

Pyae Sone Kyaw

Generative AI Engineer • Full-Stack AI Development • Azure & Responsible AI

+33 745 146 129 • pyaesonekyaw101010@gmail.com Paris, France • linkedin/pyae-sone-kyaw • github/soneeee220000
pseonkyaw.dev

PROFESSIONAL SUMMARY

Data Engineer & Generative AI Engineer with proven experience designing **end-to-end Azure data pipelines**, processing complex **unstructured data** (PDF, Word, JSON), and implementing **LLM/RAG solutions** in production. Full-stack delivery capability spanning **React / TypeScript** frontends and **Python / Django / FastAPI** backends integrated with **Azure OpenAI, Azure Functions**, and **prompt engineering** workflows including model **fine-tuning**. Currently building AI safety, compliance, and licensing systems at a Station F Paris startup; previously shipped production Azure data pipelines at a smart-mobility scale-up.

TECHNICAL SKILLS

AI / GenAI	OpenAI GPT-4o	Azure OpenAI	Prompt Engineering	RAG	LangChain	Hugging Face	FAISS
	Vector Databases	Fine-tuning	Responsible AI	Google Gemini	Genkit		
	PDF Processing	Word/DOCX	Document Intelligence	NLP	OCR		
Unstructured Data Processing							
Backend	Python	Django	FastAPI	Azure Functions	REST APIs	SQL / MySQL	Bash
Frontend	React / Next.js	TypeScript	JavaScript (ES6+)	HTML5	CSS3	Tailwind CSS	Leaflet
Cloud & DevOps	Microsoft Azure	Azure Blob Storage	Azure Storage	Azure ML Studio	Azure DevOps		
	Databricks	Docker	Git / GitHub	Vercel			
ML & Data	PyTorch	TensorFlow	scikit-learn	PySpark	Pandas	NLP	Computer Vision
Tools	Visual Studio 2022	VS Code	PyCharm	JupyterLab	Linux		

PROFESSIONAL EXPERIENCE

Data Scientist & Cloud Data Engineer | Floware — Smart Mobility Startup @ X Innovation Center Jul 2024 – Jan 2025 • Massy, France

- Designed and deployed end-to-end **Azure data pipelines** ingesting Bluetooth, computer-vision, and geospatial datasets into **Azure Blob Storage** for a real-time smart-city analytics platform.
- Built **forecasting and anomaly-detection models** surfacing mobility patterns, congestion metrics, and data-driven urban-planning insights for city-authority stakeholders.
- Improved data-processing efficiency by **35%** and reduced Azure cloud costs by **20%** through batch-architecture optimisation on **Azure ML Studio** and **Databricks**.
- Delivered pipelines in cross-functional **Agile sprints** via **Azure DevOps**, collaborating with Product Owners and Data teams to ensure solution reliability and efficiency.

Founding AI Engineer & Architect | Siloett.AI — AI Safety & Compliance Startup @ Station F Jan 2025 – Present • Paris, France

- Implemented **RAG-based document processing pipelines** handling PDF, Word, and structured documents using **FAISS** and **LangChain**, with dynamic policy enforcement and audit logging.
- Designed and implemented **responsible-AI validation layers** including content-safety filters, bias-detection checks, and compliance guardrails ensuring all AI-generated outputs meet regulatory and ethical standards.
- Built **AI safety and IP compliance systems** including audit logging, provenance tracking, and output attribution pipelines — ensuring traceable and accountable AI output lifecycles.
- Developed **prompt-engineering** and **LLM evaluation frameworks** using **Azure OpenAI GPT-4o** and **LangChain**, with systematic fine-tuning to optimise domain accuracy across compliance use cases.

KEY PROJECTS

GreenLens — Cloud Carbon Intelligence ASP.NET Core 8 • Angular 17 • Azure AI Search • Azure OpenAI • Docker
• SQLite • Playwright

- Designed and built an end-to-end **CO₂e estimation platform** with **Clean Architecture**, enabling per-resource carbon breakdowns for Azure infrastructure to support EU CSRD Scope 3 reporting compliance (effective 2026).
- Integrated **Azure AI Search** for semantic emission factor lookup across 50+ factors and **Azure OpenAI GPT-4o-mini** for AI-generated reduction strategies with 1-hour in-memory caching.
- Achieved **88 automated tests** across unit, integration, and E2E layers (xUnit, Jasmine/Karma, Playwright); deployed as a single deployable unit on **Azure App Service**.

SafeGen — Responsible AI Compliance API *Azure Functions • Azure OpenAI • RAG • FAISS • Python • React • TypeScript* 2025

- Built a serverless compliance-validation pipeline on **Azure Functions** that intercepts **Azure OpenAI** responses and evaluates them against configurable safety, bias, and regulatory rules before serving to end users.
- Developed a **React / TypeScript** monitoring dashboard displaying real-time compliance metrics, flagged outputs, and audit trails stored in **Azure Blob Storage**.
- Implemented **RAG**-based rule retrieval using **FAISS** and domain-specific compliance documents, enabling dynamic policy enforcement without redeployment.

VaxEvidence — Real-World Evidence Regulatory Platform *Next.js • React 19 • TypeScript • Supabase • REST APIs • Compliance* 2025

- Built a **full-stack regulatory compliance platform** for vaccine research — protocol builder with PICO framework, evidence library with **PubMed / ClinicalTrials.gov** API integration, and collaborative review workflows targeting **FDA/EMA** compliance.
- Developed **React 19 / TypeScript** frontend with 56+ reusable UI components and a **REST API** backend handling document exports (PDF/Word), bibliography generation, and dataset management.

DocuQuery — AI-Powered Resume & Job Match Analyser *Next.js • FastAPI • Python • TypeScript • PostgreSQL • OpenAI* 2025

- Built an **AI-powered resume–job match analysis** platform with **unstructured PDF/Word document parsing** using **OpenAI GPT-4o** and a **FastAPI** backend — users upload a resume and job description to receive a **detailed compatibility score** with category breakdowns across skills, experience, education, and keywords in under 30 seconds.
- Implemented **missing keyword identification** and **prioritised improvement recommendations**, surfacing actionable gaps between candidate profiles and job requirements via structured LLM output parsing.
- Developed a **Next.js / TypeScript** frontend with a real-time results dashboard and built a **PostgreSQL**-backed REST API handling resume parsing, session management, and score history.

HEATDEBT — Urban Thermal Equity Intelligence Platform *Next.js 14 • TypeScript • Leaflet • Google Gemini • Tailwind CSS • Vercel* 2026

- Designed and built an **urban heat-vulnerability monitoring platform** covering **14 real US Census tracts** in Montgomery, AL — interactive choropleth maps with **9 switchable data layers** (heat exposure, tree canopy, poverty rate, air quality, and more) rendered on a Leaflet dark basemap.
- Integrated **Google Gemini 2.5 Flash** via **Google Genkit** server actions to generate per-district **AI risk assessments**, prioritised intervention recommendations with budget estimates, and **EPA Environmental Justice grant narratives** — all with graceful fallback on API failure.
- Built **dual-format report generation** (PDF and Word) producing multi-page professional vulnerability reports with cover pages, neighbourhood comparison grids, and AI-driven policy sections for city planners and grant writers.

EDUCATION

M.Sc. Data Science & Network Intelligence *Sep 2023 – Dec 2024*
 Télécom SudParis — Institut Polytechnique de Paris *Paris, France*
 Relevant modules: *NLP • Computer Vision • Cloud Computing • Data Structures • Linear Algebra • Project Management*

M.Sc. Data Science & Artificial Intelligence *Aug 2022 – Dec 2024*
 Asian Institute of Technology *Bangkok, Thailand*
 Dual-degree programme with Télécom SudParis — Institut Polytechnique de Paris

B.A. Social Science Studies *Sep 2016 – Dec 2020*
 Myanmar Institute of Theology *Yangon, Myanmar*

LANGUAGES

English — Professional working proficiency

French — Beginner (A1–A2, daily conversational use in Paris)

Burmese — Native

Japanese — Conversational