Karthik Raja Kalaiselvi Bhaskar

(+1) 647-804-6600 | karthikrajabk@gmail.com | kbhaskar.com | karthikraja95 | kbhaskar

EXPERIENCE

CIBC — Toronto, Canada
Senior AI Scientist | Technical Lead — Nov 2020 – Present

- **Technical Lead** with **full autonomy** over AI/ML initiatives: presented **architectural decisions to 4 Senior Directors**, drove **3 POCs end-to-end to production**, mentored **4 Data Scientists/Engineers**, managed **5 co-ops (2 converted to FT)**, designed ML interview frameworks.
- **Led and architected Multi-Agent AI System** for Anti-Money Laundering investigations, orchestrating **7 specialized LLM agents** (Google Gemini 2.5 Pro/Flash), implementing real-time web grounding and **STR generation**, reducing analyst research time by **70%**. Instrumented with **Langfuse**.
- **Owned end-to-end** design and deployment of **enterprise-scale GenAI Chatbot** using **RAG**, vector indexing, and prompt engineering, implementing **LLM-as-a-Judge evaluation**, now serving **50,000+ employees** enterprise-wide.
- **Spearheaded** cloud migration of **ML infrastructure** to **GCP with NVIDIA GPU clusters**, **owning architecture decisions** for distributed training pipelines, achieving **92% reduction in training time** (3 days to 6 hours).
- **Independently designed and shipped** production **AI Agents** (OpenAI SDK): Talent Acquisition Agent, HR Document Generation Agent (**900+ hours saved annually**), Meeting Summarization Agent (**12 hrs/week saved per member**).
- **Led Conversational AI** initiative with **ASR** and **Neural Agent Assistance**, reducing call-center volume by **25%**. **Published at EMNLP 2022 Industry Track**.
- **Architected Semantic Search Engine** using **Transformer embeddings** with **data and model parallelism**. Deployed on **Docker, Kubernetes**.
- **Owned Strategic Workforce Prediction** system using **Gaussian Process Deep Learning** and **GNNs**, forecasting 3-5 year workforce dynamics with **uncertainty quantification**, reducing HR decision time by **40%**.
- **Led LLM fine-tuning** initiatives (**GPT-3.5, Llama, Falcon**) with **parameter-efficient fine-tuning**, achieving **30% improvement** in query accuracy. Built **DistilGPT-2** for **synthetic data generation**.
- **Owned end-to-end BERT pipeline** for client feedback analysis, reducing **80% manual labor**.

Diversio — Toronto, Canada
Machine Learning Engineer — May 2020 – Aug 2020

- Built text classification pipeline using **Transformers and LSTM** to analyze employee feedback and generate Diversity & Inclusion scores for enterprise clients.
- Designed **Deep Learning Recommendation System** to surface actionable insights, improving organizational decision-making.

TD Bank — Toronto, Canada
Data Scientist — May 2018 – Aug 2019

- Developed **Deep Learning-based threat detection system** (ProjectX) for real-time cyberattack prevention, processing **millions of events daily**.
- Built **Unsupervised Auto-Tagging** and **Automatic Rule Synthesis** using Deep Reinforcement Learning for Octavius security platform, improving defense mechanism accuracy.

Infosys (Apple Retail) — Chennai, India
Software Engineer — May 2016 – Aug 2017

- Engineered backend systems for **Apple's retail platform** serving millions of customers globally, performing root cause analysis and deploying production changes with zero downtime.

TECHNICAL SKILLS

Languages & Core: Python, SQL, C++, Git, Linux/Bash, Distributed Systems, System Design

ML/AI Frameworks: PyTorch, TensorFlow, JAX, Scikit-Learn, HuggingFace Transformers, Ray, MLflow

LLM & GenAI: LangChain, LlamaIndex, OpenAI SDK, Anthropic SDK, RAG, Vector DBs (Pinecone, ChromaDB, FAISS), Multi-Agent Systems

Infrastructure: AWS, GCP, Azure, Docker, Kubernetes, Spark, Databricks, CI/CD, Model Serving (TorchServe, Triton)

Specializations: NLP, Deep Learning, Reinforcement Learning, Graph Neural Networks, Recommender Systems, Privacy-Preserving ML

RESEARCH & PUBLICATIONS

Vector Institute & University Health Network — Toronto, Canada
Student Researcher — Aug 2020 – Jan 2021

- Built **Transformer-based multi-task classification** system for Electronic Health Records using Self-Supervised Learning and Weak Supervision on **MIMIC-III** (40K+ patient records).
- Led Team BeatCovid in **XPrize Pandemic Response Challenge**, building DL models for COVID-19 prediction, ranked **16th of 250 teams** globally.

University of Toronto Toronto, Canada
Machine Learning Researcher Sep 2018 – Jan 2020

- Designed end-to-end **Deep Learning Recommendation System** for Wolseley's e-commerce platform (100K+ products), using Matrix Factorization, Collaborative Filtering, and Neural Networks.
- Achieved **NDCG@10 of 72.4%** and **100% One-Product Hit Ratio**, deployed to production serving thousands of daily users.

Publications:

- *Bringing the State-of-the-Art to Customers: A Neural Agent Assistant Framework for Customer Service Support* – **EMNLP 2022 Industry Track**
- *Implicit Feedback Deep Collaborative Filtering Product Recommendation System* – **arXiv 2020**

SELECTED PROJECTS

Fika - Data Science Tool Open Source
End-to-End ML Automation Platform

- Built Python library automating ML workflows: **data cleaning, feature engineering, model training, and deployment** with MLFlow integration. Generates production-ready FastAPI and Docker code.

Satellite Deforestation Detection
Computer Vision & Cloud Infrastructure

- Trained **ResNet** model achieving **95.6% accuracy** and **92.7% F2 score**. Built **data flywheel** with GCP BigQuery and deployed dashboard with Streamlit.

Collaborative Game-Playing Agents
Multi-Agent Reinforcement Learning

- Implemented **Multi-Agent DDPG** with shared experience replay, **Deep Q-Networks** for navigation, and human-like chess engine using reinforcement learning.

Defense GAN & Adversarial Examples
Adversarial ML & Robustness

- Generated adversarial examples using **FGSM, DeepFool, and Carlini & Wagner** attacks. Built **Defense GAN** for robust classification against adversarial perturbations.

HyperFace - Face Analysis System
Multi-Task Learning & Computer Vision

- Built **Multi-Task CNN** using FRCNN to jointly predict facial landmarks, visibility, pose, and gender with **90.68% accuracy**.

EDUCATION

University of Toronto Toronto, Canada
M.A.Sc, Electrical & Computer Engineering (Machine Learning) – Advisors: Prof. Bo Wang, Prof. Deepa Kundur, Prof. Yuri Lawryshyn, GPA: 3.94/4.0 | Master's Research Fellowship

Anna University Chennai, India
B.Eng, Electronics and Communication Engineering, GPA: 3.74/4.0