

Mario Callejas

Engineer

411 44 590
mariothedev@gmail.com
Oslo, Norway

EXPERIENCE

HeyAudio— *AI Engineer*

Aug 2023 – Present (Remote)

- ❖ Developed cross-platform Flutter/Dart mobile apps and low-latency Python/Django backends, enabling real-time WebRTC audio processing with LiveKit.
- ❖ Built domain-specific voice assistants using LLMs, agents, and vector databases, implementing retrieval-augmented generation (RAG) for context-aware responses.
- ❖ Fine-tuned large language models using LoRA and trained custom classification models for domain-specific tasks.
- ❖ Owned MLOps pipelines for model evaluation, versioning, serving, deployment, and monitoring, balancing accuracy, latency, and inference cost.

ArribaTec— *Core Engineer*

Jan 2022 – Aug 2023 (Norway)

- ❖ Owned Microsoft Azure cloud infrastructure and DevSecOps pipelines, implementing scalable architecture, firewalls, VPN connectivity, and access controls to strengthen security, reliability, and operational efficiency.
- ❖ Improved frontend architecture and developer workflows through Redux state management, mock servers, and

GENERAL SKILLS

- Data structures and algorithms
- Computer networks
- Windows/Linux/macOS
- Bash/Make
- Vim

FRONTEND SKILLS

- HTML/Pug, CSS/SASS, Javascript/Typescript
- React, Angular, Ember

BACKEND/DEVOPS SKILLS

- Node.js/Bun, Python (FastAPI/Django)
- PostgreSQL, SQLite, MongoDB/CosmosDB, Redis, Elasticsearch, Qdrant
- Sharding, partitioning, message queues, Publish/Sub architecture, Materialized views
- Nginx/Caddy, load balancing, reverse proxying, firewalling, vpn
- HTTP, gRPC, TCP/Websockets, UDP,

automated testing, increasing maintainability and accelerating development cycles.

- ❖ Delivered secure, reliable, and performant full-stack features using React, React Native, Node.js, and MongoDB (Mongoose), applying test-driven development (TDD) with Jest.
- ❖ Built and maintained a Next.js admin dashboard for managing and monitoring self-service kiosks.

Paycertify — Senior Fullstack Engineer

August 2018 – Jan 2022 (USA)

- ❖ Shipped full-stack features across a Ruby on Rails/PostgreSQL API and Ember.js frontend, applying security best practices and comprehensive unit and acceptance testing with QUnit.

Techdata – Senior Frontend Engineer

August 2014 – August 2018 (USA)

- ❖ Lead for organization's landing page
- ❖ Led the company's design system using Atomic Design principles
- ❖ Built and scaled high-traffic Angular/.NET e-commerce applications, with a focus on performance, reliability, and maintainability

Webhooks, SSL, TLS, SSE, WebRTC (LiveKit)

- Docker Compose/Swarm, Kubernetes
- TerraForm/GitOps

CLOUD SKILLS

- AWS, Google Cloud, Digital Ocean, Firebase, Heroku, Azure

SECURITY

- XSS (stored/reflected), CSRF/SSRF, Dom Injection, SQL Injection, DDoS, Session Hijacking and Clickjacking
- Firewalls/VPNs
- Penetration testing: Nmap, Nuclei, ZAP, dalfox, katana

TESTING

- Unit testing (Mocha, Jest, Sinon, Qunit)
- E2E testing (Selenium, Nightwatch)

MACHINE LEARNING

- CNN, RNN and Transformer architectures
- Agentic frameworks: CrewAI, Agno and Langchain
- RAG: sparse and dense retrieval, chunking, ranking
- Prompt engineering (llm-as judge, CoT, ReAct, Reflexion)
- PEFT Fine-tuning

(LoRA/QLoRA)

- Model inference - vLLM
(GPU inference) * llama.cpp
(CPU inference)

- PromptFoo (prompt eng.)

- Evals & Monitoring: Ragas,
DeepEval, LangFuse & MLFlow

llama.cpp/Ollama (CPU
inference) PromptFoo (prompt
eng.)