ALEC MACLEAN GUNNY

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

EXPERIENCE

Senior Solutions Architect

NVIDIA   April 2019 – Present

- Developing APIs to accelerate inference-time execution of deep Tensorflow recommendation models by a factor of 20x on GPUs
- Collaborating and directing resources across product and engineering orgs to define and execute platform vision
- Building illustrative demos and collateral to communicate API structure, usage, and limitations

Solutions Architect

NVIDIA   September 2017 – April 2019

- Working with major consumer internet accounts to understand challenges in implementing deep learning models and infrastructure
- Collaborating with data science and infrastructure teams to build out scalable, accelerated deep learning solutions
- Synthesizing common industry issues and best practices into presentations, demos, and conference talks

Data Scientist

Children's Hospital Los Angeles   April 2016 – September 2017

- Modelling asynchronously and irregularly sampled EMR data from patient stays in the pediatric ICU to improve treatment outcomes
- Researching the use of parameterized variational inference models to forecast patient vitals under hypothetical courses of treatment
- Building visualization tools to better understand models and communicate results to physicians

Scientist

Arete Associates   August 2014 – April 2016

- Leveraging 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.E. in Engineering Physics

University of California Berkeley
 August 2010 – May 2014
Graduate with honors

LANGUAGES

Spanish
Danish

AWARDS AND HONORS

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