

Harold Triedman

email ■ github ■ linkedin ■ personal site ■ google scholar

EDUCATION

Cornell Tech

New York, NY ■ Sep 2024 – Present

- 3rd year pursuing a doctorate in Computer Science at Cornell Tech. I conduct research into the intersections of security, privacy, and AI with Vitaly Shmatikov. My focuses are on red-teaming AI systems, understanding how privacy is operationalized in practice, and explaining cryptographic notions of privacy in accessible terms.

Brown University

Providence, RI ■ Sep 2016 – Dec 2020

- B.S. in Computer Science and B.A. in History, graduated *magna cum laude* in December 2020; **GPA – 3.93/4.00**
- Relevant coursework – **CS:** Computational Theory, Algorithms for the People, Systems Security, Cryptography; **History:** History of Privacy Policy, History of Economic Thought

PUBLICATIONS (reverse chronological order)

Academic publications:

- “Agent Meltdowns: The Road to Hell Is Paved with Helpful Agents,” RDJ*, **HT***, VS ([arXiv](#), NeurIPS 2026)
- “Deep Research Agents Can Be Poisoned via User-Generated Content,” TZ, **HT**, VS ([arXiv](#))
- “Reinforcing privacy reasoning in LLMs via normative simulacra from fiction,” MF, MZC, **HT**, HN (COLM 2026, [arXiv](#))
- “What Did Elon Change? A Comprehensive Analysis of Grokipedia,” **HT**, AM ([arXiv](#))
- “Breaking and Fixing Defenses Against Control-Flow Hijacking in Multi-Agent Systems,” RDJ, **HT**, VS (ICLR 2026, [arXiv](#))
- “MillStone: How Open-Minded Are LLMs?” **HT**, VS ([arXiv](#))
- “‘Having Confidence in My Confidence Intervals’: How Data Users Engage with Privacy-Protected Wikipedia Data,” **HT**, JS, PN, RC, GK, SK, ER (CHI 2026, [arXiv](#))
- “zk-Cookies: Persistent Anonymous Credentials for the Web,” AF*, **HT***, IM ([IACR](#))
- “Multi-Agent Systems Execute Arbitrary Malicious Code,” **HT**, RDJ, VS (COLM 2025, [arXiv](#))
- “SPINACH: SPARQL-Based Information Navigation for Challenging Real-World Questions,” SL, SS, **HT**, JX, IDZ, ML (EMNLP 2024, [arXiv](#))
- “Publishing Wikipedia usage data with strong privacy guarantees,” TA, DD, MH, IJ, CL, AM, TM, TM, DP, DSM, **HT** (IPDP 2023, [arXiv](#))
- “Publishing Wikipedia Project Usage Data with Strong Privacy Protections and without Tracking,” **HT** ([PEPR](#), 2022)
- “A New Model for Weaving Responsible Computing Into Courses Across the CS Curriculum,” LC, HP, **HT**, KF ([SIGCSE](#), 2021)
- “Detecting Large Vessel Occlusion at Multiphase CT Angiography by Using a Deep Convolutional Neural Network,” MS, JV, MD, YHK, SS, **HT**, AW, HLW, AY, MJ, JB, CE, UC, GB, RM ([Radiology](#), 2020)

Selected popular press publications:

- “AI Alignment Is Censorship,” editor ([Reboot](#), 2025)
- “Threat Model,” writer ([Reboot](#), 2023)
- “Port Paradox,” editor ([Kernel](#), 2023)
- “Towards Ineffective Altruism,” writer ([Reboot](#), 2022)
- “Where Algorithmic Transparency Meets Community”, interviewee ([Reboot](#), 2021)
- “For nearly a decade, two RI opioid manufacturers made billions of opioid doses,” writer ([The Public’s Radio](#), 2021)
- “The opioid influencers,” writer ([Brown Opioid Journalism Project](#), 2020)

TECHNICAL EXPERIENCE

Student Researcher, Google

NYC ■ Feb – Aug 2026

- Conducted research on LLM post-training procedures that enhance agent safety and decrease the propensity of models to engage in meltdown behaviors. Output forthcoming.

Data Consultant, Wikimedia Enterprise

Remote ■ Oct 2024 – Present

- Building a high-quality enterprise pageview trend tool using open-source and privatized Wikimedia pageview data and consulting on privacy policy decisions. Consulting on various privacy-focused special projects pertaining to data.

Senior Privacy Engineer, Wikimedia Foundation

Remote ■ Apr 2021 – Aug 2024

- Built organizational capacity to use differential privacy in production data releases at Wikipedia (the 5th-largest website in the world, with tens of billions of monthly visits). Successfully published five separate large-scale data releases, increasing the amount of publicly-available [page view](#) and [editor](#) data by orders of magnitude, with more profitable data releases in progress. DP work has also been presented at leading [privacy conferences](#).

- Created, tested, wrote, and validated multiple privacy- and security-related policies and workflows pertaining to data management and governance, encompassing collection, [retention](#), and [publication](#) of data, including [sensitive geolocated data](#).
- Other responsibilities included regular privacy product reviews and privacy mitigations, as well as contributing to work in the realms of [responsible AI](#), open-source dataset creation, anti-surveillance, and [anti-censorship](#) tech.

Data Intern, New Hampshire Democratic Party

Remote ■ Jun – Aug 2020

- Directly shaped decision-making processes among party leadership by conceptualizing, deploying, monitoring, and assessing ML algorithms to order and prioritize tens of thousands of voter and volunteer calls among NH voters. Used BigQuery and Data Studio to build reports analyzing the effectiveness of the NHDP's campaign.

Algorithmic Recommendations Intern, The New York Times

New York, NY ■ Jun – Aug 2019

- Created a 5% increase in recipe click-through rate by independently conceptualizing, building, and validating new algorithmic approaches to recipe recommendation for 600,000+ NYT Cooking users. Worked with algorithmic and experimental design, production-scale data systems, and A/B testing.

IT and Data Security Consultant, Political Campaigns

Across the US ■ Aug 2018 – Dec 2019

- Boosted online audience size by 80% and protected digital assets as an aide on [Dan Feehan's congressional campaign](#) in the Minnesota 1st Congressional District, one of the most competitive races in the country. My focus was on organizational cybersecurity, data analytics and targeting, field organizing, and [social media content creation](#).
- Collaboratively helped develop and enact the IT and data security plan for Sunrise Movement. Admined Sunrise's GSuite, led a threat analysis audit, and helped oversee digital security upgrades for 100+ staff and volunteers.

Researcher, ELVO Machine Learning Research

Providence, RI ■ Jun 2018 – Jun 2019

- Founding member of a high-impact medical AI research team studying automated diagnosis of large strokes using machine learning in order to decrease time to surgical intervention and improve patient outcomes. Contributed to a paper [published in Radiology](#), the leading journal in the field.

WRITING and LEADERSHIP

Technology Fellow, FASPE

Germany and Poland ■ May 2022 – Jun 2022

- Part of the inaugural class of design and technology fellows participating in FASPE (Fellowships at Auschwitz for the Study of Professional Ethics). In Germany and Poland for several weeks, I studied historical examples of Nazi cooptation of technologists, architects, and designers and debated the present and future of tech ethics.

Contributing Editor, *Reboot*

Remote ■ Nov 2021 – Present

- Writer and editor for a nonprofit magazine on critical techno-optimism. Pieces include: [a deep dive into Wikipedia's algorithmic transparency efforts](#), a [critique of effective altruism and longtermism](#), a [reported piece on the politics of the energy transition in southern Louisiana](#), an [essay about the politics and psychology of privacy extremists](#).

Head Teaching Assistant for CS for Social Change

Providence, RI ■ Dec 2019 – Dec 2020

- Hired staff, developed a CS ethics curriculum, and taught a [high-level course](#). Facilitated experiential learning partnerships with nonprofits and research groups. Helped produce over 30 pieces of independent investigative data reporting on the opioid crisis in Rhode Island, with a [published piece on a local opioid factory](#) in *The Public's Radio*.

Ethics Teaching Assistant Program

Providence, RI ■ Jun 2019 – Jun 2020

- Co-wrote and designed an ethically-engaged curriculum for a 370-person deep learning class for [Brown's Responsible CS program](#). Conducted research on program effectiveness, resulting in policy recommendations for the 2020-21 school year and a paper published in the [SIGCSE Technical Symposium 2021](#).

Friday Night Jews

Providence, RI ■ Jan 2017 – Jun 2020

- Helped lead a grassroots Brown University Jewish group that facilitated conversation across the political spectrum, and also organized a national network of similar independent groups. Led a small team of students to self-fund and put on a conference for over 120 Jewish students from across the country in September 2019.

RELEVANT SKILLS and INTERESTS

- **Technical proficiencies:** differential privacy, cryptography, machine learning, cybersecurity
- **Language proficiencies:** Mandarin Chinese and Spanish
- **Certification:** IAPP Certified Information Privacy Manager (CIPM)
- **Other Interests:** Music, specifically jazz and classical piano; racial and economic social justice; technology ethics; long-distance running