

ALONSO PERALTA

Texas, USA | +1 (907) 744-8406 | alonso.pe03@gmail.com | linkedin.com/in/aperalta03 | alonsoperalta.com

EDUCATION

Texas A&M University

College Station, TX

Master of Computer Science; GPA: 3.75/4.00

Dec. 2027

- Analysis of Algorithms, Deep Reinforcement Learning, Information Storage and Retrieval, Artificial Intelligence, Cloud System Design, Machine Learning

Bachelor of Science in Computer Engineering, Minor in Mathematics; GPA: 3.65/4.00

May 2026

- Advanced Software Engineering, Data Structures & Algorithms, Database Systems

EXPERIENCE

Lead AI Engineer

Jan. 2025 – Present

ValencAI

Remote

- Own design, development, review, and task assignment for **ExtractAI** and **SDS Caller**
- Built ExtractAI to turn inconsistent chemical PDFs into structured JSON for client databases and Valenc's authoring system
- Designed a **layout-aware VLM** schema that reads SDS structure and meaning across company formats, then emits one semi-structured JSON
- Built the normalize pipeline with exact match, **TF-IDF**, **BM25**, **vector search**, aliases, cleaning, and scoped LLM reasoning
- ExtractAI matched or exceeded human data entry on **98%** of extracted fields
- Cut SDS processing from 2–3 weeks to **15 minutes** and cost from \$2k–\$3k per about 150 SDSs to **\$30**, compared to manual labor
- Shipped ExtractAI as a standalone API, reached a **couple thousand dollars** in revenue, then promoted it to a **flagship** feature
- Built SDS Caller, a **voice agent** that takes a chemical and a situation on a live call and returns the matching SDS section, such as Section 13 for disposal
- Retrieval runs over ExtractAI-structured SDS data using TF-IDF, BM25, and **semantic search**, with an orchestration layer that collects the question and formats the answer

Co-Lead Full Stack Developer

Jan. 2024 – Dec. 2024

NeuroX1

College Station, TX

- Co-led a team of 4 on **MedGet** so NeuroX1 researchers can find drug candidates instead of searching the database by hand
- Built a hybrid retrieval stack: a general NLP model plus **BioMixtral 8x22B** as the specialist for the center of excellence
- The same users found candidates **5x faster** with MedGet than with a manual database search
- Cut employee-hour cost of that search by **62%**, from about \$60/hr for a PhD researcher to about \$15/hr for a VM
- Owned the full-stack app: Python and Streamlit to lock the AI path, then React and Node for the product UI researchers use to browse and navigate
- NeuroX1 funded the work at **\$2000/month** and later put MedGet into their workflow

TECHNICAL SKILLS

Languages: Python, TypeScript, JavaScript, Bash, C++, SQL, Ruby, HTML, CSS

Frameworks: FE: React, Streamlit, Tailwind; BE: Node, Express, Ruby on Rails, OpenAPI, FastAPI, Pydantic; Cloud: Terraform, Redis, ARQ, Microservices, Logging, IAM; AI: LangChain, LangGraph, LlamaIndex, CrewAI, OpenMCP, DocAI, LLM (Cloud + Local), pgvector, RAG, Fine-Tuning, AIVoice

Developer Tools: AWS, GCP, Git, Docker, CI/CD, AI Tooling, Excel

Spoken: English, Spanish

Certificates: Cypher Query Language; AWS Architecting Solutions, Designing Data Lakes, Cloud Technical Essentials

Interests: Basketball, Gym, Billiards, Traveling, Finance, Startups, System Design, AI/ML, Quantum Tech