CHRISTOPHER K. TOKITA

christopher.tokita@gmail.com • ctokita.com • github.com/christokita

SUMMARY

I am a data scientist for good who aims to use his technical skills to create a positive impact on society. With expertise analyzing user behavior, social networks, and online ecosystems, I am a generalist data scientist that takes an interdisciplinary approach to a wide range of data science problems. I am also a full-stack data scientist that can research, develop, scale-up, and deploy a model to production.

EXPERIENCE

Phylum, Los Angeles, CA
Senior Data Scientist June **2022 – Present**
Data Scientist June **2021 –** June **2022**
Phylum is a cybersecurity startup whose platform detects risk and vulnerabilities in the open-source software ecosystem.

- Constructed an unsupervised ML model to flag anomalous behavior among software authors.
- Constructed and deployed an NLP + supervised ML model to detect online discussion of new security vulnerabilities far before they are officially documented by the US Government.
- Visualized large networks of authors + packages that comprise the open-source software ecosystem.

Princeton University, Princeton, NJ Sept **2016 –** May **2021**
Doctoral Researcher in Computational Ecology & Computational Social Science

- Used computational modeling and data science to analyze collective behavior, social networks, and information sharing in large social systems, including ant colonies and social media.
- Constructed agent-based simulation models in Python and R, including custom modules.
- Used Bayesian inference to quantify user characteristics (e.g., ideology) and treatments effects.
- Deployed digital experiments and compiled large dataset using social media APIs.

The Science and Technology Policy Institute, Washington, DC July **2014 –** June **2016**
Data Analyst

- Conducted data analysis (social network analysis, NLP, and bibliometrics) to inform decision-making in the White House Office of Science and Technology Policy and other federal agencies.

EDUCATION

Princeton University, Ph.D., Ecology & Evolutionary Biology **2016 – 2021**
Graduate Certificate, Computational Science & Engineering

Yale University, B.S., Ecology & Evolutionary Biology, *Distinction in the Major* **2010 – 2014**

TECHNICAL SKILLS

Programming Languages

Python, R, SQL (proficient), Scala (proficient), C (familiar), C++ (familiar)

Domain Knowledge

> Machine Learning: scikit-learn, PySpark.MLlib, regression, classification, clustering
> Statistics: Bayesian inference & regression (brms & pymc3), hypothesis testing
> Natural Language Processing: text similarity, topic modeling, text cleaning, regex, TF-IDF, hashing
> Deep Learning: PyTorch, fastai
> Cloud & Cluster Computing: PySpark, Spark, AWS (EMR, EC2 & S3), linux/unix/bash scripting
> Social Network Analysis & Network Science: igraph, networkx, network structure metrics
> Data Visualization: ggplot2, matplotlib, seaborn, Gephi, vector art programs
> Computational Social Science: agent-based modeling, contagions on networks, user behavior