Adem Usta

Ingénieur logiciel senior · ML / MLOps

hello@ademoverflow.com
+33 6 46 46 55 94
adem.build
Paris, France



Ingénieur logiciel depuis 2017, de l'analyse vidéo temps réel au serving de modèles sur GPU. Je conçois, construis et exploite des backends, des pipelines ML et l'infrastructure qui les porte, avec un parti pris pour le code qui reste maintenable.

EXPÉRIENCE

oct. 2024 – aujourd'hui

Staff MLOps Engineer · Integral Ad Science

Salarié · Paris, France

Développement et maintenance de Total Media Quality (TMQ), un pipeline de traitement de données temps réel pour la publicité en ligne.

- Identification des goulots d'étranglement GPU et isolation dans un service dédié
- Conception d'une architecture de model serving pour scaler les ressources GPU, mise en œuvre avec Nvidia Triton
- Mise en conformité de la CI/CD avec les contraintes IAS, ajout de tests fonctionnels et d'intégration
- Amélioration de l'observabilité avec Prometheus et Grafana, et proposition d'alertes sur les anomalies de performance et de latence

Python · FastAPI · PyTorch · Nvidia Triton · Kubernetes · AWS · Kafka · Databricks · MLflow · Prometheus · Grafana

2022 – 2024

Tech lead / Ingénieur logiciel senior · Freelance

Freelance · Hybride, Paris

Expertise en ingénierie logicielle (backend, DevOps, frontend, machine learning) pour plusieurs entreprises.

- Ondorse : développement et maintenance d'un SaaS de KYB (Know Your Business)
- NielsenIQ : développement de ScanRewards, une application mobile de collecte de tickets de caisse
- AnotherBrain : stack de monitoring pour le produit Phosphor Quality
- Context : évaluation de Ray, un framework Python de calcul ML distribué

Python · FastAPI · PostgreSQL · React · TypeScript · Celery · Ray · Kubernetes · Prometheus · Grafana

févr. 2021 – juil. 2022

Ingénieur logiciel · Boxy (ex-Storelift)

Salarié · Ivry-sur-Seine, France

Dans l'équipe Store-AI, conception et amélioration des applications de Boxy, un magasin connecté entièrement autonome.

- Suivi d'événements temps réel du comportement client, avec logs et reporting pour la business intelligence
- Intégration de capteurs de température et application IoT de suivi des températures dans le cloud
- Conception de l'architecture microservices de Boxy v2
- Mise en place de procédures qualité (linting, tests, typage statique) automatisées dans GitLab CI et des Git hooks

Python · Kubernetes · MQTT · RabbitMQ · Raspberry Pi · Arduino · Fluentd · Kibana · GitLab CI

mars 2017 – févr. 2021

Ingénieur logiciel · Context

Salarié · Paris, France

Conception et développement de logiciels d'analyse vidéo temps réel par deep learning.

- Création d'un outil interne de gestion des jeux d'images et de vidéos, et modernisation des outils d'annotation
- Conception d'une API REST d'upload et d'analyse vidéo (détection de célébrités, de logos et d'objets), avec versionnage et déploiements blue-green
- Développement d'un tableau de bord client d'analytics vidéo
- Refonte du service de calcul central et passage aux microservices sur Kafka, déployés sur Kubernetes

Python · PyTorch · Django · FastAPI · React · Angular · PostgreSQL · Kafka · Kubernetes · AWS

FORMATION

2016 **Master 2, Image et son pour les systèmes intelligents**
Université Pierre et Marie Curie (Paris VI) · mention bien

2015 **Master 1, Ingénierie de la robotique et des systèmes intelligents**
Université Pierre et Marie Curie (Paris VI) · mention très bien

2014 **Licence, Ingénierie électronique**
Université Pierre et Marie Curie (Paris VI) · mention très bien

2010 **Baccalauréat scientifique, Sciences de l'ingénieur**
Lycée Le Corbusier (Aubervilliers) · mention bien

COMPÉTENCES

LOGICIEL	Python, TypeScript, FastAPI, Pydantic, Django, Temporal, React, Svelte, React Native, Expo, PostgreSQL, MySQL, Kubernetes, Docker, AWS, VPS, CI/CD, Clean architecture, TDD
IA AGENTIQUE	Agents LLM, Évaluation d'agents, vLLM, Nvidia Triton, PyTorch, TensorFlow, Ray, MLflow, MLOps, Mise à l'échelle GPU, Modèles auto-hébergés, Vision par ordinateur
HARDWARE	Raspberry Pi, ESP32, Arduino, Conception de PCB, Capteurs, C/C++, Home Assistant, I2C, MQTT, k3s, Prometheus, Grafana
ROBOTIQUE	ROS 2, C++, Python, Perception, Vision par ordinateur, Reconnaissance de formes, Commande, Traitement du signal, Conception mécanique, Cinématique, Haptique, Réalité virtuelle

LANGUES

Français (Langue maternelle) · Turc (Langue maternelle) · Anglais (Courant)

CERTIFICATIONS

Spécialisation Machine Learning · DeepLearning.AI, Université Stanford

Spécialisation Deep Learning · DeepLearning.AI, Université Stanford