

# Abdulaziz Khader

[akhader@ucsd.edu](mailto:akhader@ucsd.edu) | [www.linkedin.com/in/abdulaziz-khader](https://www.linkedin.com/in/abdulaziz-khader) | [aokhader.vercel.app](https://aokhader.vercel.app)

## EDUCATION

University of California – San Diego

Bachelor of Science in Computer Engineering, Minor in Mathematics

Graduation: **June 2026**

**GPA: 3.7/4**

## TECHNICAL SKILLS

**Programming Languages:** Python, C/C++, C#, SQLite, Go, JS, TypeScript, Java

**ML Frameworks:** LangChain, LangGraph, Keras, TensorFlow, PyTorch, CUDA, NLTK, spaCy

**Other Frameworks:** React, NextJS, MySQL, GCP, Docker, llama.cpp, Ollama, vLLM, FastAPI, Puppeteer

## EXPERIENCE

**BMC Software**

Software Engineering Intern

Santa Clara, CA

June 2025 – Aug. 2025

- Created a **new AIOps** tool that integrates the user's **ServiceNow** data with **BMC's** existing products into a universal hub, resulting in a **17%+ projected revenue increase** and expecting **100,000+** new users.
- Utilized **LangChain** and **LangGraph** to process the customer's database by creating **CRUD tools** and exposing **ServiceNow's REST API**; hosted and optimized an **ASGI web server** for real-time updates **60%** faster.
- Built a **semantic model** to dynamically parse the user's data configurations; redesigned **1000+** lines of code into a tool factory to increase **scalability**, **efficiency**, and **usability** with other products such as **Salesforce**.

**Product Manager Accelerator**

Software Engineering Intern

Remote

Sep. 2024 – Dec. 2024

- Built a **full-stack** web app integrating **reinforcement learning** with a scalable database to evaluate player chess games and generate insights into move patterns and historical trends.
- Integrated frontend workflows with **Flask REST APIs**, **LangChain**, and **Google Gemini**, ensuring seamless delivery of chess insights to the UI and improving analysis response time for users.
- Deployed a **Supabase** database to store chess positions, reducing duplicate lookups and retrieval latency by **40%**.

**Ocelot**

AI Studio Fellow

Remote

Aug. 2024 – Dec. 2024

- Improved tone accuracy **30%** over GPT-2 baseline by building a **sentiment-conditioned** LLM on student mental health conversations, validated through human evaluation scoring **4.1/5.0** on appropriateness.
- Optimized 7B-parameter model deployment by reducing memory from 14GB to 3.5GB via **4-bit quantization** while preserving **98% model quality**; enabled **real-time inference** on local hardware via **Gradio**.
- Accelerated development cycle by **82% through CUDA optimization**, reducing training from 8 hours to 30 minutes; processed **5,600+** multi-domain training samples with **stratified evaluation** framework.

## PROJECTS

**RoboRanger — Best Edge AI & ElevenLabs Award** | *Python, C++, TFLite, llama.cpp* April 2026 – Present

- Deployed a full edge AI stack on an Arduino Uno Q, pairing a **MobileNetV3-Large** TFLite classifier (calibrated via **temperature scaling** and **TTA**) with a quantized SmolLM2-360M LLM via llama.cpp for fully offline, sub-1B-parameter inference.
- Engineered a **RAG system** using bge-small-en-v1.5 embeddings and **sqlite-vec** for vector search, designing a hybrid context strategy that injects always-on structured species metadata alongside retrieved passages and applies **cosine-similarity** thresholding in the retriever to gate hallucinations.
- Built an LLM-assisted data curation pipeline using local Llama 3.1 8B via Ollama with grammar-constrained YAML decoding and **human-in-the-loop review**; architected a dual-mode runtime falling back to a **networked ElevenLabs TTS agent** when connectivity is available.

**YouPresent** | *Unity, C#, Meta XR, Model Rigging, 3D User Interaction*

April 2025 – Feb 2026

- Built a Unity VR app for public speaking practice using **Google Gemini** to analyze user gaze, hand motion, and delivery in real time, generating personalized coaching feedback through an AI-driven virtual coach.
- Leveraged **Adobe Mixamo** to rig an interactive audience and personal coach; created recording system that captures presentations with a personalized speaker avatar and presented it at the AWE XR 2025 conference.