Sébastien Tadiello

Machine Learning Engineer

LLMs, RAG, NLP, Embedded AI, Computer Vision, MLOps

Contact

sebastientadiello@gmail.com
+33 7 60 15 96 34
https://www.linkedin.com/in/sebastien-t-ababa2128/
https://github.com/stadiello
https://stadiello.github.io/sebastien-tadiello/

Objectif professionnel

Je recherche un poste me permettant d'allier R&D appliquée et développement de solutions IA robustes, de la modélisation à l'industrialisation. Je souhaite contribuer à des projets en NLP, LLM open source, systèmes de mémoire conversationnelle, IA embarquée (TinyML), ou automatisation intelligente (BRMS, RAG).

Expériences Professionnelles

- **AI Engineer – R&D** Janvier 2024 - Présent
  Groupe Pacte Novation, Paris, France
  - Développement de systèmes TinyML sur systèmes embarqués (détection de chute, traitement audio, vision embarquée)
  - Conception d'un chatbot basé sur des LLM open source, avec RAG optimisé et function calling.
  - Recherche et développement sur l'intégration des BRMS avec des LLM pour des cas métiers réglementés.
  - Recherche et développement d'un nouveau modèle de mémoire contextuelle pour agents IA.
  - Conception de supports de formation et vulgarisation IA.
  - Développement d'une IA pour l'analyse de plans techniques (PDF, DWG) et la génération automatique de schémas électriques, via OCR, vision par ordinateur et Transformers.

  Stack : Python, C++, TensorFlow Lite, Streamlit, LangChain, Ollama, ChromaDB, Drools, Git, Docker, FastAPI, OCR, Transformers, microcontrôleurs (ARM Cortex-M), TinyML

- **Artificial Intelligence Engineer** Janvier 2025 - Mai 2025
  Mission: BNP Paribas, Paris, France
  - Mise en place d'algorithmes de machine learning appliqués à la fraude bancaire.
  - Accompagnement et formation des équipes à la création et au déploiement d'algorithmes de machine learning.

  Stack : Python, Scikit-learn, XGBoost, Jupyter, Git

- **Machine Learning Engineer** Juillet 2022 - Décembre 2023
  SCOR SE, Paris, France

- Identification de problématiques clients et développement de solutions digitales en IA.
- Développement et optimisation de pipelines de données utilisant Databricks et Azure DevOps.
- Mise en place de modèles NLP tels que BERT et GPT pour l'analyse de données textuelles.
- Déploiement d'API (FastAPI) et de modèles de machine learning en production via CI/CD (GitLab, DockerCompose).
- Déploiement de tableaux de bord interactifs via Tableau et Power BI.

Stack : Python, PySpark, Databricks, Azure DevOps, FastAPI, GitLab CI/CD, Docker Compose, Tableau, Power BI, BERT, GPT-2, Transformers

- **Data Scientist / Recherche en sciences cognitives & robotique** Octobre 2021 - Juin 2022
  UMRS 1158 Inserm-Sorbonne Université, Paris, France
  - Mise en place d'un protocole d'étude expérimentale sur la modulation de la dyspnée en interaction humain-robot.
  - Analyse de données NLP pour identifier les émotions des sujets en interaction avec un robot.
  - Développement d'algorithmes de suivi de visage et de mouvements synchrones.

  Stack : Python, OpenCV, dlib, Scikit-learn, Pandas, NLP (TF-IDF, LDA), outils de capture (robot papper, caméra, ECG), ROS

- **Data Scientist / Recherche en sciences cognitives & BCI** Février 2021 - Juin 2021
  Institut de Neurosciences Cognitives et Intégratives d'Aquitaine, Université de Bordeaux, France
  - Analyse statistique et présentation des résultats de recherche.
  - Développement d'un modèle prédictif (random forest).

Stack : Python, Scikit-learn, Random Forest, Pandas, Matplotlib, analyse statistique (ANOVA, tests non paramétriques), EEG/BCI

Formation académique

- **Certificat - Deploying TinyML (HarvardX)** 2024 – 2025
  Université Harvard (MOOC certifiant)
  Déploiement de réseaux de neurones en C++ sur microcontrôleurs (ARM Cortex-M) via TensorFlow Lite. Optimisation mémoire, quantification de modèles et détection embarquée (audio, vision, anomalies).

- **Double Master – Data Science & Intelligence Artificielle** 2021 – 2022
  Université Gustave Eiffel (ESIEE Paris) & HETIC
  Enseignements techniques : Machine Learning, Réseaux de Neurones, NLP, Cloud, DevOps, Statistiques, Text Mining, Optimisation, Big Data.
  Projets interdisciplinaires mêlant IA, traitement du langage et développement d'applications en environnement distribué.

- **Master Recherche – Sciences Cognitives** 2019 – 2021
  Université Paul-Valéry, Montpellier
  Spécialisation : Modélisation computationnelle des fonctions cognitives, neuropsychologie, démarche scientifique.
  Stage de recherche sur l'interaction humain-robot et le traitement émotionnel.

Technologies et outils

- **Langages :** Python, C++, Java, SQL, R, Bash
- **Frameworks IA/ML :** PyTorch, TensorFlow, Keras, Scikit-learn, Hugging Face Transformers
- **NLP & RAG :** LangChain, ChromaDB, Ollama, Vector Databases
- **MLOps & DevOps :** Docker, FastAPI, Git, MLflow, DVC, Vercel, Linux
- **Visualisation :** Streamlit, Power BI, Tableau
- **Modélisation :** Réseaux de Neurones Convolutionnels, Transformers, Apprentissage supervisé/non supervisé
- **Langues :** Français (natif), Anglais (professionnel)

Compétences humaines

- *Capacité d'analyse et de résolution de problèmes*
- *Travail en équipe et transmission des connaissances*
- *Synthèse et vulgarisation de phénomènes complexes*
- *Curiosité technique et veille technologique*
- *Autonomie et adaptabilité*

Projets Personnels

- **Un chatbot classique avec une interface streamlit alimentée par ollama**
  Projet de chatbot from scratch avec une interface et un RAG

  `https://github.com/stadiello/chatBot`

- **LLM integration with BRMS (Business Rule Management Systems)**
  Ce projet intègre des systèmes de gestion des règles commerciales (BRMS) et un RAG, afin d'offrir une solution de génération de texte automatisée, applicable dans différents contextes et réduisant de manière significative les hallucinations du LLM. Il s'agit d'une architecture complète disponible dans un chatBot et entièrement modulable en fonction des besoins.

  ChatBot
  Règles

- **Module de mémoire à court terme pour chatbot**
  Gérer la mémoire à court terme dans les chatbots, en utilisant une combinaison de techniques de stockage et de résumés automatiques pour optimiser le contexte conversationnel.

  GitHub

Publications, Contributions & Talks

- **Publications scientifiques**
  - Grevet, E., Forge, K., **Tadiello, S.**, et al.
    Modeling the acceptability of BCIs for motor rehabilitation after stroke: A large scale study on the general public,
    Frontiers in Neuroergonomics, 2023.
  - Desmons, C., Lavault, S., **Tadiello, S.**, et al.
    Influence d'une activité pseudo-ventilatoire chez un robot humanoïde sur les interactions humain-machine,
    Revue des Maladies Respiratoires, 2023.
  - **Tadiello, S.**, Grevet, E., et al.
    Modélisation de l'acceptabilité des procédures de rééducation post-AVC via les ICO,
    Journée Jeunes Chercheurs en Interface Cerveau-Ordinateur, 2021.

- **Blog technique personnel – IA & développement**
  stadiello.github.io
  Rédaction de tutoriels techniques sur les LLMs, RAG, mémoire des agents, DevOps (Docker, Git, FastAPI) et IA embarquée.

Dernière mise à jour : July 21, 2025