

PROFIL

Ingénieur logiciel depuis plus de 8 ans, Senior AI Engineer et responsable IA chez Neoteem (éditeur d'un ERP immobilier) depuis novembre 2025. Je conçois et mets en production des assistants IA utilisés au quotidien par les gestionnaires, avec une gouvernance stricte : données métier en lecture seule, validation humaine avant toute action engageante, évaluation systématique avant livraison. Je porte la stratégie IA auprès de la direction, j'encadre le pôle IA et je forme les équipes techniques et métier.

7

agents IA en production

-33%

de coût d'inférence

0

action engageante sans validation humaine

3 à 4

personnes encadrées

COMPÉTENCES CLÉS

- Agents et IA générative** : LangGraph, LangChain, CrewAI, agents ReAct, systèmes multi-agents, human-in-the-loop, Gemini (Vertex AI), Claude, prompt engineering.
- RAG et données** : RAG hybride (plein texte et pgvector, fusion RRF), reranking, embeddings, Qdrant, ingestion de documents (Unstructured).
- Évaluation et LLMOps** : DeepEval (LLM-as-judge), golden sets, observabilité Langfuse (traces, coûts, latence).
- Qualité logicielle** : Pytest, Vitest, Jest, mutation testing, architecture hexagonale, CI/CD.
- Back-end et cloud** : Python (FastAPI, Pydantic), TypeScript (Bun, Hono, Zod, Node.js), PostgreSQL, OpenAPI, GraphQL, Google Cloud (Cloud Run), Docker.
- Outils et front-end** : Claude Code, serveurs MCP (FastMCP), n8n, React, Next.js, Storybook.

EXPÉRIENCE PROFESSIONNELLE

Senior AI Engineer & Responsable IA · Neoteem Nov. 2025 – aujourd'hui
Éditeur d'un ERP pour les professionnels de l'immobilier · CDI, télétravail.

Stratégie et leadership

- Architecture en assistants spécialisés plutôt qu'en assistant unique : proposée à la direction et adoptée en production.
- Notes de décision pour la direction : organisation du pôle IA, choix d'architecture (recherche documentaire ou base de connaissances structurée), langage de programmation du pôle.
- Encadrement du pôle IA (3 à 4 personnes) et formation des équipes techniques et métier aux outils IA et à leurs bonnes pratiques.

3 assistants IA en production (7 agents)

- NeoChat** : répond aux gestionnaires à partir de leurs données, courriels et documents. Un agent ReAct couvre 42 vues métier avec seulement 5 outils génériques.
- NeoSupport** : répond aux questions sur le logiciel à partir de sa documentation (RAG hybride). Il demande une précision si la question est ambiguë et signale les sujets non documentés plutôt que d'inventer une réponse.
- NeoAgent** : compare les devis PDF et signale leurs incohérences (extraction multimodale), et rédige des annonces immobilières et des documents.
- Gouvernance dès la conception : données métier en lecture seule, envoi de courriel et suppression de document soumis à 5 points de validation humaine (interruptions LangGraph).

Qualité et coûts sous contrôle

- Coût d'inférence réduit de 33 % par le context caching. Chaque conversation tracée dans Langfuse, coût et latence suivis par agent.
- Aucune version livrée sans évaluation : golden set de 64 cas rejoués 5 fois (DeepEval, LLM-as-judge) et plus de 7 000 tests automatisés.

Ingénierie et outillage

- Couche de données des agents réécrite en TypeScript (Bun, Hono, Zod, PostgreSQL) à la place de fonctions SQL historiques, en architecture hexagonale vérifiée en CI.
- Score de mutation minimal de 80 % exigé avant chaque fusion, pour mesurer la capacité des tests à détecter les régressions.
- Claude Code déployé pour l'équipe (règles, sous-agents et garde-fous communs sur 5 dépôts). 2 serveurs MCP développés en interne relient les agents à la base de connaissances.

Développeur full-stack et IA · Indépendant (EI)

Déc. 2022 – nov. 2025

Mission longue chez Royal Canin, puis projets d'IA appliquée publiés en open source.

Développeur front-end senior · Royal Canin (via Ekino)

Déc. 2022 – avr. 2025

- Site de commerce en ligne déployé dans plus de 100 pays : optimisation des parcours B2B et B2C et des performances.
- Migration complète du code de JavaScript vers TypeScript, passage aux React Hooks, tests Jest et Testing Library.
- Alignement de la refonte technique sur la feuille de route avec le PMO et les product owners.

Projets d'IA appliquée, en open source sur GitHub

Mars 2025 – nov. 2025

- **mcp-server-documentation** : serveur MCP (FastMCP) qui donne aux agents un accès unifié et actualisé à la documentation de LangChain, LangGraph, Next.js et Tailwind CSS.
- **coder-crew** : 4 agents CrewAI (lead, back-end, front-end, tests) qui conçoivent, codent et testent ensemble une application, avec un LLM par rôle et des coûts suivis avec AgentOps.
- **job-booster** : application RAG (FastAPI, Next.js, LangChain, Qdrant, reranker BGE) qui rédige des courriels et des lettres de motivation à partir d'une offre d'emploi, avec plusieurs LLM au choix et des traces Langfuse.
- Autres projets : agents LangGraph et OpenAI Agents SDK, API FastAPI et Node.js en Clean Architecture, automatisations n8n.

Développeur full-stack · Schoolmouv

Mars 2022 – déc. 2022

Plateforme éducative interactive · CDI, télétravail.

- Conception et développement du MVP de zéro (React, Next.js, TypeScript), en architecture hexagonale.
- Bibliothèque de composants Storybook, API REST Express documentée avec Swagger.
- Encadrement technique de développeurs juniors et coordination avec le PMO et l'équipe produit (priorités du MVP, jalons, livrables).

Développeur full-stack · Aixia Systèmes (pour Leasys Rent)

Sept. 2018 – mars 2022

Refonte des outils numériques d'un loueur de véhicules · CDI, télétravail partiel.

- Création, dès 2018, d'une PWA React (MUI) d'analyse automatique de documents avec l'API Google Vision.
- Refonte complète du front-end et création d'un BFF GraphQL (Node.js, Apollo) devant les microservices.
- Mise en place d'un monorepo Lerna et de la CI/CD GitLab pour une équipe de 7 développeurs, puis migration des API de Symfony 3 à Symfony 5 et sécurisation.

VEILLE ET OPEN SOURCE (2026)

- **claude-forge** : générateur de configurations Claude Code vérifiées automatiquement, avec des hooks bloquants et un serveur MCP de 22 outils (FastMCP).
- **agentsdir** : outil en ligne de commande TypeScript (licence MIT) qui unifie les fichiers d'instructions AGENTS.md de Codex, Cursor et opencode, les convertit pour Claude Code et les vérifie en CI.
- **wow-forever** : assistant expert d'un jeu vidéo où le LLM ne calcule jamais, chaque chiffre venant d'un outil déterministe (serveur MCP de 8 outils, plugin Claude Code).

FORMATION ET CERTIFICATS

- **Concepteur développeur web**, titre RNCP de niveau 6 (bac + 4) · O'clock, 2018–2019.
- **Développeur logiciel**, titre RNCP de niveau 5 (bac + 2) · O'clock, 2018.
- **Anthropic Academy** : *Claude Code in Action*, *Claude with the Anthropic API*, *Introduction to Agent Skills*, *Introduction to Model Context Protocol*.
- **Hugging Face / Google** : *AI Agents Fundamentals*.

LANGUES

- **Français** : langue maternelle.
- **Anglais** : B1, lecture technique courante (documentation lue et formations suivies en anglais).