

Bhagyesh Patel

Toronto ,ON | bhagyeshpatel0703@gmail.com | +1 (416)8380459 | [LinkedIn](#) | [Portfolio](#)

EDUCATION

Honors Bachelor of Computer Science (Data-Analytics)

January 2022 – April 2025

Sheridan College • Oakville • 2025 • 3.55/4

WORK EXPERIENCE

Data Analyst – Red Pico, Mississauga, ON

March 2026 – Present

- Built automated validation rules in **Informatica** across **90%** of critical pipelines in a single quarter, catching data errors early to support downstream **ML modeling** and business reporting.
- Built governance workflows in **Dataiku** to document **model lineage**, track **bias metrics**, and maintain compliance standards across deployed **AI solutions**.
- Optimized **SQL** queries and data transformations for tables with **100k+** rows, cutting runtimes and preparing clean datasets for internal dashboards.

Generative AI Associate – Innodata Canada

August 2025 – March 2026

- Improved LLM accuracy, fluency, and coherence by **25%+** by creating and validating training datasets using **Python, NLP, and human-in-the-loop evaluation**.
- Annotated and reviewed **10K+ text samples** to align AI outputs with human values using data annotation workflows, **prompt engineering**, and **Responsible AI practices**.
- Reduced grammatical and structural errors by **30%** through syntax analysis and **NLP-based text quality checks** AI-generated text.
- Accelerated model iteration cycles by **20%** by collaborating with data scientists and engineers to refine **LLM training** specifications and data recipes, consistently meeting quality KPIs.

AI/ML Developer – Sheridan Centre for Applied AI

September 2024 – August 2025

- Developed an automated pipeline to extract fleet management KPIs from diverse plans using a locally deployed **Meta Llama 3 LLM**.
- Boosted KPI extraction accuracy and response relevance by **60%** through **advanced prompt engineering** and **few-shot learning**, eliminating model retraining.
- Deployed **BERTopic** for automated topic modeling, revealing fleet management categories and streamlining taxonomy creation, improving efficiency by **70%**
- Built a custom fleet scorecard with **NLP** and **ML**, integrating an automated pipeline to generate data-driven KPI weights using Autoencoders.

Machine Learning Data Scientist – Sheridan Centre for Applied AI

May 2024 – August 2024

- Formulated optimized data cleaning scripts for analyzing **100,000+** patient records, leveraging parallel processing frameworks to handle large datasets efficiently, and extracting actionable insights to guide stakeholder presentations.
- Improved model accuracy by **40%** by implementing a **voting technique** combining multiple machine learning models.
- Improved F1score and recall by **52.08%** and **26.6%** respectively in order to gain balanced performance matrix.
- Utilized synthetic data generation methods such as **SMOTE** and Random Oversampling to handle imbalanced datasets, doubling the recall and precision scores.

PROJECTS

FairLens - AI-Powered Hiring Solution– 2nd best capstone project award

September 2024 – April 2025

- Built **FairLens**, an AI-driven hiring platform using Python, Flask, and AWS to enable unbiased, DEI-aligned resume screening.
- Used **LLMs** for candidate-job matching and implemented **PII anonymization** to ensure fair, secure processing.
- Designed a **real-time diversity dashboard** in Python Dash to visualize workforce and recruitment funnel insights.
- Optimized performance and reduced costs by integrating **OpenAI GPT-4o Mini** and efficient AWS-based data handling.

AWS RAG-Chatbot

August 2023

- Built a **document-aware chatbot** using AWS and RAG architecture with **Claude v2 LLM**, achieving over **90%** response relevance in domain-specific queries.
- Developed a scalable document pipeline using **amazon.titan-embed-text-v1**, generating **10,000+** vector embeddings from S3-stored files for high-speed retrieval from vector databases .
- Designed a web interface with **AWS Amplify**, supporting **<500ms query response time** and seamless user interaction across devices.
- Integrated OpenSearch for real-time similarity search, **AWS Bedrock** for LLM responses, and CloudWatch for monitoring, reducing debugging time by **40%**.

EXTRACURRICULAR ACHIEVEMENTS

Peer Mentor– Sheridan College

January 2023 – April 2025

- Promoted Peer Mentor for strong leadership, managing **20** mentors and facilitating sessions for **50+** students weekly.
- Engaged in leadership and mentoring as a co-curricular role while being a student, balancing academic and team responsibilities effectively.

Research Paper Publication — Topic Modeling Enhancement using Summaries Generated by LLM Models- Proc. of International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (**ACDSA 2025**) - August 2025, Antalya-Türkiye

SKILLS

- Python, R, C++, SQL, Pandas, NumPy, Matplotlib, Seaborn, Power BI, Hadoop, Spark, Hive, MongoDB , Informatica, Alation, Dataiku, Scikit-Learn, PyTorch, TensorFlow, CUDA, LLMs ,RAG, OpenAI NLP, BERTopic, LangGraph , Autoencoders, Flask, FastAPI, CI/CD, Python Dash, Power BI (DAX, real-time dashboards), AWS Bedrock, Amazon Titan Embeddings, OpenSearch, CloudWatch, AWS (S3, Amplify, Bedrock, OpenSearch), Google Cloud, GitHub, Bitbucket, Jira, Microsoft 365