

Florent Donat-Bouillud

Data Engineer
florent.donat.bouillud@gmail.com | +33 7 82 10 58 13 | LinkedIn | GitHub

PROFIL

Ingénieur Machine Learning, diplômé dun Master en Data-IA et traitement du signal, avec des compétences en développement de solutions IA et en traitement de données et de signaux. Je souhaite mettre ces compétences au service de projets dIA embarquée et de systèmes connectés, en contribuant au prototypage, à lintégration et à la validation de solutions sur plateformes matérielles. Particulièrement intéressé par les problématiques à linterface entre intelligence artificielle, logiciel et matériel.

COMPÉTENCES

Langages : Python, SQL, C++, C
IA générative & RAG : LLM, RAG, LangChain, Chroma (base vectorielle), embeddings Hugging Face, évaluation de la recherche (Precision, Recall, MRR, Hit Rate)
Machine learning : PyTorch, TensorFlow, scikit-learn, Transformers, ONNX, OpenCV
Développement & industrialisation : FastAPI, API REST, Git/GitHub, Docker, Linux, tests (pytest)
Audio & signal : Librosa, ffmpeg, JUCE, MIDI

EXPÉRIENCE

Stagiaire R&D (6 mois) – ML pour l'improvisation musicale 2025
SCRIME, Bordeaux

- Développement de modèles génératifs LSTM et Transformers pour la génération audio temps réel : latence d'inférence, stabilité d'exécution, qualité de sortie.
- Entraînement et évaluation de modèles sur des jeux de données de musique symbolique à grande échelle.
- Conception et optimisation de pipelines d'IA temps réel pour des applications interactives (faible latence, fiabilité à l'exécution).
- Étude des contraintes de déploiement de modèles en environnement interactif à ressources limitées.

Stagiaire R&D (6 mois) – Vision par ordinateur 2024
Alten, Sèvres

- Développement d'une solution de bout en bout pour la détection d'intention de piétons en milieu urbain.
- Implémentation, évaluation et déploiement de modèles CNN et Transformers (PyTorch, TensorFlow, OpenCV) au sein de pipelines de machine learning complets.
- Optimisation de l'efficacité de calcul en vue d'un déploiement orienté production.

Stagiaire NLP / traitement de données (4 mois) 2023
France Télévisions, Paris

- Constitution de jeux de données d'entraînement pour des modèles de traduction.
- Développement de pipelines d'ingestion et de prétraitement de données NLP et de bases de données relationnelles.

PROJETS

- **Système RAG pour la recherche de connaissances sur le football**
  Pipeline RAG local (LangChain, Chroma, embeddings Hugging Face, LLM local) : ingestion et découpage d'articles Wikipédia (FIFA, Coupe du monde), recherche par similarité, évaluation du retrieval par Precision, Recall, MRR et Hit Rate.
- **VocalTwin – Analyse articulatoire de la parole et validation vidéo**
  Réalisation d'une Application web (FastAPI) qui estime des landmarks articulatoires à partir d'un fichier audio (modèle SPARC) et les anime en synchronisation avec le son. Extraction audio avec ffmpeg, mise en cache des analyses par hash de fichier, mode de validation sur vidéo de profil. Tests automatisés (pytest), documentation d'architecture, de schéma de données et de limites.
- **Détection de deepfakes audio**
  Fine-tuning et déploiement d'un modèle Transformer HuBERT pour la classification de parole authentique ou synthétique.

FORMATION

Master Ingénierie des Systèmes Intelligents – Sorbonne Université, Paris 2023–2025
IA, ML appliqué à l'audio et à la vision, deep learning, apprentissage par renforcement, NLP

Master Robotique, Commande linéaire et IA – Sorbonne Université, Paris 2022–2023

Licence Électronique, Énergie électrique et Automatique – Sorbonne Université, Paris 2019–2022

LANGUES

Français : langue maternelle **Anglais :** C1 **Allemand :** B1