** Created automated testing and deployment pipeline reducing deployment time by 70% and eliminating common integration errors.

SKILLS

Programming: Python, R, SQL, PL/SQL, C/C++, C#, Bash, JavaScript
Machine Learning / AI: PyTorch, TensorFlow, Scikit-learn, Hugging Face, LLMs, Transformers, Keras, Fine-tuning, Computer Vision, NLP, Prompt Engineering
Data Tools: Pandas, NumPy, Spark, Airflow, Databricks, Snowflake, ETL/ELT, SQL Server, Oracle, Elasticsearch, Kafka, Redis
Cloud & DevOps: AWS (S3, Lambda, SageMaker, EMR, RDS, CloudWatch), Azure (Data Factory, Databricks), GCP, Docker, CI/CD, GitHub Actions
Web & APIs: FastAPI, Flask, React, API Gateway, REST, GraphQL, Plotly, Dash, PowerBI, Tableau, Grafana, Prometheus

EDUCATION

Master of Science in Data Science | GPA: 3.6 Sept 2022 - Dec 2023
New Jersey Institute of Technology, Newark, NJ
Coursework: Big Data, Cloud Computing, Machine Learning, Deep Learning, Applied Statistics, Data Mining