MICAH J SMITH

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

EDUCATION

Massachusetts Institute of Technology, Department of EECS — 2016–2021
    Ph.D., Computer Science
    Thesis: Collaborative, Open, and Automated Data Science

Columbia University, Columbia College — 2010–2014
    B.A., Economics-Mathematics, *cum laude*

PROFESSIONAL EXPERIENCE

OpenEvidence (Cambridge, MA) — Infrastructure Lead/Machine Learning Engineer — 2022-pres.
- I own backend, site reliability, DevOps, infra, security, compliance, on-call (GCP, Django, Postgres, Python, GitHub Actions, Terraform, HIPAA) as the Technical Lead of the Core Engineering team
- Led architecture, design, backend/data implementation for complex integration of ClinicalKey AI product, delivered from prototype to GA in 6 mo., and launching into US hospitals
- Lead ML infra efforts for finetuning and inferencing large language models for healthcare
- I also do recruiting, hiring, people management, and project management for Eng team

Twitter (Boston, MA) — Machine Learning Engineer — 2021–2022
- Developed machine learning platform frameworks and infrastructure on the ML Pipelines team (TFX, Kubeflow Pipelines, TensorFlow, Beam, BigQuery/Dataflow)
- Supported training workloads for home timeline ranking and ads prediction models (100mm samples daily)
- Wisely bounced of my own free will before Elon took over

MIT LIDS, Data To AI Lab (Cambridge, MA) — Graduate Research Assistant — 2016–2021
- Researched ML systems, human-computer interaction, and databases, advised by Dr. Kalyan Veeramachaneni

Botkeeper (Boston, MA) — Machine Learning Engineer (part-time) — 2019–2021
- Led 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)

Federal Reserve Bank of New York (New York, NY) — Senior Research Analyst — 2014–2016
- Performed statistical and econometric analysis for projects including macro forecasting, monetary policy analysis, consumer expectations, and subprime mortgages (Julia, MATLAB, Stata, SQL, Python)

SKILLS

General	`Python` `TypeScript` `JavaScript` `Java` `Bash` `Julia` `C++` `C` `Scala` `MATLAB` `Haskell`
ML	`pandas` `numpy` `sklearn` `tensorflow` `OpenAI` `pytorch` `matplotlib` `seaborn` `keras` `statsmodels` `TFX` `Kubeflow`
Data	`Postgres` `MongoDB` `SQL` `Airflow` `SQLAlchemy` `Spark` `Dask` `BigQuery` `Beam` `Dataflow`
DevOps	`Python packaging` `Docker` `k8s` `GCP` `terraform` `AWS` `Heroku` `GitHub Actions/Apps` `Travis CI` `Gitlab CI`
Web	`Django` `DRF` `React` `Next.js` `MUI` `node` `jinja2` `HTML/CSS` `pelican`
Tools	`git` `GitHub` `GitLab` `Jupyter Lab/Nb/Hub` `*nix` `LaTeX` `make` `sphinx` `Bazel` `invoke`

ACTIVITIES

- Open-source development: BTB, ATM, AutoBazaar, ballet, Assemblé, DSGE.jl, FredData.jl, repolockr
- Running, biking, tennis, reading, coffee, crosswords, chess, playing with my dog Mamba

SELECTED PUBLICATIONS

- "Do we still need clinical language models?" CHIL 2023.
- "Enabling collaborative data science development with the Ballet framework." CSCW 2021.
- "The ML Bazaar: Harnessing the ML Ecosystem for Effective System Development." SIGMOD 2020.
- "Query optimization for dynamic imputation." VLDB 2017.