Keshav Arri

647-227-1538 | arri@uwindsor.ca | Ontario, Canada | LinkedIn | Portfolio | GitHub

TECHNICAL SKILLS

Languages	Python, SQL, C/C++, Java, Bash/Shell Scripting, HTML, CSS
Databases	MySQL, PostgreSQL, MongoDB, Pinecone, Weaviate , Milvus, Snowflake, Redis
AI/ML	PyTorch, TensorFlow, XGBoost, Hugging Face, LangChain, LangGraph, OpenCV, RAG
Cloud/Dev	AWS, Azure, GCP, Docker, FastAPI, React, Git, CI/CD Pipelines, RESTful APIs
Related skills	ETL Pipelines, Fine-tuning LLMs, Prompt Engineering, Data Modelling, A/B Testing, NER, OCR, Socket Programming, Data Structures & Algorithms

EDUCATION

Master of Applied Computing — Specialization: Artificial Intelligence Jan 2025 - Present

University of Windsor Ontario, Canada

- Final Semester Requires a 4 to 8 months Internship Starting Jan 2026

Bachelor of Technology in Computer Science and Engineering Sep 2020 - Sep 2024

Guru Nanak Dev Engineering College — SGPA 8.13/10 Punjab, India

EXPERIENCE

AI ML Engineer
Feb 2024 - Feb 2025

Slideoo *Bangalore, India*

- Engineered multimodal pipeline converting text, documents, websites and YouTube URLs into PowerPoint presentations, achieving 83%-time reduction to 30 seconds vs competitor's 3-min turnaround, helping secure seed funding.
- Scaled system to handle 10x traffic growth by implementing distributed architecture using OpenAI GPT and Claude model with multi-threading across Azure Web Apps, AWS Lambda and Bedrock, achieving 99% uptime.
- Improved LLM response accuracy by 25% and reduced API costs by 40% through optimized prompt engineering, few-shot learning and intelligent caching strategy using Redis and MongoDB for frequently accessed templates.
- Led a 4-member intern team to develop a prompt-based research report generation that automated 70% of manual research tasks using chain-of-thought reasoning.

Data Scientist & NLP Researcher
Jul 2023 - Jan 2024

Sabudh Foundation *Punjab, India*

- Automated 60% of manual document processing by developing intelligent OCR pipeline using Detectron2, spaCy NLP, and custom entity recognition, processing 10,000+ documents monthly with 94% accuracy.
- Achieved 91%-line segmentation accuracy on damaged ancient manuscripts successfully digitizing 10,000+ historical documents for preservation project.
- Collaborated with 8-member research team to document and present solutions to government officials, demonstrating 3x faster processing than existing solutions.

Projects

DataDialect AI | *Talk to any database using natural language* Apr 2025 – Jul 2025

- Achieved 89% query accuracy on complex multi-table joins by implementing NL2SQL system using few-shot learning and semantic similarity matching.
- Reduced database query time by 65% for non-technical users, eliminating need for SQL knowledge and saving 10+ hours weekly for business analysts.
- Integrated with 4 database types (MySQL, PostgreSQL, MongoDB, Vector Databases)

Stanford STORM API Wrapper | *Knowledge curation system* Jan 2025 – Mar 2025

- Wrapped Stanford STORM research pipeline into REST API using multi-LLM orchestration (GPT-4/Claude) via LangChain, handling 500+ daily requests with streaming responses.
- Reduced deployment time by 75% through Docker containerization and FastAPI async endpoints, enabling researchers to generate comprehensive reports in 5 minutes vs 20 minutes.

CODING COMPETITIONS

CS Demo Day Participant Apr 2025

University of Windsor Windsor, ON

- Showcased EduMetrics project to 200+ industry professionals and academic faculty.

WinHacks 2025 Jan 2025

University of Windsor Windsor, ON

- Prototyped AI-powered solution in 48-hour hackathon emphasizing rapid prototyping skills.

CERTIFICATIONS

- Dataiku Advance Designer, Developer Certificate, ML Practitioner Feb 2025
- Google Data Analytics Professional Certificate – Coursera Dec 2024
- Selenium Essential Training - LinkedIn Learning Oct 2024
- Microsoft Power BI Desktop for Business Intelligence – Udemy Mar 2023
- TCS iON Career Edge - Soft Skills Development Feb 2023