

Chakib Benziane

AI/LLM Ops | Développeur Backend | Programmeur Systèmes



consulting@blob42.xyz blob42.xyz Github Gitea Twitter

Profile

Programmeur systèmes et réseaux UNIX chevronné, spécialisé dans les opérations AI/LLM et le développement backend. Expert dans la création d'applications CLI et serveur haute performance utilisant Go, Rust et Python, avec un accent particulier sur le déploiement d'IA on-premise, les environnements air-gapped et les infrastructures auto-hébergées. Contributeur open-source dévoué avec une vaste expérience de collaboration au sein d'équipes mondiales distribuées.

Avec plus de 15 ans d'expérience en SysAdmin/DevOps, Architecture IT, OpSec et Machine Learning, j'ai fait mes preuves dans la création de valeur pour une clientèle diversifiée, allant des entreprises mondiales aux startups agiles. Ma polyvalence en développement backend et frontend, combinée à un esprit analytique rigoureux, me permet d'exceller dans les phases initiales de projets rapides et à fort impact.

Passionné par l'architecture de piles logicielles robustes et le prototypage de MVP avec des technologies de pointe ou spécialisées. Mes domaines de concentration actuels incluent le fine-tuning de LLM, l'optimisation de pipelines RAG, les modèles de diffusion et le développement d'outils pour des flux de travail IA sécurisés et hors ligne.

Experience

Ingenieur IA - Programmeur Backend Systèmes

Dec 2021 - Present

Contractuel Indépendant • Remote uniquement (Espagne)

Disponible pour des missions en freelance et contractuelles : programmeur backend / systèmes.

- Conception de systèmes backend robustes utilisant des piles Go, Rust et Python
- Architecture de piles logicielles évolutives adaptées aux besoins métier spécifiques
- Consulting expert en AI/LLM, axé sur les déploiements on-premise, le fine-tuning de modèles et l'outillage personnalisé
- Contributeur précoce de la communauté open-source LangChain, implémentation de l'outil SearxNG et de diverses fonctionnalités core

Directeur UX

Jan 2019 - Dec 2019

Client Confidentiel • Remote

Gestion de produit et consulting en design UX.

- Direction d'initiatives de gestion de produit et planification stratégique
- Exécution de recherches utilisateurs et design UX complets pour optimiser l'utilisabilité du produit
- Facilitation de la communication avec les parties prenantes et recueil précis des besoins

Designer / Développeur UI

Dec 2014 - Dec 2018

Contractuel Indépendant • Berlin (Allemagne)

Travail sur divers projets en tant que designer UI et développeur full stack.

- Développement de prototypes rapides et de MVP pour des startups en phase initiale, accélérant la mise sur le marché
- Conception de solutions pour des projets basés sur la cryptomonnaie et la blockchain
- Gestion et coordination d'équipes de développement multidisciplinaires

Développeur Full stack

Dec 2012 - Dec 2014

Freelance • Paris (France)

Création et déploiement de diverses applications web et mobiles pour des clients.

- Rédaction de spécifications techniques détaillées et architecture de designs d'applications

- Développement et déploiement d'applications web et mobiles full-stack
- Gestion du déploiement de bout en bout et maintenance système à long terme

CTO / Développeur Full stack

Dec 2012 - Dec 2014

Boxls (Startup) • Paris (France)

R&D pour une solution de sécurité basée sur le DNS.

- Développement d'un serveur DNS personnalisé utilisant Twisted Python pour le réseau haute concurrence
- Implémentation de drivers Windows Filtering Platform (WFP) en C#
- Gestion du déploiement de l'infrastructure et de la stabilité du système
- Développement d'applications full-stack utilisant le framework Django
- Mise en place de pipelines CI/CD utilisant uWSGI et Nginx

Ingénieur Systèmes / DevOps

Sep 2011 - Dec 2012

Smartjog • Paris (France)

Recherche & Développement sur la virtualisation et les systèmes de fichiers distribués.

- SRE pour une infrastructure de virtualisation à grande échelle, assurant une haute disponibilité
- R&D sur les systèmes de fichiers distribués, incluant Ceph, Sheepfs et Glusterfs
- Recherche et implémentation de solutions de virtualisation utilisant KVM, Libvirt et QEMU
- Co-développement d'un outil de gestion de VM basé sur Libvirt
- Conception d'un serveur de clés PGP LDAP entièrement conforme utilisant Python

SRE / Réseaux / DevOps

Sep 2009 - Sep 2011

Bocal (École Epitech) • Paris (France)

Dans le cadre d'un stage à temps partiel, j'ai fait partie d'une équipe gérant une infrastructure à grande échelle pour l'école.

- Gestion de l'administration réseau et système pour un campus de plus de 1 000 étudiants

- Conception d'un système dual-boot personnalisé pour répondre aux divers besoins des étudiants
- Développement d'un système de déploiement Windows automatisé basé sur Sysprep et KMS
- Création d'un pipeline de déploiement Linux personnalisé utilisant PXE et Filezilla

Projects

Contributeur Open Source précoce de LangChain

Feb 2023

- Implémentation de l'outil SearxNG, qui a depuis été intégré dans de nombreux frameworks AI en aval
- Contribution de fonctionnalités core, d'améliorations de performance et de corrections de bugs critiques lors des premières phases du projet

Autres Contributions Open Source

Jun 2012

- Contribution à un large éventail de projets open-source, avec un focus récent sur l'écosystème AI/LLM, incluant LocalAI, aichat, StableDiffusion, ChromaDB et PySyft

Instrukt: Environnement AI TUI pour agents LLM (Développeur Unique)

Jun 2023

- Activation d'instructions en langage naturel pour les agents AI
- Facilitation du développement d'agents spécialisés équipés de jeux d'outils personnalisés
- Implémentation du scan et de l'indexation automatisés des sources de données pour alimenter le question-réponse basé sur RAG

Gosuki: Gestionnaire de favoris ultra-rapide, multi-navigateurs et sans plug-in. (Développeur Unique)

Sep 2017

Un gestionnaire de favoris ultra-rapide, multi-navigateurs et sans plug-in.

- Élimination du besoin d'extensions de navigateur grâce à une intégration innovante
- Support de tous les navigateurs majeurs et de multiples profils par navigateur en temps réel

- Fourniture d'une portabilité des données flexible avec plusieurs formats d'importation et d'exportation
- Implémentation de la synchronisation multi-appareils et de l'archivage permanent des liens web

PSWatch

Jul 2025

Un surveillant de processus système et un ordonnanceur de tâches avec des règles de correspondance personnalisées.

- Fournit une surveillance des processus en temps réel pilotée par des règles de correspondance personnalisables
- Intègre l'ordonnancement de tâches et l'automatisation système

HugoBot

Feb 2018

Bot de scraping et d'agrégation de contenu pour sites web Hugo basés sur des données.

- Automatise le scraping et l'agrégation de contenu pour les sites web basés sur des données
- Génère des sorties de données optimisées compatibles avec le générateur de sites statiques Hugo

GUM

Apr 2027

Micro-bibliothèque de gestion de tâches et d'unités.

- Fournit un framework minimaliste pour la gestion et l'orchestration de tâches

codegpt-ng.nvim

May 2026

Plugin Neovim AI léger se connectant à des backends comme Ollama et OpenAI.

- Intègre la complétion de code et l'assistance propulsées par l'AI directement dans Neovim
- Supporte une large gamme de backends, incluant Ollama et OpenAI

Autres Projets

Jan 2024

Projets open source et contributions additionnels.

- Curateur de 'Awesome Explorables', une liste complète d'explications interactives et explorables

- Développement de 'aichat-ng', un fork amélioré de aichat avec des capacités avancées, disponible en tant que paquet AUR
- Création de 'tinyssh-initramfs', un initramfs minimal avec SSH pour un accès système d'urgence sécurisé

Education

Bachelor in Informatique

2009 - 2011

École d'ingénierie informatique Epitech (Paris)

Relevant Coursework: C, C++, Unix, Réseaux, Sécurité, Web, Bases de données, Conception Logicielle, Gestion de Projet

Bachelor in Mathématiques & Informatique

2007 - 2009

Université Paris 6

Certifications

Udacity Deep Learning Nanodegree

Sep 2017

Udacity

Volunteer Work

Formateur & Aide

Nov 2007 - Nov 2008

Ubuntu France

- Participation à des événements publics pour promouvoir Linux et les logiciels open-source.
- Fourniture de support gratuit aux nouveaux utilisateurs de Linux.

Skills

AI/LLM Ops

Déploiement LLM on-premise • vLLM • Llama.cpp • LocalAI • StableDiffusion •

Machine Learning & AI

LLMs • Deep Learning • PyTorch • Langchain • TensorFlow • llama.cpp •

ComfyUI • Open WebUI • ChromaDB • RAG •
N8N/LocalAGI • llama-swap • kokoro-tts •
Serveurs MCP • Gestion du cycle de vie LLM
de bout en bout • Observabilité LLM et
développement de pipelines proxy
personnalisés

Développement Backend / Systèmes

Go • Python • Rust • C • Javascript • Bash

Linux

Debian • iptables • systemd/SysVinit • pam •
Optimisation et réglage du Kernel •
optimisation des performances I/O du
système de fichiers • Architecture avancée de
systèmes de fichiers, partitionnement et
redondance • Boot/Grub/UEFI

Sécurité

SSO/LDAP • Gestion des clés API • OpSec •
air-gap

Bitcoin et Lightning Network

Bitcoin • Lightning Network • Bitcoin Core •
LND • C-Lightning

Langues

Anglais

Courant Français

Langue maternelle

Arabe

Langue maternelle

embedding/reranking • modèles de diffusion

DevOps

Docker • Kubernetes • Podman • Conception
de topologies réseau et cluster complexes •
réseaux overlay • pare-feu • Développement
d'outils CLI et de démons système
personnalisés • Compétences expertes en
dépannage • Débogage et patching de
logiciels upstream

Réseaux

Caddy • iptables/nftables • réseaux overlay •
Architectures réseau de conteneurs
sécurisées et isolées • Pratiques SRE

Écosystème Neovim

Lua • codegpt-ng.nvim • Développement
Neovim

Divers

Électronique DIY • Programmation embarquée
amateur • Arduino, ARM, ESP

Interests

**Compilateurs &
Interpréteurs**

langages de programmation

AI/LLMs & Deep Learning

Bitcoin & Lightning Network

**Électronique (Arduino,
Raspberry Pi)**