

Omar Olivares Urrutia

AI / Software Engineer — LLM Systems, Agents & Evaluation

7510043, Providencia, Santiago, Chile — open to Paris / EU (CL)



✉ omar@olivares.cl 📞 +56983810103

🐙 [ofou](https://github.com/ofou) · <https://github.com/ofou> [in](https://www.linkedin.com/in/ofou/) [ofou](https://www.linkedin.com/in/ofou/) · <https://www.linkedin.com/in/ofou/> [🐦 omarnomad](https://twitter.com/omarnomad) · <https://x.com/omarnomad>

SUMMARY

AI engineer shipping production LLM systems: agents, RAG, evals, and tool/MCP platforms. Peer-reviewed LLM evaluation work (DataEd'26: 19 models, 300 SQL cases), early engineer on a high-growth open-source coding agent (~10K GitHub stars), and production RAG for 200K+ papers. Strong in Python, FastAPI, and end-to-end ML product delivery. Open to relocating to Paris / EU.

SKILLS

ML / AI

LLM Agents ·
LLM Evaluation (LLM-as-a-Judge) ·
RAG / Retrieval ·
Fine-tuning (LoRA/QLoRA/PEFT) ·
MCP / Tool Use · Open-Source AI

Frameworks & Tools

PyTorch · Hugging Face ·
LangChain ·
OpenAI / Anthropic APIs ·
Weights & Biases · MLX

Infrastructure

Python / FastAPI · Docker ·
PostgreSQL · Vector DBs ·
AWS / GCP · CI/CD

Languages

Python · SQL · TypeScript

WORK EXPERIENCE

Senior AI Engineer at Maxxa Software SpA

Mar 2026 - Current

📍 Santiago, Chile

Fintech (credit & surety). Own production LLM platform work on the TECH team — governance, evals, observability, and agent tooling for financial products.

- Built LLM governance, evaluation, and observability for production models serving financial products
- Shipped core + enterprise LLM platform features used across internal AI products
- Built Model Context Protocol (MCP) server exposing company data and tools to AI agents and internal teams

AI Engineer at Universidad de Talca

Sept 2025 - Mar 2026

Multi-agent LLM platform for SQL education — real-time syntactic, semantic, and logical feedback on student queries.

- Designed multi-agent pipeline for granular SQL feedback (syntax, semantics, logic)
- Benchmarked frontier LLMs (Grok, GPT-4o, Claude, Gemini, DeepSeek) for tutoring quality
- Built eval suite: 19 LLMs × 300 cases with LLM-as-a-Judge + human-annotated ground truth
- Co-authored peer-reviewed paper accepted at DataEd'26 on LLM SQL error detection

AI Engineer & Developer Relations at Kilo Code

Mar 2025 - May 2025

📍 San Francisco, CA / Amsterdam

Early hire (12-person founding team) on Kilo Code, a high-growth open-source coding agent (Cline/Roo-Code lineage).

- Helped launch and scale one of the fastest-growing open-source coding agents (~10K GitHub stars)
- Built OpenRouter integrations and docs that supported #1 model ranking on OpenRouter
- Wrote developer onboarding guides and reference integrations for the agent ecosystem
- Worked with leadership on technical architecture and go-to-market (Sid Sijbrandij's team)

AI Engineer at Emergent Mind

Feb 2024 - Feb 2025

📍 North Carolina, US (Remote)

Production RAG for scientific discovery — search and analysis over arXiv-scale corpora.

- Redesigned RAG (chunking + hybrid retrieval + reranking) for 200K+ indexed papers
- Built daily arXiv ingestion/indexing pipelines and W&B evaluation harness
- Shipped FastAPI + Docker services with tracing and reliability tests for production serving

Content Developer at Neura Pod

Sept 2021 - Jan 2025

📍 San Diego, CA (Remote)

Neurotech & AI content; grew channel past 80K YouTube subscribers with multiple 200K+ view videos.

- Translated frontier BCI/AI research into accessible content for a broad technical audience

AI Evaluation Contractor at Centro Nacional de Inteligencia Artificial

Nov 2024 - Nov 2024

📍 Santiago, Chile

Model evaluation for LATAM Spanish-language LLMs.

Software Engineer at clicOH

Jul 2021 - Dec 2021

📍 Remote

Backend engineering at logistics startup (RayoApp → clicOH acquisition).

- Backend development and post-acquisition system integration
- Maintained legacy systems while migrating features to the new architecture

Software Engineering Intern at Suncast

Jan 2020 - Apr 2020

📍 Las Condes, Chile

Renewable-energy forecasting platform; helped move a monolith toward microservices.

- Refactored services and APIs; optimized database queries for forecasting workloads

PROJECTS

Graham Essays

Jul 2022 - Current

Skills: Python · Open Source

🔗 <https://github.com/ofou/graham-essays>

Automated EPUB generator for Paul Graham's essays — #1 on Hacker News, 800+ GitHub stars

- #1 on Hacker News frontpage
- Pipeline processing 200+ essays into EPUB
- 800+ GitHub stars

Lucas — AI Finance Agent

Nov 2025 - Current

Skills: Anthropic API · Agents · FastAPI

 <https://github.com/platanus-hack/platanus-hack-25-team-29>

Agentic personal finance assistant with live bank API integration (36h hackathon)

- Claude agent + FastAPI backend connected to bank accounts via Fintoc

VOLUNTEER

Software Engineer at [Neuromatch Academy](#)

Jul 2021 - Sept 2021

Software support for an inclusive computational neuroscience learning community.

EDUCATION

Postgraduate Diploma Generative AI and Large Language Models at [Centro Nacional de Inteligencia Artificial \(CENIA\)](#) Aug 2025 - Feb 2026

Major courses: LLM Fine-tuning (LoRA, QLoRA, PEFT) · Production ML (PyTorch, MLflow, W&B) · Multimodal Models & RAG Systems · Ethical AI & Bias Management

Engineer's degree Computer Science at [Universidad Católica del Maule](#)

Mar 2018 - Dec 2023

Major courses: Thesis: Fine-tuned GPT-3.5 for Chile's PAES math exam — accuracy 0.78 → 0.94 (+20%) · Data Structures & Algorithms, Databases, Distributed Systems, AI/ML

Engineer's degree (coursework) Computer Engineering at Universidad de Talca

Mar 2015 - Dec 2017

AWARDS

2nd Place — Computer Vision Hackathon at Fundación País Digital | EY

Sept 2025

Traffic monitoring from CCTV streams (congestion/anomaly detection), designed to scale across many cameras.

CERTIFICATES

IELTS Academic — Overall Band 7.5 (C1) Cambridge English

Jul 2025

PUBLICATIONS

Assessing the Ability of Large Language Models to Identify and Explain Errors in SQL Queries in DataEd'26: International Workshop on Data Systems Education Mar 2026

Empirical evaluation of 19 LLMs on SQL error detection (300 cases): LLM-as-a-Judge + human ground truth. Implications for automated tutoring systems.

Optimization of Language Models for Chile's Mathematics Higher Education Admission Test (PAES) in Universidad Católica del Maule

Feb 2024

Fine-tuned GPT-3.5 for PAES mathematics; accuracy 0.78 → 0.94 (+20%) with targeted optimization and OCR/transcription pipelines.

LANGUAGES

Spanish

Native

English

Full Professional (IELTS C1 7.5)

INTERESTS

Technology

Creative

Skills: Frontier LLMs ·
Neurotechnology / BCI ·
Open source

Skills: Music production ·
Technical writing

REFERENCES

“ *Co-founder, Kilo Code* ”

— **Matt Mazur**

“ *Founder, Emergent Mind* ”

— **Ryan Tanaka**