Victor Almeida

Linkedin | +55 79 99651-3762 | victorpy.1999@gmail.com | GitHub

Skills

- Python | SQL | AWS | LangChain | LangGraph | ETL | Kubernetes | Java | Springboot | FastAPI | SQLModel | Pydantic | OpenAI API | Microservices | JavaScript | GIT | Azure | Postgres | Docker | Lambda | S3 | CloudWatch | DynamoDB | SQS
- Back-end | PowerBI | APIs | CI/CD | RESTful Services | GraphQL | Node.js | Jenkins | Docker Compose | Serverless Architecture
- Data Science | Machine Learning | Data Analysis | Data Engineer | Cloud Service | DevOps | Deep Learning | LLM | LLM Applications | Prompt Engineer | Data Pipelines | Unstructured / Structured Data Processing | Vector Databases | Retrieval-Augmented Generation (RAG) | Model Deployment | MLOps

Experience

Johnson & Johnson - Data Scientist & IA/ML Developer 03/2023 – Present

- Built a **RAG** chatbot using **LangChain** and **LangGraph**, designing workflows with nodes for web searches, question rephrasing, and validations. Leveraged **Pydantic** and **SQLModel** for data consistency, while implementing **CI/CD** pipelines for automated testing and deployment. This solution significantly improved information accessibility, reducing over **5,000 hours** in non-field time for medical representatives, enabling more effective client interactions.
- Developed and maintained a FastAPI-based API for RAG chatbot management, integrating JWT for secure authentication and SQLModel for safe database mapping. Enabled users to create and manage custom chatbots with full control over permissions, content segregation by subjects, and AWS S3-triggered serverless data extraction pipelines. This API empowered teams to deploy chatbots quickly, personalize content, and securely manage access, enhancing agility and effectiveness across departments.
- Applied advanced **prompt engineering** and data validation techniques using **LLMs**, **OCR** with **OpenCV**, **TensorFlow** and **PyTorch** to process and extract relevant information from diverse document types.
- Developed a question-answer classifier using **Random Forest**, managing the full **ML pipeline** from data analysis to model evaluation. Utilized **Pandas**, **Seaborn**, and **Scikit-learn** for feature engineering, data transformation, and improving model accuracy through cross-validation.
- Designed and implemented serverless, event-driven workflows using **AWS Lambda**, **S3**, **DynamoDB**, and **Step Functions** for real-time, scalable data processing and automation.

Johnson & Johnson - AI Backend Developer (1y 2m) 01/2023 –03/2024

- Migrated the solution to the **LangChain** framework, developing a data extraction pipeline to handle **structured** (Excel) and **unstructured** data (PDF, DOCX, PPTX, HTML) through LangChain's **tools** and **chains**. This transition improved adaptability, enabling efficient management of diverse data sources and more accurate responses to complex queries. LangChain's modular design boosted **scalability**, **simplified maintenance**, and offered customization options, allowing teams to integrate **multiple data sources** seamlessly and support **high-impact**, **real-time interactions.**
- Served as the lead backend developer for an **AI-powered** application, creating a full data **extraction pipeline** (**ETL**) using **Python** with libraries such as **PDFMiner**, **pdfplumber**, **BeautifulSoup**, **Selenium** for web scraping, and **Aspose** for PowerPoint document extraction. This ensured the efficient handling of unstructured data in various formats.
- Designed and deployed a serverless data extraction pipeline on **Azure**, utilizing **Azure Functions**, **Blob Storage**, and Azure's embedding models. This architecture was optimized to minimize computational and financial costs, while ensuring **scalability** and integration with other Azure services.
- Implemented a similar serverless data pipeline in **AWS**, leveraging **AWS Lambda** for function execution, **Bucket S3** for storage, **CloudWatch** for monitoring, **AWS Transcribe** for speech-to-text, **AWS Translate** for multilingual capabilities, and **DynamoDB** for storage. This pipeline provided seamless integration with AWS cloud services and enhanced scalability.
- Developed a **RAG-based** chatbot by building an API with **FastAPI**, hosted on **AWS EKS** and using a microservices architecture. The chatbot retrieved embeddings from a vector database in real time, with responses generated through a large language model (**LLM**) hosted in Azure, ensuring robust and timely interaction capabilities.
- Played a key role in database architecture, designing and maintaining **SQL** databases. Performed **ETL** processes, created and optimized **tables**, **views**, **stored procedures**, and **routines**, ensuring the system could handle large volumes of data effectively and maintain high performance.

Education

Data Science Technologist 01/2023 - 06/2025

UNINASSAU - Centro Universitário Maurício de Nassau

Languages

Portuguese ..... Native
English ..... Advanced
Spanish ..... Basic

Projects

Mudi-SpringMVC Platform for Delivery Coordination (Spring Security, JPA, Thymeleaf) 11/2022

A SpringMVC-based platform connecting users for order placements and deliveries. Utilizes Spring Security for secure authentication, Spring Data JPA for database interactions, and Thymeleaf with Bootstrap for a responsive UI.

SearchVerse API for Educational Metaverse Indexing (Spring Boot, PostgreSQL, Heroku) 09/2022

Developed an API cataloging educational metaverses with keyword search functionality. Built with Spring Boot and JPA, deployed on Heroku, and featuring a frontend using HTML5, CSS3, and JavaScript ES6.

Others

Johnson & Johnson 1000 Devs Program Graduate 06/2022 – 12/2022

Completed the intensive, six-month 1000 Devs program by Johnson & Johnson and mesttra., focused on developing young professionals in backend development, Java, and key soft skills. Achieved the top score in the cohort while deepening technical skills and knowledge in agile methodologies, effective communication, and team collaboration. Gained experience working with industry mentors and connecting with peers from diverse backgrounds. This comprehensive program not only enhanced my technical abilities but also paved the way for my role at Johnson & Johnson, where I continue to contribute meaningfully.