

CINTIA I. SHINODA

DATA SCIENTIST - AI ENGINEER - DATABASE DEVELOPER

São Paulo, Brazil - cintia.builds@gmail.com

linkedin.com/in/cintiashinoda | github.com/cintia-shinoda

SUMMARY

Data Scientist focused on applied machine learning, predictive modeling and network analysis. Experience with real-world data at scale, from database design to modeling and evaluation.

PROJECTS

Astronomy FAQ Chatbot with RAG (Retrieval Augmented Generation) and a Local LLM

Brazilian Portuguese-language RAG pipeline with vector retrieval, fully local generation at no API cost and a history-aware retriever. Evaluated against a 30-question golden set, reaching 90% Recall@3. | Stack: Python, LangChain (LCEL), FAISS, sentence-transformers (all-MiniLM-L6-v2), Mistral-7B-Instruct via Ollama, Chainlit

Multi-Agent System for Orbital Collision Risk Assessment

Multi-agent architecture for screening conjunction risk between objects in orbit, combining supervised classification, graph analysis and 3D geospatial visualization | Stack: Python, Google ADK, LightGBM, NetworkX, CesiumJS

NLP Pipeline for Sentiment Analysis of Brazilian Legislative Speeches

Processing, sentiment analysis and topic extraction from São Paulo City Council's speeches using Natural Language Processing techniques | Stack: Python, NLP

SKILLS

Programming Languages: Python, SQL, R

Databases: Oracle, PL/SQL, SQL Server, MongoDB, FAISS (vector)

AI: RAG, AI Agents, LangChain (LCEL), *embeddings*, vector databases (FAISS), Local LLM (Ollama), Prompt Engineering, LLM Evaluation (*Golden Set*, Recall@k), Google ADK

Machine Learning: NumPy, Pandas, Scikit-Learn, LightGBM, NetworkX, GeoPandas

Data Engineering: Data Modeling, ETL (*Extract, Transform, Load*)

BI and Visualization: Power BI, Looker Studio, Matplotlib, Streamlit, Chainlit

Languages: Brazilian Portuguese (native) - English (advanced) - Japanese (elementary - JLPT N5) - French (elementary)

EXPERIENCE

Database Developer and Statistics - Internship | E-VAL Tecnologia, São Paulo, SP, Brazil

2016 - 2017

- Modeled and built the database infrastructure supporting software products and integrations for two major Brazilian healthcare institutions: Hospital Israelita Albert Einstein (Digital Signature Web Service) and Central Nacional Unimed (Diabetes Risk Prediction Model).
- Documented data models, schemas and integration specifications throughout the development cycle.
- Tested integrations between internal products and client systems.

Data Analyst | HandsOn.TV, Mountain View, CA, USA

2013 - 2014

- Owned Analytics for the HandsOn.TV content platform, tracking engagement across the website, Facebook, and YouTube to inform editorial decisions and distribution strategy.
- Collaborated with a Brazil-based team, working across time zones.

EDUCATION

B.Sc. in Data Science, UNIVESP

2022 - 2026

Undergraduate thesis: Vulnerability and resilience of the São Paulo public transit network: a topological analysis based on graph theory.

Specialization in Applied Mathematical Methods, UTFPR

2024 - 2025

Specialization in Artificial Intelligence & Machine Learning, FIAP

2020 - 2021

B.Tech. in Software Development, FIAP

2015 - 2017

Additional specializations in progress: Robotics, Advanced Python and AI Agents