ARSHVIR

Bengaluru, India | +91 7528953747 | arshvir882@gmail.com | LinkedIn (linkedin.com/in/arshvir-995a91203) | GitHub (github.com/avarshvir) | Portfolio

EDUCATION

Sardar Beant Singh State University | Gurdaspur, India
B. Tech in Computer Science and Engineering | 2021 - 2025
CGPA: 8.12

SKILLS

Programming Languages: Python, C, C++, Web Languages (HTML, CSS)
AI/Machine Learning: Scikit-Learn, TensorFlow, PyTorch, Transformers, Diffusion Models, NLP, YOLO, OpenCV
Databases & Knowledge Retrieval: MySQL, PostgreSQL, MongoDB, ChromaDB (Vector Databases), Neo4j, RAG
Software Engineering: DSA, Object Oriented Programming, Git, GitHub, GitHub Actions, Linux, JIRA, Software Development
MLOps & Deployment : FastAPI, vLLM, Ollama, OpenAI API, Streamlit, Gradio, Google Colab, Render

EXPERIENCE

Netscout | Bengaluru, India
Software Engineering Intern | April 2026 – Present

- **Developed an AI-powered DOM Locator Service with FastAPI backend** that accepts natural-language queries and retrieves the XPath of specific web elements using **IBM Granite Embedding** model that uses embedding-based semantic matching. Designed the system to handle complex, webpages across different frameworks with up to **32K-token context windows.**
- **Built a Machine Learning based image-diff and visual validation system for automated testing** using OpenCV and Tess4J to compare baseline and test PNGs, detect UI changes, and highlight additions, unchanged areas, and text/position changes, achieving **89% accuracy.**
- **Contributing to the backend of an intelligent automated testing harness** for automated, test execution, analysis, and validation of software.

GEE 7 Printek Private Limited | Ludhiana, India
Engineering Intern | June 2024 – August 2024

- Developed a **domain-specific RAG chatbot** using QSS-3200 digital printing machine manuals, enabling engineers and technicians to retrieve information for troubleshooting, servicing, maintenance, and error resolution.
- Developed a **C++ bulk image-to-PDF conversion utility** using **libharu**, implementing lossless image handling to efficiently generate PDF documents from large batches of images.
- Built a **JavaScript-based web application** to generate daily machine-specific access codes using the company's date-based 2260 calculation logic, with an HTML/CSS interface.

PROJECTS

NanoOCR - Optical Character Recognition [GitHub Link] | [PyPI Link] | [Website]

- Built and trained a lightweight, 4.8 million parameter Optical Character Recognition (OCR) model from scratch on Google Colab using a CRNN (CNN + BiLSTM) architecture.
- Created a memory efficient data streaming pipeline to train the model on 2.32 million images, loading files one by one to completely bypass hardware and memory limits without downloading the entire dataset at once.
- Executed a 3-phase training strategy, starting with a massive synthetic dataset and then fine-tuning the model on real-world scene text to adapt it to messy, blurry, real-life images.
- Achieved 79.20% accuracy on unseen MJSynth images and 78.37% on IIIT5k, maintaining a low Character Error Rate (CER) of 8.56% and 3.40 CTC loss.
- Published the final 18.3 MB model as a public PyPI package (pip install nano-ocr), making it instantly deployable for edge devices, embedded systems, and low-power hardware.

Visio AI (Machine Learning and Data Science Platform) [Demo Link] | [GitHub Link]

- Developed an end-to-end ML and data science software that enables users to upload dataset, automated data preprocessing & cleaning, EDA, and visualization for supervised and unsupervised learning workflows.
- Implemented training pipelines for regression, classification, and clustering algorithms, with intelligent algorithm suggestions based on target variables.
- Enabled model export functionality allowing users to download trained models for reuse in external applications.
- Integrated OpenRouter API for visual LLM to find hidden patterns of images, extending platform capabilities beyond tabular data.
- Built a no-code, user-friendly interface using Streamlit, enabling non-technical users to perform end to end ML workflows.

NGIBS – Next Gen. Intelligent Browsing System [GitHub Link] | [Website Link]

- Developed NGIBS (Next Generation Intelligent Browsing System), a privacy-focused AI research and browsing system using Python, PyQt6, LangChain, Ollama, API's and open-source LLMs for local inference.
- Designed four distinct intelligence modes: Quick Search for pretrained LLM knowledge, Live Search for real-time web retrieval, Deep Research for multi-step information gathering and synthesis, and Personalized Mode for long-term memory-based conversations.
- Engineered Live Search and Deep Research pipelines integrating DuckDuckGo API, Wikipedia, BeautifulSoup, URL content extraction, and LLM reasoning to retrieve, process, and synthesize current information.
- Implemented short-term session memory and persistent long-term semantic memory, while supporting integration of user-selected open-source LLMs through a modular local inference architecture.

CERTIFICATIONS

- AI/ML Certified Course, Coder Roots | Mohali, India (Jan 2025 - May 2025)