

# CHRISTOPHER K. TOKITA

[ctokita.com](http://ctokita.com) • [github.com/christokita](https://github.com/christokita)

## EDUCATION

---

- Princeton University**, Ph.D., Ecology + Evolutionary Biology **2016 – 2021**  
Graduate Certificate, Computational Science + Engineering  
Dissertation: “*Networks, Collective Behavior, and Information in Social Systems—From Ant Colonies to Social Media*”
- Yale University**, B.S., Ecology + Evolutionary Biology, *Distinction in the Major* **2010 – 2014**

## RELEVANT EXPERIENCE

---

- Sprout Social**, Los Angeles, CA (Remote) **2025 - Pres**  
*Senior Applied AI Scientist*  
Conducting research + development work to create AI tools that make social media more accessible, intelligent, and impactful for organizations, government, and high-profile individuals.
- > **AI Agent Evaluation**: Drove the development of evaluations for a product-wide AI agent, including further MCP tool development.
  - > **LLM-Based Message Summarization**: Built a statistical sampling pipeline in combination with an LLM model to summarize thousands of social media messages into clear, user-legible themes.
- Stripe**, Los Angeles, CA (Remote) **2023 – 2025**  
*Senior Data Scientist*  
Partnering with the Security Analytics + Detection team to help protect the internet’s economy.
- > **Web Attack Monitoring Dashboard**: Led first-ever, cross-functional (Security + Infrastructure) effort to measure attacks on Stripe’s platform. Built a company-wide dashboard covering 60% of the most dangerous web attack types, which now guides resource allocation and security investment priorities.
  - > **ML Pipeline for Labeling Vulnerabilities**: Built and productionized a model labeling software vulnerabilities with industry-standard CWE categories. Increased label coverage by 35% at 90% accuracy, significantly improving security leadership’s visibility into high-priority risks.
  - > **Dashboards + KPIs for Vulnerability Management**: Created actionable metrics and dashboard in collaboration with multiple engineering managers. Dashboard is now the source of truth in weekly meetings and quarterly operational reviews.
  - > **Insider Threat Detection Prototype**: Developed an alpha-release ML system (tree-based classifier) merging multiple behavioral signals to identify employee risk, reducing false positives for security investigations.
- Phylum**, Los Angeles, CA (Remote) **2021 – 2023**  
*Senior Data Scientist (Promoted from Data Scientist in 2022)*  
Phylum is a cybersecurity platform that uses analytics and AI to detect risk in open-source libraries used across the software industry.
- > **Bayesian Geolocation Detector**: Created first-of-its-kind stacked Bayesian model to estimate software authors’ geographic location from commit behavior, opening new threat intelligence avenues.

- > **Unsupervised Anomaly Detection:** Built an ML pipeline to flag suspicious behavior among authors whose credentials were exposed, proactively reducing risk from compromised accounts.
- > **Early Vulnerability Detection via NLP:** Developed an NLP + supervised ML system to identify online discussions of new security vulnerabilities far before official government advisories, reducing response time for threat mitigation.
- > **Leadership & Research Strategy:** Led research projects and attended team lead meetings as substitute for Director of Research when necessary.

**Princeton University**, Princeton, NJ

**2016 – 2021**

*NSF Graduate Research Fellow in Computational Ecology + Social Science*

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

- > **Digital experiments on social media:** Used Bayesian inference to A/B test data from 4,000 Twitter users and their 200,000+ followers to detect how news coverage changes people's social networks.
- > **Original simulation software:** Created custom agent-based models (in R/Python) simulating individual behavior + information sharing in complex social networks.

**IDA Science and Technology Policy Institute**, Washington, DC

**2014 – 2016**

*Data + Policy Analyst (Science Policy Fellow)*

Provided nonpartisan, data-driven analysis for the White House Office of Science and Technology Policy and federal agencies, applying NLP, social network analysis, and bibliometrics to evaluate multi-million-dollar science programs at NSF and NIH.

## TECHNICAL SKILLS + CERTIFICATIONS

---

- > **Programming Languages:** Python, R, SQL, Scala (familiar), C/C++ (familiar)
- > **Infrastructure + Platforms:** AWS, Databricks, Apache Spark (PySpark, Scala), Airflow, git/GitHub
- > **ML + Statistics:** scikit-learn, PySpark.MLlib, Bayesian Statistics, Natural Language Processing, LLMs + Generative AI (prompt engineering, evaluation, agent design, MCP), Data Visualization
- > **Domain Expertise:** Social Network Analysis, Computational Social Science, Social Media + Misinformation, Cybersecurity (anomaly detection, threat intelligence)
- > **Security Clearance:** Secret-level, Department of Defense (2014–2016)

## RELEVANT PUBLICATIONS

---

*Computational Social Science*

**Tokita CK**, Aslett K, Godel WP, Sanderson Z, Tucker JA, Nagler J, Persily N, and Bonneau R (2024). [Measuring receptivity to misinformation at scale on a social media platform](#). *PNAS Nexus*.

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

*Social Networks*

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