Pedro Henrique Goncalves de Paiva

San Francisco, CA · pedro@uni.minerva.edu · 415-336-4271 · in/pedrohgpaiva · portfolio

Professional Summary

Computational Sciences student (Data Science) who takes ambiguous and data-heavy problems from discovery to shipped products. Co-founded and leads product for an education org serving 400+ students across 14 cohorts and builds and evaluates AI/ML systems end to end, from user problem to deployed tool. Comfortable owning decisions with incomplete information and iterating quickly.

Education

Minerva University San Francisco, CA

B.Sc. in Computational Sciences, concentration in Data Science *August 2023 – May 2027*

- Relevant Coursework: Machine Learning (Finding Patterns in Data), Bayesian Statistics, Modeling Complex Systems, Algorithms & Data Structures, Probability & Statistics, Econometrics, Linear Algebra, Calculus
- Google Developer Student Club; Code2040 Fellows Program Alumni; Goldman Sachs Engineering Emerging Leaders Series

Experience

Co-Founder & Product Lead

Remote, Brazil

AURA Application Mentorship *January 2023 – Present*

- Own the product and curriculum roadmap end to end, turning an open-ended "help students get into university abroad" problem into a structured system of courses, resources, and a support platform.
- Prioritize what to build using both qualitative student feedback and quantitative funnel and outcome metrics, identifying pain points across the application journey and shipping the highest-impact improvements first.
- Translated user needs into requirements for AURA's online platform and partnered with development and design collaborators to centralize resources and student communication.
- Launched 3 courses across 14 cohorts and scaled the support model, contributing to 700+ university acceptances.

Innovations in Science & Technology Instructor and Capstone Supervisor New Haven, CT

Yale University *June 2024 – Aug 2026*

- Designed and taught seminars on recommendation systems, algorithmic bias, and digital ethics to 135 international high school students across three summer sessions.
- Paneled program-wide sessions for 650 students; oversaw 95+ students in the Psychology & Neuroscience capstone track.

Research Assistant, MateMarote: Building the Learning and Mood Dataset Buenos Aires, AR

Neuroscience Lab, Universidad Torcuato Di Tella (UTDT) *September 2025 – May 2026*

- ∉ Designed the data model, variable schema, and quality controls (double-entry and inter-rater checks) for a 15-year, ~1,300-entry research dataset.
- ∉ Built reproducible R and Python pipelines to digitize, clean, and link the data, making previously inaccessible records analysis-ready for researchers.

Teaching Assistant, Introduction to CS (Formal Analysis 50 and 51) San Francisco, CA

Minerva University *August 2024 – May 2026*

- Supported 100+ students in programming, statistics, and formal logic, diagnosing where they got stuck and adapting technical support to each need.
- Designed and piloted an oral-defense assessment workflow to evaluate reasoning and improve academic integrity.

Projects

Predicting Spotify Skips from Song Intros

- Built a full ML pipeline on 92,445 personal streaming events to test whether a song's first 10 seconds predict skips, comparing handcrafted features, frozen Audio Spectrogram Transformer embeddings, a scratch CNN, and a fine-tuned AST (PyTorch, scikit-learn, librosa).
- Designed context-only and fusion baselines as controls to test the audio result honestly, showing listening context (75.1% CV accuracy) explained far more than audio alone (61.8%), and reported the limitation rather than overclaiming.

Dengue Surveillance Dashboard

- Deployed a dashboard forecasting dengue outbreaks across all 27 Brazilian states, consolidating state-level risk information for public-health planning. Benchmarked four forecasting approaches with expanding-window backtesting to decide which model was reliable enough to ship.

RememberMe – UCBerkeley - Hack for Impact Hackathon

- Led a 4-person team to build an AI memory-support prototype for people living with dementia in 24 hours. Integrated the TwelveLabs and Vespa APIs and won Best Hack for Quality of Life and Best Use of TwelveLabs API.

Skills and Interests

Technical Skills: Python, SQL, R, C · ML/AI: transformers (AST fine-tuning), CNNs, transfer learning, model evaluation; PyTorch, scikit-learn, PyMC, ArviZ; AI APIs (TwelveLabs, Vespa) · Data: NumPy, pandas, SciPy, data modeling, pipelines, Git

Languages: Portuguese (native); English (fluent); Spanish (intermediate)