

# MICAH J SMITH

micahsmith@gmail.com ✉  
+1 805 657-2325 ☎  
www.micahsmith.com 🌐  
Boston, MA & remote 📍  
he/him 🏳️‍🌈

## EDUCATION

Massachusetts Institute of Technology, Department of EECS Ph.D., Computer Science <i>Thesis: Collaborative, Open, and Automated Data Science</i>	2016–2021
Columbia University, Columbia College B.A., Economics-Mathematics, <i>cum laude</i>	2010–2014

## PROFESSIONAL EXPERIENCE

<b>Xyla (Cambridge, MA) — Infrastructure Lead/Machine Learning Engineer</b> • Leading ML infrastructure efforts for finetuning and inferencing large language models	2022-pres.
<b>Twitter (Boston, MA) — Machine Learning Engineer</b> • Developed machine learning platform frameworks and infrastructure on the ML Pipelines team (TFX, Kubeflow Pipelines, TensorFlow, Apache Beam, GCP BigQuery, GCP Dataflow)	2021–2022
<b>MIT LIDS, Data To AI Lab (Cambridge, MA) — Graduate Research Assistant</b> • Research ML systems, human-computer interaction, and databases, advised by Dr. Kalyan Veeramachaneni	2016–2021
<b>Botkeeper (Boston, MA) — Machine Learning Engineer (part-time)</b> • Lead company-wide ML engineering efforts in transaction classification for accounting • Designed and implemented cross-client transaction embedding, automatic model retraining, ML metrics collection/storage/querying/reporting (Python, Tensorflow/Keras, scikit-learn, MongoDB, Kubernetes)	2019–2021
<b>Twitter (New York, NY) — Machine Learning Engineering Intern</b> • Designed and implemented easy-to-use hyper parameter tuning via Bayesian optimization for production ML workflows such as pRecaptcha	Summer 2018
<b>Kensho Technologies (Cambridge, MA) — Machine Learning Intern</b> • Developed time series ML model to predict trading behaviors at Treasuries desk of major US dealer	Summer 2017

## SKILLS

General

Python

TypeScript

JavaScript

Java

Bash

Julia

C++

C

Scala

MATLAB

Haskell

DS/ML

pandas

numpy

scikit-learn

tensorflow

matplotlib

seaborn

keras

statsmodels

TFX

Kubeflow

Data

MongoDB

SQL

Airflow

SQLAlchemy

Spark

Dask

BigQuery

Beam

Dataflow

DevOps

Python packaging

Docker

Kubernetes

GCP

AWS

Heroku

GitHub Actions/Apps

Travis CI

Gitlab CI

Web

Flask

pelican

jinja2

node

jQuery

Google Apps Script

Django

HTML/CSS

React

Tornado

Tools

git

GitHub

GitLab

Jupyter Lab/Nb/Hub

vim

\*nix

LaTeX

make

sphinx

VS Code

PyCharm

Eclipse

Bazel

invoke

## ACTIVITIES

- Open-source developer: [BTB](#), [ATM](#), [AutoBazaar](#), [ballet](#), [Assemblé](#), [DSGE.jl](#), [FredData.jl](#), [repolockr](#)
- Organizer/Mentor, MIT EECS Graduate Application Assistance Program
- VP Communications/Social Chair, MIT EECS Graduate Student Association
- Bartender, MIT Muddy Charles Pub
- Running, biking, tennis, basketball, reading, coffee, crosswords, chess, playing with my dog Mamba

## SELECTED PUBLICATIONS

- “Enabling collaborative data science development with the Ballet framework.” CSCW 2021. [↗](#)
- “The ML Bazaar: Harnessing the ML Ecosystem for Effective System Development.” SIGMOD 2020. [↗](#)
- “FeatureHub: towards collaborative data science.” DSAA 2017. [↗](#)
- “Query optimization for dynamic imputation.” VLDB 2017. [↗](#)