

# Hillary Sanders

hillosand@gmail.com  
503 481 0314

## Highlights:

- Seeking AI Safety research position
- 10 years of experience in machine learning research and engineering
- Lead researcher / engineer for multiple billion-sample+ deep learning products
- Author of deep learning book chapters & other publications

## EXPERIENCE

### Opendoor

Senior Research  
Scientist  
2021-2023

- Designed & conducted experiments on Opendoor's core model (house-price prediction neural network) resulting in over \$48M of total expected profit gains
- Led team of 6 researchers and engineers to re-engineer ML prod training and experimentation pipeline to improve usability, reproducibility, and uptime

### Sophos

Senior ML Researcher  
2019 - 2021  
Senior Software  
Engineer  
2018- 2019  
Data Scientist  
2017- 2018

- Developed catastrophic forgetting mitigation techniques to avoid retraining models from scratch, allowing for iterative fine-tuning over time with appx. 98% reduced forgetting effects, and ~10x reduced model training costs
- Developed novel model compression technique (variance-based Taylor series approach), resulting in appx. 30% less reduction in accuracy after 90% weight removal compared to the next-best method
- Owned research & maintenance for two deep learning multi-million parameter classification models, trained on billions of document samples
- Led engineering effort for our Endpoint Detection Response implementation (Docker, Elasticsearch, 20 separate models); also designed and maintained the production databases to support these (Elasticsearch)
- Helped design the team's main database (Redshift) with billion+ row tables; implemented a pipeline to create aggregation tables for ML training
- Crafted algorithms for developing better labels in our data

### RecipeStasher

Founder / Full-Stack  
2016 (Sep-Dec)

- Designed and built a recipe website. Key innovation over state-of-the-art was natural language recipe parser which could automatically convert unit and serving conversions across both ingredients and instructions

### Premise Data Corporation

Data Scientist  
2013 - 2016

- Developed data processing pipeline for our main product (inflation indices), as well as various other analytic pipelines (worker reliability, etc)
- Developed and implemented an optimal survey volume formulation to minimize expected variance in our inflation indices

## PUBLICATIONS

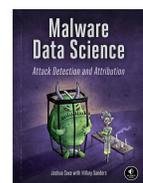
2021 **arXiv** - Magnificent Minified Models: approaches to model compression

2021 **arXiv** - Catastrophic Forgetting in the Context of Model Updates

2018 **No Starch Press (book)** - Authored the chapters of "Malware Data Science" that focus on deep learning theory and methods

2018 **IEEE** - A Deep Learning Approach to Fast, Format-Agnostic Detection of Malicious Web Content

2017 **BlackHat** - Garbage in, Garbage Out: How purportedly great ML models can be screwed up by bad data



## EDUCATION

### UC Berkeley

BS Statistics  
BS Env. Econ & Policy  
2009-2013

- Graduated Cum Laude
- Selected into competitive undergraduate research program twice:
  - Researched visualizing & tracking changes in legislative documents via NLP with graduate student Mark Huberty
  - Researched how liability affects international technology transfer

## SOFTWARE

- Daily user of **Python, ec2, Pytorch, PostgreSQL, Bash, AWS, Git**
- Frequent user of **Docker, Spark, Snowflake, Airflow**