

CHARMI CHOKSHI

+1 (514) 677-8531 | charmichokshi@gmail.com | [LinkedIn](#) | [Personal Website](#) | [GitHub](#)

Skills

Programming Languages: Python, R, SQL, Java, JavaScript, C++

Frameworks & Libraries: PyTorch, TensorFlow, Keras, LangChain, NLTK, HuggingFace, scikit-learn, Gensim, OpenCV, Numpy, Pandas, PySpark, SpaCy

Databases & Tools: GCP, AWS, Vector Database, DataBricks, Spark, PostgreSQL, Tableau, Git

Others: Docker, Kubernetes, Elasticsearch, Flask, Django, Streamlit, ReactJS, UiPath

Experience

May 2022 - November 2022

Vancouver, Canada

Research Engineer Intern at Amazon Web Services

- **Model-based False Annotation Detection:** Implemented Ensemble & Gradient Optimization technique, and conducted Loss Analysis on encoder-only, multi-modal Transformer for misannotation detection.
- **Improved data quality by 3% and model test accuracy by 1%** through re-annotation of flagged samples on 8 massive datasets. Utilized **distributed computing** for handling large-scale datasets.
- Developed end-to-end ML-assisted Table Annotation Tool with visualizations highlighting mistake regions.

January 2019 – July 2021

Ahmedabad, India

Machine Learning Engineer at Logistixian Technologies Pvt Ltd

- **Information Extraction from unstructured documents:** Deployed a generalized one-shot Learning model for extracting 80+ fields from documents, **reduced manual work from 20 to 5 minutes.**
- **Zero-shot Learning:** Observed the annotation challenges faced by the users and proposed & led information retrieval on unseen documents using NLP and CV. **Increased job annotation speed by 3X.**
- **Self-learning based feedback mechanism:** Productionized Federated Learning and Reinforcement Learning inspired model to improvise data extraction and standardization using online learning.
- **Document Group Classification:** Built an optimized and lightweight NLP model for classifying pages into 250+ categories followed by page grouping. **Changes saved 1.5 hours of work per day per user.**

May 2018 – August 2018

Ahmedabad, India

Research Trainee at Indian Space Research Organisation (ISRO)

- **Unsupervised Semantic Segmentation:** Trained an Encoder-decoder model with parallel depthwise separable CNNs on **low-resolution satellite imagery having 20+ channels.**
- Implemented Clustering and Region Growing algorithm with progressive threshold and parallel processing for post-processing and achieved SOTA results on the custom dataset.

Education

September 2021 - August 2023

Montreal, Canada

MSc in Computer Science & ML from University of Montreal & Mila

CGPA: 4.2 / 4.3

- **Electives:** Data Science, Natural Language Understanding (McGill University), ML for Climate Change
- **Teaching Assistant:** Data Structures (IFT 2015/6002), Programming 1 (IFT 1015)

August 2015 - May 2019

Ahmedabad, India

BTech in Information and Communication Tech. from Ahmedabad University

CGPA: 3.6 / 4.3

- **Electives:** Machine Learning, Computer Vision, Data Analytics and Visualization, Psychology
- **Teaching Assistant:** Advanced Data Structures and Algorithms (CSC 310), Design Thinking

Projects

DocBot: A RAG-based Chatbot to get insights from PDFs (RAG, LLM, Generative AI, PDF Q&A)

- Implemented Retrieval-Augmented Generation (RAG) framework to address limitations of LLMs, such as knowledge cut-offs and hallucinations, leveraging LangChain, FAISS, Cohere Model & Embeddings APIs.
- Introduced advanced RAG concepts including **Caching Embeddings**, **Hybrid Vector Search** using FAISS and BM25 Algorithm, and **In-memory Caching** for improved response generation efficiency.
- Deployed the bot using Streamlit having capabilities of generating answers from multiple PDFs.

Predict Pushback Time at Airports (EDA, Data Processing, Feature Engineering)

- Led a team of 3 in a competition hosted by NASA to predict the minutes until pushback of a flight.
- Utilized statistical analysis and did extensive research for developing ML models such as Random Forest and XGBoost on air traffic and weather data. **Ranked 23rd out of 408 participants.**

Clickbait Headline Detector using Advanced NLP Techniques (Pytorch, NLU, Language Generation)

- Built a system for the Webis Clickbait-17 challenge to identify Twitter posts that are clickbait in nature.
- **Pre-trained and Fine-tuned DeBERTa and ELECTRA** models and achieved an accuracy of 87.2%.
- Explored **text generation task by Fine-tuning the T5 architecture** and finding similarities between original and generated headlines for an article. Obtained 81% accuracy on the test set.

Conversational Question Answering System (Pytorch, Transformer, NLP)

- Worked on the CoQA Challenge and developed a large language model that can learn from an input passage and answer an interconnected series of questions.
- Implemented cutting-edge architectures, including **seq2seq**, **Transformer (BERT, RoBERTa)**, and **Graph-based architectures**, and achieved the best F1 score of 82.1 on the validation set.

NHL Goal Prediction (Python, Pandas, Docker, Flask)

- Built a system to predict goal probability given the shot data using Bagging/Boosting, SVM, etc. methods.
- Developed a **data analysis workflow for fetching, cleaning, and transforming data from the NHL API**, demonstrating proficiency in handling real-time data and feature engineering.
- Attained 92% accuracy on the test data, **deployed it using docker and Flask**, and **tracked model history using CometML**.

Achievements

- **Certified TensorFlow Developer**
- **Next Generation Leader of the Year, Women in IT Summit & Awards Series**
- **Excellence Scholarship Awardee** Worth \$5000, Mila
- **Canada's Developer 30 Under 30, WeaveSphere**
- **Young AI Role Model of the Year, Women in AI (North America Runner-up)**
- **Rising Star Awardee, Womxn in Data Science**
- **Google Developers Expert in ML** and **Google Cloud Champion Innovator**
- **Grand-Prize winner in Hackathon** (developed an NLP-based Job Recommender System)

Leadership and Volunteering

- **International Tech-speaker**: **Have spoken at 35+ global conferences reaching 30,000+ attendees.**
 - Have received accolades and positive feedback for exceptional communication skills.
- **ML Mentor & Trainer**
 - Tech Trainer at Public Sector Network, Canada. I deliver multi-week courses covering topics such as Generative AI, Deep Learning, Model Explainability, and Robotic Process Automation.
 - Have mentored around 2000 AI enthusiasts directly
 - Mentor at Google for Startups and Road to GDE, Google Developer
 - Mentor at Women Developer Academy, North America
 - Google DSC – Global Solution Challenge: Guided an Indonesian team of 4 to enable deaf people to communicate using AI and wearable devices who became global winners.
- **Proposal Reviewer**: DL Indabad'23, PyCon'21, PyCon'20, Mission Billion Challenge'20, West Africa