



CONTACT

@ yahyaouladchiba@gmail.com
Greater Paris, France · Remote
in linkedin.com/in/yahya-aoulad-chiba
+33 7 55 32 75 36
smaclaw.com
gh github.com/yahyaChiba
Available immediately

STACK

AI
AI agents in production & MCP tool-calling ·
Hermes Agent · OpenClaw · RAG · Vercel AI SDK ·
Anthropic / OpenAI / Gemini · Structured output
(Zod) · Agentic evaluation (model-graded +
deterministic guards) · Run tracing & replay ·
Token-cost metering · latency and cost tracked ·
Voice agents

CORE
TypeScript · Python · JavaScript · Java · C#

BACKEND
Node.js · Python (FastAPI / Flask · SQLAlchemy ·
pytest) · REST / GraphQL · Job queues & events ·
WebSockets · Java 17 / 21 · Spring Boot / Hibernate
· Kotlin

DATA
PostgreSQL (RLS, migrations, PL/pgSQL) · Redis ·
MongoDB · Supabase · SQL Server

FRONTEND
React · Next.js · React Native / Expo · Tailwind ·
ExtJS · Vitest / Playwright

DEVOPS
Docker · Kubernetes · CI/CD · Cloudflare

LEADERSHIP
Engineering Management · Architecture · Agile /
Scrum

EDUCATION

MEng, Computer Software Engineering
EHEIO, Oujda, Morocco · 2015 – 2018

LANGUAGES

English	Fluent
Arabic	Native
French	Fluent

YAHYA AOULAD CHIBA

Senior AI & Full-Stack Engineer · Tech Lead · VP of Engineering

~10 yrs

HANDS-ON ENGINEERING

7 yrs

ONE COMPANY · DEV+VP

20+

ENGINEERS LED AT PEAK

3 yrs

AI · AGENTS · RAG

PROFILE

Senior engineer, ten years in production software, the last **3 focused on AI**. I joined a regulatory-software company as an intern and left as its **VP of Engineering**, leading up to 20+ engineers at peak and driving the platform’s Java 11 to 16 and Kubernetes modernisation myself rather than delegating it. I still build: agents in front of real users, **MCP tool-calling**, RAG over each client’s own documents, and evaluation treated as part of the system, with a model grading what needs judgment and code checking what has to be exact. Java and Spring at enterprise scale, TypeScript and Python across everything I have built since.

EXPERIENCE

Independent

Remote · International

Senior AI & Full-Stack Engineer

Jul 2026 – Present

- Build and operate **smaclaw.com**, a multi-tenant AI-agent platform live with real users.
TypeScript and Python end to end: Next.js, Node and **Python (FastAPI / Flask)** services, PostgreSQL with row-level security, 100+ API routes, Stripe metered billing, and a voice agent that books over a real phone call.
- MCP tool-calling in production**: agents reach the calendar, the CRM and each tenant’s own documents through **MCP** tools, not hard-wired integrations. The platform is **agent-runtime agnostic**: I run both **Hermes Agent** and **OpenClaw** on the same tool layer, so runtimes can be compared on reliability, cost and failure modes.
- Reliability and operations**: an LLM grades every drafted reply and tool call before it leaves, and anything that must be exact is checked in code — the booking agent can only offer slots the calendar returns. Runs are traced end to end and replayable, prompts ship through the same CI/CD as the code, and I track grounded-answer rate, tool-call validity, escalation and cost per conversation.

EcoMundo

Greater Paris, France

VP of Engineering

Mar 2023 – Jul 2026

- Owned the engineering organisation — strategy, architecture, roadmap and hiring — partnering with product owners on priorities and estimations, leading **up to 20+ engineers** across three teams at peak while staying hands-on in the code.
- Delivered COSMETIC Factory, a regulatory PLM platform, end to end for enterprise clients — REACH and SDS data at volume.
- Drove platform modernisation — containerised the product onto **Kubernetes** with CI/CD, plus the **ExtJS → React** migration and **Java 11 → 16** — while the product kept shipping for enterprise clients.

Team Lead

Jan 2021 – Feb 2023

- Promoted to lead one team, then to coordinate delivery across three — growing engineers into leads and spreading domain knowledge that had been concentrated in too few heads.
- Combined hands-on full-stack development in Java and ExtJS / React with planning, estimation and mentoring.

Software Engineer

Feb 2019 – Dec 2020

- Hired after a 4-month end-of-studies internship here. Frontend developer (ExtJS 4.2) → full-stack with Java 11 and ExtJS 7, building COSMETIC Factory features.

Self-employed

Morocco

Web & Application Developer

Sep 2015 – Feb 2018

- School management platform** for a private school — class, teacher and bus timetabling, student records, fee and staff payments. *C# / WinForms*.
- Association management app** — members, activity planning, transport scheduling, certificate generation. *ASP.NET*.
- Stock & inventory management** for a poultry distribution company. *C# / WinForms*.
- Sole developer on each, requirements to deployment — SQL Server with Entity Framework throughout, while completing my engineering degree.

DOMAIN EXPERTISE

Seven years in a **regulated domain**, where satisfying the regulation *was* the product: REACH/SVHC, CLP/GHS Safety Data Sheets, cosmetics PLM. I turned it into domain models, **Drools rule engines and compliance workflows that hold up under audit** — what shipping AI with real consequences demands: traceability, explainability, human review.

Eight of 38 repositories, designed and built end to end — data model, API, frontend, and the infrastructure underneath. Multi-tenant SaaS, platform engineering, and AI features that reach production.

smmaclaw.com — White-label AI-Agent SaaS

Multi-tenant SaaS that auto-provisions a branded AI voice & chat agent per tenant, end to end.

Automated per-tenant server provisioning, 100+ API handlers, 40+ Postgres migrations with row-level security, Stripe metered billing plus a partner-commission engine, A2P-registered telephony (Twilio), and a Cloudflare Worker edge router. JWT issuance, encryption at rest, secret rotation. Unit tests plus an 11-scenario end-to-end suite.

Next.js

Postgres / RLS

Stripe metered billing

Twilio

ElevenLabs

Cloudflare Workers

Vitest

Playwright

AI Calling & Booking Backend

Turns CRM form submissions into AI-voice outbound calls, and books the outcome.

HMAC-verified webhook ingestion with idempotency keys and per-endpoint rate limiting. Voicemail detection, double-dial retry, timezone-aware scheduling, and a multi-schema Postgres store with RLS persisting calls, transcripts and bookings. Built for the failure cases, not the happy path.

TypeScript

ElevenLabs

GoHighLevel API

Postgres / RLS

HMAC webhooks

Rate limiting

Conversational AI Sales Agents

Multi-tenant chatbot backends that converse with inbound leads and book them automatically.

Layered Flask architecture (controllers → handlers → data access → SQLAlchemy) with per-organisation keys and prompts, LLM function-calling into a calendar to book, reschedule and cancel, token-cost usage metering, scheduler loops, and multi-architecture Docker CI. Refactored into a reusable library once the third tenant needed it.

Python

Flask / RESTX

SQLAlchemy

LLM function calling

Twilio

Docker Buildx

pytest

Arabic (RTL) LLM Research App

An Arabic-first AI research assistant that turns a question into sourced, scored findings.

Conversational intake → research plan → multi-source scraping → dual-LLM synthesis, on a provider-agnostic pipeline so models swap without touching product code. Structured-output intake with schema validation, then batched relevance scoring. Full right-to-left Arabic UX.

TypeScript

Vercel AI SDK

Anthropic

OpenAI

Gemini

Zod

RTL / i18n

Google Maps Lead-Extraction Platform

Three-tier system that extracts and enriches business leads from map data at scale.

A Next.js frontend, a Flask container-orchestrator driving the Docker SDK, and Playwright workers behind a job queue so runs resume, with **dedup on a stable business key** before enrichment. Per-tenant ephemeral containers, plan-limit metering and HMAC-verified payment webhooks. **A list that took ~a day of manual work now builds in under an hour.**

Next.js

Flask

Playwright

Chrome MV3

Docker SDK

GitHub Actions

pytest

k3s Scraping Cluster — Infrastructure as Code

A Kubernetes cluster that scales scraping workers with demand.

k3s on cloud hosts: a Redis job queue, a FastAPI controller, and a Horizontal Pod Autoscaler with tuned scale policies, resource requests and limits, and managed Secrets. Cluster, autoscaling policy and workers are all mine, built from scratch.

Kubernetes / k3s

FastAPI

Redis

HPA

Docker

IaC

Car-Rental Platform — Web + Mobile

A multi-tenant car-rental suite: one admin dashboard, two mobile apps, one backend.

Operator and customer apps sharing a workspace package. Realtime chat, edge functions firing push notifications on booking-status change, an offline-first action queue, role-based access control, and RPC availability checks.

Next.js

React Native / Expo

Supabase

Realtime

Offline-first

RBAC

Sentry / PostHog

Inventory & Invoicing (ERP)

Stock, invoicing and accounting for a glass-fabrication business.

Real accounting modelling rather than CRUD: an immutable inventory-moves ledger with stock derived by summation, partial-payment allocation, AR/AP aging, profit and valuation reports, and server-side PDF invoices. Spec-first — 17 numbered module specs written before the code.

TypeScript

PostgreSQL

Ledger modelling

PL/pgSQL

PDF generation

Spec-first, then tests on the paths that hurt. Design specs before the code; Vitest plus an 11-scenario Playwright suite across signup → billing → onboarding.

I build with agentic coding tools daily — Claude Code and MCP — with specs and tests around them rather than raw prompting.

Personal data as a design constraint. Row-level security everywhere rather than filtering in application code, encryption at rest, secret rotation, HMAC-verified webhooks.

Every one of these was designed, built and shipped by me — data model to deploy. The repositories are private; I am happy to walk through any of them, or read code together in a live session.

yahyaaouladchiba@gmail.com