NJIGOUH ABDOULAYE RAZAK

Ingénieur en Informatique (Cloud & IA) | Ingénierie AI Safety — RLHF, Red Teaming, Évaluation Tool-Use | Bilingue FR/EN

njigouhrazak@iut-dhaka.edu | +237 671 515 042 | Logpom, Douala, Cameroun | linkedin.com/in/njigouhabdoulaye-razak | github.com/Abdoulaye-Thespy

PROFIL

Je conçois, développe et teste des systèmes IA de bout en bout — pas seulement des prototypes. Deux ans comme AI Trainer chez Outlier : RLHF, évaluation de sécurité, vérification du tool-use/function-calling, et revue de code généré par des modèles. J'évalue également le comportement multimodal et tool-use des LLM chez TELUS Digital. Mon parcours d'ingénieur couvre l'infrastructure cloud (AWS, Docker) chez Yogaya, la direction technique d'une communauté de 100+ freelances chez FreelanceConnect, et la stratégie technologique de Global Academy, où je pilote le programme de formation IA. Bilingue français/anglais — une compétence qui se prolonge dans mon travail d'AI safety, notamment une vulnérabilité cross-lingue que j'ai découverte et documentée moi-même. Certifié AI Fluency par Anthropic Academy (finalisation en cours des modules restants).

PROJET PHARE

SafeAgent Lab — Pipeline d'AI Safety pour un assistant en production (Global Academy)

Projet personnel / professionnel — conçu, développé et testé pour Global Academy, le centre de formation que j'ai co-fondé et que je dirige.

- Construction d'un assistant RAG en français (LangChain.js, pgvector, Mistral/Llama via Ollama et API, Node.js/Express) ancré sur des données institutionnelles réelles.
- Mise en place de garde-fous d'entrée/sortie (détection d'injection de prompt, liste blanche PII, détection de fuite de prompt) et tests adversariaux automatisés avec Promptfoo et PyRIT, avec mesure réelle des taux de résistance avant/après.
- Découverte d'une vulnérabilité de fuite de prompt cross-lingue (contournement français→anglais en cours de conversation) — un angle mort peu couvert même par les outils de red teaming établis (PyRIT, Garak).
- Implémentation de MCP (Model Context Protocol) comme capacité indépendante — un serveur autonome de tools/resources/prompts — volontairement séparé du pipeline RAG pour démontrer les deux patterns clairement.

Outils de red teaming : PyRIT, Promptfoo, modèles évaluateurs open source (Llama Guard, NVIDIA NeMo Guardrails), Ollama, Docker Compose.

EXPÉRIENCE

Directeur de la Technologie et de la Pédagogie Numérique / Co-fondateur — Global Academy *Juil. 2024 – Présent · Douala, Cameroun*

- Direction de l'infrastructure technologique et de la stratégie d'un centre de formation professionnelle spécialisé dans les métiers du numérique et de la logistique.
- Pilotage du programme de formation IA pour la nouvelle année académique de l'institution : compréhension de l'IA, formulation de prompts, usage responsable.
- Conception et animation d'ateliers pratiques AWS et cybersécurité.
- Architecture et maintenance de l'assistant IA propre à l'institution (SafeAgent Lab — voir Projet phare).

AI Evaluation & Alignment Specialist — Micro1 *Août 2026 – Présent · Télétravail freelance*

- Évaluation bilingue FR/EN de l'alignement et de la sécurité des LLM (RLHF) : red teaming, tests adversariaux, robustesse, biais, hallucinations.
- Évaluation : vérification du function calling / tool-use : invocation correcte des outils, structure des arguments, validation de schéma, parsing JSON, gestion des erreurs.
- Fine-tuning ciblé et feedback technique : réduction des erreurs d'outils, calibration, qualité des réponses.
- Fine-tuning de modèles et amélioration de la qualité, de la précision et du ton des réponses.

AI Trainer — Outlier *juin 2024 – présent · Télétravail freelance*

- Évaluation de la sécurité et de l'alignement des modèles IA (RLHF) : identification des risques, sorties indésirables, tests de robustesse.
- Évaluation du function calling et du tool-use : vérification de l'invocation correcte des outils, de la structure des arguments, et du traitement des données retournées.
- Revue de code et évaluation du code généré par les modèles : correction, sécurité, gestion des cas limites, respect des bonnes pratiques.
- Fine-tuning de modèles et amélioration de la qualité, de la précision et du ton des réponses.

AI Model Evaluator — Évaluation Tool-Use & Multimodale des LLM — TELUS Digital *Mars 2024 – Présent · Télétravail freelance*

- Vérification du function calling et tool-use : invocation correcte, arguments valides, séquencement des appels approprié.
- Validation du grounding et du JSON : réponses vérifiées contre les données sources et le schéma attendu, détection des champs hallucinés.

- Classement de préférence et taxonomie des échecs : catégorisation des modes de défaillance (appels d'outils incorrects, affirmations non ancrées, violations de schéma).

Cloud & AI Practice Lead — FreelanceConnect *Oct. 2023 – Présent · Douala, Cameroun*

- Architecture de solutions cloud et IA pour des missions clients livrées via une communauté de 100+ freelances.
- Conception d'architectures cloud et IA pour des missions clients : cadrage, prototypage, MVP, mise en production et documentation technique.
- Mise en place d'infrastructures cloud scalables : conteneurisation, serverless, IaC, CI/CD, monitoring, sécurité et optimisation des coûts.
- Mentorat technique : montée en compétences d'ingénieurs sur l'infrastructure cloud, les outils IA, l'architecture et les bonnes pratiques.

Cloud Engineer — Yogaya Inc. *Sept. 2021 – Mars 2024 · Seattle (Télétravail)*

- Migration d'infrastructure on-premises vers AWS (EC2, RDS, S3) ; conteneurisation des services avec Docker ; mise en place de CloudFront/CloudWatch.
- Mise en place de supervision et d'alertes automatisées, résolution proactive des incidents d'infrastructure.

COMPÉTENCES

AI Safety & Évaluation : RLHF, red teaming, conception de garde-fous, détection d'injection de prompt & jailbreak, vérification tool-use & function-calling, revue de code de sortie modèle, taxonomie des échecs, notation par grille

Outils de red teaming : PyRIT, Promptfoo, Llama Guard, NVIDIA NeMo Guardrails, Ollama, Guardrails AI

Outillage IA / ML : LangChain.js, Mistral, Llama, MCP (Model Context Protocol), prompt engineering, RAG, QLoRA (en cours)

Cloud & DevOps : AWS (EC2, RDS, S3, Lambda, DynamoDB), Docker, CI/CD, Git/GitHub, Linux

Développement : Python, TypeScript, JavaScript, Node.js, React, Next.js

Langues : Français — Natif (C2) · Anglais — Professionnel (C1) · Espagnol — Élémentaire (A2)

CERTIFICATIONS

- Anthropic Academy — AI Fluency : Framework & Foundations, for Educators, for Small Businesses, for Builders ; Teaching AI Fluency ; AI Capabilities & Limitations ; Claude Platform 101 ; Introduction to Model Context Protocol

Finalisation en cours des modules restants du programme Anthropic Academy AI Fluency.

- Udacity — Cloud & DevOps Engineer Nanodegree
- AIX-T — Cloud Developer
- Coursera — Linux : Introduction to Shell Scripting for DevOps

FORMATION

Licence (BSc), Informatique & Ingénierie

Islamic University of Technology (IUT-OIC) 2015 – 2019

Développeur Full-Stack, Informatique — Microverse · 2020 – 2021