

Fokam Emeric

INGÉNIEUR & CHERCHEUR EN IA — NLP & TRANSPORT OPTIMAL

Ingénierie ML · IA agentique & systèmes LLM · Data engineering · Backend & intégration

☎ +237 656 62 72 38 ✉ fokamemeric@gmail.com 🌐 fokamemeric.com

📄 github.com/emeric-cyrille 🆔 ORCID 0009-0001-4232-2361 📍 Yaoundé, Cameroun



6+

APPLICATIONS EN PRODUCTION

30+

MICROSERVICES MAINTENUS

77

DÉPÔTS GITHUB

1^{er}

PISTE WIKIKGQA, ISWC 2026

01 PROFIL & RECHERCHE

Ingénieur et chercheur en IA, j'interviens sur l'ensemble du cycle de vie des systèmes d'IA, de la théorie à la production. Ma **recherche principale porte sur le traitement automatique du langage sous l'angle du transport optimal** : distances de Wasserstein, théorie des perturbations spectrales et analyse des sous-espaces propres pour mesurer comment les espaces de plongements multilingues se déplacent d'un modèle et d'une langue source à l'autre (recherche de master, laboratoire UMMISCO, IRD / Université de Yaoundé 1).

Sur le plan appliqué, je conçois des systèmes LLM qui raisonnent et vérifient leurs réponses : notre pipeline few-shot **text-to-SPARQL a remporté la piste espagnole avec mentions du challenge WikiKGQA à l'ISWC 2026** et est publié dans CEUR-WS. Je mets également en production des modèles adaptés aux langues et aux accents du Cameroun (traduction en ghomálá', reconnaissance vocale du français camerounais) et je construis les backends distribués qui les portent.

THÈMES DE RECHERCHE

Transport optimal · Géométrie de Wasserstein · Perturbations spectrales · Plongements multilingues · Apprentissage de représentations · NLP pour langues peu dotées · Reconnaissance vocale · QA sur graphes de connaissances · Raisonnement des LLM & usage d'outils

Curieux de nature, et ouvert aux questions de recherche voisines.

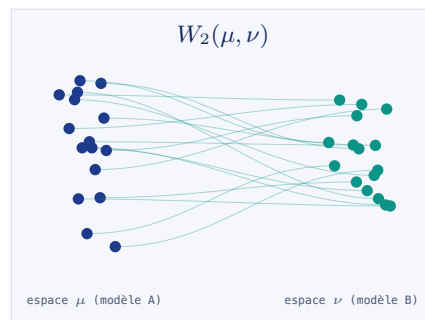


Fig. A – Plan de transport optimal entre deux espaces de plongements (illustration).

02 DOMAINES D'EXPERTISE

Recherche en IA

NLP & transport optimal

Mesurer comment le sens se déplace entre espaces de plongements, modèles et langues. Distances de Wasserstein entre distributions de plongements, théorie des perturbations spectrales et analyse des sous-espaces propres pour les représentations multilingues ; analyse sémantique par LLM et question-réponse sur graphes de connaissances (Wikidata).

Ingénierie du Machine Learning

Un modèle n'est utile que lorsque les gens peuvent s'en servir. Feature engineering, sélection et évaluation de modèles, deep learning avec PyTorch et TensorFlow, mise à disposition de modèles via des API REST ; traduction automatique neuronale, reconnaissance vocale et systèmes de recommandation en production.

IA agentique & systèmes LLM

Des systèmes LLM qui raisonnent, appellent des outils et vérifient leurs propres réponses. Orchestration d'agents avec LangChain, LangGraph et CrewAI, appel d'outils et serveurs Model Context Protocol, RAG avec LlamaIndex et bases vectorielles (FAISS, Chroma, pgvector), échantillonnage multi-candidats avec sélection guidée par l'exécution, inférence de modèles open-weight.

Data engineering & analyse de données

Des données fiables avant tout. Pipelines ETL (Pentaho), web scraping à grande échelle, modélisation de données relationnelles, capture de données d'exécution, flux de données par messages et mise en cache ; analyse exploratoire et statistique, analytique SQL, tableaux de bord (Power BI, R Shiny, React).

Développement backend & intégration

Des backends distribués qui supportent un trafic réel. Microservices et architectures événementielles (RabbitMQ), cache distribué (Redis), authentification et autorisation (JWT, OAuth 2.0), passerelles de paiement et intégration d'API tierces, documentation OpenAPI.

Outils de recherche & scientifiques	LaTeX TikZ Overleaf Jupyter Git / GitHub Hugging Face Hub Codabench Kaggle Zindi Matplotlib Plotly
Langages	Python TypeScript JavaScript SQL R PHP
Machine learning	PyTorch scikit-learn TensorFlow Hugging Face Transformers NumPy pandas SciPy Wav2Vec 2.0 / XLS-R Gradio
LLM & IA agentique	LangChain LangGraph LlamaIndex CrewAI Model Context Protocol Appel d'outils Prompt engineering Groq (LLM open-weight)
Recherche d'information & vectorielle	RAG FAISS Chroma pgvector Recherche TF-IDF SPARQL Wikidata / QLever
Data engineering	PostgreSQL MariaDB MySQL Redis RabbitMQ Scrapy Pentaho Data Integration Pipelines ETL
Backend & intégration	NestJS Node.js Django REST Framework Flask Next.js Symfony Prisma Socket.IO OpenAPI JWT / OAuth 2.0 Mobile Money
Analyse & visualisation	Power BI R Shiny Plotly Matplotlib Tableaux de bord React
Infrastructure	Docker / Compose Traefik Let's Encrypt CI/CD GitHub Actions VPS Linux Hugging Face Spaces MinIO

04 RECHERCHE & PUBLICATIONS

2025

Recherche principale

WASER: Wasserstein Subspace Embedding Representation

E. C. Kenmegne Fokam – Recherche de mémoire de master, laboratoire UMMISCO, Université de Yaoundé 1 · github.com/emeric-cyrille/WASER

Dans quelle mesure une représentation change-t-elle lorsque le même corpus est plongé avec différents modèles multilingues ou différentes langues sources ? WASER modélise les transformations géométriques des espaces de plongements par la théorie des perturbations spectrales et compare les distributions de représentations par transport optimal de Wasserstein et analyse des sous-espaces propres.

[transport optimal](#) · [distance de Wasserstein](#) · [perturbations spectrales](#) · [plongements multilingues](#)

2026

Évalué par les pairs

Lauréat de piste

TF-IDF Example Retrieval and Execution-Guided Selection for Few-Shot Text-to-SPARQL over Wikidata with Open-Weight LLMs

E. C. Kenmegne Fokam, G. E. Tela Tela – WikiKGQA 2026 @ ISWC 2026, CEUR Workshop Proceedings, vol. 4284 · ceur-ws.org/Vol-4284/paper2.pdf

Système en quatre étapes : recherche TF-IDF de trois exemples few-shot filtrés par type de requête, prompt few-shot adressé à un LLM open-weight (gpt-oss-120b, llama-3.3-70b via Groq), cinq candidats à températures variées, puis sélection guidée par l'exécution du premier candidat renvoyant une réponse non vide sur l'endpoint QLever. **Lauréat de la piste espagnole avec mentions.**

F1 macro (QALD)	EN	ES
Dev (100 q.)	0,634	0,607
Test (75 q.)	0,650	0,630

05 EXPÉRIENCE PROFESSIONNELLE



déc. 2025 –
mai 2026
Yaoundé

Développeur backend (Next.js / Node.js)

AgdTech

AgdTech — plateforme transactionnelle temps réel de plus de 30 microservices événementiels alimentés par des fournisseurs de cotes sportives.

- ◆ **Maintenance et évolution de plus de 30 microservices** : débogage, nouvelles fonctionnalités et mises en production sans rompre le flux d'événements.
- ◆ **Intégration des fournisseurs de cotes et de données de marché** (FeedConstruct, API-Sports) via un service dédié aux API tierces.
- ◆ **Communication événementielle** avec RabbitMQ, et Redis comme cache distribué.
- ◆ **Conteneurisation des services** avec Docker / Compose ; Traefik pour le routage et la sécurité ; tests unitaires et documentation des API.

Node.js · Next.js · Microservices · RabbitMQ · Redis · Docker · Traefik

sept. 2025 –
déc. 2025
À distance

Développeur backend (NestJS)

Core Extreme

Produits SaaS, dont ReserveIT, assistant de réservation IA sur WhatsApp (reserveit.cm), et une application de réservation de transats.

- ◆ **Conception et développement d'API** pour plusieurs services d'une architecture microservices NestJS.
- ◆ **Intégration de moyens de paiement**, dont Mobile Money, sur ReserveIT et l'application de réservation de transats.
- ◆ **Authentification et autorisation sécurisées** avec JWT et OAuth ; conception du schéma PostgreSQL.
- ◆ **Livraison de fonctionnalités produit** et documentation de chaque API avec Swagger / OpenAPI.

NestJS · PostgreSQL · JWT · OAuth · Mobile Money · Swagger

sept. 2024 –
janv. 2025
Douala

Data Engineer / Data Analyst

Africa Systems

Traçabilité et supervision de bout en bout des données de véhicules.

- ◆ **Scripts de capture Python** qui collectent et tracent les données d'exécution des véhicules.
- ◆ **Jobs ETL orchestrés et planifiés** avec Pentaho Data Integration.
- ◆ **API Flask** et bases PostgreSQL associées ; **tableau de bord de supervision React**.

Python · Pentaho · Flask · PostgreSQL · React · ETL

févr. 2024 –
juin 2024
Yaoundé

Data Scientist & développeur backend — stage

HS4HEALTH

Recommandation pour la plateforme gHealth Village Connect (connect.ghealthvillage.com).

- ◆ **Nettoyage et préparation des données**, conception et évaluation du modèle de recommandation IA.
- ◆ **Exposition du modèle** via des API RESTful avec Django REST Framework et **déploiement en production**.

Python · scikit-learn · Django REST · PostgreSQL

août 2023 –
oct. 2023
À distance
(France)

Développeur web (PHP / Symfony)

Freelance

- ◆ **Modélisation de la base de données relationnelle** (SQL / MariaDB), réalisation de l'interface (HTML, SASS, Bootstrap), développement des fonctionnalités Symfony et déploiement sur Platform.sh.

PHP · Symfony · MariaDB · SASS · Bootstrap

06 PROJETS CHOISIS



Ghomala' Translate
ghomala.fokamemeric.com

Traduction automatique neuronale français → ghomálá', langue bamiléké à tons : d'abord un dictionnaire sourcé de 283 entrées, puis un Seq2Seq BiLSTM avec attention à 8 têtes (29,2 M de paramètres, plongements XLM-R, ~100 ms par phrase sur CPU) ; plateforme de contribution ouverte. [PyTorch](#) · [Flask](#) · [Docker](#) · [Traefik](#)

KamerVox
Hugging Face Space

Reconnaissance vocale du français camerounais : Wav2Vec 2.0 / XLS-R affiné sur des locuteurs de Yaoundé, deux modèles (244 M et 74 M de paramètres), décodage CTC, application Gradio. Projet CAFHOMEC, UMMISCO × Université de Yaoundé 1. [PyTorch](#) · [Transformers](#) · [Gradio](#)

Pipeline KGQA agentique
github

Boucle générer-exécuter-sélectionner : le LLM écrit des requêtes SPARQL candidates, les exécute comme outil sur l'endpoint Wikidata et conserve la première réponse valide, avec repli sur la requête déterministe. [LLM open-weight](#) · [Appel d'outils](#) · [SPARQL](#)

ReserveIT
reserveit.cm

WhatsApp devient un guichet de réservation ouvert 24 h/24, 7 j/7 : un assistant IA gère la conversation, le créneau, le paiement Mobile Money et les rappels. Intégration des paiements et fonctionnalités backend. [NestJS](#) · [PostgreSQL](#) · [Mobile Money](#)

TravelLab

pour KYAD CENTER
(Guadeloupe)

Plateforme e-learning de simulation pour le BTS Tourisme : quatre espaces par rôle, éditeur de scénarios et mode examen, suivi de 33 compétences du référentiel, journaux d'audit chaînés par hachage, sauvegardes chiffrées, CI/CD. [PHP 8.2](#) · [MySQL 8](#) · [MinIO](#) · [Docker](#)

Omega

omega.fokamemic.com

Gestion des stocks et des ventes d'un bar : API NestJS + Prisma, PWA React pensée mobile d'abord, synchronisation temps réel avec Socket.IO, historique des prix, exports PDF / Excel. [NestJS](#) · [Prisma](#) · [React](#)

JOBBER & scraping de corpus

[github](#)

Spiders Scrapy sur Emploi.cm, LoumaJobs et JobInfoCamer alimentant un système de recommandation d'offres d'emploi à partir du CV ; collecte automatisée de textes bibliques pour les corpus NLP du laboratoire UMMISCO. [Scrapy](#) · [PostgreSQL](#) · [scikit-learn](#)

LangChain & LangGraph

[DataCamp](#)

Agents à appel d'outils dotés de mémoire, workflows en graphes d'états, RAG sur bases vectorielles ; ainsi que des dizaines de projets d'analyse sur données réelles. [LangChain](#) · [LangGraph](#) · [RAG](#)

07 CONFÉRENCES, ATELIERS & COMMUNAUTÉS



25–29 oct. 2026

Bari, Italie

ISWC 2026 — 25^e International Semantic Web Conference

[Présentation d'article](#) · [Lauréat de la piste WikiKGQA](#)

Conférence de référence sur le Web sémantique, les graphes de connaissances et les données liées. Présentation de notre système au premier challenge WikiKGQA : 492 paires question-SPARQL bilingues anglais/espagnol construites sur QAWiki, un instantané figé de Wikidata sur un endpoint QLever dédié, quatre pistes évaluées en F1 QALD macro sur Codabench.

27–28 mai 2025

Yaoundé (hybride)

DASSA'25 — Digital Avenues for Low-Resource Languages of Sub-Saharan Africa

[Comité d'organisation, logistique](#)

Atelier international de l'African Society in Digital Sciences (ASDS), initiative AI4AfricanLanguages, soutenu par l'Agence nationale de la recherche (ANR). Thèmes : collecte et curation de données, reconnaissance et synthèse vocales, ML explicable, multimodalité texte-audio, biais d'annotation. Responsable de la logistique. asds.africa/dassa2025

En cours

Compétitions & communautés

Challenges de data science et d'IA sur Codabench, Kaggle et Zindi ; membre de la communauté KmerAI.

08 FORMATION & CERTIFICATIONS



2020 – 2025

Master en informatique — Science des données

[Université de Yaoundé 1](#)

Mémoire de master au laboratoire UMMISCO (IRD / Université de Yaoundé 1) : transport optimal et théorie des perturbations spectrales pour comparer des espaces de plongements.

2023 – 2024

DataCamp — Parcours Data Scientist Professional (116 h)

Certifications : Data Scientist Professional · Data Analyst Professional · Data Scientist Associate · Data Analyst Associate · SQL Associate

2019

Baccalauréat C — Mathématiques & sciences physiques

[Lycée bilingue de Tougang 2](#)

09 LANGUES & LIENS



Langues

Français (langue maternelle) · Anglais (professionnel)

En ligne

fokamemic.com · github.com/emerick-cyrille (77 dépôts, personnels ou contributions) · [ORCID](#) 0009-0001-4232-2361 · huggingface.co/kfec

