Sofia Lima

slima2@alumni.cmu.edu | +1 2566527859 +34 645048690 | www.linkedin.com/in/sofia-lima
https://github.com/sofiathelima | www.fialogia.com

EDUCATION

Carnegie Mellon University Aug 2021 –May 2023
Master of Science in Computational Biology, Professional Honors, QPA: 3.62 Pittsburgh, PA
Relevant Coursework: Machine Learning, Bioinformatics, Data Integration, Bioimage Informatics, Quantitative Genetics, Computational Molecular Biology and Genomics, Biological Modeling and Simulation, Automation of Scientific Research

Vanderbilt University Aug 2015 – May 2019
Bachelor of Arts in Molecular & Cellular Biology, Minor: Neuroscience, Little John Scholar, GPA: 3.608 Nashville, TN

SKILLS

Laboratory: immunohistochemistry, confocal microscopy, PCR, western blot, electrophysiology, spectrophotometry
Programming Languages: Go, C++, JavaScript, Python, R, MATLAB, SQL, bash
Bioinformatics: bcl2fastq, bedtools, samtools, cutadapt, blast, sratoolkit, spades, bowtie2, cellranger, STAR, bwa, Seurat, biopython, Mothur, Kraken2, Alphafold, NCBI Databases, Protein Data Bank, MEME Suite, UCSC Genome Browser, IGV
Libraries & Software: cuml, sklearn, pytorch, keras, pyg open3d, trimesh, polars, pandas, matplotlib, seaborn, ollama, openwebui, PyQt, Git, Jupyter, conda, npm, ParaView, ImageJ, Audacity, LaTex, openssh, slurm, GNU parallel, sysstat, nvidia-smi, singularity, docker, wrangler, WebAssembly
Web Services: Amazon Web Services (AWS) EC2/S3/Fargate/Bedrock, CloudFlare Workers/R2/Secrets

EXPERIENCE

Research Assistant August 2025 – Present
Institute for Research in Biomedicine, López-Bigas and González-Pérez Lab Barcelona, Spain
- Migrated AWS services to on-premise high-performance computing (HPC) systems to set up a local development environment for a retrieval-augmented generation (RAG)-based chat application
- Refactored database processing to improve modularity and traceability
- Explored relational schemas to optimize storage efficiency and reduced database file size by 2X
- Wrote 100+ unit test for data fetching, processing, and exporting

Research Computing Facilitator May 2024 – May 2025
Virginia Tech, Advanced Research Computing, Division of Information Technology Blacksburg, VA
- Provided technical support for 50+ HelpDesk tickets and hosted office hours for guided troubleshooting
- Instructed workshops for connecting to the HPC resources shared by 1200+ users with a wide range of research objectives
- Installed 20+ software applications on various operating systems and different hardware microarchitectures
- Monitored central and graphics processing unit (CPU and GPU) device utilization to diagnose performance bottlenecks
- Documented best practices for parallelizing independent workloads and transferring data across storage systems
- Tested distributed training setups using dask-cuda from RAPIDS on a DGX system using the MNIST benchmarking dataset

Senior Engineer May 2022 – April 2024
Thornton Tomasetti, Applied Sciences New York, NY
- Incorporated 10+ new features and front-end functionalities for structural damage analysis software
- Generated virtual patient anatomies using a Variational Graph Autoencoder (VAE)
- Collaborated with stakeholders and internal technical teams to meet 100% of deliverable deadlines
- Presented designs with technical teams and business partners at over 20 weekly standups

Teaching Assistant January 2023 – May 2023
Carnegie Mellon University, Computational Biology Department Pittsburgh, PA
- Lead recitation for undergraduate Introduction to Computational Biology course
- Held weekly office hours for reviewing topics

Laboratory Technician October 2019 – Jul 2021
BioStrategies LC, Arkansas Biosciences Institute (ABI) Jonesboro, AR
- Designed histology experiments for studying drug delivery in 4 animal tissue types
- Presented findings in 5+ lab meetings

Research Assistant May 2016 – May 2019
Vanderbilt University, Department of Biological Sciences Nashville, TN

- Collected and analyzed immunohistochemistry and electrophysiology data to characterize newly discovered proteins in the Wnt signaling pathway during synaptogenesis at the *Drosophila* larval neuromuscular junction
- Collaborated with on campus core facilities and interdisciplinary researchers for quantifying biomarker signals *in situ*

Teaching Assistant August 2018 – May 2019
Vanderbilt University, Department of Biological Sciences Nashville, TN

- Instructed a biology laboratory course designed for non-majors
- Guided experiments

PUBLICATIONS

Lima S, *et al*. (2023). Aortas, Aneurysms, & AI report on Thornton Tomasetti company website.
Lima S, *et al*. (2022). AI Engine in Silico Trials: Aortic Graft Stent Demonstrator report on Thornton Tomasetti company website.
Acknowledged in Acosta W, Cramer CL. (2020). Targeting Macromolecules to CNS and Other Hard-to-Treat Organs using Lectin-Mediated Delivery. *Int J Mol Sci*, 21(3):971. doi: 10.3390/ijms21030971. PMID: 32024082.
Kopke D, Leahy S, Vita D, Lima S, Newman Z, Broadie K. (2020). Carrier of Wingless (Cow) Regulation of Drosophila Neuromuscular Junction Development. *eNeuro*, 7(2):ENEURO.0285-19.2020. doi: 10.1523/ENEURO.0285-19.2020.
Kopke D, Lima S, Alexandre C, Broadie K. (2017). Notum coordinates synapse development via extracellular regulation of Wingless trans-synaptic signaling. *Development*, 144:3499-3510. doi: 10.1242/dev.148130.

PROJECTS

Constructing an ML Pipeline for Transcription Start Site (TSS) Prediction Spring 2023

- Executed job scripts with high-performance computing (HPC) services provided by Pittsburgh Supercomputing Center
- Developed a deep learning pipeline for high quality genome annotation in human and mice
- Integrated publicly available experimental CAGE-Seq peak data and reference genomes
- Transformed raw DNA sequences with 4 Genomic Signal Processing (GSP) techniques
- Documented raw data acquisition, processing, model training and evaluation with coherent and concise flowcharts

Active Learning for 3D Mesh Segmentation Spring 2023

- Processed the MPI FAUST dataset of +100 3D scans as surface meshes with 12 body segments
- Compared sampling strategies in an online learning setting
- Accomplished semantic segmentation with a Graph Neural Network (GNN)
- Parallelized training with AWS EC2 instances and found a 10X speedup in runtime with 4 GPUs compared to a single device

Aortas, Aneurysms & AI Spring 2023

- Trained graph neural network (GNN) models for graph-level classification with multi-dimensional mesh data
- Analyzed performance of various model architectures with spline convolutions and pooling

AI Engine In Silico Trials: Aortic Graft Stent Demonstrator Summer 2022

- Processed finite element analysis (FEA) simulation data representing an *in Silico* aortic device implantation surgery
- Developed statistical machine learning pipelines to predict outcomes and identify key features of surgical success

Hemaglobin subunit alpha (HbA) *ab initio* Protein Structure Prediction with HP Model Spring 2022

- Implemented simulated annealing algorithm for finding the optimal scoring structure of a small protein
- Defined a scoring function based on Hydrophobic-Polar (HP) interactions
- Processed protein structure data of HbA with 141 amino acids in .pdb file format from the Protein Data Bank (PDB)
- Utilized the ProDy python package and the Visual Molecular Dynamics (VMD) program for rendering and computing distances between prediction results and experimental X-ray crystallography results from PDB
- Compared results of 100 simulations to a random baseline
- Discussed performance advantages of homology modeling with tools like AlphaFold

Discriminative Modeling of Lung Cancer Based on Gene Expression Spring 2022

- Applied various statistical machine learning methods to analyze gene expression data from The Cancer Genome Atlas (TCGA) to classify cancer subtypes lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC)

Barnes-Hut Galaxy Simulation Fall 2021

- Built a simulation of n-star galaxies using the Barnes-Hut algorithm to compute forces of gravity over time

The effects of urbanization on birdsongs Spring 2019

- Queried the Xeno-Canto database of birdsong audio recordings and analyzed frequency and rhythm using Audacity and R to make conclusions about vocal plasticity in birds in response to urbanization

CONFERENCES

Attendee July 2024
Practice and Experience in Research Computing (PEARC) Providence, RI
Attendee July 2023
3rd Nobel Turing Challenge Initiative Workshop Pittsburgh, PA

VOLUNTEERING

STEM Workshop: Introduction to Arduino Coding January 2025
Facilitator, LetsCodeBlacksburg Blacksburg, VA

- Assisted a 4-hour hands-on textual coding (C++) session for embedded systems in a microcontroller
- Facilitated safe wiring setup of simple flash-LED circuits to allow the kids age 10+ to perform the demonstration
- Explained the Blinking Chase Light example with the classic Arduino IDE as outlined in the written protocol

Elementary School Field Trip November 2019
Chaperone, ABI Jonesboro, AR

- Aided students in a laboratory procedure to extract DNA from strawberries