

# Andrew Z. Luo

MACHINE LEARNING · SYSTEMS · SOFTWARE ENGINEERING

☎ 425-241-9772 | ✉ andrew.zhao.luo@gmail.com | 🌐 andrew-zhao-luo

## Experience

### Machine Learning Systems Engineer

Seattle, WA

OCTOML, APPLIED COMPILER ENGINEERING

Mar. 2021 - Now

- Committer to TVM, an open source ML compiler. Contribute to quantization, mixed precision, and ML framework support.
- Applying TVM in SaaS product. Led an average 20% speedup in inference times in internal model zoo.

### Machine Learning Engineer

Seattle, WA

APPLE, AI/ML MACHINE INTELLIGENCE NEURAL DESIGN

Jan. 2020 - Mar. 2021

- Used quantization, sparsity, and hardware-specific knowledge to train models for Siri, Homepod, and future products
- Developing in-house solutions for training vision models and deploying/benchmarking on FPGA and ASIC environments
- Languages: *Python*. Technologies: *PyTorch, Tensorflow, CoreML, Apple Neural Engine*

### Machine Learning Engineer

Seattle, WA

XNOR.AI, MACHINE LEARNING TEAM

Aug. 2019 - Jan 2020

- Training performant computer vision models that can run on bespoke and edge hardware. Part time until Jan 2019.
- Created face identification demo showcasing XNOR's technologies to key executives at major tech companies
- Languages: *Python, C, C++*. Technologies: *PyTorch, Bazel*. textbfAcquired by Apple Jan. 2020

### Engineering Intern

San Francisco, CA

SIFT SCIENCE, CORE DATA

Jun. 2018 - Sep. 2018

- Rewrote HBase snapshot system, saving over \$1.5 million in S3 costs a year. Added BigQuery integration with HBase.

### Software Engineering Intern

Seattle, WA

FACEBOOK, ADS CORE

Jun. 2017 - Sep. 2017

- Implemented back-end statistical models to predict demographics of ad reach for customers with multi-million yearly spend

### Undergraduate Teaching Assistant

Seattle, WA

CSE312 (PROBABILITY FOR CS) AND CSE446 (INTRODUCTION TO ML), UNIVERSITY OF WASHINGTON CSE

Sep. 2018 - Jun. 2019

- Gave weekly lectures to 20-30 students, hold weekly office hours, graded homework, created answer keys and new material

## Other Projects

### The FPGA Image Convolution Photobooth

- Created algorithm to run kernel convolutions on streamed images, implemented in FPGA on Altera Cyclone V
- Integrated with camera and VGA, creating a variety of filters like Sobel edge detector, Gaussian blur, and image sharpening

## Honors

- |      |  |
|------|--|
| 2018 | <b>Honorable Mention</b> , Goldwater Scholarship     |
| 2017 | <b>Scholarship</b> , Emerging Leaders in Engineering |
| 2017 | <b>Scholarship</b> , Undergraduate Conference Award  |
| 2015 | <b>Scholarship</b> , Mary Gates Research Scholarship |

## Skills

- |                     |                          |
|---------------------|--------------------------|
| <b>Languages</b>    | Python, Java, C, C++     |
| <b>Frameworks</b>   | Sklearn, PyTorch, Django |
| <b>Technologies</b> | HBase, Linux, Airflow    |
| <b>Tools</b>        | git, bash, Bazel, ㄖㄚㄚㄚ   |

## Education

### University of Washington

Seattle, WA

DOUBLE MAJOR IN COMPUTER ENGINEERING AND BIOENGINEERING

Sep. 2015 - Jun. 2019

- **Coursework:** Machine Learning, Probability and Statistics, Real Analysis, Operating Systems, Compilers, Embedded Systems
- **GPA:** 3.95, *Summa Cum Laude*

## Publications and Patents

*Automatic Characterization of User Errors in Spirometry.* **Andrew Luo**, Eric Whitmire, James Stout, Drew Martenson, Shwetak Patel. *IEEE EMBC 2017 (Oral Presentation + Paper)*

*Compressed Neural Network Models.* US Patent App. 16/788261. James Gabriel et al. and **Andrew Luo**. Filed 13 August 2020.