# ALEC MACLEAN GUNNY

@ alec.gunny@gmail.com | San Francsico, CA | github.com/alecgunny | alecgunny.github.io

## EXPERIENCE

### Senior Solution Architect

**NVIDIA** April 2019 – Present

- Developed APIs around TensorFlow to accelerate inference-time execution of deep recommendation models by a factor of 20-40x on GPUs
- Drove recommendation project from concept to implementation, managing resources across product, engineering, and sales organizations
- Built collateral for publication, including demo walk-through and blog post

### Solution Architect

**NVIDIA** September 2017 – April 2019

- Collaborated with data science and infrastructure teams from large consumer internet companies to build scalable, GPU-accelerated deep learning systems for both training and inference
- Synthesized solutions to common industry problems into presentations, demos, and conference talks aimed at a variety of audiences

### Data Scientist

**Children's Hospital Los Angeles** April 2016 – September 2017

- Modelled asynchronously and irregularly sampled EMR data from patient stays in the pediatric ICU
- Researched parameterized structured inference models as a means to forecast patient vitals under hypothetical courses of treatment
- Built visualization tools to better interpret models and communicate results to physicians

### Scientist

**Arete Associates** August 2014 – April 2016

- Leveraged traditional signal and image processing techniques as well as machine and deep learning to solve problems ranging from vehicle classification and cancer cell detection to object tracking and system performance modelling

## CONFERENCE TALKS

### A Trip Through The NGC Tensorflow Container

**GPU Technology Conference** March 2019

- Presented an end-to-end workflow for training a deep learning model on audio data then serving it for inference at reduced precision

### Deploying Deep Neural Networks as a Service Using TensorRT and NVIDIA-Docker

**GPU Technology Conference** March 2018

- Demo on converting trained recurrent and convolutional models from multiple deep learning libraries into accelerated inference runtime engines and exposing them via an inference service

### Quick and Easy DL Workflow Proof of Concept

**GPU Technology Conference** March 2018

- Demo workflow for building embedded applications using deep learning models trained in the cloud

## UNDERGRADUATE COURSEWORK

Quantum Mechanics | Electrodynamics | Real & Complex Analysis | Dynamics | Fluid Dynamics | Materials Science | Circuits | MATLAB | Optics

## SKILLS

Python ●●●●●
Docker ●●●●●
Bash ●●●●○
C++ ●●●○○
CUDA ●●○○○

- Python Libraries

Tensorflow | Pytorch | Scikit-Learn | Pandas | Bokeh | OpenCV | Flask | TensorRT | Horovod

- Cloud Computing

AWS | Google Cloud

- Scheduling/Orchestration

SLURM | Kubernetes

## EDUCATION

### B.Sc. in Engineering Physics

**University of California Berkeley**

August 2010 – May 2014

Graduate with honors

## AWARDS AND HONORS

- Arete Spot Award Recipient
- Top 10% finisher in Kaggle Diabetic Retinopathy Detection Competition