

UMUT (HOPE) YILDIRIM

ML ENGINEER / NEW YORK CITY, NY

umut475@gmail.com

+1 (415) 792-9337

umutyildirim.com

in/umuthopeyildirim

SUMMARY

About I am a ML engineer passionate about productionizing AI. I enjoy transforming AI tools and research into practical, scalable software solutions.

Skills Python, PHP, NodeJS, JavaScript, React, NextJS, Linux, Computer Vision, Deep Learning(Transformers, LLMs), Big Data, Synthetic Data, Productivity Tools, Unity Engine(C#)

COMPETITIONS AND TECHNICAL PROJECTS

MarkAI - markai.umutyildirim.com

(Nov 2023)

- An open-source OpenAI wrapper for a RAG-based chatbot that seamlessly integrates with your documents.
- All endpoints are serverless, with the exception of the database, ensuring ease of setup and immediate use.
- Utilized technologies like NextJS, OpenAI, Vercel, Clerk, Langchain and Supabase.

Flatiron Open Source - flatironopensource.com

(Jan 2023)

- Built an open source hub for bootcamp graduates to have efficient access to course modules and lessons.
- Contributed to 70% of the front-end development and spearheaded the entire backend implementation.
- Utilized technologies like JavaScript, Vite, React, Cloudflare(Pages&Workers&R2), Education.

CUBBA: Pothole Detection System

(Jul 2020)

- Engineered a computer vision program for pothole detection using image classification on Raspberry Pi's.
- Optimized for Raspberry Pi's limited real-time CPU capabilities by multi threading and job management.
- Self-learned OpenCV, TensorFlow, Azure Cognitive Vision to develop and optimize for Raspberry Pi.

Hackathons

(2018 - Present)

- Participated in over 10 hackathons, winning seven, demonstrating innovation and technical proficiency.

WORK EXPERIENCE

Medal.TV

ML Engineer

New York City, NY, USA (April 2024 - Present)

- Working for Ferox's Highlight product, which is a screen recorder that extracts context from your screen and makes it interpretable by LLMs.
- Responsible for building RAG (Retrieval Augmented Generation) systems for Highlight
- The system will enable users to store their personal data over time, and will label it with the appropriate metadata in order to make sure it is retrieved at the right time and in the right context is the most critical piece of the product, as having data aggregated in one place is what will enable users to interact more robustly with language models

Helicone

Full-Stack Engineer

San Francisco, CA, USA (November 2023 - December 2023)

- Contributed to the development of Helicone, a platform for monitoring Large Language Models at scale, streamlining the management and analysis of LLMs.
- Led significant technical enhancements including revamping documentation, integrating a "cache_hits" table in Clickhouse, and transitioning API calls from "axios" to "fetch".
- Utilized a tech stack comprising NextJS, Cloudflare, OpenAI, Tailwindcss, Vercel, Supabase, Clickhouse, TypeScript, PostgreSQL, KV, and Cache.

Mirage

CTO

Istanbul, TR (March 2023 - November 2023)

- Developed a user-friendly web application that helps clients easily create and order large numbers of synthetic images for computer vision tasks such as image classification and object detection.
- Automated the Unity Engine with an add-on for handling image generation requests, allowing real-time synthetic image testing and facilitating bulk orders exceeding 10 million images for computer vision applications.
- Utilized technologies like JavaScript, Unity, Next.js, React, AWS, and Supabase.

PUBLICATIONS

Experimentation in Content Moderation using RWKV

Sept 2024

- Investigated RWKV model's efficacy in content moderation, leveraging its CPU-efficient architecture
- Created a novel dataset with images, videos, sounds, and text for distillation into smaller models
- Generated extensive responses (558,958 for text, 83,625 for images) using LLMs to train moderation systems
- Demonstrated RWKV's potential to improve accuracy and efficiency in content moderation, paving the way for compact, resource-efficient models

EDUCATION

Flatiron School

Manhattan, NY (Aug 2022- Nov 2022)

- Computer Science(Certificate) / Intensive web development training with Ruby(Rails), React, and JavaScript.

Isik University

Istanbul, TR (2020-2022)

- Associate Degree in Computer Programming - Associate Degree, Computer Science