Joseph Valencia

2900 NW Polk Ave, Apt 102, Corvallis, OR 97330

+1 (303) 957-6621 | valejose@oregonstate.edu | josephvalencia.github.io | josephvalencia | josephdvalencia

Education

Oregon State University (*Corvallis, OR*) Sep 2019 - Jun 2024 (Expected)

PhD Computer Science, emphasis in Artificial Intelligence

- EECS Outstanding Scholar
- Relevant Courses: Machine Learning, Deep Learning, Natural Language Processing, Convex Optimization, Matrix Analysis, Advanced Econometrics, Causal Inference

The University of Tulsa (*Tulsa, OK*) Aug 2015 - May 2019

B.S. in Computer Science, minor Economics

- Presidential Scholar
- Spring 2017 study abroad, Universidad Carlos III de Madrid

Experience

Graduate Research Assistant - **Oregon State University** (*Corvallis, OR*) Sep 2019 - Present

- Ongoing computational biology research at Hendrix Lab studying translational regulation of gene expression. Developing interpretable deep learning classifier to distinguish protein-coding messenger RNA and long noncoding RNA using Transformer architecture.
- Creating novel model interrogation strategies to identify biologically relevant sequence patterns that contribute to protein-coding potential.
- Teaching Assistant : planned curriculum and delivered hands-on studio sessions for new first-year engineering orientation course (Fall '21). Managing thirteen diverse student groups for senior software capstone (Winter '22).

Research Data Scientist Intern - **ACT, Inc.** (*Lakewood, CO*) May 2019 - Aug 2019

- Conceived of and prototyped MTLHealth, a multi-task machine learning system for detecting disturbing content in student essay responses to standardized tests.
- Integrated training data from internet toxicity datasets, trauma-related subreddits, and UK Biobank mental health studies, and adapted to small internal dataset via hybrid deep learning and logistic regression model.

Undergraduate Research Assistant - **The University of Tulsa** (*Tulsa, OK*) Sep 2017 - May 2019

- Built software tool LtrDetector for DNA LTR-retrotransposon sequences in genomes. Algorithm employs fast k-mer hashing, merging heuristics, and sequence alignment to locate repeat regions.

Summer Technology Analyst - **Deutsche Bank** (*Cary, NC*) Jun 2017 - Jul 2017

- Completed data ingestion pipeline for fraud prevention team and led backend programming of financial adviser chatbot prototype for intern project.

Manuscripts

- **Valencia, Joseph**, and Erin Yao, "MTLHealth: A Deep Learning System for Detecting Disturbing Content in Student Essays.", 2021. Abstract accepted at APA convention 2020, cancelled due to COVID-19. arXiv:2103.0429
- **Valencia, J.D.**, Girgis, H.Z, "LtrDetector: A tool-suite for detecting long terminal repeat retrotransposons de-novo", 2019. BMC Genomics 20, 450

Recent Projects

Satisfiability Approaches to Predicting RNA Pseudoknots Github Link

- Class project implementing minimum free energy methods for predicting RNA secondary structure with pseudoknots using pySMT SAT solver library.

Computational Biology Camp Github Link

- Remote instruction of middle school science camp on computational biology for Hendrix Lab outreach. Designed and delivered interactive coding tutorial teaching students to study biological sequences using Python.

Fair Principal Component Analysis for Correcting Population Stratification Github Link

- Class project testing suitability of fairness aware PCA variants for correcting population stratification in GWAS studies.

Technical Skills

Languages	Python, C++, R
Data Science	PyTorch, NumPy, Pandas, scikit-learn
Bioinformatics	BioPython, EMBOSS, BLAST, bedtools
Operating Systems	Linux, MacOS