

Harshkumar Patel

+1 (437) 421 3950 | Theharsh1@outlook.com | Toronto, Ontario

[Portfolio Website](#) | [LinkedIn Profile](#) | [GitHub Profile](#)

Education & Certifications

- Postgraduate Diploma – **Artificial Intelligence & Machine Learning** | Lambton College, Toronto, ON | Jan 2022 – Sept 2023
- Bachelor of Engineering – **Computer Engineering** | Gujarat Technological University, India | Aug 2016 – Aug 2020
- **Machine Learning Specialization by Stanford**, Coursera
- **Deep Learning Specialization by Stanford**, Coursera

Skills

- **Programming Languages:** SQL, Python, R, JAVA, C++
- **Python Libraries:** TensorFlow, Pandas, OpenCV, Matplotlib, Seaborn, PyTorch, SciPy, Scikit-Learn, Keras, XGBoost, NLTK, Flask, Django, Requests, urllib3, Pillow, Scrapy, BeautifulSoup4, PyAutoGUI, Pywinauto, Selenium
- **Machine Learning & Deep Learning Skills:** Supervised, Unsupervised & Semi-Supervised Algorithms, Data and Quantitative Analysis, Predictive Modelling, Natural Language Processing, Data Visualization, Image processing, Audio Processing, Model Development
- **Cloud & Big Data:** Azure, AWS, Apache Spark, Docker, MLflow
- **Data Visualization Tools:** Tableau, Power BI, Excel
- **Databases:** MySQL, MariaDB, Oracle, SQL Server,
- **Methodologies:** SOLID Principles, Programming Design Patterns, Agile, Test-Driven Development
- **Operating Systems:** Linux, Unix, macOS, Windows
- **Professional Skills:** Data Storytelling & Visualization, Strong Problem-Solving, Critical Thinking, Attention to the details

Experience

Data Scientist: [Dataannotation.tech](#) | Remote | Jan 2024 – Present

- Analyzed and optimized **large model training datasets** to improve **accuracy** and **performance** across multiple applications.
- Performed **data cleaning**, **feature validation**, and **exploratory data analysis** techniques to ensure data precision and anomalies
- Tracked the progress of **multiple AI models** by monitoring the **training** and **testing** accuracy graphs.
- **Improved 20%** efficiency by using appropriate **hyperparameters** and choosing the right **algorithm**.
- Presented **findings and insights** through visuals and contributed to data-driven improvements in annotation workflows.

Research Scientist Intern: Microsoft (Nuance), Montreal | Mar 2023 – Aug 2023

- Spearheaded the implementation of **Speech Completion** for **Microsoft Teams Meetings** and **PowerPoint Presentations**, elevating **real-time speech recognition** by **80%** and integrating **GPT-4** for outcome **prediction** in a **Gradio UI**.
- Utilized **Python**, **Gradio**, **Azure API**, **GPT-4 API**, **Azure Cognitive Speech Service**, **JavaScript**, and **ReactJS** technologies.
- Designed and implemented a **CI/CD pipeline** using **GitHub Actions**, reducing deployment failures by **40%** and shortening the release cycle from **2 days to 1 day**.
- **Awarded Regional Winner at Microsoft Global Hackathon 2023** for developing a **real-time speech summarization system**, improving accessibility and user experience.

Data/Business Analyst: [Welltech Minerals](#), India | May 2020 – Sept 2021

- Collected, cleaned, and analyzed large datasets using **Excel**, **SQL**, and **Python** to identify business KPIs meeting the business goals.
- Boosted the **data processing time by 40%** through the extensive use of **python data manipulation libraries** and **writing optimized SQL queries**, which resulted in a **15% increase** in business growth.
- Improved **requirement clarity by 30%** by developing **Use Cases**, **Process Flows**, and **Sequence Diagrams** using **Rational Rose**, ensuring alignment between technical and business teams.
- Increased **reporting efficiency by 25%** by automating **data extraction and transformation pipelines**, resulting in faster and more reliable business insights.
- Developed **interactive dashboards and reports in Tableau** to visualize KPIs, operational trends, and financial performance, improving business decision-making and stakeholder visibility.

Deep Learning Scientist: [SoftVan](#), India | May 2019 – Apr 2020

- Assisted in training and deploying **multiple deep learning models**, improving classification **accuracy from 70% to 92%**.
- Automated **monitoring scripts** that reduced downtime in image and text **processing pipelines by 25%**.
- Trained and evaluated **CNN models** for pattern detection and recognition, improving **model accuracy by 98%**.
- Participated in code reviews, bug fixing, and optimization, applying **Agile methodology** and collaborating with teams.
- Employed **Flask**, **Django**, **Jinja2**, and **OpenCV** for the backend, and **HTML**, **CSS**, and **JavaScript** for the interactive frontend of the website.