

Jonathan O. Tellechea

Los Angeles, CA | tellecheajonathan@gmail.com
jonathantellechea.com | linkedin.com/in/jotellechea | github.com/jotellechea

Education

University of California, Santa Cruz

Santa Cruz, CA

Bachelor of Science, Physics

2017 – 2020

Thesis: [Prospects for discovery of ttHH production at the High-Luminosity LHC using Boosted Decision Trees](#)

Technical Skills

Programming Languages: Python, Bash, Swift, LaTeX.

Machine Learning & Data Science: PyTorch, TensorFlow, Keras, Scikit-Learn, Pandas, NumPy, Boosted Decision Trees, Deep Neural Networks, Transformers, feature engineering.

Scientific Computing & Tools: ROOT, uproot, awkward-array, Coffea, Matplotlib, Seaborn, Docker, Linux, High Throughput Computing clusters, XRootD.

Experience

High Energy Physics Lab, University of Virginia

2021 – 2023

Postbaccalaureate Researcher

Charlottesville, VA

- Engineered Python analysis pipelines over CMS Open Data in NanoAOD format, pulling datasets from CERN EOS cloud storage via XRootD and processing events with Coffea, awkward-array, uproot, NumPy, and Matplotlib to reconstruct W boson invariant mass and missing transverse energy and validate published cross-section measurements.
- Delivered a columnar Coffea-based workflow as a replacement for CentOS 7 CMSSW Docker containers, which had blocked outside users, lowering the technical barrier to public CMS Open Data analysis as part of an outreach initiative.
- Architected a characterization and stress-testing framework for Silicon Photomultipliers, operating a CAEN FERS digitizer to acquire time-of-flight, time-over-threshold, and signal-to-noise measurements across a grid of candidate sensors and tracking per-device dark current growth in a controlled enclosure to drive component selection for MIP Timing Detector integration.

Santa Cruz Institute for Particle Physics, UC Santa Cruz

2020 – 2021

Machine Learning Researcher

Santa Cruz, CA

- Spearheaded the transition from Boosted Decision Trees to Deep Neural Networks in TensorFlow and Keras for rare-signal detection against overwhelming, highly imbalanced background classes, improving statistical significance 174 percent from 0.35σ to 0.96σ .
- Architected a configurable DNN with parameterized depth, neuron count, and dropout rate driven by a single training function, enabling systematic hyperparameter search across layer configurations, batch sizes, and regularization strengths with EarlyStopping and ModelCheckpoint to prevent overfitting and persist best weights.
- Built a CSV-based experiment tracking system logging model configuration, runtime, AUC, average precision, confusion matrix entries, maximum signal significance, and signal and background yields, enabling reproducible comparison across hundreds of training runs.

Santa Cruz Institute for Particle Physics, UC Santa Cruz

2019 – 2020

Undergraduate Research Assistant

Santa Cruz, CA

- Engineered a Boosted Decision Tree classifier with Scikit-Learn AdaBoost to separate a rare signal channel from dominant background processes, raising signal significance 35 percent from a 0.26σ rules-based baseline to 0.35σ at a model AUC of 0.86.
- Replicated the published cut-based analysis of ATLAS Public Note ATL-PHYS-PUB-2016-023 as a performance baseline, reproducing the released plots in Python, ROOT, NumPy, and Matplotlib to validate the pipeline before any model was applied.
- Engineered high-level features from raw kinematic variables, normalized inputs with StandardScaler, and produced feature importance analysis identifying b-tagged jet counts and average pseudorapidity separation as the primary drivers of classifier discrimination.
- Automated ingestion and preprocessing of simulated datasets exceeding 100,000 events with Python and Bash scripts that cleaned, concatenated, and flattened nested ROOT records into tabular form, submitting jobs over SSH and SCP on the SCIPP Tier 3 High Throughput Computing cluster.

Projects

chess-transformer | *PyTorch, Gradio, Hugging Face, Stockfish*

- Ran a 2x2 ablation of model capacity against data scale across four trained checkpoints, measuring legal-move and fully-legal-game rates over 500 generated games each: neither larger capacity nor a 1.2 million game corpus moved results alone, while the combination reached 99.1 percent legal moves and 51.8 percent fully-legal games, roughly 12 times baseline.
- Caught and fixed three evaluation-infrastructure defects that had corrupted reported results, including mid-epoch loss averaging and castling false-positives in the legality checker, and confirmed learned board state with linear probes on the residual stream measured against a random-initialized control.

resume-tailor | *Python, Streamlit, Ollama, LaTeX*

- Diagnosed a repeatable data-attribution defect in a LangChain and ChromaDB retrieval pipeline where recursive chunking severed source paragraphs from their role headers, tested two retrieval-side fixes, and eliminated the failure entirely by replacing retrieval with full-context injection once the document was shown to fit the context window.
- Shipped a Streamlit application that generates a compile-ready tailored resume plus a per-requirement keyword coverage report, tightening prompt constraints from observed failure modes and removing a numeric match score that proved non-reproducible across runs.

Additional Employment

Managed Career Solutions (contracted to Los Angeles Public Library) 2024 – Present

Technical Operations Specialist

Los Angeles, CA

- Functioned as the operational intermediary between patrons, branch staff, and Central IT, triaging and escalating cross-platform issues across Mac, PC, iOS, Android, and ChromeOS and accelerating resolution of systemic failures individual departments could not resolve.
- Delivered bilingual English and Spanish remote support by phone and webchat across all library branches, managing a distributed fleet of Chromebooks and cellular hotspots and serving as subject matter expert and trainer for third-party digital services.