

Ezéchiél SAWADOGO

Traitement automatique des langues · Modèles de langue · Apprentissage profond

sawadogoezechiel09@gmail.com · 06 13 19 77 99 · Noisiel (77) — Mobilité Île-de-France

LinkedIn : linkedin.com/in/ezechiel-sawadogo-483079291 · GitHub : github.com/ezechiel-sawadogo-nlp · Hugging Face : huggingface.co/Ezeck01

PROFIL

Étudiant en M2 Langue et Informatique à Sorbonne Université (major de promotion en M1), à la recherche d'un stage de fin d'études de 6 mois à partir de mars 2027 en NLP et modèles de langue. Double compétence sciences du langage et informatique : conception de pipelines complets (structuration de données multilingues, annotation, fine-tuning, évaluation quantitative), avec une attention particulière à la qualité des données et à la reproductibilité.

EXPÉRIENCES PROFESSIONNELLES

Stagiaire Recherche & Développement : CNRS / MSHS de Poitiers

Consortium Huma-Num ARIANE · Juin – Septembre 2026 · Encadrement : Anaïs Chambat

- Conçu et déployé un outil de visualisation interactive (D3.js) de projets d'édition numérique (2 vues croisées, filtres dynamiques, version responsive), mis en production sur GitLab Pages.
- Développé une chaîne de traitement Python complète : extraction CSV, normalisation, contrôle qualité automatisé, structuration bilingue FR/EN de 13 projets et 38 outils.
- Refondu un questionnaire de recensement de 29 items. Mission menée en autonomie avec cycles de recette réguliers auprès de la responsable scientifique.

Stagiaire : Laboratoire ELLIADD, Université de Franche-Comté

Pôle Discours, Langues, Sociétés · Avril – Juin 2025 · Encadrement : Laura Muller & Manon Him-Aquili

- Co-construit le Glossaire de l'étudiant-e bisontin-e (ouvrage imprimé) : collecte du lexique étudiant, structuration des entrées et rédaction des définitions.
- Réalisé une veille scientifique et des dossiers thématiques pour la revue Hermès (Cairn, HAL, OpenEdition, Persée) ; synthèse critique de productions académiques.

PROJETS

Détection de discours haineux en français — Fine-tuning de CamemBERT

- Fine-tuning de camembert-base sur FrenchHateSpeechSuperset (42 660 exemples) avec PyTorch et Hugging Face Transformers.
- Diagnostic d'un effondrement du modèle sur la classe majoritaire (89 % d'exactitude, rappel nul) ; rééquilibrage par sous-échantillonnage → rappel à 0,81. Arbitrage précision/rappel documenté.

Reconnaissance de caractères dans des monogrammes de sceaux byzantins

Mémoire de M1 : 18/20 · Sorbonne Université · Encadrement : Victoria Eyharabide

- Pipeline semi-automatique de bout en bout (annotation → détection → classification → évaluation). Travaux acceptés à DH2026 (Daejeon, juillet 2026).
- YOLOv8s : mAP@50 $\approx$ 0,778 (validation croisée 5 plis, 351 images, 49 classes).
- ResNet18 : 96,66 % d'exactitude, F1-macro 0,937 ; comparaison avec ViT et DINOv2-Small.

Évaluation d'OCR et traitement de corpus

- Comparaison Kraken / Tesseract sur corpus d'imprimés du XVIIe siècle (projet Antonomaz) : CER/WER, similarité de Jaccard.
- Analyse lexicométrique d'un corpus de presse : distributions zipfiennes, détection de néologismes, log-vraisemblance.

COMPÉTENCES

Langages : Python, Java, C++, Bash

Deep Learning : PyTorch, YOLOv8, ResNet, ViT, DINOv2, scikit-learn

NLP / TAL : Hugging Face Transformers, CamemBERT, spaCy, NLTK, annotation & NER

Données : pandas, XML/TEI, RDF, JSON, CSV, contrôle qualité, corpus bilingues

Outils : Git, GitLab CI, LaTeX, D3.js, Hugging Face Hub

Métriques : mAP, F1, matrices de confusion, validation croisée, CER/WER

Langues : Français (langue maternelle) · Anglais (B2)

FORMATION

Master Langue et Informatique : Sorbonne Université

2025 – 2026 · Major de promotion de M1 sur les deux semestres

Licence Sciences du Langage : Université de Franche-Comté, Besançon

2023 – 2025

Études de linguistique : Université Joseph Ki-Zerbo, Ouagadougou (Burkina Faso)

2019 – 2023

PUBLICATION & DIVERS

Publication : Sawadogo E., Barry M. D., Eyharabide V. Digital Curation and Datafication of Byzantine Seal Monograms, DH2026, Daejeon, juillet 2026.

Écriture : Fondateur de Mots & Pouvoir, publication en ligne sur le langage et le pouvoir (substack.com/@motspouvoir).

Emploi étudiant : Équipier polyvalent, McDonald's La Défense (2025 – 2026).