Mathias Lécuyer

Post-Doctoral Researcher Microsoft Research NY

Areas

Software systems, privacy, security, machine learning, statistics, causal inference.

Education

2013–2019 **PhD in Computer Science**, *Columbia University*, New York, *GPA 4.0*.
- Thesis: *Security, Privacy, and Transparency Guarantees for Machine Learning Systems*.
- Advisors: Roxana Geambasu, Augustin Chaintreau, and Daniel Hsu.

2011–2012 **Master of Science in Computer Science**, *Columbia University*, New York, *GPA: 3.90*.

2008–2011 **Ingénieur Diplômé**, *Ecole Polytechnique*, Paris, France, *GPA: 3.90*.
Major in Computer Science, multidisciplinary training in Physics, Enonomics, and Biology.

2006–2008 **Classes Préparatoires in Science**, *Lycée Louis-le-Grand*, Paris, France, *GPA: 4.0*.
Intensive preparation in Mathematics and Physics for competitive admission into top French universities.

Work Experience

2019-Current **Post-Doctoral Researcher**, *Microsoft Research*, New York.

2013-2019 **Research Assistant**, *Columbia University*, New York.

Summer 2017 **Research Intern**, *Microsoft Research*, New York.
- Applied reinforcement learning to systems, with a focus on off-policy evaluation [3].

2012–2014 **Teaching Assistant**, *Columbia University*, New York.
- Distributed Systems. Instructor: Roxana Geambasu.
- Data Journalism. Instructor: Mark Hansen.
- Computer Networking. Instructor: Augustin Chaintreau.

Summer 2014 **Freelance**, *Floatingapps*, New York.
- Developed a Rails application for rental data visualization, using Cassandra, Elasticsearch, d3, and React.
- Built an iOS application for email visualization as timelines in Objective-C.

Summer 2013 **Freelance**, *Milky*, Paris, France.
- Built an extended iOS application for *Le Grand Journal*, a show of major French TV Channel Canal+.

2009–2012 **Cofounder**, *Pionid*, Paris, France.
Developed the main product, an iOS group messaging application.

Apr–Jun 2011 **Junior Consultant**, *Atos*, Paris, France.
Worked on archiving as a service.

Research

My research addresses the new system challenges and opportunities introduced by the data and artificial intelligence revolutions. While data-driven systems can yield social and economic benefits, they also open new security and privacy threats, and their opacity and unpredictability can undermine users' trust. To address these challenges, I design, implement, and evaluate rigorous, theory-backed systems that are both practical and provide provable guarantees of security, privacy, and statistical soundness. To provide these guarantees, my system designs leverage theory from statistics, machine learning, causal inference, and differential privacy.

Current projects **PixelDP: Certified Defense Against Adversarial Examples in Machine Learning**.
Adversarial examples that fool prediction models are a new class of attacks introduced by machine learning deployments. I developed PixelDP, the first certified defense that both offers a guaranteed level of robustness against these attacks and scales to large models and datasets, such as Google's Inception on the ImageNet dataset [1]. PixelDP's design leverages differential privacy, a theory from the privacy domain.

+1 (917) 355 4318 • mathias@cs.columbia.edu • http://mathias.lecuyer.me

Sage: Minimizing Data Exposure in Machine Learning Applications.
Machine learning ecosystems collect vast amounts of personal information and make it widely accessible within a company. I am developing new data protection abstractions to enable these intended use cases while minimizing data exposure to internal or external leaks [2, 4]. These abstractions rely on differential privacy, statistical tests, and training set reduction techniques form machine learning.

Sunlight: Increasing the Transparency of Machine Learning Systems.
Today's web services are opaque black boxes that leverage users' data for targeting and personalization with little end-user control or accountability. I built a new set of tools that increase users' visibility into how they are being targeted [8, 6]. The tools leverage causal inference to detect which pieces of input data cause targeting, machine learning heuristics to scale to a large number of input data, and statistical methods to rigorously assess the validity of results.

Sayer: Counterfactual Evaluation of Systems.
Machine learning and statistics also present new opportunities to solve traditional systems challenges in improved ways. For example, deploying good load balancing or resource allocation policies often requires an explicit evaluation, which can be heavy and delay reaction to changing environments. I developed Sayer, a generic tool for counterfactual evaluation of systems policies, enabling the statistically sound evaluation of these policies without actually running them [3]. Sayer builds on concepts from causal inference and reinforcement learning.

Past projects

Impact of the Sharing Economy.
In a controversy about Airbnb's impact on cities, three reports used seemingly contradictory aggregate statistics about the occupancy and revenue distribution to arrive at opposite conclusions. To inform the debate, I implemented a reliable way to estimate the occupancy and revenues of Airbnb's hosts [5]. The contradictory conclusions could all be explained by a variant of the inspection paradox.

Synapse: Heterogeneous-Database Replication.
Synapse [7] is an heterogeneous-database replication system that lets programmers of complex, multi-service applications share data across services running on distinct database engines, in real time, and with solid consistency semantics. We deployed Synapse at a New York City startup.

Dispatch: Secure and Private Reporting for Citizen Journalists.
Dispatch allows journalists and citizen reporters to publish their work pseudo-anonymously, using an identity-based encryption scheme. It also provides censorship-resiliant functionality with Bluetooth message passing when no other connection is available.

Publications

[1] Mathias Lécuyer, Vaggelis Atlidakis, Roxana Geambasu, Daniel Hsu, and Suman Jana. Certified robustness to adversarial examples with differential privacy. In *Proceedings of the IEEE Symposium on Security and Privacy (Oakland)*, 2019.

[2] Mathias Lécuyer, Riley Spahn, Kiran Vodrahalli, Roxana Geambasu, and Daniel Hsu. Privacy accounting and quality control in the sage differentially private ML platform. In *Proceedings of the ACM Symposium on Operating Systems Principles (SOSP)*, 2019.

[3] Mathias Lécuyer, Joshua Lockerman, Lamont Nelson, Siddhartha Sen, Amit Sharma, and Aleksandrs Slivkins. Harvesting randomness to optimize distributed systems. In *The Seventeenth ACM Workshop on Hot Topics in Networks (HotNets)*, 2017.

[4] Mathias Lécuyer, Riley B. Spahn, Roxana Geambasu, Tzu-Kuo Huang, and Siddhartha Sen. Pyramid: Enhancing selectivity in big data protection with count featurization. In *Proceedings of the IEEE Symposium on Security and Privacy (Oakland)*, 2017.

[5] Mathias Lécuyer, Max Tucker, and Augustin Chaintreau. Improving the transparency of the sharing economy. In *Proceedings of the International World Wide Web Conference (WWW)*, 2017.

[6] Mathias Lécuyer, Riley B. Spahn, Giannis Spiliopoulos, Augustin Chaintreau, Roxana Geambasu, and Daniel Hsu. Sunlight: Fine-grained targeting detection at scale with statistical confidence. In *Proceedings of the ACM Conference on Computer and Communications Security (CCS)*, 2015.

[7] Nicolas Viennot, Mathias Lécuyer, Jonathan Bell, Roxana Geambasu, and Jason Nieh. Synapse: New data integration abstractions for agile web application development. In *Proceedings of the European Conference on Computer Systems (EuroSys)*, 2015.

[8] Mathias Lécuyer, Guillaume Ducoffe, Francis Lan, Andrei Papancea, Theofilos Petsios, Riley Spahn, Augustin Chaintreau, and Roxana Geambasu. XRay: Increasing the web's transparency with differential correlation. In *Proceedings of the USENIX Security Symposium*, 2014.

Invited Talks

2020 *Towards a Practical Differentially Private Machine Learning Platform*, Northwest Data Science Seminar Series.

2019 *Privacy Accounting and Quality Control in the Sage Differentially Private ML Platform*, University of California Berkeley, Security Seminar.

2019 *Security, Privacy, and Transparency Guarantees for Machine Learning Systems*, University of British Columbia.

2019 *Security, Privacy, and Transparency Guarantees for Machine Learning Systems*, University of California Los Angeles.

2019 *Security, Privacy, and Transparency Guarantees for Machine Learning Systems*, Microsoft Research New York.

2018 *Certified Robustness to Adversarial Examples with Differential Privacy*, University of Washington, Security Seminar.

2018 *Certified Robustness to Adversarial Examples with Differential Privacy*, University of California Berkeley, Security Seminar.

2018 *Certified Robustness to Adversarial Examples with Differential Privacy*, Google Brain.

2018 *Harvesting Randomness for Counterfactual Evaluation of Systems*, Stanford, NetSeminar.

2018 *Certified Robustness to Adversarial Examples with Differential Privacy*, Stanford, Security Lunch.

2014 *XRay: Enhancing the Web's Transparency with Differential Correlation*, University of Washington and Microsoft Research Summer Institute.

2012 *Dispatch: weaving a safe web of news*, Columbia Journalism & Technology Breakfast.

Media Coverage

2016 *The data republic*, "To safeguard democracy, the use of data should be made as transparent as possible," The Economist.

2016 Priya Kumar, *When Was the Last Time You Read a Privacy Policy?*, Slate.com.

2015 Tom Simonite, *Probing the Dark Side of Google's Ad-Targeting System*, MIT Technology Review.

2014 Steve Lohr, *XRay: A New Tool for Tracking the Use of Personal Data on the Web*, The New York Times.

Service

2020/2021 Program Committee, USENIX Security 2021.

2020 External Reviewer, International Conference on Machine Learning 2020.

2019 Guest Lecture, graduate class "Decentralized Security: Theory and Systems" (CS 294-163).
Instructor: Raluca Ada PopaUniversity of California Berkeley

2020 External Reviewer, Security & Privacy 2020.

2020 Program Committee, Eurosys 2020.

2019/2020 Program Committee, USENIX Security 2020.

2019 Reviewer, Journal of Machine Learning Research.

2019 Program Committee, ACM Symposium on Cloud Computing (SoCC) 2019.

2018 Program Committee, Systems for Machine Learning Workshop at the Conference on Neural Information Processing Systems, http://learningsys.org/nips18/.

2017 Reviewer, AISys Workshop at the ACM Symposium on Operating Systems Principles, http://learningsys.org/sosp17/.

2016 Reviewer, ACM Transactions on Internet Technology.

2016 & 2014 Co-designed and delivered a programming workshop for the Annual Engineering Exploration, organized by the Society of Women Engineers for New York City high-school female students.