

# Max Hoffman

maximilian.wolfgang1@gmail.com | 414.491.4294 | <https://www.maxh.ai>

## PROFESSIONAL SUMMARY

---

Academic researcher trained in statistics, data science, and the architecting and tuning of machine learning models. Transitioned from medical school to software engineering with industry experience in productionizing distributed data systems and pipelines at scale, building full-stack web applications, and automating engineering operations with cloud and CI tooling.

Niche background in using cross-disciplinary skills to model data systems at individual-service and company-wide levels to build systems that let individuals do the work of thousands, and support company-wide data efficiency that bridges the work of segmented data science/engineering teams. Contributed to early-stage startup diligence and investment.

### Areas of Expertise:

- Model design, training, and implementation with Tensorflow, Pytorch, and SparkML.
- Data processing and integration with Python and Spark Scala.
- Data workflow automation with Airflow, Luigi, and Snakemake.
- Custom Python, Java, and Scala libraries for data infrastructure tooling.
- Infrastructure automation and CI/CD with AWS, Terraform, Docker/Singularity, and Jenkins.
- Monitoring and tracking with Prometheus and Grafana.
- Full-stack web applications with React and Vue.

### Skills:

Python2/3 | Scala | Java | Bash | Spark | Vim | Jupyter | Pandas | Numpy | Chartify | SparkML  
JTS | Tensorflow | Pytorch | XGBoost | SKLearn | Sonnet | SWIG | ONNX | Springboot | Maven | Sbt |  
KubeFlow | Sagemaker | MLLeap | TFServing | Hadoop | HDFS | Hive | Beam | Docker | Singularity | PBS |  
Slurm | Jenkins | AWS | Terraform | Postgres | MySQL | Aurora | Dynamo | Mongo | JS | ES6 | React | Redux

## EXPERIENCE

---

### Factual Inc. – Software Engineer

February 2019 – April 2020

- Our team's training data was small and hand-collected, leading to biased and poorly generalizing models. I wrote a processing system and CI workflow that asof-joined consumer transactions with our location and place datasets to generate artificial training rows. This boosted the size of our training data 50-fold and smoothed model artifact and bias.
- External engineering services lacked interfaces for inter-team use, so I designed and wrote a Python/Scala library to exposed ML logic and database tooling. This let our team quickly test and integrate those services into our production products.
- Sales engineering and data insight teams regularly experienced footfall dashboard turbulence that disrupted customer experiences. I refactored the visit data processing workflow to unwind self-referential data loops. This removed a house-of-cards-like experience from the customer end and gave my team a fast way to end-to-end test how our changes affected footfall metrics.
- Product and sales teams regularly had to decide between data consistency (which customers liked) with accuracy (which customers do not see) without quantitative support from engineering. I designed a system of "attachment segment maps": 2D geographical shapes that compute per-POI attachment-zones for dozens of cities, and built complementary visualization dashboards, and metrics pipelines to measure attachment-zone consistency over time as place data, shape data, and place-attachment model versions changed. This system helped product and engineering teams communicate quantitatively regarding trade-offs between accuracy improvements and customer happiness.

- Product teams wanted immediate updates to footfall dashboards but our data infrastructure required a week window to process and deliver data. My team and I designed a forecasting model that estimated per-POI footfall visits for a week into the future based on historical and categorical information. This model filled the gap in our processing time, providing current estimates while true values were being generated.

### **Contrary Capital – Venture Partner**

January 2018 – May 2019

- Backed by Tesla, Reddit, and Facebook’s early founders, Contrary Capital is a student-led network of venture partners to provide resources and funding to like-minded student-founders across the US. I contributed technical and biomedical expertise towards screening and funding portfolio companies while in medical school. I continue to support the Contrary network.

### **Washington University Medical School – Machine Learning Engineer**

May 2018 – December 2018

- Re-implemented a series of historical models to organize quantum-mechanic property prediction literature benchmarks.
- Optimized a novel graph-recurrent architecture that outperformed state of the art accuracy by 10-fold in key metrics.
- Designed a parallel featurization framework for the lab’s training-lifecycle library.

## **RESEARCH**

---

### **Washington University**

#### **Swamidass Lab**

May 2018 – December 2018

- Predicted quantum chemical properties of small molecule drug targets with graph recurrent networks (manuscript submitted).

#### **Wessel Lab**

August 2017 – December 2017

- Re-implemented dynamics-parsing algorithm (SINDY) to classify visual spike train patterns (in mice) with fewer and more human-interpretable features (manuscript submitted).

### **Pomona College**

#### **Sazinsky Lab**

January 2015 – May 2016

- Discovered structural and chemical features of an *S. aureus* biofilm protein responsible for bacterial respiratory colonization and infections (wet lab).

#### **Shtylla Lab**

May 2015 – May 2015

- Modeled and visualized *C. elegans* mitotic spindle and motor mechanics through the asymmetric first division in Matlab.

#### **Fryer Lab**

Jan 2013 – Aug 2013

- Modeled the evolutionary dynamics of chemical reactions with Python and Matlab.

## **EDUCATION**

---

### **Washington University School of Medicine**

August 2017 – May 2018

### **Pomona College – B.A. Chemistry & Mathematics**

May 2016

Summa Cum Laude

Pomona Scholar 2012-2016

11.9/12 GPA