

Bernardo Uziel Franco Torres

AI Engineer - Python Developer

Contact

Phone

+52 993-112-3961

Email

uziel.francotorres@gmail.com

City

Mérida, Yucatán, México

GitHub

github.com/Aplex2723

Education

2021 - 2025 (In-Progress)

**Information Technologies and
Digital Business**

Anahuac Mayab

2018 - 2021

High School

Colegio Latinoamericano de Tabasco

Expertise

- Azure
- Python
- Databases
- Azure AI Services
- Google Cloud
- AWS
- LangChain
- LLM's
- Networking & Security

Languages

**English - Complete professional
competence**

Spanish - Native

AI Engineer specializing in **GenAI, LLMs, RAG, and Fine-Tuning**, with a focus on developing chatbots and intelligent virtual assistants. With over 4 years of experience building AI-driven microservices, I manage the entire development lifecycle, from architectural design to deployment. Throughout my career, I have worked with multiple technologies and programming languages, enabling me to tackle complex problems with efficient and scalable solutions.

Proficient in cloud computing with expertise in IaaS platforms such as **Azure, AWS, and GCP**. I specialize in designing robust and scalable solutions tailored to project requirements. I am a strong advocate of teamwork, promoting open communication and knowledge sharing to foster innovation. Passionate about solving complex problems, I excel at tackling new challenges and contributing to the advancement of AI technologies that redefine human-machine interactions.

Experience

Dec 2023 - Present

Sonata Software - Full time - Guadalajara, Jal. Remote

Technical Support Engineer - AI services

In charge of providing support, guidance, and technical consulting to Microsoft's strategic and large enterprise customers in the area of **Azure AI Services** (Cognitive Services), including all their Artificial Intelligence services (Document Intelligence, Speech Services, Copilot, AI Search, OpenAI, etc.). This includes supporting these services through REST APIs and the Azure SDK, ensuring fast responses to technical issues and meeting customer expectations to achieve desired satisfaction.

Clear, fluid, and empathetic communication in English with customers and the team is employed to understand problems and needs, providing the best possible solutions or architectures to fulfill customer requirements.

Achievements:

- **DNA Award - Run for It** (Jul 2024):

For ensuring quick responses and significantly improving customer satisfaction, achieving **35% positive feedback** across the region and team.

- **DNA Award - Play to Win** (Sep 2024):

For excelling in task management and resolving customer issues with impressive efficiency, making a significant impact on customer satisfaction and experience.

- **3 Times in a Row Star of the Quarter**

Sep 2023 - Jun 2024

Bibli.ai - Part time - Austin, Texas. Remote

Python AI Back-end Developer

Responsible for the development and implementation of serverless microservices in **Microsoft Azure Functions (Serverless Computing)** for an AI tutor platform that creates written and audio stories to help users learn English by answering their questions and doubts (regarding the stories). This included the deployment of these services. To enhance the AI tutor's capabilities, implemented Retrieval-Augmented Generation (RAG) techniques with Pinecone vector databases to provide the generated stories directly to the AI tutor. This ensured highly accurate and relevant question-and-answer responses to learners' inquiries, improving their overall learning experience.

Developed REST APIs with Python for the service development using multiple libraries/APIs such as OpenAI, Twilio, Azure SDK, etc. Implemented artificial intelligence services within multiple microservices, including Whisper, GPT model's, and Azure Speech Services, to facilitate interactive and personalized learning experiences.

Utilized non-relational databases (NoSQL) with Azure Storage, managing containers and blobs for document management through the Azure SDK, ensuring efficient data handling for the AI tutor application.

Implemented CI/CD with GitHub Actions and DevOps practices within an Agile methodology framework, optimizing the microservices development lifecycle and ensuring effective validation testing for the AI-based learning platform.

Oct 2019 - Present

codey-tech.com - Mérida, Yucatán

Freelance AI Developer & PM

In charge of performing, coordinating, and managing the technological aspects of my business, I currently help startups and clients meet and achieve their goals using AI services. My focus is on problem-solving and the development of high-level technological projects to deliver the best possible products to customers. I have successfully brought projects from scratch to production using multiple PaaS platforms, with a particular emphasis on AWS and Azure. By leveraging services such as Amazon Bedrock, EC2, Azure App Services and other Cloud offerings, I ensure the infrastructure is robust and scalable, capable of supporting any project's demands.

I specialize in the development of virtual assistants, leveraging cutting-edge technologies such as **LangChain** (LangGraph) and **AutoGen** to integrate various large language model (LLM) agent architectures. This ensures the effective fulfillment of client expectations. Additionally, I design and develop intelligent chatbot systems capable of performing specific tasks within a system, such as user registration and facilitating interactions between different tools. This includes architecting these agents or microservices using the latest advancements in artificial intelligence technologies.

I have extensive experience working with both open-source Large Language Models (LLMs) and Small Language Models (SLMs), including Phi, Gemini, Mixtral, and OpenAI's GPT series. My expertise encompasses managing large volumes of documents using **Retrieval-Augmented Generation (RAG)** techniques, which enhance information retrieval and processing through vector databases such as Pinecone and Chroma. Additionally, I am skilled in fine-tuning models to meet specific client requirements, ensuring optimal performance and relevance across diverse applications.

Jun 2024 - March 2025

BAMX Mérida - Part time - Mérida, Yucatán

Principal Software Engineer

Led the design and implementation of backend and operational systems, focusing on translating client needs into technical specifications and actionable solutions. Developed an inventory management system for BAMX using advanced tools such as Django REST API, API-based microservices design, and PostgreSQL, ensuring security, scalability, and accessibility. Utilized AWS services including **EC2, Route 53, and VPC** to architect a robust infrastructure tailored to BAMX's requirements. Additionally, contributed to frontend development with React Native to deliver a seamless user experience.

- **Client Liaison and Specification Development:** Served as the main point of contact, transforming client requirements into clear technical specifications.
- **Backend and API Design:** Designed secure and scalable backend architectures using Django REST API and PostgreSQL, implementing best practices for RESTful API development to ensure maintainability and integration ease.
- **AWS Infrastructure Implementation:** Leveraged AWS services (EC2, Route 53, VPC) to build a high-availability, scalable infrastructure that supports BAMX's solutions effectively.
- **Frontend Development and Integration:** Assisted in developing the frontend with React Native, ensuring cohesive integration between backend services and the user interface.
- **Team Coordination and Quality Assurance:** Coordinated the multidisciplinary team to meet project deadlines and maintain quality standards, implementing security and usability features tailored to diverse user needs.

Certifications



Azure OpenAI Specialist

Cloud Academy - Jan 2024



Cognitive Services Specialist

Cloud Academy - Jan 2024



Data, ML, and AI in Google Cloud

Google Cloud - Nov 2022



Convolutional Neural Networks in TensorFlow

DeepLearning.ai - May 2023



Artificial Intelligence with Machine Learning in Java

Oracle - Nov 2023



Cloud Computing Foundations

Duke University - Nov 2024



IBM AI Developer

IBM - Sep 2022



Networking and Security in Google Cloud

Google Cloud - Sep 2022



Software Engineering: Software Design and Project Management

Hong Kong University - Nov 2024