

Washington X. Quevedo

AI Engineer · Applied ML, LLM Agents & XR

Quito, Ecuador (UTC-5) · Open to remote · wxquevedo@inmersoft.com · linkedin.com/in/wxquevedo · github.com/wx-quevedo · Google Scholar: scholar.google.com/citations?user=zPNPhwkAAAAJ

Summary

AI engineer who takes models from experiment to production: on-device computer vision, LLM-based agents and tooling, and inference services running on modest infrastructure. Currently building the AI identification module of a Stanford University marine research platform, live on the web with its mobile app in development. Also an XR developer and published researcher: 30+ peer-reviewed papers, 640 citations, h-index 14, a registered software IP and a national research award. Founder of Inmersoft (2018), with 200+ delivered projects across AI, XR, web and mobile.

Experience

Stanford University — Software & AI Consultant (Buscapez project)

07/2026 – present · Remote · contracted through Inmersoft as an official Stanford supplier

- Built the species-identification module for a tuna-tagging research platform: a 4-class YOLO detector trained in PyTorch (mAP50 0.96 for detection on validation), exported to ONNX for the server and to ExecuTorch for offline on-device inference on iOS and Android.
- Found that the mobile runtime distorted input images and fixed it with letterboxing; species classification keeps improving as the team validates more real photos.
- Deployed a FastAPI + ONNX Runtime inference service in production: 49 ms per image and 70 MB of RAM on a 2-core VPS shared with other systems.
- Shipped the human-in-the-loop data pipeline: assisted annotation with mandatory human review, bulk dataset upload with hash-based deduplication, and YOLO export for incremental retraining.
- Delivered the web platform (Laravel, React), live in production; the React Native/Expo mobile app is in development.

Laboratorios Bagó del Ecuador — Digital Solutions & AI Developer

10/2020 – present · Quito (pharmaceutical company)

- Built an AI-assisted platform for the product managers' yearly creative event (MKT Attack): schedules the sessions, shows who presents and the time left, and gives teams product information as context for ideas; 8 product managers, 12 campaigns.
- Developed vSpace, the medical-visit platform used by the sales force (literature, visit logging, promotions, interactive apps), including 20 embedded mini-games with 400 users and 20,000 monthly interactions.
- Delivered a group VR session on 30 synchronized Meta Quest headsets: one Unity app (360° guided tour, meditation, interactive game) controlled from a Windows panel over a local network, with per-headset battery status; all 360° video shot and edited in-house.
- Built TrackingApp, a logistics platform for promotional material (Laravel 12, Inertia, React): 3 roles plus a public portal, PDF waybill parsing, Excel import and scheduled notifications, covered by 246+ automated tests.
- Built the MKT Attack event web app (live schedule, QR-based uploads, live ranking), a corporate metaverse and 12 augmented-reality filters.

Arlift — Platform Engineer (contract)

10/2025 – present · Remote · QHSE compliance for an oil & gas operator

- Built a QHSE compliance platform with a Laravel API and a React SPA: RBAC, workflow engine, audit logging, org chart (React Flow), geospatial maps (Leaflet) and 5 compliance modules, covered by 308 tests.

Inmersoft — Founder & Principal Engineer

09/2018 – present · Quito

- Own engineering company with 200+ delivered projects: VR/AR for engineering, medicine and entertainment, digital twins of drones and robots for control-algorithm testing, Unity ML-Agents, and hardware integration (Leap Motion, Emotiv, Arduino, IoT).

Sansar (Linden Lab) — Game Developer (contract via BairesDev)

2022 · Remote

- Built avatar interaction mechanics and optimized shaders and materials in a social VR platform, reducing memory usage by ~40%.

Earlier roles

- **Merge** — AR Research Engineer (2020): led the migration research from Vuforia to Unity AR Foundation with unified ARKit/ARCore object recognition.
- **WebCreek** — Unity 3D Developer (2019 - 2020).
- **Inter-American Development Bank** — Platform Developer, GoHelp platform (2016 - 2017).
- **Universidad de las Fuerzas Armadas ESPE** — VR/AR Research Analyst (2016 - 2018).

Selected Projects

Localia — private generative-AI platform (localia.cc)

- End-to-end pipeline from a web app (Laravel, React) through a Redis queue to remote GPU workers (RTX 4090) running ComfyUI models (FLUX 2, Wan 2.2, LTX), with results stored in Cloudflare R2.
- Guaranteed result delivery with Redis BLMOVE and pending-message recovery; real-time progress over WebSockets.

Local LLM infrastructure and multi-agent development workflow

- Self-hosted llama.cpp server running Gemma (26B QAT) at 130-175 tokens/s on an RTX 4090 with tool calling, shared across the local network.
- Orchestrator/coding-agent workflow with Claude Code and MCP integrations (ClickUp, Figma, browser), plus a custom planning kit that tracks phases, estimates and real agent time per task.

Gazetrack — webcam gaze tracking (paper in preparation)

- 3.97° overall (2.51° in the central region) on a 49-inch ultrawide with free head movement, versus the 4°-8° typical of webcam systems. Fully local processing.

Skills

AI / ML	Python, PyTorch, Ultralytics YOLO, ONNX Runtime, ExecuTorch, computer vision, model export and on-device inference, ComfyUI (FLUX, Wan, LTX)
LLMs & agents	Anthropic Claude API (primary), OpenAI and Google Gemini APIs, Claude Code, MCP, tool calling, local LLM serving (llama.cpp), multi-agent workflows
Backend	FastAPI, Laravel, REST APIs, Redis, PostgreSQL, MySQL, SQLite, WebSockets
Frontend & mobile	React, TypeScript, Inertia, React Native / Expo
Infrastructure	Linux (Ubuntu VPS), Nginx, Supervisor, systemd, Git, Cloudflare R2
XR & games	Unity, C#, VR/AR (Meta Quest, ARKit/ARCore), Unity ML-Agents

Research & Recognition

- 30+ peer-reviewed publications (Springer: AVR, ISVC; ACM ICETC; CISTI) · 640 citations · h-index 14 · i10-index 21 (Google Scholar, Oct 2026). Peer reviewer.
- Most cited: Unity3D-MatLab simulator in real time for robotics applications (88 citations); Virtual reality system for training in automotive mechanics (first author, 50 citations).
- Registered software IP: Virtual 2D Laser Sensor, Ecuadorian Intellectual Property Institute, QUI-052058 (2017).
- First place, SENESCYT National Awards (Galardones Nacionales), 2016.

Education

B.Eng. in Electronics and Instrumentation Engineering — Universidad de las Fuerzas Armadas ESPE, Ecuador (2009 - 2015)

Languages

Spanish (native) · English (professional working proficiency)