



Nicolas Ha

Place of birth: Nantes, France

Nationality: French, British

Contact

📍 Paris, France (**Home**)

✉ hanicolas74@gmail.com

🌐 github.com/nha

🌐 stackoverflow.com/u/1327651

🌐 linkedin.com/in/nicolasha

🌐 nicolas-ha.com

ABOUT ME

Senior software engineer, 12+ years, now building AI systems: small specialized AI models, agentic orchestration and workflows, RAG, MCP tooling, agent memory.

I have led a backend team on a pipeline serving 1.3B views a month, helped stand up new teams inside a bank going for its licence, and shipped a product alone from an empty repo to the App Store. Hands-on throughout: I still write the code.

SKILLS

AGENTIC AI & LLM SYSTEMS

LangChain / LangGraph | MCP servers and tools | RAG and hybrid search | Agent memory | Multi-agent workflows | Agent evaluation and benchmarking | Embeddings and local inference (ONNX, llama.cpp) | Vector search

LEADERSHIP

Technical leadership | Leading and mentoring engineers | Standing up new teams | Hiring and running contractors | Architecture ownership

PROGRAMMING LANGUAGES

Python | C++ | Clojure | TypeScript | JavaScript

PLATFORMS & INFRASTRUCTURE

Node.js | React | Docker | Kubernetes | AWS | CI/CD | Git/GitLab

DATA & MESSAGING

PostgreSQL | Redis | SQLite | FoundationDB | ElasticSearch | MongoDB | Kafka | Pulsar | RabbitMQ

HOW I WORK

Benchmark-driven iteration | Scalable, reliable production systems | Testing (unit, integration, e2e, generative)

WORK EXPERIENCE

France

Founder & Lead AI Engineer: Klar (klar.im)

01/2026 - Current

In production on the App Store: on-device, privacy-preserving spam filter for Apple Mail.

- Trained the classifier in Python: a four-class head on a frozen XLM-RoBERTa encoder, quantised to run on the Mac's own GPU in under 50ms a message, with no network access.
- Built the C++17 engine it runs on (llama.cpp/GGML) behind a stable C API, with Node and Swift bindings, a Postfix milter and Docker end-to-end tests.
- AI tooling: custom memory and retrieval (MCP, LangChain, RAG, agent hooks and skills), cross-model adversarial review and supervisor/worker system.

Narval (GM NFT) Paris, France

Staff developer

06/2025 - 12/2025

Joined as early engineer (team of 4) when no product existed.

- Built custodian integrations (BitGo, Zodia) enabling institutional custodians to access DeFi (Aave staking, TON transactions), with GNAP/OAuth authn/authz on an OPA policy engine.
- Built transaction decoding/proxy infrastructure for on-chain transaction validation (EIP-7730).

Griffin Bank (BaaS) London, United Kingdom

Senior software engineer

04/2022 - 05/2025

Built platform components a UK banking licence depended on, in a heavily regulated environment.

- Led several new platform initiatives end-to-end, and helped build out new engineering teams as the org scaled toward its licence.
- Implemented secure admin authorization, a webhooks architecture, and passkey-based multi-factor authentication.
- Designed the queuing and processing systems, audit logging, scheduled-job infrastructure, and HTTP message signing behind it.

● **TurtleQueue (SaaS) London, United Kingdom**

Independent Developer

2019 - 2022

Led development of TurtleQueue, a Pulsar-compatible message broker SaaS platform.

- Took the product 0-to-1: designed and built the entire technology stack from scratch; secured R&D tax incentives.
- Hired and led a small team of contractor engineers (Poland) to scale delivery.
- Ran freelance contract engineering for MailOnline in parallel.

● **MailOnline (Media) London, United Kingdom**

Backend team lead software engineer

2016 - 2019

Promoted from Senior Engineer to Backend Team Lead, managing a team of experienced developers.

- Architected and scaled the core publishing pipeline to 1.3B views a month.
- Implemented ad placement that increased revenue; delivered real-time analytics that reduced feedback cycles.

● **Go Find It (Startup mobile apps) France**

Software engineer

2015 - 2016

Built a full-stack mobile CMS for luxury hotels.

- Ran the pipeline from development through DevOps to App Store publishing for a team of 10 mobile developers, with Python SSH deployment tooling.

● **Apizee (Startup web telecommunications) France**

Software engineer

2014

First backend engineer hire.

- Took a prototype 0-to-1: connectors bridging WebRTC customers with legacy phone systems, which secured a multi-year enterprise contract with Orange.
- Technology demoed at Mobile World Congress Barcelona, on HP's booth.

● **Siléane (Robotics SMB) France**

Intern

2012

- Prototyped 0-to-1 a controller for a Fanuc industrial robotic arm driven by Microsoft Kinect vision, so non-technical users could program movement paths by gesture.

● **Dassault Aviation (equipments division) France**

Intern

2010

- Developed an automated debugging bench to stress-test military avionics flight-control circuit boards, enabling 24-hour testing cycles that cut testing time by 50%.

EDUCATION & TRAINING

● **2011 - 2014** Lannion, France

Master's degree in computer science ENSSAT

Student representative. Erasmus exchange in Germany, with classes taught in German.

● **2010 - 2011** Edinburgh, United Kingdom

Bachelor of Engineering in Electronic and Computer Engineering Edinburgh Napier University

Coursework included C, embedded/real-time systems design, and state machines.

LANGUAGE SKILLS

MOTHER TONGUE(S): French | **OTHER LANGUAGE(S):** English