

# Emily Costa

## Software Engineer

 [emilyjcosta5.github.io](https://github.com/emilyjcosta5)

 [/emilyjcosta](https://www.linkedin.com/company/emilyjcosta)

 [/emilyjcosta5](https://github.com/emilyjcosta5)

## Work Experience

### Graduate Student Researcher

#### Northeastern University - Boston, MA

August 2020 - Present

- Apply machine learning and statistical inference to analyze I/O patterns in HPC workloads to make trend predictions and improve job scheduling based on probing data of large-scale clustered file systems of various configurations
- Design solutions for resolving I/O issues affecting production high-performance computing systems and data-intensive computational workloads
- Develop an open-source Python packages that enables system users and admins to improve software performance through congestion identification

### Machine Learning Software Developer Intern

#### Oak Ridge Nat'l Lab - Oak Ridge, TN

June 2021 - August 2021

- Trained large-scale deep learning models implemented in a Python library which use BERT from the Huggingface Transformers API to perform computational phenotyping and enable automate cancer report classification
- Built custom software stack with ROCm/HIP that ports software to new GPU-based supercomputer architectures, such as Frontier and Aurorora
- Profiled and analyzed model pretraining performance to enable usage of Frontier, the first exascale supercomputing system

### C/C++ Software Developer Intern

#### Florida International University - Miami, FL

June 2019 - August 2019

- Optimized and revamped file compression for improving I/O performance and reliability in scalable mass spectrometry proteomic software in order to
- Implemented novel method using HDF5 to improve computational workflow

### Python Software Developer Intern

#### Oak Ridge Nat'l Lab - Oak Ridge, TN

June 2019 - August 2019

- Developed a computational framework for scaling a Bayesian inference algorithm that integrated into open-source scientific Python packages
- Automated software tests with support continuous integration to increase code reliability and quality of the complex packages

## Software & Frameworks



**Best** BASH, Dask, Git, Java, Linux, Numpy, Pandas, Python



**Better** C/C++, Deepspeed, Pytorch, R, Scikit-Learn



**Good** Docker, HuggingFace, Julia, noSQL, SQL, SQLite, XML

## Education

### Northeastern University

August 2020 - May 2022

Computer Engineering, M.S.

#### · Relevant Coursework:

Computer Systems, Database Management, Data Structures, and Machine Learning

#### · Funding & Distinctions:

GEM Full Fellowship, and Research Assistantship

### Florida International University

August 2017 - July 2020

Mathematics, B.S.

## Projects

### Command my Stocks


January 2020

 [/cmdmystocks](https://github.com/emilyjcosta5/cmdmystocks)

- Created an open-source Linux command line user interface to easily customize trading bots on Google Cloud Platform
- Launched an algorithm to trade based on the magnitude of stock price changes

### Smoky Mountain Challenge


July 2019

 [/datachallenge2](https://github.com/emilyjcosta5/datachallenge2)

- Implemented a convolutional neural network, ResNet-50, that classifies diffraction patterns to reduce the workload of compute and data intensive software
- Balanced and managed >600 GB using dynamic processing techniques and SMOTE to generate psuedo-images

### OmiCloud

March 2019

 [/OmiCloud](https://github.com/emilyjcosta5/OmiCloud)

- Developed and implemented algorithms that rapidly detect surfaces and moving objects using 3-D points collected by a Microsoft Kinect to identify a fallen person and alert emergency medical services