



Ali
Ncibi

DATA SCIENTIST

CONTACT

- 🏠 Paris
- 📍 Île-de-France
- ✉ contact@alincibi.page
- 📞 +33 7 84 89 08 23
- 🌐 <https://alincibi.page>

LANGUES

- Anglais Maîtrise professionnelle (C1)
- Français Bilingue
- Arabe Langue maternelle

PUBLICATIONS

- A. Ncibi, Towards Domain-Robust Activity Recognition using Textual Representations of Binary Sensor Events, 18th International Conference on Agents and Artificial Intelligence (ICAART), Marbella, Spain, March 2026. [PDF](#)
- A. Ncibi, L. Bouganim, P. Pucheral, FlowAR: A Framework for Data-Driven Development of Human Activity Recognition Systems using Binary Sensors, 21st International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT), Lucca, Italy, June 2025, pp. 286–288. [PDF](#)
- A. Ncibi, L. Bouganim, P. Pucheral, FlowAR: une plateforme uniformisée pour la reconnaissance des activités humaines à partir de capteurs binaires, 25e Conférence francophone Extraction et Gestion des Connaissances (EGC), Strasbourg, France, January 2025, pp. 503–510. [PDF](#)

Docteur en IA et Ingénieur Data Science — De l'analyse exploratoire de données à la conception, l'optimisation et le déploiement de modèles.

FORMATIONS

Doctorat en Informatique — Machine Learning & IA 12/2022 - 06/2026

UVSQ / Université Paris-Saclay Versailles, France

Conception de pipelines structurés et évaluation de stratégies de représentation sémantique pour la reconnaissance automatique des activités humaines à partir de capteurs ambiants. Thèse soutenue en juin 2026.

Diplôme d'Ingénieur en Sciences des Données (Double Diplôme) 09/2019 - 06/2021

Télécom Paris, Institut Polytechnique de Paris Palaiseau, France

Optimisation numérique, statistiques, apprentissage statistique, vision par ordinateur, traitement du langage naturel, exploration de graphes, programmation, et sciences humaines et sociales.

Diplôme d'Ingénieur en IT & Télécoms 09/2017 - 06/2019

Ecole Nationale d'Ingénieurs de Tunis (ENIT) Tunis, Tunisia

Electronique, transmission et traitement des signaux, techniques de mise en réseau, recherche opérationnelle, programmation, bases de données et administration système.

EXPÉRIENCES

Doctorant 12/2022 - 06/2026

UVSQ / INRIA Saclay, France

- Structuration, normalisation et analyse exploratoire de 13 jeux de données temporels issus de foyers connectés, couvrant des périodes d'enregistrement de 14 jours à 7 mois.
- Conception, entraînement et comparaison de modèles classiques (Random Forest, Decision Trees, KNN) et d'architectures de Deep Learning pour données séquentielles (CNN, RNN, LSTM, Transformers, sous PyTorch Lightning), dédiés à l'extraction de représentations, à la segmentation et à la classification d'activités.
- Mise en place d'un protocole rigoureux de validation croisée *Leave-One-Day-Out* avec métriques adaptées au déséquilibre de classes, et optimisation bayésienne des hyper-paramètres via processus gaussiens (Scikit-Optimize).
- Développement d'une plateforme d'expérimentation et de visualisation sous Streamlit, conteneurisée avec Docker et déployée sur GCP (Cloud Run) pour le suivi collaboratif des résultats.

Ingénieur Data Science / NLP (CDD) 11/2021 - 05/2022

SouthPigalle Paris, France

- Conception d'un schéma d'annotation (IOB/BIO) et constitution d'un corpus de *slot filling* par croisement de templates de phrases et d'une base de données relationnelle.
- Spécialisation du modèle RoBERTa (Hugging Face, Flair) pour l'extraction de slots et optimisation des hyper-paramètres via Optuna.
- Conteneurisation de l'inférence (FastAPI, Docker) et déploiement sur AWS (EC2) pour intégration directe dans le chatbot.

Ingénieur de Recherche — Stage de fin d'études 03/2021 - 09/2021

INSERM Clamart, France

- Définition d'un schéma d'annotation TLINK et constitution sous Doccano d'un corpus annoté à partir de comptes rendus médicaux non structurés, dédié à la reconstruction de chronologies cliniques.
- Adaptation du modèle de langue CamemBERT (PyTorch, Hugging Face, Flair) pour l'étiquetage de concepts médicaux et la classification de leurs relations temporelles.
- Réglage fin du pipeline d'apprentissage (architectures d'étiquetage/classification, weight decay, learning rate) et développement d'une interface de visualisation interactive des chronologies patients (Streamlit, spaCy).

Langages Python, SQL

ML & Deep Learning Modélisation (Scikit-learn, SciPy), frameworks DL (TensorFlow, PyTorch, PyTorch Lightning), écosystème NLP (Hugging Face) et optimisation d'hyperparamètres (Optuna, Scikit-Optimize)

Data Viz Streamlit, Matplotlib/Seaborn, Plotly

Cloud & Infrastructure Google Cloud Platform (GCP), Amazon Web Services (AWS)

Containerisation Docker (images/containers)

CI/CD & Versioning Git (GitHub, GitLab), GitHub Actions, GitLab CI/CD, pipelines de tests et déploiement

Collaboration Documentation (LaTeX, Markdown, Office, Lucidchart), méthodologies agiles (Scrum, Kanban, backlog), outils (Jira, GitHub Projects, Notion)

Environnements Linux, Windows

Projet Open-Source 06/2022 - 08/2022

Personnel Paris, France

- Conception, développement et packaging d'une bibliothèque Python (Relex) dédiée à l'extraction de relations sémantiques (*Semantic Relation Extraction*).
- Implémentation d'une architecture modulaire articulée autour de PyTorch et Flair NLP pour l'entraînement et l'inférence de modèles de langue spécialisés.
- Structuration du projet pour distribution (Setup/PyPI), rédaction des guides d'utilisation et exemples d'usage. (<https://pypi.org/project/relex/>)

Projet de Recherche 09/2020 - 02/2021

LTCI Lab. (Télécom Paris) Palaiseau, France

- Étude du transfert de style d'accompagnements musicaux MIDI et adaptation du modèle Groove2Groove (architecture Seq2Seq) ciblée sur les motifs de batterie (*drum kits*).
- Conception des pipelines d'entraînement TensorFlow sur données synthétiques et sur le jeu de données benchmark Lakh MIDI Dataset (LMD).
- Définition de métriques d'évaluation objectives (caractéristiques musicales/rythmiques) et conduite de tests utilisateurs subjectifs.

Projet de Recherche 02/2020 - 04/2020

LTCI Lab. (Télécom Paris) Palaiseau, France

- État de l'art des architectures Deep Learning et analyse comparative des fonctions de coût (loss functions) dédiées au recalage d'images médicales (ex: Mutual Information, Local Normalized Cross-Correlation).
- Implémentation TensorFlow de différentes variantes de losses, entraînement du pipeline et validation expérimentale des performances de recalage sur le dataset MICCAI BraTS.