Jafet Brito Perez

Ingénieur IA · Intégration de LLM et automatisation

Sudbury (Ontario), Canada · À distance (heure de l'Est) · (249) 532-1984

jafetbrito028@protonmail.com LinkedIn GitHub jafetbrito.com jafetbrito.art jafetbrito.blog

PROFIL

Développeur avec plus de 10 ans d'expérience en développement logiciel, aujourd'hui centré sur la mise en production de l'IA. Chez Amazon, j'ai construit le pipeline automatisé d'évaluation des LLM d'Alexa et standardisé leur affinage. Créateur d'Oliver Academy, une plateforme d'apprentissage gamifiée dont le tuteur IA, les personnages et le mode agent fonctionnent grâce à une passerelle que j'ai construite pour plus de 7 fournisseurs de modèles (Claude, OpenAI, Gemini, DeepSeek et d'autres). Je développe aussi des bots WhatsApp et Telegram et des automatisations pour les entreprises. Espagnol natif, anglais professionnel. Disponible pour des missions à distance ou un emploi.

COMPÉTENCES

Modèles et IA : API Anthropic Claude, API OpenAI, Gemini, DeepSeek, Groq, OpenRouter, Ollama (modèles locaux), prompt engineering, agents avec outils, conception multi-agents, Claude Code, LangChain, LangGraph, LangSmith, Hugging Face, AutoGen, CrewAI, DSPy, Llama 3, RAG, fine-tuning, PyTorch, TensorFlow, Ray, MLflow, Pinecone, Amazon SageMaker, Google Vertex AI

Automatisation : n8n, scripts Python, WhatsApp (whatsapp-web.js), API Telegram Bot, API Notion, automatisation des e-mails

Backend et données : Supabase (PostgreSQL, Row Level Security, Auth, Edge Functions en Deno/TypeScript), Node.js, API REST, MySQL, MongoDB

Frontend et déploiement : React 19, Vite, TypeScript, Tailwind CSS, Git/GitHub, Vercel, Hostinger, Docker/Podman, Linux

EXPÉRIENCE

Fondateur et seul développeur — Oliver Academy Juin 2026 – Aujourd'hui

Plateforme d'apprentissage gamifiée et bilingue (espagnol/anglais) avec plus de 20 cours : IA, cybersécurité, Python, Bash, droit, médecine, philosophie, nahuatl et plus. Démo en direct sur demande.

`React 19 · Vite · Supabase · Deno Edge Functions · Three.js · Phaser · Tailwind 4 · PWA`

- Création d'une passerelle IA côté serveur (Supabase Edge Function) qui sert plus de 7 fournisseurs via une seule interface : 4 adaptateurs natifs et une voie compatible OpenAI.
- Conception « apportez votre propre clé » : la clé de chaque utilisateur est stockée sous sécurité au niveau des lignes et n'est jamais journalisée ; le passage des appels côté serveur a résolu les blocages CORS du navigateur.
- La passerelle alimente le tuteur IA (« L'Oracle d'Oliver »), des personnages IA dans la salle de classe 3D et un mode agent qui utilise des outils.

- Création de cours d'IA (prompt engineering, méthode d'étude avec NotebookLM, Claude pour les seniors) et d'un jeu de duel de prompts.
- Assistant de réunion IA qui enregistre les sessions Zoom, Teams, Meet et en présentiel (AWS Lambda, S3) et utilise Llama 3 pour générer des notes, des e-mails de suivi et des tâches CRM, avec relecture humaine.
- Agents de recommandation avec Llama 3 affiné, génération augmentée par récupération sur Pinecone et un système de débat multi-agents avec LangGraph.
- Cadre décentralisé d'apprentissage par renforcement multi-agents (MARL) avec Ray/RLlib et PettingZoo, utilisant QMIX et MADDPG.
- Connecteurs MCP pour les CRM Salesforce, Wealthbox, Redtail et Microsoft Dynamics.

Fondateur et développeur — Rubic's Digital Solutions Sept 2023 – Aujourd'hui

- Je dirige mon studio, **rubicsdigitalsolutions.ca** : j'ai conçu et développé son site et je réalise des projets web, d'automatisation et de sécurité pour de petites entreprises.
- Création d'« Athena », mon assistant d'opérations IA : agents sur modèles de langage, scripts Python et API Notion pour les tâches, la recherche et les listes de prospects.

Développeur senior IA et full-stack — Amazon Févr 2021 – Juil 2023

- Pipeline automatisé d'évaluation de LLM comparant chaque semaine plus de 100 variantes de modèles et de prompts ; réduction de 95 % de la validation manuelle et contrôle des lancements d'Alexa+ sur 20 indicateurs.
- Standardisation de l'affinage des LLM internes, améliorant 8 indicateurs conversationnels.
- Conception d'un système de recommandation de contenu multi-appareils déployé sur 6 expériences vocales et à écran.
- Pilotage de la vision technique à 2 ans d'Alexa Subscriptions sur 8 lignes de produits et plus de 50 ingénieurs.

Développeur full-stack — Tecsolt Févr 2018 – Déc 2020

- Pipelines de données événementiels avec Kafka et Redis Streams pour l'analytique et le suivi des comportements.

Développeur full-stack junior — Laserfiche Janv 2016 – Déc 2017

- Intégration d'API de recommandation par apprentissage automatique dans les applications clients.

Développeur web et designer — Waa.mx (2022 – 2024) : développement et maintenance de sites et de boutiques en ligne de clients, du design visuel au front-end et au back-end.

Développeur web et designer — Libra Digital (2021 – 2022) : création de sites d'événements et de campagnes comme TEDx Cuernavaca 2022 et Feria del Terror (WordPress, PHP, JavaScript) ; SEO/SEA et support technique.

Développeur logiciel — Raxione · Développeur web — ITSoftAndHardware (2020 – 2021) : développement front-end, documentation de modules et tests qualité.

AUTRES PROJETS

- **Congrès des 28** (système multi-agents) : 28 agents IA spécialisés répartis en 4 chambres (stratégique, technique, créative et psychosociale), chacun avec sa personnalité et ses compétences en JSON, qui débattent et planifient des décisions d'entreprise.

- **PromptForge et Base de Gatos** (prompt.base-de-gatos.online) : outils de prompt engineering pour construire des prompts structurés et sémantiques, concevoir des agents et gérer un coffre-fort de prompts classés et notés, avec import/export.
- **Bot de prospection WhatsApp** (Node.js, whatsapp-web.js) : messages personnalisés et campagnes vers des prospects à partir de fichiers CSV.
- **Bitácora Troula** (API Claude, Vercel serverless) : planificateur d'itinéraires en camping-car avec IA ; un proxy garde la clé côté serveur.
- **jafetbrito.blog** : blog technique bilingue avec 17 articles de fond, dont des guides complets sur le prompt engineering, Claude, NotebookLM et l'apprentissage de l'IA depuis zéro.

FORMATION ET LANGUES

- Licence en intelligence artificielle — UAEM, Morelos, Mexique (2011 – 2015)
- Harvard CS50x (informatique) et CS50P (Python) · Platzi (plus de 50 cours : web, Python, IA, cybersécurité) · Claude Code (2026) · TryHackMe (en cours)

Langues : Espagnol (natif) · Anglais (professionnel) · Français (notions)