

# Ram Senthamarai

[ram.senth@gmail.com](mailto:ram.senth@gmail.com)

<http://www.linkedin.com/in/ramsenth>

<https://github.com/ram-senth>

## Professional Profile

---

15+ years of software engineering experience in distributed systems, backend engineering, data engineering, ML engineering and big data. Most recent 5 years focused on large scale feature engineering, experimentation, machine learning, ML infrastructure and backend services. Seeking a technical leadership role at the crossroads of software engineering and data science.

## Experience

---

### Amazon

Palo Alto, CA

#### Senior Software Engineer (L6)

Jun 2021 - Jul 2023

- Reduced the cost of serving visual search queries by 30% using a Vision Transformer based classification model to control downstream call fan-out
- Improved the visual search relevance score by about 8 percent through tuning of ensemble of models
- Productionized multimodal object detection model that uses text and image resulting in about 4% improvement to visual search result relevance
- Responsible for designing and building inference systems, running experiments and providing cross team technical leadership across these 3 projects

### Uber

Palo Alto, CA

#### Senior Software Engineer II (5b)

Apr 2018 - Feb 2021

- Implemented MinHash and Locality Sensitive Hashing based system to identify duplicates in 500k user signups each day in near real-time, saving Uber about 1 million USD a month
- Improved availability from 98% to 99.98%, reduced p99 response time by 75% and increased the QPS capacity by 3x for the graph query web service
- Implemented 3 feature engineering pipelines to publish, in near-real time, 55 graph based features that are used by 8 machine learning models to improve fraud detection rate by 4.5 basis points
- Evangelized Cypher query language for graph based data exploration and feature discovery and drove implementation on top of in-house HDFS based analytical graph database

### Lumity

San Mateo, CA

#### Founding Team Member, Principal Software Engineer

Jul 2014 - Feb 2018

- Owned the backend and the APIs for health plan recommender and shop-for-benefits experience
- Architected benefit package system for complex and dynamic rules around rates, premiums, payroll deductions, waiting periods, eligibility changes, and contributions

### Xperi (previously Rovi)

Santa Clara, CA

#### Software Architect

Mar 2009 - May 2014

- Defined the architecture for Rovi ad delivery network for connected TVs and devices
- Lead ad network implementation and multiple client integrations

## Education

---

### University of California, Berkeley

Dec 2023

Master of Information and Data Science (MIDS)

**Relevant Courses** Applied Machine Learning, Machine Learning at Scale, Statistics For Data Science, Natural Language Processing with Deep Learning, Computer Vision, Research Design and Applications for Data and Analysis, Modern Data Applications and Data Visualization.

## Skills and Tools

---

**Languages:** Java, Python, Scala, Javascript, R

**Data Stores:** PostGres, MySql, Cassandra, Hive, Redis

**Cloud Platforms:** Amazon Web Services (AWS), Google Cloud Platform (GCP), Databricks/Azure

**Distributed Computing:** MapReduce, Hadoop, Spark, Flink, Kafka

**ML Tools:** Jupyter Notebook, D3.js, Tableau, Pandas, Numpy, Matplotlib, TensorFlow, PyTorch, LangChain, scikit-learn, HuggingFace

## Key Projects

---

### AI Based Conversational Tutor

Oct 2023 - Dec 2023

Capstone Project

UC Berkeley

[Web based chatbot](#) for interactive Spanish learning using fine tuned BERT for error classification and LLM for response generation. Hosted on AWS Lambda, Amplify and SageMaker. One of the **6 projects selected out of 41 for showcase**.

### Interactive Story Generation For Kids

Oct 2022 - Dec 2022

Natural Language Processing with Deep Learning Course

UC Berkeley

[Fine tuned large language models](#) (LLMs) for Natural Language Generation (NLG) task of story generation.

### Distracted Driver Detector

Oct 2022 - Dec 2022

Computer Vision Course

UC Berkeley

[Built and evaluated](#) VGG16, RandomForest and Logistic Regression models powering an iOS app for active detection of distracted driving.

### Flight delay prediction

Mar 2022 - Apr 2022

Machine Learning at Scale Course

UC Berkeley

Databricks/Azure, PySpark and MLLib based logistic regression model, trained on terabytes of airlines, airports and weather data, for predicting flight delay.

## Training

---

### [Deep Learning Specialization](#)

Aug 2022

DeepLearning.AI (through Coursera)

### [Machine Learning](#)

Jul 2019

Stanford University (through Coursera)