Óscar A. Chávez Ortiz (Website: chavezoscar009.github.io)

chavezoscar009@gmail.com — (951) 801-9296 — LinkedIn: chavezoscar009 — GitHub: astroscar09, chavezoscar009

Technical Skills

Languages:	Python (**Expert**), SQL (**Intermediate**), Git/GitHub (**Intermediate**), R (**Basic**),
ML Frameworks:	scikit-learn, TensorFlow, Keras, PyTorch, XGBoost, sbi
ML Techniques:	Regression, Classification, Neural Networks (NN), Convolutional Neural Networks (CNN), Random Forest, XGBoost, Simulation-Based Inference (SBI)
Bayesian/Statistical Methods:	Bayesian Analysis, MCMC, Uncertainty Quantification, Model Evaluation, Hypothesis Testing
Visualization:	Matplotlib (**Expert**), Seaborn (**Expert**), Tableau (**Basic**)
Data Tools:	Pandas, NumPy, SciPy, Astropy, h5py, Photutils, emcee
Communication:	Public Presentations, Scientific Writing, Media Commentary, Outreach Talks (Astronomy on Tap, undergraduate organizations)

Work Experience

Graduate Student Researcher August 2020 – Present
University of Texas at Austin, Austin, TX

Scalable ML Pipeline for Feature Engineering and Model Optimization
Goal: Develop scalable ML/AI pipelines to predict key properties from a large dataset with high accuracy and robust uncertainty estimates.

1. Designed and implemented a scalable data workflow to process a large observational dataset ($\geq$ 1,000,000), engineer predictive features using physically motivated quantities, and used simulation based inference (SBI) to infer key system properties. This included automated data cleaning and sub-selection, parameter estimation, model training, and evaluation.
2. Conducted hyperparameter optimization and uncertainty calibration to deliver a high-performing model (90% accuracy) capable of reproducing key system properties with robust error estimates.
3. Developed a custom framework for others to use the trained model in a user friendly way via a pip-installable package. Ensuring ease of use and proper documentation for those not familiar with ML/AI techniques.

Bayesian Model Selection and Parameter Estimation
Goal: Evaluate which combination of physical processes best explains the observed features of a high-redshift galaxy by comparing a suite of candidate models and quantifying the evidence for each.

1. Applied Bayesian model comparison to select the most likely model from a suite of models, using the Bayesian evidence to guide inference and comparison.
2. Executed large-scale parameter estimation to infer key system properties, including black hole mass, star formation rate, and other astrophysical features, with uncertainty quantification via Bayesian modeling.
3. Developed reproducible workflows for model evaluation, enabling systematic benchmarking and comparison across a complex parameter grid.
4. Interpreted results to inform theoretical scenarios for early galaxy and black hole formation, connecting model outputs to physically meaningful conclusions.

Bayesian Inference & Predictive Modeling Pipeline
Goal: Develop a scalable Bayesian inference pipeline to extract physical properties from complex multi-dimensional datasets, quantify uncertainties, and identify key patterns in observational data.

1. Designed and implemented a Bayesian inference pipeline to analyze large-scale observational datasets, integrating multiple data sources into a single algorithm to uncover key drivers of signal detectability.
2. Built custom statistical modeling tools with *emcee* and performed feature engineering to identify correlations between inferred parameters and detection outcomes.
3. Automated data ingestion, cross-matching, and analysis workflows, enabling reproducible and scalable research published in a peer-reviewed journal (*Astrophysical Journal*).

Pilot Study: Feature-Observable Correlation Analysis
Goal: Assess whether meaningful correlations between features and target observables could be identified from a minimal dataset, establishing a proof-of-concept for scalable ML pipelines.

1. Developed an end-to-end data analysis pipeline, including data ingestion, cleaning, and feature selection, to identify high-quality samples for further modeling.
2. Integrated multiple datasets and applied statistical techniques (Spearman correlation, bootstrapping) to quantify relationships and associated uncertainties.
3. Implemented custom Python code for feature extraction, data modeling, and validation, uncovering initial correlations that informed the design of a larger-scale ML pipeline for predictive modeling and feature engineering.

End-to-End Pipeline Performance Evaluations
Goal: Assess and correct for biases in our data processing pipeline to ensure accurate and complete measurements from observational datasets.

1. Developed an automated synthetic data generation and framework to measure pipeline completeness and detect failure modes.
2. Diagnosed and resolved pipeline issues through systematic testing, improving end-to-end data processing reliability.
3. Applied completeness corrections to imaging and spectral features, ensuring that the final results were physically meaningful and unbiased for downstream modeling.

Personal Projects

Auditory Health Care Analysis

1. Developed a deep learning pipeline to classify raw .wav audio files as indicative of pain or not, converting audio to spectrograms for model input.
2. Built a custom neural network from scratch, then leveraged a pre-trained Hugging Face model to improve classification performance using transfer learning.
3. Evaluated model generalization with randomized and patient-level splits, identifying a significant performance drop (75% $\rightarrow$ 60%) when testing on unseen patients, highlighting the importance of robust evaluation strategies.
4. Conducted end-to-end workflow including data preprocessing, model training, hyperparameter tuning, and performance analysis to inform future medical ML applications.

Bayesian Model Selection and Parameter Estimation

1. Led a team of four to compare multiple ML models (Neural Networks, XGBoost, Linear Regression, Random Forest) on a shared dataset, coordinating methodology and ensuring consistent evaluation metrics.
2. Consolidated results from diverse modeling approaches, designing a unified evaluation framework to benchmark model performance across techniques.
3. Mentored team members in statistical modeling, ML evaluation, and reproducible coding practices, fostering collaboration and accelerating project progress.

Education

Ph.D in Astronomy — Spring/Summer 2026
University of Texas at Austin, Austin, TX

Bachelors of Art in Astrophysics — December 2019
University of California Berkeley, Berkeley, CA