

# GAURAV MUDBHATKAL

gmudbhat@gmail.com · [Github](#) · [LinkedIn](#)

## SUMMARY OF SKILLS

**Languages and Tools:** Python, SQL, Docker, PostgreSQL, Kubernetes, AWS, Argo, Cypher, Neo4j, Elastic, Spark

**Machine Learning:** PyTorch, HuggingFace, Scikit-Learn, OpenAI, LangChain, Metaflow, LlamaIndex, TensorFlow

**Soft Skills:** Agile, Problem Analysis, JIRA, Confluence, Notion, ServiceNow, Trello

## EXPERIENCE

### **Limbik**

*Data Scientist*

Jun 2022-Present

Kitchener, ON

- Improved [PFI](#) models' performance by 20% by implementing a two-tower model, leveraging OpenAI's CLIP embeddings for multi-modal social media data and PyTorch autoencoders for user representation
- Architected a hierarchical PostgreSQL database for efficient data storage and retrieval, integrated with a Python-based pipeline for monthly survey data ingestion
- Orchestrated MLOps flow for [Decipher](#), leveraging Metaflow to handle model training and monitoring workloads with Docker and Kubernetes
- Simplified data retrieval and analysis by enabling natural language querying capabilities for PostgreSQL database using LlamaIndex and OpenAI embeddings
- Engineered data ingestion pipeline with Cypher and Python to ingest survey data into a Neo4j graph database for optimized storage and retrieval

### **Thomson Reuters**

*Machine Learning Intern*

Sep 2021-Dec 2021

Toronto, ON

- Tuned and evaluated document classification models after conducting surveys across various departments
- Implemented model deployment pipelines, integrating machine learning services within the Azure DevOps framework

### **Manulife**

*Data Engineer Co-op*

Jan 2021-Apr 2021

Waterloo, ON

- Facilitated data warehousing in Azure Data Lake by writing Spark scripts in Azure Databricks
- Coordinated the migration of ~30 repositories to GitHub with ServiceNow, JIRA, and Confluence

### **KPMG**

*Data Engineer Co-op*

Jul 2020-Aug 2020

Waterloo, ON

- Delivered subsidies to clients affected by COVID-19 as part of the Canada Emergency Wage Subsidy program
- Calculated upwards of \$2,000,000 in eligible subsidies by wrangling client payroll data using workflows in Alteryx

### **RBC**

*Data Scientist Co-op*

Sep 2019-Dec 2019

Toronto, ON

- Generated Doc2Vec embeddings for ~10,000 documents after preprocessing & tokenizing text using NLTK
- Achieved 96% accuracy & 82% precision through a neural network-based multi-label text classifier with Keras

## EDUCATION

### **University of Waterloo**

Honors, Bachelor of Applied Sciences

Management Engineering, Computing Option

Waterloo, ON

- Relevant coursework: Software Engineering, Data Science,

## PROJECTS AND EXTRACURRICULARS

### **Conrad School of Entrepreneurship**

May 2020

- Conducted and pitched research on the storage & analysis of digital evidence data by Law Enforcement for a chance to win \$15,000 in funding through the Problem Pitch Competition