Harpreet Matharoo

harpreet.matharoo@outlook.com | LinkedIn | Github | Website | (289)-776-6747 | Toronto, ON

ABOUT

Seasoned data professional and machine learning engineer with 7 years of experience in developing data management solutions. I recently graduated from Georgia Tech with a Master's in Computer Science with a Machine Learning specialization. I bring a unique blend of skills in Python software development, and deep learning with the PyTorch framework, complemented by hands-on proficiency in model deployment on AWS using Sagemaker, Lambda, EC2, and S3.

TECHNICAL SKILLS

Languages	**:** Python, SQL, C#
Frameworks / Libraries	**:**.HuggingFace, Scikit-Learn, Chalice, PyTorch, Ray, MLFlow, Sagemaker
Infrastructure	**:** Docker, Jira, Git, GitHub Actions, AWS Lambda, AWS EC2, AWS CloudFormation
Machine / Deep Learning	**:** RAG, DPR, Transformers, GPT, CNN, OCR
Certifications	: AWS Cloud Practitioner

EXPERIENCE

ML Alchemy Solutions, Aug 2023 – Present
Machine Learning Engineer
Project 1 - Resume Job Matching

- Building a semantic job matching application for a recruiting agency on **AWS** for an increased response rate from the users.
- Implemented a serverless **ETL** pipeline that routinely scrapes new job posting using **AWS Lambda** and **Eventbridge**, transforms and loads the data in **Opensearch.**
- Tested and compared various retrieval systems such as **Dense Passage Retrieval** and **Retrieval-Augmented Generation**.
- Built **CI/CD** using **GitHub Actions** and **Chalice** to automate testing and deployment of the incremental improvement to the job search program.

Project 2 - Potency Evaluation of Marijuana Products

- Exploratory data analysis of the psychological effects for a marijuana marketing agency to evaluate the potency of various marijuana products.
- Analyzed the brainwave data of the participants with appropriate visualization techniques and Bayesian approach using tools such as **PyMC3, Matplotlib** and **Pandas.**

CAD Microsolutions Inc., Toronto, ON, **Mar 2019 – Present**
Senior Applications Engineer / Software Developer

- Implemented, troubleshot, and enhanced existing product data management solutions with **SQL** backbone serving a customer base in the manufacturing industry. Implemented solutions that improved user productivity by over 41%.
- Created an add-in in **C#** that parses through the data, automatically fixes broken references based on predefined rules, and validates its quality.
- Implemented integration tools for seamless integration between the databases of ERP and PDM systems.
- Developed an add-in in **C#** that automates release package creation post-design approval workflow for a customer and notifies relevant parties upon completion. Also, created an **SQL** report that a user can use to review the package creation task queue.
- Automated engineering drawing template standards integration by developing an event-listening add-in in **C#** that automatically checks and updates the standards when a user opens a legacy drawing.

Upchain Inc., Toronto, ON, May 2018 – Feb 2019
Customer Success Engineer

- Provided technical support, user training, drafted proposals for feature enhancements in **Jira,** and implemented the Upchain product life cycle management system for the early adopters.
- Led product adoption initiatives that played a crucial role in the acquisition by a Fortune 500 company (Autodesk).

CAD Microsolutions Inc., Toronto, ON, **Aug 2016 – May 2018**
Applications Engineer

- Provided technical support and user training for computer-aided engineering solutions. Solved the highest number of customer cases(above 700 cases on average) annually amongst a team of 7 support engineers.

EDUCATION

Georgia Institute of Technology, Atlanta, Georgia, **Dec 2023**
MS in Computer Science, **GPA: 4.0/4.0**

McMaster University, Hamilton, Ontario, **May 2016**
MASc in Mechanical Engineering

Indian Institute of Technology, Delhi, **May 2013**
B.Tech in Mechanical Engineering

Selected Coursework: Natural Language Processing, Deep Learning, Machine Learning, Reinforcement Learning, Bayesian Statistics, Artificial Intelligence, High-Performance Computing

PROJECTS

Translating Handwritten Equations to LaTeX - Web, GitHub

- Implemented a Seq2Seq architecture with attention for decoding handwritten equations. A **CNN** encodes the input image, which is translated by a series of **gated recurrent units**.
- Achieved expression recognition accuracy of 0.52 and word recognition accuracy of 0.82.
- Developed an end-to-end open source OCR web application on **Streamlit**.

Skills: **PyTorch | Streamlit**

Interactive Tutorial for the Transformer Architecture - Web, GitHub

- Built an interactive application to expose the inner workings of the **Transformer** architecture.
- Employed multi-view visualization, stateful components created in **React.js** and **D3.js**.
- Connected the visualization to the inference harness of a **transformer** trained on a neural machine translation task with **PyTorch**. Tested inference hosting on **AWS EC2** and **AWS Lambda**.

Skills: **React | D3 | Javascript | PyTorch | AWS**

Transforming Visual QA Task into Lexical QA Task

- Transformed VQA task into Lexical QA to enable zero-shot transfer to reduce inference cost.
- Modeled the problem as (a) **Extractive QA** and (b) **Generative QA** with captions combined with questions and answers.
- Implemented ML pipeline with **HuggingFace** that generates captions using **Vision Transformer** and **GPT-2** in series.

Skills: **PyTorch | HuggingFace | Transformers**

Frontend Application to Communicate with ChatGPT API - Github

- Built a vanilla chat interface that connects with **ChatGPT API** and lets the user communicate with the agent.
- The front end of the application is created in **React.js**.

Skills: **React | Javascript | OpenAI API**

Google Research Football Environment

- Trained football-playing multi-agent in a simulator, comparing performance against baseline agents trained with independent PPO. Achieved a win rate of 0.41 compared to the base rate of 0.32.
- Implemented central critic for faster learning and demonstrated collaboration between individual agents through reward shaping

Skills: **PyTorch | RLlib | Ray | Gym**

VOLUNTEER EXPERIENCE

Reinforcement Learning Group Co-Lead, Cohere for AI, June 2023 - Present

- Increased community engagement by hosting bi-weekly study group sessions and invited distinguished speakers to talk about RL research.

AWARDS AND HONORS

- **Graduate and Research Fellowship Worth $ 28,000/yr,** McMaster University, **2013-2016**
- **Ranked 603 out of 400,000 Students,** Joint Entrance Exam for admission to Indian Institutes of Technology, **2009**