

# Stephen Bottos

778-966-8141 | bottos.steve@gmail.com | linkedin.com/in/stephen-bottos | github.com/stevebottos

## EXPERIENCE

<b>Lead Machine Learning Engineer</b> — Kibeam, Inc. <i>Oakland, CA (Remote)</i>	Aug. 2023 – Present
<ul style="list-style-type: none"><li>Architected multimodal agentic systems integrating vision, text, and audio (Langchain, MLFlow, Huggingface, OpenAI, Gemini, ElevenLabs) with custom spatial reasoning layers and RAG-enabled reflection patterns, reducing content generation turnaround time from days to under an hour across production workflows.</li><li>Built end-to-end MLOps infrastructure (AWS ECS, Airflow, MLflow, Redshift) supporting 100+ models/week iteration and deployment of 1,000+ production models across tens of thousands of deployed units.</li><li>Enhanced ML data labeling with Vision Transformers for automated soft-labeling, eliminating manual annotation. Optimized lightweight CNNs for ESP32 edge inference through int8 quantization and custom layers.</li><li>Led cross-functional ML strategy, defining technical roadmaps and research priorities to align agentic automation initiatives with product vision across engineering and product teams.</li></ul>	
<b>Senior Machine Learning Engineer</b> — Plainsight, Inc. <i>San Diego, CA</i>	Aug. 2021 – Aug. 2023
<ul style="list-style-type: none"><li>Designed, trained, and deployed custom PyTorch model architectures for multimodal applications (vision, video, text, LiDAR) supporting accounts up to \$10M. Scaled cloud workflows in GCP using VertexAI and custom Docker containers for training and RestAPI deployment.</li><li>Developed internal platform using Vector Databases (Milvus, BigQuery, ClickHouse) for automated train/validation/test set splitting through deduplication and similarity inspection. Extended this to catalog and auto-annotate millions of images, improving model performance by 10-20%.</li></ul>	
<b>Machine Learning Engineer</b> — alwaysAI, Inc. <i>San Diego, CA</i>	June 2020 – Aug. 2021
<ul style="list-style-type: none"><li>Designed and deployed production Computer Vision models (TensorFlow, PyTorch) with quantization and pruning for real-time edge inference. Built AWS ECS-based training/inference platform handling ETL pipelines and model deployment.</li><li>Invented patented object tracking and re-identification system (<a href="#">US11915434B2</a>) enabling cross-camera tracking for physical analytics, deployed in \$1M+ contracts still in production today.</li></ul>	
<b>Machine Learning Engineer</b> — Qimia, Inc. <i>San Diego, CA</i>	Sep. 2019 – June 2020
<ul style="list-style-type: none"><li>Led data engineering and ML for large-scale advertising optimization across billions of records using Apache Spark. Developed spill detection and inventory analysis systems for robotics deployments in retail environments.</li><li>Built PySpark/Scala data pipelines and predictive models on AWS Redshift, delivering insights through custom dashboards and Docker-based applications.</li></ul>	
<b>Machine Learning Researcher</b> — University of Windsor <i>Windsor, ON</i>	Jan. 2018 – Aug. 2019
<ul style="list-style-type: none"><li>First-authored three publications related to eye-gaze tracking and signal processing (<a href="#">Researchgate</a>). Thesis: <i>Statistical Methods to Measure Reading Progression Using Eye-Gaze Fixation Points</i> (<a href="#">link</a>).</li></ul>	

## EDUCATION

<b>M.Sc, Electrical and Computer Engineering</b> — University of Windsor <i>Windsor, ON</i>	Jan. 2018 – Aug. 2019
<b>B.Eng, Mechanical Engineering</b> — University of Windsor <i>Windsor, ON</i>	Sep. 2012 – Aug. 2016

## TECHNICAL SKILLS

<b>ML/AI:</b> LLMs (RAG, prompt engineering, LoRA/QLoRA, prefix tuning), Multimodal Models (CLIP, BLIP, VideoLLava, custom implementations), Vision Models (CNNs, ViTs, VJepa2), Agentic Systems, Vector Databases, Hierarchical Reasoning
<b>Frameworks:</b> PyTorch, TensorFlow, Langchain, MLflow, Huggingface, PySpark, Scikit-Learn, OpenCV, TensorRT, ONNX
<b>Infrastructure:</b> AWS (ECS, MWAA, ECR, Redshift), GCP, Azure, Docker, Airflow, Git/GitHub CI/CD
<b>Languages:</b> Python, C/C++, SQL, Scala, MATLAB