



Billy Girboux

Architecte IA & Systèmes Agentiques · Tech Lead Full-Stack · 18 ans d'exp.

Mobilité France entière · Ouvert au déménagement +33 6 51 01 99 46
contact@billygirboux.fr billygirboux.fr linkedin.com/in/billy-girboux

PROFIL

Architecte logiciel et Tech Lead Full-Stack avec plus de 18 ans d'expérience, dont les dernières années centrées sur l'**IA générative et les systèmes agentiques en production**. J'ai entraîné un LLM **from scratch**, industrialisé des boucles d'apprentissage autonomes, opéré du **serving GPU** (vLLM, CUDA) et conçu des architectures **RAG** avec garde-fous. J'ai évolué jusqu'au rôle de **CTO** au sein de startups SaaS, pilotant aussi bien les équipes engineering que la vision technique et produit.

Je couvre l'ensemble de la chaîne : dataset, entraînement, inférence, backend (Java/Spring, Node.js, Python), infrastructure et déploiement. Rompu aux environnements à forte contrainte, je suis aussi à l'aise pour structurer une architecture que pour descendre dans le code.

Ce qui m'intéresse dans l'IA n'est pas le POC mais l'**industrialisation** : le différenciant n'est presque jamais le modèle, c'est la chaîne autour : dataset, coût d'inférence, latence, fallback quand le modèle se trompe, et capacité à écarter les usages IA qui n'apportent rien.

COMPÉTENCES TECHNIQUES

IA, LLM & AGENTIQUE

- **Entraînement** : from scratch (PyTorch), fine-tuning qLora
- **Serving & inférence** : vLLM, Ollama, CUDA, GPU dédié
- **Systèmes agentiques** : boucles autonomes, orchestration, garde-fous
- **RAG**, prompt engineering, évaluation de modèles
- **Industrialisation** : pipelines de données, MLOps, coûts d'inférence

BACKEND

- **Java** : Spring Boot, JPA
- **Node.js** : Express, APIs REST
- **Python** : PyTorch, pipelines IA, scripting
- **API RESTful**, messagerie async.

FRONTEND & MOBILE

- **JavaScript / TypeScript** : React, Ember.js
- **Flutter / Android** : cross-platform
- Déploiement Google Play / App Store

INFRASTRUCTURE & DEVOPS

- **Virtualisation** : Proxmox, ESXi
- **Réseau & sécurité** : OPNSense
- **Conteneurisation** : Docker, Kubernetes
- **Administration Linux** (Debian)
- **CI/CD**, automatisation infra
- **Monitoring** : Grafana

DONNÉES & ARCHITECTURE

- **SQL** : MariaDB, MySQL, PostgreSQL
- **Elasticsearch, Redis**
- **Messagerie** : Kafka, RabbitMQ
- **Architecture** SOA / microservices / monolithique

LANGUES & DIVERS

- Français : langue maternelle
- Anglais : professionnel
- Documentation technique
- Formateur Java / JS / Android

RÉALISATIONS IA & SYSTÈMES AGENTIQUES

Philidor · LLM entraîné from scratch Publié

2026

Modèles 51M et 142M de paramètres · Entraînement sur GPU grand public

Entraînement d'un modèle de langage de zéro, sans lui fournir aucune règle du jeu, pour mesurer ce qu'il déduit seul. Pipeline de données complet : 89,3 M de parties lues, 11 M conservées après filtrage (12,4 % de rétention, ~800 M de tokens), protocole d'évaluation sans fuite train/val. Résultat : **98 victoires sur 100 face à un modèle 681 fois plus gros** (Qwen3.6-35B), soit +798 points d'Elo. Modèle déployé en bot sur Lichess : 691 parties classées contre des joueurs réels, 1719 Elo en rapide, zéro coup illégal. Démarche, mesures et intégralité du code publiés et reproductibles.

Article : billygirboux.fr/fr/blog/modele-ia-echecs-weekend · Bot : lichess.org/@/philidor-142M · Code : github.com/geekourson/philidor

PyTorch Dataset engineering CUDA Évaluation Spécialisation vs taille

Curvy · régression symbolique, du nuage de points à la formule En ligne

2026

Transformer encodeur-décodeur 31,6 M de paramètres · Démo publique servie sur GPU

Modèle qui rend la **formule mathématique** d'une courbe tracée à la main. Parti pris d'architecture : le réseau prédit une structure, un optimiseur numérique remplit les constantes. Décodage sous **masque d'arité** : 0 % de sorties invalides par construction. Jeu de test figé avant entraînement pour mesurer de la généralisation et non de la restitution : **68,4 % de formules jamais vues retrouvées** (plafond mesuré 79,1 %) et **2,7x mieux qu'un polynôme en extrapolation**. Gain final de 4,9 points obtenu sans réentraîner (beam search élargi, initialisation spectrale, cache clé/valeur), pour une latence médiane de 779 ms en service. Code publié sous licence MIT.

Article : billygirboux.fr/fr/blog/retrouver-la-formule-dune-courbe · Démo : billygirboux.fr/curvy · Code : github.com/geekourson/curvy

PyTorch Transformer Beam search Optimisation d'inférence Service GPU Open source

Hello Wellington · boucle agentique de spécialisation continue

En production

2025

Application IA conversationnelle · Fondateur & développeur solo

Chaque conversation alimente le dataset et déclenche un réentraînement **qLora sans intervention humaine** : boucle autonome de spécialisation du modèle sur les usages réels. Infrastructure GPU dédiée (CUDA), serving **vLLM**, personas IA personnalisables, backend Java/Spring, mobile Flutter, DevOps complet. Choix structurant : auto-héberger le modèle plutôt que dépendre d'une API tierce, pour maîtriser coûts unitaires, données et réentraînement.

qLora · vLLM · CUDA · RAG · Prompt Engineering · Flutter · Java Spring · Docker

Teelt.io · IA générative embarquée dans un SaaS RH

2022 · 2026

En tant que CTO · Du cadrage à la production, avec l'équipe engineering

Identification des cas d'usage à valeur réelle (et arbitrage de ceux à écarter), architecture **RAG sur les contenus clients**, prompt engineering et garde-fous, mise en production et suivi des coûts.

EXPÉRIENCES PROFESSIONNELLES

CTO

2022 · Juil. 2026

Teelt.io · Solutions SaaS Onboarding & Offboarding · Télétravail

Direction technique complète. Conception de la plateforme, gestion de l'infrastructure, encadrement technique et architecture. Intégration de l'IA générative dans le produit (RAG, prompt engineering, garde-fous). Poste ayant pris fin suite à la liquidation judiciaire de la société en juillet 2026.

Lead Développeur & Co-fondateur

2018 · 2022

Fragmnt.io / Soyoutrip · Solutions informatiques · Sophia Antipolis

Création de Soyoutrip : application de planification de voyages avec optimisation géographique des itinéraires. Architecture complète, développement full-stack, déploiement en production.

Lead Technique / Architecte Logiciel · CIO

2014 · 2018

Option Way · Agence de voyage en ligne · Sophia Antipolis

Création from scratch de l'application web. Stack complet : backend Java/Spring, frontend Ember.js, batches. Développement d'un agent de voyage virtuel intelligent. Référent technique pendant 4 ans.

Ingénieur Intégrateur

2013

Login People · Authentification forte · Sophia Antipolis

Développement d'une solution d'authentification forte (technologie ADN du numérique). Outils intranet et extranet.

Ingénieur Mobile Android

2012

AndroMC Systems · Accessoires mobiles · Nice

Application Android couplée à un accessoire USB : éthylotest intelligent avec services de accompagnement intégrés.

Ingénieur Étude et Développement

2008 · 2012

Humanis / Aprionis · Mutuelles · Blois / Malakoff

Développement applicatif : tarificateurs, extranets, outils de gestion de contrats.

PROJETS PERSONNELS

Munr · plateforme SaaS d'observabilité d'infrastructure

En production

2025

Fondateur & développeur solo · [munr.io](#)

Conception et développement en solo d'une plateforme d'observabilité hébergée en France : logs, métriques, alertes et erreurs centralisés dans un seul outil. Architecture complète, développement full-stack, **agent système** déployé sur les machines supervisées, infrastructure et billing. Alternative souveraine à Datadog, non soumise au CLOUD Act.

Observabilité · SaaS · Docker · Linux · Full-Stack · Infra · Souveraineté numérique

Seryz & Maison Nourson · produits livrés en solo, l'IA comme levier

En production

2026

Seryz : journal musculation et HIIT iOS/Android (React Native), 100 % hors-ligne, 200+ exercices, 6 langues. Maison Nourson : deux jeux publiés sur Steam (Java/libGDX), du moteur de rendu à la publication. Livrés seul, en déléguant les parties mécaniques aux LLM et en gardant la main sur l'architecture et les arbitrages produit.

FORMATION

PÉRIODE	DIPLÔME / SPÉCIALITÉ	ÉTABLISSEMENT	LIEU
2006 · 2008	Ingénieur en Informatique	Polytech Nice-Sophia Antipolis	Nice
2004 · 2005	Classes Préparatoires MPSI / MP	Lycée Louis Thuillier	Amiens
2000 · 2003	Baccalauréat Mathématiques	Institution du Saint-Esprit	Beauvais