

# Zakariae El Mernissi

✉ zakariaeelmernissi@gmail.com ☎ +212663363170 🌐 Zakariae el mernissi 📅 08/08/2000 🔗 zakariae.live

## Professional Experience

<b>Data Scientist &amp; Artificial Intelligence Engineer, Capgemini TS</b> - <b>Member of the GenAI Lab – Designing Innovative Generative AI Solutions</b> - <b>AI-Powered HR Platform</b> – Intelligent CV Management System: Multi-agent GenAI solution built with <b>Google's Agent Development Kit (ADK)</b> Smart document parsing with entity extraction and metadata enrichment. Semantic search using vector embeddings and retrieval via <b>Vertex AI</b> . User-friendly interface built with <b>Streamlit</b> . - <b>Tech stack:</b> Python, GCP ADK, Gemini Flash 2.0, Vertex AI, Cloud Storage, PostgreSQL, Markdown, Streamlit - <b>Supply Chain Analyzer</b> – GenAI-Powered Logistics Intelligence System: Cloud-native GenAI system for analyzing international supply chain constraints. Document search pipeline powered by <b>Amazon Titan</b> (for embedding) and <b>OpenSearch Service</b> (for semantic retrieval). Final AI responses generated using <b>Claude 3.5 Sonnet</b> via <b>Amazon Bedrock</b> , enriched with references. Front-end experience delivered through a secure <b>ReactJS web app</b> hosted on <b>Amazon S3</b> + <b>CloudFront</b> with <b>Amazon Cognito</b> authentication. - <b>Tech stack:</b> ReactJS, Python, AWS Bedrock, Claude 3.5 Sonnet, Amazon Titan, OpenSearch, AWS Lambda, Amazon API Gateway, Amazon S3, CloudFront, Cognito - <b>Computer System Validation (CSV) Platform</b> – Pharmaceutical Industry : Multi-agent GenAI system designed for automated validation and compliance analysis in pharmaceutical environments. Specialized agent roles for document analysis and regulatory compliance verification. Intelligent document parsing and search capabilities using <b>AI Search</b> and <b>MarkitDown</b> tools for comprehensive file processing. Automated validation workflows with real-time analysis and reporting for regulatory compliance. Seamless CI/CD integration with <b>GitHub Actions</b> for continuous deployment and testing. <b>ReactJS</b> frontend deployed on <b>Azure web app</b> - <b>Tech stack:</b> Python, Azure AI Services, GPT-4.1 mini, CrewAI Tools, AI Search, MarkitDown, GitHub Actions, Azure Web Apps, ReactJS	09/2024 – present Casablanca, Morocco
<b>Data Scientist &amp; Artificial Intelligence Engineer · Stage PFE, Capgemini TS</b> - <b>Codebase Analysis &amp; Documentation Generator</b> – Intelligent Code Understanding System: Dual-path processing system combining vector embeddings and knowledge graph construction for comprehensive codebase analysis. Automated directory and file extraction with entity identification for code components and relationships mapping. Semantic code search using <b>Pinecone</b> vector database for contextual code retrieval. Knowledge graph representation in <b>Neo4j</b> database for code relationship. Text-to-Cypher query for interacting with code graph structures. - <b>Tech stack:</b> Python, Pinecone, Neo4j, Gemini 1.5 pro, google embeddings, Recursive chunking, Langchain.	03/2024 – 08/2024 Casablanca, Morocco
<b>Data scientist · Stage PFA, Office National des Chemins de Fer (ONCF)</b> - <b>Data Analysis &amp; System Migration:</b> Analysis of supplier and item data from the Procurement Information System. <b>ETL pipelines</b> for migration preparation. Analytical dashboards with Python ( <b>Pandas</b> , <b>NumPy</b> ) for procurement insights visualization. - <b>Tech stack:</b> Python, Pandas, NumPy, Microsoft SQL Server, Power bi.	08/2023 – 09/2023 Rabat, Morocco
<b>Full-stack Developer · Stage PFE, Société Tangéroise de Maintenance (STM)</b> Development of a web application for taxi booking.	04/2022 – 06/2022 Tangier, Morocco
<b>Stage d'apprentissage, Institut National de la Recherche Agronomique</b> Use of MATLAB programming language in the study of goat behavior.	04/2021 – 06/2021 Tangier, Morocco
<b>Business Analyst · Saisonnier, M'nar Park</b> Collection and organization of multi-source data for strategic analysis. Development of reports and dashboards optimizing operational performance.	06/2019 – 08/2019 Tangier, Morocco

# Freelance

---

## Machine Learning Engineer

Remote

### - Personalized Learning Analytics System

- Built a machine learning system that analyzes student behavior on an e-learning platform.
- Used student onboarding forms (ILS and HEXAD) to determine learner profiles and motivations.
- Designed a tracking system for course progress, quiz attempts, and interactions with motivational messages.
- Implemented feedback loops using "Like/Dislike" reactions to messages, enabling adaptive encouragement strategies.
- Provided actionable insights on student engagement and personalized learning experiences.
- **Tech stack:** Python, Pandas, Scikit-learn, Streamlit, SQLite (or relevant), JSON, Matplotlib

## GenAI Engineer

Remote

### - AI Agent for Commercial Property Ownership Research

- Developed an intelligent AI agent system for automated commercial property ownership research and due diligence
- Built multi-source data integration pipeline connecting public records, property databases, and legal document repositories
- Implemented entity extraction and relationship mapping using GPT-4 for property ownership chains and corporate structures
- Created automated web scraping workflows with intelligent data validation and cross-referencing capabilities
- Designed knowledge graph representation using Neo4j to visualize complex ownership relationships and property portfolios
- Built semantic search capabilities for property records using vector embeddings and contextual retrieval
- Developed comprehensive reporting system with automated property analysis, risk assessment, and ownership timeline generation
- Integrated compliance checks for regulatory requirements and due diligence standards
- **Tech stack:** Python, OpenAI GPT-4, Neo4j, BeautifulSoup, LangChain, Faiss, FastAPI, PostgreSQL

### - Intelligent Book Recommendation Platform – AI-Powered Literary Discovery System

- Built robust backend service architecture using Python and FastAPI for scalable recommendation engine deployment.
- Designed and implemented comprehensive RAG (Retrieval-Augmented Generation) pipeline using LangChain for contextual book discovery Integrated vector database system (Weaviate/Pinecone) for efficient storage and semantic querying of book embeddings and user preferences.
- Developed intelligent conversational assistant using OpenAI GPT-4o with context-aware prompt engineering for personalized literature recommendations Created sophisticated "taste graph" algorithm combining semantic search results with deep user preference analysis and reading history Implemented advanced embedding models for book content analysis, genre classification, and user preference mapping.
- Built real-time recommendation system with PostgreSQL integration for user profile management and recommendation tracking.
- Designed scalable microservices architecture supporting high-volume concurrent user interactions and recommendation queries.
- **Tech stack:** Python, FastAPI, LangChain, OpenAI GPT-4o, Pinecone, PostgreSQL, Semantic Search, RAG Architecture, Embedding Models, Microservices.

### - English Accent Analyzer – AI-Powered Multilingual Audio Intelligence System

- Built comprehensive video-to-audio processing pipeline supporting YouTube URLs and direct video file formats.
- Implemented advanced audio preprocessing with noise reduction using Deep Filternet for enhanced accent detection accuracy.
- Integrated Wav2Vec2 transformer model via Hugging Face for sophisticated accent classification (British, American, Australian, etc.)
- Created confidence scoring system with AI-powered natural language explanations using OpenRouter API integration.
- Designed user-friendly interface with real-time processing capabilities and detailed analytics dashboard.
- Automated multimedia content analysis with seamless video download and audio extraction workflows
- **Tech stack:** Python, Streamlit, Hugging Face Transformers (Wav2Vec2), MoviePy, LibROSA, Deep Filternet, yt-dlp, OpenRouter API (Mistral), Audio Processing Libraries

# Technical Skills

---

## Generative AI

- **LLMs:** OpenAI GPT models, Google Gemini , Claude, Mistral & others
- **Model Integration:** LiteLLM, Portkey, OpenRouter, Langgraph
- **LLM Frameworks:** LangChain, Langgraph, LlamaIndex
- **Multi-Agent AI :** CrewAI, Google ADK, Autogen, Semantic Kernel,
- **Vector Databases:** Vertex AI Vector Search, Pinecone, AWS OpenSearch Service, Weaviate, Qdrant, ElasticSearch , Azure AI Search
- **Chunking:** Chonkie (Framework), Recursive chunking, Semantic chunking, context-aware
- **Parsing :** Docling, Markitdown, Unstructured, pdfplumber, PyMuPDF

## Cloud Platforms & AI Services

- **Google Cloud Platform:** Vertex AI, Cloud Run, Cloud Storage, BigQuery, GCP ADK .
- **Amazon Web Services:** Bedrock, Lambda, API Gateway, S3,
- **Microsoft Azure:** AI Services, Web Apps, Azure OpenAI, Azure Blob Storage

## Python Development & APIs

- **Programming:** Python, SQL
- **API Development:** FastAPI, Flask
- **UI Frameworks:** Streamlit, Gradio
- **Package Management:** Poetry , virtual environments

## Database Management

- Mongodb, Neo4j, PostgreSQL

## DevOps & Infrastructure

- **CI/CD:** GitLab CI/CD , Github actions
- **Version Control:** Git
- **Infrastructure as Code:** Terraform

## AI Evaluation & Testing

- **Evaluation Frameworks:** DeepEval, RAGAS, TruLens, MLFlow
- **Metrics:** Faithfulness, answer relevancy, retrieval accuracy, hallucination detection

# Certifications

---

- |   |                                   |   |                                |
|---|-----------------------------------|---|--------------------------------|
| - OCI Generative AI Professional – Oracle | - Microsoft Azure AI Fundamentals | - AWS Partner: Generative AI Essentials | - Google Cloud : Generative AI |
|---|-----------------------------------|---|--------------------------------|

# Hackathon

---

Google cloud GenAI Hackathon 2024

Microsoft Generative AI Hackathon 2025

AWS Generative AI Hackathon 2024

# Education

---

<b>Master's in Mobility and Big Data</b> , <i>Faculty of Sciences and Technologies</i>	09/2022 – 07/2024 Tangier, Morocco
<b>Bachelor's in computer engineering</b> , <i>Faculty of Sciences and Technologies</i>	10/2021 – 07/2022 Tangier, Morocco
<b>Diploma of Scientific and Technical University Studies (DEUST): MIP</b> , <i>Faculty of Sciences and Technologies</i>	09/2018 – 06/2021 Tangier, Morocco

# Languages

---

**English: Advanced** — EFSET English Certificate (CEFR Level C1) • **French: Advanced**