

CHRISTOPHER K. TOKITA

ctokita.com • github.com/christokita

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**

RELEVANT EXPERIENCE

- Phylum**, Los Angeles, CA (Remote) **2022 – Pres**
Senior Data Scientist **2021 – 2022**
Data Scientist

Phylum is a cybersecurity platform that uses analytics and AI to detect risk in the open-source software ecosystem, which now powers over 90% of software.

- Constructed and deployed a stacked Bayesian ML model that can estimate a software author's geographic location from their commit behavior. This is the first-of-its-kind in the field of cybersecurity.
- Constructed an unsupervised ML model to flag anomalous behavior among software authors whose passwords may have been exposed during a data breach event.
- 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.
- Led research projects and attended team lead meetings as substitute for research manager when necessary.

- Princeton University**, Princeton, NJ **2016 – 2021**

NSF Graduate Research Fellow in Computational Ecology & Social Science

Used computational modeling and data science to analyze collective behavior, social networks, and information sharing in complex social systems, ranging from ant colonies to social media.

- Constructed agent-based simulation models in Python and R, including custom modules.
- Used Bayesian inference to quantify social media user characteristics (e.g., ideology) and treatments effects.
- Deployed digital experiments and compiled large datasets using social media APIs.

- IDA Science and Technology Policy Institute**, Washington, DC **2014 – 2016**

Data & Policy Analyst (Science Policy Fellow)

Established by Congress in 1990, The Science and Technology Policy provides policy analysis to the White House Office of Science and Technology Policy and federal agencies.

- Evaluated the outcomes of multi-million-dollar grant programs at the NSF and NIH using social network analysis, NLP (probabilistic topic modeling), and bibliometrics to.
- Created novel statistical metric to quantify underrepresented minority college graduation rates in STEM.
- Consistent interaction and communication with Federal Agencies, including interviews, presentations, and meetings.

RELEVANT PUBLICATIONS

Computational Social Science

Tokita CK, Aslett K, Godel WP, Sanderson Z, Tucker JA, Nagler J, Persily N, and Bonneau R. Measuring and mitigating belief in misinformation at the scale of the social media ecosystem. **In Prep**

Tokita CK, Guess AM*, and Tarnita CE*. Polarized information ecosystems can reorganize social networks via information cascades. *Proceedings of the National Academy of Sciences*. **2021**

Social Networks

Tokita CK, and Tarnita CE. Social influence and biased interactions can drive emergent behavioural specialization and modular social networks across systems. *Journal of the Royal Society Interface*. **2020**

TECHNICAL SKILLS & CERTIFICATIONS

Programming Languages

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

Security Clearance

Secret-level clearance, Department of Defense, 2014 – 2016

Data Science 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
- > **Cloud & Cluster Computing**: PySpark, Apache Spark, AWS (EMR & S3)
- > **Data Visualization**: ggplot2, matplotlib, seaborn, Gephi, vector art programs

Domain Knowledge

- > **Social Network Analysis & Network Science**: community detection, network structure metrics
- > **Computational Social Science**: agent-based modeling, contagions on networks, collective behavior
- > **Online communities**: social media, misinformation, user behavior
- > **Cybersecurity**: anomalous user behavior and packages in open-source software ecosystem