Resume · maxence.pro

Maxence Foulon

AI Product Engineer · Founder of Selvren

From research and UX to design, apps, backend and AI — I design it, build it and ship it.



Le Mans, France · French native · English C1 · Remote today · moving to Korea

WhatsApp +33 6 19 32 62 26

Email maxence@selvren.com

KakaoTalk ID Maxence.foulon

GitHub github.com/Takinggg

Looking for AI Product Engineer · Full-stack developer — a team or startup that needs someone who builds · remote + regular 2–3-month stays in Korea

11 yrs	~10+	5+	58 GiB
shipping code, since age 12	freelance clients	SaaS launched	LLM run on a 48 GB Mac

Summary

I design and ship whole products alone: research and UX, UI, iOS and Android apps (Expo), backend, data, infra and AI.

I run AI coding agents like a team and own the architecture, rules, merge gate and review (Selvren: 3,182 commits in 11 weeks, 11,901 tests).

Down to the inference engine: MoEVM (llama.cpp) runs a 58 GiB model on a 48 GB Mac, 3.1× faster with identical output.

Paid to code at 12, then around ten freelance UI/UX clients. Few diplomas; the work is the proof.

Experience

2026 – now — **Selvren** · Founder & CEO
B2B AI knowledge platform with source-verified answers; product and architecture

2025 – now — **Freelance** · AI Product Engineer
SouverainIA built solo, work with WebAPI, 5+ SaaS launched

2024 – 2025 — **Freelance (sole trader)** · UI/UX Designer
Around ten clients · Wiim AI, 6 months as the only designer

2023 – 2024 — **Unixo (IT & telecom)** · Apprentice → 4-month contract
Kept on after the CESI apprenticeship ended

2021 – 2023 — **Le Mans Université IT** · Apprentice (systems & network)
Support, GLPI service desk, PC deployment server, network and servers

2015 – — **Minecraft servers** · Plugin and mod developer
First paid code at 12

Education

2023 – 2024 — **CESI engineering school** · Left after 6 months to build — Engineering degree (Bac+5) · apprenticeship

2021 – 2023 — **EKOD (CCI Le Mans Sarthe)** · Graduated — Systems & Network Technician (Bac+2) · apprentice at Le Mans Université

2018 – 2021 — **Lycée Saint-Joseph La Salle** · With honours — Bac Pro · Digital Systems (SN)

Projects in depth

Skills

Research · UX
User interviews, personas, user flows, edge cases, Mobbin benchmarking, FigJam

UI · Front end
Figma, design systems, React, Vue, Next.js, Astro, TanStack Start, Motion

Mobile
Expo, React Native, iOS, Android

Backend
TypeScript, Node.js, Fastify, AdonisJS, Python, REST APIs, auth & 2FA, Stripe

Data
PostgreSQL, PostGIS, pgvector, Redis / Valkey, S3, migrations, data modelling

Infra
Linux, networks, Docker, VPS, nginx, Cloudflare

AI
RAG with citations, Mistral, local LLMs (Ollama, MLX), llama.cpp, agent orchestration, Claude Code, Codex

Selected projects

Selvren

B2B AI knowledge platform with source-verified answers

2026 – now · Founder & CEO

Teams ask AI domain experts that read their own private documents, and every claim is checked against its sources before it is shown. Available as a dashboard, a private API and MCP.

- Wrote the product direction and 28 architecture decision records; AI coding agents implement, behind my merge gate and review
- Fastify modular monolith (14 modules), PostgreSQL 17 + pgvector, row-level security, 143 migrations
- Ingestion: 12 formats, antivirus scan, originals encrypted with AES-256-GCM, Mistral embeddings
- Verified-claim answers: numbers, dates, conditions and negations checked against sources; abstains when they fall short
- Self-hosted European infrastructure and observability (VPS in Germany and Lithuania, Grafana, Loki, Tempo, GlitchTip)

3,182 commits in 11 weeks · **11,901 tests in the merge gate, none skipped** · **private retrieval recall@6 0.65 → 0.96** · **open-source build kit (TypeScript SDK + React, MIT)**

TypeScript · Fastify · PostgreSQL · pgvector · Mistral · React · TanStack Start · Clerk · Docker

MoEVM

MoE expert paging for llama.cpp

2026 · Personal R&D · open source

A llama.cpp extension that runs Mixture-of-Experts models larger than GPU memory on Apple Silicon: the experts in use stay on the GPU, the rest are paged from the SSD.

- Expert sidecar packing with a SHA-256 index, LRU RAM arena and prefetch
- Graph rewrite onto a persistent Metal pool, logit-parity tooling

gpt-oss-120b (58 GiB) on a 48 GB Mac's GPU at 8.72 tok/s (median of 5)

Qwen3-Coder-Next 6.41 → 19.90 tok/s (3.1× vs llama.cpp op-offload), bit-identical logits · **draft PR #25309 on llama.cpp**

github.com/Takinggg/moevm (MIT)

C++ · Metal · llama.cpp · GGUF · benchmarking

watch-mcp — An MCP server that lets AI watch videos and screen recordings

2026 · Open source · solo build

Lets Claude, Codex and Cursor understand MP4/MOV videos: keyframes picked where the picture changes, packed into timestamped storyboards, plus a local Whisper transcript in one timeline. WatchRecorder, its macOS app: press ⌥⇧R, show and tell, your AI gets the recording.

- MCP server (TypeScript): adaptive keyframes, timestamped storyboards, zoom/frame drill-down, bundled ffmpeg
- Local speech-to-text (Whisper on Metal) with word timestamps; the .watch bundle keeps screen, voice and clicks on one clock
- macOS app (Swift, ScreenCaptureKit): hotkey recording, a library, and the prompt handed to the AI app in use

Built and shipped in one night · **92 tests (66 Node, 26 Swift)** · **a 2-min 1080p recording watched in ~10 s, transcript included (M5 Pro)** · **github.com/Takinggg/watch-mcp (MIT)**

TypeScript · Swift · MCP · Whisper · ffmpeg · ScreenCaptureKit

Where? — An iOS app for planning outings, set in Seoul

2026 – now · Own product · end to end

An iOS-first Expo app that turns “let’s go out” into a real plan, alone or with friends. The map stays under clouds until you actually go there.

- Product rules, UX flows, visual direction, the iOS app, the backend and the AI build pipeline — all mine

38 end-to-end journeys · **344 flow steps** · **262 recovery branches** · **19 backend modules**

Expo · React Native · TypeScript · Fastify · PostGIS · Mapbox

SouverainIA — An AI executive assistant inside WhatsApp

2026 · Client · solo build

Watches important messages, prepares replies and generates documents such as quotes, right inside WhatsApp.

- Product, backend, security (TOTP 2FA) and deployment, built alone

1,749 automated tests

Python · WhatsApp API · Docker · Redis · nginx

SyncCast — Keeps esports casters on the same frame

2025 · Own product · beta

Keeps casters on the same frame across Twitch and YouTube, with leader-based resyncs on a shared server clock and its own HLS relay.

- Real-time sync, HLS relay (streamlink + FFmpeg → R2), roles, live drawing

268 commits in 8 days · **landing redesigned to explain the backend**

Next.js · Ably · Supabase · hls.js · FFmpeg · Cloudflare R2

More experience and projects

2026	**Bastide** Co-founder · brand, offer, site and SEO content for AI assistants hosted in France for SMEs
2026	**Naiad** Partner project · product, UX and app for connected aquariums (QR product recognition, live tank dashboard, AI assistant)
2024 – 2025	**Wiim AI** 6-month contract · the only designer, from UX thinking to the entire app
2021 – 2023	**Le Mans Université IT** User support, GLPI service desk, PC deployment server and workstation installs, network and server maintenance