# Mustafa Shaikh

mustafanshaikh@gmail.com | linkedIn/mustafa-shaikh-716ba8140 | shaikh58.github.io

Machine learning practitioner focusing on object tracking with experience in computer vision, deep learning and classical ML. I have 5 years of industry experience in data science working closely with business stakeholders. I can dive deep into the details, and I have also lead projects end to end, from identifying a problem and designing a solution, to implementation and delivery.

## EDUCATION

**M.S. Electrical and Computer Engineering**                    San Diego, CA | 2022 - 2024
UNIVERSITY OF CALIFORNIA, SAN DIEGO
 Specialization: Robotics and Intelligent Systems

**B.A.Sc. Engineering Science**                                Toronto, ON | 2013 - 2017
UNIVERSITY OF TORONTO

## WORK EXPERIENCE

**RESEARCH ENGINEER** | SALK INSTITUTE FOR BIOLOGICAL STUDIES              San Diego, CA | 2024 -

**Key Skills:** PyTorch, Lightning, Multi-object tracking, deep learning, encoder/decoder transformers, ResNet, DETr, rotary embeddings, segmentation, Kalman filters, Hungarian matching, video encoding, ffmpeg

**Open source project:** https://dreem.sleap.ai

- Co-developed a transformer-based **multiple object tracker** for microscopy and animal videos, achieving **>98%** tracking accuracy
- Acquired and processed public cell tracking video datasets 50x larger than in-house data
- Created a **pretrained microscopy model** that **improved metrics** by **>40%**, and beats state of the art by **>8%**
- **Optimized inference** code to reduce **GPU memory** and **runtime** by **40%** with only 5% decrease in accuracy
- Developed new **global tracking accuracy** metric to address gaps in existing multi-object tracking metrics
- Developed a **metrics pipeline** to enable fast model evaluation and development

**GRADUATE RESEARCHER** | EXISTENTIAL ROBOTICS LAB, UC SAN DIEGO       San Diego, CA | 2023 - 2024

**Key Skills:** Model predictive control, control barrier functions (CBF), casADi, CVX, Extended Kalman Filter, RRT*, collision avoidance, ROS, Jackal robot, LiDAR, HectorSLAM

**Publication:** *"Control Strategies for Pursuit-Evasion Under Occlusion Using Visibility and Safety Barrier Functions"* accepted at the *IEEE International Conference on Robotics and Automation (ICRA) 2025*
Paper: https://arxiv.org/abs/2411.01321
Project Website: https://existentialrobotics.org/VisibilityControl/

- Implemented a **model predictive controller** with **control barrier** constraints for a robot to track a target
- Implemented **Extended Kalman Filter** to estimate target's position and velocity from camera detections
- Demonstrated **>95% tracking** in **real world experiments** with Jackal robot

**SR. DATA SCIENTIST** | WALMART CANADA                                  Toronto, ON | 2019 - 2022

**Key Skills:** Natural Language Processing - Spacy, Named Entity Recognition, BERT, human-in-the-loop systems, AutoML, PySpark, SQL, MLOps, Pandas, Google Cloud Platform, Airflow

**Project lead - Automated Attribute Assignment** (2021-2022)
**Goal:** Improve search quality for customers by automatically populating product data for 3rd party sellers

- Developed **named entity recognition pipeline** to learn **context-aware** features from product descriptions; led to **>$1MM CAD revenue increase** annually by populating features for over 500,000 items
- Recognized need for **high quality custom annotated data**; pitched, acquired and integrated a **human-in-the-loop** annotation tool (Prodigy) with active labelling
- **Coordinated Jr. Data Scientist**, and guided the implementation of an **asynchronous orchestration layer**
- Worked closely with business stakeholders to guide problem framing, roadmap, execution and production support

Other Projects (Apr. 2019 to Jun. 2021)

- Developed and deployed **hierarchical model factory** to categorize 3rd party vendor items on walmart.ca; increased **categorization rate from 90% to 97%** which increased product views for previously 'unfindable' items
- Lead developer for **grocery substitutions recommendation** engine; **300bps improvement in customer satisfaction**
- Created and maintained **fulfillment centre forecast** to optimize labour; **>90%** accuracy up from 75% led to **$1MM CAD annual labour savings**

**SR. DATA ANALYST** | Shoppers Drug Mart                       Toronto, ON | 2017 - 2019

Key Skills: Data engineering, systems engineering, performance metrics, worfklow improvements

- Developed **ETL pipeline** and a large SQL application that **transforms event triggers** into key pharmacy **workflow metrics** such as customer wait times, rework rates and labour heatmaps, to enable store performance reporting
- Used metrics to create business cases and guided rollout of **$1MM CAD in workflow improvements**

## PROJECTS

**DEEP LEARNING AND SOFTWARE** ⤤                        UC San Diego, 2023 - 2024

Key Skills: PointNet, instrinsics, odometry, depth images, SIFT, point cloud registration, PyTorch, C++, RAII, smart pointers, design patterns, templates, variants, STL, LLM finetuning, workflows, Model Context Protocol

- Created an **LLM workflow** with **finetuned Llama 3.2 1B Instruct** that retrieves research papers from Arxiv based on user chat queries. Integrated with Cursor using **Model Context Protocol**
- Developed **PointNet-based** approach to solve the **point cloud registration** problem, and achieved performance close to **Iterative Closest Point** on basic trajectories in Eden dataset
- (C++) **String library** with underlying buffer manager. **Achieved 25% lower memory usage than C++ std::string** for common string operations: append, replace, insert, erase, search
- (C++) Graph-based **in-memory JSON datastore** with query capability
- (C++) **Compressing archive tool** with add, extract, retrieve capability

## SKILLS

Python, C++, SQL, PyTorch, JAX, Numpy, Pandas, Keras, CVX, CasADi, OpenCV, Open3D, GCP, PySpark, Hadoop, Airflow, ROS, Adobe Illustrator