Pearl Natalia

pearlnnatalia@gmail.com | pearlnatalia.com | linkedin.com/in/pearlnatalia | github.com/pearl-natalia

Education

University of Waterloo Sept 2023 - Apr 2028
Bachelor of Software Engineering (Honours), Co-op *Waterloo, ON*

- **Involvement:** WATonomous (Self-Driving Car Design Team), Women in Engineering (Sponsorship Director)
- **Coursework:** Data Structures and OOP (C, C++), Database Management (SQL), Compiler Design (C++)

Technical Skills

Languages: Python, Java, SQL (Postgres), C, C#, C++, JavaScript, TypeScript, HTML/CSS
Tools: AWS, MongoDB, Snowflake, Databricks, Neo4j, Docker, Google Cloud, Modal, Jenkins, Trino, Terraform, Amplitude
Frameworks: React, Angular, Flask, Node.js, Next.js, Django, Bootstrap, Spring Boot, ROS
ML Libraries: TensorFlow, PyTorch, scikit-learn, NumPy, Pandas, OpenCV, Matplotlib, Keras, MLflow

Experience

Machine Learning Intern Sept 2025 – Present
The Globe and Mail *Toronto, ON*

- Built a deep-learning retention-offer model with Amplitude data and target encoding, projecting **$300k/year** in churn savings
- Engineered a Slackbot (Databricks, Anthropic, AWS) to surface and send journalists personalized briefings, **saving 3h/week**
- Created a **call agent** (Salesforce, Anthropic, Amplitude) to generate tailored offer justifications for sales teams on churn calls
- Led end-to-end projects, collaborating with Legal, Sales, and Editorial teams to design solutions that supported key objectives

Software Engineering Intern Jan 2025 – Apr 2025
PointClickCare *Toronto, ON*

- Built a **Grafana** dashboard with **Docker** to visualize **Trino** pipeline metrics, generating **100k datapoints** for **stress testing**
- Developed a Spring Boot tool to convert SQL queries to Trino format via syntax trees, enabling **scalable** federated querying
- Collaboratively optimized batch sizes to reduce latency in migrating **millions** of health records across **distributed databases**
- Documented iterative progress in **Confluence** following **Agile methodologies** across all projects

Machine Learning Engineer Sept 2024 – Aug 2025
Mimrr *New York City, NY*

- Developed **agents for enterprise clients** to automate meeting planning/scheduling and legacy code updates using **Neo4j**
- Created a **C# chatbot** and implemented persistent memory using **context-free grammar**, to extract and store user data
- Improved **LLaMA's text generation** by integrating Outlines for more structured and desired chatbot outputs

Web Developer Intern May 2024 – Aug 2024
Generis Global Partners *Toronto, ON*

- Automated data cleansing via a web app (CLIP embeddings, web scraping) hosted on **AWS EC2**, saving **20h/week**
- Rebuilt 15 websites to **increase traffic and SEO by 35%** using **Google Analytics** and marketing team mockups
- Built a Chrome extension to parse emails via **Flask** for new website content, **accelerating development time by 73%**

Projects

Mental Health Assistant | *React, Modal, Pandas, NumPy, Scikit-Learn, OpenAI embeddings, GPT-4 API, SQLite*

- Created a responsive **React web app** with **SQLite** for fast data retrieval, and **hosted ML models** on Modal
- Designed a **RAG pipeline** to generate accurate counselor advice by retrieving similar patient cases from a vector database
- Trained an anxiety prediction model with **Naive Bayes**, **Random Forest classifiers** and **TF-IDF** vectorization
- Built a **semantic search** over **14k** counseling dialogues using embeddings and optimized results via LLM-based preprocessing

AI Wearable for SOS (HackThe6ix) | *React, Flask, Auth0, MongoDB, Gemini Flash, Hume AI, Mapbox*

- Led the hardware integration of a SOS device to capture video and audio recording with a Raspberry Pi and Flask
- Generated **contextual summaries of video frames** via **Gemini Flash and Hume AI APIs**, highlighting critical threats
- Developed a **CNN-based audio classification model** with spectrograms using TensorFlow to classify violent audio clips

AI-Powered Dash Cam | *OpenCV, Numpy, HERE Maps, Pandas, Scikit-Learn, React, SQLite*

- Developed **low-latency geolocation algorithms** to detect vehicle speed/turns, road violations and speed cameras
- Implemented real-time traffic light and road sign detection from live dashcam footage using YOLOv8, Flask and FFmpeg
- Improved **CNN model** accuracy on low-quality frames via relative positioning, color masking and Hough transforms